

An investigation of the roles of **GATA4** and its target genes in cardiac differentiation of iPS Cells

Nicola Graham

Cardiff University

PhD

2023



Summary

Heart development and homeostasis are controlled by the collective action of numerous transcription factors (TFs), that come together to form complex and dynamic gene regulatory networks (GRN) that define and maintain cellular identity. Disruption of the cardiac GRN is associated with the development of congenital heart defects (CHDs), and cardiovascular disease (CVD) in later life. Yet, our knowledge of the TFs that contribute to the GRN and their target genes is incomplete. The TF GATA4 has been established as a core member of the cardiac GRN with roles throughout heart development and the adult heart. Therefore, the genes regulated by GATA4 are likely involved in these processes too. Using *Xenopus laevis* embryos to model cardiac development two genes were identified and validated: *tbx2* and *prdm1*, as Gata4 targets. However, it was not known whether this regulatory relationship was conserved in human cardiac development and whether these genes themselves are required for normal heart development.

The work presented in this thesis addresses this with the creation of *GATA4*, *TBX2*, and *PRDM1* mutant human iPSC lines using CRISPR-Cas9 gene editing. The results presented demonstrate that GATA4 is absolutely essential for the formation of cardiomyocytes (CMs) from iPSCs, which has not previously been examined in depth in a human model of cardiomyogenesis. It is also demonstrated that *TBX2* and *PRDM1* are differentially regulated in the absence of GATA4, indicating this regulatory relationship is likely conserved between *Xenopus* and humans. Furthermore, it is demonstrated that *TBX2* and *PRDM1* are modulators of the cardiac transcriptome. Separately the role of GATA6, a TF that is closely related to GATA4, is investigated. GATA6 has recently been shown to be essential for iPSC-CM differentiation. The results presented herein add to this by showing that removal of a GATA6 regulatory domain disrupts normal iPSC-CM differentiation. Together these investigations expand our knowledge of what nodes in the cardiac GRN are required for normal cardiac development, and therefore what elements may factor into the development of CHDs and CVDs.

Declaration

This thesis is being submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is it being submitted concurrently for any other degree or award (outside of any formal collaboration agreement between the University and a partner organisation).

I hereby give consent for my thesis, if accepted, to be available in the University's Open Access repository (or, where approved, to be available in the University's library and for inter-library loan), and for the title and summary to be made available to outside organisations, subject to the expiry of a University-approved bar on access if applicable.

This thesis is the result of my own independent work, except where otherwise stated, and the views expressed are my own. Other sources are acknowledged by explicit references. The thesis has not been edited by a third party beyond what is permitted by Cardiff University's Use of Third Party Editors by Research Degree Students Procedure.

Word Count = 63,817

Acknowledgements

First and foremost, thank you to the British Heart Foundation and its donators for the generous funding provided for this PhD studentship, and I hope that this project will make a meaningful contribution to the field. This process has been challenging and upon entering University as a little undergrad a PhD is not something I imagined myself doing. Therefore, I must also express my gratitude for the many people whose confidence and support gave me the idea I could do this and that have inspired me to persevere.

To my supervisor, Dr. Branko Latinkić thank you for entrusting me with this project. Your enthusiastic support, guidance, and encouragement have been invaluable especially when experiments have been particularly uncooperative, and during the writing up section of this thesis. Special thanks also goes to Dr. Pavel Kirilenko who has contributed a great deal to my practical and technical knowledge and given his time to produce some of the molecular reagents used in this thesis. Thank you to Dr. Ewan Fowler who conducted calcium imaging of the TBX2 lines generated during this work and provided the MatLab script used for their analysis. Additionally, thanks goes to Dr. Peter Watson who gave his time to help me image calcium dynamics in the GATA4 lines. I would also like to extend my gratitude to Dr. Neil Rodrigues and Dr. Leigh-Anne Thomas for their help learning flow cytometry, and Dr. Chris Denning for providing the Rebl Pat iPSC line used throughout.

To my fellow Cardiff BIOSI PhD students and friends thank you for the many trips to Brodie's, without this continuous supply of caffeine and company I would surely have given up many many times. I am also thankful to my friends outside of university who have been a friendly ear for all complaints throughout the years, and who provide great advice and fun distractions. I am also grateful for Igor Camara and family who have lent their support in tough times. To my family, thanks must first be given to my number one fan, Susan Graham. Thank you for your unconditional love and un-wavering support for everything that I do. Also, for making me, cheers! To my second biggest fan Andrew, you may have a far better metabolism than me and a stable job, but I have totally won at school. Of course, I must also express my appreciation for Murphy who has been a loyal writing companion and all around very good boy.

Finally, this thesis is dedicated in loving memory to Vincent Michael Graham who is missed more than he will ever know. Every day is filled with gratitude for the memories made with you, that guide me through this world and give me the faith needed to persevere.

List of non-standard abbreviations

ARVC	Arrhythmogenic right ventricular cardiomyopathy
AVC	Atrioventricular canal
AVE	Anterior visceral endoderm
AVSD	Atrioventricular septal defect
BSA	Bovine serum albumin
CHD	Congenital heart defects
CICR	Calcium induced calcium release
CM	Cardiomyocyte
CP	Cardiac progenitor
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
CVD	Cardiovascular disease
DBD	DNA binding domain
DCM	Dilated cardiomyopathy
DVE	Distal visceral endoderm
E	Mouse embryonic day
ExE	Extraembryonic ectoderm
FDHM	Fluorescence duration at half maximum
FHF	First heart field
GoF	Gain of function
GRN	Gene regulatory network
HCM	Hypertrophic cardiomyopathy
HF	Heart failure

HMT	Histone methyl transferase
ICM	Inner cell mass
iCM	Induced cardiomyocyte
IF	Immunofluorescence
iPSC	Induced pluripotent stem cell
LoF	Loss of function
MBT	Mid-blastula transition
mESC	Mouse embryonic stem cell
NMD	Nonsense mediated decay
NS	Nuclear speckles
OFT	Outflow tract
PE	Primitive endoderm
PS	Primitive streak
PTA	Patent truncus arteriosus
PTC	Premature termination codon
RNP	Ribonucleoprotein
RT-PCR	Reverse transcription polymerase chain reaction
sgRNA	Single guide RNA
SHF	Second heart field
SMC	Smooth muscle cell
SR	Sarcoplasmic reticulum
TBE	T-box element
TE	Trophectoderm
TF	Transcription factor

tracrRNA Trans-activating CRISPR RNA

UCN Unmet clinical need

VE Visceral endoderm

VSD Ventricular septal defect

WB Western blotting

List of Tables

Chapter 1 – Introduction

Table 1.1 A summary of the primary phenotypes caused by GATA factor knockout in mice.

Table 1.2 A summary of results from GATA4, 5, and 6 loss of function models

Chapter 2 – Materials and methods

Table 2.1 Forward and reverse primer sequences, covering a broad range of mycoplasma species.

Table 2.2 Cycling conditions for amplification of mycoplasma DNA.

Table 2.3 crRNA sequences used for CRISPR Cas9 gene editing

Table 2.4 Antibodies used for Flow Cytometry

Table 2.5 Secondary Antibodies used for flow cytometry for detection of unconjugated primaries

Table 2.6 Commonly used PCR cycling conditions for RT-PCR

Table 2.7 Primers used for RT-PCR

Table 2.8 Primers used for RT-PCR analysis of *PRDM1* expression, with isoform specific primers.

Table 2.9 Primers used for gDNA PCR screening of CRISPR-Cas9 mutants, and for Sanger sequencing

Table 2.10 Antibodies used for Immunofluorescence

Table 2.11 Cell stains used for Calcium Imaging and immunofluorescent staining of fixed cells

Chapter 3 - Differentiation of human iPS cells into cardiomyocytes requires cardiac transcription factor Gata4

Table 3.1 A summary of the results obtained from immunofluorescent staining of WT and *GATA4*^{-/-} 1 and *GATA4*^{-/-} 2 cells at day 32 of differentiation.

Chapter 4 - Tbx2 is a modulator of the CM differentiation program

Table 4.1 A summary of the gDNA loci flagged during analysis of the WT and *TBX2* mutant lines.

Table 4.2 Transcripts affected by the additional deletions present in the *TBX2*^{c.700_761} mutant.

Table 4.3 Differentiation efficiency is seemingly unaffected by mutation of *TBX2*, but cell condition is impacted

Table 4.4 Differentially regulated genes of interest identified at day 10 in the *TBX2* mutants

Table 4.5 A survey of calcium handling gene expression in day 10 *TBX2* mutant cells

Chapter 5 - Prdm1 is dispensable for the differentiation of iPSCs into cardiomyocytes

Table 5.1 Summary of karyotyping results for PRDM1 mutants

Chapter 6 - Deletion of amino acids 311-337 from a predicted transcriptional activation domain of Gata6 leads to the production of hypertrophic cardiomyocytes

Table 6.1 GATA6 variants with known congenital heart defect associations located in the deletion region (residues 311-337)

Table 6.2 A summary of the genomic integrity of the WT and *GATA6*^{c.928-1010delInsA} lines

Table 6.3 Possible GATA6 target genes at day 3-4 of iPSC-CM differentiation

Table 6.4 Possible GATA6 target genes at day 8-10 of iPSC-CM differentiation

Supplemental Data

Table S1. A summary of the plasmids used to generate inducible cell lines.

Table S2. A list of primers used to assess plasmid integration into the AAVS1 locus.

Table S3. A selection of the maturation media formulations tested

List of figures

Chapter 1 – Introduction4

Figure 1.1 Cardiovascular deaths by subtype in the UK, 2021

Figure 1.2 Establishment of asymmetries in early mouse development

Figure 1.3 Migration of cardiac progenitors during development

Figure 1.4 Interconnectivity between the cardiac gene regulatory network

Figure 1.5 An overview of heart tube formation and the cardiac fields

Figure 1.6 A comparison of sequence similarity between GATA TFs 1-6

Figure 1.7 The divergence of the T-box family

Figure 1.8 A linear depiction of PRDM1 and its interactors

Figure 1.9 An overview of the milestones of iPSC to CM differentiation

Figure 1.10 A summary of possible maturation techniques

Figure 1.11 An overview of the CRISPR-Cas9 system from *Streptococcus pyogenes*

Figure 1.12 A summary of the questions this thesis aims to address

Chapter 2 – Materials and methods

Figure 2.1 A visual summary of the CDM3 and GiWi protocols published by Burrige *et al.* 2014, and Lian *et al.* 2012.

Figure 2.2 Cell measurements and classification system

Chapter 3 - Differentiation of human iPS cells into cardiomyocytes requires cardiac transcription factor Gata4

Figure 3.1. CRISPR targeting strategy and the identification of GATA4 mutants

Figure 3.2 *GATA4* mRNA is reduced in mutants and no detectable protein is produced

Figure 3.3 GATA4 mutant cells maintain pluripotency

Figure 3.4 GATA4 mutant iPSCs fail to beat

Figure 3.5 The expression of sarcomeric genes is decreased in WT and GATA4^{-/-} cells

Figure 3.6 GATA4 mutants form fewer cells positive for CM markers

Figure 3.7 Staining for sarcomeric proteins reveals dysfunctional organisation in GATA4^{-/-} mutants

Figure 3.8 Reduced calcium flux and expression of calcium handling genes in GATA4 mutant cells

Figure 3.9 Whilst staining for CM markers is decreased in GATA4^{-/-} cells, staining for mesenchymal markers is increased

Figure 3.10 Classification of the cell membrane morphologies of WT and GATA4^{-/-} TNNT2 positive cells

Figure 3.11 RT-PCR for markers of iPSC-CM milestones

Figure 3.12 The dynamic expression of *PRDM1* and *TBX2* during iPSC to CM differentiation is changed in the absence of GATA4

Chapter 4 - TBX2 is a modulator of the CM differentiation program

Figure 4.1 Derivation of TBX2 mutants using CRISPR-Cas9 gene editing

Figure 4.2 The edited TBX2 cell lines maintain pluripotency

Figure 4.3 Expression of TBX2 and closely related factor TBX3 in WT and mutant iPS cells during cardiac directed differentiation

Figure 4.4 Unexpected exon skipping in TBX2 mutants restores TBX2 ORF and maintains expression

Figure 4.5 A comparison of the DNA binding domains of WT and mutant TBX2 with close relative TBX3

Figure 4.6 Embryos injected with mutant *TBX2* resemble un-injected controls

Figure 4.7 TBX2 mutant iPS cells form CMs

Figure 4.8 TBX2 mutant iPS cells form CMs at a similar efficiency to WT but display hypertrophy

Figure 4.9 TBX2 mutant cells are more often disorganised

Figure 4.10 Calcium signalling is dysfunctional in the *Tbx2* mutants

Figure 4.11 A comparison of the genes found to be upregulated in *Tbx2*^{c.700_761del} cardiomyocytes

Figure 4.12 Analysis of calcium handling gene expression alterations in *Tbx2* mutant CMs

Figure 4.13 Transcriptional changes in the *Tbx2* mutants reveals an upregulation of cardiac development and disease relevant genes

Figure 4.14 *Tbx2* is a modulator of the cardiac GRN that is required for production of WT-like CMs

Chapter 5 - PRDM1 is dispensable for the differentiation of iPSCs into cardiomyocytes

Figure 5.1 A linear depiction of PRDM1 and its interactors

Figure 5.2 Generation of PRDM1 mutants using CRISPR-Cas9

Figure 5.3 PRDM1 expression in WT and mutants

Figure 5.4 The features of pluripotency are maintained in PRDM1 mutant iPSCs

Figure 5.5 CM formation is slightly accelerated in PRDM1 mutants

Figure 5.6 Differentiation efficiency is unaffected in PRDM1 mutant lines

Figure 5.7 PRDM1 cells form WT-like CMs

Figure 5.8 Analysis of transcriptional changes in PRDM1 mutants

Figure 5.9 PRDM1 represses alternative gene programs in early cardiomyogenesis

Chapter 6 - Deletion of amino acids 311-337 from transcriptional activation domain 2 of GATA6 leads to the production of hypertrophic cardiomyocytes

Figure 6.1 Creation of a GATA6 mutant iPSC line using CRISPR-Cas9 gene editing

Figure 6.2 The GATA6 mutant protein lacks residues 311-337

Figure 6.3 *GATA6*^{c.928_1010delInsA} mutant iPSCs maintain pluripotency

Figure 6.4 The GATA6 mutant has increased intrinsic transcriptional activation

Figure 6.5 Differentiation efficiency is unaffected in the mutants but TNNT2 expression per cell is increased

Figure 6.6 The cardiomyocytes formed from *GATA6*^{c.928_1010delInsA} iPS cells are hypertrophic and disorganised

Figure 6.8 Analysis of transcriptional changes in the *GATA6*^{c.928_1010delInsA} cells at day 10

Figure 6.8 Analysis of transcriptional changes in the *GATA6*^{c.928_1010delInsA} cells

Table 6.3 Possible GATA6 target genes at day 3-4 of iPSC-CM differentiation

Table 6.4 Possible GATA6 target genes at day 8-10 of iPSC-CM differentiation

Figure 6.9 The expression of markers of CM differentiation are largely unperturbed in the *Gata6*^{c.928_1010delInsA} line

Figure 6.10 Residues 311-337 of cardiac TF GATA6 are required for normal CM development.

Chapter 7. Overview of findings and future directions

Figure 7.1 A summary of the conclusions produced during this thesis

Supplemental Data

Figure S1. PCR based screening identifies A4, A6, and A7 as potentially double positive clones for pTRE-*PRDM1* integration

Figure S2. Glucose is necessary for cell survival in a high fatty acid environment.

Figure S3. Maturation media has mixed effects on CM phenotype

Figure S3. A comparison of the morphology between beating and non-beating HFOs

Figure S4. The addition of WT cells to GATA4 null iPS cell cultures does not improve their cardiomyogenic differentiation potential

Figure S5. IF staining for Gata6 shows it to be homogenously distributed throughout the nucleus of WT and *GATA6*^{c.928_1010delInsA} CMs

Figure S6. Most GATA4 null cells fail to form any sarcomeres

Figure S7. Quality control data for WT and TBX2 mutant iPSC-CM differentiation RNA samples

Figure S8. Quality control data for the RNA-seq data generated from WT and PRDM1^{-/-} mutant lines during iPSC-CM differentiation

Figure S9. Quality control data for WT and *GATA6*^{c.928_1010delInsA} mutant iPSC-CM differentiation RNA samples

Figure S10. There is no significant difference in the proportion of TNNT2 positive and MYBPC3 positive cells produced within each cell line

Table of Contents

Summary	I
Chapter 1 - Introduction	- 1 -
1.1 Thesis chapters overview	- 1 -
1.2 Cardiovascular disease in the UK	- 2 -
1.2.1 Insights into development from Congenital Heart Defects	- 4 -
1.3 An overview of cardiac development	- 5 -
1.3.1 Early embryonic development	- 5 -
1.3.2 Patterning of the early embryo through the establishment of a morphogen gradient	- 6 -
1.3.3 Heart tube formation and the heart fields	- 11 -
1.4 Genetic regulation of cardiac development	- 15 -
1.4.1 Interactivity and cooperativity between members of the cardiac GRN	- 15 -
1.4.2 GATA4 is a core member of the cardiac GRN	- 16 -
1.4.3 The origins of GATA4 and the GATA family of transcription factors	- 19 -
1.4.4 The roles of GATA factors 1 to 3 in haematopoiesis and beyond	- 20 -
1.4.5 GATA factors 4, 5, and 6 in cardiac development	- 24 -
1.4.6 GATA4 in human models of cardiac development	- 26 -
1.4.7 Expanding the Gene Regulatory Network	- 29 -
1.5 TBX2	- 29 -
1.5.1 The T-box family of transcription factors	- 29 -
1.5.4 T-box factors in cardiac development	- 30 -
1.5.2 TBX2 in AVC formation	- 31 -
1.5.3 Are there earlier roles for TBX2 in cardiac development?	- 31 -
1.6 PRDM1 and cardiac development	- 32 -
1.6.1 The PR/SET and zinc finger containing family of transcriptional regulators	- 32 -
1.6.2 Known roles for PRDM1 in cardiac development	- 33 -
1.7 Modelling cardiac development with human iPS cells	- 34 -
1.7.1 Resetting differentiation potential	- 35 -
1.7.2 Differentiating iPSCs into CMs	- 36 -
1.7.2 Multiple factors contribute to the immature phenotype of iPSC-CMs	- 37 -
1.7.3 CRISPR-Cas9 gene editing as a tool for studying cardiac development	- 39 -
1.8 Project Aims & Hypotheses	- 42 -
Chapter 2 - Materials and methods	- 44 -
2.1 Induced Pluripotent Stem Cell (iPSC) maintenance	- 44 -

2.1.1 iPSC culture	- 44 -
2.1.2 iPSC passaging.....	- 44 -
2.2 Gene editing with CRISPR-Cas9	- 44 -
2.2.1 Preparing cells for nucleofection	- 44 -
2.2.2 PCR based detection of mycoplasma	- 45 -
2.2.3 crRNA design.....	- 46 -
2.2.4 Formation of guide RNAs.....	48
2.2.5 Ribonucleoprotein complex (RNP) formation	48
2.2.6 Nucleofection.....	48
2.2.7 Selection of RNP positive cells via fluorescence activated cell sorting (FACS).....	48
2.2.8 PCR based screening of cells.....	49
2.3 Quality checking edited lines	49
2.3.1 Assessing pluripotency.....	49
2.3.2 Karyotyping	49
2.4 Cardiac Differentiation.....	50
2.4.1 Chemically defined Wnt Modulation (CDM3) Protocol.....	50
2.4.2 GiWi Differentiation Protocol	51
2.4.3 Metabolic Selection and Maintenance	51
2.4.5 Dissociation of young (day 7-10) CMs	51
2.4.6 Dissociation of older (>10 days) CMs.....	51
2.4.6 Heart forming organoid (HFO) differentiation	52
2.4.7 Dissociation of HFO's	52
2.5 Phenotypic analysis of iPSCs and CMs	53
2.5.1 Morphology and Beating	53
2.5.2 Flow Cytometry for Analysis of CM Markers.....	53
2.5.3 Flow Cytometry for Pluripotency Marker POU5F1.....	54
2.5.4 RNA Extraction and cDNA Synthesis.....	56
2.5.6 RT-PCR.....	56
2.5.7 Immunofluorescence (IF).....	60
2.5.8 Characterisation of CMs by IF	60
2.5.9 Whole mount immunofluorescence for HFO's.....	62
2.5.10 Calcium Imaging.....	62
2.5.11 Western Blotting.....	63
2.6 RNA-seq analysis	- 65 -
2.7 Statistical analysis	- 65 -

Chapter 3 – Cardiac transcription factor GATA4 is required for iPSC-CM differentiation .	- 66 -
3.1 Introduction	- 66 -
3.1.1 GATA4 in cardiac development	- 66 -
3.1.2 Incomplete redundancy between GATA4 and GATA6	- 67 -
3.1.3 Studying GATA factors in human iPSC-CM differentiation.....	- 68 -
3.1.4 Aims.....	- 69 -
3.2 Results.....	- 71 -
3.2.1 Creation of two trans-heterozygous GATA4 mutants using CRISPR-Cas9 gene editing	- 71 -
3.2.2 GATA4 protein is not expressed in mutants	- 72 -
3.2.3 GATA4 mutant cell lines maintain pluripotency	- 74 -
3.2.4 GATA4 mutants fail to form functional cardiomyocytes	- 75 -
3.2.5 GATA4 is needed for expression onset and the maintenance of a selection of sarcomeric genes	- 77 -
3.2.6 Positive Staining for CM Specific Proteins Reduced in GATA4 Mutants	- 78 -
3.2.7 The GATA4 ^{-/-} TNNT2 ⁺ cells lack sarcomeres	- 80 -
3.2.8 GATA4 mutant cells produce prolonged calcium transients	- 82 -
3.2.9 The number of mesenchymal cells is increased in the GATA4 ^{-/-} mutant.....	- 84 -
3.2.10 A sub-set of GATA4 ^{-/-} TNNT2 ⁺ cells display abnormal cell membrane morphology	- 86 -
3.2.11 No significant disruption of the expression of other cardiac GRN factors was noted.....	- 88 -
3.2.12 The expression of GATA4 target genes TBX2 and PRDM1 is altered in the GATA4 mutants.....	- 90 -
3.3 Discussion.....	- 92 -
3.3.1 The inability of GATA4 ^{-/-} cells to beat is underscored by low differentiation efficiency and defective sarcomere formation.....	- 92 -
3.3.2 Does the decrease in TNNT2 and MYBPC3 expression extend to other sarcomeric markers?	- 94 -
3.3.3 Sluggish calcium handling is likely a cause of contractile insufficiency and may be indicative of a fate change	- 95 -
3.3.4 Alternative cell fates in GATA4 ^{-/-} lines and the potential for aberrantly specified cell types	- 96 -
3.3.5 When does the process of cardiomyogenesis start to go wrong in the GATA4 ^{-/-} cells?.....	- 97 -
3.3.6 Conclusions	- 98 -
Chapter 4 –TBX2 is a modulator of the CM differentiation program	- 100 -
4.1 Introduction	- 100 -

4.1.1 TBX2 acts as a transcriptional repressor in AVC development.....	- 100 -
4.1.2 The TBX2 expression pattern during development suggests cardiac activity prior to its role in the AVC	- 101 -
4.1.3 Analysing the requirement for TBX2 during iPSC-CM development.....	- 102 -
4.2 Results.....	- 103 -
4.2.1 Editing of <i>TBX2</i> locus with CRISPR-Cas9	- 103 -
4.2.2 TBX2 mutant iPS cell lines retain pluripotency.....	- 104 -
4.2.3 No obvious detrimental changes to genome integrity in TBX2 edited lines ...	- 106 -
4.2.4 TBX2 is expressed at similar level in WT and mutant iPSC-CMs.....	- 107 -
4.2.5 Expression of TBX2 is rescued by exon skipping.....	- 108 -
4.2.6 DNA binding domain structure is disrupted in the TBX2 mutant cell lines	- 110 -
4.2.7 The TBX2 mutant proteins are unstable in <i>Xenopus laevis</i> embryos	- 112 -
4.2.8 TBX2 mutant cell lines produce functional cardiomyocytes	- 114 -
4.2.9 Differentiation efficiency is seemingly unaffected in the TBX2 mutants but cells appear hypertrophic	- 114 -
4.2.10 Fibril alignment and sarcomere formation is abnormal in TBX2 mutant cardiomyocytes.....	- 117 -
4.2.11 Calcium handling is disrupted in the TBX2 mutant CMs	- 119 -
4.2.12 RNA-sequencing of TBX2 mutant cells reveals early upregulation of genes associated with cardiac stress	- 122 -
4.2.13 Identification of miss regulated calcium handling genes in day 10 CMs.....	- 124 -
4.2.14 Gene ontology enrichment analysis reveals upregulation of cardiac development and disease genes in TBX2 mutants.....	- 126 -
4.3 Discussion.....	- 129 -
4.3.1 Production of normal cardiomyocytes requires WT expression of TBX2.....	- 129 -
4.3.2 How does CM heterogeneity affect the phenotype of TBX2 mutant cells?....	- 130 -
4.3.3 TBX2 mutant CMs display characteristics of various CVDs	- 131 -
4.3.4 Abnormal calcium handling: cause or effect?	- 132 -
4.3.5 Different levels of dysfunction between the mutants	- 134 -
4.3.6 Conclusions	- 135 -
Chapter 5 - PRDM1 is dispensable for the differentiation of iPSCs into cardiomyocytes.....	- 136 -
5.1 Introduction	- 136 -
5.1.1 Transcription factor PRDM1 as a recruiter of chromatin modifying enzymes-	136 -
5.1.2 PRDM1 in cardiac development	- 137 -
5.2 Results.....	- 140 -
5.2.1 Editing of the <i>PRDM1</i> locus using CRISPR-Cas9 gene editing.....	- 140 -

5.2.2 Analysis of PRDM1 expression in WT and mutant lines	142 -
5.2.3 PRDM1 mutant lines retain pluripotency	144 -
5.2.4 Formation of contracting CMs is slightly accelerated in PRDM1 mutants	147 -
5.2.5 Differentiation efficiency is unaffected in the PRDM1 mutants	149 -
5.2.6 PRDM1 mutants form wild type-like cardiomyocytes	150 -
5.2.7 RNA-Seq analysis of PRDM1 ^{-/-} cells reveals roles in early gene regulation ..	152 -
5.3 Chapter 5 - Discussion	155 -
5.3.1 In the absence of PRDM1 CM commitment is accelerated	155 -
5.3.2 Preliminary transcriptome analysis for PRDM1 ^{-/-} mutants reveals upregulation of non-cardiac gene programs	156 -
5.3.3 Differences in PRDM1 isoform expression do not seem to impact CM development	157 -
5.3.4 Conclusions	158 -
Chapter 6 – Deletion of amino acids 311-337 from a predicted transcriptional activation domain of GATA6 leads to the production of hypertrophic cardiomyocytes	159 -
6.1 Introduction	159 -
6.1.1 Roles for GATA6 in early development	159 -
6.1.2 Studying GATA6 in human iPSC models of development	160 -
6.1.3 Roles for GATA6 in later stages of cardiac development	161 -
6.1.4 Project Rationale	161 -
6.2 Results	162 -
6.2.1 Creation of a GATA6 mutant iPS cell line using CRISPR-Cas9 gene editing	162 -
6.2.2 The deletion affects a CHD associated region in transcriptional activation domain 2	163 -
6.2.3 Genomic integrity and pluripotency are maintained in the GATA6 mutant ...	165 -
6.2.3 The deletion of amino acids 311-337 may lead to increased GATA6 transcriptional activation	168 -
6.2.4 Differentiation efficiency and beating onset is unchanged in the GATA6 mutant	169 -
6.2.5 <i>GATA6</i> ^{c.928_1010delInsA} expression leads to the production of hypertrophic CMs .-	171 -
6.2.6 Identification of early transcriptional changes during cardiac differentiation of <i>GATA6</i> ^{c.928_1010delInsA} mutant cells	173 -
6.2.7 Analysis of gene expression changes in <i>GATA6</i> ^{c.928_1010delInsA} CMs	175 -
6.2.8 Identification of likely direct targets of GATA6 regulation	177 -
6.2.9 The expression of iPSC-CM differentiation milestone markers is largely unchanged in the GATA6 mutant	179 -
6.3 Discussion	180 -

6.3.1 Deletion of amino acids 311-337 of GATA6 increases its transcriptional activation ability.....	- 180 -
6.3.2 iPSC-CM differentiation efficiency is unchanged in the GATA6 mutant, but CM phenotype is affected	- 181 -
6.3.3 Exploration of early transcriptional changes in GATA6 mutants.....	- 182 -
6.3.4 Indications of dysfunctional metabolism in the GATA6 mutant cells	- 183 -
6.3.5 Conclusions	- 184 -
Chapter 7 – Overview of Findings and Future Directions.....	- 185 -
7.1 The general limitations of iPSC-CM models.....	- 185 -
7.1.1 Inter- and intra-line variability in iPSC-CM differentiation efficiency	- 185 -
7.1.2 How does iPSC-CM maturity affect the phenotypes observed?	- 185 -
7.1.3 Technical and time-frame limitations.....	- 186 -
7.2 GATA4 is essential for human iPSC-CM differentiation.....	- 186 -
7.2.1 Reproducibility of the phenotype in different iPSCs	- 187 -
7.2.2 When does cardiomyogenesis fail in the GATA4 null lines & can it be rescued?	- 187 -
7.2.3 Is a failure in cardiac progenitor proliferation responsible for the phenotype observed in the GATA4 null lines?	- 189 -
7.2.4 If not CMs, what do GATA4 null cells become?	- 189 -
7.3 PRDM1 and TBX2 are regulated by GATA4 in iPSC-CMs.....	- 190 -
7.4 TBX2 is required for the formation of normal CMs, likely by fine tuning the cardiac differentiation program	- 190 -
7.4.1 – The creation of novel TBX2 lines requires modification of the CRISPR targeting strategy used.....	- 191 -
7.4.2 What are the drivers of pathology in the TBX2 lines?	- 191 -
7.4.3 How is proliferation affected in the TBX2 lines?	- 192 -
7.5 PRDM1 is not essential for iPSC-CM differentiation.....	- 192 -
7.5.1 How does knock-out of PRDM1 affect inter-line differentiation efficiency variability?.....	- 192 -
7.5.2 Is repression of PRDM1 required for differentiation?	- 193 -
7.6 Increased GATA6 transcriptional activity results in the production of hypertrophic iPSC- CMs	- 193 -
7.6.1 How does deletion of residues 311-337 affect GATA6 protein stability and localisation?	- 194 -
7.6.2 Sensitivity to hypertrophy in the GATA6 mutant line	- 195 -
7.6.3 Examining the activity of GATA6 in other developmental contexts.....	- 195 -
7.7 Concluding remarks	- 196 -
References	- 198 -

Supplementary Data - 211 -

Chapter 1 - Introduction

1.1 Thesis chapters overview

Heart development is an exquisitely regulated process whose understanding is fundamental to the interpretation and treatment of congenital heart defects (CHDs) and cardiovascular diseases (CVDs) which have a profound effect on the quality of life of many people as a leading cause of death worldwide (Bhatnagar et al. 2016; Ma et al. 2020; Ahmad and Anderson 2021). Cardiac development is driven forward by the interaction of multiple TFs to form a cardiac gene regulatory network (GRN), and disruption of this network is associated with the development of CHDs and CVDs (Tan et al. 2002; Fahed et al. 2014). Previous research has highlighted cardiogenic TF GATA4 as a key mediator and nexus of the cardiac GRN. This has been established in various model organisms, but substantial differences between species in the requirement for GATA4 and closely related factors GATA5 and 6 has been noted. How applicable the role of GATA4 in these models is to human cardiac development has yet to be fully explored, however with the advent of iPSCs, and iPSC-CM differentiation protocols the process of cardiomyogenesis can now be studied in a human genetic background *in vitro*.

The work presented in **CHAPTER 3** aims to help clarify GATA4's role and position in human cardiac development by examining the effect of a loss of GATA4 expression during iPSC to CM differentiation, with a mutation created using CRISPR-Cas9 genome engineering, a method that will be used throughout. Another aim of this work is to expand our knowledge of the cardiac GRN by investigating genes that have been identified and validated as *Gata4* targets in *Xenopus laevis* (Latinkić *et al.* data not published) but that have not received adequate research attention in the field of cardiac development thus far. These genes are *TBX2* and *PRDM1*. **CHAPTER 4** will focus on *TBX2*, a cardiac TF with known roles in promoting formation of the atrioventricular canal (AVC) and outflow tract (OFT) through the repression of chamber myocardium genes (Christoffels et al. 2004; Harrelson et al. 2004). The results presented in Chapter 4 allude to an earlier role for *TBX2* during iPSC-CM development, where disruption of *TBX2* results in the formation of abnormal CMs. **CHAPTER 5** looks at the role of *PRDM1*, a TF which is involved in the development of a diverse range of other cell types and is often characterised as a transcriptional repressor (Bikoff et al. 2009). Herein, *PRDM1* is shown to play a modest modulatory role in iPSC-CM

differentiation, where it may be required for efficient repression of alternative gene programs early in this process. Finally, **CHAPTER 6** investigates GATA6, a closely related family member of GATA4. Expression of GATA6 is essential for iPSC-CM differentiation, and no CMs are formed in its absence (Sharma *et al.* 2020). The GATA6 iPSC line presented here carries a deletion in a predicted regulatory domain of the protein. This deletion seems to increase its transcriptional activity and this change in the protein sequence is found to be incompatible with the formation of normal CMs. Thus, this chapter offers new insights into the regions of GATA6 required for normal function of the protein and consequently sheds light on the roles of GATA6 during iPSC-CM differentiation.

1.2 Cardiovascular disease in the UK

The heart pumps blood through the circulatory system allowing it to perform the myriad of functions that are dependent on robust circulation. The contractile cells of the heart, cardiomyocytes (CMs), are formed early in embryogenesis and begin to beat at approximately day 22 of human development continuing throughout life (Später *et al.* 2014). Postnatally there is limited production of new CMs; the annual turnover rate is estimated to be 1-1.9% in those under 20 years of age, and this rate decreases as we age and by 75 years the rate is ~0.45% (Bergmann *et al.* 2009). This low rate of CM production means much of the heart is composed of CMs formed in the embryo and in the first few years of life. The average life expectancy in the UK has gradually increased over the last few decades. Latest figures for life expectancy in the UK are 79 years for males, and 82.9 years for females (Office for National Statistics 2020). This means even at a resting rate of 80 bpm the average human heart will beat over 3.4 billion times in a lifetime. Therefore, it is unsurprising that with these increasing gains in life expectancy, we have seen and will likely continue to see increases in the number of people developing and living with cardiovascular disease (CVD), as the amount of continued stress the cardiovascular system is exposed to over a lifetime increases.

Cardiovascular disease is a broad term encompassing all diseases affecting the heart and vascular system (figure 1.1), collectively CVDs are responsible for 25-27% (~170,000) of deaths per year in the UK. Of these 25-28% will be in those under the age of 75 (Bhatnagar *et al.* 2016; The British Heart Foundation 2023). Clinical interventions for CVD have improved the prognosis for many, however at this time the 5-year survival rate for persons

diagnosed with heart failure (HF) is just 26.2% (Taylor et al. 2019), demonstrating the large unmet clinical need for treatments, and cardiovascular research.

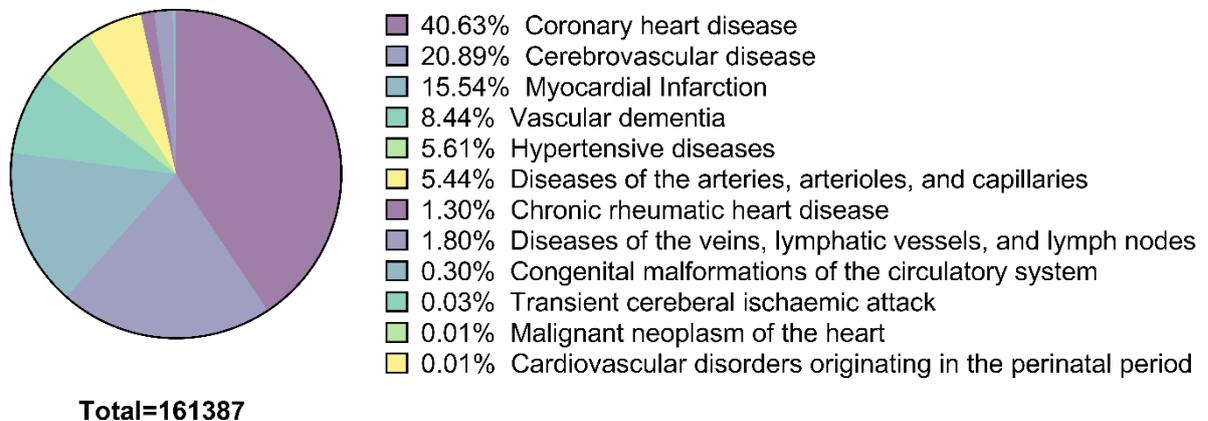


Figure 1.1 Cardiovascular deaths by subtype in the UK, 2021

Within the UK 161,387 deaths where cardiovascular disease was the primary cause were reported in 2021. The proportion of these deaths belonging to each subtype of cardiovascular disease (CVD) are shown above. This data is taken from the BHF (2023) Analysis of official UK mortality data.

Heart failure is a chronic condition and can be defined as a deterioration in cardiac function such that the heart can no longer meet the physiological demands of the body (Groenewegen *et al.* 2020). HF can be triggered by many of the CVD subtypes referenced in figure 1.1 that compromise heart function. As an example, 20-30% of patients who survive a myocardial infarction will develop HF within 1 year of the event (Hung *et al.* 2013; Sulo *et al.* 2016; Qureshi *et al.* 2018). During a myocardial infarction the oxygen supply to the heart is restricted, which can lead to a loss of CMs in the ischaemic area via apoptosis, autophagy, or necrosis. This is followed by a period of inflammation and immune cell infiltration to clear the damaged tissue and scar formation as a form of repair (Nakamura and Sadoshima 2018; Zimmer *et al.* 2019). Increased strain on the remaining CMs induces compensatory hypertrophic CM remodelling. In a pathological context this remodelling is characterised by changes in sarcomere structure, metabolism, mitochondrial function, an increase in reactive oxygen species, and the reactivation of foetal cardiac genes (Nakamura and Sadoshima

2018). These changes can subsequently lead to further CM loss and a decrease in function progressing to HF. Current treatments for HF aim to alleviate the dysfunction and slow the progression of the disease but cannot replace lost myocardium. Hence, significant attention has been focused on advancing myocardial regenerative therapies. These approaches include efforts to reprogram non-CM cardiac cells into CMs, stimulate innate CM regenerative capacity, and the delivery of CMs derived exogenously (Eschenhagen *et al.* 2017; Tzahor and Poss 2017). The future success of these approaches requires further understanding of what is needed for normal cardiac development.

1.2.1 Insights into development from Congenital Heart Defects

Congenital heart defects (CHDs) are a consequence of abnormal cardiac development *in utero* and the term CHD most commonly refers to morphological defects that overtly affect the structure of the heart that are apparent at birth. The understanding of CHDs represent a significant unmet clinical need in its own right with CHDs affecting 0.8-0.9% of live births (Van Der Linde *et al.* 2011; Triedman and Newburger 2016; Liu *et al.* 2019). Additionally, many of the genes and pathways that are dysregulated in CHDs are also dysregulated in CVDs including HF where reactivation of foetal cardiac genes occurs (Tan *et al.* 2002; Fahed *et al.* 2014), therefore by examining the genes that contribute to CHD we can gain an understanding of what is needed for normal development which may also aide our interpretation of these common diseases of later life.

CHDs are thought to have a strong genetic component. Evidenced by a recurrence rate in the families of CHD sufferers of around 11% (Ellesøe *et al.* 2016), although, it is worth noting that this varies between studies and the type of CHD being investigated (Fesslova *et al.* 2011; Ellesøe *et al.* 2016; Yokouchi-Konishi *et al.* 2019). In a study by Blue *et al.* using a panel of 57 known CHD risk genes, a genetic cause could be established for 31% of patients, all with a family history of CHD. It was noted that regardless of whether an established disease-causing variant was found that the patients generally had a higher number of rare variants of unknown significance in the 57 known CHD risk genes analysed (Blue *et al.* 2014). This panel included genes relevant to cardiac development such as *GATA4* and *GATA6*, that are also regulators of cardiac hypertrophy and HF (Pikkarainen *et al.* 2003; Oka *et al.* 2006; Van Berlo *et al.* 2010). Collectively, these studies demonstrate the overlap between the gene networks that regulate cardiac development and disease. Furthermore, the results

published by Blue *et al.* also underscore the value of investigating further the roles of genes known to be important for cardiac development, so that the effects of potential disease causing variants can be understood.

Another source of CHDs is *de novo* mutations, which are non-inherited spontaneously occurring mutations (Rahbari *et al.* 2016; Mohiuddin *et al.* 2022), these account for an additional 8-10% of CHD cases (Zaidi *et al.* 2013; Jin *et al.* 2017). Whilst again supporting a genetic basis for CHDs these figures demonstrate that there remains a gap in what can currently be explained by genetics for CHD risk. This is likely due to these diseases being polygenic, whereby a combination of variants in multiple genes may modify disease penetrance and severity, and that environmental factors can also impact cardiac development and health (Cerrone *et al.* 2019; Zhang *et al.* 2021). For these cases we need to expand our knowledge of the cardiac gene program and its interacting factors. This approach of widening the net is displayed in a paper by Sevim-Bayrak *et al.* where they were able to identify, using exome sequencing, mutations in 23 novel genes that share a biological function with known CHD causing genes and thus these genes may alter these pathways modifying disease penetrance (Sevim Bayrak *et al.* 2020). Further research has found that mutations in genes involved in more general aspects of cellular development, such as histone modifying enzymes and splicing factors, also have an association with CHD (Zaidi *et al.* 2013; Homsy *et al.* 2015). These studies demonstrate that cardiac development is influenced by many genetic factors and highlights the need for further research to understand this.

1.3 An overview of cardiac development

1.3.1 Early embryonic development

The core genes and signalling pathways that instruct cardiac development are highly conserved between humans, mice, other vertebrates, and *Drosophila* (Olson 2006). Naturally, there are variations between these organisms in early development but the sequence of events that take place is similar (Sissman 1970; Kojima *et al.* 2014; Krishnan *et al.* 2014; Lansford and Rugonyi 2020; Zhai *et al.* 2022). Therefore, although the focus of this thesis is human heart development this overview will primarily draw on studies of early murine development due to the wealth of information available from this model organism, complimented with studies in alternative models where applicable.

Embryonic development is initiated at fertilisation with the formation of a totipotent zygote that will give rise to the embryo and extra embryonic tissues. Following several rounds of cleavage, the zygote forms a ball of cells referred to as a blastula that will compact to form a morula. Compaction stimulates the partitioning of the cells into two lineages; the internal cells of the morula become the inner cell mass (ICM) which will form the epiblast/embryo proper and primitive endoderm (PE), whilst the external cells become trophectoderm (TE) that gives rise to the ectoplacental cone and extra embryonic ectoderm (ExE) (Takaoka and Hamada 2012). Around ~E3.5, cavitation creates a fluid filled cavity called the blastocoel, introducing an asymmetry with one side of the ICM juxtaposed with the TE, and the other facing into the blastocoel (Takaoka and Hamada 2012) (figure 1a). The ICM cells proliferate forming a mass of transiently pluripotent cells, identifiable by their expression of pluripotency factors NANOG, POU5F1, and SOX2 (Osorno and Chambers 2011; Takaoka and Hamada 2012). As the cells proceed through differentiation and exit pluripotency these factors will be down regulated.

Some of the first cells to do this within the ICM are PE cells, which are identifiable by their expression of GATA6 (Chazaud et al. 2006). This is the earliest essential role for GATA6 in development, demonstrated by a failure to form PE in GATA6 null mice, resulting in lethality at ~E4.5 (Morrissey *et al.* 1998; Koutsourakis *et al.* 1999; Zhao *et al.* 2005). The PE cells are initially specified in a 'salt and pepper' pattern before migrating to the border between the epiblast and the blastocoel (figure 1a) (Chazaud et al. 2006; Takaoka and Hamada 2012). Following this the ectoplacental cone derived from the TE mediates implantation and the PE comes to surround the epiblast and ExE in two layers, an outer parietal layer and an inner visceral endoderm (VE) layer that maintains contact with the ICM which will henceforth be referred to as the epiblast. GATA4 is partially responsible for the formation of the VE, and in its absence VE formation is defective, which is lethal by ~E7.5-E10.5 in mice (Molkentin et al. 1997). The establishment of the VE is an integral step in patterning of the early embryo and is prerequisite to successful cardiac development. The importance of the VE will be made clear below.

1.3.2 Patterning of the early embryo through the establishment of a morphogen gradient

Initially the inductive signalling factor NODAL is expressed throughout the epiblast (Brennan et al. 2001; Papanayotou et al. 2014). As the epiblast grows and elongates

the proximal distal axis is accentuated (figure 1.2b), forming an egg cylinder in mice or a flattened disc in other amniotes including humans (Krishnan *et al.* 2014; Lansford and Rugonyi 2020). This elongation firmly establishes embryonic asymmetry. Reciprocal signalling circuits between the ExE and the proximal epiblast create an area of high Nodal and Wnt signalling activity, whilst in the distal end of the epiblast these signals are repressed. High nodal signalling is maintained in the proximal epiblast by a combination of factors. Firstly, the expression of proprotein convertases *FURIN* and *PCSK6* (previously known as *PACE4*) becomes restricted to the ExE by E5.5 (Mesnard *et al.* 2006). These enzymes cleave *NODAL* forming the mature signalling peptide. Secondly, *NODAL* signals from the proximal epiblast induce expression of *BMP4* (Ben-Haim *et al.* 2006). *BMP4* acts to induce *WNT3* in the proximal epiblast, which subsequently induces *NODAL* expression (Ben-Haim *et al.* 2006), creating a positive induction loop.

NODAL can induce *BMP4*, *FURIN*, and *PCSK6* even in its proprotein form (Ben-Haim *et al.* 2006), however mice carrying cleavage resistant pro-*NODAL* or KO of *FURIN*^{-/-} and *PACE4*^{-/-} developmentally arrest during gastrulation (Beck *et al.* 2002; Ben-Haim *et al.* 2006). This shows that although pro-*NODAL* can instigate the signalling loop, mature *NODAL* is required to attain robust signalling capable of inducing normal gastrulation. Lastly, Wnt signalling may also stabilise *NODAL* expression through inhibition of *GSK3B* which would otherwise mark *SMAD3*, a downstream effector of Nodal signalling, for degradation (Guo *et al.* 2008). In the distal epiblast *NODAL* activates expression of Wnt and Nodal antagonists such as, *CER1*, and *LEFTY*, in the adjacent VE. *CER1* and *LEFTY* subsequently diffuse back into the epiblast inhibiting the Wnt and Nodal signalling pathways. Out of reach of the positive reinforcing signals emanating from the ExE (Mesnard *et al.* 2006) this region becomes an area of low Wnt and Nodal signalling and is referred to as the distal visceral endoderm (DVE). This combination of factors establishes a proximal to distal morphogen gradient, imparting positional information on the cells along this axis which will impact their fate.

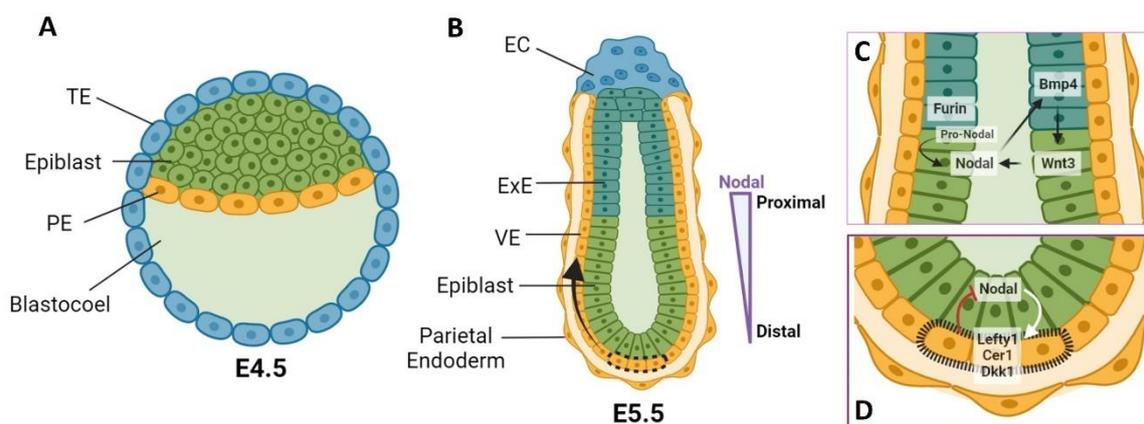


Figure 1.2 Establishment of asymmetries in early mouse development

(A) A blastocyst stage embryo in which the inner cell mass has separated into epiblast and primitive endoderm (PE) cells. With the PE forming a border between the epiblast and the blastocoel. The trophoctoderm (TE) surrounds the blastocyst and later forms the ectoplacental cone (EC), and the extra embryonic ectoderm (ExE). (B) By E5.5 the embryo has elongated establishing the proximal distal axis and the PE has formed two layers; parietal endoderm, and visceral endoderm (VE). Migration of distal VE cells (indicated by the dotted outline) to the future anterior side of the embryo, beginning at E5.5, establishes the anterior side of the embryo. The migration route of the cells to the future anterior side is indicated by an arrow. (C) NODAL can induce BMP4 in the ExE, which then induces Wnt signalling in the proximal epiblast that induces NODAL signals in this region of the epiblast. The expression of FURIN from the ExE adds to this by cleaving pro-nodal to its active form. (D) In the distal epiblast nodal signalling from the epiblast induces expression of LEFTY, CER1, and DKK1 in the distal VE, which due to its distance from the ExE is not in receipt of Nodal or Wnt reinforcing signals. Therefore, leading to the establishment of a proximal distal gradient of NODAL and WNT expression. Figure 1.2 was created using BioRender and was adapted from (Arnold and Robertson 2009).

At around E5.5-6 DVE cells expressing HHEX tagged with GFP can be visualised migrating to the future anterior side of the embryo, where they now form the anterior visceral endoderm (AVE) (Rodriguez et al. 2005; Hoshino et al. 2015). The expression of Wnt and Nodal inhibitors from the AVE restricts active Wnt and Nodal signalling to the posterior side of the embryo. This adds another dimension to the morphogen gradient previously established, with the highest levels of Wnt and Nodal activity now in the proximal posterior epiblast, and the lowest levels in the distal anterior epiblast. Activation of the Nodal and Wnt signalling pathways is necessary for gastrulation and mesodermal specification. This is demonstrated by the failure to form a PS in Nodal locus deletion mutants and by the failure to form mesoderm in WNT3^{-/-} KO mice (Conlon et al. 1994; Haegel et al. 1995). Furthermore, inhibitory signals from the AVE are required to restrict PS formation to the posterior of the embryo, demonstrated by ectopic formation of >1 PS when LEFTY1 and CER1 are knocked out in mice (Perea-Gomez et al. 2002). In accordance with this, the PS forms shortly after AVE formation at around E6.25 in the proximal posterior epiblast where active Nodal and Wnt signals are at their highest.

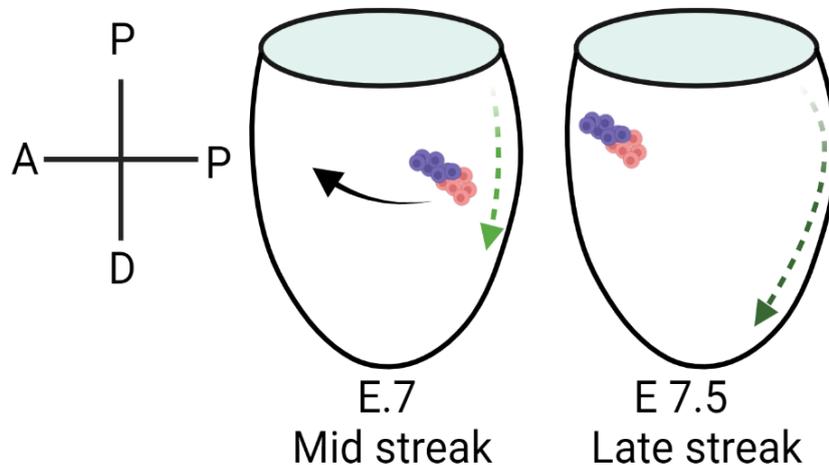


Figure 1.3 Migration of cardiac progenitors during development

The cells that will form the heart initially reside in the posterior epiblast. The cells ingress at the mid-streak stage and move as two bilateral patches to the anterior of the embryo, reach their destination by the late streak stage. The extending primitive streak at the posterior is depicted by green arrows. The purple cells are used to represent the first heart field progenitors that ingress first, shortly followed by the cells that will form the second heart field in pink. Figure created using BioRender and adapted from (Ivanovitch et al. 2021).

Prior to gastrulation the cells that give rise to the heart are located at the posterior of the embryo, they will ingress at the mid-streak stage as two bi-lateral patches to the anterior ventral side of the embryo (Ivanovitch et al. 2021) (figure 1.3). Due to their position along the proximal distal axis of the PS and their time of ingression they are exposed to moderately high levels of Wnt and Nodal signalling prior to gastrulation, and it is these graded signals received before migration that is thought to determine their fate (Ivanovitch et al. 2021). In support of this, lineage tracing studies have shown that allocation of cells from the mesoderm to a cardiac fate happens before or shortly after the start of gastrulation (Devine et al. 2014; Ivanovitch et al. 2021). The cells quickly progress from mesoderm to cardiac mesoderm, and during this time mesoderm can be identified by its expression of conserved T-box TF TBXT, more commonly known as T or Brachyury (Wilkinson et al. 1990; Kispert and Herrmann 1994) which subsequently induces MESP1 (David et al. 2011). Lineage tracing of MESP1+ descendants show these cells contribute to all areas of the heart, in addition to some other mesodermal cell types. MESP1 is also required for migration of cells from the cardiogenic mesoderm through the streak, and thus is considered one of the earliest markers of multipotent cardiogenic mesoderm (Saga *et al.* 1999; Yoshida *et al.* 2008; Chan *et al.* 2013).

The movement of the cells to the anterior of the embryo brings them in close range of the AVE, which remains a source of Wnt signalling inhibitors (Kimura-Yoshida et al. 2005; Hoshino et al. 2015). This second phase of Wnt inhibition has been shown to be important for cardiac differentiation. This is illustrated well by the deletion of CTNNB (otherwise known as β -catenin) a positive downstream effector of Wnt signalling in the AVE, this results in the formation of multiple hearts in mice, demonstrating active Wnt signalling has a repressive effect on cardiac development at this time (Lickert et al. 2002). This has been further tested in mESCs where Wnt stimulation to form mesoderm followed by Wnt pathway inhibition is necessary and sufficient for successful cardiac differentiation (Naito et al. 2006, Buijtenlijk et al. 2020). Biphasic Wnt signalling is something that has also been exploited to produce CMs *in vitro* from human iPS cells, and subsequently these protocols have been used in this thesis (Lian et al. 2013; BurrIDGE et al. 2014).

1.3.3 Heart tube formation and the heart fields

Following these movements, the early cardiac progenitors (CPs) now sit in their correct position, towards the ventral side of the embryo. The embryo will subsequently undergo ventrolateral folding around the endoderm to form the primitive foregut. The heart forming regions are carried along with this and will fuse at the midline to form a crescent shape trough caudal to the differentiating cranial mesoderm (Peters *et al.* 1994; Christoffels and Jensen 2020). Cardiac progenitors and differentiating CMs are identifiable at this time through their expression of TFs such as GATA4, TBX5, and NKX2-5 that drive cardiogenesis forward, inducing the expression of structural genes such as *TNNT2*, *MYPBC3*, and *ACTN2* that are necessary for the identity and functionality of a CM (Ieda *et al.* 2010; Schlesinger *et al.* 2011; Bruneau 2013; Jain and Epstein 2018). This region also contains differentiating endocardial cells that initially form bilateral primitive heart tubes. Fusion of these tubes at the midline and their enveloping by CMs within the crescent creates a single primitive heart tube (Christoffels and Jensen 2020).

The cardiac crescent or heart forming region contains almost all the cells necessary for heart development. However, the boundaries of the heart forming region and whether this region is one field or many has been a source of debate (Abu-Issa *et al.* 2004; Moorman *et al.* 2007). Early labelling studies in chick and mice demonstrated that the cells of the outflow tract (OFT) and right ventricle are added after the formation of the linear heart tube from a population of *Fgf10* expressing cells located in the splanchnic mesoderm (Kelly *et al.* 2001; Mjaatvedt *et al.* 2001; Waldo *et al.* 2001). These cells are medial to the 'classical heart field' at E7.5 and can then be located dorsally and anteriorly to the heart tube after its formation. These results confirmed earlier embryological observations made in mice (Virágh and Challice 1973) and cell labelling experiments in chick, that indicated cells outside of the heart tube, originating from the splanchnic mesoderm, contribute to its growth (Arguello *et al.* 1975). Collectively these studies have led to the hypothesis that there are at least two pools of cardiac progenitors within the heart forming region. The 'classical heart field', henceforth referred to as the first heart field (FHF), whose cells commit early to form the linear heart tube and will later give rise to the left ventricle and contribute to the atria, and the second heart field (SHF), whose cells are added progressively to the heart tube to form the right ventricle, OFT, and the atria (Kelly and Buckingham 2002; Kelly *et al.* 2014).

The existence of two distinct populations was further investigated by Cai *et al.* 2003. This study showed that ISL1 null mice fail to form an OFT, right ventricle, and produce less atrial tissue, identifying it as a possible marker for the SHF. Staining for ISL1 at the cardiac crescent stage (E7-7.5) seemingly confirms this, showing it is expressed in the splanchnic mesoderm in a domain contiguous with but exclusive of the MYL7 positive FHF, reminiscent of the staining seen for FGF10 (Kelly *et al.* 2001). Furthermore, lineage tracing revealed ISL1+ descendants give rise to the OFT and right ventricle, some of the atria, and a small proportion of the left ventricle (Cai *et al.* 2003). Demonstrating the SHF contribute to more of the heart than previously thought. Subsequently it was noted that ISL1 is expressed in all cardiac progenitors prior to its restriction to the SHF (Prall *et al.* 2007; Ma *et al.* 2008). Therefore, bringing into question whether these cells are distinct fields, or are essentially the same field of cells at different stages of differentiation.

The temporal differences between the prospective fields was confirmed using clonal analysis of MESP1 expressing cells at different time points, which as noted earlier contribute to all areas of the heart and thus MESP1 marks both fields (Saga *et al.* 1999). Pulsed labelling of MESP1+ cells at the onset of gastrulation (E6.25) primarily labels the left ventricle and atria, whilst labelling at E7.25 marks the OFT and IFT (Lescroart *et al.* 2014). Similar results were generated by Devine *et al.*, who in addition demonstrated that labelling TBX5+ and MEF2CA+ populations from E6.5 specifically labels either the FHF or SHF respectively (Devine *et al.* 2014). Thus, confirming that these populations are temporally distinct from one another as they ingress through the PS at different times and that they are transcriptionally distinct from one another before they reach the cardiac crescent through their differential expression of TBX5 and MEF2C.

Recent studies using scRNA-seq on multiple time points has allowed for the transcriptional trajectory of the FHF and SHF to be explored in detail (Tyser *et al.* 2021; Zhang *et al.* 2021). In these studies, it has been noted that the SHF progenitors differentiate into CMs via two trajectories; one with an intermediary resembling FHF CPs, the other is distinct from this and represents a SHF specific route. The existence of this distinct route supports the idea that the two populations are distinct, whilst the common route shared by the FHF and some SHF progenitors and their eventual convergence on a similar CM

transcriptional state acknowledges the overlap in function of these cells and their close association with one another during development.

This study also yielded the identification of another field of cells, that expresses genes associated with the first and SHF, but can be uniquely identified by their expression of MAB21L2 and low NKX2-5 levels. Labelling of the MAB21L2+ cells at E8 has shown this field lies at the rostral border of the FHF and extends to the extraembryonic/embryonic border. MAB21L2+ its descendants give rise to cells in the FHF and SHF as well as the proepicardium. Due to the location of this field next to the FHF it has been named the Juxta-cardiac field (Tyser *et al.* 2021). Similar results have been obtained by Zhang *et al.* who traced these cells using HAND1 expression as a marker (Zhang *et al.* 2021). These results expand the boundaries of the heart forming region and reveal the variety of cell states within it that contribute to heart formation. A summary of the genes that can be used to identify each region is shown in figure 1.4.

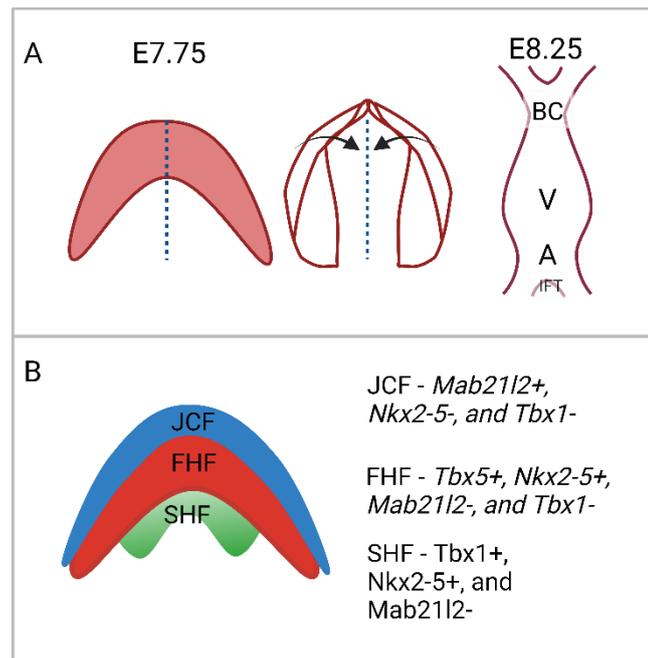


Figure 1.4 An overview of heart tube formation and the cardiac fields

(A) Around E7.75 the bilateral cardiac fields (red) will start to fuse at the midline (blue dotted line). Ventrolateral folding of the embryo folds the heart forming field towards the midline. Fusion of primitive endocardial tubes within the heart forming region and the covering of these with cardiomyocytes leads to the formation of a single linear heart tube around E8-8.25. The early heart tube contains a primitive ventricle (V) that gives rise to the left ventricle, the future atrium (A), the inflow tract (IFT), and bulbous cordis (BC) from which the right ventricle, outflow tract, aorta and pulmonary trunk develop. (B) A representation of the heart fields that contribute to heart tube formation and elongation. The first heart field (FHF, red), at E7.5 the second heart field (SHF, green) is medial to this, and at the rostral border of the FHF sits the newly discovered juxtacardiac field (JCF, blue). Markers of these regions are also shown. This figure was created using Biorender and adapted from figures presented by Christoffels and Jensen 2020, and Tyser *et al.* 2021.

Shortly after formation of the heart tube the future atrial and ventricular portions acquire a chamber specific gene expression pattern (Christoffels et al. 2000; Tyser et al. 2016). Expression of a subset sarcomeric genes can be seen even at the cardiac crescent stage (E7.5) and beating can be seen in the early heart tube at E8-9 in mice (Chen et al. 2010; Tyser et al. 2016; Tyser and Srinivas 2020). The heart tube continues to elongate through the addition of proliferative SHF cells, unequal proliferation within the heart tube

and the biased addition of the SHF progenitors is thought to contribute to rightward looping of the linear heart tube (Rana *et al.* 2013; Saijoh and Hamada 2020). The heart tube will subsequently undergo septation and valvulogenesis through the contribution of muscular septae, mesenchymal, and endocardial cells, to create the intricate 4 chambered structure of the adult heart. These morphogenic processes are discussed in further detail by (Christoffels *et al.* 2000; Butcher and Markwald 2007; Rana *et al.* 2013; Anderson *et al.* 2014; Saijoh and Hamada 2020). After birth heart size increases further, as the CMs formed during gestation grow by hypertrophy and mature (Buijendijk *et al.* 2020).

1.4 Genetic regulation of cardiac development

The steps of cardiac development are guided by a host of transcription factors (TFs), some of which have been mentioned above. These TFs make up a large proportion of the CVD and CHD risk genes discussed previously. TFs are proteins possessing one of a variety of DNA binding domains (DBD) that typically recognise a 6-12 bp DNA motif (Wunderlich and Mirny 2009; Lambert *et al.* 2018). They also possess effector domains that allow them to mediate protein-protein interactions with other TFs or chromatin modifiers, to activate or repress target gene expression. This allows for the formation of large, specialised TF complexes, recognising specific combinations of motifs (He *et al.* 2011; Junion *et al.* 2012; Luna-Zurita *et al.* 2016). This complexity is necessary to determine binding specificity as the DNA motifs recognized by a given TF is likely to be relatively common throughout the genome. The additive nature of these interactions also allows for discrete modulation of transcription to be achieved.

1.4.1 Interactivity and cooperativity between members of the cardiac GRN

The core network of TFs that orchestrate cardiac development, includes GATA4/5/ 6, NKX2-5, HAND1, 2, TBX2/ 5/ 20 and proteins belonging to the MADS domain family such as MEF2 and SRF (Olson 2006; Bruneau 2013). Evidence from germline and conditional mouse knockout models has shown that the absence of any one of these factors results in an array of severe cardiac defects that almost invariably lead to embryonic lethality, with the exception of GATA5 to be discussed in more depth later (section 1.4.4) (Lin *et al.* 1997; Srivastava *et al.* 1997; Firulli *et al.* 1998; Morrisey *et al.* 1998; Riley *et al.* 1998; Koutsourakis *et al.* 1999; Tanaka *et al.* 1999; Bruneau *et al.* 2001; Molkenin *et al.* 2002; Watt *et al.* 2004; Niu *et al.* 2005; Stennard *et al.* 2005). This highlights the vulnerability of the cardiac GRN to

perturbation. However, it should also be noted that loss of a single member does not entirely halt cardiomyogenesis. Thus, in this regard the cardiac GRN is somewhat resilient.

Invariably these factors have additional roles during development outside of cardiogenesis where they bind regulatory sites of non-CM genes and regulate their expression. It has been determined that combinatorial binding is needed to establish and maintain their cardiac specific binding patterns and to produce robust expression of cardiac genes (Durocher et al. 1997; Morin 2000; Hiroi et al. 2001; Habets et al. 2002a; Maitra et al. 2009; Ang et al. 2016a; Kinnunen et al. 2018). The utilisation of ChIP-seq has illustrated just how widespread combinatorial regulation of transcription by members of the cardiac GRN is during heart development (He et al. 2011; Schlesinger et al. 2011; Luna-Zurita et al. 2016; Akerberg et al. 2019; Robbe et al. 2022). Analysis of TBX5, NKX2-5 and GATA4 chromatin occupancy in CPs and CMs demonstrates that co-occupancy by all three factors (TBX5, NKX2-5, and GATA4) is associated with the highest expression of CP and CM specific genes. Loss of one cardiac factor (NKX2-5 or TBX5), or a combination of both leads to the loss of the other factors from CP and CM specific binding sites, and their relocation to non-cardiac sites (Luna-Zurita et al. 2016). Supporting the notion that accurate and robust control of gene expression by TFs is co-dependent and co-operative, the other ChIP-seq experiments referenced above agree with this conclusion and explore the binding patterns of additional members of the cardiac GRN.

1.4.2 GATA4 is a core member of the cardiac GRN

GATA4 may be considered a key member of the cardiac GRN. Firstly, it is expressed throughout cardiac development and in adult CMs where it is a key mediator of homeostasis and hypertrophy (Pikkarainen et al. 2003; Bisping et al. 2006; Oka et al. 2006). As discussed, during these processes GATA4 binding activity is associated with the upregulation of many CM specific genes and the repression of alternative gene programs (Schlesinger et al. 2011; He et al. 2014; Ang et al. 2016b; Luna-Zurita et al. 2016; Akerberg et al. 2019; Robbe et al. 2022). Furthermore, in humans several mutations in *GATA4* are associated with the development of CHDs and CVDs (Garg et al. 2003; Rajagopal et al. 2007; Wang et al. 2013; Yang et al. 2013). All these factors point to a central role for GATA4 in heart development. Early evidence for the importance of GATA4 in cardiomyogenesis comes from studies using P19 mouse carcinoma cells. This cell line can be induced to form CMs through the addition

of DMSO, knockdown of GATA4 inhibits this process blocking CM formation (Grépin et al. 1995). Furthermore, over expression of GATA4 in P19 cells significantly increases the number of CMs formed and can achieve this even in the absence of DMSO (Grépin et al. 1997). In this model these findings suggest that GATA4 is essential for CM formation and is sufficient to drive that process.

In contrast to the above, the earliest attempts to study GATA4 *in vivo* in mouse knockout models failed to identify a central role for GATA4 in CMs themselves. Loss of GATA4 was found to be lethal by E7.5-10.5, due to defects in ventral body wall closure and heart development (Kuo et al. 1997; Molkentin et al. 1997). The GATA4 null mice presented with cardia bifida where two cardiac patches or tubes form, identifiable by their expression of CM differentiation markers such as MYH6, NPPA, and MYL2 (Molkentin et al. 1997) although beating was not observed. Furthermore, the injection of GATA4^{-/-} mESCs into otherwise WT embryos revealed the GATA4 null cells could contribute CMs to an apparently normal heart (Kuo et al. 1997). This combination of results suggested the primary cause of the cardiac defect was in the migration of the heart precursors, proposed to be because of a requirement for GATA4 in other tissues such as VE and the associated defects in ventral body wall closure. This requirement was later confirmed by the observation that embryos generated through tetraploid complementation where wild type VE is present can successfully form a linear heart tube (Watt et al. 2004). However, further defects in cardiac development were observed, such as hypoplastic ventricular myocardium, disrupted looping, septation, and a failure to form any proepicardium.

A study by Zeisberg *et al.* 2005 demonstrates that the defects observed by Watt *et al.* are in part caused by a direct requirement for GATA4 in developing CMs. In this study, Cre recombinase (Cre) was placed under the control of the endogenous promoters for *Nkx2-5* or *Myh6* deleting *Gata4*^{fl/fl} in early CPs and established CMs respectively. Deletion of GATA4 in CPs often resulted in the formation of a single left ventricular chamber, indicating right ventricle formation had failed. Furthermore, the left ventricular myocardium was found to be hypoplastic, recapitulating some of the morphological defects seen by Watt *et al.* 2004. Later deletion using *Myh6* driven Cre also resulted in hypoplasia of the myocardium. Further analysis revealed that proliferation was reduced at both stages partially explaining these phenotypes (Zeisberg et al. 2005). These findings establish that

GATA4 has an intrinsic role in CP and CM proliferation that is necessary for normal heart development. Furthermore, direct binding and induction of cell cycle genes; *Cdk4*, and *Ccnd2*, by GATA4 has been demonstrated in CMs revealing the mechanism underlying the reduction in proliferation that was observed by Zeisberg *et al.* (Rojas et al. 2008).

In addition to its intrinsic role in proliferation, studies using ectopic expression of GATA4 show it can drive the formation of beating tissue in certain settings. In *Xenopus laevis* ectodermal explants that would otherwise form epidermis, Gata4 overexpression results in the formation of beating tissue and the induction of sarcomeric genes such as *myl2*, *myh6*, and *tnni3* (Latinkic et al. 2003). In other models and developmental contexts additional factors are required to achieve the formation of beating tissue. For example, overexpression of GATA4 in combination with TBX5 and chromatin remodelling subunit SMARCD3 can induce ectopic beating tissue in the mouse epiblast mesoderm (Takeuchi and Bruneau 2009). Accordingly, the expression of various CM specific genes necessary for beating, such as MYL7 and ACTC1 were induced as a consequence of this. In fully differentiated cell types such as rat tail-tip fibroblasts overexpression of GATA4 in combination with TBX5 and MEF2C is required for the formation of beating tissue, with GATA4 specifically required to induce the expression of TNNT2, a pan-myocardial sarcomeric gene essential for contraction (Ieda et al. 2010). GATA4 has also been utilised for the transdifferentiation of mouse and human fibroblasts with varying efficiencies (Song et al. 2012; Nam et al. 2013). Additional transdifferentiation studies are reviewed extensively by Adams *et al.*, where it is noted that GATA4 is one of the most common factors included in transdifferentiation experiments (Adams et al. 2021). That GATA4 is often found to be necessary to drive cardiogenesis in these experiments supports the conclusion that it is a potent cardiogenic factor, establishing GATA4 at the core of the cardiac GRN (see figure 1.5).

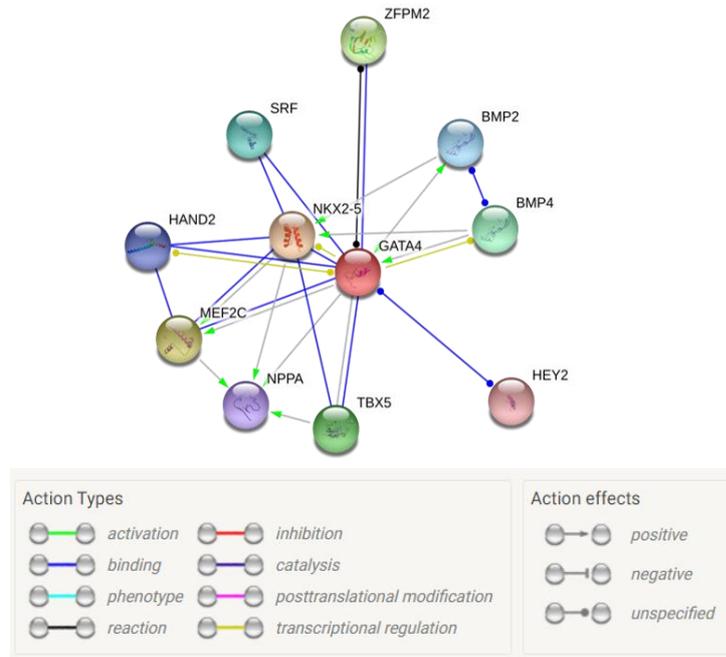


Figure 1.5 Interconnectivity between the cardiac gene regulatory network

This network showing GATA4 at the centre shows just 10 of the genes that GATA4 is known to interact with. Showing that GATA4 physically binds to and/or regulates the transcription of many of the other core components of the cardiac GRN. Diagram made using the STRING database (Szklarczyk *et al.* 2019).

1.4.3 The origins of GATA4 and the GATA family of transcription factors

GATA4 is a member of the GATA family of TFs, which in the human genome is composed of GATA factors 1-6. These factors all possess a well conserved GATA DBD made up of two zinc fingers which recognise the DNA motif (A/T)GATA(A/G). Although the DBD is highly conserved between the factors, the carboxy and amino termini have diverged (see figure 1.6). Through their sequence similarity, and their expression profiles the GATA family of TFs can be split into two sub-families; GATA1-3 and GATA4-6 (Lowry and Atchley 2000; Peterkin *et al.* 2005; He *et al.* 2007; Lentjes *et al.* 2016). This division is further substantiated by the analysis of the genomes of two simple deuterostomes *Branchiostoma floridae* and *Saccoglossus kowalevskii* which have rudimentary hearts. This identified two GATA genes: one is ancestral to GATA1-3 and the other to GATA4-6 (Moller and Philpott 1973; Gillis *et al.* 2009; Kaul-Strehlow and Stach 2013). The common evolutionary origin of the GATA TFs may explain the potential for redundancy between them, as will be discussed below.

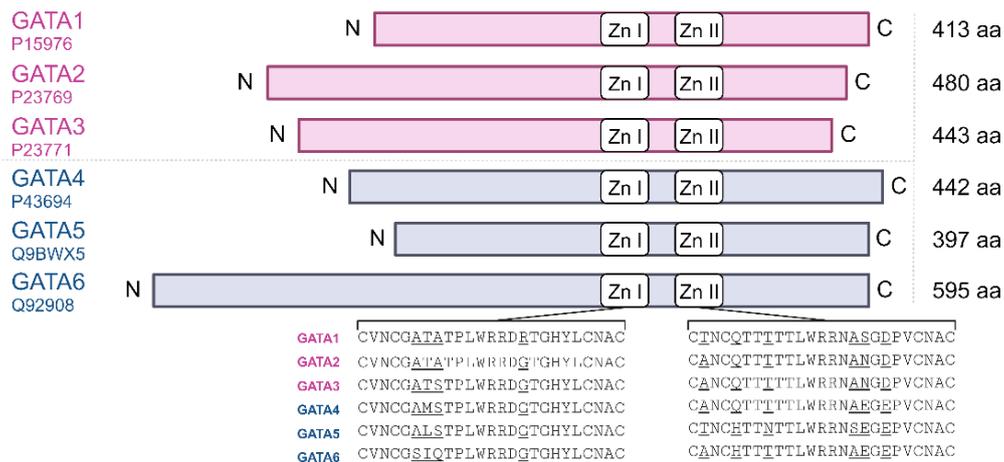


Figure 1.6 A comparison of sequence similarity between GATA TFs 1-6

Each of the human GATA factors is represented above. Their zinc fingers have been aligned using the T-Coffee sequence alignment tool (Notredame *et al.* 2000) and the sequence differences between the factors within these regions have been underlined. This highlights the high degree of sequence conservation within the DNA binding domain between the factors. However, outside of the DNA binding domain the length of the N and C termini of the proteins are variable in size, reflecting the sequence divergence between the factors in these domains. This figure was created using BioRender and was adapted from Lentjes *et al.* 2016 and Tremblay *et al.* 2018.

1.4.4 The roles of GATA factors 1 to 3 in haematopoiesis and beyond

GATA factors 1, 2, and 3, are largely devoted to haematopoietic development, but have some direct and indirect links to cardiac development as will be discussed. GATA1 is the founding member of the GATA family and is required for erythroid maturation, with GATA1^{-/-} mice succumbing to severe anaemia around E10.5 due to a lack of definitive erythroid cells (Pevny *et al.* 1991; Pevny *et al.* 1995). Following these studies, that demonstrated the early lethality of GATA1 null mice, conditional knockout models were utilised to investigate additional roles for GATA1 later in development. Defining roles for GATA1 in haematopoietic cells, such as, megakaryocytes, dendritic cells, basophils, and eosinophils (Ohneda and Yamamoto 2002). No direct role for GATA1 in cardiac development has been identified; however, both the haematopoietic system and cardiovascular system have mesodermal origins and develop in close proximity to one another. Further to this, the MESP1+ population of cells that give rise to cardiac mesoderm in mice also gives rise to

haematopoietic cells (Chan *et al.* 2013). Repression of GATA1 by cardiogenic TF NKX2-5 has been found to be important in directing mesodermal cells to a cardiac fate. In NKX2-5 null mice the erythroid gene program is derepressed in CPs, accordingly overexpression of NKX2-5 represses these erythroid genes including GATA1. Therefore, it seems these factors antagonise one another to help direct the early segregation of the blood and cardiac lineages (Caprioli *et al.* 2011).

Similarly, to GATA1 loss of GATA2 expression is embryonic lethal by ~E10 in mouse development due to severe anaemia. Both primitive and definitive erythrocytes are lacking in these embryos, this was attributed to a requirement for GATA2 in the expansion of multipotent haematopoietic progenitors (Tsai *et al.* 1994; Tsai and Orkin 1997). GATA2 also has important roles in the development of mast cells (Tsai *et al.* 1994; Tsai and Orkin 1997), the kidneys, urogenital system (Zhou *et al.* 1998), various neuronal cell types (Zhou *et al.* 2000; Craven *et al.* 2004; Kala *et al.* 2009; Willett and Greene 2011), and in the trophoblast where it shows redundancy with GATA3 (Home *et al.* 2017). Once again, much like GATA1, GATA2 may be linked to cardiogenesis indirectly. In human iPSC-CM differentiation knockout of GATA2 increases CM differentiation efficiency whilst overexpression of GATA2 reduces it (Castaño *et al.* 2019). Furthermore, in RNA-seq and ChIP-seq analysis conducted by Castaño *et al.*, they observed direct binding and repression of cardiac genes such as *TBX3*, by GATA2 in haematopoietic progenitor cells. Echoing the results seen for GATA1 and suggesting these factors can both promote haematopoietic fates at the expense of myocardial fates.

In haematopoiesis GATA3 is implicated in the differentiation of the T cell lineage (Ting *et al.* 1996; Zhu *et al.* 2004), thus is also commonly referred to as a haematopoietic factor. However, loss of GATA3 expression has a multitude of effects on mouse development, some of which imply it has direct roles in cardiac development. GATA3 null embryos are recoverable up to E11, and present with a wide range of defects including haemorrhaging, renal insufficiency, hypoplastic mandibles, adrenal gland insufficiency, and cardiac structural defects affecting the endocardial cushions and OFT (Pandolfi *et al.* 1995). Development can be partially rescued in GATA3 null mice through the expression of GATA3 in the sympathoadrenal lineages, or by the supplementation of pregnant dams carrying GATA3^{-/-} mice with noradrenaline intermediaries, that would usually be produced by the

adrenal glands (Lim *et al.* 2000; Moriguchi *et al.* 2006). Despite these remedies that alleviate the adrenal gland insufficiency seen the mice still present with structural cardiac defects. Expression of GATA3 in the developing heart appears to be restricted to endocardial cells therefore it is likely these defects are due to a requirement for GATA3 in endocardial cells. In summary, although GATA factors 1-3 have important roles in haematopoiesis, they, like the other members of their family (GATA4-6), have additional roles in development and show the potential for redundancy with one another in some contexts, as will be discussed below for GATA4-6. Furthermore, while GATA1 and GATA2 are indirectly linked to cardiac development, the direct roles of GATA3 in cardiac morphogenesis have received little attention.

Table 1.1 A summary of the primary phenotypes caused by GATA factor knockout in mice.

For further studies relating to GATA4, 5, and 6 see table 1.2.

Gene	Earliest lethality	Cause of lethality	References
GATA1	E10-11	Failure to form mature erythrocytes. Leading to severe anaemia.	(Pevny <i>et al.</i> 1991; Pevny <i>et al.</i> 1995)
GATA2	E10-11.5	Hematopoietic progenitor expansion fails. Leading to severe anaemia.	(Tsai <i>et al.</i> 1994; Tsai and Orkin 1997)
GATA3	E11-12	Defects in many organ systems observed. The most immediate cause of lethality seems due to noradrenaline insufficiency, exacerbated by cardiac morphogenic defects.	(Pandolfi <i>et al.</i> 1995; Raid <i>et al.</i> 2009)
GATA4	E8-9.5	Initially due to defective visceral endoderm formation.	(Kuo <i>et al.</i> 1997; Molkenin <i>et al.</i> 1997; Watt <i>et al.</i> 2004)
GATA5	Viable	-	(Molkenin <i>et al.</i> 2002)
GATA6	E4.5	Defective primitive endoderm formation.	(Morrissey <i>et al.</i> 1998; Koutsourakis <i>et al.</i> 1999; Zhao <i>et al.</i> 2005)

1.4.5 GATA factors 4, 5, and 6 in cardiac development

As detailed in section 1.4.3, GATA4/5/6 have a common evolutionary origin in animals with rudimentary hearts suggesting an ancestral GATA factor was once sufficient for all cardiac roles. In the invertebrate *Drosophila melanogaster* this seems to be the case. In *Drosophila* the GATA TF *pannier* is most closely related to GATA4. Early studies in *Drosophila* showed that knockout of *pannier* leads to a failure in cardial cell production (Gajewski et al. 1999). Conversely over-expression of *pannier* leads to superfluous production of cardial cells, an effect that can be phenocopied by the overexpression of mouse GATA4 (Gajewski et al. 1999) demonstrating that in this scenario GATA4 can take on all roles in cardiogenesis. However, in mice which like humans have 6 GATA factors knockout of both GATA4 and GATA6 is needed to produce acardia (Zhao et al. 2008), consistent results have been obtained in Zebrafish and *Xenopus* (Holtzinger and Evans 2007; Haworth et al. 2008; Sam et al. 2020), suggesting that the potential for redundancy between these factors has been conserved.

In WT mouse embryos GATA4 and 6 are co-expressed in many cell types throughout the heart including the myocardium, whereas GATA5 expression is largely restricted to endocardium and endocardial cells (Nemer and Nemer 2003). In agreement with this GATA5^{-/-} mice develop bicuspid aortic valve, which can be phenocopied by endocardial specific deletion of GATA5 (Molkentin et al. 2002; Laforest et al. 2011), confirming in the mammalian heart that GATA5 is primarily required for endocardial development. CMs are the largest contributors to heart volume in humans (Litviňuková et al. 2020) yet the relative contribution of GATA4 and GATA6 to their development has not been clearly resolved.

As discussed previously, knockout of GATA4 or GATA6 in mice results in embryonic lethality due to defects in the development of extraembryonic tissues such as the PE and VE (Kuo et al. 1997; Molkentin et al. 1997; Koutsourakis et al. 1999). Tetraploid complementation experiments where GATA4 and/or 6 are knocked out in the epiblast specifically, and the extraembryonic tissues are derived from WT mESCs have been used to circumvent this. These studies suggest that knockout of GATA4 rather than GATA6 has a more detrimental on cardiac development. The GATA4 null embryos generated presented with an array of cardiac defects, such as a dilated heart tube lacking chambers, and hypoplastic ventricular myocardium with reduced trabeculation. In these mice, there is also

a complete failure to form the proepicardium, suggesting that GATA4 may play a pivotal role in the recently uncovered juxtacardiac field from which the proepicardium forms (Watt et al. 2004). In comparison, GATA6 null embryos have relatively normal hearts, with discrete reductions in trabeculation recorded for some embryos (Zhao et al. 2005). In both the GATA4 and 6 knockout mice CMs are still formed, and the expression of cardiac progenitor markers *Nkx2-5*, *Hand 1/2*, *Srf*, and definitive CM genes such as *Myh6/7*, and *Myl2/7* are also maintained despite them containing GATA responsive elements (Watt et al. 2004; Zhao et al. 2005). That expression of these genes is maintained and that CMs are still formed is attributed to the potential for GATA4 and GATA6 to compensate for one another, as they are both able to bind (A/T)GATA(A/G) motifs thus can potentially regulate the same targets. Indeed, upregulation of GATA6 has been observed in multiple studies when GATA4 is knocked out (Kuo et al. 1997; Molkenin et al. 1997; Watt et al. 2004). In support of this, heterozygosity for GATA4 or 6 in mice does not lead to any apparent heart malformations, whereas compound heterozygosity results in myocardial thinning and septation defects (Xin et al. 2006a). This suggests that if a certain level of either GATA4 or GATA6 is maintained cardiac development can proceed normally. However, when *Gata6* cDNA is used to replace the endogenous *Gata4* coding sequence, heart development is still abnormal (Borok et al. 2016), strongly suggesting that GATA4 has critical and unique roles in mammalian cardiac development.

Further studies reveal that GATA4, 5, and 6 have different levels of cardiogenic activity. Overexpression of *Gata4* in *Xenopus* animal cap explants otherwise fated to become ectoderm is sufficient to induce the expression of cardiac specific genes and produces spontaneously beating tissue (Latinkic 2003). Injection of *gata5* mRNA induces cardiac gene expression by a comparable amount to *gata4*, but *gata6* mRNA does so to a lesser extent and induces the expression of liver specific markers also, suggesting that *Gata4* and 5 have higher cardiogenic activity in *Xenopus* (Latinkic 2003). Other evidence for factor specific roles comes from analysis of the interaction of GATA4 or GATA6 with another cardiogenic TF, TBX5. Compound heterozygous GATA4/TBX5 mutants harbour more severe and penetrant cardiac defects than those seen in GATA6/TBX5 mutants. Moreover, expression of the *Myh6* gene is maintained in GATA6/TBX5 mutants but not GATA4/TBX5 mutants revealing a requirement for GATA4 in the expression of *Myh6* that cannot be

compensated for by GATA6 (Maitra et al. 2009). Therefore, there are significant differences in the apparent abilities of GATA4/5/6 to compensate for one another in these different model organisms. The most compelling evidence for a non-redundant role for GATA4 in human cardiac development is the occurrence of CHDs in humans with highly penetrant GATA4 CHD associated haploinsufficiency mutations (Prendiville et al. 2014). This implies there is a low level of tolerance for a decrease or alteration in GATA4 activity in humans, which is perhaps not captured in the mouse models discussed. The discordance between human and mouse models will be discussed in further detail below.

1.4.6 GATA4 in human models of cardiac development

Study of the CHD causing GATA4 G296S human mutation in mouse and iPSC-CM models identifies integral roles for GATA4 in cardiac development and highlights the differences between species. This mutation prevents GATA4 from physically interacting with TBX5 and is associated with the occurrence of a range of atrial and ventricular septal defects in humans, of a severity often requiring surgery (Garg et al. 2003). In contrast, heterozygous mouse models carrying the corresponding mutation appear grossly normal, with a proportion showing mild defects in post-natal atrial septal closure. A comparable phenotype to that seen in humans is only seen in homozygous mice (Misra et al. 2012). This suggests the human cardiac GRN is more sensitive to disruption of GATA4 activity, therefore assessing the activities of GATA4 in a human genetic background is a necessity. Further, analysis using patient-derived iPSCs-CMs showed that recruitment of GATA4 to cardiac enhancers by TBX5 was impaired as a result of the mutation, and there was a failure to repress ectopic endothelial gene programs in these cells. Indicating interaction between GATA4 and TBX5 is required for the maintenance of stable cell identity in CMs. Furthermore, the CMs generated displayed defects in contractility, physiology, and sarcomeric organisation (Ang et al. 2016). These studies demonstrate that human iPSC-CM models may provide further insights into the phenotypic changes associated with human disease-causing mutations than possible from mouse models. Supporting the use of these models to further scrutinise the roles of GATA factors in cardiac development.

Results for human iPSC GATA4 and GATA6 knockout lines have recently been published. The findings from these papers initially suggests that the requirement for GATA6 in iPSC-CM differentiation is greater than for GATA4 (Sharma et al. 2020; Gonzalez-Teran et

al. 2022). In the report published by Sharma et al. loss of GATA6 expression in iPSCs was shown to be incompatible with the formation of CMs, with no beating or TNNT2 positive cells observed. In contrast to this, GATA4 knockout iPSCs were able to form CMs with some deficiencies in differentiation noted, such as delayed differentiation, reduced differentiation efficiency, and slower beating (Gonzalez-Teran et al. 2022). Given the information described in the previous paragraphs regarding the relative contributions of GATA4 and 6 to cardiac development in mice this is a somewhat surprising result but may be due to stage specific requirements for each factor. In the GATA6 knockout line, expression of *MESP1* was found to be down regulated at day-4 suggesting early deficits in cardiac mesoderm formation. Information regarding the expression of cardiac mesoderm markers is not available for the GATA4 line, however as beating CMs were formed it can be assumed the cells have successfully passed through this stage. RNA-seq data for day-6 cardiac progenitors (CP) generated from the GATA4 null lines was available and this data demonstrates that loss of GATA4 had a significant effect on the CP transcriptome. An array of cardiogenic TFs such as *TBX2/5/18/20*, and *MEF2C* were found to be down-regulated as was the expression of an array of CM specific genes. Therefore, supporting a central role for GATA4 as a regulator of the cardiac GRN in this model, but perhaps at a later stage than GATA6. Further research is needed to clarify these points.

Table 1.2 A summary of results from GATA4, 5, and 6 loss of function models

The studies referenced below summarise numerous mouse loss of function models for GATA4, 5, and 6. Available data from relevant human iPSC loss of function models is also included and are highlighted in blue. AVSD = Atrial ventricular septal defect, DORV = double outlet right ventricle, and BAV = Bicuspid aortic valve. Table adapted from Borok *et al.* 2016

Genotype	Lethality	Phenotypic features	References
<i>Gata4</i> ^{-/-}	E8.5-9.5	Ventral folding fails. Cardia bifida.	Molkentin <i>et al.</i> 1997
<i>Gata4</i> ^{+/-}	Viable	No apparent abnormalities	Kuo <i>et al.</i> 1999 Molkentin <i>et al.</i> 1997
<i>Gata4</i> ^{-/-} , tetraploid complementation	E10.5	The right ventricle, proepicardium and septum transversum mesenchyme fail to form. Looping fails and the myocardium is hypoplastic.	Watt <i>et al.</i> 2004
<i>Gata4</i> ^{V217G/V217G} (disrupts interaction with FOG2)	E13.5	AVSD and DORV. Hypoplastic ventricles.	Crispino <i>et al.</i> 2001
<i>Gata4</i> ^{fl/fl} , <i>Nkx2.5</i> -Cre (deletion in cardiac progenitors)	E10.5	The right ventricle fails to form, hypoplastic myocardium, defects in endocardial cushion formation.	Zeisberg <i>et al.</i> 2005
<i>Gata4</i> ^{fl/fl} , <i>Mef2c</i> -Cre (deletion in SHF)	E12.5	Right ventricle hypoplasia.	Rojas <i>et al.</i> 2008
<i>Gata4</i> ^{fl/fl} , <i>Myh6</i> - or <i>Myh7</i> -Cre	Not lethal	Reduced basal cardiac function, unable to instigate compensatory hypertrophy in response to pressure overload leading to heart failure.	Oka <i>et al.</i> 2004
<i>GATA4</i> ^{G296S/G296S} (disrupts interaction with TBX5)	N/A iPSC-CM	The CMs formed are disorganised and display impaired contractility, calcium handling, and metabolic activity. Cardiac gene expression is less robust, whilst endocardial gene expression is derepressed.	Ang <i>et al.</i> 2016
<i>GATA4</i> ^{-/-}	N/A iPSC-CM	Reduced iPSC-CM differentiation efficiency. Slower beat rate.	Gonzalez-Teran <i>et al.</i> 2022
<i>Gata4</i> ^{-/-} <i>Gata6</i> ^{-/-} , tetraploid complementation	E10.5	Acardia	Zhao <i>et al.</i> 2008
<i>Gata4</i> ^{Gata6/Gata6}	E12.5	Reduced septum transversum mesenchyme and proepicardium formation, thin myocardium, AVSD, and hypocellular endocardial cushions.	Borok <i>et al.</i> 2016
<i>Gata5</i> ^{-/-}	Viable	Viable and fertile, genitourinary defects	Molkentin <i>et al.</i> 2002
<i>Gata5</i> ^{-/-}	Viable	Higher incidence of BAV (~25%)	LaForest <i>et al.</i> 2011
<i>Gata5</i> ^{fl/fl} Tie-Cre	Viable	Higher incidence of BAV (~21%)	LaForest <i>et al.</i> 2011
<i>Gata6</i> ^{-/-}	E5.5-7.5	Defects in primitive endoderm formation.	Morrissey <i>et al.</i> 1998 Koutsarakis <i>et al.</i> 1999
<i>Gata6</i> ^{-/-} , tetraploid complementation	E10.5	Liver development fails. Small reduction in ventricular trabeculation in some embryos.	Zhao <i>et al.</i> 2005
<i>Gata6</i> ^{+/-}	Viable	Higher incidence of BAV (~56% in males, and 27% in females).	Garibeh <i>et al.</i> 2018
<i>GATA6</i> ^{-/-}	N/A	No CMs formed.	Sharma <i>et al.</i> 2020

1.4.7 Expanding the Gene Regulatory Network

In addition to providing clarity on the role of GATA4 in cardiac development, it is also important to identify its downstream target genes and understand their contribution to the cardiac GRN. Previous work in our lab using over expression of Gata4 in pluripotent *X. laevis* animal cap explants as a model of cardiogenesis (Latinkic 2003) combined with an RNA-seq screen generated an extensive list of candidate target genes for Gata4. Two such genes are *tbx2* and *prdm1* and these were validated further by a combination of methods. RT-PCR analysis was used to confirm the trends observed in the RNA-seq screen, verifying *tbx2* as a gene that is consistently upregulated following Gata4 activation, whereas *prdm1* is downregulated. CHIP RT-PCR confirmed Gata4 binds at selected regulatory regions for both genes, indicating direct regulation of these genes by Gata4. MO mediated knockdown of *tbx2* and over expression of *prdm1* through targeted injection both interfere with normal cardiac development, demonstrating that the regulated expression of these genes is needed for normal cardiac development to take place in *Xenopus* (Latinkic et al, data not published). What is currently known about TBX2 and PRDM1 in cardiac development is explored below.

1.5 TBX2

1.5.1 The T-box family of transcription factors

TBX2 belongs to the T-box superfamily of TFs, that has origins before the evolution of multi-cellularity (Sebé-Pedrós et al. 2013). In accordance with this T-box factors are ubiquitous across the animal kingdom and function in many developmental contexts. The number of T-box factors has expanded with organism complexity: in humans 17 have been identified (Papaioannou 2001; Papaioannou 2014). This expansion has allowed for diversification of their roles: in humans Tbx factors are necessary for pluripotency, germ layer formation, organogenesis, and limb development, amongst other roles (Papaioannou 2014b). Consequently, mutations in these genes are associated with a wide range of developmental diseases. They are related to one another through their DBD of 180-200 amino acids referred to as the T-box, which recognizes the consensus half site AGGTGTGAAA. Analysis of sequence divergence in the T-box domain allows them to be grouped into 5 sub-families (see figure 1.7). TBX2 belongs to the TBX2 sub-family, which also includes TBX3, TBX4, and TBX5. Within this group TBX2 and -3 are most closely related to one another (95.48% similarity within their T-box domain, 59.2% overall), and TBX4 to TBX5 (93.89% similarity between their T-box domain, 55.73% overall). High conservation of

residues within the DBDs of TBX2 and 3, means they will likely recognise the same DNA motifs and where co-expressed have the potential to regulate the same target genes. When considering the entirety of the proteins, there has been significant divergence between TBX2 and 3 outside of the DBDs and this likely affects which proteins they can interact with, which may dictate differences in the targets these genes regulate.

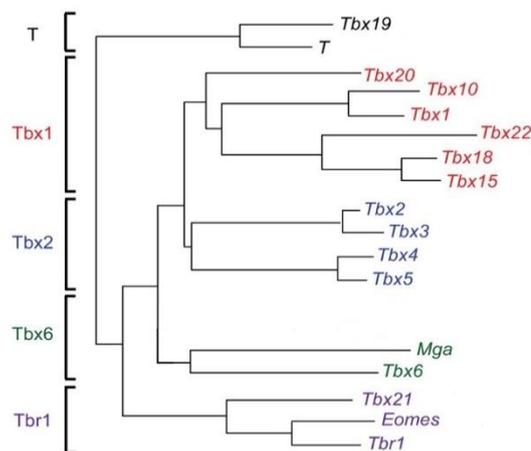


Figure 1.7 The divergence of the T-box family

The 17 human T-box factors grouped into their 5 sub-families based on sequence similarity between their DNA binding domains. Adapted from (Papaioannou *et al.* 2014).

1.5.4 T-box factors in cardiac development

In the context of heart development, the most well studied of the Tbx factors is TBX5 which is a co-factor of GATA4 and a member of the core cardiac GRN as reviewed previously. Mutations in TBX5 are strongly associated with Holt-Oram Syndrome (HOS). The primary feature of this syndrome is upper limb deformities relating to TBX5's role in limb bud patterning, often with accompanying cardiac abnormalities including conduction abnormalities, Tetralogy of Fallot, atrial and ventricular septal defects (Basson *et al.* 1997; Packham 2003; Mcdermott *et al.* 2005). In *Xenopus laevis* expression of a dominant negative version of Tbx5 results in a heartless or severely reduced heart phenotype (Horb and Thomsen 1999). In mice, TBX5^{-/-} mutants die early in development, as heart development arrests at E9.5 with formation of a hypoplastic heart tube that fails to undergo looping (Bruneau *et al.* 1999; Bruneau *et al.* 2001). Its strong association with HOS which presents in 1:100,000 live births, and the dramatic phenotypes seen in loss of function studies means TBX5 is well placed as a key cardiac TF and therefore has been a focus of research attention.

1.5.2 TBX2 in AVC formation

Numerous other T-box factors have important roles during heart development including GATA4 target TBX2 and closely related factor TBX3 (Greulich et al. 2011). Current evidence places TBX2 in opposition of other T-box factors, like TBX5. As a transcriptional activator TBX5 induces the expression of a host of myocardial genes including *Nppa* and *Gja5* broadly throughout the myocardium (Steimle and Moskowitz 2017). TBX5 is also expressed in the atrioventricular canal (AVC), overlapping with expression of TBX2 and TBX3. Expression of TBX2 and 3 has been shown to be necessary for repression of chamber myocardial genes such as *Nppa* and *Gja5* in the AVC (Christoffels et al. 2004; Hoogaars et al. 2008). In primary atrial and ventricular CM cultures, co-transfection with TBX2 significantly reduces the ability of TBX5 and NKX2-5 to synergistically activate the *Nppa* promoter (Habets et al. 2002b) showing TBX2 can compete for TBE binding sites in this context and suggesting it may perform this function in the AVC. This is a role TBX3 may also be able to fulfil based on its high degree of homology to TBX2 and its overlapping expression pattern. Further research seems to confirm this as knockout of TBX2 and 3 is needed to abolish the chamber restricted expression pattern of NPPA and GJA5 in mice (Singh et al. 2012). Implying that either TBX2 or 3 can mediate repression of chamber development in the AVC to some extent.

1.5.3 Are there earlier roles for TBX2 in cardiac development?

The events required for AVC formation occur during morphogenesis after the heart fields have been specified and the heart tube formed; however, the expression pattern of TBX2 throughout development suggests the potential for earlier roles. TBX2 expression is first detected at E8 in mice within the cardiac crescent just prior to the formation of the early heart tube; its expression is later restricted to the regions that will form the AVC, and OFT (Christoffels et al. 2004; Aanhaanen et al. 2009). This expression pattern is also observed in chick and zebrafish (Yamada et al. 2000; Ribeiro et al. 2007), and in agreement with mouse studies in zebrafish *Tbx2* is required for the repression of chamber specific genes in non-chamber myocardium (Ribeiro et al. 2007), demonstrating conservation of this later function. In another study in zebrafish, it was noted that knockdown of *Tbx2* resulted in embryos with an underdeveloped or absent ventricle, supporting the hypothesis that *Tbx2* has an earlier fundamental role in FHF myocardium development (Fong et al. 2005). Other evidence for this comes from *Xenopus* where closely related factor *Tbx3* was shown

to encourage mesendoderm development (Weidgang et al. 2013). In this study the investigators note redundancy with Tbx2, as MO mediated knock down of either resulted in delayed gastrulation. Formation of mesoderm and its patterning to form cardiogenic mesoderm are processes that happen in quick succession and that likely overlaps with the process of gastrulation, therefore it is possible Tbx2 has a role to play in these early stages.

Information regarding TBX2 in human cardiac development is very sparse, however, there is one report of two families with TBX2 mutations causing a DiGeorge like syndrome, which is usually associated with mutations in TBX1 (Liu et al. 2018b). A range of cardiac defects are seen in DiGeorge Syndrome, but often the OFT is affected (Lee and van den Veyver 2017). Deletion of TBX1 in the cardiogenic mesoderm of mice results in reduced proliferation of the SHF cells that give rise to the OFT, the mice subsequently presented with OFT defects comparable to those seen in humans (Zhang et al. 2006). Further analysis suggests this is likely due to a reduction in FGF10 signalling downstream of TBX1 which would otherwise induce proliferation of the SHF. *In vitro* TBX1 can bind a conserved T-box element in the *Fgf10* promoter, that can also be activated by TBX5 (Xu et al. 2004), suggesting an overlap in targets for these factors, as could be the case for TBX2 based on its ability to compete with TBX5 for TBE binding sites in other contexts (Habets et al. 2002b). However, it should be noted that other genes are affected in DiGeorge Syndrome, so the observation of a DiGeorge like syndrome in a few individuals carrying TBX2 mutations may be coincidental. Further research is needed to establish the roles of TBX2 in cardiac development so that cases like this may be interpreted properly.

1.6 PRDM1 and cardiac development

1.6.1 The PR/SET and zinc finger containing family of transcriptional regulators

PRDM1 belongs to the PR/SET and Zinc Finger containing family of TFs (Bikoff et al. 2009). There is little conservation between the factors that are related through their PR domain, which resembles the SET domains found in histone methyl transferases (HMTs). The SET domain of an ancestral gene is thought to have merged with multiple ancestral zinc finger genes to create the family which possess variable numbers of Zinc Fingers, from none in PRDM11 to seventeen in PRDM15 (Fumasoni et al. 2007; Vervoort et al. 2016). PRDM1 contains 5 Zinc fingers which enable it to recognize the core sequence GAAAG (Kuo and Calame 2004; Mould et al. 2015). Despite having a SET-like domain most Prdm proteins lack

HMT activity, including PRDM1 which instead exerts its generally repressive effects on gene expression through interactions with a wide variety of chromatin modifying binding partners (Bikoff et al. 2009; see figure 1.8). This variety of binding partners enables it to fulfil multiple roles during embryonic development and as a mediator of adult stem cell pools. For example, it is needed for primordial germ cell specification, plasma cell differentiation, the maturation of enterocytes, and sebaceous gland progenitor pool maintenance (Horsley et al. 2006; Kallies et al. 2007; Bikoff et al. 2009). The potential for PRDM1 to participate in the cardiac GRN has not been fully explored.

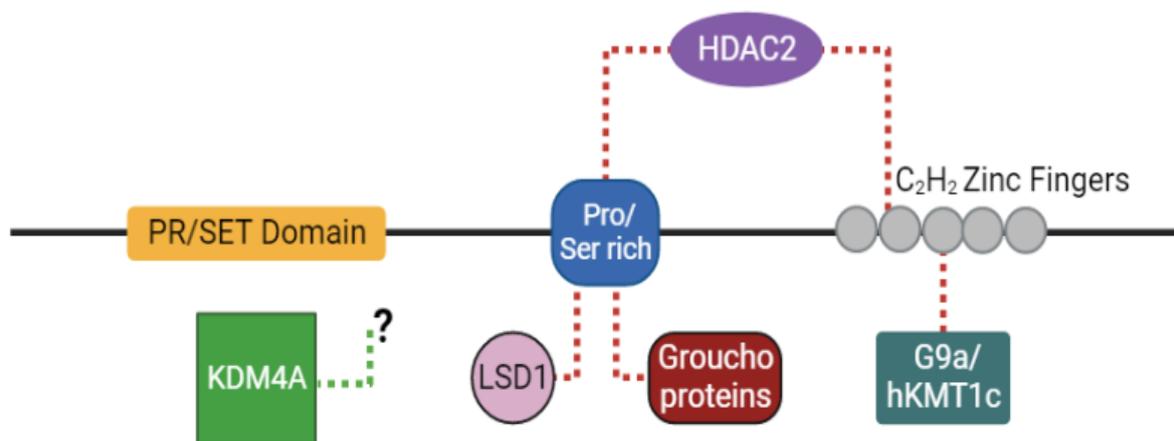


Figure 1.8 A linear depiction of PRDM1 and its interactors

Prdm1 has 3 characteristic domains. A PR/SET domain that lacks histone methyltransferase (HMT) activity, a proline/serine rich domain, and a DNA binding domain (DBD) made up of 5 C₂H₂ zinc fingers. The domains required for direct interaction with repressive chromatin modifiers are connected by red dotted lines. An interaction with a positive chromatin modifier KDM4A is also shown, although the area of PRDM1 that it interacts with is currently unknown. Figure created using BioRender and adapted from Bikoff *et al.* 2009.

1.6.2 Known roles for PRDM1 in cardiac development

A role for PRDM1 in the development of the SHF has been identified. The SHF contributes primarily to the formation of the right ventricle, ventricular septum, and OFT. Lineage tracing studies in mouse embryos show that cells expressing PRDM1 contribute primarily to heart regions derived from the SHF, with a smaller population of cells contributing to the left ventricle that is derived from the FHF (Robertson et al. 2007).

Consistent with this, conditional knockout of PRDM1 in mice causes right ventricle hypoplasia and OFT defects that can be attributed to a reduction in proliferation of SHF progenitors. Targeted deletion of PRDM1 in the MESP1 positive mesoderm that generates the FHF and SHF also results in PTA and VSDs (Vincent et al. 2014). In contrast deletion of PRDM1 using a *Mef2c* driver, which targets late cardiac progenitors, does not lead to severe defects (Vincent et al. 2014). The widespread contribution of PRDM1 expressing cells to the heart, and the severity of the defects seen when it is removed early in development suggest an integral role for PRDM1 in early cardiogenesis.

In other cell types where the role of PRDM1 has been investigated a common theme is PRDM1 acting as a repressor to block the progression of differentiation. For example, in the mouse sebaceous gland and mammary luminal adult stem cells it suppresses the fully differentiated gene program, and cell proliferation, with its down regulation necessary for the formation of a proliferative progenitor pool that will subsequently re-populate the respective ducts (Horsley et al. 2006; Ahmed et al. 2016). It is possible PRDM1 could play a similar role in early cardiac specification, in preventing the expression of alternative gene programs, and stabilising the early decision to acquire a cardiac mesoderm fate.

1.7 Modelling cardiac development with human iPS cells

The data generated in animal models that has been discussed above has provided great insight into cardiac development, however there is a growing appreciation for the discordance between animal models and humans. The study of the human CHD causing G296S mutation as discussed in section 1.4.5 is a good example of these differences and shows how iPSC-CM models can be utilised to generate further insights. Patient derived iPSC CMs have now been used to successfully model and re-capitulate the features of an array of congenital cardiomyopathies *in vitro* including hypertrophic cardiomyopathy (Seeger et al. 2019), dilated cardiomyopathy (Briganti et al. 2020) and long QT syndromes (Liang et al. 2016). These examples support the utility and validity of using iPSCs to model human cardiomyogenesis and disease. However, as within any model system there are some drawbacks to using iPSCs. The pros and cons of the model are explored in further detail below.

1.7.1 Resetting differentiation potential

To understand the limitations of iPSC models it is necessary to consider how an iPSC is formed. Induced pluripotent stem cells are derived from somatic cells, most commonly blood cells or skin fibroblasts. Being somatic cells, they have acquired a restricted differentiated state stabilised by the accumulation and refinement of epigenetic marks during development. The generation of iPSCs requires a reset of the epigenome to a state that is comparable to that seen in the pluripotent cells of the ICM. Inducing cells to undergo this change has been the subject of extensive research for many decades.

The ability to reverse somatic cell fate to totipotency was first demonstrated in 1962 by John Gurdon through the process of somatic cell nuclear transfer (SCNT) (Gurdon 1962), a technique developed by Briggs and King a decade earlier (Briggs and King 1952; King and Briggs 1954). Using the nuclei of intestinal epithelial cells from feeding tadpoles as donor nuclei Gurdon was able to generate fertile adult frogs by transferring them into enucleated *Xenopus* eggs, with a success rate of ~1.5% (Gurdon 1962; Gurdon and Uehlinger 1966). These experiments confirmed that; (1) somatic cells retain the genetic information necessary to form a new organism, (2) that differentiation is reversible, and (3) that *Xenopus* egg cytoplasm contains unknown factors capable of orchestrating this process. It was noted in these experiments that the success rate of SCNT declines as the developmental age of the donor nuclei increases (Gurdon *et al.* 1958; Gurdon 1962; Gurdon *et al.* 1975). A phenomenon also observed by Briggs and King who had previously demonstrated tadpoles could be formed using SCNT of nuclei from blastula to late gastrula stage embryos in *Rana pipiens* (Briggs and King 1952; King and Briggs 1954). These observations combined with the low success rate of SCNT, indicate that although cells can be induced to de-differentiate, they are resistant to doing so and this becomes more difficult as differentiation proceeds.

Subsequently in 2006, Yamanaka *et al.* defined a cocktail of pluripotency associated TFs; OCT4 (POU5F1), SOX2, KLF4 and MYC, that when ectopically expressed are capable of inducing the dedifferentiation of mouse and human somatic cells to a pluripotent state (Takahashi and Yamanaka 2006; Takahashi *et al.* 2007). Similarly, to SCNT the success rate of reprogramming is low (<1%). Reflecting the difficulty in overcoming the epigenetic barriers put in place as differentiation proceeds. Numerous methods for improving reprogramming efficiency have been developed such as, the inclusion of alternative reprogramming factors,

and small molecules often targeting epigenetic modifiers. These methods are reviewed extensively by (Rony *et al.* 2015). Once established iPSCs exhibit many traits associated with pluripotency, such as endogenous expression of pluripotency genes, unlimited self-renewal, the potential to form all 3 germ layers, and contribute to the germline in chimeric mice (Takahashi *et al.* 2007). Thereby confirming that reprogramming has been successful.

However, numerous studies have reported the retention of some somatic cell epigenetic marks after re-programming (Marchetto *et al.* 2009; Kim *et al.* 2011; Lister *et al.* 2011; Ohi *et al.* 2011), as is also the case for SCNT (Ng. and Gurdon 2008). Aberrant changes in epigenetic signature may also be picked up through the re-programming process (Lister *et al.* 2011). This has the potential to introduce a bias in differentiation potential and create variation between lines. Genetic background and how the cells are handled can also introduce variation. Steps can be taken to reduce and account for this variation. For example by conducting tests of pluripotency quality to ensure complete reprogramming and genetic testing to ensure genomic integrity is maintained (Volpato and Webber 2020). In some circumstances this variation between lines may also be beneficial in understanding how genetic background affects the penetrance of disease-causing mutations. The most effective way to account for variation is to conduct experiments using multiple iPSC lines, although this is both costly and time consuming.

1.7.2 Differentiating iPSCs into CMs

Despite these challenges, many protocols exist for highly efficient and reproducible iPSC-CM differentiation (Mummery *et al.* 2012; Lian *et al.* 2013; Burridge *et al.* 2014b) and these protocols have facilitated the studies referenced in section 1.7. Analysis of gene transcription throughout the development of these cells show that they re-capitulate the steps seen in early cardiac development (Burridge *et al.* 2014b; Van Den Berg *et al.* 2015; Churko *et al.* 2018) (see figure 1.9). The CMs formed spontaneously beat and display excitation-contraction coupling dynamics that resemble that of early CMs (Germanguz *et al.* 2011). However, there are some caveats to this, it is important to note that the CMs formed *in vitro* from iPSC's are phenotypically closer to embryonic CMs than adult CMs (Van Den Berg *et al.* 2015). This means this model of CM development is limited in some respects, and likely does not capture the process of maturation that takes place during the transition from embryonic to postnatal and adult life.

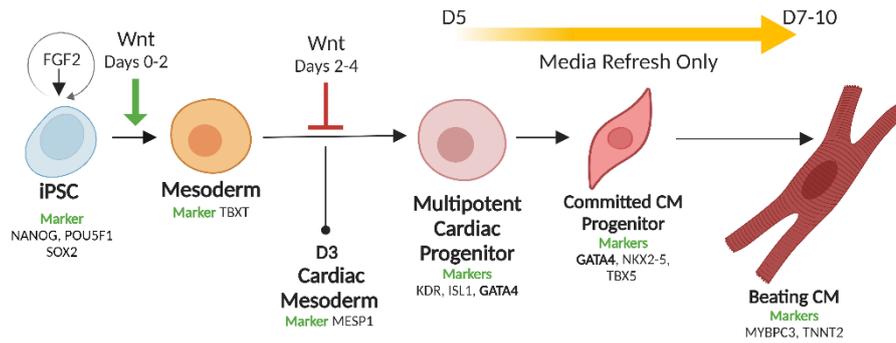


Figure 1.9 An overview of the milestones of iPSC to CM differentiation

Using bi-phasic Wnt modulation protocols such as those devised by Burridge *et al.* 2014 and Lian *et al.* 2012 it is possible to produce beating CMs from iPSCs within 7 days. The progress of differentiation can be followed by analysing the expression of known marker genes for each stage, as shown above. This sequence of gene expression changes resembles what is seen during *in vivo* cardiomyogenesis. Figure created using BioRender and adapted from Burridge *et al.* 2014.

1.7.2 Multiple factors contribute to the immature phenotype of iPSC-CMs

Some of the signs of immaturity observed in iPSC-CMs include a flattened morphology and reduced sarcomeric organisation, these features are also apparent in cultured adult primary CMs which undergo partial dedifferentiation when they are cultured *in vitro* (Bird *et al.* 2003). Therefore, the lack of maturity seen in iPSC-CMs is thought to be due to the absence of several factors and cues that are normally present *in vivo*. These factors include (1) **Time**; *in vitro* beating CMs can be formed within 7 days. Limited information from human heart samples, suggests the process of CM maturation continues for years after birth. For example, it takes 6 years for gap junctions to become localised to intercalated disks, an adult CM phenotype (Angst *et al.* 1997). (2) **Physical cues**; at birth there is a sharp increase in cardiac load, CMs respond to an increase in pressure and stretch by increasing their contractile force through sarcomeric remodelling (Konhilas *et al.* 2002; Yang *et al.* 2016; Guo and Pu 2020). (3) **Metabolic milieu**; the availability of metabolites changes significantly after birth. The primary method of energy production switches from glycolysis in neonates, to oxidative phosphorylation after birth, correlating with a 40-fold increase in fatty acid availability which provides an energy rich substrate to meet increasing CM output (Makinde *et al.* 1998a). (4) **Hormones and Signalling**; There is also an increase in

exposure to hormones that encourage maturation such as triiodothyronine (T₃) and cortisol (Makinde et al. 1998a). (5) **Cell-cell interactions**; iPSC-CM differentiation protocols have been designed with the aim of producing relatively pure populations of CMs (≥98% with purification) (Burrige et al. 2014b). *In vivo* CMs interact with various cell types, such as fibroblasts, smooth muscle cells, endocardial cells, and immune cells (Litviňuková et al. 2020). These other cell types can have a significant effect on CM phenotype. For example, fibroblasts have been demonstrated to effect Ca signalling in CMs, and to participate in reciprocal paracrine signalling that contributes to CM hypertrophy (Cartledge et al. 2015).

All of the factors listed above contribute to the phenotypic state of CMs in the heart, and the development of methods to improve maturation status based on these is an area of active research. Some of the approaches used are summarized in figure 1.10, but none have yet delivered CMs that score well in all aspects of maturation. Despite these limitations iPSC-CM differentiation can recapitulate large aspects of early cardiomyogenesis and thus provide a viable and useful platform for studying this process.

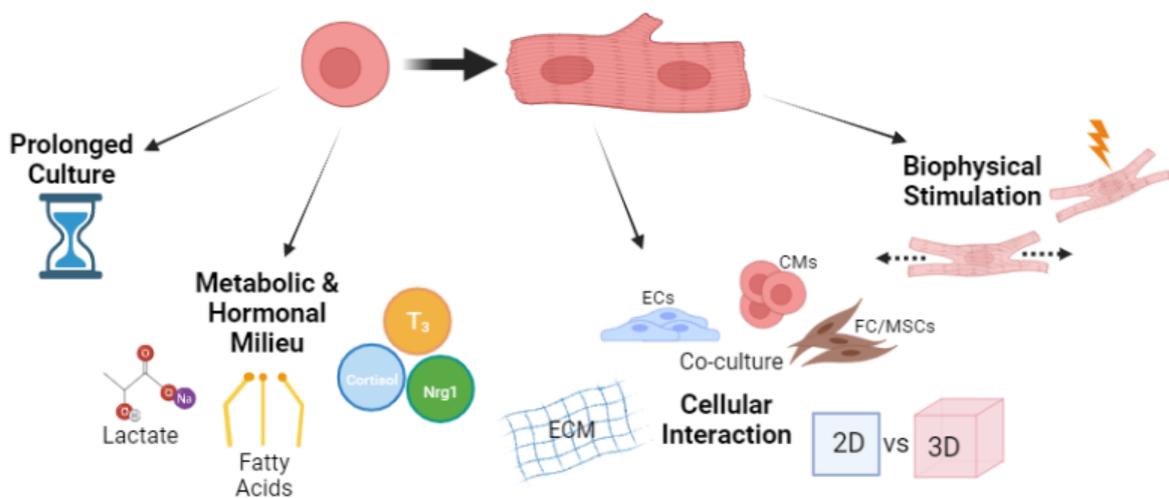


Figure 1.10 A summary of possible maturation techniques

The techniques displayed above all aim to create a more enriched *in vitro* environment that more closely recapitulates the elements that a cardiomyocyte is exposed to *in vivo*. The papers referenced here provide examples of the expected outcome of using these techniques - (Kolker et al. 2012; Yang et al. 2014; Jung et al. 2016; Y. et al. 2017; Correia et al. 2018; Dias et al. 2018; Yang et al. 2019; Feyen et al. 2020; Huang et al. 2020; Ni et al. 2021).

1.7.3 CRISPR-Cas9 gene editing as a tool for studying cardiac development

The arrival of technologies such as the CRISPR-Cas9 gene editing system have made genome editing readily accessible. Although there are some restrictions in the sites that can be chosen, the range of CRISPR-Cas9 systems now available makes it possible to induce almost any mutation. For a current and comprehensive review of the technologies available, please see (Jianli Tao *et al.* 2023; Li *et al.* 2023). The CRISPR-Cas9 system is originally a form of adaptive immunity in bacteria and archaea, the system has been co-opted for use in other organisms including human cells. The first hints of the existence of the CRISPR Cas system came in 1987 by the lab of Ishino, with the observance of repetitive elements at the 3' end of *iap*, an unrelated gene being investigated in *Escherichia coli* (Ishino *et al.* 1987). A comparative search of these sequences between species was conducted and revealed them to be widespread in bacteria and archaea, suggesting they may have a functional role (Mojica *et al.* 2000).

Following this steady progress was made to elucidate the role of these elements in bacteria and archaea as a form of adaptive immune response against bacteriophages and viruses (Makarova *et al.* 2006; Barrangou *et al.* 2007). In bacteria this system is made up of a CRISPR array – a section of the genome containing a cluster of repetitive sequences interspaced with short stretches of non-repetitive sequences referred to as spacers, these non-repetitive sequences originate from contact with invading pathogens. Preceding the array of spacers are trans-activating RNA (tracrRNA) sequences which contain homologous regions that match the repetitive sequences surrounding the spacers. Therefore, when the spacers are transcribed, the tracrRNA can hybridise with these regions to form a guide RNA (gRNA). Cleavage at the 5' end of the spacer reduces its length to 20 nt's which is the length that is now commonly used for artificial crRNA design. The tracrRNA component of the guide hybridises with a Cas nuclease, such as Cas9 to form a ribonucleoprotein complex. This enables the gRNA to guide the nuclease to cleave any invading RNA/DNA with a complimentary sequence, leading to the destruction of invading pathogens (Barrangou and Doudna 2016) (see figure 1.11).

Since the elucidation of the mechanism, it has been demonstrated that components of the system can be used in heterologous systems (Sapranaukas *et al.* 2011). It was then demonstrated that the user may design their own guide sequence to induce cleavage of a

desired target (Gasiunas et al. 2012; Jinek et al. 2012). CRISPR-Cas9 gene editing is now routinely used as a research tool to study diverse biological questions in many different models. Due to the high efficiency and accuracy of editing and the ability to predict off-target effects CRISPR-Cas9 has also been used successfully in clinical trials to treat diseases such as β thalassemia (Frangoul et al. 2021) and Duchenne Muscular Dystrophy (Min et al. 2019). For the project presented herein, this system provides a reliable and amenable system to target the genes discussed previously (*GATA4*, *GATA6*, *TBX2*, and *PRDM1*). A two-guide strategy will be used to introduce two double strand breaks into the genes listed, with the intention of creating frameshifting knockout mutations. Once generated the lines will then be used to investigate the roles of these genes in cardiogenesis through the use of established iPSC-CM differentiation protocols.

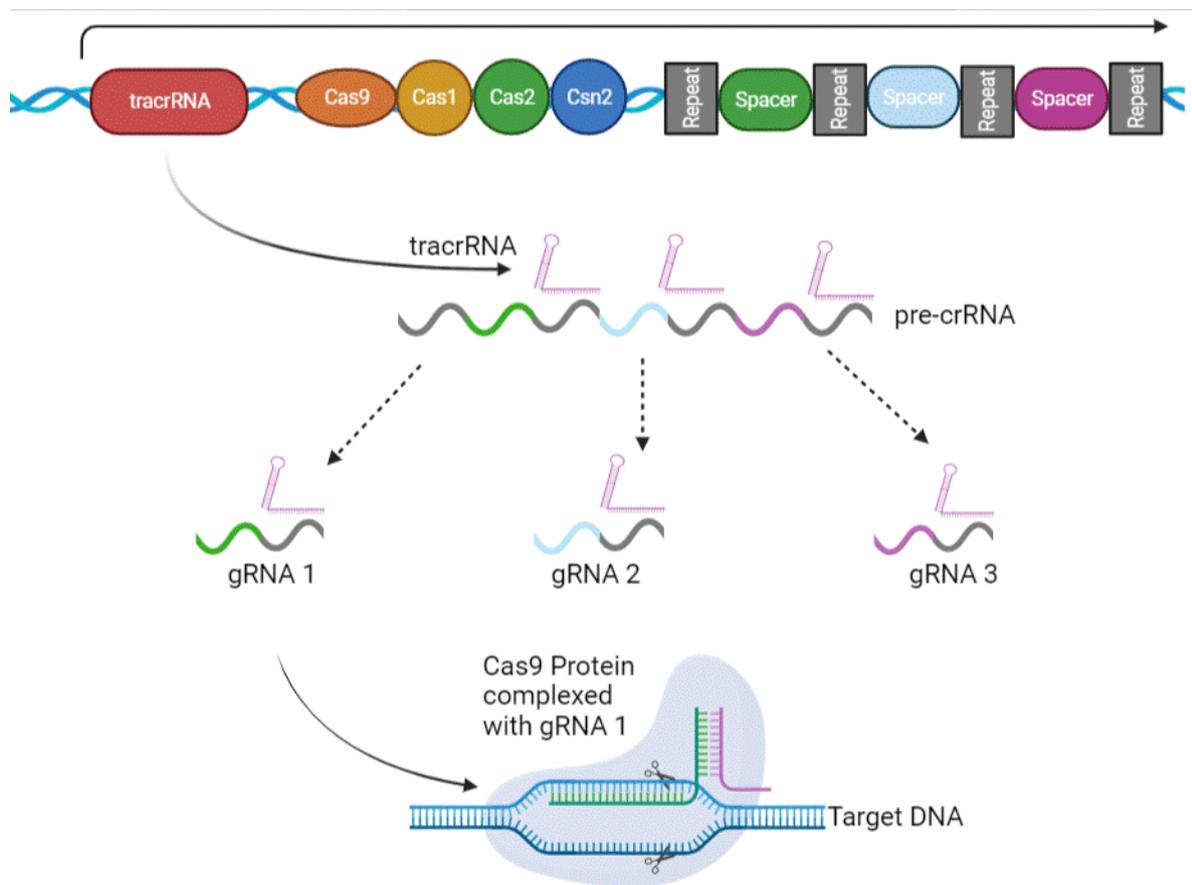


Figure 1.11 An overview of the CRISPR-Cas9 system from *Streptococcus pyogenes*

In bacteria such as *S. pyogenes* the CRISPR-Cas9 system consists of the CRISPR locus, containing a *tracrRNA*, and the Cas enzymatic genes including Cas9 that the *tracrRNA* will hybridise with. There are also ‘spacer’ elements derived from exposure to invading pathogens. These are interspersed with repetitive elements that will be transcribed with the spacers to form a crRNA. The repetitive regions of the crRNA are complimentary to the *tracrRNA* region, and thus can bind to them to form a gRNA that will guide the Cas9 enzyme to its targets. In artificial systems the gRNA is designed by the user. This figure was created using BioRender and is adapted from a figure published by Jiang and Doudna 2017.

1.8 Project Aims & Hypotheses

There is a significant amount of evidence available about the role GATA4 plays in heart development and homeostasis in model organisms, but how well these findings relate to human cardiac development is unclear. This thesis aims to address this through the creation of GATA4 null human iPSC lines using CRISPR-Cas9 gene editing. The cardiogenic potential of these cells will then be assessed using established Wnt modulation-based iPSC-CM differentiation protocols (Burrige et al. 2014b; Burrige et al. 2015), and the phenotype of these cells will be explored using a mixture of imaging and molecular laboratory techniques. Furthermore, as a key member of the cardiac GRN, that in certain contexts is sufficient to drive the process of cardiogenesis, the genes that GATA4 regulates are also of interest as they are also potential mediators of heart development. Thus, the second aim of this thesis is to determine if regulation of *TBX2* and *PRDM1* by GATA4 is conserved in this model.

Following this, the functionality of the target genes themselves during cardiac development will be assessed. For *TBX2*, a positive target of Gata4 regulation in *Xenopus*, the goal is to generate null lines using CRISPR-Cas9 gene editing once again. These lines will be used to determine whether *TBX2* expression is necessary for normal iPSC-CM differentiation. In the case of *PRDM1*, null lines will also be created, to address if it is required at the earliest stages of cardiac development. In addition, during this project the intention is to create a *PRDM1* gain of function line to enable *PRDM1* to be forcibly expressed outside of its usual time window. This will enable the requirement for GATA4 mediated repression of *PRDM1* in cardiac development to be assessed. A summary of these aims is shown in figure 1.12.

The final aim of this thesis was initially to create GATA6 null lines and compound heterozygous GATA4 and GATA6 null lines. To assess the role of GATA6 in iPSC-CM development and to examine the potential for redundancy between these closely related factors. Since the commencement of the project a GATA6 null human iPSC cell line has been reported in the context of cardiac development. However, the line reported herein contains a deletion in a predicted transcriptional activation domain of GATA6 that helps to inform what regions of the protein are required for its function in normal CMs.

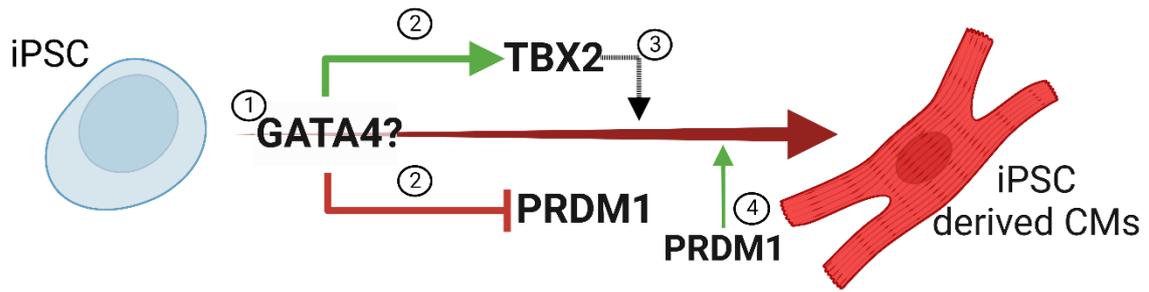


Figure 1.12 A summary of the questions this thesis aims to address

(1) Is GATA4 required for iPSC-CM differentiation and what are its roles during this process? (2) Does GATA4 induce TBX2 and repress PRDM1 expression in this model? (3) Is TBX2 expression required for normal cardiac development? (4) Is repression of PRDM1 required for iPSC-CM differentiation? Another aim to be addressed that is not depicted is to examine whether PRDM1 is required at the earliest stages of heart development during cardiac mesoderm specification which is when PRDM1 is normally expressed. The final aim to be investigated, that has not been depicted in the figure, is to examine what regions of GATA6 are required for normal cardiomyogenesis.

Chapter 2 - Materials and methods

Unless otherwise stated all chemicals were obtained from Merck, and all cell culture reagents and materials from ThermoFisher or FisherScientific.

2.1 Induced Pluripotent Stem Cell (iPSC) maintenance

2.1.1 iPSC culture

Rebl Pat iPSCs were kindly given to the lab by Professor Chris Denning, Nottingham University. iPSCs were maintained in Corning™ T25 cell culture treated flasks coated with 5 µg/mL recombinant human vitronectin (VTN-N) in Dulbecco's phosphate-buffered saline without calcium and magnesium (DPBS^{-Ca-Mg}) for 1 hour at room temperature (RT) or with Geltrex™ supplied at 12-18 mg/mL and then diluted 1:400 in DMEM/F12 Media, to give a concentration of between 30-45 µg/mL for coating at 37°C for 1 hour. iPSCs were fed with Essential 8™ Flex Media (E8F) or B8 media (DMEM F-12 (cat no. 1320-033), Insulin 20 µg/mL, ascorbic acid-2-phosphate 200 µg/mL, Optiferrin 20 µg/mL (BioOrbyt), NaSe 20 ng/mL, FGF2-G3 50 ng/mL (NorthWestern University), TGF-β1 1 ng/mL (Peprotech), HEPES pH7.4 15 mM, EmbryoMax Nucleosides 0.5x (Merck), and incubated at 37°C in a humidified atmosphere with 5% CO₂.

2.1.2 iPSC passaging

When 70-85% confluence was reached, the cells were passaged. To achieve this the cells were dissociated by first washing with DPBS^{-Ca-Mg}. The cells were then incubated with ReLeSR™ for 1 minute at RT, the solution was then removed, leaving only a thin layer of ReLeSR™, then incubated for a further 6 minutes at RT before adding DPBS^{-Ca-Mg} to dislodge colonies. The cell suspension generated was then triturated to generate smaller cell aggregates. For regular passaging 1:8-1:12 of the cell suspension is then placed into a coated T25 flask with E8F. If the cell aggregates generated during trituration were smaller than expected 2.5-5 µM Y27632 (HelloBio) was added to the E8F media for the first 24 hours following passage.

2.2 Gene editing with CRISPR-Cas9

2.2.1 Preparing cells for nucleofection

Prior to nucleofection cells were sub-cultured at a higher seeding density of approximately 1:3-1:5 following passaging, with the colonies then reaching 70-80%

confluence within 2-3 days. A minimum of 30 minutes before nucleofection E8F media was replaced with fresh E8F 10 μ M Y-27632. All cell lines in culture and to be nucleofected were tested regularly to confirm the absence of mycoplasma infection as this is known to negatively impact gene editing outcomes, the method for this is described below.

2.2.2 PCR based detection of mycoplasma

All primer sequences used are shown in table 2.1. PCR reactions were conducted in a volume of 25 μ l, to each was added 0.1 μ M of forward and reverse primer mix (table 1) (Eurofins, Wolverhampton, UK), 1X MyTaq red reaction buffer, 0.5U KAPA Taq Polymerase (both KAPA Biosystems), and 2 μ L of cell culture supernatant. A sample of cell culture supernatant was also subjected to a harsher lysis treatment prior to PCR amplification, to ensure the detection of any intracellular mycoplasma. This treatment was conducted as follows; to 80 μ L of cell culture supernatant 1 μ L of 19 mg/mL Proteinase K (Roche) was added, and ~15-20 μ L of 50% Chelex[®] 100 sodium solution (Merck). The samples were then incubated for 90-120 minutes at 55°C and then at 95°C for 5 minutes. 2 μ l of this treated cell culture supernatant was added to the PCR reaction. An MJ Mini Thermal Cycler Machine (Biorad) was used for cycling, and the cycling conditions are shown in table 2.2.

Table 2.1 Forward and reverse primer sequences used for mycoplasma detection.

Forward primers	
Myco-5-1	CGCCTGAGTAGTACGTTCGC
Myco-5-2	CGCCTGAGTAGTACGTACGC
Myco-5-2	TGCCTGAGTAGTACATTCGC
Myco-5-2	TGCCTGGGTAGTACATTCGC
Myco-5-5	CGCCTGGGTAGTACATTCGC
Myco-5-6	CGCCTGAGTAGTATGCTCGC
Reverse Primers	
Myco-3-1	GCGGTGTGTACAAGACCCGA
Myco-3-2	GCGGTGTGTACAAAACCCGA
Myco-3-3	GCGGTGTGTACAAAACCCGA

Table 2.2 Cycling conditions for amplification of mycoplasma DNA.

Step	Temp (°C)	Time (min)
Initial denaturation	95	2
Low Specificity – repeat for 5 cycles	94	0.5
	50	0.5
	72	0.5
High Specificity, repeat for 30 cycles	94	0.25
	56	0.25
	72	0.5
Store	4	

2.2.3 crRNA design

crRNAs were designed using the DeskGen Cloud tool which was hosted and supported by DeskTop Genetics. The tool provided an on-target activity score based on work by (Doench et al. 2016), and an off-target score for specificity based on work by Click or tap here to enter text.(Hsu et al. 2013). Guides were selected based on these scores to maximize activity and minimize the risk of off-target effects. As there are other tools available and there is little consensus between them, guides were also selected using the CCTop tool designed by (Stemmer et al. 2015). Two pairs of guides were selected for each gene. Each pair was intended to be used together to induce a deletion of a number of nucleotides that is not divisible by three between each double strand break (DSB), to cause a frameshift mutation.

Table 2.3. crRNA sequences used for CRISPR Cas9 gene editing

Gene	Gene Location	PAM	Protospacer Sequence	Distance Between Cut Sites (bp)
<i>GATA6</i>	Chr18 22,168,942 - 22,203,028	AGG	GACGCCTCAGCTCGACACGG	109
<i>GATA6</i>	Chr18 22,168,942 - 22,203,028	GGG	AGCTCGACAGGTTGCCCCCG	
<i>TBX2</i>	Chr17 61,399,395 - 61,409,966	CGG	GCTCGCACTATGTGGAAGCG	71
<i>TBX2</i>	Chr17 61,399,395 - 61,409,966	GGG	CGGCGATGAAGTCGGTCTCC	
<i>PRDM1</i>	Chr 6 106,085,819 - 106,110,439	TGG	CTCTCCCCGGGAGCAAAACC	121
<i>PRDM1</i>	Chr 6 106,085,819 - 106,110,439	CGG	AGGCTTCACTACCCTTATCC	
<i>PRDM1</i>	Chr 6 106,085,819 - 106,110,439	AGG	ATTGTCAGCTCTCCGGGATA	196
<i>PRDM1</i>	Chr 6 106,085,819 - 106,110,439	TGG	CATTAAAGCCGTCAATGAAG	
<i>TBX2</i>	Chr17 61,399,395 - 61,409,966	CGG	GCTCGCACTATGTGGAAGCG	61
<i>TBX2</i>	Chr17 61,399,395 - 61,409,966	AGG	GTCGGTCTCCGGGAACACGT	
<i>GATA4</i>	Chr8 11707781 - 11709011	AGG	TGTGGGCACGTAGACTGGCG	94
<i>GATA4</i>	Chr8 11707781 - 11709011	CGG	GCGTCCGGAGGCGCCTCGGG	
<i>GATA6</i>	Chr18 22,168,942 - 22,203,028	GGG	CCCCACGTAGGGCGAGTAG	82
<i>GATA6</i>	Chr18 22,168,942 - 22,203,028	GGG	GGCCGACAGCGAGCTGTACT	
<i>GATA4</i>	Chr8 11707781 - 11709011	GGG	CGGCCATGTAAGCCGGGTAG	68
<i>GATA4</i>	Chr8 11707781 - 11709011	GGG	TGCAGGACCGGGCTGTGCGAA	

2.2.4 Formation of guide RNAs

Guide RNAs for CRISPR-Cas9 experiments were purchased as crRNAs (Integrated DNA Technologies, California, USA) containing the 20 nt target specific protospacer region, and a 16 nt region that hybridizes with the 5'-550 ATTO tracrRNA (Integrated DNA Technologies). The tracrRNA carries a fluorescent tag (5' 550-ATTO) and another region that allows hybridization with the Cas9 enzyme. Each crRNAs was combined with a tracrRNA in equimolar concentrations in IDTE Nuclease free buffer (30 mM HEPES (pH 7.5), 100 mM potassium acetate). Then heated to 95°C for 2 minutes, and gradually cooling to RT.

2.2.5 Ribonucleoprotein complex (RNP) formation

Alt-R S.p. HiFi Cas9 Nuclease V3 was diluted in Cas9 storage buffer (10 mM Tris-HCl (pH 7.4), 300 mM NaCl, 0.1 mM EDTA, 1 mM DTT) to a concentration of 6.2 µg/µL. Equal amounts of each sgRNA complex and the Cas9 enzyme were then mixed, to give 10 µL of RNP complex, and incubated together for 10 minutes at RT.

2.2.6 Nucleofection

When the sub-culture of iPSCs had reached 70-80% confluence they were dissociated. Firstly, the cell layer was washed with 5 mL of DPBS^{-Ca-Mg+} for five minutes at RT twice. Then incubated with 5 mL of StemPro Accutase (ThermoFisher) for 5-10 minutes at 37°C until a single cell suspension was formed. The cells were passed through a 20 µm cell strainer (ThermoFisher), then centrifuged at 100g for 5 minutes. The supernatant was removed and 1×10^6 cells were re-suspended in 100 µL of P3 Primary Cell Buffer per nucleofection (Lonza). To this 10 µL of RNP complex was added, and the suspension was moved to a nucleocuvette (Lonza). Electroporation was carried out using a Lonza 4D Nucleofector with the CA137 program. Post-nucleofection the cells were allowed to recover for 10 minutes at 37°C and then recovered to 1-well of a 6-well plate coated with Geltrex and fed with E8F supplemented with 1x RevitaCell Supplement (ThermoFisher) and 100 U/mL penicillin and 100 µg/mL streptomycin (PenStrep).

2.2.7 Selection of RNP positive cells via fluorescence activated cell sorting (FACS)

The cells were allowed to recover for 24 hours and then dissociated using Accutase as before. The top 15% brightest cells were selected using the BD FACS Aria™ Fusion (BD Biosciences), by Mark Bishop of the European Cancer Stem Cell Institute (ESCRI, Cardiff University)), and sorted as single cells into a 96-well plate coated with Geltrex, and a 10 cm²

dish at a density of 50 cells/cm², and maintained in E8F 10 μM Y27632 until cells showed attachment to the substrate and colonies started to form. Colonies were then picked for DNA extraction and genotyping, successfully edited colonies were cultured, expanded and frozen.

2.2.8 PCR based screening of cells

Cells were grown for approximately 10-14 days to allow colony formation, ~75% of the colony was removed for genomic DNA (gDNA) extraction and the remaining ~25% was replica plated. gDNA was extracted from colonies by lysis in 80 μL of 0.19mg/mL Proteinase K (Roche) in PBS, to this was added 15 μL of 50% Chelex[®] 100 sodium solution (Merck). The samples were then incubated for 90-120 minutes at 55°C and then at 95°C for 5 minutes. Samples were subject to 35-40c of PCR amplification using genotyping primers shown in appendix table S3, reactions included 5% dimethyl sulphoxide (DMSO). The resulting products were analysed by agarose gel electrophoresis, and promising amplicons sent for sequencing by Eurofins Genomics. For this, the fragments were purified using a PCR Purification Kit (Qiagen). The concentration of purified PCR product was measured using a Nanodrop™, and samples were then diluted to 5 ng/mL in ddH₂O, with 2μl of 1mM Forward or Reverse primer and sent for sequencing. The results were analysed using Synthego ICE (Hsiao et al. 2018).

2.3 Quality checking edited lines

2.3.1 Assessing pluripotency

Pluripotency was assessed by confirming expression of pluripotency genes Nanog, Pou5f1 and Sox2 using a mixture of methods including; RT-PCR, flow cytometry, and Immunofluorescence. These methods are described in further detail in section 2.5.3.

2.3.2 Karyotyping

Genomic DNA was extracted from >2 million iPSCs using the Blood and Tissue DNA Extraction Kit (Qiagen), following the manufacturers guidance. This gDNA was then karyotyped using the Infinium Global Screening Array-24, covering 654,027 markers across the genome. Microarray processing and analysis was handled by Alexandra Evans at the MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University.

2.4 Cardiac Differentiation

WT and mutant iPSCs were differentiated using one of two commonly used iPSC-CM differentiation protocols, both are based on Wnt modulation. These are described below and summarized in figure 1.1.

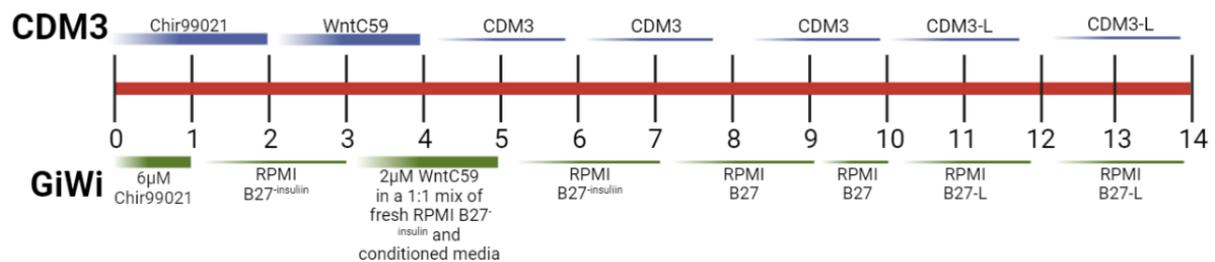


Figure 2.1 A visual summary of the CDM3 and GiWi protocols published by Burrige *et al.* 2014, and Lian *et al.* 2012.

These protocols deviate slightly from the published protocols through the use of an alternative Wnt inhibitor; WntC59.

2.4.1 Chemically defined Wnt Modulation (CDM3) Protocol

The first is a protocol designed by Paul Burrige that has been shown to work with >200 iPSC lines (Burrige et al. 2015). This protocol uses a simple chemically defined media made up of 3 components, henceforth referred to as CDM3. The success of cardiac differentiation is highly dependent on seeding density iPSC to CM differentiation was only carried out when iPSCs had shown a steady growth rate over a minimum of four passages and therefore would grow predictably. For differentiation cells were seeded at a density of 14×10^4 - 17.5×10^4 cells per cm^2 , in wells coated with 1:400 Geltrex and expanded to 75-80% confluency in E8F medium before differentiation was started, taking approximately 4 days. To initiate differentiation cells were changed to CDM3 (RPMI 1640, 500 $\mu\text{g}/\text{mL}$, 213 $\mu\text{g}/\text{mL}$ L-Ascorbic acid-2-phosphate) with 6 μM CHIR99021 (HelloBio) for 48 hours. Following this media was changed to CDM3 with Wnt Inhibitors; either 2 μM WntC59, or 5 μM KY0211 and 5 μM XAV939 (HelloBio) for 48 hours. After these treatment stages cells were maintained in CDM3 media alone and changed every 2 days until beating was noted. All differentiation media also contained Penicillin – Streptomycin 50 U/mL (ThermoFisher), and was pre-warmed to 37°C.

2.4.2 GiWi Differentiation Protocol

This protocol has been adapted from (Lian et al. 2013). For differentiation iPS cells were seeded at a density of 14×10^4 - 17.5×10^4 cells per cm^2 , in wells coated with Geltrex (5 $\mu\text{g}/\text{ml}$) and expanded to 75-80% confluency in E8F medium before differentiation was started, taking approximately 4 days. To start differentiation the cell culture media was changed to RPMI B27^{-insulin} (ThermoFisher) with 6 μM CHIR99021 for 24 hours. The media was then removed and replaced with RPMI B27^{-insulin} for 24 hours. Following this the spent media was removed and mixed 1:1 with fresh RPMI B27^{-insulin} to make conditioned media. To the conditioned media was added 2 μM WntC59, and cells were maintained in this for a further 48 hours. After this the media was refreshed every 2 days with RPMI B27^{-insulin}, and then with RPMI B27 on day 7.

2.4.3 Metabolic Selection and Maintenance

Following differentiation using the CDM3 protocols described above cells were switched to CDM3-L (RPMI 1640 no glucose, 500 $\mu\text{g}/\text{ml}$ Fraction V BSA, 213 $\mu\text{g}/\text{ml}$ L-Ascorbic acid-2-phosphate, 5 mM Sodium DL-Lactate). Media was changed every two days for a total of four days. For cells differentiated using the GiWi protocol RPMI B27-L (RPMI 1640 no glucose, 1x B27, 5 mM Sodium DL-Lactate) was used instead. Once again, the media was changed every two days for a total of 4 days. Upon finishing selection CMs derived from either protocol were then maintained in CDM3 with media changes every 2-3 days.

2.4.5 Dissociation of young (day 7-10) CMs

The cell layer was washed with DPBS^{-Ca-Mg} for 5 minutes. This solution was then removed, and TrpLe Express was added to the cell layer. The cells were incubated in this for 5-10 minutes at 37°C. The TrpLE was then neutralized by the addition of an equal amount of serum containing media. The dissociated cells in solution were then gently triturated 5-10 times using a 5 mL stripette and filtered through a 100 μm Cell Strainer (ThermoFisher) to ensure a single cell suspension was formed. The solution was centrifuged at 400 rcf for 5 minutes, and the supernatant was removed, the cell pellet was resuspended in CDM3 with 10% Heat Inactivated Foetal Bovine Serum (HI-FBS) to give $\sim 1 \times 10^6$ cells/mL for replating.

2.4.6 Dissociation of older (>10 days) CMs

For older CM cultures a longer dissociation was necessary. Firstly, the cell layer was washed with Hanks Buffered Saline Solution without Ca and Mg (HBSS, Merck). This was

then removed and pre-warmed dissociation solution (HBSS, 200 U/mL Collagen Type II (Worthington Biotech), 1 mM HEPES pH 7.4, 10 μ M Y37632 (HelloBio)) was added and the cells incubated at 37 °C for 3-3.5 hours. The cell solution was then triturated, filtered, and resuspended as above, ready for re-plating.

2.4.6 Heart forming organoid (HFO) differentiation

iPSC's were seeded at a density of 10,000 cells per well in an ultra-low adhesion round bottomed 96-well plates (VWR) in E8F with 5 μ M Y27632. The plate was then centrifuged at 400 rcf for 3 minutes to bring the iPSCs to the bottom of the well, encouraging aggregation, with the aim of producing just one aggregate per well. The media was refreshed with E8F the following day. 24 hours after this the media was removed and 20 μ l of undiluted geltrex ~15mg/mL was placed on top of the aggregates and allowed to solidify at 37 °C for 15 minutes. Fresh E8F was then placed on top. After this process cells were differentiated as per the GiWi or CDM3 protocols described in section 2.4.1 and 2.4.2. Successful differentiation into heart forming organoids (HFO's) was determined by the presence of beating before day 15 of differentiation. Metabolic selection and maturation were not carried out for these experiments.

2.4.7 Dissociation of HFO's

HFO's were pooled together into groups of 5, this was to ensure there was a visible cell pellet in later steps. To begin dissociation, they were incubated in 1 mL of dissociation solution (HBSS, 200U/mL Collagen Type II (Worthington Biotech), 1 mM HEPES pH 7.4, 10 μ M Y37632 (HelloBio)) for - hour at 37°C. An equal amount of culture media was added to dilute this solution and then cells were centrifuged at 400 rcf for 5 minutes. They were then resuspended in TrpLE Express for 5 minutes at RT, then gently triturated no more than five times to break the aggregates into small clumps. They were then incubated for a further 5 minutes, to generate a single cell suspension. An excess of culture media was added to dilute the TrpLE and then cells were spun down at 400 rcf for 3 minutes. The cell pellet was re-suspended in CDM3 with 10% HI-FBS. The equivalent of 1 aggregate was plated onto a 13 mm coverslip for IF and allowed to attach. The media was changed a day later, and a minimum of 48 hours recovery time was allowed before analysing any re-plated cells.

2.5 Phenotypic analysis of iPSCs and CMs

2.5.1 Morphology and Beating

The cells were monitored continuously throughout their differentiation for morphological changes and for beating, indicating the presence of CMs. Routine brightfield pictures and videos of cells in culture were taken using the Leica Dmi1 microscope, and MC170HD camera at 30 fps. An Olympus IX71 motorised inverted microscope, fitted with a Hamamatsu ORCA Flash 4 CMOS Camera for imaging at up to 100 fps was used for beating analysis.

2.5.2 Flow Cytometry for Analysis of CM Markers

On day 10-12 the cells were assessed for their expression of cardiac markers. Cells were dissociated as described in section 2.4.6. The cells were dissociated using collagenase type II in PBS without Ca and Mg for 1.5 hours, this was then removed from the cell layer, and trypsin-EDTA 0.25% was applied to the cell layer for no more than ten minutes, cells were gently triturated to form a single cell solution, and then sieved through a 100 μ M cell sieve (FisherScientific), the trypsin solution was then neutralized using an equal amount of Neutralisation Media (DMEM/F12 media 10% FBS 2 mM Ascorbic Acid). Cells were spun down at 100g for 3 minutes, and the supernatant removed. The cells were then fixed using 4% paraformaldehyde for ten minutes at RT, 10 mL of PBS without Ca²⁺ Mg²⁺ was added to dilute the PFA, and the solution was then spun down 300g for 5 minutes. To permeabilise the cells they were treated with 0.1% Triton X-100 at RT for 10 minutes, then spun down at 300g for 5 minutes. The cells were then resuspended in 1 mL of HI-FBS, which was then diluted with approximately 14ml of PBS without Ca²⁺ Mg²⁺. The cells were re-suspended in FACS Buffer (PBS without Ca²⁺ Mg²⁺ 2% HI-FBS). Then split into appropriate aliquots for staining and spun down at 500g for 5 minutes. They were re-suspended in 100 μ L of FACS Buffer containing anti-Ctnt-TNNT2 (REA400, Miltenyi) or isotype control-fitc (REA293, Miltenyi) as per the manufacturer's instructions, also table 4. The cells were incubated on ice in darkness for 30 minutes, with agitation. They were then spun down at 300g for 5 minutes and re-suspended in FACS Buffer to wash, then passed through a 100 μ M cell sieve, spun down once more, and resuspended in 200-300 μ L of FACS buffer ready for analysis. Cells were analysed using the BD LSR Fortessa (provided by the European Cancer Stem Cell

Research Institute (ESCRI), Cardiff University, and maintained by Mark Bishop, ESCRI, Cardiff University).

2.5.3 Flow Cytometry for Pluripotency Marker POU5F1

Staining for pluripotency markers was carried out as above with some alterations. iPSCs were dissociated by washing the cell layer in DPBS^{-Ca-Mg} for 5 minutes. This was then removed and followed by incubation of the cells with TrpLE for 5-10 minutes at 37°C. The TrpLE was then diluted using an excess of DPBS^{-Ca-Mg}. The dissociated cells in solution were then gently triturated 5-10 times using a 5 mL stripette and filtered through a 40 µm Cell Strainer (ThermoFisher) to ensure a single cell suspension was formed. The solution was centrifuged at 400 rcf for 5 minutes, and the supernatant was removed, fixation, permeabilization, and staining were carried out as above. The primary antibodies used for iPSC staining were Rabbit polyclonal IgG anti-OCT4/POU5F1 (ab19857), anti-SOX2 (ab97959), and Isotype Control ab171870 detailed in table 4. All are unconjugated, therefore secondary antibody incubation was needed. For this after processing with primary antibodies as above, the cells were incubated with Goat anti-rabbit – 488 (A-11008, ThermoFisher) in FACS buffer for a further 30 minutes, then spun down at 400 rcf for 3 minutes and resuspended in 1 mL of FACS buffer to wash. They were then once again spun down and resuspended in 200-300 µL ready for analysis.

Table 2.4 Antibodies used for Flow Cytometry

ID/Catalogue Number	Manufacturer	Target Protein	Host	Clonality	Isotype	Permeabilisation	Dilution
303112	BioLegend	Pecam1 (CD31)	Mouse	monoclonal	IgG1, k	0.1% Triton-X-100	1 µL/1x10 ⁶ cells
400130	BioLegend	Isotype Ctrl	Mouse	monoclonal	IgG1, k	0.1% Triton-X-100	1 µL/1x10 ⁶ cells
REA400	Miltenyi	Ctnt-fitc	Recombinant	monoclonal	IgG1	0.1% Triton-X-100	1 µL/1x10 ⁶ cells
REA293	Miltenyi	Keyhole Limpet Heamocyanin	Recombinant	monoclonal	IgG1	0.1% Triton-X-100	1 µL/1x10 ⁶ cells
ab171870	Abcam	Isotype Ctrl	Rabbit	polyclonal	IgG	0.1% Triton-X-100	1:100
ab19875	Abcam	Pouf51	Rabbit	polyclonal	IgG	0.1% Triton-X-100	1:100

Table 2.5 Secondary Antibodies used for flow cytometry for detection of unconjugated primaries

ID/Catalogue number	Manufacturer	Host	Target	Conjugate	Dilution
A-11008	ThermoFisher	Goat	Rabbit	Alexa-488	1 in 500
A-11012	ThermoFisher	Goat	Rabbit	Alexa-594	1 in 500

2.5.4 RNA Extraction and cDNA Synthesis

RNA was first extracted from the cells using the acid-guanidinium thiocyanate phenol chloroform method (Chomczynski and Sacchi 1987). For cDNA synthesis 1-2 µg of RNA, 0.5 mM dNTPs and 0.1 µg/µL of random primers (both Invitrogen) were mixed to a volume of 12 µL and heated to 65°C for 10 minutes and then chilled on ice. 1X Reverse Transcriptase reaction buffer, Ribolock 20U was added along with 200 U of RevertAid reverse transcriptase (All from Thermofisher) to all, except negative control samples, to a volume of 20 µL. The samples were incubated at RT for 10 minutes followed by 42°C for 1-hour, 45°C for 10 minutes and then 70°C for 10 minutes.

2.5.6 RT-PCR

All primer sequences used were designed using NCBI's primer blast program for (Ye et al. 2012), and are listed in tables 7-9. These primers were designed to amplify regions spanning at least one exon-exon boundary, therefore ensuring no false positive signal due to gDNA contamination. The primer pairs were designed to anneal at 58°C, this was confirmed with an annealing temperature gradient test. The selection of cycle numbers for each primer pair was identified by examining the signal produced across a range of cycles and selecting a cycle at which amplification is linear. Commonly used housekeeping gene *GAPDH* was selected for use as a normalisation control to find the appropriate amount of cDNA to use in further PCR reactions. PCR reactions were conducted in a volume of 25 µL, to each was added 0.1 µM of forward and reverse primer (Eurofins), 1X MyTaq red reaction buffer, 0.5U KAPA Taq Polymerase (both KAPA Biosystems), and the appropriate volume of cDNA. An MJ Mini Thermal Cycler Machine (Biorad) was used for cycling conducted as shown in table 6, with steps 2-4 repeated a variable number of times depending on the gene being amplified. RT-PCR results were visualized using 1-3% agarose gel electrophoresis. 0.2 ng/ml 1kb plus ladder was loaded in the first lane of each gel for size comparison with bands.

Table 2.6 Commonly used PCR cycling conditions for RT-PCR

Step	Purpose	Temp (°C)	Time (min)
1	Denaturation	95	3
2	Denaturation	95	0.5
3	Annealing	58	0.5
4	Extension	72	0.5
5	Final extension	72	4
6	Storage	4	-

Table 2.7 Primers used for RT-PCR

Gene	Forward	Reverse
ACTA2	CCTATCCCCGGGACTAAGAC	GGTACTTCAGGGTCAGGAT
CACNA1C	TCAGCTCTAACAAACAGGTTTCG	AAGCCCCATAAGCAGTCATCT
CDH5	CACCCAGACCAAGTACACAT	ATGTATCGGAGGTCGATGGT
GAPDH	GGGCTGCTTTTAACTCTGGT	CATCGCCCCACTTGATTTTG
GATA4	GTCCAGACGTTCTCAGTCA	GGCCTCCTTCTTGCTATCC
GSC	AGTGGAGGTCTGGTTTAAGAAC	TTACAGCTCCTCGTTCCTCT
HAND2	AGTGCATCCCCAACGTA CC	TTCAAGATTTCGTTTCAG CTCCT
ISL1	TGTTACCAACTGTACAACCA	CATTTGATCCCGTACAACCTGA T
MESP1	CTATATCGGCCACCTGTCCG	CAGTCTGCCAAGGAACCACT
MYBPC3	AGCTCCACTTCATGGAG GTC	TGTCAAACACCCACTCA TCG
MYH6	GGAAGAGTGAGCGGCCATCAAGG	CTGCTGGAGAGGTTATTCTCTCG
MYH7	GCCAACACCAACCTGTCCAAGTTC	TGCAAAGGCTCCAGGTCTGAGGGC
NANOG	AAGGCAAACAACCCACT TCT	GGCATCCCTGCGTCACA

NKX2-5	AGACCGAAAAGAAAGCCTGAAA	GGCAGAGAGACGCTTGGTAA
POU5F1	TCCCTTCGCAAGCCCTCAT	TGACGGTGCAGGGCTCCGGGGA
RYR2	AGCCATTCTGCAAGACTCAC	GCTGCGTTTGATGCTTTCAT
SOX17	ATGCGGGATACGCCAGTGAC	TTCCACGACTTGCCCAGCAT
SOX2	CAGCGCATGGACAGTTA CG	AGTAGGACATGCTGTAG GTGG
TBX1	GGGTACTCCCAATCCTATTCTG	ATGGGTGAGGGGTGTGTAG
TAGLN2	GCTACCTGAAGCCGGTGTCC	TAGCACCGTGCCATCCTTGA
TBX2	GCTGACGATTGCCGCTATAA	TGCATGGAGTTTAGGATGGTGAAG
TBX3	GTCGGGAAGGCGAATGTTT	GAAAGTGACGACTTTGGACATC
TBX5	ACTGGGGTATGCCTGGTAAT	CAAGGTTCTGCTCTCCAAC
TNNI1	TCCACAACACCAGGGAG ATTA	AACATCTTCTCCGGCC TTC
TNNI3	CTCCAACACCGCGCTT ATG	CCGTGATGTTCTTGGTG ACT
TNNT2	TCGAGGCTCACT TTGAGAAC	TTTTCCGCTCTG TCTTCTGG
PRDM1	GCCACATGAATGCCAGGTCT	GGTGAACCTTGAGGCTACAG

Table 2.8 Primers used for RT-PCR analysis of *PRDM1* expression, with isoform specific primers.

Gene	Forward	Reverse
<i>PRDM1</i> 201/ CCDS34505	ACTCTGTGGTGG GTTAATCGG	GACTGCTCTGTG TTTGTGTGAGA
<i>PRDM1</i> 203/ CCDS5054	CCTCCAGTGTTG CGGAGA	CCTCCAGTGTTG CGGAGA
<i>PRDM1</i> 201, 203, and predicted transcripts	GCCACATGAATGCCAGGTCT	GGTGAACCTTGAGGCTACAG

Table 2.9 Primers used for gDNA PCR screening of CRISPR-Cas9 mutants, and for Sanger sequencing

Gene	Forward	Reverse
<i>GATA4</i>	GGAAGCTGCGGCCTACAG	CCCTCGACAGGGCTCAAGA
<i>TBX2</i>	CTCCCCGCAGACCATCCTAAA	ACATCTCCATCAGAACGACGG
<i>PRDM1</i>	TCACCACTTCATTGACGGCTTT	GTTTTTCCTCTCCTAGTCCCTAGC
<i>GATA6</i>	GGCCGACAGCCCTCCATA	CCCTTACCTGCACTGGGA

2.5.7 Immunofluorescence (IF)

CMs or iPSCs were dissociated cells as described previously and then resuspended to a concentration of 1×10^6 cells/mL, 40 μ L of cell solution was then plated onto 13mm glass coverslips coated with 1:400 Geltrex. The cells were allowed to attach and recover for a minimum of 48 hours before processing. The cells were fixed in 4% PFA for 10 minutes, followed by permeabilization with 0.25% Triton-X-100 PBS without Ca^{2+} Mg^{2+} . Alternatively, fixation and permeabilisation was carried out using 100% ice cold methanol for 15 minutes at 4°C. Regardless of fixation and permeabilization method used, blocking was carried out with PBS without Ca^{2+} Mg^{2+} 10% Serum (Serum was matched to the host organism of the secondary antibody) for 1-hour at RT. Incubation with antibodies summarized in table 10 in blocking buffer was conducted overnight at 4°C. The samples were then washed five times for 15-minutes at RT with PBS- 0.1% Triton-X100, then incubated with 1:500 Goat Anti-Mouse Alexa Fluor-488 antibodies or anti-rabbit depending on the primary antibody used (ThermoFisher) in blocking buffer for 2-hours in darkness, or 1:1000 if incubated overnight. Hoechst 33342 at 1 μ /mL was used to counterstain nuclei by incubation for 15 minutes, followed by 3 five-minute washes in PBS without Ca^{2+} Mg^{2+} . This was followed by mounting in Vectashield hard set mounting media with Phalloidin TRITC. Alternatively, HCS Cell Mask Orange was used at 1x to stain cell membranes and nuclei for 30 minutes, followed by 3 five minute washes in PBS without Ca^{2+} Mg^{2+} , mounting was then carried out using Vectashield with DAPI (Vector Laboratories) to stain the nuclei and coverslips were mounted face down on glass slides and sealed. A Zeiss LSM880 Confocal Microscope was used for imaging and FIJI/Image J Software (Schneider et al. 2012) for image analysis and editing.

2.5.8 Characterisation of CMs by IF

Following IF staining and imaging CM phenotype was classified as demonstrated in figure 2.2. Parameters such as cell area, circularity, aspect ratio, and number of nuclei were measured using FIJI/ImageJ software (figure 2.2 a-c). The organisation and overall condition of the CMs was also categorized following the criteria shown in figure 2.2 d. Class I represents the most organised cells and class IV represents the most disorganized. If a cell does not meet one of the criteria in class I, it moves down a class and so on.

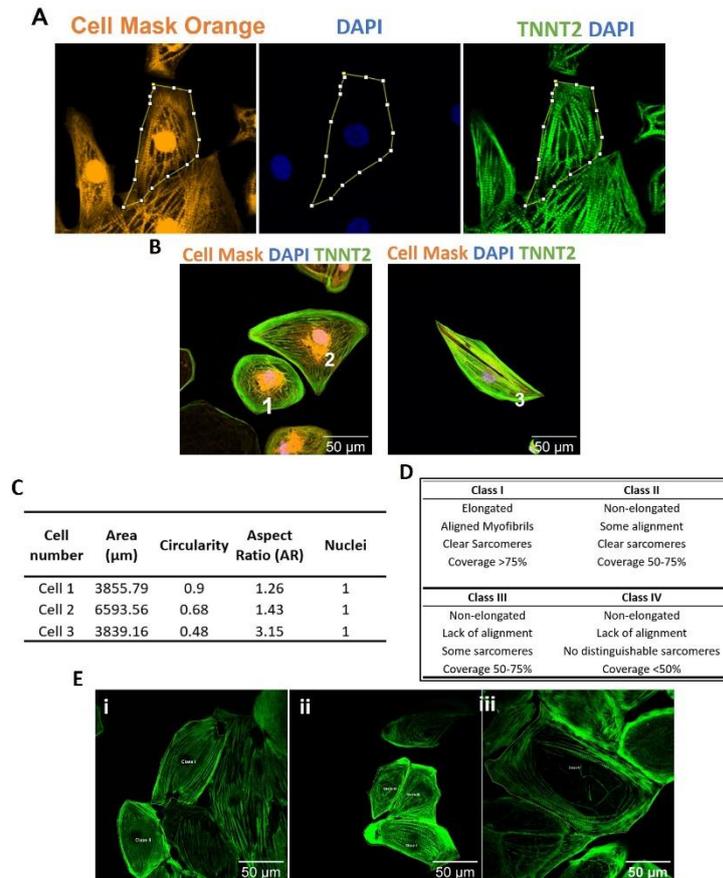


Figure 2.2 Cell measurements and classification system

(A) To determine cell area, circularity, and aspect ratio the perimeter of the cell was outlined as shown in figure 2.2 a, using HCS Cell mask orange to delineate the cell membrane. A note of the number of nuclei within this outlined region was noted. (B) An example of a selection of cells that have been assessed using this method. (C) A summary of the results from the cells shown in B. (D) A table detailing the criteria used to assign cells to a classification. Coverage refers to how much of the cell area is filled with TNNT2 staining. (E i-iii) Examples of cells that align with each classification.

2.5.9 Whole mount immunofluorescence for HFO's

HFO's were fixed using 4% PFA pH 7.4 for 1.5 hours at RT with gentle agitation. Permeabilisation was carried out using PBS with 1.5 % Triton-X-100 for 30 minutes. The HFO's were blocked using blocking solution (PBS 10% goat serum and 0.1 % Triton-X-100) for three hours at RT. Primary antibodies were added to blocking solution for overnight incubation at 4 °C, with gentle agitation. The organoids were then washed 3 times in PBS 0.1 % Triton-X-100 for 30 minutes each time. Then secondary antibodies in blocking solution were added and this incubation was carried out for 48 hours at 4 °C. Hoechst 33342 was added at 5 µg/mL PBS 0.1 % Triton-X-100 for 30 minutes. Then the HFO's were washed as carried out for primary antibody incubation. Antibody concentrations used are detailed in table 7.

For light sheet microscopy the HFO's were drawn into size 3 (inner diameter 1.5 mm) light sheet capillaries and embedded in a solution of 4 % low melting point agarose in distilled water. For imaging a Zeiss Lightsheet Z.1 was used the chamber was filled with PBS. Image acquisition parameters were generally set at: scan mode = frame, frame size = 1024 x 1024, averaging = 1, bit depth = 8. Single side illumination was used, and the organoids were imaged 4 times with 90 degrees of rotation between each stack of images, these were then compiled to give a composite 3D image.

2.5.10 Calcium Imaging

For the Gata4 mutants' calcium imaging was conducted using the widefield setting on the Olympus IX71 microscope. Before imaging the cells were loaded with 1 µM Fluo8-AM or 1 µM Fluo4-AM in cell culture media for 30 minutes at 37 °C. This media was then removed and Tyrode's solution with 1mM HEPES pH 7.4 (Merck, T2397) was added. The cells were then imaged by brightfield microscopy for beating and then with fluorescent imaging for calcium flux.

The change in fluorescence over time was plotted using FIJI/Image J Software (Schneider et al. 2012). MatLab R2012b was used to extract information regarding the amplitude and length of any detected calcium transients, using a custom script provided by Dr. Ewan Fowler, Cardiff University.

2.5.11 Western Blotting

Western Blots were carried out using standard methods, as described in (Towbin et al. 1979). Cells were lysed in RIPA Buffer (50 mM Tris-HCl pH8, 150 mM NaCl, 0.1 % SDS, 0.5% Na Deoxycholate, 1 % NP-40, and 1x Protease Inhibitor Complete Mini (Roche)). Protein concentration was measured using the Pierce BCA Protein Assay Kit from ThermoFisher, or loading volume was normalized by cell number. The protein extract was mixed with 2x Laemmli Loading Buffer (BioRad) and boiled for 10 minutes before gel loading, 20 µg of total protein was used per well. Bands were resolved via polyacrylamide gel electrophoresis with 2cm of 5% stacking gel and 5cm of 9% resolving gel, the samples were then transferred to polyvinylidene difluoride membrane (Millipore). 5% Skimmed Milk in TBS Tween 20 (TBS-Tw: 5 mM Tris-Base (Fisher) pH 7.4, 20 mM NaCl, 0.1% Tween 20) was used to block the membranes for 1-hour. For detection of Tbx2 skimmed milk concentration was reduced to 1 % to improve signal intensity. Incubation with primary antibodies was carried out overnight in blocking buffer at 4°C with constant rotation. Following this the membranes were washed with TBS-Tween 3 times for 10 minutes each time. Secondary antibodies were then applied to the membranes, these were diluted in 5% skimmed milk TBS-Tween or for detection of Tbx2 1% skimmed milk TBS-Tween. Secondary antibody dilutions are as follows; 1:10,000 for goat anti-mouse-HRP (Merck-Millipore), 1:50,000 for rabbit anti-mouse-HRP antibodies (Santa Cruz), and 1:10,000 for Goat anti-rabbit-HRP antibodies (Abcam). The membranes were once again washed, this was conducted 3 times for 10 minutes each time in TBS-Tween to remove excess antibody. Clarity Western ECL (Biorad) was used for development followed by exposure of the membranes to Amersham Hyperfilm (Merck). The film was developed using developer and fixer from WolfLabs using standard methods.

Table 2.10 – Antibodies used for Immunofluorescence

Clone ID/Cat no.	Manufacturer	Target protein	Host	Isotype	Staining concentration	Fixation	Permeabilisation
ab19875	Abcam	POU5F1	Rabbit	IgG	1:2000	4% PFA	0.25% Triton-X-100
ab97959	Abcam	SOX2	Rabbit	IgG	1:1000	4% PFA	0.25% Triton-X-100
12G10	DSHB	TUBA1A (α -Tubulin)	Mouse	IgG1	1 μ g / mL	4% PFA	0.25% Triton-X-100
ab32575	Abcam	ACTA2 (α smooth muscle actin)	Rabbit	IgG	1:200	Methanol	Methanol
MA5-12960	Thermofisher	TNNT2	Mouse	IgG1	1:600 (1:100 for 3D cultures)	4% PFA	0.25% Triton-X-100
H-112	SantaCruz Biotech	GATA4	Rabbit	IgG	1:500	4% PFA	0.25% Triton-X-100
AF1700	R&D	GATA6	Goat	IgG	1:500	4% PFA	0.25% Triton-X-100
sc-137180	SantaCruz Biotech	MYBPC3	Mouse	IgG1k	1:500	4% PFA	0.25% Triton-X-100
06-570	Millipore	Histone H3 (p-ser 10)	Rabbit		1:1000	4% PFA	0.25% Triton-X-100
M-38	DSHB	Pro-COL1A1	Mouse	IgG1	1 μ g /mL	4% PFA	0.25% Triton-X-100
CH1	DSHB	TPM1 (Tropomyosin 1)	Mouse	IgG1	0.5 μ g /mL	4% PFA	0.25% Triton-X-100
E7	DSHB	β -Tubulin	Mouse	IgG1	0.5 μ g /mL	4% PFA	0.25% Triton-X-100
AMF-17b	DSHB	VIM (Vimentin)	Mouse	monoclonal	IgG1	1 μ g/mL	Methanol

Table 2.11 Cell stains used for Calcium Imaging and Immunofluorescent staining of fixed cells

ID/Cat no.	Manufacturer	Name	Target Stained/Purpose	Working concentration
H32713	ThermoFisher	HCS Cell Mask Orange	Cell membrane, cytosol, nucleus	1x
H-1200	VectaShield	Vectashield Antifade Mounting Media with DAPI	Nuclei	-
H-1600	VectaShield	Vectashield Hard Set Mounting Media with Phalloidin-TRITC	Filamentous Actin	-
21083-AAT 20550-AAT	Stratech	Fluo-8AM and Fluo-4AM	Calcium Imaging	1 μ M

2.6 RNA-seq analysis

RNA extractions were carried out as detailed in section 2.5.4. These samples were then made into cDNA libraries and sequenced by Angela Marchbank, The Genome Research Hub, Cardiff University. Bioinformatics and DEseq2 analysis was provided by Antonios Tselingas, Cardiff University and Prof. Peter Kille, Cardiff University. Antonios also provided some of the Heatmaps presented in this thesis, and this is indicated where appropriate. Gene Ontology enrichment analysis was carried out using ShinyGO v0.77 and STRING v11.5 for network analysis (Szklarczyk et al. 2019; Ge et al. 2020).

2.7 Statistical analysis

Unless otherwise stated all graphs and statistical analysis were created using GraphPad Prism v9.3.1. For normally distributed data where a comparison of more than two groups has been made this has been done using an ordinary one-way ANOVA. For non-parametric data a Mann-Whitney test has been used. When comparing just two groups an unpaired t-test has been utilised for parametric data, or a Mann-Whitney test for non-parametric data.

Chapter 3 – Cardiac transcription factor GATA4 is required for iPSC-CM differentiation

3.1 Introduction

The specification of pluripotent cells from the ICM to a cardiac fate and their subsequent differentiation into CMs requires the integration of multiple signals and the combinatorial action of a conserved set of TFs. Early in development graded signalling within the embryo sub-divides the mesoderm to form cardiac mesoderm, from which amongst various other cardiac structures, the myocardium will form. GATA4 is a TF that is expressed in the earliest stages of this process (Heikinheimo et al. 1994). As cardiac differentiation proceeds CPs from the mesoderm will coalesce to form the linear heart tube, which will subsequently undergo looping and septation to form the adult heart. GATA4 continues to be expressed broadly throughout the heart during this time, and then in the adult heart (Nemer and Nemer 2003; Bisping et al. 2006; Oka et al. 2006). Various studies have demonstrated crucial roles for GATA4 in CM development and homeostasis, this has established GATA4 as a central node within the cardiac GRN and these studies will be considered below. The results presented in this Chapter aim to add to this knowledge base by using iPSC-CM differentiation to model the requirement for GATA4 in human cardiomyogenesis, for which little information is available.

3.1.1 GATA4 in cardiac development

The first knockout experiments in mice demonstrated a role for GATA4 in early cardiac morphogenesis. In mice loss of GATA4 expression is lethal by E9.5, due to requirements for GATA4 in the extraembryonic endoderm. These mice also display, cardia bifida, where the migration of CPs and their fusion at the midline to form the linear heart tube fails (Kuo et al. 1997; Molkentin et al. 1997). Similar results have been seen in other model organisms, such as *Xenopus*, where knockdown of Gata4 results in a small increase in acardia but the primary defect observed is cardia bifida (Peterkin et al. 2007; Haworth et al. 2008). In the two patches of cells, that have failed to fuse, expression of CM markers such as; *MYH6*, *NPPA*, *MYL2*, and *MYL4* are detectable by *in situ* hybridisation indicating that specification and differentiation of some CMs has taken place (Kuo et al. 1997; Molkentin et al. 1997; Haworth et al. 2008). Work by Watt *et al.* using tetraploid complementation demonstrated that the cardia bifida phenotype can be overcome if GATA4 null embryos are

supplied with GATA4 positive extraembryonic endoderm. Suggesting these early aspects of cardiac development rely on GATA4 expression in extraembryonic endoderm rather than in CPs themselves. However, the presence of a variety of other cardiac defects in the GATA4 null embryos such as hypoplastic ventricular myocardium, disrupted looping, septation, and the failure to form any proepicardium indicates cardiac development has not proceeded normally (Watt et al. 2004). Thus, GATA4 likely does have intrinsic roles in the cardiac cell types formed during heart development.

A selection of other studies have demonstrated that GATA4 does indeed have roles in the development of cardiac progenitors themselves. Deletion of GATA4 specifically in CPs using Cre recombinase driven by *Nkx2-5* results in lethality by E11.5, due to severe hypoplasia of the myocardium, this is in addition to deficient endocardial cushion development which causes chamber septation to fail. Re-capitulating a number of the defects observed by Watt *et al.* These deficiencies were partially explained by reduced CP proliferation (Zeisberg et al. 2005), and further study has revealed the mechanisms behind this. Deletion of GATA4 in the SHF using *Mef2c* as a driver of Cre expression results in right ventricular hypoplasia as a consequence of reduced proliferation. In this study GATA4 was shown to directly regulate and induced the expression of cell cycle regulators such as *Cyclin D2*, *A2*, and *Cdk4* in the SHF progenitors (Rojas et al. 2008). Later studies in *Xenopus* revealed that Gata4 directly interacts with Cyclin D2 to induce CM specific genes such as *myl7* and *myh6* (Yamak et al. 2014). Taken together these studies show that GATA4 promotes proliferation and differentiation of CPs. Furthermore, genome wide ChIP studies demonstrate GATA4 binding at the enhancers and promoters of an array of CM specific genes during development and in the adult heart (He et al. 2014; Ang et al. 2016; Luna-Zurita et al. 2016; Akerberg et al. 2019), as well as binding at non-CM genes to repress expression of ectopic gene programs (Ang et al. 2016; Robbe et al. 2022). Collectively, these studies suggest that GATA4 plays a substantial role in regulating the cardiac GRN in development.

3.1.2 Incomplete redundancy between GATA4 and GATA6

GATA6 is a close relative of GATA4 and there is considerable overlap in their expression domains in the developing and adult heart (Morrisey et al. 1996; Kuo et al. 1997; Molkenin et al. 1997; Nemer and Nemer 2003; Zhao et al. 2008). Due to these similarities

and their ability to recognise the same (A/T) GATA (A/G) DNA motif it has been proposed that they may regulate each other's targets and in doing so have the potential to compensate for the loss of the other factor. In support of this, heterozygous mutation of either GATA4 or 6 results in the production of phenotypically normal mice. Yet, compound heterozygosity for GATA4 and 6 results in septation defects, and myocardial thinning (Xin et al. 2006b). Suggesting that if a certain level of GATA4 or 6 is maintained that cardiac development can proceed normally.

However, this likely oversimplifies the situation, and support for individual roles for each factor has also been documented. It is known that GATA4 interacts and co-occupies DNA sites with other cardiogenic TFs and this makes binding specificity possible, if these interactions are disrupted then binding specificity is lost (Luna-Zurita et al. 2016; Robbe et al. 2022). One such TF is TBX5, comparison of GATA4/TBX5 compound heterozygous mice with GATA6/TBX5 compound heterozygous mice, identifies individual roles for GATA4 in CM development and transcriptional regulation. Firstly, the cardiac defects noted in the GATA4/TBX5 mutants are more severe and penetrant than those in GATA6/TBX5 mutants. Moreover, expression of *Myh6* is maintained in GATA6/TBX5 mutants but not in GATA4/TBX5 mutants, revealing a requirement for GATA4 in the activation of *Myh6* expression that cannot be compensated for by GATA6 (Maitra et al. 2009). Similarly, if the *Gata4* coding sequence (CDS) is replaced with *Gata6* cDNA; but the *Gata4* genomic locus is otherwise unchanged, mice present with a range of severe defects including ventricular hypoplasia, reduced trabeculation, and abnormal endocardial cushion and pro-epicardium formation (Borok et al. 2016). The phenotype documented by Borok *et al.* is reminiscent of that reported by Watt *et al.*, jointly these studies make it clear that any redundancy between the two factors that has been observed previously is incomplete.

3.1.3 Studying GATA factors in human iPSC-CM differentiation

Further to the above, a recent study published by Sharma *et al.* examining the requirement for GATA6 in iPSC-CM differentiation shows it is indispensable for this process. This stands in contrast to what is observed in mouse embryos where GATA6 null mice, show only minor cardiac defects (Morrissey et al. 1998; Koutsourakis et al. 1999; Zhao et al. 2005). This raises the question of whether the requirement for GATA4 during human iPSC-CM

differentiation will also differ to what has been documented in model organisms such as mice and *Xenopus*.

Investigation of the disease causing GATA4 G296S mutation in mouse and human models of development supports this notion. This mutation occurs heterozygously in human patients and results in a range of atrial and ventricular septal defects, of a severity often requiring surgery (Garg et al. 2003). In contrast, heterozygous mouse models carrying the corresponding mutation appear grossly normal, with a proportion showing mild defects in post-natal atrial septal closure (Misra et al. 2012). However, CMs generated from patient derived iPSCs appear abnormal and display contractility, and physiology defects (Ang et al. 2016a). This example suggests that human cardiac development is more sensitive to GATA4 mutation and provides a good illustration of how iPSC-CM models have been instrumental in uncovering species differences. Furthermore, this example shows how this model can provide insights into disease pathogenesis that go beyond what is observable in mice. A recent publication by Gonzalez-Teran *et al.* documents the creation of a GATA4 null human iPSC line in which differentiation efficiency is reduced and beating is delayed. Suggesting GATA4 influences this process but is not essential. The results presented here explore this further with the creation and characterisation of additional GATA4 null iPSC lines and examination of the expression of two of its target genes; *TBX2* and *PRDM1*.

3.1.4 Aims

There is a significant amount of evidence available from model organisms that GATA4 is essential for normal cardiac development, how well these findings relate to human cardiac development is unclear. The role of GATA4 in this process will be addressed in this Chapter through the creation of GATA4 null human iPSC lines using CRISPR-Cas9 gene editing. The cardiogenic potential of the lines created will be assessed using established Wnt modulation-based iPSC-CM differentiation protocols (BurrIDGE et al. 2014b; BurrIDGE et al. 2015). Following this the phenotype of the cells formed will be explored using a mixture of imaging and molecular laboratory techniques.

As discussed, GATA4 is expressed early in cardiac development, and thus in addition to its known roles directly regulating the expression of a selection of committed CM markers it can exert its effects at earlier stages. However, identification of these early targets is far from complete. As outlined in section 1.4.2, GATA4 is sufficient to drive cardiomyogenesis in

Xenopus laevis pluripotent explants (Latinkic 2003). Other work in our lab has combined this method with RNA-seq to identify early differentially regulated genes that respond to GATA4 expression. Two novel targets, *prdm1* and *tbx2*, were identified and using CHIP RT-PCR were validated as direct targets of Gata4 regulation in *X. laevis* (Latinkic *et al.* data not published). It is unknown whether this relationship is conserved in humans, therefore another aim of this work is to clarify this, by assessing the expression of these genes in the GATA4 null lines created.

3.2 Results

3.2.1 Creation of two trans-heterozygous *GATA4* mutants using CRISPR-Cas9 gene editing

CRISPR-Cas9 gene editing is a versatile tool for introducing mutations into DNA (Jiang and Doudna 2017). Here, CRISPR Cas9 was used to edit the *GATA4* gene locus using two guides to target exon 2, its first coding exon (see figure 3.1a). The intended cut sites were placed 68 bp away from each other with the intention of deleting the intervening bases and creating a frameshifting mutation that should stimulate nonsense mediated decay of any transcripts produced from the *GATA4* locus. Two mutants were identified, both produce two amplicons that are lower than the WT (figure 3.1b), indicating the lines carry different mutations on each allele. Sanger sequencing (figure 3.1c), followed by SyntheGo ICE signal deconvolution for each band revealed one to be an 11 bp deletion, and the other a 67 bp deletion (see figure 3.1c), both are frameshifting mutations and therefore are expected to result in diminished *GATA4* mRNA and an absence of protein expression.

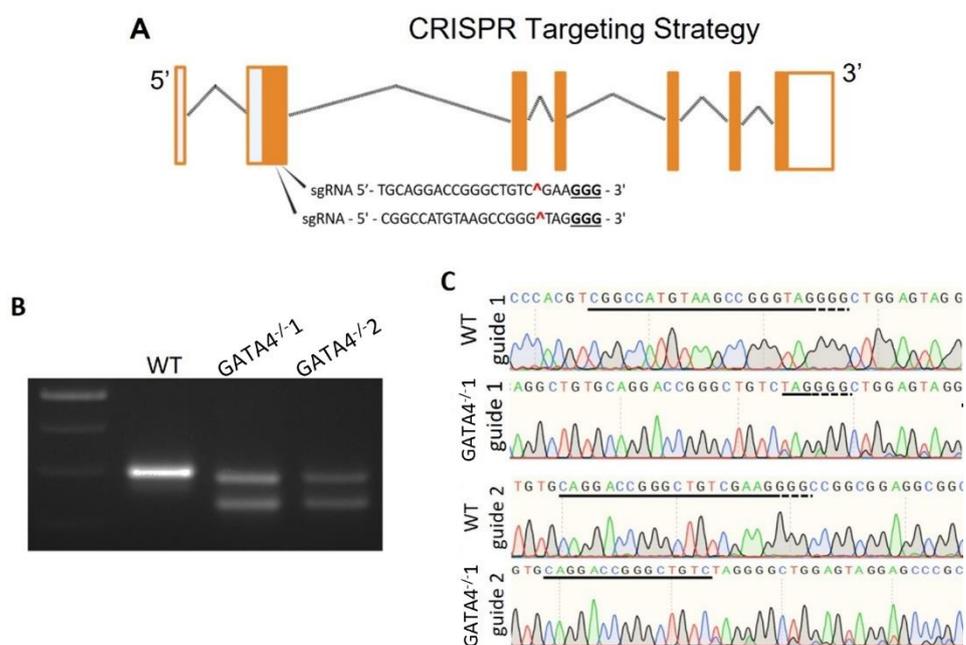


Figure 3.1 CRISPR targeting strategy and the identification of *GATA4* mutants

(A) Schematic representation of the *GATA4* locus. Boxes are used to represent exons, filled boxes marking coding regions, and hollow boxes marking non-coding regions. Introns are shown as dotted lines. (B) Gel separation of the PCR products from WT and mutant gDNA surrounding the target site. (C) Sanger Sequencing trace files for the WT and *GATA4*^{-/-} amplicons.

3.2.2 GATA4 protein is not expressed in mutants

WT cells were differentiated using the CDM3 protocol, an established method for cardiac directed differentiation using Wnt modulation (Burridge et al. 2014). *GATA4* expression was analysed during differentiation by RT-PCR (figure 3.1b), and the pattern found to be consistent with published data for similar differentiation protocols (Churko et al. 2018). *GATA4* is expressed from the onset of differentiation at a low level, and starts to increase at day 4, coinciding with the establishment of the cardiac progenitor pool (figure 3.2a). To determine how *GATA4* expression had been affected by the mutations, WT and mutant iPSCs were analysed at day 4 and day 10 of differentiation. At day 4 RT-PCR analysis shows reduced levels of *GATA4* in the mutants (figure 3.2c), the same trend can be seen at day 10 (figure 3.2d). Furthermore, *GATA4* protein can readily be detected in WT day 4 whole cell protein extracts by Western blotting, in contrast no *GATA4* signal was detected in the mutant lines (figures 3.2e). Immunofluorescence conducted on day 12 samples shows that *GATA4* localises to the nucleus of WT cells, whereas no signal can be seen in the *GATA4*^{-/-}1 mutant. Together these readouts confirm loss of *GATA4* protein expression in these lines.

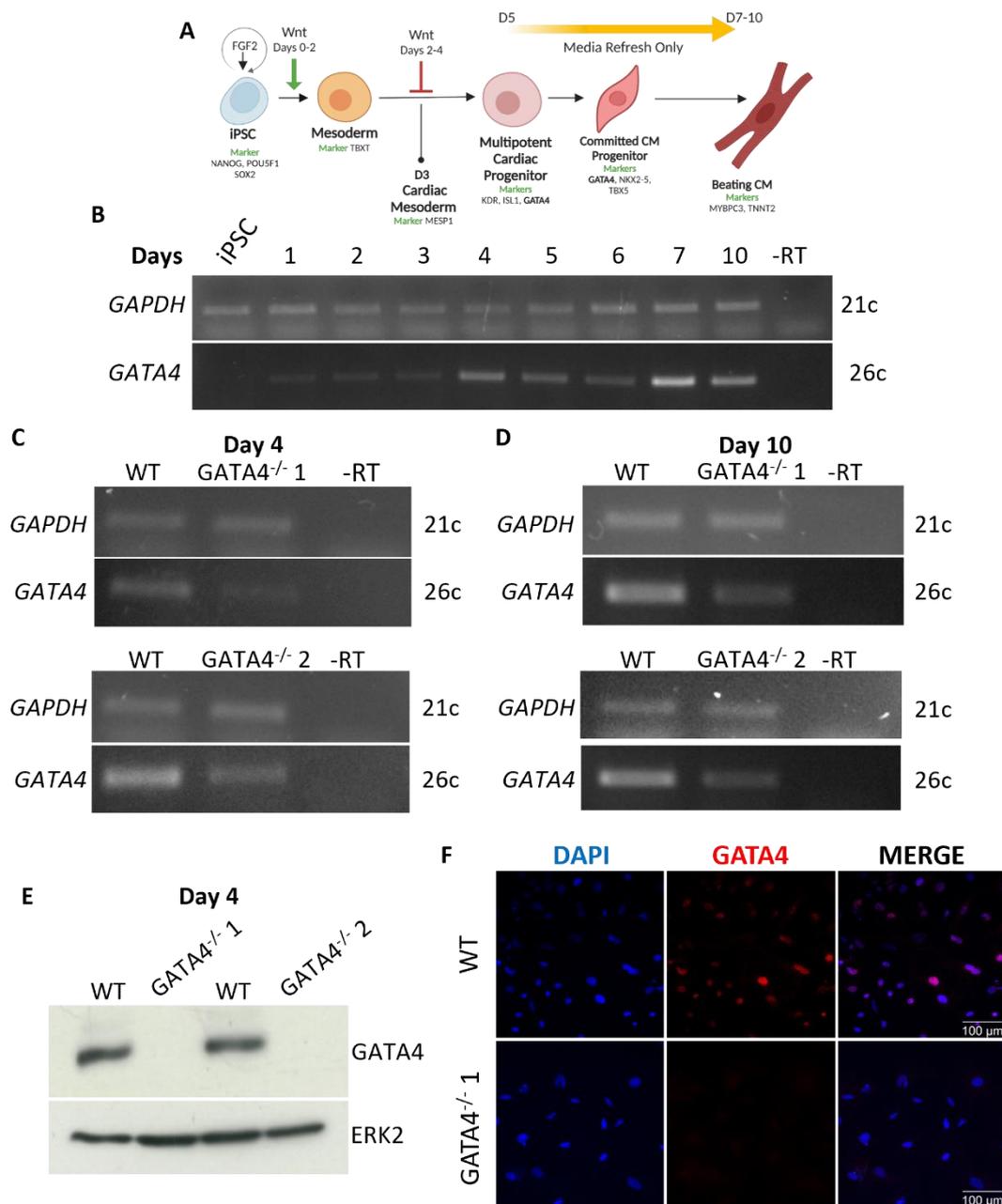


Figure 3.2 *GATA4* mRNA is reduced in mutants and no detectable protein is produced

(A) A diagram for reference showing some of the key steps and markers expressed during iPSC-CM differentiation. (B) Expression of *GATA4* during a WT iPSC-CM differentiation analysed by RT-PCR, *GAPDH* was used for normalisation. (C) Expression of *GATA4* was analysed by RT-PCR at day 4, (D) and day 10 of differentiation for WT and mutants. (E) Corresponding Western blot images show *GATA4* protein levels in the samples at day 4 (n = 4) ERK2 was used as a control for protein loading. (n = 3) (F) Immunofluorescence was also performed on 12 WT and *GATA4*^{-/-} 1 samples, showing *GATA4*⁺ cells in red. Cells were counterstained with DAPI (blue) (n = 3).

3.2.3 GATA4 mutant cell lines maintain pluripotency

During editing iPSCs are exposed to harsh conditions that can affect genomic integrity, this can subsequently affect pluripotency precluding them from further use. To ensure the clones selected are of a similar quality to WT despite the mutation in the *GATA4* gene they now carry, several quality checks were carried out. Firstly, the morphology of the colonies was surveyed, and no differences were seen between WT and the mutants. All form rounded tight compact colonies, in which it is hard to distinguish single cells, which is indicative of pluripotent stem cell colonies (figure 3.3a). IF staining showed the expression of pluripotency factors POU5F1 and SOX2 was comparable between WT and mutants (figure 3.3b). Therefore, the cells are of a suitable quality for further use.

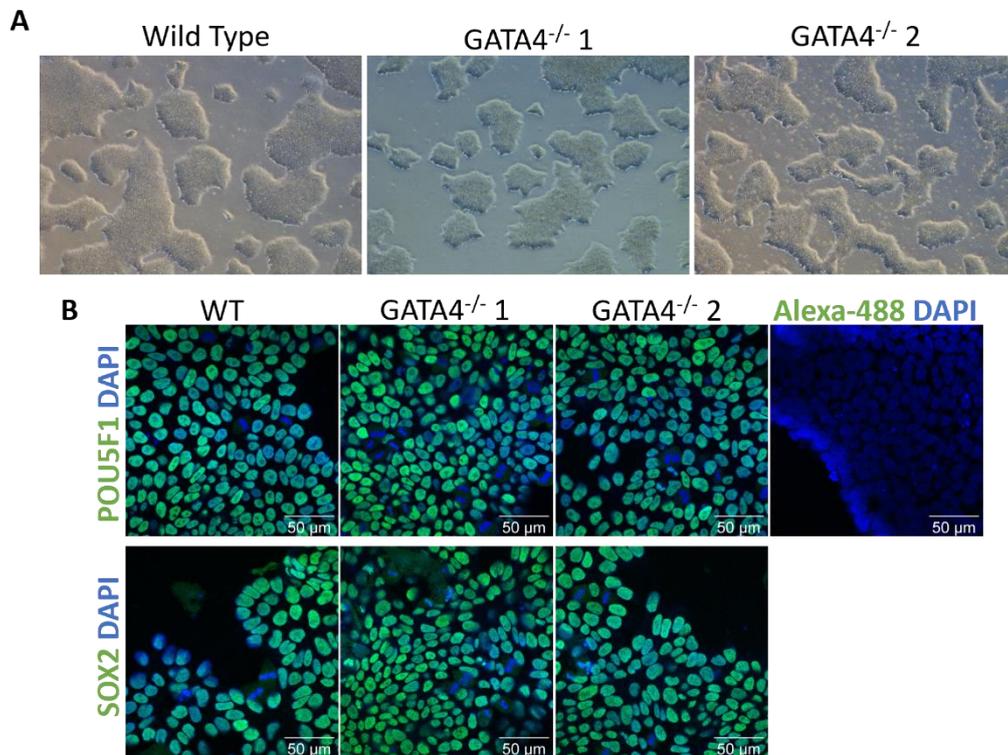


Figure 3.3 GATA4 mutant cells maintain pluripotency

A) Bright field images of WT and GATA4 Mutant iPSC colonies at 4x magnification. (B) POU5F1 and SOX2 expression detected by immunofluorescence in WT and GATA4 mutant cell lines. 63x magnification used. N = 2.

3.2.4 GATA4 mutants fail to form functional cardiomyocytes

To test the role of GATA4 in human iPSC-CM differentiation, WT and the GATA4 mutant iPSCs were differentiated using the CDM3 protocol (Burrige *et al.* 2014) and monitored for beating from days 7-32 (supplemental videos 1-6). Across 12 different experiments the proportion of beating wells for WT varied from 30%-100%, with an average of $68.50\% \pm 7.51$ SEM ($n = 153$ wells, across 12 independent differentiations). No beating was seen in the wells of the mutant cell lines, GATA4^{-/-1} ($n = 139$ wells, across 12 independent differentiations), GATA4^{-/-2} ($n = 66$ wells, across 6 independent differentiations) (figure 3.4b). Despite the lack of beating seen in the mutants, the gross morphology of the cells produced appears similar to WT cells. At day 12 robust cellular networks can be seen in the WT wells, that beat synchronously (figure 3.4a, supplemental video 1). Networks appearing similar to these are also observable in the mutants, but no beating is seen (figure 3.4a). At day 32 at 10x magnification following re-plating single cells can be discerned (figure 3.4a), again no beating was detected.

In a pilot experiment, the WT and mutant cells were differentiated using an alternative 3D 'cardiac organoid' model of differentiation. The iPSCs were aggregated and embedded in extracellular matrix as per the protocol published by (Drakhlis *et al.* 2021), then differentiated following the CDM3 protocol used throughout. Agreeing with the results seen in 2D, in this model 5/13 WT organoids were beating by day 15, in contrast to 0/19 beating HFO's for GATA4^{-/-1} (figure 3.4c, and supplemental video 7-8).

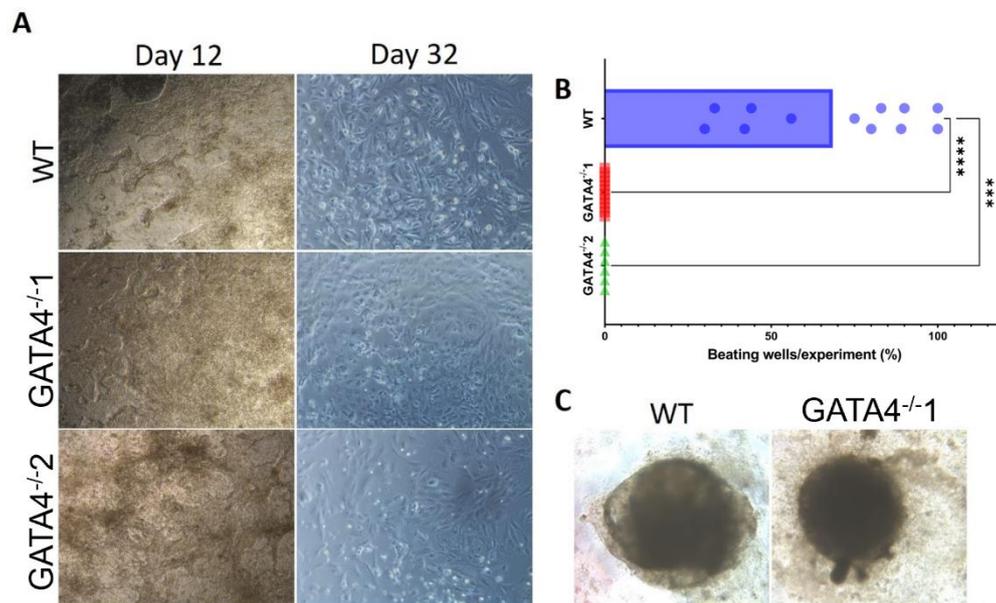


Figure 3.4 GATA4 mutant iPSCs fail to form beating cells

(A) The first two column show representative brightfield images of WT, GATA4^{-/-1}, and GATA4^{-/-2} cells at day 12 of a CDM3 differentiation at 4x magnification. The final two column shows brightfield 10x images of cells at day 32 of differentiation. (B) A summary of the mean average number of beating wells observed per experiment for WT (n=153 wells, across 12 independent experiments) and GATA4^{-/-} lines (GATA4^{-/-1}; n=139 wells, across 12 independent differentiations, and for GATA4^{-/-2}; n=66 wells, across 6 independent differentiations). Statistical test = one-way ANOVA. P-values; ***≤0.01, and ****≤0.001. (C) Representative images of WT and GATA4^{-/-1} Heart Forming Organoids cultured until day 32, shown at 4x. N = 1.

3.2.5 GATA4 is needed for the expression onset and maintenance of a selection of sarcomeric genes

GATA4 is known to bind to and induce expression of a number of sarcomeric components, such as; *MYH6*, *MYH7*, and *TNNI3* (Molkentin et al. 1994; Hasegawa et al. 1997; Latinkic 2004; Tomoya Sakamoto 2022). These genes are also specific to CMs and thus can be used as another method for determining if any CM-like cells are present in the GATA4 null cell cultures. The expression of these genes was analysed by RT-PCR at days 10 and 32 of differentiation in WT and the GATA4 mutants, in addition to the expression of *TNNT2* a commonly used as an indicator of differentiation efficiency (Lian et al. 2012; Burridge et al. 2014) and *MYBPC3* which are both pan-myocardial markers. The GATA4 target genes were consistently down regulated at both time points, as were *TNNT2* and *MYBPC3* (figure 3.5a). The exception to this pattern was *TNNI1*, whose expression seems unaffected, or even increased in comparison to WT at day 10 (figure 3.5a). When examined again at day 32 this trend in *TNNI1* expression is not maintained, and it follows the same pattern as the other markers assayed (figure 3.5b). These results demonstrate a general downregulation in the expression of definitive CM markers.

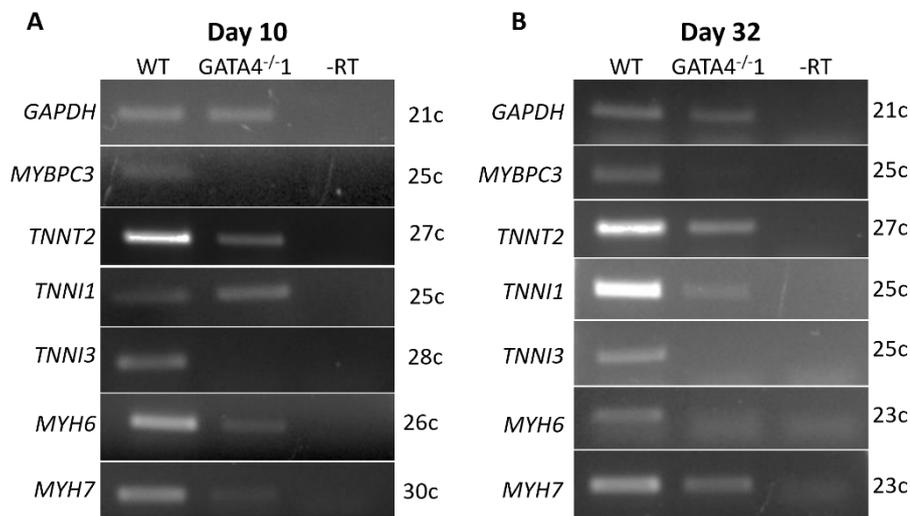


Figure 3.5 The expression of sarcomeric genes is decreased in WT and GATA4^{-/-1} cells

(A) Expression of a selection of sarcomeric genes was analysed using RT-PCR following 10 days of directed differentiation using the CDM3 protocol, *GAPDH* is used as a normalisation control, n = 2. (B) The cells were maintained until day 32 and expression of the genes was analysed again by RT-PCR, again *GAPDH* is used as a normalisation control, n = 2.

3.2.6 Positive Staining for CM Specific Proteins Reduced in GATA4 Mutants

The phenotype of the GATA4^{-/-} cells was further characterised at a single cell level by immunofluorescence (IF). This analysis showed that far fewer TNNT2 positive cells (TNNT2+) were formed from the GATA4^{-/-} iPSCs (see table 3.1), on average the number of TNNT2+ WT cells was 64.25% ± 10.87 (n=4), for GATA4^{-/-} 1 14.50% ± 7.14 (n=4), and GATA4^{-/-} 2 7.25% ± 8.29 (n=4). The number of cells positive for MYBPC3 was lower again for the GATA4 mutants. WT cultures showed 84% ± 1, GATA4^{-/-} 1 1.68% ± 1.68, and GATA4^{-/-} 2 0% ± 0 (for all lines n=3). However, when TNNT2 and MYBPC3 staining within each cell line was compared, no significant difference was seen between the two (see supplemental figure 10). To assess if the expression level of these genes was reduced in individual cells mean fluorescence was measured per cell, showing that on average GATA4^{-/-}1 cells were 2.49 times lower than WT (see table 3.1 and figure 3.6b). This analysis was also performed for GATA4^{-/-} 2, and a 1.99-fold decrease was observed but sample number was low (n=8 cells). This analysis was not performed for MYPBC3 staining in either line, due to insufficient numbers of positive cells.

Table 3.1 A summary of the results obtained from immunofluorescent staining of WT and GATA4^{-/-} 1 and GATA4^{-/-} 2 cells at day 32 of differentiation.

The differentiation efficiency for the WT and GATA4^{-/-} lines was calculated based on positivity for TNNT2. For the WT and GATA4^{-/-}1 this was calculated across four independent biological repeats, whilst three repeats were used for evaluation of the GATA4^{-/-} 2 cell line. For all other metrics three biological repeats were used for all cell lines. The mean averages are presented as the Mean ± SEM, and in this table, ‘n’ represents the number of cells examined for each metric across the biological repeats conducted.

	Cell Line		
	WT	GATA4 ^{-/-} 1	GATA4 ^{-/-} 2
TNNT2+ cells	64.25% ± 10.87, n=935	14.50% ± 7.14, n=642	7.25% ± 8.29, n=475
TNNT2+ cells with sarcomeres	92.33% ± 1.45, n = 89	6.33% ± 1.33, n = 90	15.00% ± 15.00, n = 13
Mean TNNT2 Fluorescence/Cell	876.2 ± 42.17, n = 81	351.3 ± 39.53, n = 90	438.8 ± 95.21, n = 8
MYBPC3+ cells	84% ± 1, n = 290	1.68% ± 1.68, n = 396	0% ± 0, n = 266
MYBPC3+ cells with sarcomeres	87.67% ± 8.2, n = 116	6.5% ± 6.5 n = 8	0% ± 0 n = 0

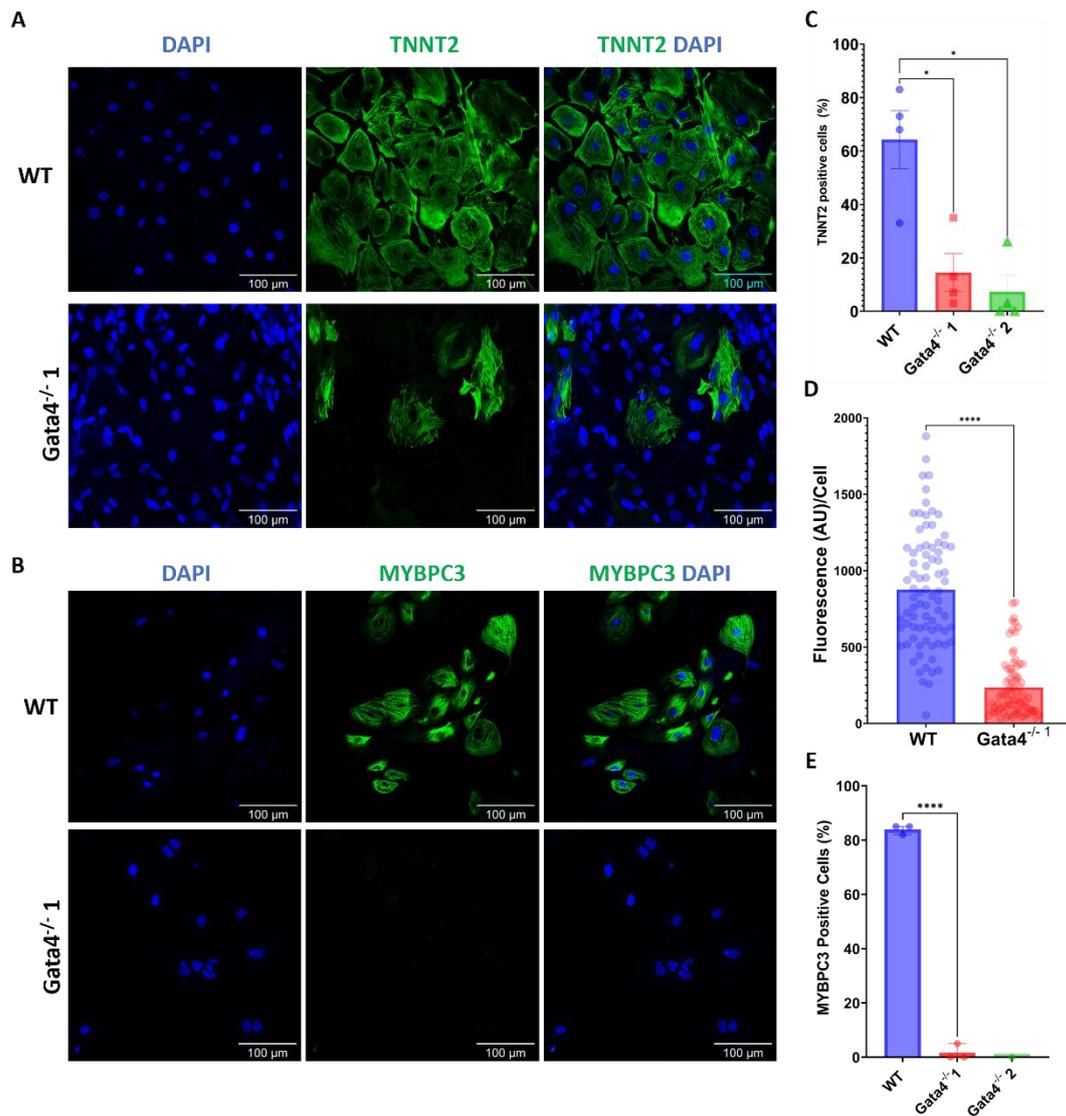


Figure 3.6 GATA4 mutants form fewer cells positive for CM markers

(A) Staining for TNNT2 (green) in WT and GATA^{-/-1} mutant cells at day 32 of cardiac directed differentiation. DAPI (blue) was used to counterstain nuclei, n=4. Representative examples shown. (B) Staining for MYBPC3 (green) in the same cells, DAPI was once again used to counterstain. (C) A comparison of the proportion of TNNT2 positive cells present between cell lines, details of the number of cells analysed and biological repeats conducted are given above in table 3.1. (D) TNNT2 mean fluorescence per cell for WT (n = 81 cells, across 3 independent experiments) and Gata4^{-/-1} (n = 90 cells, across 3 experiments). (E) A comparison of the proportion of TNNT2 positive cells present between cell lines, details of the number of cells analysed and biological repeats conducted are given above in table 3.1. A one-way ANOVA was used for all statistical analysis except for panel E, where the WT and GATA4^{-/-1} line have been compared using a student's t-test, due to only one repeat being available for GATA4^{-/-2} with the MYBPC3 marker. P-values: * ≤ 0.05, and **** ≤ 0.0001.

3.2.7 The GATA4^{-/-}TNNT2⁺ cells lack sarcomeres

There are far fewer cells expressing CM specific sarcomeric proteins, and the rare positives contain far less TNNT2 protein, thus providing a likely explanation why mutant cells do not beat. However, as a small proportion still express these proteins, albeit at a reduced level, closer inspection of the cells was warranted. Higher resolution pictures of cells at day 32 of differentiation show clear striations in the WT cells when they are stained for TNNT2, MYBPC3, or F-ACTIN (figure 3.7ai, bi, and ci), revealing different components of the sarcomere. In the GATA4^{-/-} cell line striations could only be seen in 6.33% ± 1.33 of TNNT2 positive cells. An example of a GATA4^{-/-} cell with striations is shown in figure 7e. This cell is the least dysfunctional cell that was observed across all IF experiments for the GATA4 null lines (n = 4), but still has a large proportion of fibrils which lack striations. Further examples of cells with the resemblance of some striations are given in S6, all have considerably worse sarcomere structure than the cell shown in figure 3.7e.

Furthermore, no striations were observed in any of the MYBPC3 positive cells, only diffuse fibrils were seen. ACTIN co-localised with TNNT2 and MYBPC3 staining in WT cells, but no F-ACTIN striations were seen in the Gata4 mutants, confirming the results seen for TNNT2. To eliminate the possibility that this is due to an overall problem with the assembly of structural proteins in the cell staining for another cytoskeletal protein α -tubulin was carried out. Organisation of α -tubulin between the WT and mutants was indistinguishable (figure 3.7 di-iv). Indicating sarcomere formation or maintenance is specifically affected by the absence of GATA4. This will contribute to the lack of beating seen and could be due to the reduced TNNT2 expression seen in the TNNT2⁺ GATA4 null cells analysed (see figure 3.6d) or due to the absence of other unidentified sarcomeric components missing in the GATA4 null cells.

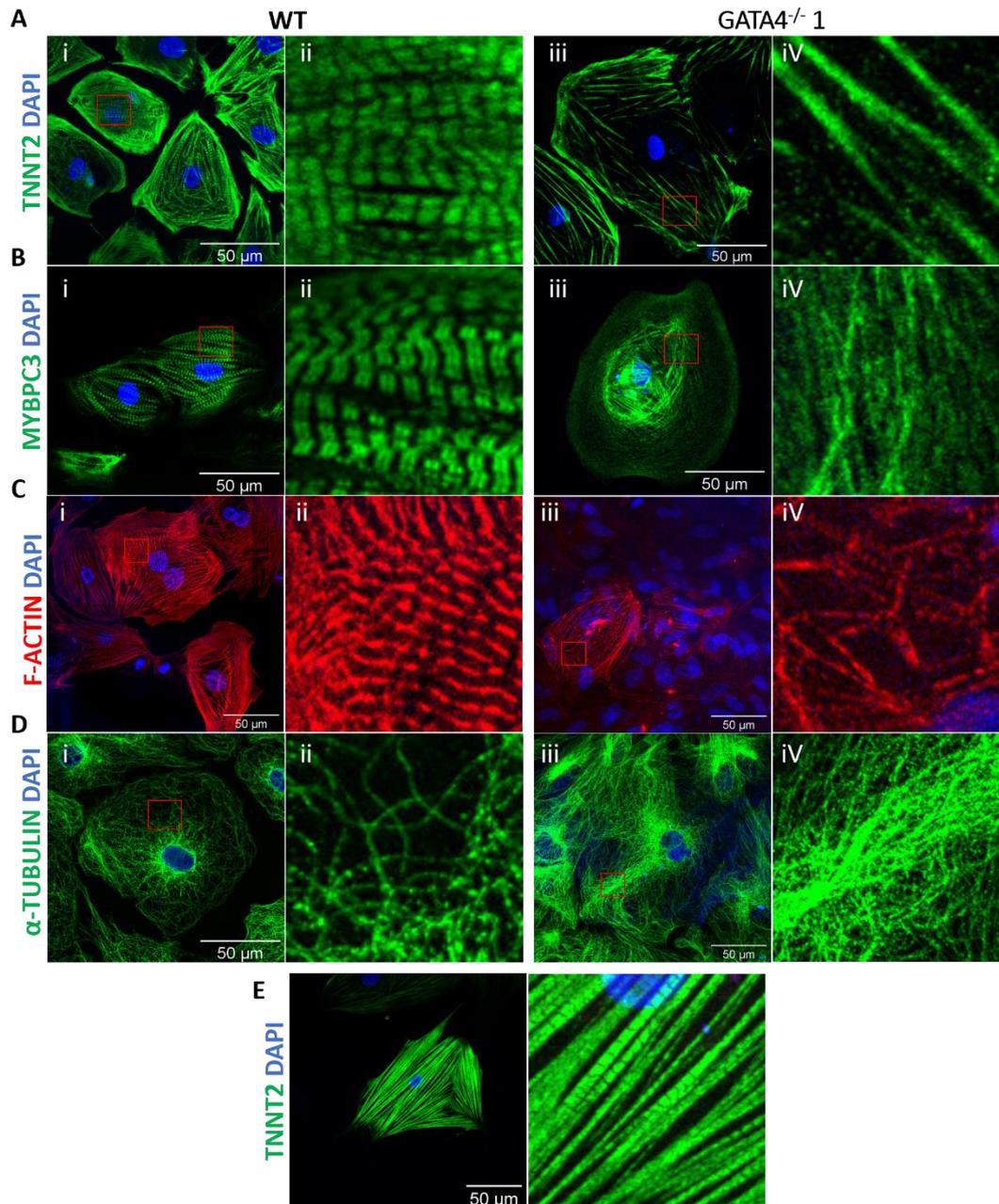


Figure 3.7 Staining for sarcomeric proteins reveals dysfunctional organisation in $GATA4^{-/-1}$ mutants

IF staining was carried out on WT and $GATA4^{-/-1}$ lines at day 32 and images taken at 40x with an oil immersion lens. Panels (A-C) show staining for sarcomeric components; TNNT2, MYBPC3, and Filamentous Actin (F-ACTIN). Panel (D) shows staining for non-sarcomeric cytoskeletal protein α -TUBULIN. Images labelled ii and iv show cropped images enlarged so that any striations present can be more easily observed. In panel (E), a $GATA4^{-/-1}$ cell with some striations is shown, this was the least dysfunctional cell observed among the $GATA4^{-/-}$ cells analysed. A red arrow is used to highlight a fibril with striations, and a white arrow to highlight a fibril without striations.

3.2.8 GATA4 mutant cells produce prolonged calcium transients

In addition to, the expression of CM specific markers and the assembly of sarcomeres another defining characteristic of CMs is the highly regulated calcium flux that drives contraction, which can be detected before the onset of beating (Tyser et al. 2016). Therefore, this was measured to determine if a defect in calcium handling was also contributing to a loss of functionality in the mutants. This analysis was conducted at day 32 and matched samples were used to determine efficiency through IF detection of TNNT2. For the WT cells differentiation efficiency ranged from 47.26-83% (3 biological repeats), from <1%-13% for the GATA4^{-/-}1 mutant (3 biological repeats), and <1-26% for GATA4^{-/-}2 (2 biological repeats). Images were taken using a widefield microscope to capture a field of cells stained with Fluo-8AM to monitor changing calcium levels over time. In the WT cultures there are many patches of cells undergoing calcium flux, whereas few cells are exhibiting this in the GATA4^{-/-} cultures (figure 3.8a). When plotting the change over time for independent patches of cells fast and regularly spaced transients are observed in the WT cells (figure 3.8b), with a full cycle taking 1.31-1.70 secs (n=4 wells). In the GATA4 mutants although beating cannot be detected by eye (see supplemental video 9-11) protracted calcium transients can, taking 8.22-11.55 seconds to reach a peak and return to baseline, and having a lower amplitude than the WT cells (figure 3.8b).

To understand why this may be the expression of two key calcium handling genes was inspected using RT-PCR. In the mutants the expression of *CACNA1C* an L-type Ca²⁺ channel responsible for calcium influx at the cell membrane is lower, as is the expression of *RYR2* which is necessary for Ca signalling amplification through calcium induced calcium release (CICR) from the sarcoplasmic reticulum (figure 3.8c). This may help to explain the slow calcium influx into the cells which was on average 5.95 times longer in the mutants, and the lower overall amplitude of the calcium transients recorded. This dysfunctional calcium signalling likely contributes to the lack of beating detected in the GATA4 null lines.

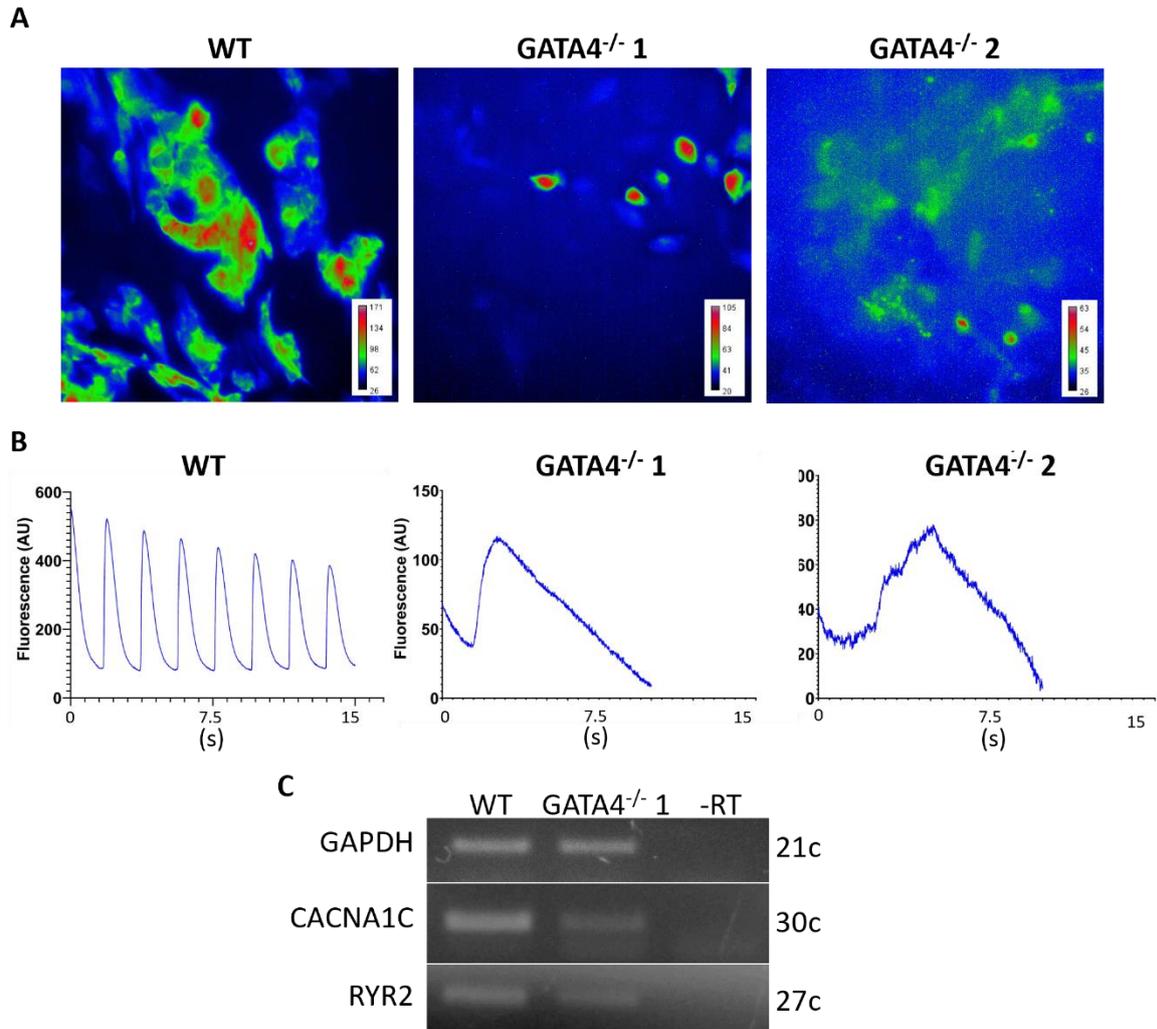


Figure 3.8 Reduced calcium flux and expression of calcium handling genes in GATA4 mutant cells

(A) The fluctuations in Fluo-8AM fluorescence intensity in a field of WT, GATA4^{-/-} 1, and GATA4^{-/-} 2 cells are shown as RGB pseudo-coloured projections. Red areas being pixels with the highest standard deviation/variation over time, and blue being the lowest. (B) Calcium transients plotted over time from selected ROIs for each cell line. For each cell line n= 9, with 3 wells analysed across 3 independent experiments. (C) The expression of calcium handling genes measured by RT-PCR, n = 1.

Both calcium imaging and RT-PCR was conducted on day 32 samples.

3.2.9 The number of mesenchymal cells is increased in the $GATA4^{-/-}$ mutant

The vast reduction in the number of TNNT2 and MYBPC3 positive cells in the mutants, leaves the question of the identity of the remaining cell types. During cardiac directed differentiation multiple other cardiovascular cell types may be formed such as epicardial cells, fibroblasts, endocardial, and smooth muscle cells (Grancharova et al. 2021; Jiang et al. 2021). Therefore, these cell types are the most likely candidates for the remaining cell types. Staining was performed for markers of some of these cell types and compared to the number of cells positive for TNNT2 in the same experiment.

Firstly, in the preliminary experiment analysed the number of TNNT2 positive cells formed from the $GATA4^{-/-}$ iPS cells was very low at <1%, whereas 47% of WT cells were positive for TNNT2 (see figure 3.9c). The remaining cell types were identified through their expression of Vimentin (VIM), Pro-collagen type I (COL1A1), and α -Smooth muscle actin (ACTA2). VIM as a marker identifies a range of non-myocardial cell types including fibroblasts, smooth muscle cells, endothelial cells, and pericytes (Doppler et al. 2017). In WT cultures 65% of cells were positive for VIM, whereas 97% of $GATA4^{-/-}$ cells were positive for VIM (figure 3.9a). Suggesting most of the cells formed from the $GATA4$ mutant had adopted a non-myocardial mesenchymal fate.

Amongst the non-myocardial cells staining for pro-collagen type I was used to identify fibroblasts. The number of cells positive for COL1A1 was similar for both the WT (68%) and $GATA4^{-/-}$ (72%) lines (figure 3.9b). However, it should be noted that other cell types can be a source of collagen deposition under certain conditions, including cardiomyocytes so this should not be considered a definitive marker of fibroblasts (Heras-Bautista et al. 2019). Smooth muscle cells were identified through positive staining for ACTA2. Low ACTA2 expression was also seen in some CMs, and these were not included in the total. 48% of the WT cells stained positively for ACTA2 in comparison to 78% of the $GATA4^{-/-}$ cells (figure 3.9c).

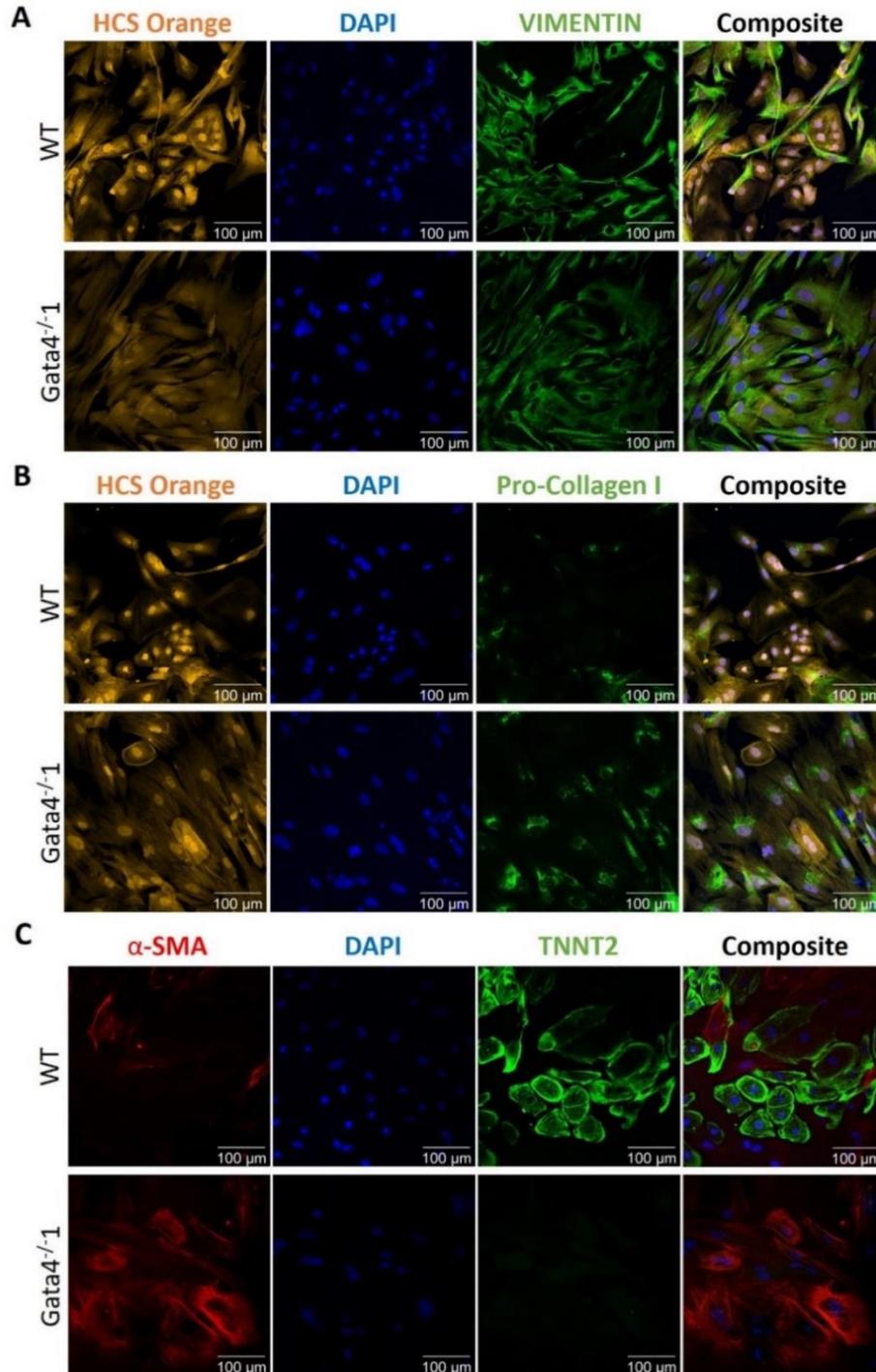


Figure 3.9 Whilst staining for CM markers is decreased in GATA4^{-/-1} cells, staining for mesenchymal markers is increased

WT and GATA4^{-/-1} cells were differentiated to day 32 using the CDM3 cardiogenic differentiation protocol, and then stained by IF. (A) VIMENTIN (green), HCS orange and DAPI (blue) used to counterstain. (B) Pro-collagen I (green), HCS orange, and DAPI (blue) were used to counterstain. (C) ACTA2/ α -SMA (red), TNNT2 (green), and DAPI to counterstain. N = 2.

3.2.10 A sub-set of GATA4^{-/-} TNNT2⁺ cells display abnormal cell membrane morphology

Of the cells present that stained positively for TNNT2 or MYBPC3, a proportion exhibited overall cell morphologies that are not typically associated with CMs. Examples of these morphologies are given in figures 3.10 a-c. *In vivo* CMs exhibit an elongated shape that allows them to stack together (Sever 2000). As has been shown throughout (see figure 3a-b in section 3.2.6) CMs derived from iPSCs vary from *in vivo* CMs considerably, taking on a variety of shapes. However, neither CMs derived *in vivo*, or *in vitro* generally exhibit extensive branching processes as has been observed in some of the GATA4 null CM-like TNNT2 positive cells. In the GATA4 mutants, cells presenting with convoluted or highly convoluted edges were common being observed in 35.2% of the TNNT2 positive cells present, in comparison to 12.5% of TNNT⁺ cells in WT cultures (figure 3.10c). These results demonstrate a majority of the GATA4^{-/-} cells are exhibiting a cellular phenotype associated with non-CM cell types, whilst expressing CM markers such as TNNT2 and MYBPC3 suggesting these cells may have a somewhat unstable cellular identity.

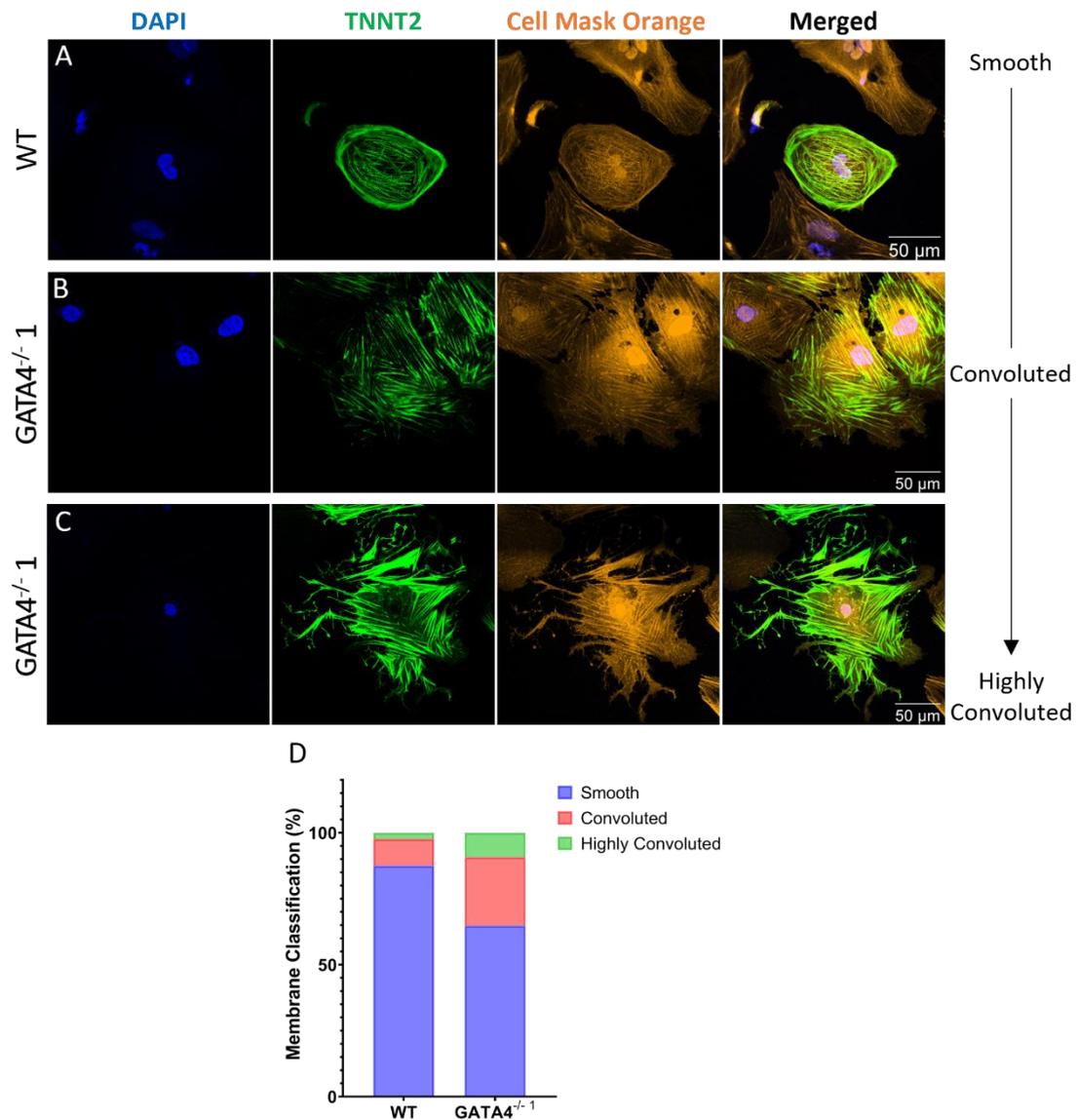


Figure 3.10 Classification of the cell membrane morphologies of WT and GATA4^{-/-1} TNNT2 positive cells

(A) A WT cell displaying a smooth membrane morphology. (B) A GATA4^{-/-1} cell showing a more complex membrane morphology (convoluted). (C) A GATA4^{-/-1} cell with many cell membrane projections (highly convoluted). (D) The membrane morphologies of TNNT2 positive cells were assigned to smooth, convoluted, or highly convoluted classes based on their resemblance to the examples shown in A-C. WT cells analysed = 88, GATA4^{-/-1} = 55, across n = 4 experiments. The proportion of cells in each class were compared using Fisher's exact test, yielding a p-value of 0.0093. All cells were analysed at day 32, imaged at 40x with an oil immersion lens, and were stained for TNNT2 (green), HCS Orange that stains the plasma membrane, nuclear membrane, and cytosol (Orange), and counterstained with DAPI (blue) for nuclei staining.

3.2.11 Expression of cardiac GRN member *TBX5* is disrupted in *GATA4* null cells

In order to determine if there is earlier disruption to the cardiac GRN samples were taken for RT-PCR analysis at earlier time-points throughout development. Currently, due to time constraints, this has only been conducted for the *GATA4*^{-/-}1 line but it will be extended to both lines and additional repeats in future work. In the *GATA4*^{-/-}1 line low expression of *GATA4* was observed from day 2 of differentiation in agreement with the RNA-seq dataset published by (Churko et al. 2018). This stage aligns with the expression of mesodermal marker *TBXT*, and cardiac mesoderm marker *MESP1*, both of which seemed unaffected in the *GATA4* mutant. Endodermal marker *GSC* is transiently expressed during iPSC-CM differentiation and was found to be down-regulated in the mutant.

Cardiac progenitor marker *NKX2-5* is regulated by *GATA4* during cardiac development and acts as a co-activator with *GATA4* in the expression of CM specific genes (Durocher et al. 1997; Brown et al. 2004). The pattern of *NKX2-5* expression was maintained in the mutant, but expression was lower than normal at Day 10 (figure 3.11b). Expression of *TBX5* another member of the cardiac GRN that is essential for normal cardiac development (Horb and Thomsen 1999; Bruneau et al. 2001; Steimle and Moskowitz 2017), was found to be greatly decreased in the *GATA4*^{-/-}1 line throughout the differentiation protocol. These results suggest that the earliest steps in cardiac development; mesoderm and cardiac mesoderm formation proceed normally in *GATA4* null cells but that the expression of cardiac progenitor genes like *TBX5*, and thus cardiac progenitor specification are negatively affected by the absence of *GATA4* expression.

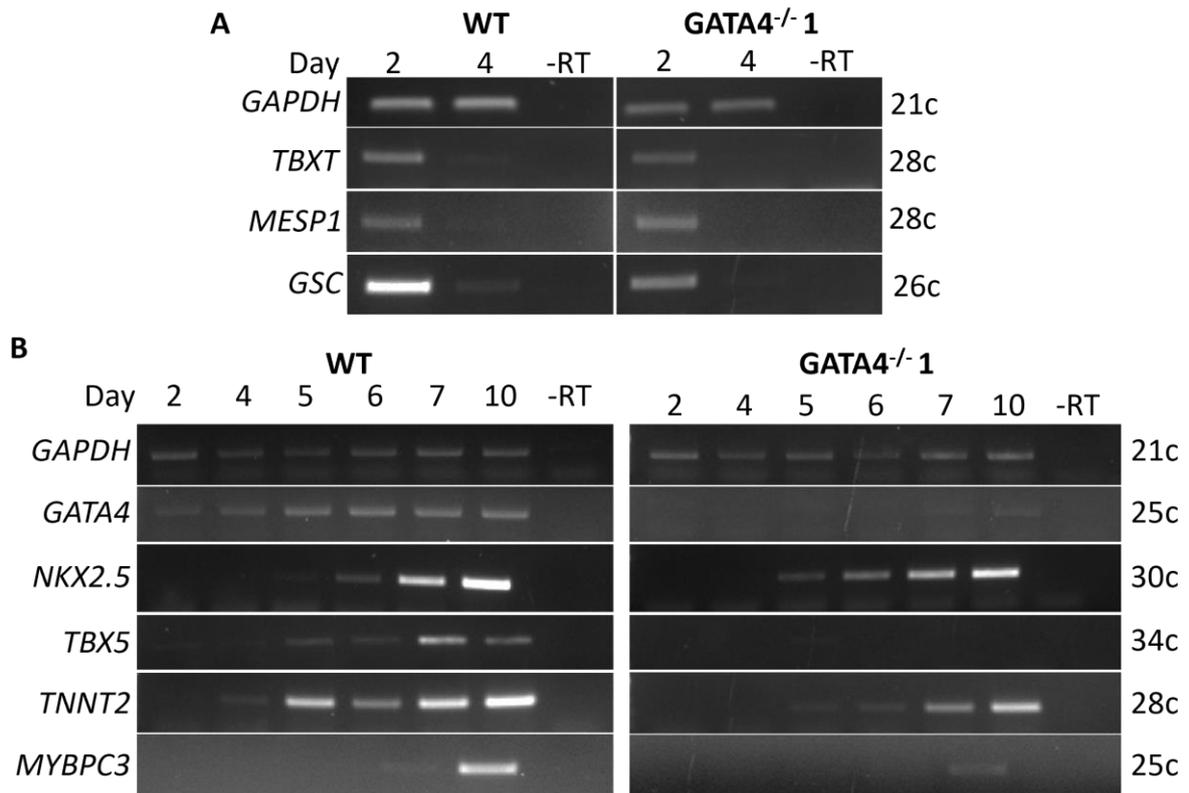


Figure 3.11 RT-PCR for markers of iPSC-CM milestones

(A) Analysis of the expression of *TBXT*, *MESP1*, and *GSC* at days 2 and 4 of differentiation in WT and GATA4^{-/-} 1 mutant cells (n = 2). (B) Expression of cardiac GRN member *NKX2-5* throughout differentiation alongside *GATA4*. Expression of *TNNT2* and *MYBPC3* is shown as well to show the progression of terminal differentiation (n = 2).

3.2.12 The expression of GATA4 target genes *TBX2* and *PRDM1* is altered in the GATA4 mutants

Using *Xenopus laevis* as a model for cardiogenesis our lab identified two target genes of GATA4 regulation, *TBX2* and *PRDM1*. In this model GATA4 is needed for proper induction of *TBX2* expression, and for repression of *PRDM1* (data not published). To determine if these genes are likely targets of GATA4 in human cardiac development their expression pattern was analysed during iPS cell to CM differentiation in comparison to GATA4's. During iPS cell differentiation *GATA4* is expressed at a low level during the earliest stages of cardiac development; corresponding with the establishment of mesoderm, and quickly followed by cardiac mesoderm specification during days 1-3 of differentiation (figure 3.12a-b). Its expression increases at day 4 coinciding with the early cardiac progenitor stage, remaining high and peaking around the time when beating becomes detectable (day 7-12) (figure 3.12a-b). In contrast, *PRDM1* mRNA expression peaks at day 3, and decreases when *GATA4* expression initially peaks at day 4. Conversely expression of *TBX2*, becomes prominent only after the peak in *GATA4* expression (see figure 3.12b). This expression pattern suggests that regulation of these targets by GATA4 is likely conserved in this model.

Further to this, if *PRDM1* and *TBX2* are indeed targets of GATA4 during iPSC-CM differentiation it would be expected that their expression would be affected in the *GATA4*^{-/-} lines described herein. Indeed, in cells lacking GATA4 protein, *PRDM1* expression is higher than normal at day 4 when its expression should be repressed and correspondingly at day 10 *TBX2* expression is lower in the mutant cells. This data is consistent with the hypothesis that *PRDM1* is a negative and *TBX2* a positive target of GATA4 transcriptional regulation.

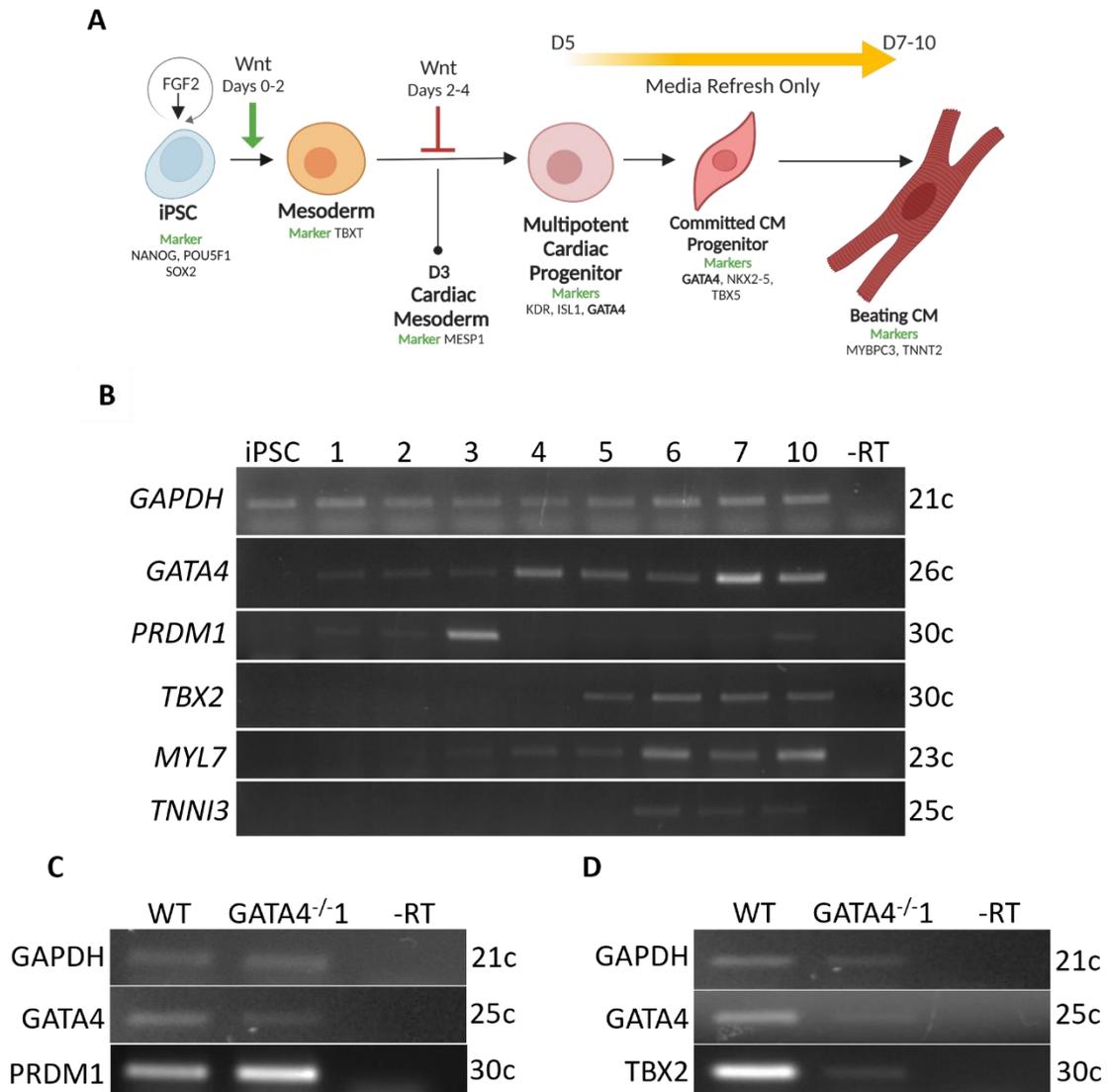


Figure 3.12 The dynamic expression of *PRDM1* and *TBX2* during iPSC to CM differentiation is changed in the absence of *GATA4*

(A) An overview of the milestones of iPSC to CM development and some markers expressed at relative stages. (B) Expression of key markers of differentiation progression throughout differentiation are shown. *GATA4* expression is highest during the cardiac progenitor stage, coinciding with decreasing *PRDM1* expression, and increasing *TBX2* expression. N = 3 (C) RT-PCR data showing the expression of *GATA4* in mutant cells at day 4 is reduced, and the expression of *PRDM1* is higher. (D) RT-PCR for day 10 samples showing *GATA4* expression is lower in mutants, as is *TBX2*, n = 2. *GAPDH* has been used as a normalisation control.

3.3 Discussion

The results presented in this chapter show that GATA4 is essential for the formation of functional CMs from iPS cells. Previous work in mouse, *Xenopus*, and zebrafish models has established integral roles for GATA4 during CM development (Watt et al. 2004; Zeisberg et al. 2005; Holtzinger and Evans 2007; Haworth et al. 2008; Borok et al. 2016; Sam et al. 2020) the effect on cardiac development is generally more profound when more than one GATA factor is removed (Holtzinger and Evans 2007; Haworth et al. 2008; Sam et al. 2020). In murine development specifically a block to cardiac differentiation is only observed when GATA4 and GATA6 are both knocked out, however GATA4^{-/-} GATA6^{-/-} mESC lines can still generate beating EBs at a far reduced efficiency (Zhao et al. 2008). Therefore, the results presented here are somewhat unexpected, as although the cells formed from the GATA4^{-/-} lines display some elements of a differentiated phenotype such as the expression of CM specific genes in a reduced proportion of cells present, they reproducibly do not beat. These results stand in contrast to a recent report by Gonzalez-Teran *et al.* 2022 of another GATA4^{-/-} iPS cell line in which it was shown that although deficits in differentiation efficiency were present beating CMs could be formed after a delay. Further characterisation of the GATA4^{-/-} lines presented in this thesis demonstrates major phenotypic deficiencies that provide a strong biological basis for the lack of beating in these lines. These results provide insights into the requirement for GATA4 in a human model of cardiac development, that are likely applicable to aspects of cardiomyocyte development *in vivo*, and this is discussed below.

3.3.1 The inability of GATA4^{-/-} cells to beat is underscored by low differentiation efficiency and defective sarcomere formation

Two factors identified during the project that likely explain the failure of GATA4^{-/-} cells to form beating CMs are reduced differentiation efficiency and defective sarcomeres formation. Reduced efficiency in this study was determined by the presence or absence of TNNT2 expression by IF. TNNT2 is a pan-myocardial gene that is required for beating and proper sarcomere formation (Sehnert et al. 2002; Nishii et al. 2008), which is routinely used as an indicator of iPSC-CM differentiation efficiency (Lian et al. 2012; Burridge et al. 2014). Here, the highest number of TNNT2⁺ cells observed in a GATA4^{-/-} line differentiation was 35%, which is not far from the average efficiency of ~35-40% observed by Gonzalez-Teran *et al.* in their GATA4^{-/-} iPS line/s. However, even in this differentiation where the number of TNNT2⁺ cells was relatively high no beating was detected. This failure to beat was highly

reproducible being observed across 12 independent differentiations which produced 139 technical repeats for GATA4^{-/-} 1, and 66 for GATA4^{-/-} 2. Closer inspection of the myofibrils formed in the mutant cell lines revealed that very few (6.33%) of the TNNT2⁺ cells present contained any form of striations that would indicate the presence of properly formed sarcomeres an essential requirement for beating, therefore explaining the lack of contraction in these cells.

In the cells that do express TNNT2, deficient sarcomere formation might be in part due to reduced expression of TNNT2 in individual cells. The level of TNNT2 expressed per cell was determined using IF and was found to be 2-2.5 fold lower on average in the TNNT2⁺ GATA4 null cells in comparison to TNNT2⁺ WT CMs. This result demonstrates that expression of TNNT2 is less robust in GATA4^{-/-} cells that have acquired a CM-like fate. Further evidence that these CM-like GATA4 null cells are less robustly activating the cardiac gene program comes from analysis of another pan-myocardial sarcomeric marker, MYBPC3 in these cells. The number of MYBPC3⁺ cells detected by IF was consistently lower than TNNT2 for the same differentiation, with the proportion of MYBPC3 positive cells ranging from 0-5% in the mutants in comparison to the 7-35% range seen for TNNT2. The discordance between the two markers was not statistically significant. However, as these are both pan-myocardial markers, it would be expected that within the same experiment that the same proportion of cells will stain positively for each marker. That they do not match suggests that a proportion of the TNNT2⁺ Gata4^{-/-} cells could be failing to activate expression of MYBPC3. This in addition to reduced levels of TNNT2 per cell indicate an overall reduction in the robustness of cardiac GRN activation in the cells due to the absence of GATA4.

Further to the above, the discrepancy between the two markers could be a true observation or may be due to differences in antibody sensitivity. To clarify these results and address this issue further IF should be conducted with an antibody against a different MYBPC3 antigen, that will also preferably allow for dual staining to be conducted, so that TNNT2 and MYBPC3 levels can be compared within the same cell. If confirmed this may be an example of different levels of dependency on GATA4 mediated transcriptional regulation between the two genes, despite both being bound by GATA4 in foetal and adult mouse hearts (Akerberg et al. 2019). This brings about the question of whether there are other CM

components that have a higher dependency on GATA4 regulation and are therefore missing from the GATA4 null lines which is subsequently contributing to the non-functional stick like myofibrils that were observed in these cells. This is considered further below.

3.3.2 Does the decrease in TNNT2 and MYBPC3 expression extend to other sarcomeric markers?

It should be noted that MYBPC3 is not essential for sarcomere assembly: in MYBPC3 Loss of function mutant mice sarcomeres are still formed and mice are viable but exhibit significant levels of cardiac hypertrophy (Harris et al. 2002; Seeger et al. 2019), consistent with MYBPC3's role as a modulator of cardiac contraction rather than an integral sarcomeric component. Therefore, reduced expression of this protein should not have such a pronounced effect on sarcomere condition as may be the case here. There are numerous proteins that contribute to the structural integrity of the sarcomere (Boateng and Goldspink 2008). It would be interesting to extend the analysis conducted here to other sarcomeric genes to determine which components are reliant on GATA4 regulation, and therefore responsible for the phenotype observed. This has been partially assessed here using RT-PCR, consistent with *MYH6*, *MYH7*, and *TNNI3* as positive targets of transcriptional regulation by GATA4 (Molkentin et al. 1994; Hasegawa et al. 1997; Latinkic 2004; Tomoya Sakamoto 2022), these genes were down-regulated in the GATA4^{-/-} 1 line. Therefore, the absence of beating may be a consequence of a failure to activate these genes in the TNNT2+ CM-like cells. Another viable method for addressing this issue is to subject the samples to scRNA-seq. This technique allows for inter- and intra-cell differences in gene expression to be resolved for TNNT2 and MYBPC3, and for the extension of this analysis to all CM components.

There were exceptions to this trend with sarcomeric gene *TNNI1* expressed at a comparable if not higher level in the GATA4^{-/-} line at day 10. Thus, this gene is seemingly independent of GATA4 regulation at this stage. Comparison of the genes analysed by RT-PCR with the extensive cardiogenic TF CHIP-seq data set generated by Akerberg *et al.* 2019 shows that GATA4 binds to the promoter region of all other sarcomeric genes analysed in foetal and adult mouse hearts with the exception of *TNNI1*, although GATA4 binding is seen at intronic and intergenic sites surrounding the gene. It is well known that TFs can regulate their targets from a distance and motifs do not need to be present in proximal promoter regions for regulation to take place, but this pattern is interesting nonetheless and is a

possible mechanism for the differential regulation seen. The promoter region of *TNNI1* is usually bound by alternative cardiogenic TFs such as; *MEF2A/C*, and *TBX5*, it is therefore possible that one or more of these factors is sufficient for transcription of *TNNI1* at the near WT levels observed in the *GATA4* null lines generated herein.

3.3.3 Sluggish calcium handling is likely a cause of contractile insufficiency and may be indicative of a fate change

In CMs contraction is coupled to calcium flux, and observable calcium flux precedes physical contraction *in vivo* and *in vitro* (Tyser et al. 2016; Churko et al. 2018). In the *GATA4* null lines calcium flux was observed but was seen in far fewer cells and those cells that did produced long protracted transients with a lower amplitude than normal. These same cells were imaged by brightfield microscopy before collecting calcium imaging data and no beating was observed (see supplemental videos S9-S14). These results may indicate reduced expression of calcium handling genes in any *TNNT2*⁺ CM-like cells produced from the *GATA4* null lines, which would follow the results observed for the other CM markers. In support of this *GATA4* is known to regulate the expression of calcium handling gene *NCX1* (Cheng et al. 1999; Koban et al. 2001), and the splicing of various calcium handling genes including *RYR2* and *CACNA1C* (Zhu et al. 2022). However, RT-PCR demonstrated only a modest decrease in *RYR2* and *CACNA1C* expression in the overall population of cells produced from the *GATA4*^{-/-} lines.

Alternatively, the reduction in the proportion of cells experiencing calcium flux may be due to the reduced differentiation efficiency of the *GATA4* null lines, whilst the low amplitude protracted transients exhibited by some cells may be emanating from non-CM cell types that also express genes such as *RYR2* and *CACNA1C*, explaining why their expression levels are somewhat normal. For the cells used in this analysis differentiation efficiency was determined in matched samples by IF staining, the number of *TNNT2*⁺ cells present ranged from 0-26% for the mutants. Therefore, few CM-like cells were present in the population. A preliminary experiment to examine the expression of markers of alternative non-CM cell fates in the mutant lines, revealed that a large majority of the cells present stained positively for smooth muscle marker *ACTA2* (Aikawa et al. 1993). SMCs also exhibit calcium transients, they are generally slower than those seen in CMs and can show large variations in length, form, and regularity (Berridge 1997; Kotlikoff 2003; Iyer et al. 2015). Therefore, it is likely these protracted transients are originating from some of many

SMCs present in the GATA4^{-/-} cultures. Further repeats will be required to demonstrate this conclusively. To achieve this and to match calcium handling properties with cell identity, calcium imaging should be complemented by staining for CM and SMC specific markers on the same group of cells.

3.3.4 Alternative cell fates in GATA4^{-/-} lines and the potential for aberrantly specified cell types

As discussed, the number of TNNT2⁺ CM-like cells produced from the mutant lines was reduced, and staining for other cell type markers revealed an increase in non-CM cell types. This preliminary experiment only includes one repeat and therefore must be interpreted with caution. Nevertheless, these results indicate a potential fate change in the GATA4^{-/-} cell line. Vimentin (VIM) was used as a broad marker of mesenchymal cell types which includes fibroblasts, SMCs, endothelial cells, and pericytes (Doppler et al. 2017). These cell types are also found in the human heart and follow the same early developmental pathway as CMs (Kattman et al. 2006; Moretti et al. 2006; Litviňuková et al. 2020). Therefore, unsurprisingly they are usually found at low levels following iPSC-CM differentiation (Cui et al. 2019; Jiang et al. 2021) In the GATA4^{-/-}1 mutant 96.97% of the cells were positive for VIM, indicating that almost all cells present have acquired a mesenchymal fate. ACTA2 was used to identify smooth muscle cells in the population, and these made up 78% of the cells present in the null line, indicating the large majority had acquired a SMC fate. It was thought a proportion of the remaining mesenchymal cells would likely be fibroblasts therefore Pro-COL1A1 was used to identify these cells based on fibroblasts being the primary source of extracellular collagen deposition in the heart (Cowling et al. 2019). However, the number of Pro-COL1A1 expressing cells was higher than anticipated in both WT and GATA4 mutant cells. Other cell types can be a source of collagen deposition under certain conditions, including SMCs and CMs (Schram et al. 2010; Zhu et al. 2013; Heras-Bautista et al. 2019) therefore, pro-COL1A1 alone is not a strong enough discriminator and other fibroblast markers will be needed to confirm these results.

In addition to the general switch of cells to a mesenchymal cell fate there was also the observation of cells with properties of more than one cell type in the GATA4^{-/-} lines. These cells stained positively for TNNT2, but their overall cell morphology was more consistent with that of activated myofibroblasts which should be TNNT2 negative. Activated myofibroblasts have a striking stellate shape with dendritic extensions (Hinz et al. 2007; Seo

et al. 2020), making them distinct from the CMs formed from iPS cell differentiation which generally have smooth edges. The presence of these cells suggests that transcriptional state may be unsteady in the *GATA4*^{-/-} cells with the expression of usually mutually exclusive genes. In support of this a recent paper by Robbe *et al.* 2022 demonstrated that *GATA4* physically interacts with repressive chromatin modifier *CHD4* recruiting it to non-CM gene regulatory sites in mouse myocardium. They analysed this further with the deletion of a *GATA* site in the modulatory region of SMC gene *MYH11*, which lead to its upregulation in CMs (Robbe et al. 2022), Misexpression of *MYH11* in CMs has previously been shown to result in sarcomere disarray, and results in a significant decrease in ventricular output (Wilczewski et al. 2018). This emphasises the point that CM function and phenotype is not just a factor of maintaining robust cardiac GRN expression, but also the repression of alternative gene programs. Another example that underscores this point comes from Ang *et al.* 2016 in their study of the *GATA4* G296S mutation that disrupts the physical interaction between *GATA4* and *TBX5* in human iPS cells. Loss of this interaction lead to the loss of *GATA4* and *TBX5* binding at cardiac enhancers, and their relocation to ectopic sites, which was associated with attenuation of the cardiac GRN, and upregulation of ectopic gene programs including endothelial markers. Accompanying defects in physiology and morphology in the *GATA4* G296S mutant CMs was recorded (Ang *et al.* 2016). It follows that the gene programs repressed by *CHD4* and *TBX5* in concert with *GATA4* would be aberrantly expressed in the *GATA4* null mutants presented in this thesis, and that this likely contributes to the phenotype described herein. In future experiments this can be addressed using RNA-seq to determine the overlap between genes differentially expressed in the lines presented herein, and those published by Ang *et al.* 2015 and Robbe *et al.* 2022.

3.3.5 When does the process of cardiomyogenesis start to go wrong in the *GATA4*^{-/-} cells?

In development expression of *GATA4* can be detected at a low-level before mesoderm has been established pre-gastrulation (Pilon et al. 2008) but mesoderm is still established in *GATA4* null embryos (Kuo et al. 1997; Molkenkin et al. 1997). Therefore, it was not expected that this would be disrupted in the mutants. At day 2 corresponding roughly with the mesoderm stage in iPS cell development RT-PCR revealed the expression of mesodermal marker *TBXT* was unchanged in the mutant as was expression of cardiac mesoderm marker *MESP1*. Expression of *GATA4* rises at day 4 of iPS cell differentiation

(Burridge et al. 2014a; Churko et al. 2018) corresponding with the expression of early cardiac progenitor markers such as *NKX2-5* which was expressed in the mutant line at a slightly lower level. GATA4 has been implicated in the activation of *NKX2-5* previously (Brown et al. 2004), but it does receive other positive regulatory inputs at this time from genes such as *MESP1* which was expressed at a level comparable to WT. Even though its expression is lower the presence of this gene indicates that some cardiac specification is taking place. GATA4 and *NKX2-5* are known to physically interact and subsequently act as co-activators to induce the expression of terminal differentiation markers such as *NPPA*, *MYH6*, and *MYH7* (Durocher et al. 1997; Lee et al. 1998). Therefore, loss of GATA4 as a co-activator may contribute to the reduced expression of CM markers seen previously.

3.3.6 Conclusions

The results presented in this chapter build on our current understanding of the requirement for GATA4 in cardiac development, demonstrating that in this iPS cell line *GATA4* is required for cardiac differentiation. The severity of the phenotype observed here confirms previous observations of penetrance differences between mouse and human models (Ang et al. 2016; Sharma et al. 2020). The intricacies of GATA4's role in human cardiac development still requires further exploration and confirmation in other iPSC lines, but the results presented thus far provide exciting new avenues of research to follow up that may inform our understanding of the cardiac GRN, and more broadly how transcriptional regulation affects cell fate. The results also suggest that regulation of *PRDM1* and *TBX2* by GATA4 during cardiomyogenesis is conserved between *Xenopus* and humans. The question remains if these genes themselves contribute to iPS-CM differentiation downstream of GATA4, this is explored in the following chapters

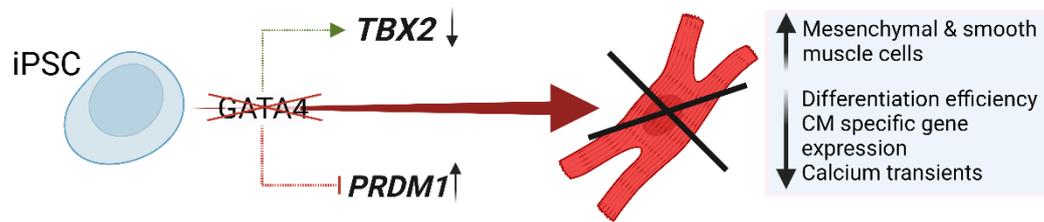


Figure 3.14 A visual summary of the findings of Chapter 3

The results presented in Chapter 3 demonstrate that GATA4 is required for iPSC-CM differentiation. In its absence no beating is seen, differentiation efficiency is severely reduced, CM specific gene expression is decreased, and calcium transients are fewer and slower, whilst the number of mesenchymal and smooth muscle cells is increased. The expression of GATA4 target genes; *TBX2* and *PRDM1* are dysregulated in the GATA4 null lines. However, it remains to be seen whether GATA4 directly regulates these genes in this model.

Chapter 4 –TBX2 is a modulator of the CM differentiation program

4.1 Introduction

The previous chapter demonstrated that GATA4 is needed to generate functional CMs from human iPS cells, and in its absence a number of genes are dysregulated. One such gene is *TBX2* which is downregulated in the absence of GATA4, in agreement with its identification and validation as a positive target of GATA4 regulation during *X. laevis* cardiomyogenesis (Latinkic *et al.* data not published). Positive regulation of *TBX2* in this context suggests it is likely a contributor to the cardiac GRN. However, the current understanding of *TBX2* in cardiac development is limited to its role in the development of non-chamber AVC and OFT myocardium (Harrelson *et al.* 2004; Singh *et al.* 2012). This chapter aims to expand on this knowledge base and identify the possibility of broader roles for this TF in myocardial development by scrutinising the requirement for *TBX2* during iPSC to CM differentiation.

4.1.1 *TBX2* acts as a transcriptional repressor in AVC development

In its established role in the AVC *TBX2* represses the expression of chamber myocardial genes, such as; *NPPA*, *GJA5*, and *SMPX*, as demonstrated in mouse embryonic development by the expansion of their expression into the AVC when *TBX2* is knocked out (Harrelson *et al.* 2004), and their loss from the chamber myocardium when *TBX2* is overexpressed in these areas (Christoffels *et al.* 2004). These results have established *TBX2* as a transcriptional repressor, and biochemical studies by Habets *et al.* provide an explanation for how this is achieved. *TBX2* can compete with *TBX5*, a transcriptional activator, for TBE binding sites in the *NPPA* promoter to repress transcription (Habets *et al.* 2002b). *TBX5* is expressed broadly throughout the cardiac crescent, before becoming restricted to the left ventricle, atria, and AVC in mice by E9 (Bruneau *et al.* 1999; Christoffels *et al.* 2004; Steimle and Moskowitz 2017), whereas *TBX2* expression is restricted to the AVC at this time (Gibson-Brown *et al.* 1998; Christoffels *et al.* 2004; Harrelson *et al.* 2004; Aanhaanen *et al.* 2009). Therefore, it is thought *in vivo* that *TBX2* may displace *TBX5* from active promoters. Additionally, proteomics data from lung tissue demonstrates that *TBX2* associates with an array of repressive chromatin modifiers (Lüdtke *et al.* 2021). Thus, it is probably through a combination of the displacement of activating TFs, and the recruitment

of repressive chromatin modifiers that TBX2 is able to repress transcription. However, it should be noted in endocardial development TBX2 has been demonstrated to activate the expression of genes relevant to endocardial cushion formation; TGFB1 and HAS2, when it is ectopically expressed throughout the heart (Shirai et al. 2009), therefore TBX2 may act as a transcriptional activator in certain situations.

4.1.2 The TBX2 expression pattern during development suggests cardiac activity prior to its role in the AVC

An early step during cardiac development is the coalescence of cardiac progenitors residing in the cardiac crescent to form the linear heart tube (Vincent and Buckingham 2010). In mouse and chick development TBX2 expression can first be detected in the cardiac crescent and linear heart tube, before becoming restricted to the AVC (Yamada et al. 2000; Aanhaanen et al. 2009), suggesting it may have earlier roles in the development of the cardiac progenitor pool. In agreement with this, work in zebrafish has identified a requirement for Tbx2 in promoting proliferation of ventricular CMs in the linear heart tube, with proliferation reduced specifically in this chamber when Tbx2 is knocked down (Sedletcaia and Evans 2011). However, contrasting results have been observed in mice where knockout of Tbx2 has had no effect on chamber myocardium proliferation (Harrelson et al. 2004; Aanhaanen et al. 2011), and in another study ectopic overexpression of Tbx2 in mice from the cardiac crescent stage and then throughout the linear heart tube actually resulted in reduced proliferation affecting all areas of the heart tube (Dupays et al. 2009). This suggests that in contrast to what was observed in zebrafish, Tbx2 in fact has a negative effect on proliferation. Furthermore, constriction of the AVC requires Tbx2 and suppression of proliferation: AVC fails to constrict in Tbx2 mutants (Cai et al. 2005; Ribeiro et al. 2007), adding further evidence that perhaps Tbx2 has a negative effect on proliferation.

The requirement for TBX2 in cardiomyogenesis has yet to be examined in a human model of heart development. During the differentiation of WT iPS cells the expression of cardiac progenitor markers such as ISL1 and GATA4 is seen approximately 24-hours before TBX2 expression starts to rise, also aligning with a rise in the expression of other cardiac progenitor markers such as NKX2-5 and TBX5 (Churko et al. 2018;). This pattern of expression supports the proposition that TBX2 makes an earlier contribution to the cardiac GRN than previously determined from its characterisation in AVC development, and

suggests that iPSC-CM differentiation is a valid model for determining the wider contribution of TBX2 to myocardium development.

4.1.3 Analysing the requirement for TBX2 during iPSC-CM development

In order to examine the requirement for TBX2 in this model two mutant iPS cell lines have been created using CRISPR-Cas9 gene editing to disrupt the DNA binding domain of TBX2. As discussed above, in the AVC myocardium TBX2 represses the expression of chamber myocardial genes and potentially represses early CM proliferation. Therefore, in these experiments mutation of TBX2 may be expected to result in increased expression of chamber myocardial genes and increased CM proliferation, both are factors that may be expected to promote CM differentiation. The results presented herein do not support this hypothesis; instead, they implicate TBX2 as an essential modulator of CM phenotype with mutation of TBX2 having detrimental effects on CM morphology and physiology. RNA-seq analysis has then been utilised to explore early changes in the transcriptional profile of the cells that may drive this dysfunctional phenotype. These results are explored in detail below.

4.2 Results

4.2.1 Editing of *TBX2* locus with CRISPR-Cas9

To examine whether *TBX2* is needed for normal cardiac development, CRISPR Cas9 gene editing was used to mutagenise the *TBX2* locus. *TBX2* is encoded by 7 exons; the gRNAs were targeted to exon 3 (figure 4.1 a) which encodes a portion of the DNA binding domain. The guides were placed 61 bp apart with the aim of creating a frameshifting deletion (figure 4.1 a). PCR screening of gDNA taken from single colony isolates showed the expected ~61 bp drop in PCR band size (figure 4.1 b). Sanger sequencing of these mutants revealed one clone with the desired 61bp frameshifting deletion (*TBX2*^{c.700_761del}), and another with a 61 bp deletion and a 1 bp (T) insertion (*TBX2*^{c.700_761delInsT}) (figure 4.1 c), that leaves the coding sequence in frame but removes a portion of the DNA binding domain.

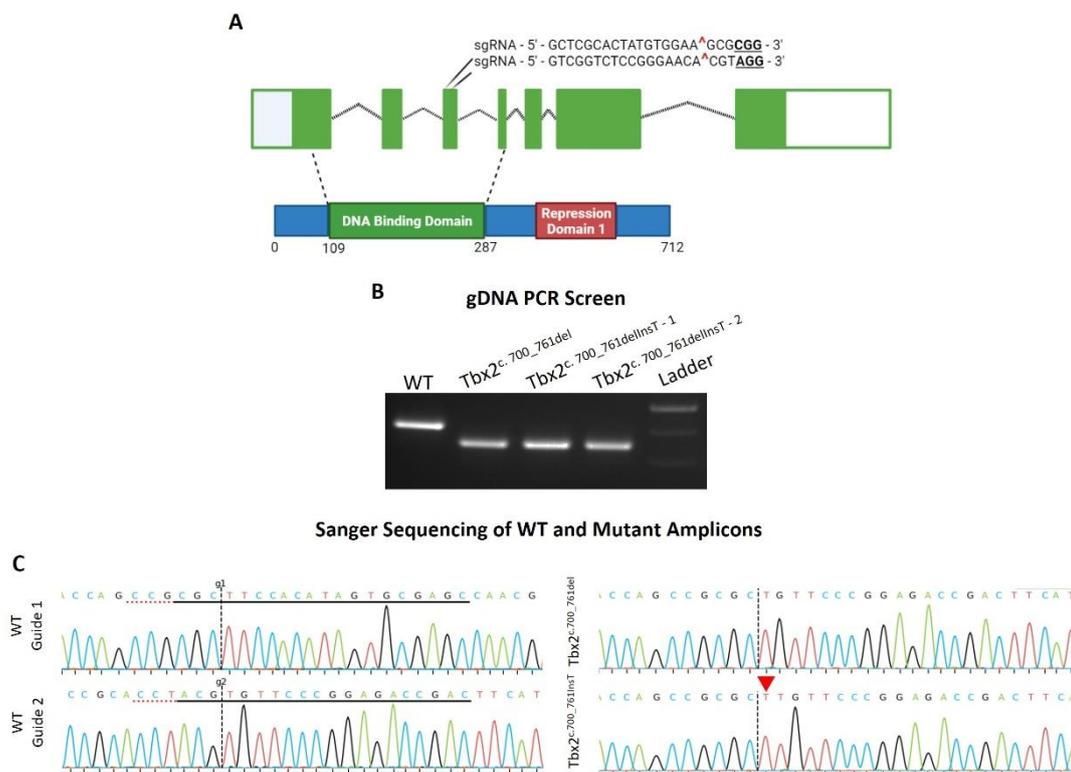


Figure 4.1 Derivation of *TBX2* mutants using CRISPR-Cas9 gene editing

(A) The two guide RNAs for *TBX2* editing were targeted to exon 3, which encodes a section of the *TBX2* DNA binding domain, the exons that contribute to this domain are mapped onto the protein sequence below. (B) gDNA PCR amplicons of the region surrounding the deletion target site, in WT and mutants. (C) Sanger sequencing files showing the sequence surrounding the guide sites for the WT sequence, in comparison to the sequence following editing in the ^{c.700_761del}, and *TBX2*^{c.700_761delInsT} mutants.

4.2.2 TBX2 mutant iPSC cell lines retain pluripotency

Following successful gene editing, it is necessary to ensure the mutant iPSCs can still maintain pluripotency for downstream applications. The TBX2 mutants form rounded compact colonies characteristic of iPSCs (figure 4.2 a). IF demonstrates expression of pluripotency transcription factors POU5F1 and SOX2 is present in the mutants and restricted to the nucleus as expected (figure 4.2 b). Flow cytometry was conducted to quantify this at the population level, the gating strategy used to identify cells stained positively for POU5F1 is shown in figure 4.2 c. This isotype control demonstrates there is a high degree of background staining present, however positive staining can clearly be distinguished as there is little overlap between them. Taking this background into account the flow cytometry analysis shows POU5F1 is expressed in >97% of the population in all cell lines (figure 4.2 d). Furthermore, RT-PCR which gives a lower resolution view of the overall population, shows that *POU5F1*, *SOX2*, and another pluripotency factor *NANOG*, are robustly expressed in the mutants. Together these results strongly indicate pluripotency is maintained in the mutants.

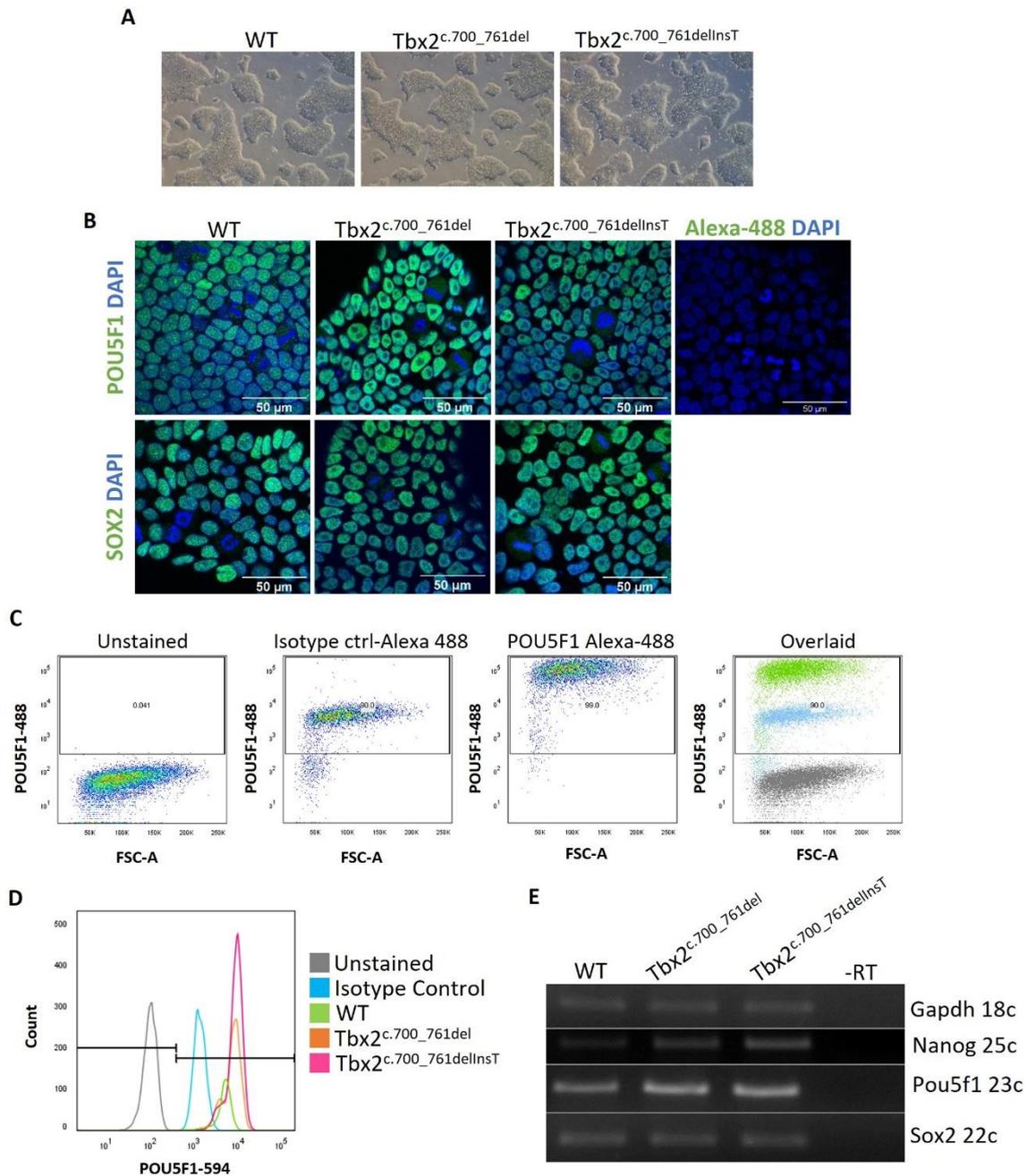


Figure 4.2 The edited TBX2 cell lines maintain pluripotency

(A) Brightfield 4x images of WT and mutant iPSC colonies. (B) Confocal (63x) images of WT and TBX2 mutant iPSCs stained for POU5F1 and SOX2 (green), and counterstained with DAPI (blue). (C) A demonstration of the flow cytometry gating strategy used to identify POU5F1 positive cells. Unstained (grey), isotype control (blue), and positively stained population (green). (D) Histogram showing the proportion of cells staining positively for POU5F1 by flow cytometry for WT and TBX2 mutant cell lines. (E) Expression of pluripotency factors *NANOG*, *POU5F1*, and *SOX2* analysed by RT-PCR. *GAPDH* is used as a normalisation control.

4.2.3 No obvious detrimental changes to genome integrity in *TBX2* edited lines

CRISPR Cas9 induces double strand breaks, which can lead to large chromosomal rearrangements (Tao et al. 2023) . To confirm genomic integrity had been maintained the genomes of the edited lines were compared to that of the WT Rebl Pat parental line using a microarray covering 654,027 markers across the human genome. Microarray processing and analysis was handled by Alexandra Evans at the MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University. This analysis demonstrated that the edited lines were comparable to WT, with the exception of *TBX2*^{c.760-761}, which has an additional deletion on chromosome 13 (table 4.1). This region contains two lncRNAs AL390964 and AL442636, both have no known functional significance (table 4.2).

Table 4.1 A summary of the gDNA loci flagged during analysis of the WT and *TBX2* mutant lines. *TBX2*^{c.760-761} was found to carry an additional mutation (bold) to WT.

Cell Line	Location	Event	Size (bp)	Transcripts from affected region
WT	chr2:139414906-139884468	Duplication	469,563	Contains lncRNA AC016710, pseudogene MRPS18BP2, small nucleolar RNA SNORA72, pseudogene RPL9P13, and lncRNA AC078851.
WT	chr19:20605360-20700135	Deletion	94,776	Contains lncRNA AC010636, pseudogene AC010329, and ZNF626.
<i>TBX2</i> ^{c.700-761del}	chr2:139414906-139884468	Duplication	469,563	Contains lncRNA AC016710, pseudogene MRPS18BP2, small nucleolar RNA SNORA72, pseudogene RPL9P13, and lncRNA AC078851.
<i>TBX2</i>^{c.700-761del}	chr19:20605360-20700135	Deletion	94,776	Contains lncRNA AC010636, pseudogene AC010329, and ZNF626.
<i>TBX2</i>^{c.700-761del}	chr13:5450330-54970922	Deletion	467,593	Contains lncRNA AL390964 and AL442636 - both have no known significance.
<i>TBX2</i> ^{c.700-761delInsA}	chr2:139414906-139884468	Duplication	469,563	Contains lncRNA AC016710, pseudogene MRPS18BP2, small nucleolar RNA SNORA72, pseudogene RPL9P13, and lncRNA AC078851.
<i>TBX2</i>^{c.700-761delInsA}	chr19:20605360-20700135	Deletion	94,776	Contains lncRNA AC010636, pseudogene AC010329, and ZNF626.

Table 4.2 Transcripts affected by the deletion in the *TBX2*^{c.700-761} mutant.

Location	Event	Transcripts	Protein Coding	Function
chr13:5450330-54970922	Deletion	AL390964	No, lncRNA	No known function
chr13:5450330-54970922	Deletion	AL442636	No, lncRNA	No known function

4.2.4 TBX2 is expressed at similar level in WT and mutant iPSC-CMs

To determine the effect of the mutations on TBX2 expression the cells were differentiated using the CDM3 protocol and expression at the RNA and protein level were assayed, using RT-PCR, and Western Blotting (WB) respectively. If TBX2 has indeed been knocked out, a decrease in the RNA signal is expected, and ideally no detectable protein should be seen. However, in the mutants *TBX2* was still robustly expressed at the RNA level (figure 4.3 a-c), whilst WB to determine protein levels in these samples indicates that TBX2 protein may still be produced. However, it should be noted that these WB results are not currently conclusive due to overall low signal intensity and the high amounts of background staining present.

In parallel, the expression of a closely related factor *TBX3* was also analysed by RT-PCR. *TBX3* is co-expressed with *TBX2* in the AVC, where redundancy between the two factors has been noted (Singh et al. 2012) and upregulation of *TBX3* as a compensatory mechanism for reduced *TBX2* activity may be expected (Rossi et al. 2015). This regulatory relationship might operate outside AVC as well. However, our data suggests that *TBX3* expression seems unaffected in the mutant cells.

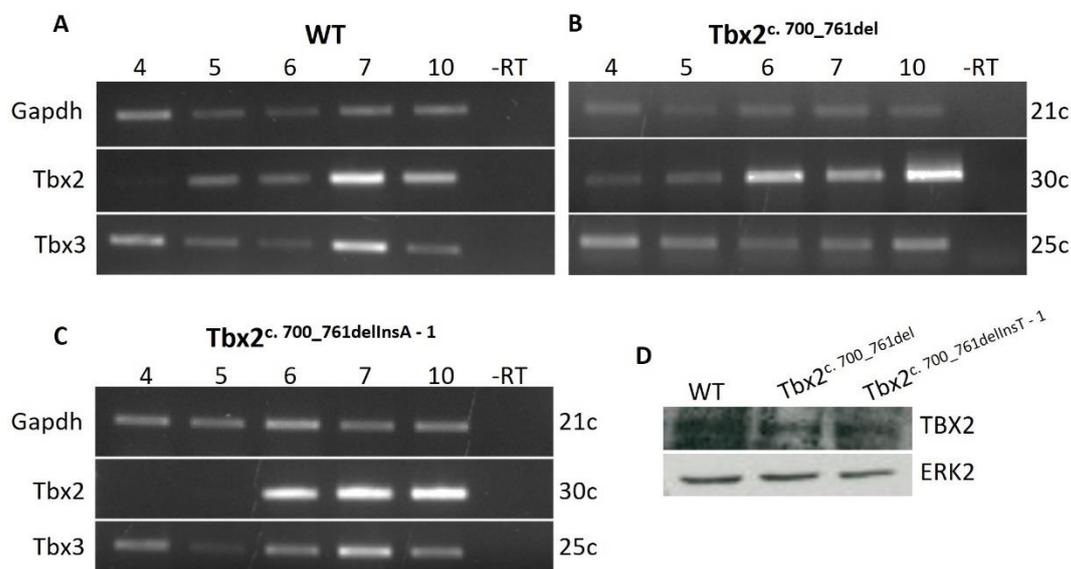


Figure 4.3 Expression of TBX2 and closely related factor TBX3 in WT and mutant iPSC cells during cardiac directed differentiation

(A-C) Expression of *TBX2* and *TBX3* during CDM3 cardiac directed differentiation in WT, *TBX2^{c.700_761del}*, and *TBX2^{c.700_761delInsA}*. (D) Western Blots showing the presence of TBX2 protein in WT and mutant cells at day 10 of a CDM3 differentiation, n = 1. ERK2 is used as a loading control.

4.2.5 Expression of TBX2 is rescued by exon skipping

The gDNA mutation in the *TBX2*^{c.700_761del} mutant should introduce a premature termination codon (PTC) into the coding sequence, that should stimulate nonsense mediated decay of the mRNA transcript (Hug et al. 2015), consequently causing a reduction or complete loss of protein expression. However, the previous results showing unchanged mRNA levels suggest that this may not be the case. Therefore, further analysis of the cDNA being produced was undertaken. RT-PCR amplification crossing the targeted region produces a 589 bp amplicon in WT, whilst in the *TBX2*^{c.700_761del} and *TBX2*^{c.700_761delInsA} mutants 528 bp and 529 bp, respectively.

For the *TBX2*^{c.700_761delInsA} mutant the expected 529 bp amplicon was detected, this was in addition to a lower band. This lower band was also detected in the *Tbx2*^{c.700_761del} mutant as the only mRNA species being produced (figure 4.4 a). The size of this lower band is clearly much smaller than expected therefore, Sanger sequencing was employed to confirm the identity of these bands. This confirmed what was predicted for the 529 bp amplicon present in the *TBX2*^{c.700_761delInsA} mutant and revealed that the lower band carries a 147 bp deletion (figure 4.4 b -c). This deletion encompasses the entirety of exon 3 which recreates the reading frame by fusing exons 2 and 4. The single band present in *TBX2*^{c.700_761del} carries the same exon 2-4 fusion, rescuing expression of TBX2 through exon skipping and creating an amplicon of 442 bp that aligns with the pattern seen using gel electrophoresis (figure 4.4 a).

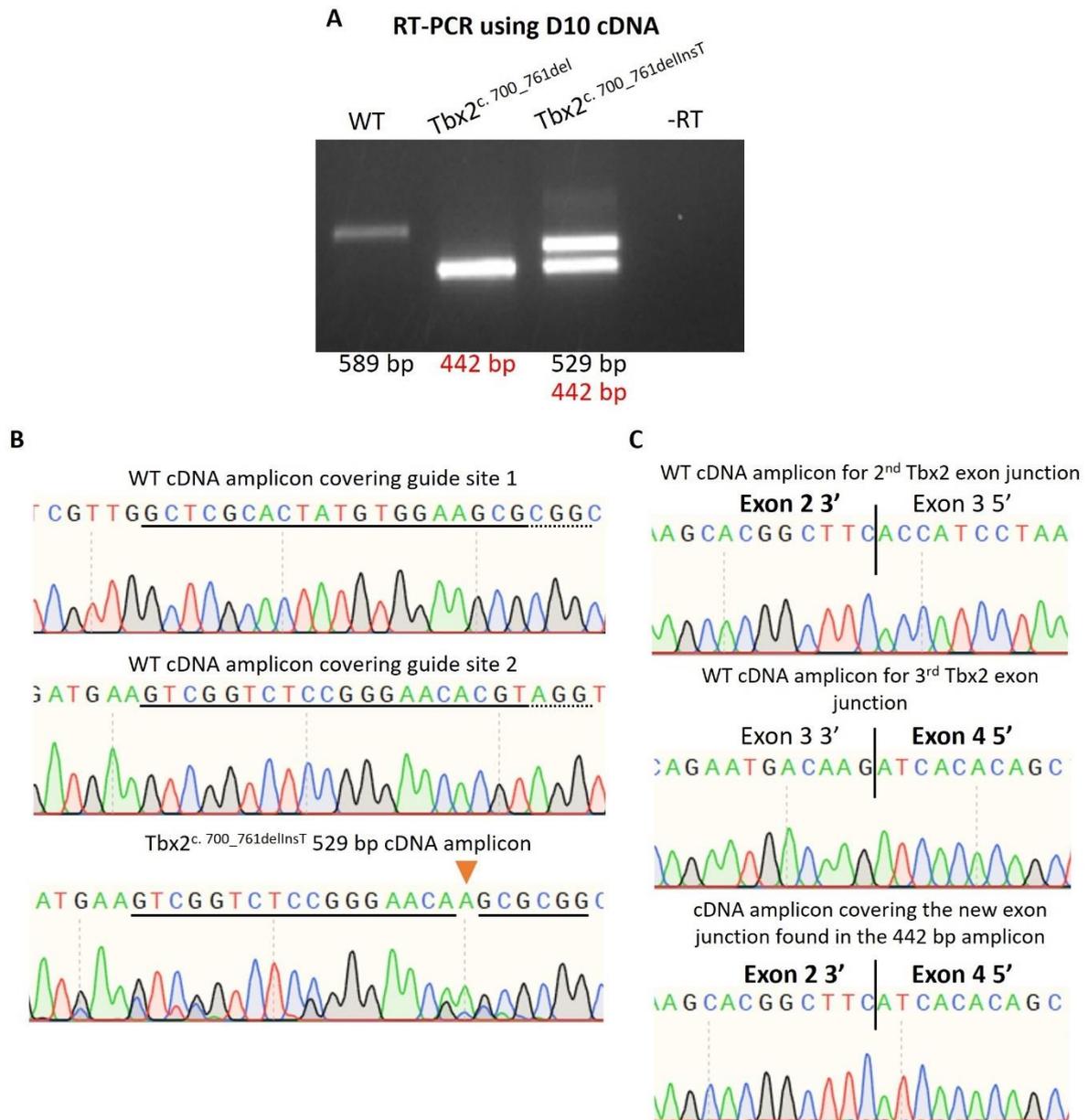


Figure 4.4 Unexpected exon skipping in *TBX2* mutants restores the *TBX2* ORF and maintains expression

A) RT-PCR conducted on day 10 CDM3 samples for WT and mutants, amplifying the region covering the mutation site. The size of the bands (bp) is noted underneath. (B-C) The amplicons formed from the WT and mutant lines for *Tbx2* were gel purified and sent for Sanger sequencing. (B) shows the WT band (589 bp) in comparison to the 442 bp *TBX2*^{c.700_761delInsT} mutant band, CRISPR gRNA binding sites are underlined, and the inserted base present in the mutant is indicated by an orange arrow. (C) Sequencing of the lower 449 bp band found in both mutants shows exon skipping. Exon junctions are indicated by a black line, for the WT and mutant amplicons.

4.2.6 DNA binding domain structure is disrupted in the TBX2 mutant cell lines

TBX2 is still produced in the mutants (figure 4.3 b), however large portions of the DNA binding domain (DBD) will be missing as a result of the mutations introduced and the exon skipping taking place. Structural studies have resolved how the DBD of other T-box TFs such as TBX5, and 3 bind their DNA targets (Coll et al. 2002; Stirnimann et al. 2010). There is no available structure for TBX2's DBD, but for its closest relative TBX3 with which it shares 95.48% DBD homology, the structure has been determined using NMR at a resolution of 1.7 Å (Coll et al. 2002). A comparison of the TBX2 and TBX3 sequences demonstrates this similarity (figure 4.5 a). The dotted red box shows the amino acids missing when exon 3 is skipped (NP_005985.3 p. T222_K270del), and those shown in red and underlined indicates those missing when the 60 bp deletion is present in the TBX2 mutants (NP_005985.3 F234_T253del). The conserved residues in these regions that are known to directly interact with the DNA sequence in the TBX3 model and are indicated by the blue arrows. These directly interacting residues are lost when exon 3 is skipped, but not in the 60 bp deletion. However, the deletion of the residues is still likely to have a major effect on the overall DNA binding domain fold. A surface rendering of the DBD of two TBX3 monomers, binding to a palindromic TBE site is shown in figure 4.5 b. The regions conserved between TBX3 and TBX2 are shown in green, on top of this the deletion regions are shown in red. It can be seen from this that a significant proportion of the DNA binding domain will be lost from the protein whether produced from mRNA carrying a 60 bp deletion, or the exon 2-4 fusion described in figure 4.4.

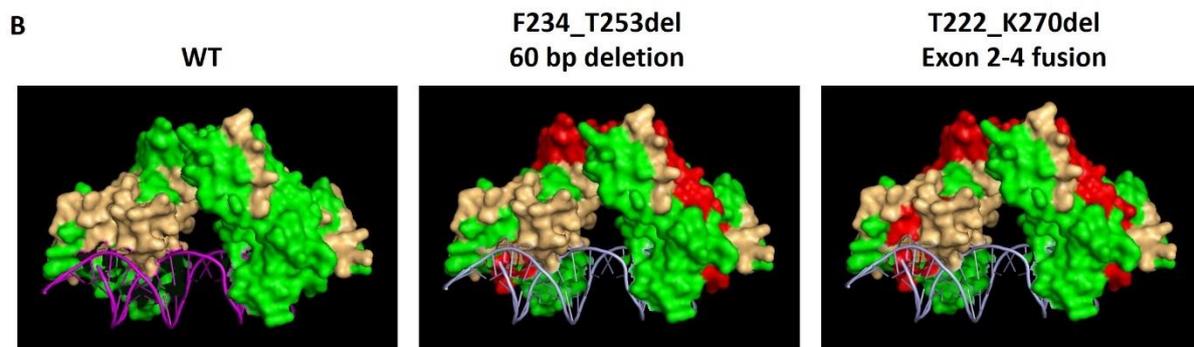
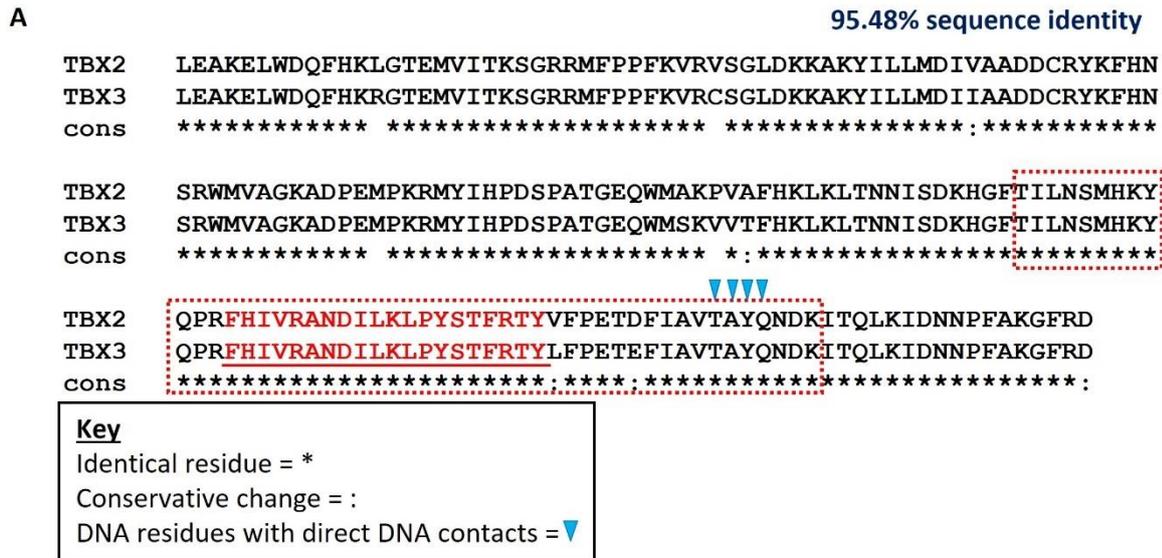


Figure 4.5 A comparison of the DNA binding domains of WT and mutant TBX2 with close relative TBX3

(A) An alignment of the amino acids making up the DNA binding domains (DBD) of TBX2 and TBX3. Non-conservative amino acid changes are indicated by a space. The deleted residues lost when exon 3 is skipped are shown within the red dotted box. The 20 amino acids missing from Tbx2^{c.700_761delInsA} when the 60 bp deletion is present are shown in red and are underlined. (B) The surface structure of the TBX3 DBD bound as a dimer to a palindromic DNA sequence containing T box elements (blue) is shown. Conserved residues between TBX2 and TBX3 shown in green. When mRNA carrying the 60 bp deletion is translated amino acids F234_T253 will be lost from the DBD. When exon 3 is skipped creating a fusion of exons 2-4, amino acids T222_K270 will be lost. These deleted residues are shown in red.

Images produced using PyMol and PDB file 1H6F (Coll et al. 2002).

4.2.7 The *TBX2* mutant proteins are unstable in *Xenopus laevis* embryos

When *TBX2* is uniformly overexpressed in *Xenopus laevis* embryos it causes gastrulation defects resulting in severe axial abnormalities (Cho et al. 2017). This provides an amenable assay for assessing the activity of the *TBX2* mutants created here. If the DBD of the mutants has been rendered ineffective we would expect embryos injected with mutant *TBX2* mRNA to have a milder phenotype, resembling un-injected controls. Embryos were injected uniformly at the 2- or 4-cell stage with WT or mutant *TBX2* mRNAs tagged with hemagglutinin (HA), these embryos then analysed at stage 35. Embryos showing only deformations in the size of the eye or tail were classified as mild, whereas those that had failed to extend their AP axis, and/or showed bifurcation of the tail were classified as severe. A small proportion of the un-injected control embryos showed mild abnormalities 8%, whilst 1% showed severe abnormalities, the rest were considered normal (n = 101) (figure 4.6 a-b). The embryos injected with WT *TBX2* mRNA were severely abnormal in most cases 96.8% ((normal-3.2%, mild-21.5%, severe 75.3%, n = 93). The number of abnormal embryos seen in those injected with mutant *TBX2* mRNA was slightly higher than in the un-injected controls, however in contrast to WT *TBX2* mRNA a large proportion of the embryos injected with h*TBX2*^{c.700_761del} mRNA were normal (normal 88.7%, mild 10.3%, severe 1%, n = 97), as were those injected with h*TBX2*^{c.700_761delInsA} mRNA (normal 79.6%, mild 12.2%, severe 8.2%, n = 49) (figure 4.6 c). This suggested that the proteins had indeed been rendered ineffective.

However, WB confirmed that the mutant proteins were not detectable in samples taken at stage 10.5, suggesting the results above are simply a consequence of degradation of the mutant proteins. Further examination of this revealed that the mutant proteins can be detected in stage 8.5 embryos which is around the time of mid-blastula transition (MBT). Therefore, it seems that mutant proteins are translated and accumulate prior to MBT but shortly after MBT when zygotic transcription is initiated the mutant proteins are degraded. To further examine the fate of the mutant proteins the injected embryos were analysed at stage 9 by IHC using an anti-HA tag antibody to specifically detect injected proteins rather than endogenous expression. Staining intensity was overall lower in those injected with mutant proteins, and where staining was present it was more diffuse, in contrast to the clear spots seen in WT injected embryos, which is presumed to represent nuclear accumulation of the protein, as would be expected for a TF (figure 4.6 e). That only a low level of diffuse

staining is observed in embryos injected with *TBX2* mutant mRNA suggests the mutant proteins are being degraded and excluded from the nucleus and supports what was seen by WB and the characterisation of these mutations as loss of function mutations.

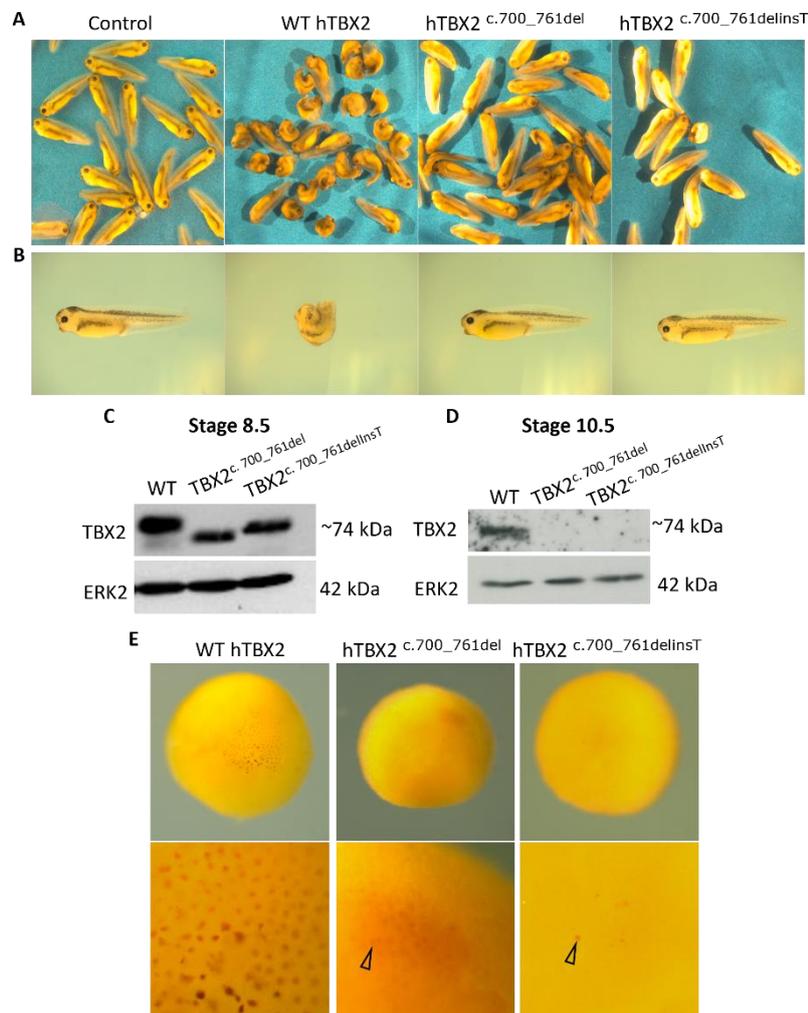


Figure 4.6 Embryos injected with mutant *TBX2* resemble un-injected controls

(A) 2x Brightfield images showing stage 35 *Xenopus laevis* embryos, injected uniformly at the 2-4 cell stage with a WT *TBX2* (n = 93, across n = 2 experiments), *hTBX2*^{c.700_761del} (n = 97), or *hTBX2*^{c.700_761delInsA} mRNA (n = 49), in comparison to un-injected controls (n = 101). (B) Single pictures displaying representative example embryos from each treatment group. (C) WB showing the expression of WT and mutant *TBX2* in samples taken at stage 8.5. (D) WB showing the expression of WT and mutant *TBX2* in samples taken at stage 10.5. ERK2 is used as a loading control for C and D. (E) Immunohistochemical staining in stage 9 embryos using an anti-HA antibody to detect HA tagged WT and mutant *hTBX2* proteins. N = 1.

Embryo injections and IHC staining was conducted by Dr. Branko Latinkic.

The mRNAs for injection were prepared by Dr. Pavel Kirilenko.

4.2.8 TBX2 mutant cell lines produce functional cardiomyocytes

The effect of the mutations characterised above on the formation of CMs was assessed using the Wnt modulation differentiation protocols described in (Burrige *et al.* 2014, and Lian *et al.* 2012). The expression profile of *TBX2* suggests it is not necessary for the onset of differentiation, as it rises after the expression of cardiogenic TFs such as *GATA4*, and after the expression of some sarcomeric genes is established (Churko *et al.* 2018). In agreement with this beating cells were formed from the mutants and through brightfield microscopic analysis of the cells no obvious differences in appearance were detectable (figure 4.7 a-b). However, a small difference in beating onset was noted, with the cells formed from the *TBX2*^{c.700_761delInsT} line started beating 0.44 days earlier than WT (figure 4.7 c). These results indicate the *TBX2* mutations have not had a major effect on differentiation rate.

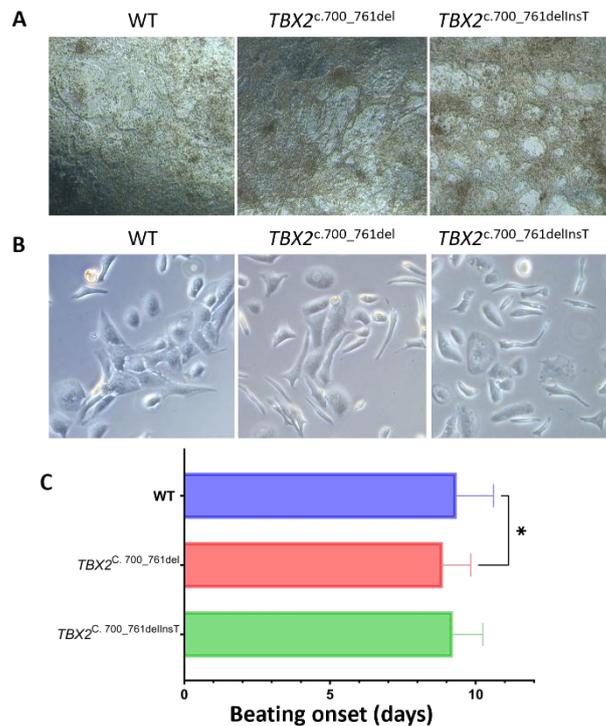


Figure 4.7 TBX2 mutant iPS cells form CMs

(A) This panel shows beating CMs generated using the CDM3 protocol following 10 days of differentiation. Cells are shown at 5x mag. (B) The same cells at day 32 after re-plating and at 10x magnification allowing individual cells to be discerned. Corresponding videos for A-B are available in supplemental data. (C) Quantification of beating onset in WT cells and mutants between days 7-12 of differentiation. The average beating onset for each line was as follows; WT - 9.34 days n = 113 wells, *TBX2*^{c.700_761delInsT} - 8.9 days n = 117 wells, and for *TBX2*^{c.700_761delInsT} - 9.23 days n = 118 wells. Across n = 11 independent repeats. A one-way ANOVA was used for statistical comparison. * = p-value ≤ 0.05.

4.2.9 Differentiation efficiency is seemingly unaffected in the TBX2 mutants but cells appear hypertrophic

As stated above the onset of beating and macroscopic inspection of the cells formed from the TBX2 mutants did not reveal any substantial differences. This is an indication that differentiation is not affected by the TBX2 mutations. Differentiation efficiency was inspected further using IF at day 32. Antibodies against pan-myocardial marker TNNT2 revealed no significant difference in differentiation efficiency between WT and the TBX2 mutants (figure 4.8 a-b). However, from this overview of the population it was apparent that a proportion of the mutant cells were considerably larger than observed in the WT population. Measurement of the cells revealed that on average the CMs formed from the *TBX2*^{c.700_761} and *TBX2*^{c.700_761delInsT} cell lines were 2.27-2.30 times larger respectively than WT CMs (figure 4.8 c). Other aspects of cell morphology that are known to be reflective of CM maturation status and condition were measured such as circularity and the number of nuclei per cell (Bray et al. 2008; Ribeiro et al. 2015; Haftbaradaran Esfahani and Knöll 2020; Karbassi et al. 2020). The mutants were found to be on average 6.75-9.23% more circular than their WT counterparts (figure 4.8 d and table 4.3). However, no significant difference was seen in the number of number of nuclei per cell (table 4.3).

To determine when these differences become apparent these measurements were also recorded for day 12 WT and *TBX2*^{c.700_761} CM samples in a single preliminary experiment. These results revealed that a difference in cell size is apparent at this time, with the TBX2 mutant CMs were found to be 1.56-fold larger than the WT average. The TBX2 mutant cells were smaller at day 12 than there day 32 counterparts, whereas the WT cells were found to be of a similar size at both time-points. This suggests that the *TBX2*^{c.700_761} become more divergent from the WT line over time (figure 4.8 c).

Table 4.3 Differentiation efficiency is seemingly unaffected by mutation of TBX2, but cell condition is impacted

The values given are averages from n number of cells across 3 biological repeats \pm SEM where applicable. The values given here correspond to the graphs in figure 4.8.

		Cell Line		
		WT	TBX2 c.700_761del	TBX2 c.700_761delInsT
Day 32	TNNT2+ cells (%)	62.00% \pm 7, n=649	66.00% \pm 9.23, n=871	68.33% \pm 12.44, n=448
	Area (μm^2)	2553 \pm 131.10, n = 105	5817 \pm 465.60, n = 82	5886 \pm 346.10, n= 132
	Circularity (AU)	0.726 \pm 0.011, n = 110	0.7746 \pm 0.016 n = 82	0.793 \pm 0.009, n = 140
	Nuclei (Mono-. Bi-, Poly-nucleated (%))	81, 18, and 1, n = 110	79, 20, and 1, n = 82	84, 15, and 1, n = 140
Day 12	Area (μm^2)	2667 \pm 246.70, n = 23	4171 \pm 402.90, n = 33	-
	Circularity (AU)	0.655 \pm 0.025, n = 23	0.657 \pm 0.212, n = 33	-
	Nuclei (Mono-. Bi-, Poly-nucleated (%))	90, 10, and 0, n =23	94, 6, and 0, n =33	-

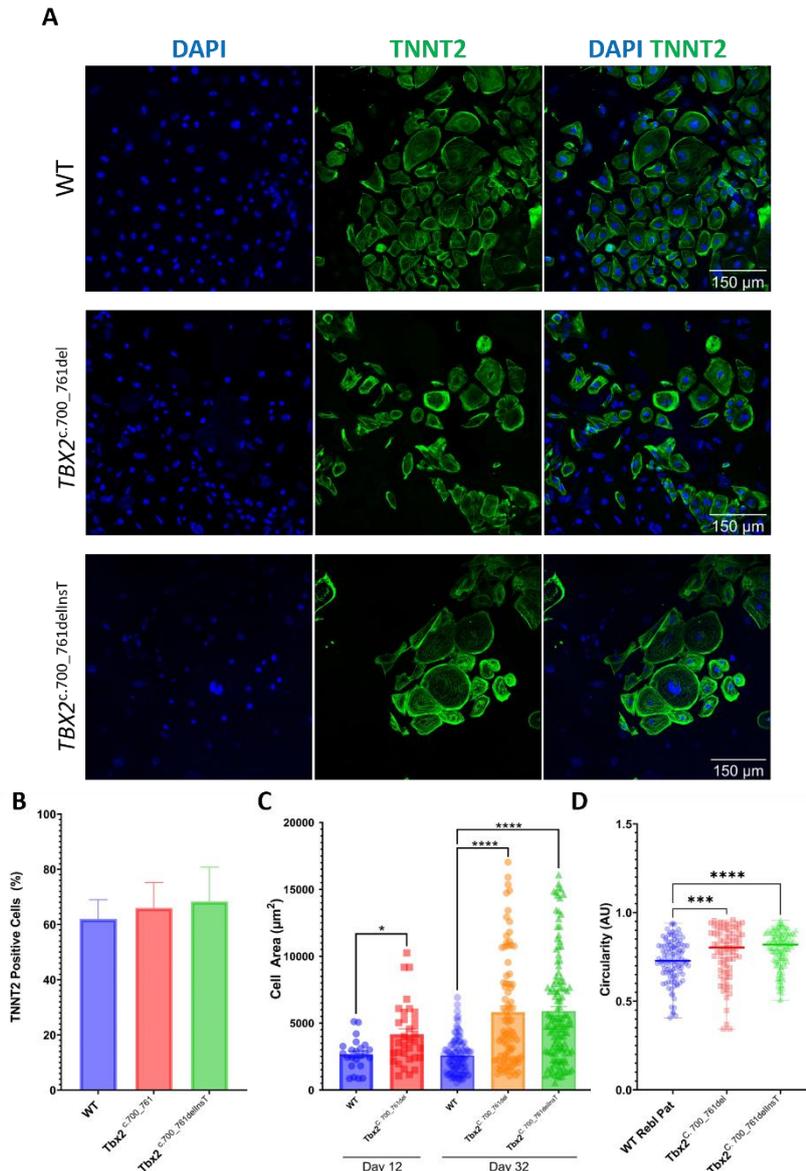


Figure 4.8 TBX2 mutant iPSCs form CMs at a similar efficiency to WT but display hypertrophy

(A) Representative fields of WT and TBX2 mutant day 32 CMs are shown at 40x magnification. The cells were stained for TNNT2 (green) to identify CMs, and counterstained with DAPI (blue) to identify all cells. (B) Across 3 biological repeats the proportion of TNNT2 positive cells in the total cell population was recorded. The values are given as a percentage \pm SEM and are as follows; WT 62% \pm 7.00, TBX2^{c.700_761del} 66% \pm 9.23, and TBX2^{700_761delInsT} 65% \pm 12.66. (C) The surface area of the CMs was measured across 3 differentiations at day 32 and 1 differentiation at day 12. (D) Circularity measurements for the day 32 cells (n = 3). For figures B-D the number of cells analysed is given in table 4.2. A one-way ANOVA has been used for all statistical comparisons shown. P-values; * \leq 0.05, *** \leq 0.001, and **** \leq 0.0001.

4.2.10 Fibril alignment and sarcomere formation is abnormal in TBX2 mutant cardiomyocytes

The above observations reveal the TBX2 mutant CMs to be hypertrophic. *In vivo* this may be considered a physiological adaptive condition in some contexts (Nakamura and Sadoshima 2018). However, closer inspection of the cells revealed that fibrils were often highly disorganised, and the appearance of sarcomeres was also altered, indicating this is a pathological form of hypertrophy. Representative examples of day 32 hCMs generated from the WT and TBX2 mutant lines along with a magnified view of their sarcomeres is shown in figure 4.9 a.

A classification system adapted from Ang *et al.* 2016 that integrates multiple CM phenotypic parameters was used to quantify the overall condition of the cell population. Examples of the cells belonging to each class are given in figure 4.9 c and the criteria for classification are described in the accompanying table. In the WT population a variety of cells belonging to these classifications are present, 59.8% were found to belong in classes I or II which represents the more organised cells in the population. The remainder align with the classification criteria for class III and class IV which represent the disorganised cells in the population (figure 4.9 b), which is expected due to generally immature and heterogenous nature of iPSC-CM cultures (Ang *et al.* 2016a; Mosqueira *et al.* 2018). With this said a shift was observed in the TBX2 mutant cells: the number of organised cells (classes I and II) was just 25.8% in the *TBX2*^{c.700_761del} CMs with the remaining majority found to be disorganised (class III and class IV). A smaller shift was seen in the second mutant (*TBX2*^{c.700_761delInsT}), still more disorganised cells were observed than in the WT with 49.3% of the cells determined to be organised and 50.7% in the disorganised classifications of the CMs. This smaller difference may be of note when considering the *TBX2*^{c.700_761delInsA} mutant retains a larger portion of its DBD. Nonetheless, the observation of an increase in disorganisation in combination with the increase in cell size implies that both TBX2 mutant cell lines produce CMs that exhibit pathological hypertrophy.

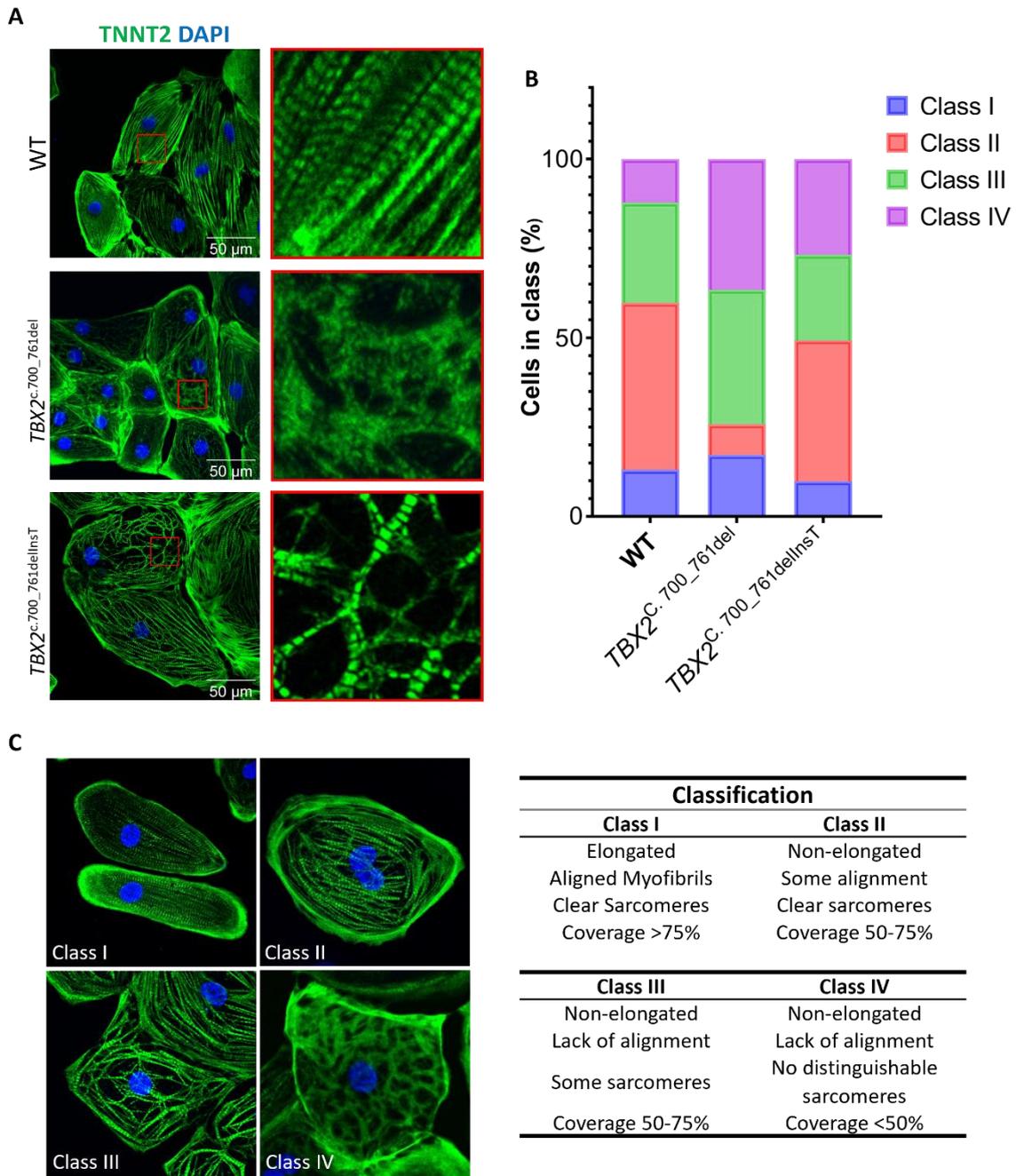


Figure 4.9 TBX2 cells are more often disorganised

(A) An example of the cells formed from WT and TBX2 mutant lines at day 32 of differentiation. The cells are stained with TNNT2 (green), and cell nuclei are counterstained with DAPI (blue). A magnified area of the myofibrils in these cells is also shown. (B) The number of cells that belong in each class was quantified over 3 biological repeat WT $n = 107$, $TBX2^{c.700_761del}$ $n = 93$, and $TBX2^{c.700_761delInsT}$ $n = 142$. A Chi-squared test was used for comparison and yielded a p-value of <0.0001 . (C) Representative example cells for each classification are shown, the associated table gives a breakdown of each classification.

4.2.11 Calcium handling is disrupted in the TBX2 mutant CMs

Calcium flux is coupled to contraction in CMs and is necessary for this function, it is also an indicator of CM health with altered calcium handling associated with multiple cardiomyopathies including cardiac hypertrophy (Crocini and Gotthardt 2021). Therefore, given the previous observations of dysfunction in the TBX2 mutant CMs calcium handling was also examined. Fluo4-AM dye was used to monitor calcium flux in day 32-35 WT and TBX2 mutant CMs, and representative trace files for each are shown in figure 4.10 a-c. The amplitude of the calcium peaks was notably higher in the mutant CMs, and there was a larger range of amplitudes observed indicating that there is a wider heterogeneity in the Ca handling properties of the mutant CMs (figure 4.10 d), which may correlate with the variance in morphological appearance of the cells. Time to peak was also found to be significantly longer in the *TBX2*^{c.700_761del} mutant (figure 4.10 e). FDHM and time to decay were higher in the mutants but not found to be significantly different (figure 4.10 f-g).

Another feature that was noted from visual inspection of the calcium trace files was the appearance of calcium sparks in between transients in WT and mutant cells. Calcium sparks indicate the release of calcium from ryanodine receptor/s in the sarcoplasmic reticulum (SR). These sparks are usually associated with calcium influx into the cell, activating RYR2-mediated Ca release which amplifies the signal creating a calcium transient that is coupled to contraction (Cheng et al. 1993). However, in some circumstances spontaneous sparks may be observed in between transients. Spontaneous sparks have been observed in iPSC-CMs previously and their occurrence was affected by SR calcium load (Zhang et al. 2013). In WT cells sparks were seen in 16.68% (n=18) of the trace files, versus 40% (n = 30) *TBX2*^{c.700_761del} mutant and 31.25% (n=19) in the *TBX2*^{c.700_761del} mutant (figure 4.10H). Despite the observed variances among the lines, these differences failed to reach statistical significance when compared using a Chi-squared test.

Nonetheless, the results presented in this figure demonstrate that aspects of the calcium handling properties of the TBX2 mutants such as amplitude are changed in comparison to the WT cells, such that the TBX2 mutant cells are regularly exposed to higher than normal intracellular calcium levels.

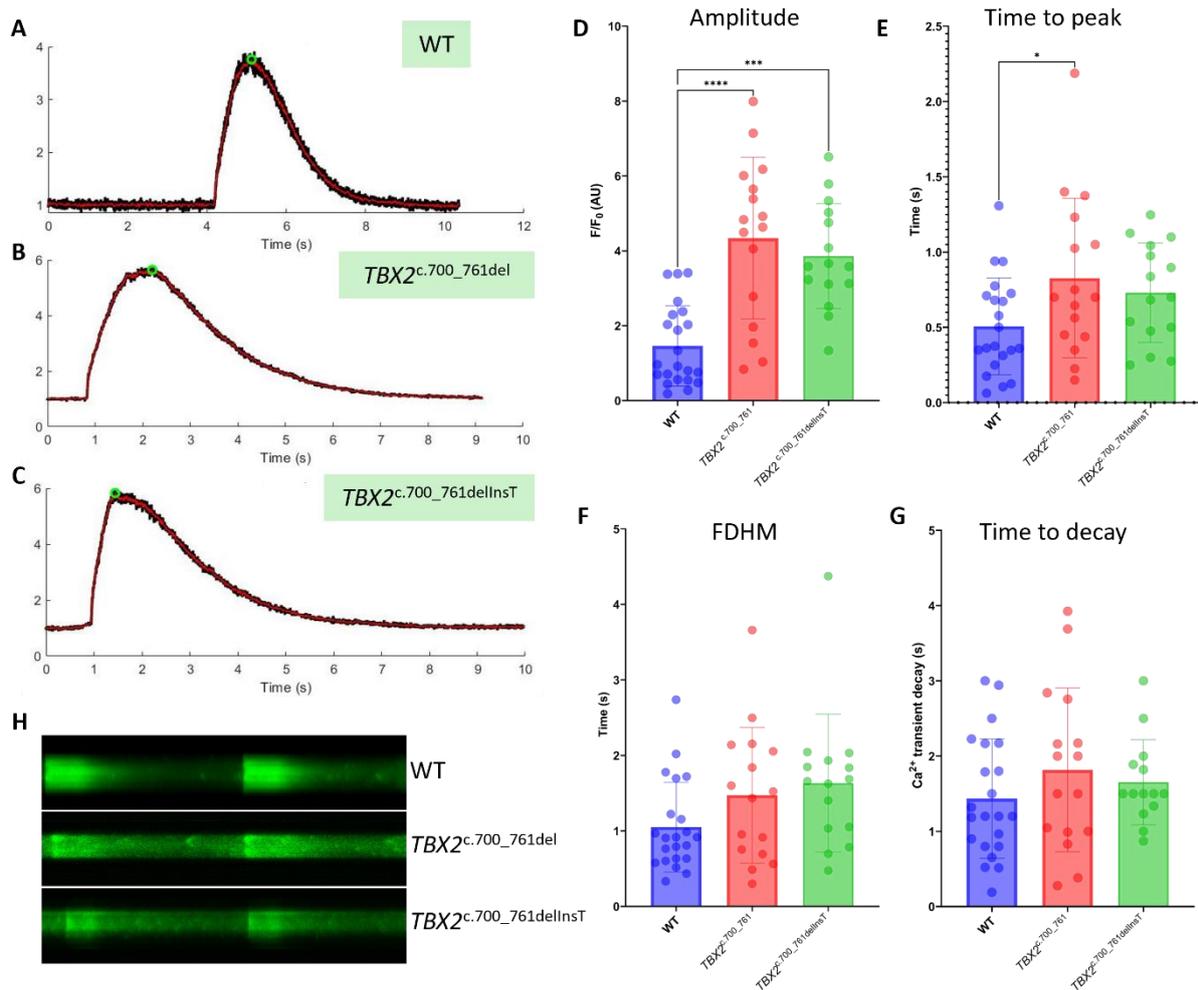


Figure 4.10 Calcium signalling is dysfunctional in the TBX2 mutants

All cells were imaged between days 32-35 of development. Fluo-4AM was used to detect calcium transients in these cells. (A-C) Representative calcium transients for each line are shown and the (D) average amplitude, (E) time to peak, (F) fluorescence duration at half maximum (FDHM), and (G) time to decay for each cell line have been plotted. For figures D-G a one-way ANOVA has been used for comparison and the p-values are as follows; * ≤ 0.05 , *** ≤ 0.001 , and **** ≤ 0.0001 . WT $n = 22$, *TBX2*^{c.700_761del} $n = 16$, and for *TBX2*^{c.700_761del} $n = 15$. (H) These images show fluorescence over time in one region of interest, these images have been subjected to mean filtering of 2 to reduce noise enabling the small ‘sparks’ between transients to be more easily visualised.

The MatLab script used to generate this data and the imaging of the cells for this analysis were generated by Dr. Ewan Fowler, Cardiff University.

4.2.12 RNA-sequencing of TBX2 mutant cells reveals early upregulation of genes associated with cardiac stress

To further explore the phenotype of the cells RNA-sequencing was conducted on samples taken at days 7 and 10. These days align approximately with the peak of *TBX2* expression as observed in (Churko et al. 2018), and the onset of beating (figures 4.3 and 4.7) respectively. Two biological repeats for each cell line were submitted for sequencing. Initially a list of differentially regulated genes with a log₂FC value of ± 2 and a p-adjusted value < 0.05 was generated, to give a list of genes where the changes are relatively large and consistent. With these criteria applied 125 genes were found to be differentially expressed in the *TBX2*^{c.700_761} cell line at day 7, versus 26 in the *TBX2*^{c.700_761delInsT} line, 17 of these genes were found to be co-regulated between the lines (figure 4.11 a). Similarly, at day 10 using the strict criteria described above few genes were found to be differentially regulated; 33 differentially regulated genes were identified in the *TBX2*^{c.700_761} cell line versus 7 in the *TBX2*^{c.700_761delInsT} line, 3 of which were co-regulated in both lines (figure 4.11b).

The heatmaps below display the genes differentially regulated in the *TBX2*^{c.700_761} line, and their expression in WT and *TBX2*^{c.700_761delInsT} cells at day 7 and 10 (figure 4.11 a-b). It is noticeable from these graphs that there is a considerable amount of variation between the two repeats, this is also evident from the use of principal component analysis (see supplemental figure S7). The variability between these samples makes it difficult to detect consistent changes between the samples, lowering the number of genes identified as differentially expressed when the above criteria is applied. Nonetheless, some of the genes identified through this analysis are of interest and may be relevant to the phenotypic differences observed between the WT and *TBX2* mutants. For example, increased *NPPA* and *NPPB* levels are associated with heart failure (HF) progression in humans (Tan et al. 2002), both were upregulated at day 10 in the *TBX2* mutants (see table 4). *VCAM1* is another gene that has been associated with HF and immune cell infiltration of the myocardium. (Troncoso et al. 2021; Wang et al. 2021a), and was upregulated at day 10. These changes may indicate early signs of stress in the *TBX2* mutant cells.

Table 4.4 Differentially regulated genes of interest identified at day 10 in the *TBX2* mutant lines

The genes below were found to be differentially regulated in the *TBX2* mutant lines at day 10, when looking for a Log2FC value of ≥ 2 in the *TBX2*^{c.700_761del} line and a p-adjusted value ≤ 0.05 .

Ensembl ID	Gene ID	<i>TBX2</i> ^{c.700_761}	<i>TBX2</i> ^{c.700_761delInsT}
		Average log2FC	Average log2FC
ENSG00000175206	<i>NPPA</i>	5.2	2.5
ENSG00000120937	<i>NPPB</i>	3.3	1.2
ENSG00000162693	<i>VCAM1</i>	3.3	2.2

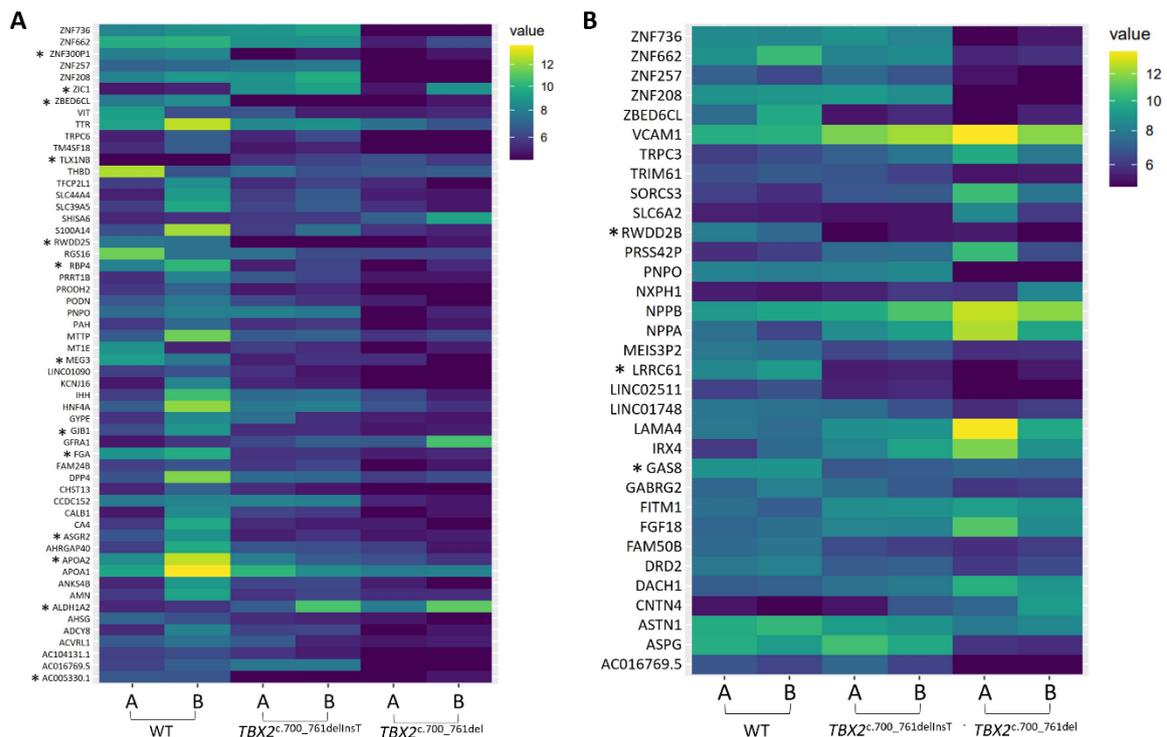


Figure 4.11 A comparison of the genes found to be upregulated in *TBX2*^{c.700_761del} cardiomyocytes

Cardiomyocytes were generated using the CDM3 differentiation protocol and samples taken at day 7 and day 10 were submitted for RNA-sequencing. Differentially expressed genes in this figure were identified using DEseq2. Following this genes that displayed a log2FC of ≥ 2 and had a p-adjusted value of < 0.05 in the *TBX2*^{c.700_761del} line at day 7 or 10 were selected for display. The genes found to be dysregulated in both *Tbx2* mutant lines on a given day are indicated with an asterisk. (A) Genes differentially regulated at day 7. (B) Genes differentially regulated at day 10. **DEseq2 analysis and heatmaps were generated by Antonios Tselingas, Cardiff University.**

4.2.13 Identification of miss regulated calcium handling genes in day 10 CMs

In an effort to uncover early transcriptomic changes in calcium handling genes that could contribute to the differences in calcium physiology observed in the mutants at day 32, the expression of genes assigned to the ‘WP calcium regulation in cardiac cells’ (WP536) gene set from the MSig Database were analysed. Those with a $\log_2FC \geq 1$ were included for analysis. Through this several calcium handling genes and signalling pathway interactors that modify calcium handling were identified as differentially regulated (table 5). Two of these genes, *CACNA1C* and *RYR2* were followed up for further analysis and their expression was examined at days 7-10 and day 32. *CACNA1C* encodes an L-type Ca²⁺ channel responsible for calcium influx at the cell membrane, and *RYR2* is necessary for Ca signalling amplification through calcium induced calcium release (CICR) from the sarcoplasmic reticulum. Therefore, both of these genes if upregulated have the potential to lead to the higher calcium transient amplitudes observed in the TBX2 mutant CMs at day 32 (see figure 10). RT-PCR analysis at multiple timepoints revealed their expression level to be dependent on the day analysed. At day 32 it seems that *CACNA1C* is slightly downregulated and *RYR2* expression is slightly upregulated. These changes are small and may simply reflect variability in the processing of these samples or the samples themselves. For small changes like this further repeats are necessary to improve confidence in the reliability of these results.

Table 4.5 A survey of calcium handling gene expression in day 10 TBX2 mutant cells

The expression of genes belonging to the WP536 gene set from the MSig database with a \log_2FC of ≥ 1 in the *TBX2*^{c.700_761del} line are displayed below. Those showing a consistent pattern of change in both lines are indicated with an Asterix and are shown in bold.

Gene symbol	Description	Log ₂ FC	
		<i>TBX2</i> ^{c.700_761del}	<i>TBX2</i> ^{c.700_761delInsT}
ADCY5 *	Adenylate cyclase	4.0	1.2
GJA3 *	Gap junction protein also known as Cx46	3.2	2.2
PLN *	Regulator of SERCA activity	1.7	2.0
CHRM4 *	G-protein coupled receptor	1.6	1.9
CACNA1C	L-type plasma membrane Ca channel	1.6	0.2
PRKCH *	Ca sensitive protein kinase C	1.2	1.5
RYR2	Sarcoplasmic reticulum calcium release	1.2	0.2
ATP1B1 *	Plasma membrane Na/K ATPase driven exchanger	1.0	1.0
RGS16	GTPase activating protein	-1.1	-0.3
ADCY2	Adenylate cyclase	-1.2	-0.1
GJA4	Gap junction protein also known as Cx37	-1.3	1.0
CHRM3	G-protein coupled receptor	-1.3	-0.3
GNG2	Guanine nucleotide binding protein	-1.5	-0.9
ADCY8 *	Adenylate cyclase	-1.8	-1.1
PKIB	cAMP dependent protein kinase inhibitor B	-2.2	-0.6
SLC8A3	Sodium calcium exchanger	-2.3	0.0
ADRA1D	Adrenergic receptor	-2.9	-0.4

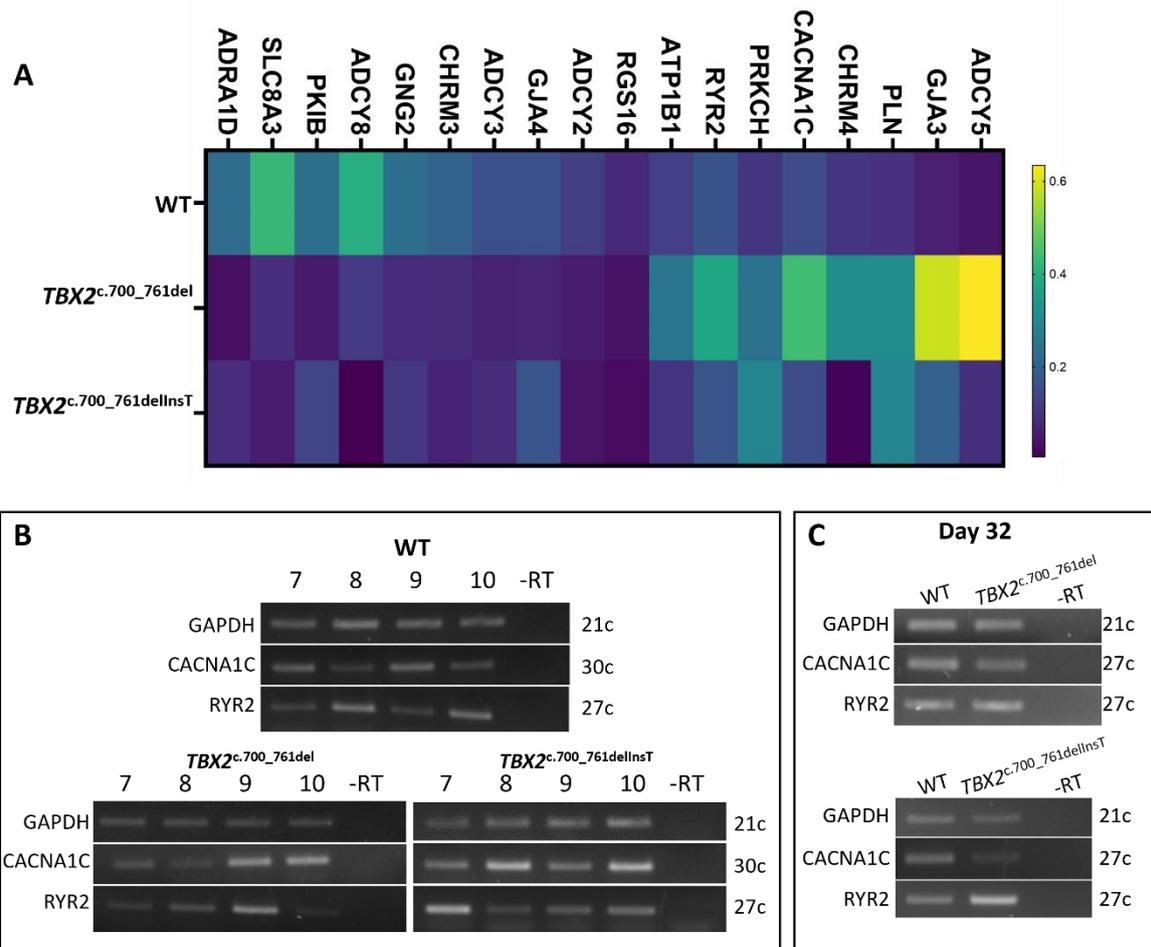


Figure 4.12 Analysis of calcium handling gene expression alterations in *TBX2* mutant CMs

(A) Heatmap showing expression values for Ca handling genes at day 10 in the WT and *TBX2* mutants. Values have been averaged between the two repeats. (B) RT-PCRs for *RYR2* and *CACNA1C* at days 7-10 and (C) day 32 corresponding with when calcium imaging was conducted for these cell lines. N = 2.

4.2.14 Gene ontology enrichment analysis reveals upregulation of cardiac development and disease genes in TBX2 mutants

To explore the data generated fully the less restrictive selection criteria used to identify calcium handling gene expression changes was used to explore the entire dataset. Through this at day-7 81 genes were found to be upregulated in both TBX2 lines and 519 down-regulated (see figure 4.13 a). At day-10 168 were found to be upregulated in both TBX2 lines and 175 down-regulated. It should be noted whether using relaxed or strict filtering criteria that more genes were found to be differentially expressed in the *TBX2*^{c.700_761} line. However, based on the premise that genes found to be differentially regulated in both are likely to be more sensitive to TBX2 regulation these genes were used for Gene Ontology (GO) enrichment analysis.

The list of genes found to be upregulated at day-7 were associated with terms related to cardiac development, such as cardiac chamber development and morphogenesis (GO terms: 0003208, 0003206, 0003231, and 0003205). This theme continued at day-10 with the list of upregulated genes found to be associated with terms including cardiac muscle development and contraction (GO terms: 0048738 and 0060047). Terms related to neurogenesis were also flagged in the day 7 upregulated gene set, which would not be expected based on the relatively high differentiation efficiencies seen for the TBX2 mutants. However, closer inspection of these neural terms shows that a number of the genes listed overlap with the cardiac and muscle gene sets, and includes genes such as *MEF2C*, *HEY1*, *RARB*, *PROX1*, and *JAG1*. The day 10 upregulated gene set revealed a number of associations with cardiac disease relevant pathways, as well as cardiac contraction and adrenergic signalling pathways. The STRING network created using these terms shows that the genes found to be upregulated that are involved in these pathways and demonstrates significant overlap between these pathways. The list of down-regulated genes for days 7 and 10, were associated with terms related to extracellular matrix organisation, migration, cell adhesion and regulation of kinase activity.

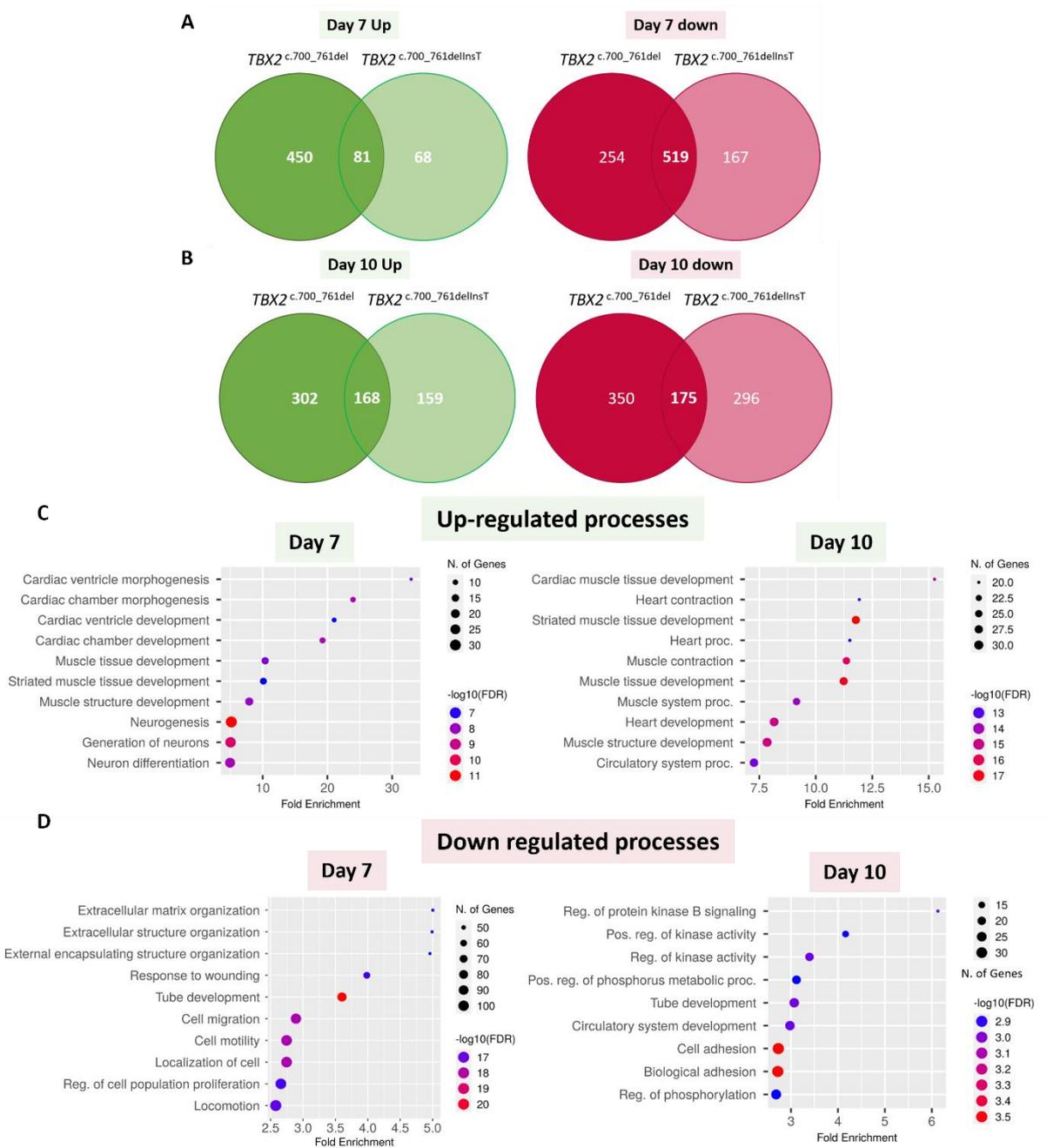


Figure 4.13 Continues on the next page

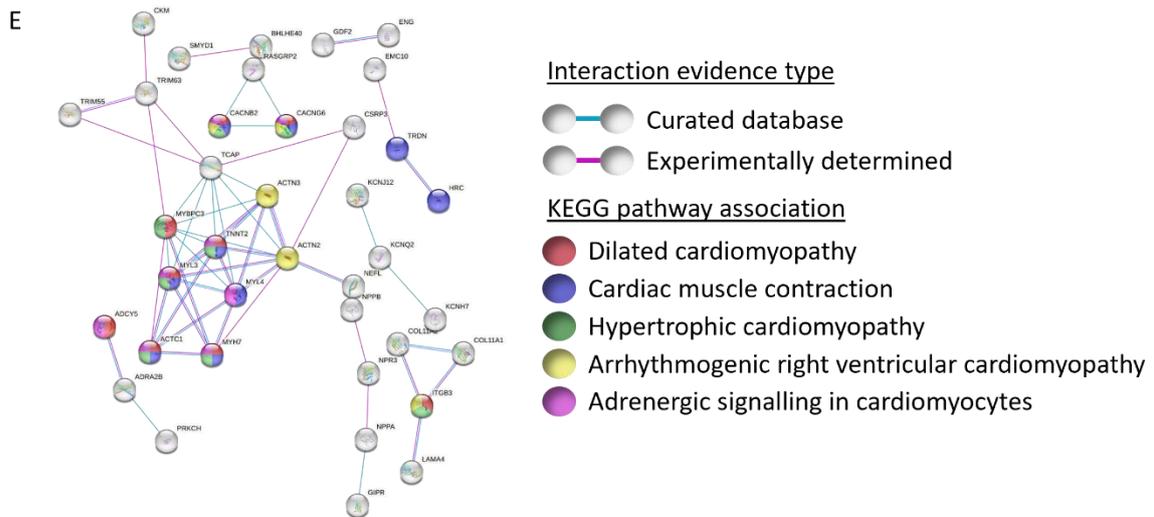


Figure 4.13 Transcriptional changes in the TBX2 mutants reveals upregulation of cardiac development and disease relevant genes

(A-B) Using a \log_2FC value of ≥ 1 a list of up- and downregulated genes was generated for each mutant at days 7 and 10. The overlap between these gene lists for each mutant is displayed using graphs created using Venny v2.1 (Oliveros 2021). (C-D) The overlapping genes found to be dysregulated in both mutants were subject to GO term enrichment analysis using ShinyGO (Ge et al. 2020), and the top 10 associated terms are reported. (E) Once again using ShinyGO a list of KEGG pathways and associated genes upregulated at day 10 was generated. The STRING network (Szklarczyk et al. 2019) displayed, shows the pathways these genes belong to and the interactions between their protein products.

4.3 Discussion

The actions of TBX2 in cardiac development have been well characterised in the context of AVC and OFT formation (Christoffels et al. 2004; Harrelson et al. 2004; Aanhaanen et al. 2011; Singh et al. 2012). TBX2 is broadly expressed in the linear heart tube before becoming restricted to these specialised tissues, indicating TBX2 may have a broader effect on early cardiac development outside of the OFT and AVC. The results presented herein show using iPSC-CM differentiation to model cardiomyogenesis that TBX2 is required for the formation of normal CMs. Mutation of the *TBX2* locus resulted in the formation of a high number of abnormal CMs. The phenotype of these cells is detailed below as well as possible reasons for its development.

4.3.1 Production of normal cardiomyocytes requires WT expression of TBX2

The mutants created herein carry DBD disrupting mutations, given their location they are proposed to constitute loss of function mutations. In addition, currently available data on protein fate indicate the mutant protein may be unstable, which will likewise contribute to a loss of function. The TBX2 mutant lines generated CMs at a similar rate and efficiency to their WT counterparts. Differentiation was confirmed by the observation of beating and the expression of CM specific markers such as TNNT2. These results agree with previous observations in mouse KO models that show chamber CM differentiation proceeds in the absence of TBX2 (Harrelson et al. 2004; Singh et al. 2012). In these studies, the defects recorded primarily affected the AVC and OFT, whilst no significant effects were detected in the development of atrial or ventricular myocardium. The Wnt modulation method of cardiac differentiation used herein primarily produces ventricular-like CMs, with some atrial and nodal like cells also present (Lian et al. 2013; BurrIDGE et al. 2014b; Galdos et al. 2022). In contrast to the studies by Harrelson and Singh *et al.* a significant effect on the phenotype of these myocardial cells was noted.

The TBX2 mutant CMs displayed hypertrophy, myofibril disarray, and abnormal calcium handling properties, qualities indicative of myocardial pathology. The *in vivo* mouse studies discussed have invariably analysed TBX2 loss of function in fixed whole hearts or sections and no calcium handling analysis was carried out. These methods may lack the resolution needed for the phenotypic changes described in the TBX2 mutant lines described herein to be observed. The onset of beating and the morphology of the beating

sheets/networks formed from the TBX2 mutant cells resemble that of WT cultures at a macro level. It is only when these cells are examined on an individual basis that elements of the phenotype such as sarcomeric disorganisation can be appreciated, offering a potential reason for why this was unnoticed in previous studies.

An element of the phenotype that should be easily detectable in whole hearts or heart sections is hypertrophy, yet this has not been observed previously (Harrelson et al. 2004; Singh et al. 2012). A possible explanation for this is that the most abnormal hypertrophic cells may be eliminated *in vivo*. If this was the case an increase in apoptosis, and possibly myocardial hypoplasia would be expected in these hearts, this has not been noted in the mouse knockout models presented thus far (Harrelson et al. 2004; Singh et al. 2012). However, this is worth closer investigation as understanding whether these abnormal CMs would be maintained during organogenesis is important for understanding how a loss of TBX2 function may contribute to human disease. In the papers presented by Harrelson *et al.* and Singh *et al.* in the embryos where AVC constriction has failed the presumptive chambers of the heart are dilated, and oedema is present. This is proposed to be a result of circulatory insufficiency due to defective AVC and OFT formation. The dilation and oedema present are likely to have secondary effects on the myocardium that will cloud interpretation of the myocardial phenotype, whereas the iPSC-CMs formed herein are not exposed to these symptoms. These considerations might provide an alternative explanation for the differences in phenotypic presentation seen.

4.3.2 How does CM heterogeneity affect the phenotype of TBX2 mutant cells?

The observance of phenotypic heterogeneity within the TBX2 mutant population raises an interesting question: why are some CMs in the population badly affected by the TBX2 mutation and others not? It is clear even in the WT population the CMs present are heterogenous in their morphology, and physiology. This is not unexpected as multiple studies have demonstrated that CMs formed using similar protocols exhibit heterogenous transcriptional states, morphologies, and physiological profiles (Yechikov et al. 2016; Churko et al. 2018; Cuomo et al. 2020). The heterogeneity is proposed to be the result of CMs within the population exhibiting different levels of maturation, or their allocation to different sub-types. A scRNA-seq study conducted by Galdos *et al.* demonstrated that iPSC-CMs formed via the Wnt modulation protocols used herein are biased towards a left

ventricular fate when their transcriptional state is considered (Galdos et al. 2022). However, when multiple parameters of the CM phenotype are considered, it is clear iPSC derived CMs do not align strictly to a particular CM subtype due to their immaturity (Kane and Terracciano 2017). Even within the human adult heart sub-specification of ventricular and atrial CMs is seen (Litviňuková et al. 2020). This makes it difficult to answer from the data presently available whether the difference between the WT-like and abnormal TBX2 mutant cells is stochastic, due to a requirement for TBX2 in maturing cells or in the development of a particular CM subtype. The first step to addressing this would be to determine if TBX2 expression is heterogenous within the CM population, currently the data reported here only provides a measurement of bulk TBX2 expression, but this could be measured in individual cells through the use of scRNA-seq and IF. Furthermore, simple modifications to the differentiation protocols used herein are available that allow for the derivation of more atrial, nodal, or conduction-like CM sub-types, as well as protocols for CM maturation (Cyganeck et al. 2018; Lyra-Leite et al. 2022; Prodan et al. 2022; Wiesinger et al. 2022). Applying these protocols and then assessing if TBX2 expression is upregulated in cells headed for a certain developmental pathway could go some way to answering the questions posed.

4.3.3 TBX2 mutant CMs display characteristics of various CVDs

Although the mutations in TBX2 are having a negative effect on CM phenotype it is not currently clear how. In 2D culture CMs are not experiencing the same stressors as they would in the heart such as stretch and haemodynamic pressure that may trigger hypertrophy *in vivo* (Stansfield et al. 2014; Nakamura and Sadoshima 2018). Hypertrophic growth, myofibril disarray, and abnormal calcium handling, as seen in the TBX2 mutant cells, are all hallmarks of pathological hypertrophy in humans which can progress to heart failure (Stansfield et al. 2014). These characteristics have been demonstrated in iPSC-CM models of genetic and pharmacological stimulated cardiomyopathies previously (Sun et al. 2012; Tanaka et al. 2014; Broughton et al. 2016; Wyles et al. 2016; Mosqueira et al. 2018; Zhou et al. 2019). The changes observed in the TBX2 mutant lines were primarily determined at day 32, however an increase in average cell size was notable even at day 12 by IF. Furthermore, RNA-seq analysis of samples taken at day 7 and 10 demonstrated transcriptional changes are present at an early stage and this likely contributes to the visual phenotypic changes noted at days 12 and 32. In the TBX2 mutants, genes associated with cardiac development

processes were over-represented at both timepoints, and at day 10 upregulation of cardiac stress hormones *NPPA* and *NPPB* was observed. Additionally, KEGG pathway analysis identified an association of the upregulated genes at day 10 with several disease pathways including hypertrophic cardiomyopathy. This suggests these cells are under stress shortly after the commencement of beating, instigating a hypertrophic gene program early in their development, which likely contributes to the deterioration in cellular condition noted at day 32.

Interestingly at day 7 terms related to cardiac chamber morphogenesis and development were found to be upregulated in our RNA-seq datasets. The genes within these terms included *MEF2C*, *PROX1*, *FOXC1*, and *JAG1*, which have roles in the differentiation and maturation of ventricular myocardium. For example, *MEF2C* is a member of the core cardiac GRN, and is known to guide ventricular CM differentiation in cooperation with other cardiac TFs such as *NKX2-5* (Vincentz et al. 2008; Materna et al. 2019). *PROX1* expression has been shown to be important for proper sarcomere formation, myofibril alignment, and physiological foetal CM hypertrophy (Risebro et al. 2009). *FOXC1* is needed for the formation of trabeculated myocardium, which is severely diminished in *FOXC1* deficient zebrafish embryos (He et al. 2022), whereas *JAG1* knock out in mouse myocardium results in the formation of highly trabeculated left ventricular myocardium that fails to compact (D'Amato et al. 2016). All of these genes therefore seem to have a positive effect on ventricular myocardium development being required for its progressive differentiation and maturation. Therefore, it is possible that certain aspects of maturation are accelerated or there is an over activation of these processes in the *TBX2* cells, that could be leading to dysfunction in these cells, and that *TBX2* acts in the early stages of CM differentiation to dampen the expression of these genes ensuring measured progression through differentiation.

4.3.4 Abnormal calcium handling: cause or effect?

Another possible stimulus for CM hypertrophy is dysfunctional calcium handling. Contraction is coupled to calcium flux and calcium can also activate pro-hypertrophic gene programs; thus calcium handling, CM function and health are inextricably linked (Molkentin 2013). Calcium amplitude was increased in both mutants, and time to peak was significantly increased in the *TBX2*^{c.700_761del} mutant. Furthermore, there were more spontaneous

calcium sparks detected in between transients, an increase in spontaneous sparks can happen as the result of increase SR calcium concentration (Lukyanenko et al. 2001). These changes indicate that calcium handling in the cells is disrupted and mean that over time intracellular levels of calcium are higher for longer periods of time.

When calcium levels are high calmodulin can bind Ca^{2+} ions causing conformational changes that allow it to activate the Ca^{2+} /calmodulin dependent kinase II signalling cascade, and the Calcineurin-NFAT pathway. Both pathways can activate transcriptional programs that promote hypertrophic pathological remodelling (Molkentin et al. 1998; Molkentin 2013). Several calcium handling genes and genes that modulate these pathways were found to be dysregulated in the mutants therefore calcium signalling disruption is a possible driver of pathology in these cells. TFs NFATC2 and 3 are important mediators of calcium induced hypertrophic remodelling, and both have been demonstrated to be essential for this process (Wilkins et al. 2002; Bourajjaj et al. 2008). Therefore, we may expect to see nuclear accumulation of Nfat TFs such as these if calcium is a driver of this process, and this is something that could be easily assessed using IF (Molkentin et al. 1998; Molkentin 2013). Furthermore, it has been demonstrated using calcineurin inhibitors cyclosporin A and FK506 that blocking these pathways can alleviate the hypertrophic response (De Windt et al. 2001; Molkentin 2013), as can the use of Verapamil and Carvedilol which can reduce calcium influx levels and thus lower Ca signalling pathway activation (Wyles et al. 2016). Implementation of these treatments in the TBX2 mutant CMs is another method that could be used to confirm if calcium dysregulation is a mechanism of disease pathogenesis in these cells.

It is not clear at the transcriptional level how a lack of TBX2 may be responsible for the changes in calcium handling. The insights gathered from RNA-seq analysis suggest there is deregulation of some calcium handling genes in the TBX2 mutants, but expression level of these genes was variable between the repeats and the direction of these changes in gene expression are dependent on the day analysed. Comparison of the genes flagged as differentially regulated with datasets available via ChIP-Atlas shows that TBX2 binds sites surrounding; *CHRM4*, *ATP1B1*, *PKIIB*, *ADRAID*, and *PRKCH*. However, it should be noted that these results were generated using mammary, hepatocyte, and neuroblastoma lines, how relevant this is to CM development is not clear.

4.3.5 Different levels of dysfunction between the mutants

Of the data discussed above the $TBX2^{c.700_761delInsT}$ mutant, by some metrics, exhibits a milder phenotype than the $TBX2^{c.700_761del}$ mutant. There is a significant amount of variation between differentiations even for WT cells, thus further repeats are required to resolve these differences. However, of course these two mutants are molecularly different, and this should be taken into consideration also. The mutants retain different amounts of their DBD, with $TBX2^{c.700_761del}$ missing 49 amino acids including 4 DNA contacting residues, whereas the $TBX2^{c.700_761delInsT}$ mutant is missing 20 amino acids, none of which are predicted to directly contact DNA, based on comparisons with a model of close relative TBX3 bound to DNA (Coll et al. 2002). Despite these differences, it is thought due to the effect the deletions will have on the DBD fold and the observation that both proteins are unstable in *Xenopus* that DNA binding activity will be lost in both. These mutant proteins may well be unstable during iPSC-CM differentiation also, however, repeats of the WB conducted here with alternative antibodies that can provide a clearer result is needed to improve confidence in this.

The IHC staining conducted in *Xenopus* provides further evidence that the mutant proteins are likely unstable and may also be unable to reach the nucleus. The mutant proteins were undetectable after MBT by WB, however staining of the *Xenopus* embryos by IHC at stage 8.5-9 at the onset of MBT revealed small regions of mutant protein expression. The amount of staining present was decreased and was diffuse in those injected with mutant mRNAs, in comparison to the punctate staining observed in many cells in the WT $TBX2$ injected embryos. Based on these preliminary results it appears translocation of the mutant protein into the nucleus is impaired. A nuclear localisation signal (RREKRK) has been characterised in TBX3, and this motif is absolutely conserved in TBX2 (Carlson et al. 2001; Abrahams et al. 2008). This motif is still present in the mutants however it is found at positions 292-297 in TBX2 which is close to deleted regions in the mutant proteins; $TBX2^{c.700_761del} = 222_270 del$, and $TBX2^{c.700_761delInsT} = 234_253del$. The effects of these mutations on the protein fold may affect recognition of the nuclear localisation signal based on their proximity to the mutations. Alternatively, the proteins may be retained in the endoplasmic reticulum as a result of protein misfolding subsequently activating the unfolded protein response. This would tally with the reduced expression levels observed via IHC, and their decrease to undetectable levels when measured by WB after MBT. The fate of

the mutant TBX2 proteins will need to be definitively confirmed in the human iPSC-CM samples possibly using an alternative TBX2 antibody to achieve this.

4.3.6 Conclusions

The work conducted here documents a previously unappreciated role for TBX2 as a modulator of the cardiac GRN. Although CM formation is possible in the absence of WT TBX2, the CMs formed are abnormal and exhibit signs of pathology. What drives this dysfunction is at present unclear. As TBX2 is known as a transcriptional repressor and there was an observed upregulation of chamber myocardial genes in these lines, it is supposed that over-activation of these programs may be responsible for this pathology. Abnormal calcium handling is another possible driver of dysfunction in these cells. Furthermore, it is not clear if loss of TBX2 has the same effect on all CMs within the population. The steps required to answer these remaining questions are outlined in Chapter 7.

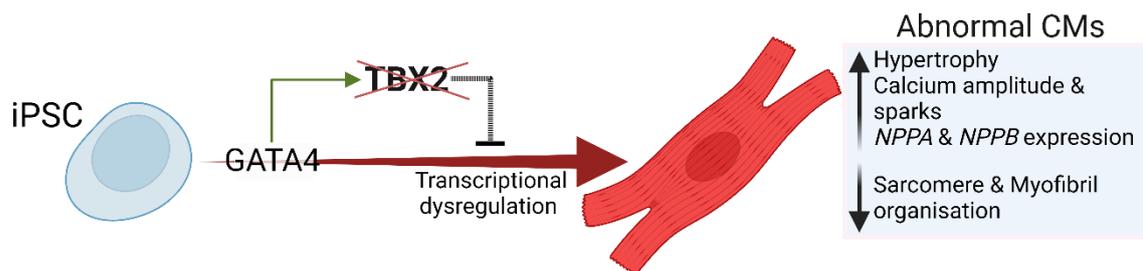


Figure 4.14 TBX2 is a modulator of the cardiac GRN that is required for production of WT-like CMs

From the data previously generated by our lab in *X. laevis* (data not published) and displayed in Chapter 3 it seems that *TBX2* is a positive target of GATA4 transcriptional regulation and is activated at the cardiac progenitor stage. The results presented in Chapter 4 suggest that TBX2 activity during iPSC-CM differentiation is necessary for the formation of WT-like CMs and that TBX2 may act as a necessary attenuator of the cardiac GRN.

Chapter 5 - PRDM1 is dispensable for the differentiation of iPSCs into cardiomyocytes

5.1 Introduction

Cell fate is established and maintained through the collaboration of multiple TFs that form a GRN. Regulated combinatorial and sequential action of multiple nodes within the GRN is necessary to attain specific transcriptional outputs. Most TFs will modulate transcription by recruiting chromatin modifiers to their location, and during development this helps to establish the appropriate chromatin landscape for the next TFs in the cascade of gene regulation. The integration of chromatin landscape, and the complement of TFs expressed at a given stage allows for precise modulation of cell fate decisions during development, and in mature adult cells these components act as a unique signature to maintain a stable phenotype. When these GRN's are disrupted developmental disorders and disease ensue, therefore identifying what TFs take part in these networks and understanding the role they play in this process is fundamental for our understanding of normal development and disease.

5.1.1 Transcription factor PRDM1 as a recruiter of chromatin modifying enzymes

PRDM1 is a TF with many roles during development, it has primarily been characterised as a transcriptional repressor. Structurally PRDM1 is made up of a PR/SET domain that lacks histone methyltransferase activity, a proline/serine rich domain, and 5 C₂H₂ Zn fingers that constitute the DBD (Bikoff et al. 2009; di Tullio et al. 2022)(see figure 5.1). PRDM1 directly interacts with repressive chromatin modifiers, such as; LSD1, Groucho family proteins, HDAC2, and G9a/hKMT1c (EHMT2), through its DBD and/or its PR domain, allowing for the construction of multiprotein repressive complexes at sites throughout the genome (Ren et al. 1999; Yu et al. 2000; Györy et al. 2003; Su et al. 2009). Recently, co-regulation of gene expression by PRDM1 and a positive chromatin modifying enzyme KDM4A has been observed during neuronal and sensory neural progenitor development (Prajapati et al. 2019). This is one of few reported incidences of PRDM1 associating with an activating chromatin modifier. The ability of PRDM1 to act as scaffold for the recruitment of a variety of chromatin modifiers, allows it to participate in multiple development processes, including pivotal roles in plasma cell, primordial germ cell, sensory neuron, sensory vibrissae, and enterocyte maturation (Bikoff et al., 2009; Mould et al., 2015; Prajapati et al., 2019). Most

of these roles have been well characterised, however little is known about PRDM1 in the context of heart development.

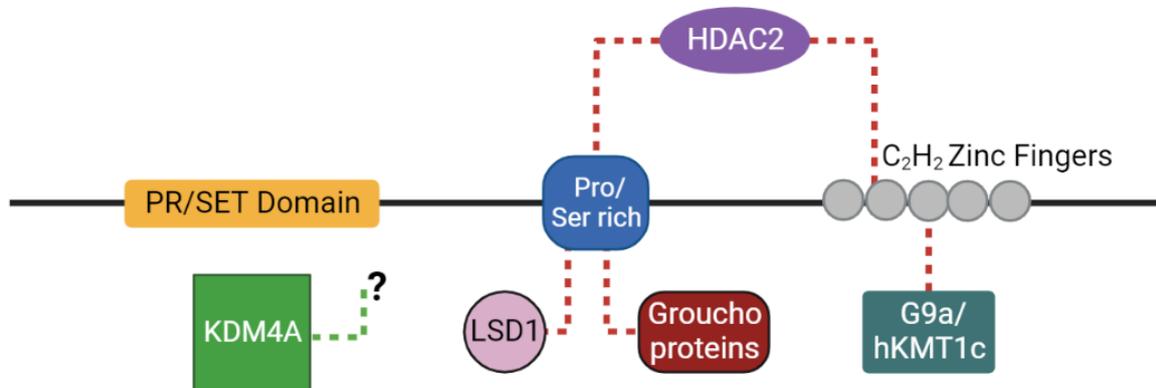


Figure 5.1 A linear depiction of PRDM1 and its interactors

Prdm1 has 3 characteristic domains. A PR/SET domain that lacks histone methyltransferase (HMT) activity, a proline/serine rich domain (PS), and a DNA binding domain (DBD) made up of 5 C₂H₂ zinc fingers. The domains required for direct interaction with repressive chromatin modifiers are connected by red dotted lines. An interaction with a positive chromatin modifier KDM4A is also shown, although the area of PRDM1 that it interacts with is currently unknown. Figure is adapted from Bikoff *et al.* 2009.

5.1.2 PRDM1 in cardiac development

Evidence from mice suggests that PRDM1 is a participant in the cardiac GRN, with fate mapping results demonstrating that PRDM1⁺ cells give rise to a large proportion of the right ventricle, OFT, and ventricular septum, they also make a smaller contribution to the atria (Robertson *et al.* 2007), which are all SHF derivatives (Van Vliet *et al.* 2012; Devine *et al.* 2014). A small patch of PRDM1⁺ cells are also present in the apex of the left ventricle a FHF derivative (Robertson *et al.* 2007). Corresponding with this pattern, in PRDM1^{-/-} mice a range of cardiac defects associated with these regions are seen, for example persistent truncus arteriosus (PTA), ventricular septal defects (VSD), and a more pronounced apex (Robertson *et al.* 2007). Targeted deletion of PRDM1 in the cardiac mesoderm using the MESP1 promoter as a driver, also results in PTA and VSDs, but an effect on the heart apex is not noted (Vincent *et al.* 2014). In contrast to the MESP1 driver, deletion of PRDM1 using a

MEF2C driver, which targets the cardiac progenitor stage later in development, does not lead to severe defects (Vincent et al. 2014). The widespread contribution of PRDM1 expressing cells to the heart, and the severity of the defects seen when it is removed early in development suggest an integral role for PRDM1 in early cardiogenesis.

From the limited published data available it seems that PRDM1 could play a similar role in humans. Firstly, in human iPS cell to CM development expression of *PRDM1* peaks at day 3 coinciding with the establishment of cardiac mesoderm (Chapter 3, figure 3.12 and Churko *et al.* 2018), the earliest step in cardiac development. In addition to this, PRDM1 homozygous deletion as part of a microdeletion on Chr 6q21 has been associated with PTA in one patient (Shaheen et al. 2015), a phenotype that was observed in PRDM1^{-/-} mice. In other cell types where the role of PRDM1 has been investigated a common theme is PRDM1 acting as a repressor to block the progression of differentiation. For example, in the mouse sebaceous gland and mammary luminal adult stem cells it suppresses the fully differentiated gene program and cell proliferation, this suppression is required to maintain the progenitor pool that will subsequently re-populate the respective ducts (Horsley et al. 2006; Ahmed et al. 2016). It is possible PRDM1 could play a similar role in early cardiac specification, in preventing the expression of alternative gene programs, and stabilising the early decision to acquire a cardiac mesoderm fate.

According to this view, for progression through differentiation and commitment it should then be necessary that PRDM1 expression is downregulated (Horsley et al. 2006; Ahmed et al. 2016). Unpublished data from our group also appears to fit this model. Overexpression of GATA4 can guide *Xenopus* pluripotent animal cap explants to form spontaneously beating tissue. In an RNA-seq screen to identify early gene expression changes in response to GATA4 overexpression, *PRDM1* was identified and characterised as a direct target of GATA4 transcriptional repression by CHIP RT-PCR. This result suggests its repression is needed for differentiation and commitment. In addition to this, overexpression of PRDM1 in whole *Xenopus* embryos is not compatible with normal heart development. Regulation of *PRDM1* by GATA4 seems to be conserved in humans given that *PRDM1* expression is down regulated at day 4 (early cardiac progenitor), coinciding with increased *GATA4* expression during iPS cell to CM differentiation, and that it is upregulated in a *GATA4* null background (Chapter 3, figure 3.12).

Therefore, the aim of this section of work was to assess if regulated expression of PRDM1 is needed for normal cardiac development. Multiple attempts were made to produce a stable PRDM1 gain of function line so that PRDM1 could be forcibly induced outside of its usual expression window. This was unsuccessful but the approach taken to derive these lines and accompanying preliminary data are described in supplemental figure S1. Nonetheless, as expression of this gene is at its highest at day-3 corresponding with the expression of cardiac mesoderm markers such as *MESP1* (Churko et al. 2018), it was thought that PRDM1 may have functions at this early stage of cardiac development. Therefore, CRISPR-Cas9 gene editing was employed to create human iPS cells carrying LoF mutations in the *PRDM1* gene locus and the effect of these mutations on CM formation has subsequently been evaluated.

5.2 Results

5.2.1 Editing of the *PRDM1* locus using CRISPR-Cas9 gene editing

Creation of *PRDM1* mutant lines was achieved using CRISPR Cas9 gene editing. The primary consensus coding sequence (CCDS5054) for *PRDM1* is encoded by 7 exons and is shown in figure 5.2 a. Several other transcripts may be produced from this locus therefore, the gRNAs were targeted to an exon common to all protein coding transcripts. Exon 4 of the primary CDS is used by all protein coding transcripts derived from the *PRDM1* locus except *PRDM1-204*, a predicted transcript with no recorded evidence of *in vivo* transcription. Therefore, given its utilisation by all major transcripts it was decided exon 4 would be an optimal target. The guides were placed 121 bp apart with the aim of creating a frameshifting deletion. PCR screening of gDNA taken from single isolates identified two mutants with the expected ~121 bp drop in PCR band size, and another mutant that forms an amplicon of similar size to WT (figure 5.2 b). Sanger sequencing of these mutants confirmed the 121 bp deletion in two of the mutants and revealed a 1 bp adenine insertion at the cut site for sgRNA 1 in the other (figure 5.2 c).

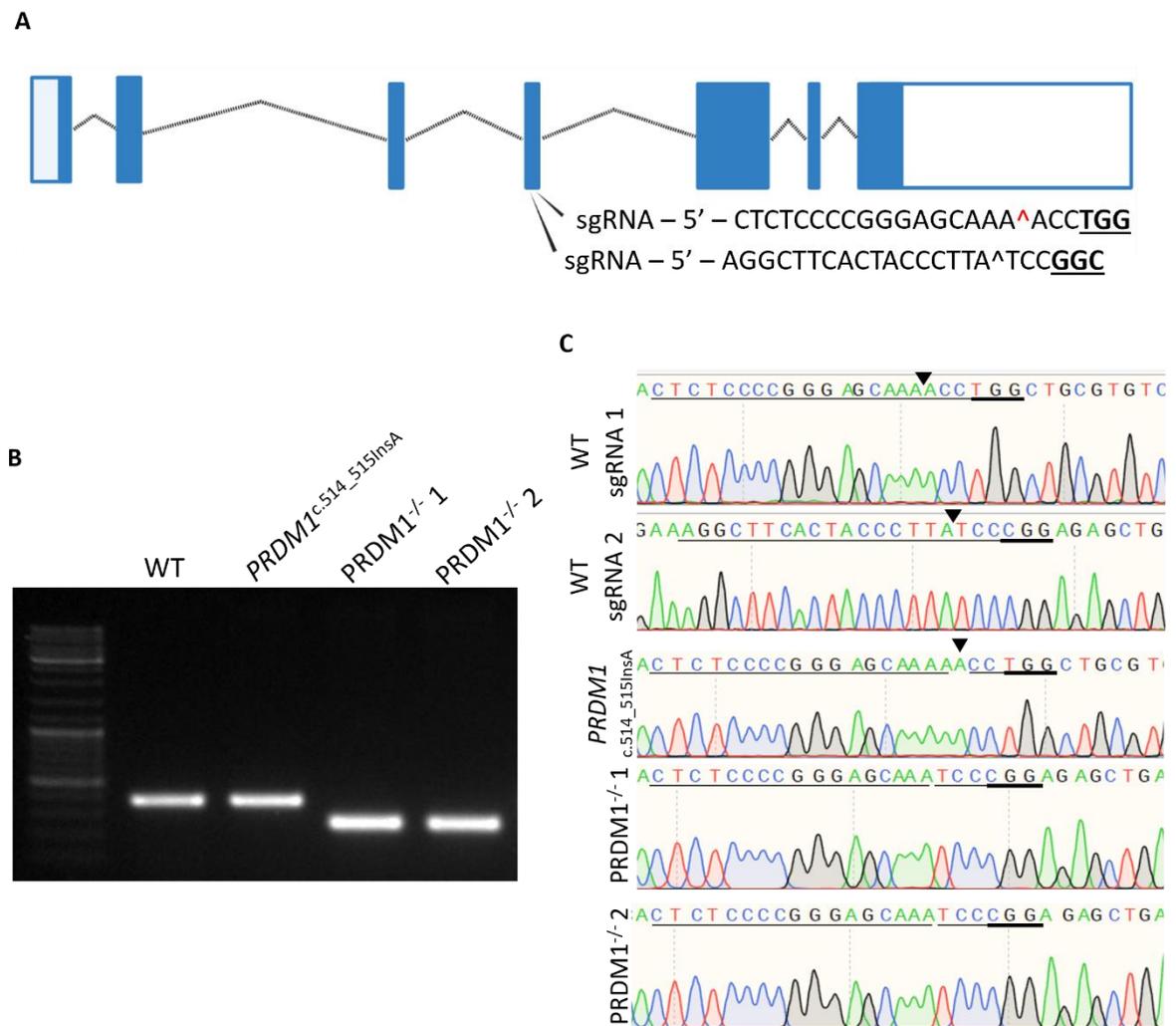


Figure 5.2 Generation of PRDM1 mutants using CRISPR-Cas9

(A) A diagram representing the *PRDM1* primary transcript (CCDS5054). Exons are represented as boxes, with coding area filled, and non-coding regions empty. Introns are represented by dotted lines and sgRNA cut sites placed at exon 4 are indicated. (B) Gel image showing the amplification of the region surrounding the edit site for WT and candidate mutants. (C) Sanger sequencing trace files for the regions covering sgRNA cut site 1, and 2 in WT and mutants. The protospacer sequence is underlined, and PAM sequences are underlined in bold. Cut sites are indicated by black arrows, and the inserted base in the *PRDM1*^{c.514_515InsA} mutant is indicated by a red arrow.

5.2.2 Analysis of PRDM1 expression in WT and mutant lines

The mutations described above all result in a frameshift introducing premature termination codons into the sequence, and thus are expected to induce nonsense mediated decay of the mRNA transcripts resulting in no protein being produced (Hug et al. 2015). Western blotting (WB) was used to determine if this was the case, prior to this the expression of PRDM1 during development was examined in a WT background to select the optimal day for analysis. The mRNA expression pattern of PRDM1 during WT cardiac differentiation is known and the peak is seen at day 3 of development, however protein levels peak at day 4 (figure 5.3 a). Therefore, day 4 was henceforth used for comparison between WT and mutant cells. For the mutants a loss of protein was only seen in those carrying the 121 bp deletion ($PRDM1^{-/-}$). In $PRDM1^{c.514_515InsA}$ protein was still detected (figure 5.3 b). This does not correspond with what is expected from the gDNA sequence, therefore sequencing of the gDNA was repeated and was also conducted on day 3 cDNA samples for WT and the $PRDM1^{c.514_515InsA}$ mutant to ensure accuracy. The gDNA sequencing validated what was previously seen, and sequencing of the mutant cDNA confirmed the mutation is present in the mRNA (figure 5.3 c).

Distinct transcripts that could be affected were investigated using RT-PCR. The two major isoforms of *PRDM1* that are translated are PRDM1 203/PRDM1 α (CCDS5054) which is 891 aa long and the shorter form is PRDM1 201/PRDM1 β (CCDS34505) at 691 aa long. The shorter form has a truncated PR/SET domain which has been shown to reduce its ability to repress target gene transcription (Györy et al. 2003). RT-PCR demonstrated that in WT the predominant isoform being expressed is *PRDM1 201/PRDM1 α* the short form, in the $PRDM1^{c.514_515InsA}$ line expression of this transcript appears to be decreased or absent (figure 5.3d). However, expression of the long form (*PRDM1-203*) is comparable between WT and the mutant, how this transcript escapes nonsense mediated decay is unclear. In the $PRDM1^{-/-}$ line primers detecting both transcripts were used for RT-PCR, and these results show PRDM1 expression is present at day 1 of differentiation and is then silenced by day-3 confirming both transcripts have been lost (figure 3e). The antibody used for detection of PRDM1 protein by WB recognises a motif present in both isoforms, however it was not possible to reliably detect multiple isoforms by WB even in the WT line. The most prominent band that is regularly detected migrates at ~95 kDa, thus this most likely represents the long (89 kDa) form of the protein.

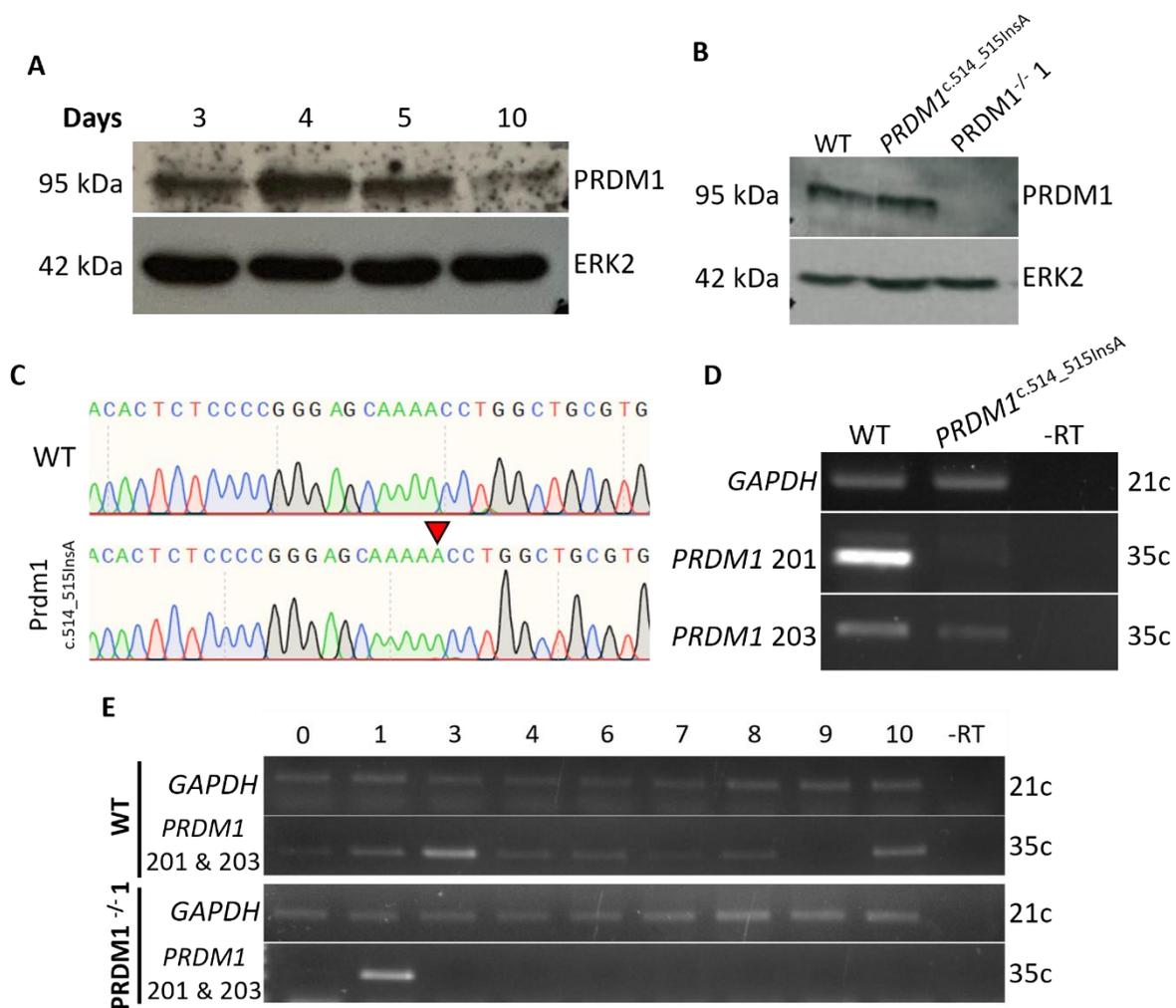


Figure 5.3 PRDM1 expression in WT and mutants

(A) The profile of PRDM1 protein levels in WT differentiation were determined using western blotting of day 3, 4, 5, and 10 samples. ERK2 was used as a loading control. (B) Western blot showing PRDM1 levels in WT and mutant lines at day 4, again ERK2 was used as a loading control, $n = 3$. (C) Sanger sequencing trace files of WT and *PRDM1*^{c.514_515InsA} cDNA demonstrates that the inserted adenine residue is reflected in the mRNA, and this is indicated by the red arrow. (D) RT-PCR for the expression of specific *PRDM1* isoforms in day 4 samples from WT and *PRDM1*^{c.514_515InsA} mutants, $n = 2$. (E) RT-PCR at different days throughout iPSC-CM differentiation for the WT and *PRDM1*^{-/-} line, to detect expression of PRDM1 α (CCDS5054) and β (CCDS34505). GAPDH has been used as a normalisation control, $n = 2$.

5.2.3 PRDM1 mutant lines retain pluripotency

Examination of the PRDM1 mutant lines was undertaken to test their ability to maintain pluripotency. Brightfield images at 4x magnification show rounded compact colonies characteristic of iPSCs (figure 5.4a). IF was used to confirm the expression of pluripotency TFs; POU5F1 and SOX2 in mutants. As can be seen in figure 5.4b both factors are robustly expressed in the mutant lines, and staining is restricted to the nucleus of non-dividing cells, as expected. Flow cytometry was conducted to quantify this at the population level, showing that POU5F1 is expressed in >97% of the population in all cell lines (figure 5.4d). In addition to this, RT-PCR which also gives an overall population view, shows that for two of the mutants the expression of *POU5F1*, *SOX2*, and another pluripotency factor *NANOG*, is maintained in the mutants (figure 5.4c). Together these results confirm that pluripotency is preserved in the mutants, suggesting that their physiology has not been fundamentally altered by gene editing.

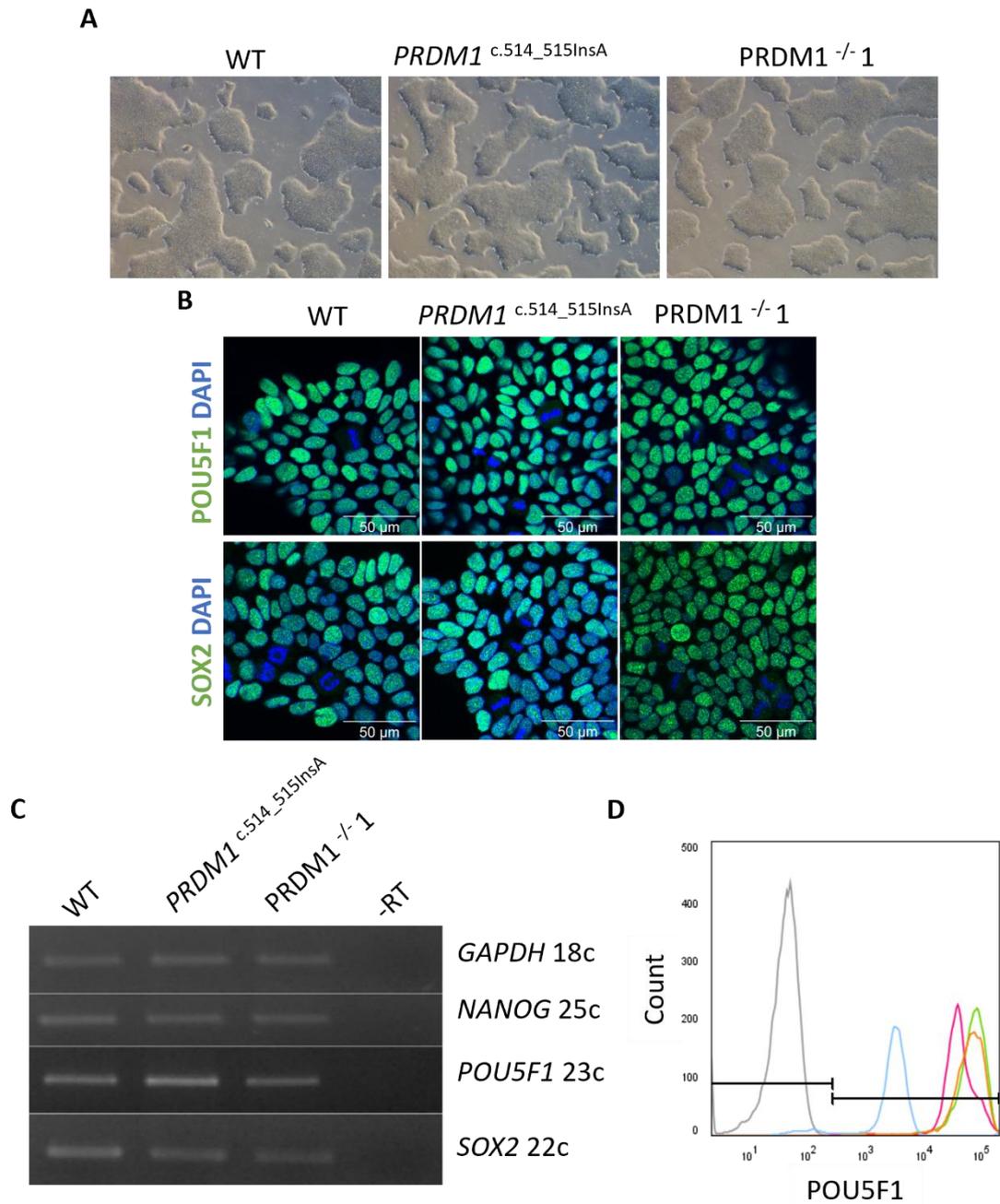


Figure 5.4 The features of pluripotency are maintained in *Prdm1* mutant iPSCs

(A) Brightfield 4x images of WT and mutant iPSC colonies. (B) Confocal images of WT and *Prdm1* mutant iPSCs stained for POU5F1 and SOX2, and counterstained with DAPI. Scale is indicated. (C) RT-PCR for the expression of pluripotency factors *NANOG*, *POU5F1*, and *SOX2* and loading control *GAPDH* in WT and mutant iPSCs. PCR cycles are indicated. (D) A graph showing the proportion of cells stained positively for POU5F1 by flow cytometry for WT and mutant samples.

During gene editing genomic integrity may also be compromised due to the introduction of potentially mutagenic double strand breaks and by the selection stress exerted by single cell cloning. Genomic integrity post editing was confirmed using a microarray covering 654,027 markers across the human genome (see section 2.3.2). No differences between the WT Rebl Pat parental line and the mutants were seen (Table 5.1).

Table 5.1 Summary of karyotyping results for PRDM1 mutants

No changes in genomic integrity were observed between WT and the mutants, but the features flagged as different from the reference genomes used by the PennCNV software.

Cell Line	Location	Event	Size	Transcripts from affected regions
Rebl Pat WT	chr2:139414906-139884468	Duplication	469,563	lncRNA AC016710, pseudogene MRPS18BP2, small nucleolar RNA SNORA72, pseudogene RPL9P13, and lncRNA AC078851.
Rebl Pat WT	chr19:20605360-20700135	Deletion	94,776	lncRNA AC010636, pseudogene AC010329, and ZNF626.
<i>PRDM1</i> c.514_515InsA	chr2:139414906-139884468	Duplication	469,563	Contains lncRNA AC016710, pseudogene MRPS18BP2, small nucleolar RNA SNORA72, pseudogene RPL9P13, and lncRNA AC078851.
<i>PRDM1</i> c.514_515InsA	chr19:20605360-20700135	Deletion	94,776	Contains lncRNA AC010636, pseudogene AC010329, and ZNF626.
PRDM1 ^{-/-} 1	chr2:139414906-139884468	Duplication	469,563	Contains lncRNA AC016710, pseudogene MRPS18BP2, small nucleolar RNA SNORA72, pseudogene RPL9P13, and lncRNA AC078851.
PRDM1 ^{-/-} 1	chr19:20605360-20700135	Deletion	94,776	Contains lncRNA AC010636, pseudogene AC010329, and ZNF626.

5.2.4 Formation of contracting CMs is slightly accelerated in PRDM1 mutants

PRDM1 mRNA expression is highest at day 3 aligning with the establishment of cardiac mesoderm during iPSC to CM differentiation. This may be considered the earliest step in cardiac development, and it was hypothesised that loss of PRDM1 expression at this time may negatively affect cardiac mesoderm formation which would subsequently adversely affect CM formation. To examine how the loss of PRDM1 expression affects iPSC-CM differentiation the mutant lines were differentiated to CMs using the CDM3 and GiWi protocols. PRDM1 mutant iPSC cells formed CMs at a similar efficiency to WT (figure 5.5b), and representative images of the cells at day 32 are shown in figure 5.4a, for corresponding videos that show beating see supplemental video files. As the cell lines were able to form CMs this demonstrates that PRDM1 expression at day 3 is not necessary for the formation of cardiac mesoderm, as the cells are able to proceed through development.

It was noted for the PRDM1 mutants that the onset of beating is earlier on average. Across 12 experiments *PRDM1*^{c.514_515delInsA} CMs started beating 1.17 days earlier (n = 118), and for PRDM1^{-/-} mutants examined over 8 experiments beating started around 2.10 days earlier (n=48), this necessitates that the expression and translation of sarcomeric genes must start earlier in the PRDM1 mutants to accommodate this. Analysis of the expression of a selection of sarcomeric genes by RT-PCR confirmed that generally expression of these genes does start earlier. However, the expression level of these markers does not exceed that seen in WT and for some markers despite starting earlier their expression is lower overall, eg. *MYH7* and *TNNI3*. These results suggest differentiation is proceeding at an accelerated rate in comparison to the parental line, but that differentiation efficiency is not increased because of this.

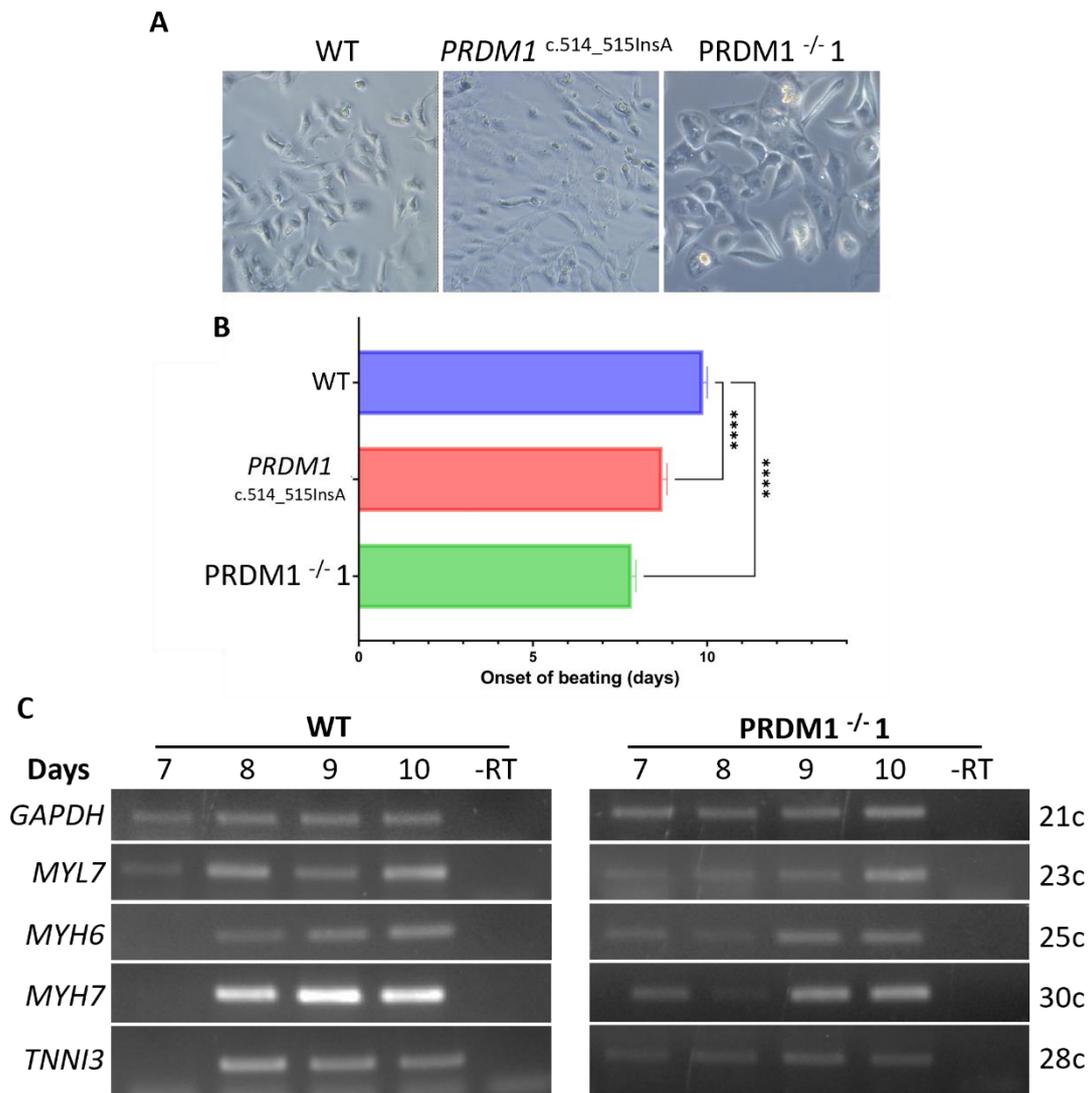


Figure 5.5 CM formation is slightly accelerated in *PRDM1* mutants

(A) Brightfield images of WT and *PRDM1* mutant CMs at day 32 of differentiation, taken at 10x magnification. (B) The results of daily monitoring of WT and mutant cells for beating from days 7-12. WT wells = 115, *PRDM1*^{c.514_515InsA} wells = 118. *PRDM1*^{-/- 1} wells = 48/ across 8-12 experiments. (C) Expression of a selection of sarcomeric genes analysed by RT-PCR in a WT and *PRDM1*^{-/- 1} differentiation, N = 2.

5.2.5 Differentiation efficiency is unaffected in the PRDM1 mutants

An accelerated rate of differentiation suggests the cells are committing to a CM fate earlier which could impact differentiation efficiency. Earlier commitment may reduce the amount of time available for proliferation of the cardiac progenitor pool, meaning less CMs will be formed. RT-PCR analysis revealed conflicting results between markers; *MYL7* and *MYH6* levels were comparable between the WT and PRDM1^{-/-} line at day 10, but expression of *MYH7* and *TNNI3* was slightly lower. Therefore, differentiation efficiency was examined at a higher resolution using IF. WT and PRDM1 mutant cells were examined at day 32 for expression of pan-myocardial marker TNNT2 (figure 5.6a). This revealed no statistical difference between the WT and mutant lines in the number of CMs formed (figure 5.6b). However, the amount of TNNT2 per cell was found to be 26% lower on average in the PRDM1^{c.514_515A} line (figure 5.6c).

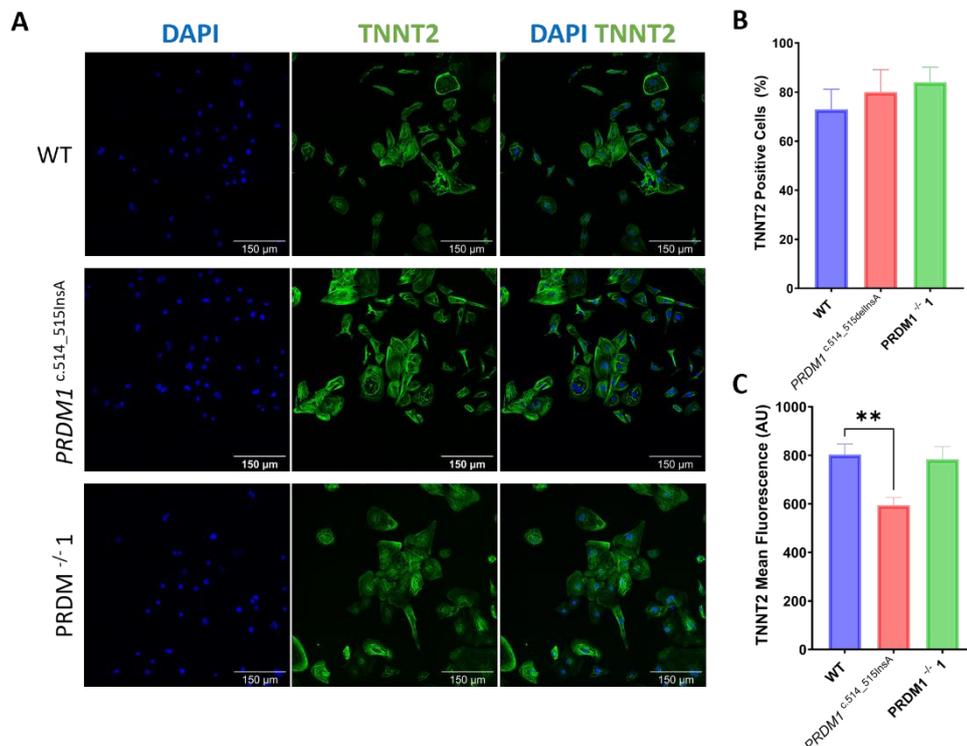


Figure 5.6 Differentiation efficiency is unaffected in PRDM1 mutant lines

(A) WT and PRDM1 mutant lines were differentiated into CMs using the CDM3 or GiWi protocols, then stained for pan-myocardial marker TNNT2 (green). Nuclei were stained with DAPI (blue). Representative examples of a field of cells for each cell line are shown at 20x magnification. (B) The proportion of cells positive for TNNT2 for each line is as follows; WT 73% ± 16.27 n = 576, PRDM1^{c.514_515InsA} 80.00% ± 18.35 n = 673, and for PRDM1^{-/-} 84.00% ± 12.30 n = 367. Figure legend continues on the following page.

Figure 5.6 continued (C) Mean fluorescence per cell was plotted for WT (n = 92), *PRDM1*^{c.14_515InsA} (n = 81), and *PRDM1*^{-/-} (n = 75). For B-C the data is shown as the mean average \pm SEM. A one-way ANOVA was used for comparison. Only statistically significant results are indicated, with p-values as follows: ** \leq 0.05.

5.2.6 *PRDM1* mutants form wild type-like cardiomyocytes

To fully inspect and characterise the phenotype of the cells formed multiple aspects of CM morphology were examined and scored using a classification system adapted from Ang *et al.* 2016. This system integrates parameters such as cell elongation, myofibril alignment, and sarcomere clarity to determine the quality of the CMs formed. Examples of cells belonging to each class are given in figure 5.7b and the criteria for classification are described in the accompanying table. Classes I and II represent organised cells, versus class III and class IV which represent more disorganised cells. From this analysis it was determined that the *PRDM1* mutants form a mix of CMs that are similar in condition to WT cells. A slight increase in the number of cells belonging to classes I and II was noted for the *PRDM1* mutants. Specifically, these percentages were 60% for WT, 62% for *PRDM1*^{c.514_515InsA}, and 68% for *PRDM1*^{-/-}. Representative images for each line are shown in figure 5.7a. Elongation although covered in the classification system was also explored quantitatively. In the adult heart CMs are anisotropic, and therefore elongation or a higher aspect ratio is considered an indicator of maturity in iPSC-CM cultures (Yang *et al.* 2014). No significant difference between WT and *PRDM1*^{-/-} cells was observed, whereas the *PRDM1*^{c.514_515InsA} cells showed a small but significant (p-value <0.05) 1.1 fold increase in elongation (figure 5.7e).

Additional indicators of CM quality and maturation status were examined including cell area and the number of nuclei per cell. No significant difference in cell area was observed between WT (n = 184) and *PRDM1*^{c.514_515InsA} (n = 164) cells, however *PRDM1*^{-/-} cells (n = 106) were found to be 1.4-fold larger on average (p-value < 0.001) (figure 5.7d), whereas no significant difference was observed in the proportion of mono-, di-, or polyploid cells between all lines. Together these results together demonstrate no consistent improvement or deterioration in the condition of the *Prdm1* mutant cells in comparison to WT.

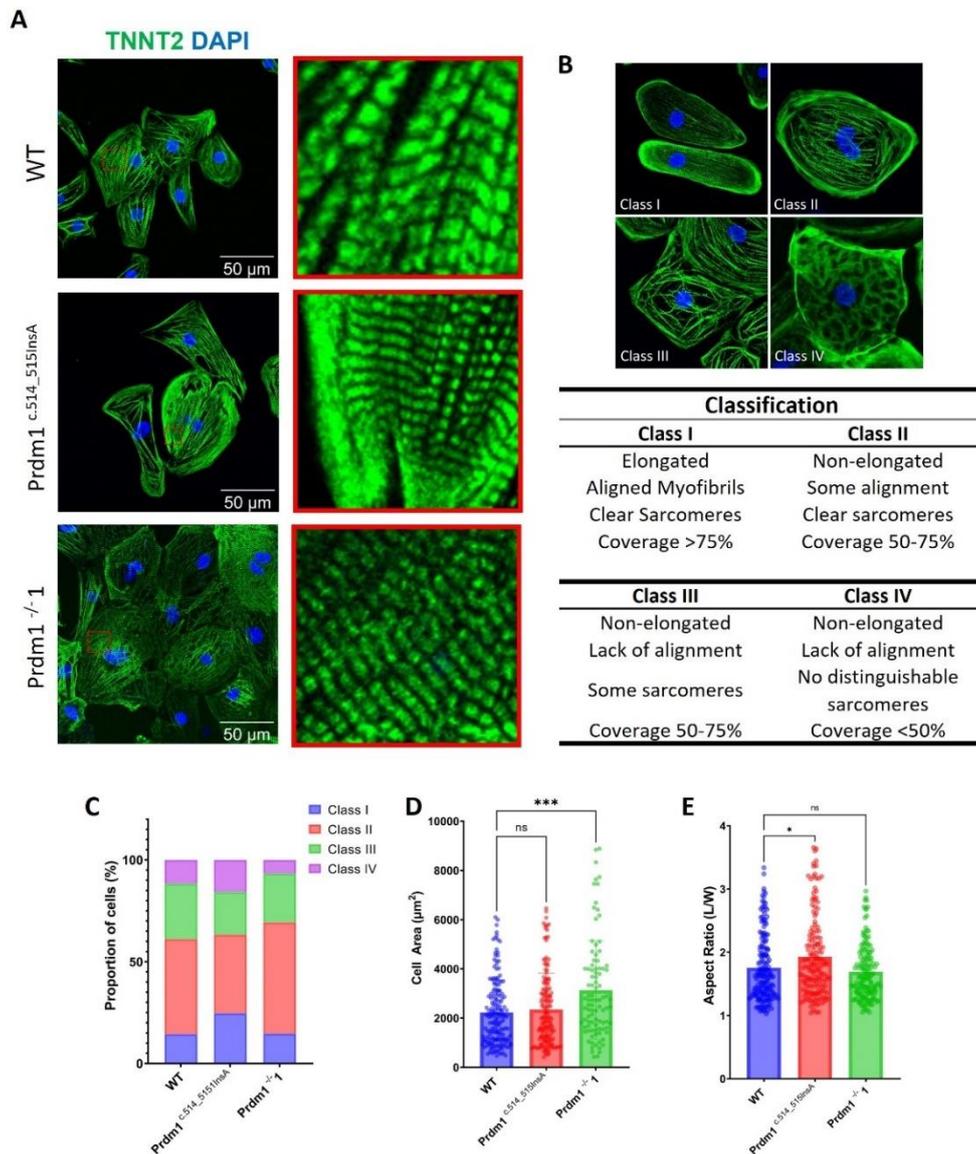


Figure 5.7 PRDM1 mutant iPS cells form WT-like CMs

(A) Staining for sarcomeric protein TNNT2 (green) in day 32 hCMs formed from WT and PRDM1 mutant cells, counterstained with DAPI (blue). Shown next to a zoomed image showing sarcomeric and fibril condition in the cells. (B) Example cells for each classification are shown, the associated table gives a breakdown of the criteria used to assign a cell to a class. (C) The proportion of cells in each class, WT $n = 180$, $PRDM1^{c.514_515InsA}$ $n = 158$, and $PRDM1^{-/-}$ $n = 150$. Statistical test used = Chi-squared, p -value = 0.0077. (D) The surface area of the cells was measured across the differentiations, WT cells were on average $2225 \mu\text{m}^2 \pm 101.30$, $PRDM1^{c.514_515InsA}$ cells were $2352 \mu\text{m}^2$ on average ± 119.20 , and $PRDM1^{-/-}$ cells were $3134 \mu\text{m}^2$ on average ± 206.20 . (E) Aspect ratio was measured for these same cells; WT – 1.75 ± 0.04 , $PRDM1^{c.514_515InsA}$ 1.93 ± 0.05 , and $PRDM1^{-/-}$ cells were 1.69 ± 0.04 . For figures D-E a one-way ANOVA was used for statistical comparison. P-values: * ≤ 0.05 , and *** ≤ 0.001 .

5.2.7 RNA-Seq analysis of PRDM1^{-/-} cells reveals roles in early gene regulation

RNA-seq was conducted for the WT and PRDM1^{-/-} line to determine if loss of PRDM1 had affected the transcriptome of these cells. Samples were taken at day 3 and day 7. Day 3 was chosen as *PRDM1* mRNA expression is highest at this time as reported by Churko *et al.* 2018 in their RNA-seq dataset, and PRDM1 protein levels peak at day 4 (figure 5.3a). Therefore, it is assumed that the earliest and largest effects on gene expression will be around this day. Day 7 was chosen as another time point to attempt to capture gene changes relating to the early onset of beating in the PRDM1^{-/-} mutant line. Only one repeat for each sample was sequenced due to limited resources, therefore statistical analysis could not be run on this data and timepoint selection was also constrained because of this. Nonetheless, the results have been analysed and the differences between the WT and PRDM1 mutant line are outlined below.

As introduced previously PRDM1 is a transcriptional repressor that is expressed most prominently at day 3-4. In agreement with this more genes were found to be differentially regulated at day 3 (1033) than at day 7 (355). Furthermore, at day 3 more genes were found to be upregulated (959) than downregulated (224) in the PRDM1 null lines, consistent with PRDM1's role as a repressor. Gene Ontology (GO) term enrichment analysis revealed that of the day 3 upregulated genes a significant number were associated with neuronal projection morphogenesis (GO: 0048812), and other terms related to neuronal development (figure 5.7b). Down-regulated terms at this time include GO: 0061311 'cell surface receptor signalling involved in heart development', which includes neural crest associated gene *MSX1* and early cardiac inducing factor *BMP4* (Ren et al., 2011).

Cardiac mesoderm marker *MESP1* was found to be downregulated at day 3, as was early cardiac progenitor marker *ISL1*, which is expressed in the cardiac lineage from the cardiac crescent stage in mice (Cai et al. 2003; Sun et al. 2007). Expression of *MESP1* and *ISL1* was also examined by RT-PCR at further days during differentiation (figure 5.8d). *MESP1* expression was similar between WT and the mutant at day 2 but was down-regulated at day-3. *ISL1* was found to be mildly downregulated throughout development. This indicates the expression of some cardiac specific genes is shifted slightly from its usual pattern at this early timepoint. However, the expression of additional cardiac lineage genes such as *GATA4* and *HAND2* was comparable between the WT and PRDM1^{-/-} line.

At day 7, as previously noted fewer genes are differentially expressed however the enriched GO: BP terms follow a similar trend to those seen at day 3, with those associated with synaptic transmission upregulated, and some cardiac associated GO: BP terms (Cardiac chamber morphogenesis, GO: 0003206) down regulated. The genes belonging to GO: 0003206 includes a several myocardial genes needed for CM function, such as *MYL7*, *MYH6*, and *MYH7*, which have also been examined by RT-PCR (Figure 5.5c) and where there was only observable down-regulation of *MYH7*. Taken together these results show a down-regulation of CM developmental genes at both time-points but given that the PRDM1 mutant lines produced WT like CMs it does not seem this initial down-regulation has a significant impact on the ability of the PRDM1^{-/-} cells to form CMs.

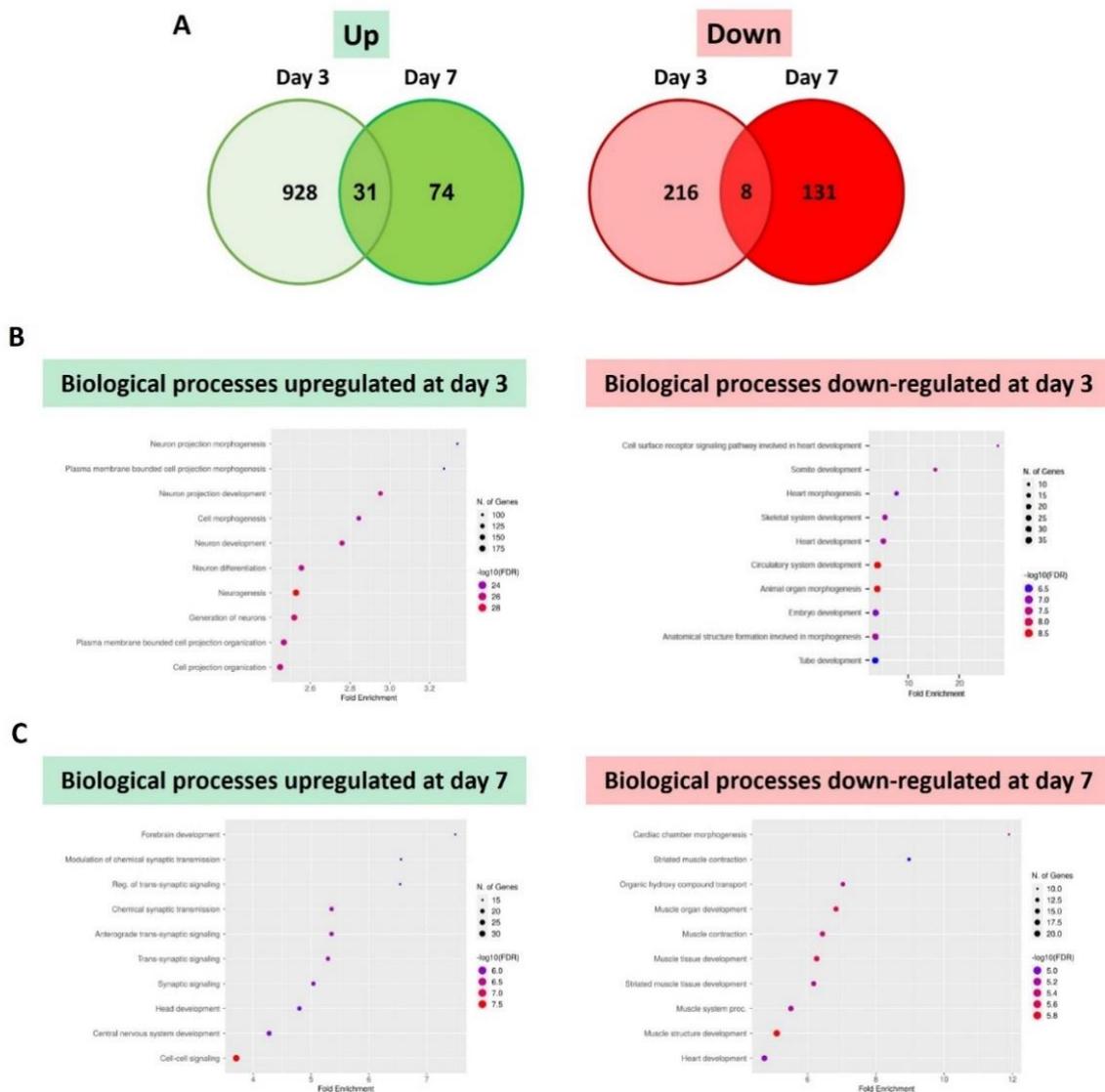


Figure 5.8 continues on the following page.

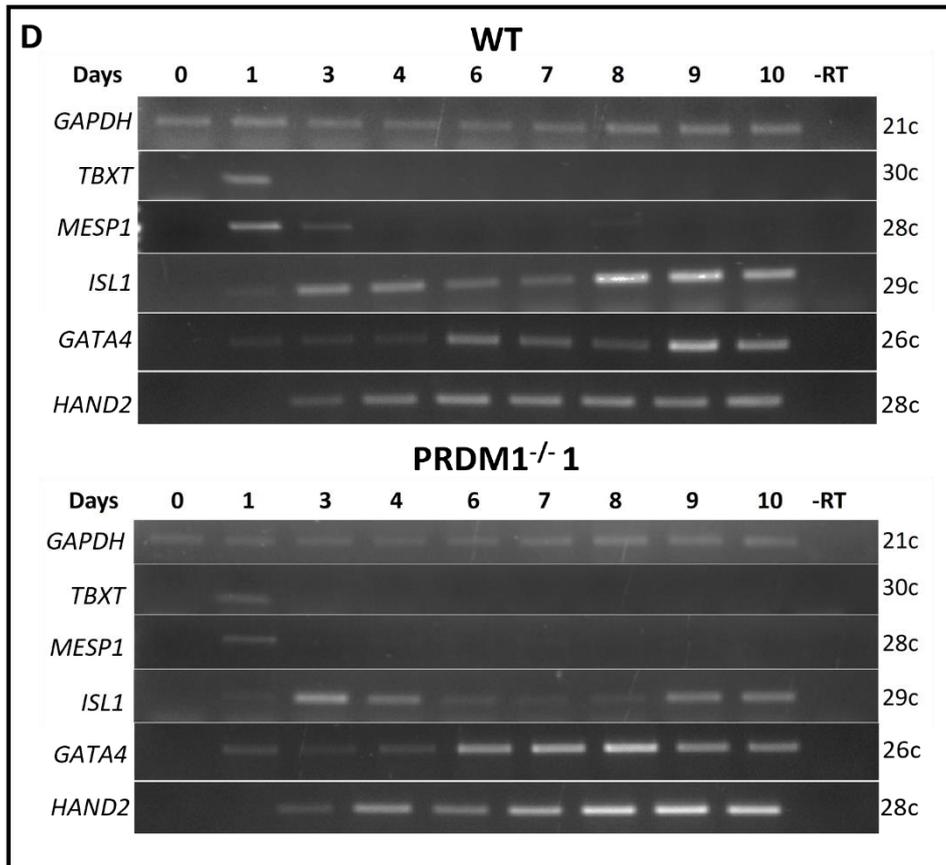


Figure 5.8 Analysis of transcriptional changes in PRDM1 mutants

(A) Comparison of the number of genes differentially regulated at days 3 and 7 in WT and PRDM1^{-/-} samples, n=1. (B) Analysis of the enriched GO:BP (Gene Ontology: Bio-processes) up-regulated and down regulated in the PRDM1^{-/-} lines at day 3, (C) and day 7. (D) RT-PCR analysis was conducted for mesoderm marker *TBXT*, cardiac mesoderm marker *MESP1*, and cardiac mesoderm and second heart field marker *ISL1* in WT and PRDM1^{-/-} cells during differentiation. *GAPDH* was used for normalisation.

5.3 Chapter 5 - Discussion

The information available for the role of PRDM1 in cardiac development is limited to steps after the onset of differentiation and in addition the requirement for PRDM1 in a human model of cardiomyogenesis has not previously been explored. Studies in mice show that PRDM1+ cells from the cardiac mesoderm and SHF give rise to a considerable proportion of the heart, and when it is knocked out at the cardiac mesoderm stage mice present with a range of OFT and AVSDs (Robertson et al. 2007; Vincent et al. 2014). Therefore, it was hypothesised that PRDM1 may be required during the early stages of iPSC-CM differentiation. From the investigations conducted herein it has been demonstrated that WT-like CMs can be formed from cells lacking one or all PRDM1 isoforms, thus PRDM1 expression is not essential for cardiomyogenesis under the conditions tested. However, the observation of a small but significant acceleration in the onset of differentiation and changes in the underlying gene program suggests that PRDM1 has a discrete effect on the cardiac GRN that may have implications for cardiac development. These results add to the understanding of how PRDM1 contributes to normal cardiac development, and the changes noted are explored in further detail below.

5.3.1 In the absence of PRDM1 CM commitment is accelerated

During this study beating was detected in WT cells as early as day 8, but on average started around day 9.89. For the PRDM1^{-/-} mutant the onset of beating was on average 2 days earlier than typically recorded for the WT parental line, and 1 day earlier for the PRDM1^{c.514_515InsA} mutant. Thus, it is accelerated in comparison to the parental line but this is not exceptionally early when compared with other studies and other WT cell lines (Lian et al. 2012; Burr ridge et al. 2014b; Burr ridge et al. 2015; Elorbany et al. 2022). Further comparison of the mutant and WT line by RT-PCR demonstrated the earlier onset of contraction was preceded by precocious expression of sarcomeric genes that would naturally be needed for this accelerated beating phenotype to present, suggesting the PRDM1 mutant cells commit to a cardiac fate at an accelerated rate. There are multiple examples of PRDM1 repressing differentiation in other developmental circumstances, for example; in mammary luminal and sebaceous gland stem cells, in migrating primordial germ cells, and enterocyte maturation (Horsley et al. 2006; Harper et al. 2011; Ahmed et al. 2016; Yamashiro et al. 2016). Therefore, in this situation it may be that the absence of PRDM1

leads to earlier CM beating through a lack of repressive activity during iPSC-CM differentiation allowing precocious instigation of the cardiac GRN.

In the other developmental contexts mentioned PRDM1 represses differentiation and in doing so helps to maintain cells in a progenitor state, therefore it was postulated that loss of PRDM1 may have a negative effect on maintenance of the cardiac progenitor pool. If this was the case, we may expect to see a reduced number of CMs formed but no negative effect on differentiation efficiency was seen in the cells when they were examined by IF for expression of CM marker TNNT2. Transcriptomic analysis of the cells flagged early cardiac progenitor marker *ISL1* as being down regulated in the mutant. *ISL1* is initially a marker of the FHF, and later becomes restricted to the SHF (Cai et al. 2003; Van Vliet et al. 2012). Following formation of the linear heart tube from the FHF the heart grows through the addition of cardiac progenitors from the *ISL1* positive SHF, and thus these cells are essential for normal heart development (Van Vliet et al. 2012). Here, *ISL1* expression was found to be lower throughout differentiation in the PRDM1^{-/-} mutant when examined by RT-PCR, suggesting there may be an effect on both the FHF and SHF cell populations.

In Prdm1^{fl/fl} Cre-Mesp1 mice PRDM1 is deleted at the cardiac mesoderm stage and therefore both FHF and SHF derivatives have the potential to be affected. However, formation of the left ventricle, a FHF derivative is apparently normal in these mice whereas defects are observed in the ventricular septum and OFT (Vincent et al. 2014). These features are both derived from the ISL1+ SHF, suggesting that the SHF is more sensitive to loss of PRDM1 than the FHF. The Wnt modulation protocols used herein have been found to yield primarily (~90%) ventricular like FHF derived CMs (Galdos et al. 2022). Organoid differentiation protocols such as those devised by (Drakhlis et al. 2021; Galdos et al. 2022) have shown the potential to produce a greater proportion of SHF derivatives. This is one method that could be used to examine if one field is more affected by the loss of PRDM1 than the other in human cardiomyogenesis.

5.3.2 Preliminary transcriptome analysis for PRDM1^{-/-} mutants reveals upregulation of non-cardiac gene programs

Another theme that was noted from the RNA-seq data was the de-repression of non-CM gene programs in the PRDM1^{-/-} mutant at both day 3 and day 7. The genes upregulated were generally associated with neural GRN terms. This corresponds with experiments from

zebrafish which shows that overexpression of Prdm1 during their development results in reduced anterior development affecting the forebrain, eyes, and anterior mid-brain (Wilm and Solnica-Krezel 2005) indicating high levels of Prdm1 expression interfere with the activation of the GRNs governing neural development. Despite these early transcriptome changes in the PRDM1^{-/-} mutant there is no indication that any neural cell types are present in the differentiated population, which at day 32 mostly contained CMs with on average 84% of cells positive for TNNT2. Therefore, these upregulated neural terms likely represent an early failure to completely repress the expression of these genes during CM development, whilst further inductive steps not present in a CM differentiation protocol would be needed to firmly establish a neural fate. Further to this point PRDM1 is required later in neural development for the sequential specification of neural, neural crest and sensory progenitor cells in chick (Prajapati et al. 2019). Therefore, timely and regulated expression of Prdm1 seems to be an important factor in the development of neural lineages.

5.3.3 Differences in PRDM1 isoform expression do not seem to impact CM development

Of the lines created in this study, the PRDM1^{-/-} line was shown to be a null line, however the *PRDM1*^{c.514_515InsA} line still expressed the long form of PRDM1 (PRDM1 α /CCDS5054). No consistent differences in the phenotype between these cell lines was seen in the format in which they have currently been tested. The different forms of the protein do however have different properties, the short form of PRDM1 (PRDM1 β /CCDS34505) is a less effective transcriptional repressor than the long form (PRDM1 α /CCDS5054) (Györy et al. 2003). The consequences of this have been explored in multiple cancer models. These studies have found that expression of the short form is induced in myeloma, colorectal cancer and in lymphocytes following Epstein Barr Virus infection which can cause lymphoma (Vrzalikova et al. 2012; Liu et al. 2018a; Romero-García et al. 2020). Activation of the long form of PRDM1 was associated with apoptosis in myeloma cells, whereas the short form was not (Romero-García et al. 2020). Suggesting the long form may act as a tumour repressor. In contrast to this in colorectal tumour organoids when the short form of PRDM1 was over-expressed cell proliferation rate decreased and the PRDM1 β (short-form) expressing cells were eliminated from the organoid (Liu et al. 2018a). Thus, suggesting that like the long form the short form may also act as tumour repressor in other cell types. These studies suggest PRDM1 expression can modulate apoptosis, proliferation, and can act as a cue for cell elimination but that these scenarios may be dependent on the isoform of

PRDM1 expressed. The lines created herein may be useful for further examining these scenarios in other developmental contexts.

5.3.4 Conclusions

The results presented in this chapter give some insights into the contribution of PRDM1 to cardiac development. From the points addressed here it is clear that loss of PRDM1 does not have a major impact on the outcome of cardiac development in human iPSC-CMs, despite small but tangible differences in beating onset and gene expression profile in the mutants during their development. Taken together we can conclude that PRDM1 is contributing to the cardiac GRN, however the loss of this factor is compensated during development such that normal CMs can still be produced. It is not clear if a failure to down regulate PRDM1 expression would be detrimental to CM development, but this is likely possible given its identification as a negative target of GATA4 regulation. Examining ectopic expression of PRDM1 during iPSC-CM differentiation would be a clear next step to further investigate the effects of PRDM1 on cardiac differentiation. The approaches required to achieve this along and additional future directions for this work are discussed in Chapter 7.

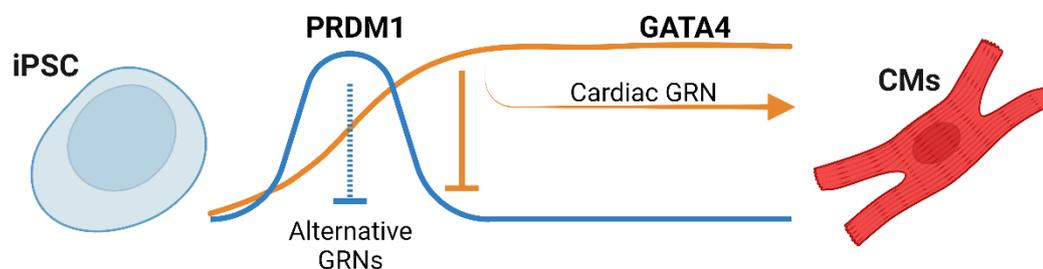


Figure 5.9 PRDM1 represses alternative gene programs in early cardiomyogenesis

During iPSC to cardiomyocyte (CM) differentiation PRDM1 represses the expression of alternative gene programs (eg. neural) early in the process of development. GATA4 expression rises and represses *PRDM1*, whilst activating cardiac specific gene expression. When PRDM1 is knocked out the expression of alternative gene regulatory networks (GRNs) is de-repressed, but a lack of additional factors means these GRNs are not activated robustly enough to induce a fate change. It remains to be determined whether prolonged PRDM1 expression would also lead to repression of the cardiac GRN, but the slight acceleration of differentiation in the PRDM1 mutants suggests this may be the case.

Chapter 6 – Deletion of amino acids 311-337 from a predicted transcriptional activation domain of GATA6 leads to the production of hypertrophic cardiomyocytes

6.1 Introduction

In Chapter 3 it was established that GATA4 is absolutely required for iPSC-CM differentiation. In this final results chapter, the aim is to investigate the role of the closely related TF GATA6 in this process. Mutations in the GATA6 locus are associated with a range of human CHDs and with the formation of dilated cardiomyopathy later in life, identifying GATA6 as a CHD and CVD risk gene (Kodo et al. 2009; Maitra et al. 2010; Sun et al. 2012; Wang et al. 2012; Zheng et al. 2012; Xu et al. 2018; Whitcomb et al. 2020). This indicates that GATA6 is an important contributor to the human cardiac GRN. However, few studies have investigated the roles of GATA6 in a human model of cardiac development. What is known about this factor is considered below.

6.1.1 Roles for GATA6 in early development

Like all members of the GATA family of TFs GATA6 possess a DBD made up of two zinc fingers that recognise (A/T) GATA (A/G) motifs throughout the genome (Lentjes et al., 2016; Peterkin et al., 2007). The earliest requirement for GATA6 in development is in the formation of primitive endoderm, which precedes establishment of the VE, as reviewed in section 1.3.1 (Morrissey et al. 1998; Koutsourakis et al. 1999). Following this GATA6 is expressed in a variety of mesoderm and endoderm derived tissues including the heart, intestines, liver, lungs, and pancreas (Laverriere et al. 1994; Suzuki et al. 1996; Morrissey et al. 1997a; Decker et al. 2006). Throughout development the expression domain of GATA6 often overlaps with GATA4, this includes structures such as the heart and pancreas. This combined with their ability to recognise the same DNA motif, through their highly conserved DBDs, gives them the potential to recognise and regulate one another's targets. The significance of this is considered below.

To circumvent the requirement for GATA6 in primitive endoderm development and to investigate its later roles researchers have utilised tetraploid complementation (Zhao et al. 2005). This is akin to what has been performed for GATA4 to bypass its role in VE formation, a derivative of primitive endoderm (Watt et al. 2004). These studies revealed that GATA6 is dispensable for early heart formation in mice, but loss of GATA4 leads to a

range of cardiac defects (Watt et al. 2004; Zhao et al. 2005). Interestingly replacement of the GATA4 coding sequence with *GATA6* cDNA, is a viable substitute for GATA4 in visceral endoderm formation (Borok et al. 2016). However, cardiac defects of a severity similar to that seen by Watt *et al.* are still observed, therefore indicating that GATA6 can compensate for a loss of GATA4 expression in some circumstances but only offers a partial rescue of the cardiac phenotype.

6.1.2 Studying GATA6 in human iPSC models of development

In other developmental contexts it seems that GATA4 and 6 exhibit complete redundancy with one another. GATA4 or GATA6 single knockout mice are able to form a normal pancreas, homozygous knockout of both GATA4 and 6 is needed to block pancreas formation (Carrasco et al. 2012; Xuan et al. 2012). This suggests that either factor can take on all roles in this developmental context. However, this finding does not extend to human development. In humans, heterozygous LoF mutations in GATA6 are known to cause pancreatic agenesis (Stanescu. et al., 2015; Suzuki et al., 2014). Indicating that human pancreatic development is more sensitive to a reduction in GATA6 activity than what has been observed in mice. Using human iPSCs it has been possible to explore the requirement for GATA6 during the formation of β -islet like pancreatic cells. This work has revealed that human pancreatic development is very sensitive to both GATA6 and 4 levels. Removal of one GATA6 allele impairs definitive endoderm formation, a precursor step to pancreatic progenitor formation, whereas GATA4^{-/-} iPSCs are able to form definitive endoderm but progenitor formation is reduced. Furthermore, deletion of either allele impairs pancreatic progenitor differentiation (Shi et al. 2017). Thus, using this model these researchers have identified overlapping and individual roles for GATA4 and 6 at different time-points within pancreatic development, that were not apparent in mice. It is possible that this is also the case for cardiac development.

A report detailing the cardiogenic potential of a GATA6^{-/-} iPSC cell was recently published (Sharma et al. 2020). This report indicates that in contrast to mouse development (Zhao et al. 2005; Zhao et al. 2008), loss of GATA6 expression is not tolerated in human iPSC-CM development. A decrease in *MESP1* expression at day 4 of differentiation indicates there are likely early deficiencies in cardiac development in the absence of GATA6. Further work is required to identify later roles for Gata6 during iPSC-CM differentiation.

6.1.3 Roles for GATA6 in later stages of cardiac development

Despite the limitations of mouse models it should be noted that they have been instrumental in identifying later roles for GATA6 in cardiac development, that have informed how we understand human disease. For example, deletion of GATA6 at the cardiac progenitor stage using an NKX2-5 Cre driver leads to AVSDs formation in mice (Van Berlo et al. 2010), whereas PTA can be induced in mice by the deletion of GATA6 from smooth muscle and neural crest lineages (Lepore et al. 2006). These studies identify roles for GATA6 in the SHF and recognise its contribution to non-CM cell types in the heart. Furthermore, in adult mice it has been shown that like GATA4, GATA6 is also a mediator of adaptive hypertrophy in response to pressure overload (Liang et al. 2001; Oka et al. 2006) and overexpression of GATA6 alone is sufficient to cause hypertrophy in older mice (Liang et al. 2001; Van Berlo et al. 2010). These results demonstrate that GATA6 likely has diverse roles throughout cardiac development and further study is required to determine what these are.

6.1.4 Project Rationale

The results presented by Sharma *et al.* 2020 demonstrate that GATA6 plays a more significant role in human cardiac development than anticipated based on studies in model organisms. Warranting further exploration of GATA6's role in iPSC-CM differentiation models. Knockout of GATA6 leads to early deficiencies in cardiac mesoderm formation with no CMs produced, thus limiting the exploration of later roles. Rather than a knockout mutation the GATA6 line created in this project carries a deletion of amino acids 311-337, that based on its homology to GATA4 is predicted to be a transcriptional activation domain (Morrisey *et al.* 1997). Allowing for assessment of how this region of the protein contributes to GATA6 function during iPSC-CM development.

6.2 Results

6.2.1 Creation of a *GATA6* mutant iPSC cell line using CRISPR-Cas9 gene editing

At the start of this project in 2018 the initial aim was to create a *GATA6*^{-/-} line, however, since then Sharma *et al.* 2020 demonstrated through the creation of their *GATA6*^{-/-} lines that *GATA6* is essential for iPSC-CM differentiation. The targeting strategy used was set with the intention of answering the initial aim of this project. Therefore, a two-guide strategy was used to target exon 2 of the *GATA6* gene locus (figure 6.1a). The guides were placed 82bp apart to create a frameshifting mutation that would induce NMD and thus create a null line. A likely mutant was identified using gDNA PCR (figure 6.1b), following sanger sequencing it was revealed that the mutant carried the intended 82bp deletion, but with the insertion of 1 adenine residue (figure 6.1c). This insertion brings the CDS back into frame, therefore this is likely not a null mutation but creates a mutant form of *GATA6* that is distinct from what was previously published by Sharma *et al.* 2020.

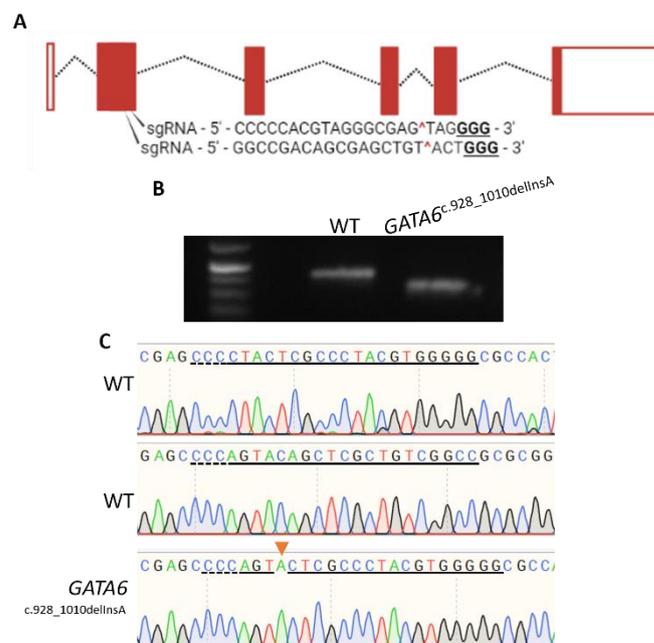


Figure 6.1 Creation of a *GATA6* mutant iPSC line using CRISPR-Cas9 gene editing

(A) The *GATA6* locus with the sgRNAs used for editing. Boxes represent exons with coding regions filled, whereas non-coding regions are unfilled. Introns are represented by dotted lines. (B) Separation of the PCR products generated from amplification of the region covering the edit site in the WT and *GATA6* mutant. (C) Sanger sequencing trace files of the WT and *GATA6*^{c.928_1010delInsA} RT-PCR amplicons. PAM sites are indicated by dotted lines, whilst protospacers are solidly underlined. The remnants of the sgRNA target sequences are indicated for the mutant line, and the inserted base is indicated with an orange arrowhead.

6.2.2 The deletion affects a CHD associated region in transcriptional activation domain 2

Given that the mutation leaves the CDS in frame, it was expected that mRNA and protein would still be produced from the *GATA6* locus. This was confirmed using RT-PCR analysis of day 10 iPSC-CMs, demonstrating that *GATA6* expression in the mutant is comparable to WT (figure 6.2a). Sanger sequencing confirmed the mutation seen in the gDNA was present in the cDNA, likewise WB confirms that protein is produced and a decrease in protein size is seen (figure 6.2b). It should be noted here that the apparent increase in *GATA6* protein was not consistent between repeats (n = 2) and requires further investigation. Based on the cDNA sequence from the *GATA6*^{c.928_1010delInsA} line amino acids residues 311_337 will be missing. These residues belong to a predicted transcriptional activation domain and are highlighted in figure 6.2c. The importance of this domain in the aetiology of ASDs is suggested by the number of mutations associated with ASD cases in this region such as; G321R, H330_H333 del, H331Q, H331_H333 del, H332R, H333del, H333dup, and H332_H333del. One of these mutations (H333del) was found to be associated with two cases of ASD and PTA (table 6.1).

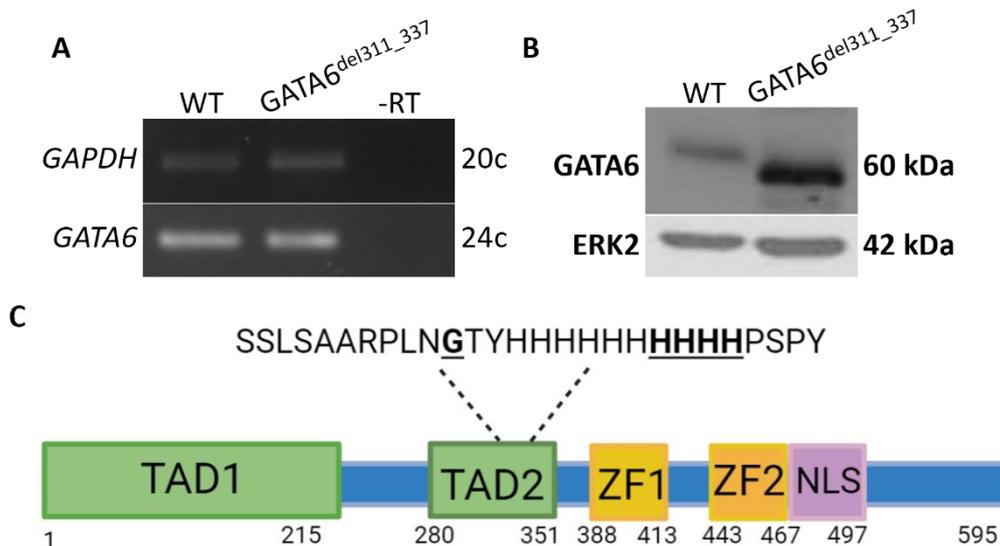


Figure 6.2 The *GATA6* mutant protein lacks residues 311-337

(A) RT-PCR for *GATA6* expression in WT and the *GATA6*^{928_1010delInsA} mutant, the iPSCs were differentiated into hCMs using the CDM3 protocol, and samples taken at day 10. (B) Western blot showing *GATA6* expression in WT and mutant day 10 hCMs. *ERK2* has been used as a normalisation control, n = 2. (C) An overview of *GATA6*'s protein structure, with the deleted residues highlighted above. (TAD=transcriptional activation domain, ZF=Zinc Finger, NLS=nuclear localisation sequence). Residues associated with atrial ventricular septal defects are underlined in bold.

Table 6.1 GATA6 amino acid variants with known congenital heart defect associations located in the deletion region (residues 311-337)

A summary of the GATA6 amino acid variations that have been identified in patients with atrioventricular septal defect 5 (AVSD 5), and persistent truncus arteriosus (PTA) that belong to the deletion region created in the GATA6 mutant iPSC line.

Data taken from the NCBI ClinVar database (Landrum et al. 2018).

Protein Change	Associated Condition	# of reports	NCBI Accession #
p.His331_His333del	AVSD 5	1	VCV000835877
p.His330_His333del	AVSD 5	1	VCV000655336
p.Gly321Arg	AVSD 5	1	VCV001000801
p.His332Arg	AVSD 5	1	VCV000540133
p.His333dup	AVSD 5	1	VCV000404060
p.His332_His333del	AVSD 5	1	VCV000240131
p.His333del	AVSD 5 & PTA	2	VCV000129132, VCV000412723
p.His331Gln	AVSD 5	1	VCV001497308

6.2.3 Genomic integrity and pluripotency are maintained in the GATA6 mutant

Given the non-null nature and location of the mutation as described above and its association with CHDs such as AVSD 5 and PTA, the line was deemed worthy of further investigation. The first step in this process was to confirm that pluripotency and genomic integrity has been maintained after editing. Expression of WT GATA6 has previously been implicated in the maintenance of pluripotency in iPSC colonies (Nakanishi et al. 2019), therefore it is possible that disruption of this gene may affect pluripotency. Maintenance of pluripotency was analysed using a combination of methods. Brightfield microscopy images show that the WT and mutant lines form tightly packed rounded colonies characteristic of iPSCs (figure 6.3 a). The expression of pluripotency factors POU5F1 and SOX2 in the WT and *GATA6*^{c.928_1010delInsA} lines was confirmed using IF, which showed expression of these factors was restricted to the nucleus of non-dividing cells as expected (figure 6.2b). Flow cytometry was then used to quantify this at the population level. The proportion of POU5F1+ cells was similar between the WT and mutant iPSCs (both >97%). Finally, RT-PCR analysis shows that mRNA expression of pluripotency factors POU5F1, SOX2, and an additional pluripotency factor NANOG is indistinguishable from WT (figure 6.3b).

In addition to concerns about pluripotency, the introduction of dsDNA breaks through CRISPR-Cas9 has the potential to cause larger genomic rearrangements (Jianli Tao et al. 2023). To confirm genomic integrity had been maintained the genomes of the edited lines were compared to that of the WT Rebl Pat parental line using a microarray covering 654,027 markers across the human genome. No changes between the WT and GATA6 mutant line were detected (table 6.2).

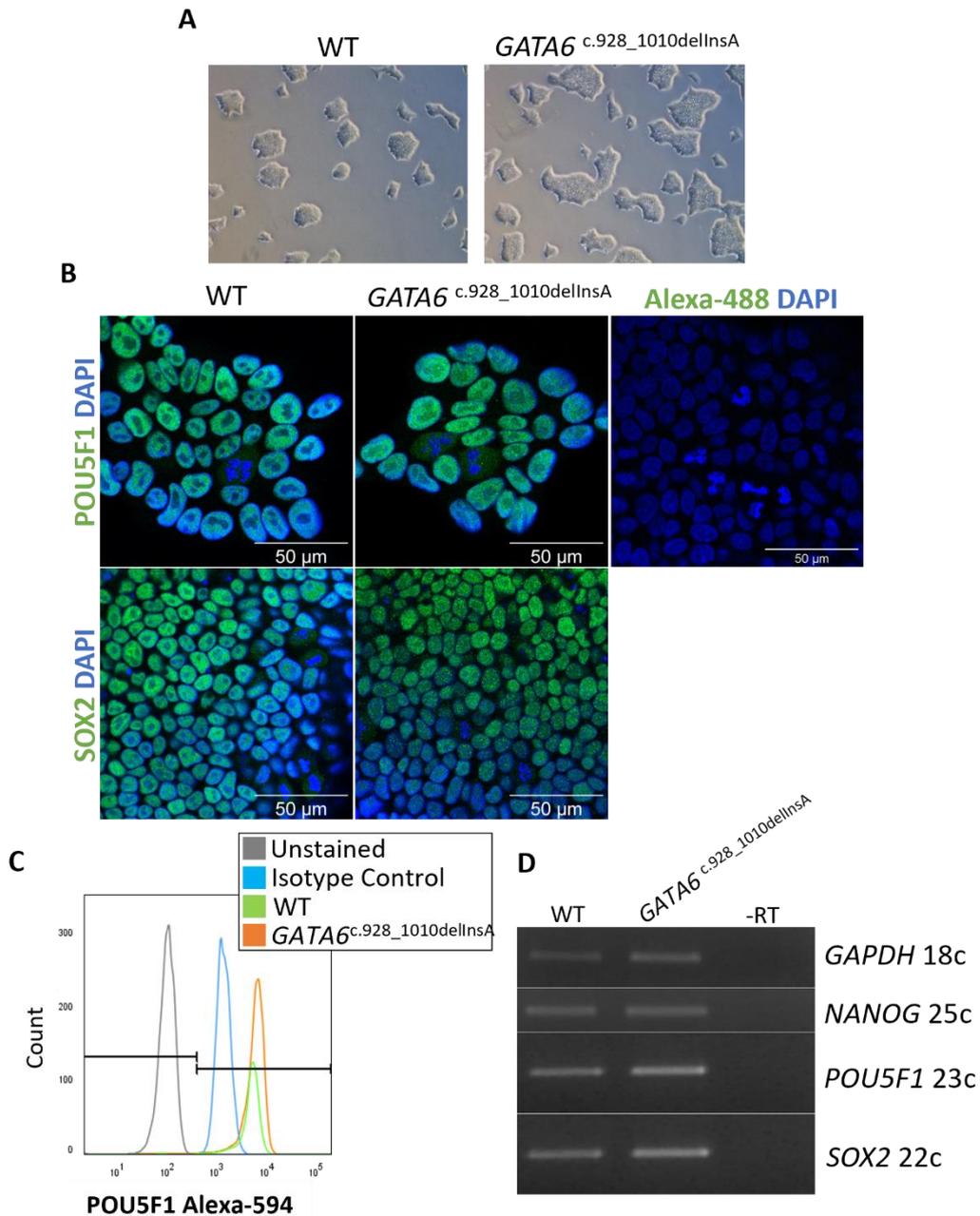


Figure 6.3 *GATA6*^{c.928_1010delInsA} mutant iPSCs maintain pluripotency

(A) Brightfield 4x images showing the rounded compact colonies formed by both WT and *Gata6*^{c.928_1010delInsA} line. (B) Immunofluorescence showing the expression and localisation of pluripotency markers POU5F1 and SOX2 to nuclei of WT and *GATA6*^{c.928_1010delInsA} iPSC cultures. (C) Flow cytometry for WT and the *GATA6* mutant samples are as follows; unstained (grey), isotype control (blue), WT iPSCs (green), *GATA6*^{c.928_1010delInsA} iPSCs (orange). (D) RT-PCR showing the expression of pluripotency markers (*NANOG*, *POU5F1*, and *SOX2*) in WT and *GATA6*^{c.928_1010delInsA} iPSCs.

Table 6.2 A summary of the genomic integrity of the WT and *GATA6*^{c.928_1010delInsA} lines

A summary of any genomic variation that was flagged by the PennCNV software. No differences between WT and the *GATA6* mutant line were seen.

Cell Line	Location	Event	Size	Transcripts from affected regions
Rebl Pat WT	chr2:139414906 -139884468	Duplication	469,563	Contains lncRNA AC016710, pseudogene MRPS18BP2, small nucleolar RNA SNORA72, pseudogene RPL9P13, and lncRNA AC078851.
Rebl Pat WT	chr19:20605360 -20700135	Deletion	94,776	Contains lncRNA AC010636, pseudogene AC010329, and ZNF626.
<i>GATA6</i> ^{c.928_1010delInsA}	chr2:139414906 -139884468	Duplication	469,563	Contains lncRNA AC016710, pseudogene MRPS18BP2, small nucleolar RNA SNORA72, pseudogene RPL9P13, and lncRNA AC078851.
<i>GATA6</i> ^{c.928_1010delInsA}	chr19:20605360 -20700135	Deletion	94,776	Contains lncRNA AC010636, pseudogene AC010329, and ZNF626.

6.2.3 The deletion of amino acids 311-337 may lead to increased GATA6 transcriptional activation

GATA6 can act as a transcriptional activator and repressor in different cellular contexts, and this is likely dependent on the presence of different combinations of co-activators and co-repressors (Charron et al. 1999; Yin and Herring 2005; Ghatnekar and Trojanowska 2008). Here, the ability of the GATA6 mutant protein to activate transcription was assessed using *Xenopus laevis* embryos as previously conducted by (Gallagher et al. 2012). Embryos were injected with WT or *GATA6*^{c.928_1010delInsA} synthetic capped mRNA, a reporter plasmid containing firefly luciferase under the control of a GATA motif, and a plasmid constitutively expressing *Renilla* luciferase under the control of the TK promoter (RL-TK), which served as normalisation control. Embryos injected with the plasmids alone provided a baseline. In comparison to baseline WT GATA6 leads to a 4.3-fold average increase in luciferase activity. For the GATA6 mutant transcriptional activation was more pronounced being on average 4-fold higher than WT GATA6 or 17-fold higher than the baseline (n = 5). Therefore, it seems the mutant GATA6 produced in this study has increased transcriptional activity. It should be noted that although the embryos were injected with the same amounts of mRNA it has not possible to confirm that protein levels between the samples were the same. Thus, it cannot be determined at this time whether this increase is due to increased intrinsic transcriptional activity or an increase in protein stability.

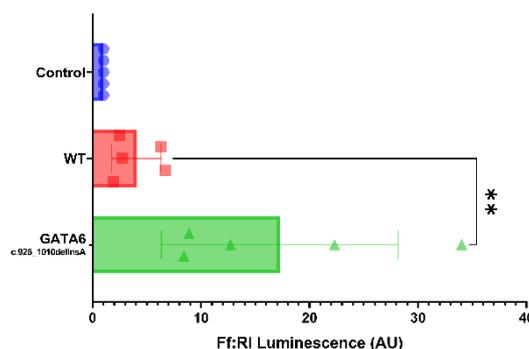


Figure 6.4 Transcriptional activity is likely increased in the GATA6 mutant

The transcriptional activity of WT and *GATA6*^{c.928_1010delInsA} was assessed through co-injection of their mRNAs into *X. laevis* embryos, 30 pg each of 2XGATALuc a luciferase reporter driven by two GATA elements, and TK-RL (Promega). Injection of the plasmids alone was used as a control. N = 5. A Student's t-test was used for statistical comparison between the samples. P-value ** ≤ 0.005 .

6.2.4 Differentiation efficiency and beating onset is unchanged in the *GATA6* mutant

A previous report established a positive association between *GATA6* expression in iPSCs and CM differentiation efficiency (Yoon et al. 2018). Analogous to this, the *GATA6* mutant presented here demonstrates increased transcriptional activity. Therefore, it was hypothesised that the mutation may have a positive effect on differentiation efficiency. Yoon *et al* 2018 based their conclusion primarily on the numbers of beating foci present in embryoid bodies formed from iPSCs with different *GATA6* expression levels. In this Chapter, differentiations were conducted in 2D using the CDM3 and GiWi Wnt modulation protocols (Lian et al. 2013; BurrIDGE et al. 2014b), in this format beating starts in isolated foci or often as a single synchronous sheet or network. Through the observation of these different beating patterns, it is possible to estimate differentiation efficiency at a low resolution. The cells were monitored for beating from day 7-12 across 14 experiments, and no consistent differences in beating were perceptible. Similarly, no significant difference in beating onset was detected, with WT cells starting to beat at day $9.14 \pm \text{SEM } 0.11$ and the *GATA6*^{c.928_1010delInsA} cells starting to beat at day $9.34 \pm \text{SEM } 0.11$ (figure 6.5 b).

Immunofluorescence detection for pan-myocardial marker, TNNT2 was used to determine differentiation efficiency with greater resolution. On average the proportion of CMs present per experiment was higher for the mutant line at $72.2\% \pm \text{SEM } 9.33$ TNNT2+ cells versus $57.71\% \pm \text{SEM } 9.65$ for WT (figure 6.4 a and c). However, this difference was not found to be significant ($p\text{-value} = 0.321$), possibly due to the large variation in differentiation efficiency seen generally for iPSC-CM differentiation. Additionally, when these values are considered in comparison to the levels seen in other experiments, conducted during this thesis (see section 5.2.5), the *GATA6*^{c.928_1010delInsA} efficiency rates do not stand out. During this analysis the amount of TNNT2 per cell was also recorded, and this did yield a significant result ($p\text{-value} < 0.0001$), with the mutant having 1.9-fold more TNNT2 signal per cell, than WT (figure 6.5 d), suggesting that the mutation present in the *GATA6*^{c.928_1010delInsA} line may be having other effects on CM phenotype.

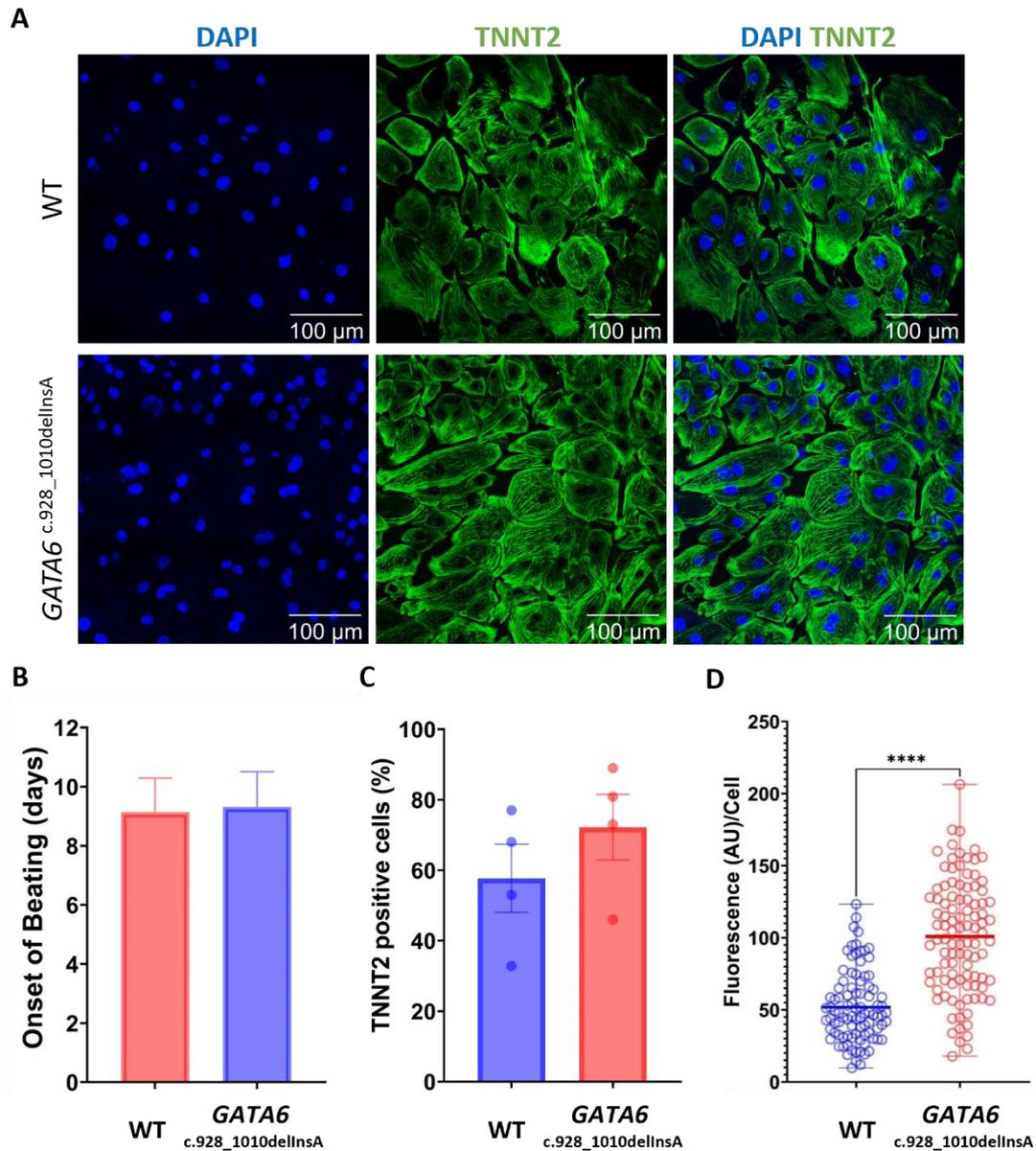


Figure 6.5 Differentiation efficiency is unaffected in the mutants but TNNT2 expression per cell is increased

(A) A panel showing the expression of TNNT2 (green) in a field of WT and *GATA6*^{c.928_1010delInsA} CMs at day 32, DAPI (blue) has been used to counterstain. (B) A graph showing the first day on which beating was noted for both lines. Recorded between days 7 – 12 of differentiation. WT n = 111 wells, and for *GATA6*^{c.928_1010delInsA} n = 125 wells across 14 independent differentiations. (C) The number of TNNT2 positive cells from IF images like those shown in panel A, was analysed at day 32 across 4 different experiments. (D) TNNT2 mean fluorescence measured per cell for WT n = 92 and for *GATA6*^{c.928_1010delInsA} n = 100.

6.2.5 *GATA6*^{c.928_1010delInsA} expression leads to the production of hypertrophic CMs

Upon closer inspection of the cells at a higher magnification, it became apparent that the mutant line exhibits a greater prevalence of large disorganised cells. Staining was conducted for two pan-myocardial markers; TNNT2 and MYBPC3 in day 32 WT and *GATA6*^{c.928_1010delInsA} CMs. TNNT2 has been used to quantify the differences seen, with MYBPC3 staining following a similar trend (see figure 6.5 a-b). On average the *GATA6* mutant cells (n = 199) were 1.52-fold larger than their WT counterparts (n = 192) (figure 6.6 d). The cells were assigned to classes based on criteria adapted from Ang *et al.* 2016, described previously in sections 4.1.10, and 5.2.5. Classes I and II represent more organised cells within the population, whilst classes III and IV represent more disorganised cells (figure 6.6 e). In the WT population 57% of the CMs present aligned with classes I-II, and 43% with classes III-IV. The *GATA6* mutant CMs were found to be more disorganised on average, with 30% of the CMs present aligning with classes I-II, and 70% aligning with classes III and IV. No significant differences were found in the other metrics of cellular condition measured such as circularity and number of nuclei.

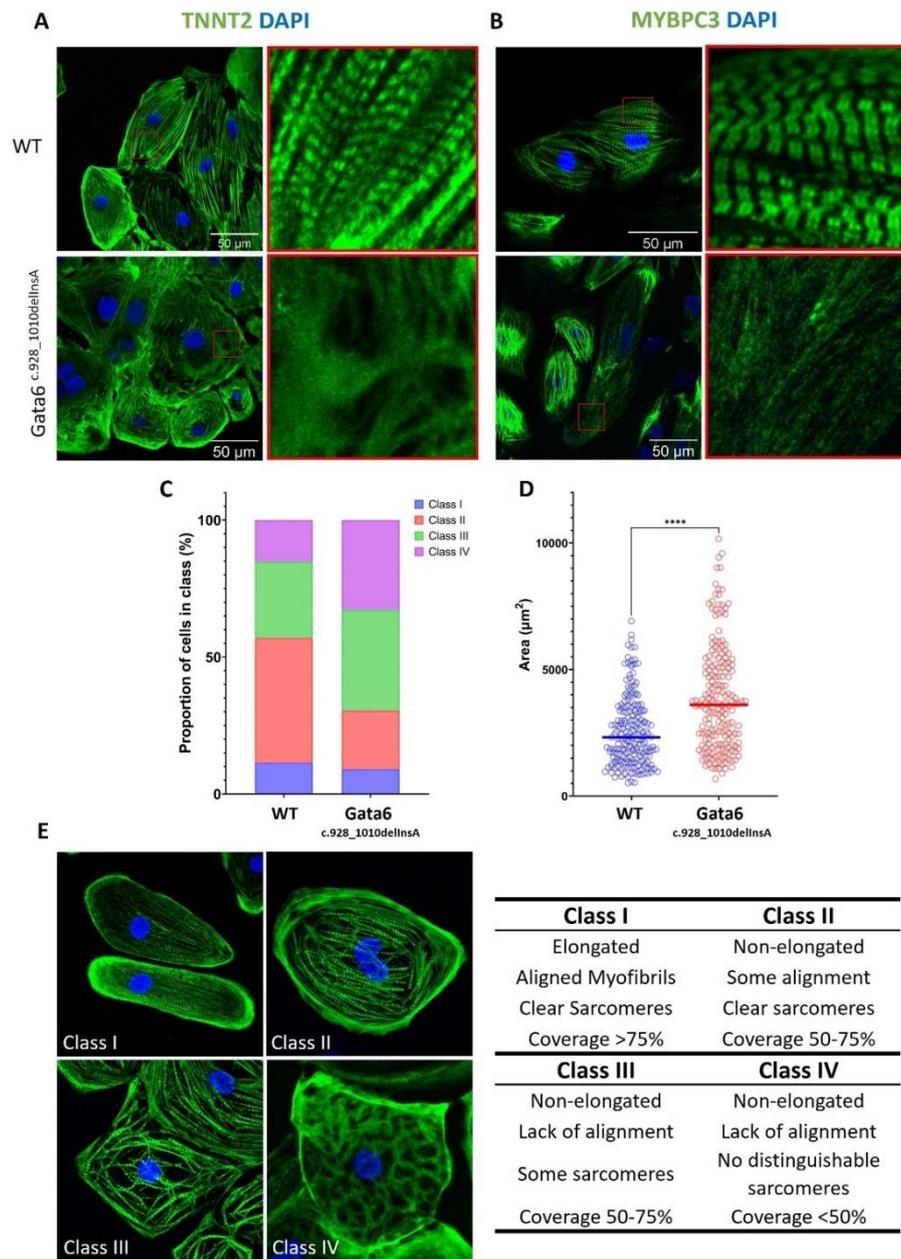


Figure 6.6 The cardiomyocytes formed from *GATA6*^{c.928_1010delInsA} iPS cells are hypertrophic and disorganised

Representative images of WT and GATA6 mutant cells after 32 days of differentiation, with an enlarged image of sarcomere staining shown. (A) IF staining for TNNT2 (green) and counterstained with DAPI (blue), (B) Staining for MYBPC3 (green) and DAPI (blue). (C) Categorisation of the CMs produced from the WT (n = 184), and *GATA6*^{c.928_1010delInsA} (n = 177) lines based on the parameters shown in table 3. Statistical test used = Chi-squared, p-value ≤ 0.0001. (D) Cell area measurements for the WT and GATA6 mutant line. WT n = 199 cells, *GATA6*^{c.928_1010delInsA} n = 192 across 4 independent experiments. Statistical test used = Student's t-test. **** = p-value ≤ 0.0001.

6.2.6 Identification of early transcriptional changes during cardiac differentiation of *GATA6*^{c.928_1010delInsA} mutant cells

Given the defects seen in the *GATA6*^{c.928_1010delInsA} CMs bulk RNA-seq was conducted to identify underlying transcriptomic changes that could drive the disorganised hypertrophic phenotype seen at day 32. Day 3 was selected as an early time point corresponding with the establishment of cardiac mesoderm in this model, this time point also occurs ~1 day after the initial peak of *GATA6* expression at the RNA level (Churko et al. 2018). Therefore, the changes recorded at day-3 should identify more direct changes in gene expression. Day 10 was chosen as an end time-point to analyse the effect on the resulting differentiated cell population. A threshold of ± 1 log₂ FC was used to identify differentially expressed genes.

A considerable number of genes were found to be differentially regulated at both time-points, however the effect seemed to be more acute early in differentiation demonstrated by the considerably larger number of genes differentially regulated at day 3 (1,435) in comparison to day 10 (284). Analysis of the genes differentially regulated at day 3 revealed an activation of alternative GRNs associated with neuron guidance (GO terms; 0010975, 0031175). These terms include genes belonging to the semaphorin family such as SEMA3E/F, and SEMA6A. These genes encode signalling molecules that can be membrane bound or secreted, and act as attractive or repulsive cues to guide axon growth, however this activity is not restricted to neurons and these genes can also act as cues for migrating cells and are therefore not neuron specific (Epstein et al. 2015; Sun et al. 2018).

The top terms associated with the down-regulated genes were related to Oxidative phosphorylation (GO terms; 0007005, 0006119) (figure 6.7 b). The genes associated with these terms were investigated further, and closer inspection of these expression changes reveals they are not consistent between repeats (figure 6.7 d). Nonetheless the possibility of a switch in metabolic regulation at this early stage of development was investigated further, by looking at genes involved in glycolysis. A number of genes that contribute to glycolysis were found to be upregulated at day 3 and are shown in figure 6.7 d. These changes were found to be more consistent between the repeats. Importantly, the variation in *GAPDH* expression, an integral component of the glycolysis pathway which has been used as a normalisation control throughout, was low.

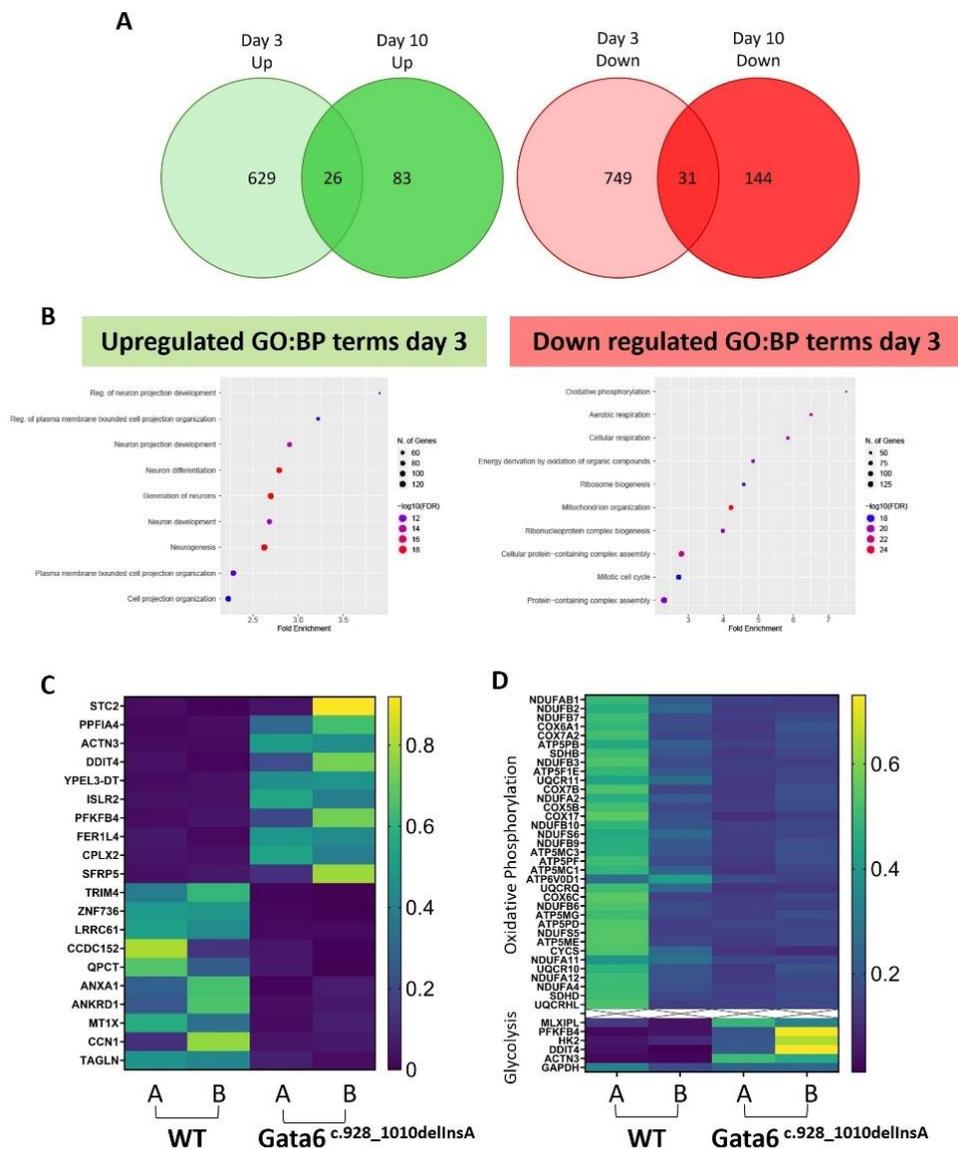


Figure 6.7 Analysis of transcriptional changes in the *GATA6*^{c.928_1010delInsA} cells

(A) Venn diagrams generated using Venny v 2.1.0 showing the overlap between the genes upregulated and down-regulated at day-3 and day-10. A cut of value of ≥ 1 Log₂FC was used to generate the lists of differentially expressed genes. (B) Gene Ontology: Bioprocess (GO:BP) terms associated with the lists of upregulated and downregulated genes generated for day 3 are shown. Dot-plots were created using ShinyGO v 0.77. (C) A heatmap showing the top 10 genes that were identified as upregulated and down regulated at day 3. (D) Analysis of genes associated with oxidative phosphorylation and glycolysis that were found to be differentially expressed on average.

6.2.7 Analysis of gene expression changes in *GATA6*^{c.928_1010delInsA} CMs

At day 10, the number of genes differentially expressed is lower than at day 3, but gene expression is still affected considerably. Upregulated GO Bio-process terms included those related to the fasciculation/bundling of sensory neurons (GO:0097155). The genes associated with this term were EPHA3 and 4, which recent studies demonstrate have important cardiac roles. EPHA3 KO mice display AVSDs and display a reduced number of migrating endocardial cells which is required for endocardial cushion formation and septation (Stephen et al. 2007), whereas KO of EPHA4 has been associated with atrial hypertrophy (Li et al. 2021). Other upregulated terms related to soft palate and skeletal development; these terms actually include several ECM genes. Indeed, if the list of upregulated terms is analysed instead for GO: Cellular component term enrichment then we can see that there is an upregulation in ECM associated genes in the GATA6 line at this time. Furthermore, the top 10 upregulated genes at this time includes POSTN an ECM gene that is upregulated in hypertrophy and heart failure (figure 6.8 c). The down-regulated terms included those related regulation of phospholipid interactions.

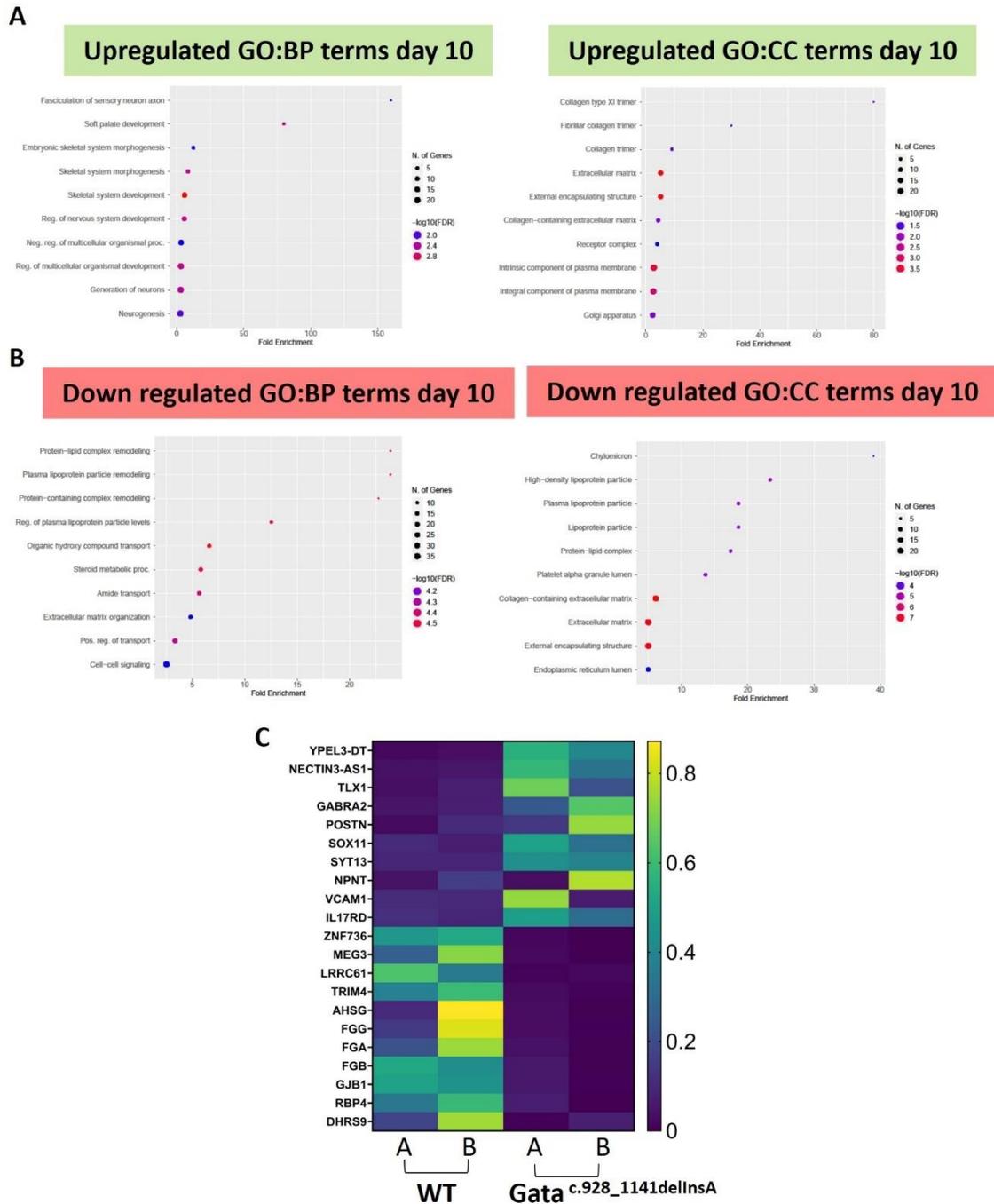


Figure 6.8 Analysis of transcriptional changes in the *GATA6*^{c.928_1010delInsA} cells at day 10

(A) Gene Ontology: Bioprocess (GO: BP) terms associated with the lists of up- and down-regulated genes generated for day 10 samples. (B) Gene Ontology: Cellular Component (GO:CC) terms associated with the lists of up- and down-regulated genes generated for day 10 samples. (C) The top 10 genes that were on average upregulated or downregulated in the *GATA6*^{c.928_1010delInsA} at day 10. The dot-plots shown were created using ShinyGO v 0.77.

6.2.8 Identification of likely direct targets of GATA6 regulation

With the aim of finding direct changes of GATA6 action the differentially regulated genes identified in this experiment were compared with data generated by Sharma *et al.* 2020. On the basis that the *GATA6*^{c.928_1010delInsA} mutant seems to show higher transcriptional activation activity our list of upregulated genes at day 3 was compared with the genes found to be down-regulated in the *GATA6*^{-/-} lines generated by Sharma *et al.* 2020 and with the binding patterns of WT GATA6 at day-4, as determined by ChIP-seq. For the day-10 samples down-regulated terms from the Sharma paper at day 8 and 12 were considered, along with binding pattern of WT GATA6 at day 8. The genes that satisfied these criteria are displayed below, and represent likely targets for GATA6 during cardiac development, that may be of interest for further analysis.

Table 6.3 Possible GATA6 target genes at day 3-4 of iPSC-CM differentiation

The genes listed below were all found to be upregulated at day 3 in the *GATA6*^{c.928_1010delInsA} line. These genes were also found to be differentially regulated at day-4 in the *GATA6*^{-/-} lines generated by Sharma *et al.* 2020. Additionally, GATA6 binding in WT cells was detected in the vicinity of these genes using ChIP-seq by Sharma *et al.* at day-4. These attributes highlight these genes as potential direct targets of GATA6 regulation around this time during iPSC-CM differentiation.

Gene ID	Description	Log2FC	Function and relation to cardiac development	Reference
SFRP5	Secreted frizzled related protein 5	3.3	SFRP5 is expressed at the cardiac crescent stage. SFRP5 + descendants give rise to the sinus venosus, left ventricle, atria and OFT.	(Fujii et al. 2017)
TBX6	T-box transcription factor 6	2.2	Temporary expression can induce mesoderm and cardiac mesoderm, whereas prolonged activity induces skeletal muscle differentiation.	(Sadahiro et al. 2018)
ADCY5	Adenylate Cyclase 5	1.9	Calcium sensitive adenylate cyclase. One of two major forms expressed in the heart.	(Ho et al. 2010)
CFC1	Cryptic family 1	1.7	A co-receptor for nodal signals. KO of this gene results in heterotaxia.	(Gaio et al. 1999)
PDGFRA	Platelet derived growth factor alpha	1.5	A cardiac fibroblast marker, that is required for the migration of epicardial derived fibroblasts during development.	(Smith et al. 2011; Ivey et al. 2019)

AMER3	APC membrane recruitment protein 3	1.4	Acts as a positive modulator of Wnt signalling in colorectal cancer cells.	(Brauburger et al. 2014)
FOXH1	Forkhead Activin Signal Transducer-1	1.3	A Nodal signalling responsive TF, important for mesendoderm induction, and later for the induction of MEF2C in the SHF.	(Watanabe and Whitman 1999; von Both et al. 2004)
LRRC32	Leucine rich repeat containing 32	1.2	Shown to act as membrane tether for TGFβ1 in other cellular contexts.	(Tran et al. 2009)

Table 6.4. Possible GATA6 target genes at day 8-10 of iPSC-CM differentiation

The genes listed below are all upregulated at day 10 in the *GATA6*^{c.928_1010delInsA} cell line. These genes were also found to be differentially regulated at day-8 and 12 in the *GATA6*^{-/-} lines generated by Sharma *et al.* 2020. Additionally, GATA6 binding in WT cells was detected in the vicinity of these genes using ChIP-seq by Sharma *et al.* at day-8. These attributes highlight these genes as potential direct targets of GATA6 regulation around this time during iPSC-CM differentiation.

Gene ID	Description	Log2FC	Function and relation to cardiac development	References
SOX11	SRY-box TF 11	2.3	Required in both mesodermal and neural crest cells for normal OFT formation.	(Paul et al. 2014)
CLIC5	Chloride intracellular channel 5	2.0	Mitochondrial membrane channel. Upregulation is associated with atrial fibrillation.	(Ponnalagu et al. 2016; Jiang et al. 2017)
TBX1	T-box factor 1	1.9	Required for normal OFT morphogenesis.	(Zhang et al. 2006; Vincent et al. 2014)
CORO6	Coronin 6	1.1	Actin binding protein. Reduced expression of CORO6 is seen in failing hearts.	(Hinger et al. 2021)

6.2.9 The expression of iPSC-CM differentiation milestone markers is largely unchanged in the *GATA6* mutant

To compliment the analysis of a limited number of samples by RNA-seq, RT-PCR was carried out on a broader range of time-points. These results are displayed below and show that expression of genes that mark developmental milestones in iPSC-CM differentiation are generally unchanged in their expression pattern and level, with the exception of *NKX2-5* which is upregulated earlier in the mutant line. The expression of *GSC* an early mesendoderm marker, and *SOX17* a marker of definitive endoderm are also upregulated in the mutant. These RT-PCRs have only been carried out in one repeat, and upregulation of these genes at day 3 and/or day 10 was not consistent between the two replicates used for RNA-seq, thus these results should be interpreted with caution until additional repeats can be carried out.

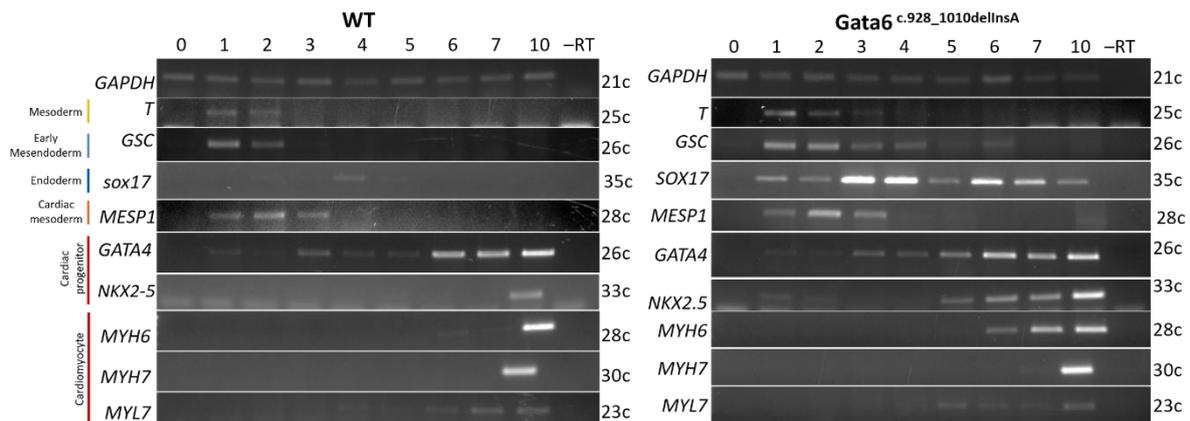


Figure 6.9 Expression of iPSC-CM differentiation markers is largely unchanged in the *GATA6* mutant line

RT-PCR was used to examine the expression of a number of genes during differentiation of the WT and *GATA6*^{c.928_1010delInsA}. *GAPDH* was used as a normalisation control.

6.3 Discussion

GATA6 has previously been characterised as a contributor to the cardiac GRN in development and homeostasis in model organisms, and most recently has been shown to be essential for cardiac differentiation from human iPSC cells. The results presented here provide further insight into the contribution of GATA6 to this process, with the creation of a mutant lacking amino acids 311-337 of transcriptional activation domain 2. The results currently available suggest that this deletion of residues likely leads to increased Gata6 transcriptional activity and through this it has been demonstrated that iPSC-CMs are sensitive to GATA6 activity levels, with this version of the protein leading to the formation of disorganised and hypertrophic CMs. Further steps for analysis of this region and how the mutation of this region may lead to the formation of dysfunctional CMs is explored below.

6.3.1 Deletion of amino acids 311-337 of GATA6 increases its transcriptional activation ability

The deletion described in this project removes 27 amino acids (aa 311-337) from a predicted transcriptional activation domain (aa 280-351) of GATA6 (Morrisey et al. 1997; Sharma et al. 2020). This deletion was found to increase transcriptional activation from a GATA-site driven reporter, which indicates this region within the domain may be required to attenuate GATA6 activity. The region happens to be histidine rich containing a string of 10 uninterrupted histidine residues. Interestingly discrete deletions, duplication, and replacement of these histidine residues has been found in 7 patients with ASD and 1 patient with ASD and PTA (Landrum et al. 2018). The association of these diseases with this cluster of residues in combination with the results presented here, suggests that this region plays an important role in regulating GATA6 activity.

Poly-histidine tracts are more commonly found in TFs (Salichs et al. 2009) yet little information is available regarding their function. The poly-histidine tracts of YY1, HOXA1, POU4F2 have been analysed and were all shown to be necessary for the sequestering of these TFs in nuclear speckles (NS) (Salichs et al. 2009; Wang et al. 2021b). NS have been associated with transcriptional activation and increased splicing efficiency, due to the observation of an accumulation of enhancer regions, active RNA pol II, and splicing factors in these regions (Ilik and Aktaş 2022; Bhate et al. 2023). However, an early study which utilised Bromo-UTP to label transcriptionally active regions of the nucleus, found that active regions do not necessarily associate with NS (Cmarko et al. 1999). More information is required on

the role of NS in transcriptional regulation and how TFs are shuttled between active and inactive domains of transcription. However, it is possible that nuclear localisation of GATA6 is altered in the mutant created here, and that this could subsequently alter GATA6 activity. This may also be applicable to the more discrete mutations observed in patients, and the creation of lines carrying these mutations would be a logical step in understanding how this region regulates GATA6 activity, whilst also providing results that will be more applicable to patients.

6.3.2 iPSC-CM differentiation efficiency is unchanged in the GATA6 mutant, but CM phenotype is affected

As has been previously mentioned higher expression of GATA6 in iPSCs has been associated with higher iPSC-CM differentiation efficiency (Yoon et al. 2018). Thus, as the GATA6 mutant seems to have increased transcriptional activity akin to higher expression of GATA6, we would expect to see an increase in differentiation efficiency. However, no significant increase was noted. There are multiple factors that contribute to differentiation efficiency, and although the protocols used can reproducibly yield highly pure CM cultures (Lian et al. 2012; BurrIDGE et al. 2014b), there is still considerable variation between differentiations. Some of the transcriptional changes observed through RNA-seq analysis at day-3, that will be discussed shortly, suggest this mutation may have a positive effect on CM differentiation. However, more repeats are likely needed to determine if there is an effect on differentiation efficiency to overcome the variability seen.

Conversely an effect on CM phenotype was readily notable. The *GATA6*^{c.928_1010delInsA} mutant CMs displayed increased cell size and increased TNNT2 staining at day 32, which are indicators of cardiac hypertrophy. This correlates with previous studies that have demonstrated GATA6 is required for hypertrophy in response to pressure induced overload in mice, and that over-expression of GATA6 alone is sufficient to stimulate hypertrophy in aged mice (Liang et al. 2001; Van Berlo et al. 2010). A similar level of disorganisation and hypertrophy was seen in the *TBX2*^{c.700_761del} line when analysed at day 32. In this line an increase in cell size was detectable at day-12 as was an increase in the expression of CM stress such as *NPPA* and *NPPB* (Nakamura and Sadoshima 2018). The RNA-seq conducted for the GATA6 mutant line at day-10 did not detect an increase in these markers, nor was there a significant change in the expression of sarcomeric genes that may be expected when hypertrophy is present. This may mean that hypertrophy is happening at a later stage in the

GATA6 mutant. Examination of the GATA6 mutants at earlier time-points by IF is required to clarify when hypertrophy begins in these cells and may help in understanding what the potential triggers are for this change.

6.3.3 Exploration of early transcriptional changes in GATA6 mutants

Gata6 mRNA expression can be reliably detected during iPSC-CM differentiation from around day-2, aligning with the expression of mesendoderm markers such as *TBXT*, *EOMES*, and *GSC* (Churko et al. 2018). Placing GATA6 at a timepoint where it has the potential to affect the earliest stages of cardiac development. The results generated by Sharma *et al.* 2020, suggest knockout of *GATA6* in iPSCs does not negatively impact early mesendoderm formation, as demonstrated by similar or higher levels of the expression of *TBXT*, *EOMES*, and *GSC* in these lines. The results reported here somewhat agree with this, expression of *TBXT* was comparable to WT suggesting mesoderm formation proceeds normally, but prolonged expression of early mesendoderm marker *GSC*, and increased expression of *SOX17* suggest that the propensity for definitive endoderm formation is increased in this line. These changes will need to be verified in further repeats but are in agreement with previous observations that GATA6 is required for and can promote the formation of definitive endoderm (Cai et al. 2008; Guye et al. 2016; Chia et al. 2019; Heslop et al. 2021). Despite these changes expression of cardiac mesoderm marker *MESP1* in the *GATA6^{c.928_1010delInsA}* line was undisrupted.

In addition to the above, RNA-seq analysis flagged several other differentially expressed genes that may be of interest in these early stages of differentiation. A paper published by Sadahiro *et al.* 2018 demonstrates that *TBX6* can be a positive inducer of cardiac development. In this study *TBX6* was overexpressed in mESCs and was able to induce mesoderm and subsequently cardiac mesoderm differentiation, when applied from days 0-3 of differentiation. Extension of *TBX6* expression beyond this time induced skeletal muscle development (Sadahiro et al. 2018). They also observed higher expression of *PDGFRA* when *TBX6* was induced during this cardiogenic window, correspondingly an upregulation of *PDGFRA* was observed in the RNA-seq results generated herein. At day-10 GO: BP analysis revealed an upregulation of skeletal muscle terms however closer inspection revealed the genes associated with these terms did not include any definitive markers of skeletal muscle. Rather they contained genes related to ECM deposition, as well as TFs *SOX11* and *TBX1*

which in the context of a cardiac differentiation may represent the activation of gene programs related to OFT specification (Paul et al., 2014; Xu et al., 2004; Zhang et al., 2006). This may be a genuine change as OFT formation is a process GATA6 is known to be active in through its induction of SEMA3C in the OFT and underlying myocardium in mice (Kodo et al. 2009). If and how increased GATA6 activity would affect this process is unclear. *SFRP5* was another gene found to be upregulated at day three. This gene is expressed at the cardiac crescent stage in mice and demarcates most of the cells that will give rise to the heart, the contribution of these cells to the left ventricle, interventricular septum, and the OFT is particularly strong (Fujii et al. 2017). Taken together these gene expression changes would suggest a net positive effect on cardiac development at day three.

6.3.4 Indications of dysfunctional metabolism in the GATA6 mutant cells

At day 10, there was an upregulation of ECM genes; *COL9A2*, *COL11A1*, *COL13A1*, and *POSTN*. Increased ECM deposition is associated with pathological remodelling in cardiac hypertrophy (Nakamura and Sadoshima 2018; Vigil-Garcia et al. 2021), *POSTN* in particular has been associated with HF (Stansfield et al. 2009; Dixon et al. 2019). The down-regulated terms at this time were also revealing. A number of pathways related to lipoprotein transport and metabolism were flagged. This may be a specific related to the location of the GATA6 mutation (deletion of aa 311-337). Residues 227-331 of GATA6 have been shown to interact with PPARA in myoblast cells, where this interaction induces SLC2A4 (previously known as GLUT-4) expression increasing glycolysis (Yao et al. 2012). However, PPARA is also a modulator of the expression of components of ketogenic, lipogenic, lipid and fatty acid transport, cholesterol, and fatty acid metabolic pathways (Han et al. 2017). It is not clear whether interaction of GATA6 with PPARA is required for modulation of metabolism in CMs. However, metabolic changes are a known catalyst for cardiomyopathy (Chong et al. 2017; Towbin and Jefferies 2017; Lopaschuk et al. 2021), so this should be followed up in future experiments with comparison of glycolysis and oxidative phosphorylation rates in these cells (Readnower et al. 2012). Furthermore, the switch from glycolysis to lipid metabolism in CMs is associated with CM maturation *in vivo* and in iPSC-CMs and it is possible disruption of the interaction between GATA6 and PPARA could affect maturation (Makinde et al. 1998b; Correia et al. 2018; Horikoshi et al. 2019; Feyen et al. 2020). Therefore, it would also be interesting to see if this process is also disrupted in the cells. In summary, abnormal

metabolic handling in the *GATA6*^{c.928_1010delInsA} cells is a possible mechanism through which dysfunction and hypertrophy could occur.

6.3.5 Conclusions

To conclude, the results presented here provide insights into the roles of amino acids 311-337 of transcriptional activation domain 2 of GATA6 in iPSC-CMs. The RNA-seq data presented highlights changes at early stages that may positively contribute to cardiac differentiation, and later metabolic changes that may negatively affect CM phenotype. Further work is needed to determine how this mutation affects GATA6 localisation, and to determine if this region is an important mediator of protein-protein interactions that may regulate activity of GATA6.

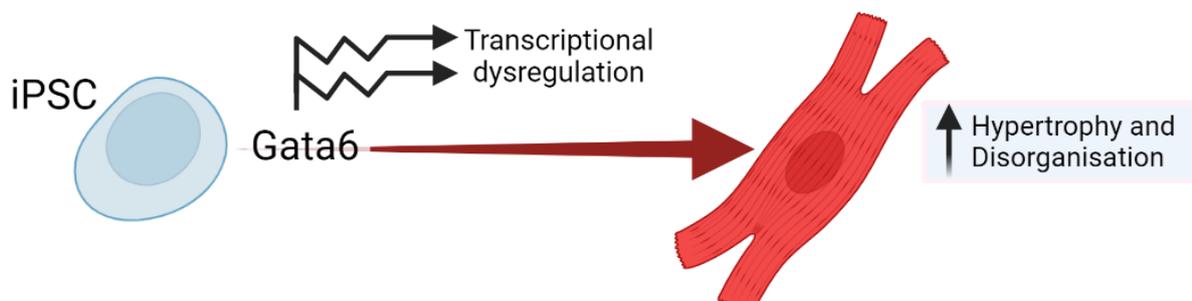


Figure 6.10 Residues 311-337 of cardiac TF GATA6 are required for normal CM development.

Residues 311-337 belonging to transcriptional activation domain 2 of GATA6 seem to be required for normal iPSC-CM differentiation. Initial results suggest loss of these residues results in increased GATA6 transcriptional activity, and this was associated with the dysregulation of a number of genes. These changes were associated with the formation of hypertrophic disorganised CMs. Indicating that iPSC-CM differentiation is sensitive to GATA6 activity levels, and that this section of the protein may be responsible for maintaining normal levels of GATA6 activity.

Chapter 7 – Overview of Findings and Future Directions

Successful cardiac development relies on the proper collaboration of numerous TFs in a GRN. Due to the intricate nature of these interactions this process often and easily goes awry, resulting in the development of CHDs and CVDs. Despite significant research into this subject, it can be challenging to identify an exact cause in some cases. This is likely attributable to the gaps that remain in our understanding of what factors feed into the cardiac GRN and the mechanisms through which they act. The results presented in this thesis provide insights into the relative contributions of TFs GATA4, TBX2, PRDM1, and GATA6, to the cardiac GRN, using an iPSC-CM model of cardiac differentiation. The highlights of these findings will be outlined in this chapter along with the limitations of this study. The steps needed to address these limitations and avenues for future research are also discussed.

7.1 The general limitations of iPSC-CM models

7.1.1 Inter- and intra-line variability in iPSC-CM differentiation efficiency

There are certain limitations of the model that are applicable to all lines created. It has been known for some time that iPSC lines exhibit biased differentiation potential, inevitably this will affect the results of any experiments conducted (Bock et al. 2011; Ohno et al. 2013; Bouma et al. 2017). Some of the factors that contribute to this phenomenon include genetic background variation, donor tissue site, and the re-programming method used. An extensive study of 711 human iPSC lines from the hiPSCI bank estimated that genetic variation between donors may account for as much as 46% of the variation seen in the expression level of lineage defining genes (Kilpinen et al. 2017), making inter-line genetic variation a significant source of variability. It has been recommended based on these reasons that any experiments using iPSCs should be conducted in 4-6 independent parental lines (Germain and Testa 2017; Volpato and Webber 2020). However, the time and cost required to generate multiple lines can be prohibitive. In a number of reports just one parental cell line is utilised as is the case here (Sharma et al. 2020; Gonzalez-Teran et al. 2022).

7.1.2 How does iPSC-CM maturity affect the phenotypes observed?

In addition to inter-line variability another firmly established limitation of iPSC-CM differentiation as a model of cardiomyogenesis is the immaturity of the CMs formed. This limits the phenotypic characterisation of all lines created in this work, as it is possible that more mature CMs will exhibit further phenotypic differences that may be more

physiologically relevant. Alternatively, some of the mutants created may fail to mature properly which would be informative. Therefore, a line of future work will be directed towards adopting different models of maturation (the factors associated with maturation *in vivo* and methods for improving CM maturation are reviewed in section 1.7.2). Some of these approaches have already been trialled and are presented in figure S2-3.

7.1.3 Technical and time-frame limitations

Extensive effort and time was required to produce the GATA4^{-/-} lines presented in Chapter 3. Subsequently, this has reduced the amount of time available to analyse these lines and some elements could not be completed within the time frame of the project. For example, the analysis of *PRDM1* and *TBX2* expression in the GATA4^{-/-} 1 line was limited to days 4 and 10 respectively. For completeness it would be ideal to analyse their expression throughout differentiation with a selection of other time points, and to extend this analysis to the GATA4^{-/-}2 line. These patterns should then be confirmed at the protein level using WB.

Furthermore, RT-PCR has been used throughout as a method for identifying and confirming trends in gene expression levels between samples through qualitative assessment of band intensity. These results have been sufficient to draw conclusions and meet the aims laid out herein. However, this could be improved through the utilisation of a quantitative method for gene expression analysis such as quantitative RT-PCR or more extensive RNA-seq sampling. These methods allow for exact quantification of the transcripts present and have a higher dynamic range, thus can detect lower abundance targets (Klein 2002; Wang *et al.* 2009). Therefore, this would allow for more precise conclusions to be drawn.

7.2 GATA4 is essential for human iPSC-CM differentiation

Extensive work using model organisms has established that GATA4 is central to the cardiac GRN, controlling many aspects of embryonic cardiac development and homeostasis. An examination of the requirement for GATA4 in human iPSC-CM differentiation was delivered in Chapter 3 with the creation of 2 GATA4 null iPSC lines that reproducibly failed to form beating CMs. The phenotype reported here is more severe than that described for mice, and another GATA4 null iPSC line (Watt *et al.* 2004; Gonzalez-Teran *et al.* 2022). However, the thorough examination of these lines provided in Chapter 3, demonstrating

reduced differentiation efficiency, defective sarcomere formation, and abnormal calcium handling in the GATA4^{-/-} cells supports the conclusion that functional CM formation requires GATA4. Re-affirming the conclusion that GATA4 is an essential node within the human cardiac GRN.

7.2.1 Reproducibility of the phenotype in different iPSCs

Based on the difference between the results published here and those published by Gonzalez-Teran *et al.* the question of how inter-line variability affects phenotypic presentation is of particular importance for GATA4. Although cardiac development was disrupted in the line published by Gonzalez-Teran *et al.* they found that beating CMs could be produced where the lines created here failed to do so. Therefore, the creation of GATA4 null mutations in additional iPSC cell lines is essential to resolve the cause of these differences. Another aspect in which the methods used to derive the results detailed in Chapter 3 diverge from Gonzalez-Teran *et al.* 2022 is in the differentiation method used. Currently, the results reported in Chapter 3 have been generated using the CDM3 protocol, whereas Gonzalez-Teran *et al.* have used the GiWi protocol published by Lian *et al.* 2012. Both differentiation methods rely on bi-phasic Wnt modulation for CM induction, the primary difference between them is in the basal media used. The GiWi protocol uses RPMI 1640 + B27, a supplement containing an array of compounds that support cell culture, whereas the CDM3 protocol utilises RPMI 1640 with just two components of the B27 supplement: ascorbic acid-2-phosphate and bovine serum albumin. This limited list of components has been shown to be sufficient to support CM differentiation in >200 iPSC lines (Burrige *et al.* 2015). As both protocols generate CMs at a similar efficiency, and target modulation of the same pathways this difference is not expected to lead to significantly different results. Nonetheless this should be addressed by direct comparison of differentiation of the lines via both protocols in case there is an effect on the quality of the cells produced. This is something that is actively being addressed.

7.2.2 When does cardiomyogenesis fail in the GATA4 null lines & can it be rescued?

GATA4 is expressed throughout cardiac differentiation and in adult CMs; where it regulates the transcription of cardiac progenitor and CM specific genes and is a known mediator of CM hypertrophy and survival in adult CMs. Thus, it is a TF that fulfils many roles in CM development and homeostasis. Examining at what point CM differentiation goes wrong in the GATA4 null lines reported here has begun with RT-PCR analysis, determining

that differences between the WT and GATA4 null lines start to emerge at the cardiac progenitor stage. The number of genes analysed was limited and should be extended further with the use of RNA-seq to comprehensively analyse the transcriptomic changes occurring throughout differentiation. Sequencing of RNA samples covering days 3-10 of differentiation would capture the establishment of the cardiac progenitor pool and their differentiation into CMs. Alternatively single cell RNA-seq could be used to achieved this with increased resolution. Numerous scRNA-seq papers have recently emerged showcasing the utility of this technique to study the trajectory of single cells during cardiac development *in vivo* and in iPSC-CM models (Churko *et al.* 2018; Grancharova *et al.* 2021; Tyser *et al.* 2021; Zhang *et al.* 2021; Galdos *et al.* 2022). Furthermore, although CMs can be derived from iPSCs with high efficiency other cell types will be present, such as stromal cells, SMCs, endothelial cells, and epicardial like cells (Churko *et al.* 2018; Grancharova *et al.* 2021). GATA4 is expressed in a number of these cell types, where it also regulates their development and homeostasis (Watt *et al.* 2004; Dittrich *et al.* 2021). These non-CM cell types also impact cardiac development; therefore, it would be of interest to study these in the GATA4 null lines presented herein alongside the CM population. Single cell RNA-seq would facilitate the monitoring of these different trajectories, with the additional benefit of potentially identifying any novel trajectories or cell types that arise, this is addressed further in section 7.2.4

To study the requirement for GATA4 past this time will require the creation of additional tools. The insertion of an inducible GATA4 allele into the null lines would be a suitable way to achieve this, enabling the examination of GATA4 at different stages of development and in adult CMs. This also provides an important control. If GATA4 is essential for iPSC-CM differentiation, then re-introducing its expression should rescue the GATA4 null phenotype characterised here. Previous attempts to create inducible lines are detailed in figure S1. The approach trialled used the Tet-on-system, the Tet-activator and target genes under control of the Tet response element were delivered using separate plasmids and were targeted to the AAVS1 locus (Sim *et al.* 2014). Whilst the plasmid containing the target gene was successfully integrated the construct carrying the Tet-activator was not. It is believed this may be due to a reduced selection pressure for integration of the Tet-activator construct based on the antibiotic resistance gene used. The strategy for future attempts will

be to place the Tet-activator and gene of interest in the same construct, with a strong selection cassette and reporter allele for robust separation of positive clones, as has been done in (Akhtar et al. 2018).

7.2.3 Is a failure in cardiac progenitor proliferation responsible for the phenotype observed in the GATA4 null lines?

Something that has not been addressed in this project thus far is how proliferation is affected in the GATA4 null lines. Proliferation in cardiac progenitors is in part controlled by GATA4 (Zeisberg et al. 2005; Rojas et al. 2008; Singh et al. 2010; Misra et al. 2014; Yamak et al. 2014). The reduced efficiency of CM formation observed could well be due to reduced proliferation of cardiac progenitors, therefore this should be investigated. Methods for analysing proliferation are well established and in combination with staining for stage specific markers it should be possible to resolve how proliferation is affected at each stage of development and this may go some way to explaining the phenotypic differences observed.

7.2.4 If not CMs, what do GATA4 null cells become?

Preliminary IF analysis revealed many of the GATA4^{-/-} cells present were positive for VIM and ACTA2, markers of mesenchymal and SMCs respectively. This evidence suggests the absence of GATA4 expression has resulted in a cell fate switch. The preliminary IF analysis conducted was limited, having been performed on one differentiation with a small number of cell markers, and thus provides only a partial insight into the phenotype. Bulk RNA-seq or scRNA-seq would be suitable methods for examining this comprehensively. scRNA-seq is preferable as it is clear there is heterogeneity within the GATA4 population between; cells that express TNNT2 and reach a CM-like fate, those that express TNNT2 and exhibit abnormal morphologies reminiscent of activated myofibroblasts, and those that are TNNT2 negative. These different cell types likely represent a spectrum of dysfunctional transcriptional regulation that may be masked by bulk RNA-seq thus making scRNA-seq the preferred method of analysis. Once candidate cell types and reliable markers have been identified these changes can easily be confirmed by IF.

Another possibility is that the reduction in differentiation efficiency observed is due to decreased proliferation or apoptosis of CM precursors. This is not something that has been addressed in this project thus far, but GATA4 does have a documented role in cardiac progenitor proliferation, and as a regulator of pro-survival mechanisms in established CMs

(Kobayashi et al., 2006; Misra et al., 2014; Rojas et al., 2008; Singh et al., 2010; Yamak et al., 2014; Zeisberg et al., 2005). Methods for analysing proliferation and apoptosis are well established and in combination with staining for stage specific markers it should be possible to resolve how these processes are affected at each stage of development and this may go some way to explaining the phenotypic differences observed.

7.3 PRDM1 and TBX2 are regulated by GATA4 in iPSC-CMs

In *X. laevis* overexpression of GATA4 is sufficient to drive cardiomyogenesis in pluripotent ectodermal explants (Latinkic 2003). Prior to the onset of the work presented herein, the *Xenopus* model of cardiac development referenced above was used to identify a set of genes that are regulated by GATA4 and thus may comprise additional nodes within the cardiac GRN. The GATA4 null iPSC lines have been utilised here to indicate if two of these target genes, TBX2 and PRDM1, are targets of GATA4 in human cardiomyogenesis. Both genes were found to be dysregulated in the null lines suggesting this regulatory relationship is conserved between *X. laevis* and human CM differentiation in vitro. The GATA4 inducible lines produced to meet the questions raise in section 7.2.2 can be used to further address this question, with up-regulation of TBX2 expected when GATA4 is induced and repression of PRDM1.

7.4 TBX2 is required for the formation of normal CMs, likely by fine tuning the cardiac differentiation program

Chapters 4 and 5 investigate whether GATA4 target genes (*TBX2* and *PRDM1*) themselves are functional nodes within the cardiac GRN. *TBX2* was found to be a positive target of GATA4 regulation in *Xenopus* and to be downregulated during in the GATA4 null lines described in Chapter 3. Through the creation of two *TBX2* loss of function lines carrying DBD disrupting mutations, it was demonstrated that *TBX2* is required for the formation of normal CMs. The CMs formed from the *TBX2* mutants displayed various features of CM pathology, such as hypertrophy, myofibril disorganisation, altered calcium handling properties, and increased expression of genes associated with cardiac distress and disease. Analysis of early transcriptional changes in these cells at day 7, prior to the onset of beating for these lines, revealed that several genes related to cardiac chamber development, morphogenesis, and muscle development were upregulated. *TBX2* has previously been identified as a repressor of chamber gene expression during AVC formation, a later step in cardiac development (Christoffels *et al.* 2000; Harrelson *et al.* 2004; Ribeiro *et al.* 2007). The

results presented suggest that TBX2 has an important role in dampening the expression of these pathways more broadly in early CMs. It is uncertain at this time if this relates to all cells within the population or a subset fated to give rise to non-chamber CM sub-types. However, the protocols used primarily forms a mixture of chamber-like CMs, therefore the working hypothesis is that TBX2 attenuates the expression of these genes in chamber like cells before its later role in the AVC. Further investigation is required to determine how excessive activation of these genes results in a pathological state.

7.4.1 – The creation of novel TBX2 lines requires modification of the CRISPR targeting strategy used

As discussed in section 7.1.1 the creation of mutants in different parental lines is a necessary next step to solidify the reproducibility of these results. Before embarking on producing further lines the CRISPR targeting strategy used should be re-considered. The CRISPR guides used herein were targeted to exon 3, which resulted in variable degrees of exon skipping between the mutants that consequently bring the coding region back into frame. Surprisingly exon skipping is not an uncommon outcome after CRISPR-Cas9 gene editing as demonstrated in a screen by (Tuladhar et al. 2019). From this study it is suggested that asymmetric exons may be targeted to overcome this, as even if skipped they will still leave the coding sequence out of frame, and therefore this strategy is more likely to result in nonsense-mediated decay. Any future lines will be created with this in mind. This strategy could also be complimented with RNAi based knockdown approaches, this would allow further validation of the results and could be used to examine the short-term effects of TBX2 silencing at different stages of iPSC-CM differentiation.

7.4.2 What are the drivers of pathology in the TBX2 lines?

Possible drivers for the pathological phenotype observed in the TBX2 mutants were explored using IF, calcium imaging, and RNA-seq. Analysis by IF showed that TBX2 mutant cells display hypertrophy from an early stage (day 12), which worsens as the cells develop (D32), with abnormal calcium handling noted in the D32 cells. RNA-seq analysis at day 10 revealed dysregulation of some calcium genes at this stage, and this was identified as possible driver of pathology. To test this calcium imaging should be conducted at multiple stages throughout differentiation to capture any early changes. Furthermore, it is unclear if TBX2 directly regulates these genes due to the absence of iPSC-CM based ChIP-seq data sets for TBX2 binding. Therefore, this is another avenue for further research.

7.4.3 How is proliferation affected in the TBX2 lines?

An element that was not assessed during this study was the effect of TBX2 mutation on proliferation of the cells. The observation of lower proliferation in the AVC coinciding with TBX2 expression and that TBX2 can repress expression of *Mycn*, a pro-proliferative gene, has led to the characterisation of TBX2 as a repressor of proliferation (Christoffels et al., 2004; Harrelson et al., 2004; Ribeiro et al., 2007; Singh et al., 2012). However, as outlined in section 4.1.2 there are conflicting reports on the effect of TBX2 loss or gain of function on proliferation in the heart. Furthermore, in various malignancies TBX2 has been associated with a pro-proliferative anti-senescence phenotype (Vance et al. 2005; Zhu et al. 2014; Crawford et al. 2019). Therefore, the activity of TBX2 in proliferation is highly context dependent and requires independent assessment in this iPSC-CM differentiation model.

7.5 PRDM1 is not essential for iPSC-CM differentiation

PRDM1 was identified as a negative target of GATA4 transcriptional regulation in *X. laevis* and was found to be upregulated in the GATA4 null iPSC lines created herein. This suggests that regulation PRDM1 may be required for iPSC-CM differentiation. This question has been answered through the creation of PRDM1 null lines, one lacking all PRDM1 isoforms (PRDM1^{-/-}) and the other lacking expression of the short form of the protein (PRDM1^{c.514_515InsA}). Both lines were found to exhibit accelerated differentiation onset, thus apparently supporting the view that GATA4 mediated repression of PRDM1 may promote cardiac differentiation. From preliminary RNA-seq analysis of the PRDM1^{-/-} line it was noted that there was a de-repression of various neural gene programs at early stages of differentiation, when PRDM1 expression and presumably its activity would usually be at its highest. Thus, it is proposed from these findings that PRDM1 may have an early role in repressing the expression of alternative differentiation programs, and then subsequently in modulating the instigation of the cardiac gene program. This early de-regulation did not seem to have a negative impact on the CMs formed, which appeared to resemble WT cells in most aspects.

7.5.1 How does knock-out of PRDM1 affect inter-line differentiation efficiency variability?

The acceleration of beating onset in both PRDM1 lines created in Chapter 5 provides an interesting avenue for further research. As already expressed, the lines created here all stem from the same parental line. This line reliably produces good CM yields, and knockout

of PRDM1 seemed to accelerate this process. Given PRDM1's characterisation as a transcriptional repressor, and the earlier activation of CM specific genes in these lines it follows that PRDM1 may act as an inhibitor of this process. Thus, it would be interesting to see if knockout or knockdown of PRDM1 could have this effect in other lines, perhaps improving cardiogenic differentiation potential in lower efficiency lines. This role may also be applicable to the production of induced CMs (iCMs). Trans-differentiation of alternative cell types into iCMs has been touted as a possible method for intrinsic cardiac repair, but it is generally an inefficient process, and it is difficult to achieve complete re-programming (Lyra-Leite et al. 2022). It is possible that PRDM1 could act as a barrier to re-programming, therefore its removal may improve efficiency. Furthermore, given the widespread expression of PRDM1, and the upregulation of neural genes observed here these points could be applicable to other differentiation pathways, the iPSC lines established here provide a useful tool for studying this.

7.5.2 Is repression of PRDM1 required for differentiation?

If the hypothesis given above is true and PRDM1 acts as a barrier to differentiation, then over-expression of this gene should delay differentiation. This would be best addressed through the creation of cell lines carrying an inducible form of PRDM1 using the methods reference in section 7.2.2. These lines could be used to ask if PRDM1 expression is maintained beyond its usual time window during iPSC-CM differentiation, does that lead to delayed differentiation rather than the accelerated differentiation seen in the knockout lines. This would provide more solid support for the hypothesis that GATA4 mediated repression of PRDM1 is necessary for differentiation, which was an original aim of this thesis.

7.6 Increased GATA6 transcriptional activity results in the production of hypertrophic iPSC- CMs

In Chapter 6 the actions of another established cardiogenic TF, GATA6, were examined during human iPSC-CM differentiation. A study by Sharma *et al.* 2020 demonstrated GATA6 is absolutely required for the formation of CMs from iPSCs. The results presented here add to this by showing that iPSC-CM differentiation is sensitive to the level of GATA6 activity. The mutant line ($GATA6^{c.928_1010delInsA}$) created carries a deletion of amino acid residues 311-337. This portion of the protein seems to be required to attenuate GATA6 activity as demonstrated by an increase in GATA6 transcriptional activity. The

GATA6^{c.928_1010delInsA} mutation resulted in the production of pathological CMs that displayed hypertrophy and myofibril disorganisation. Mutations in this region of the protein are associated with the development of CHDs (see table 6.1 (Landrum et al. 2018)), furthermore excessive *GATA6* activity has been shown to promote hypertrophy in established CMs (Liang et al. 2001; Van Berlo et al. 2010). Thus, based on the data presented and these observations it seems that this region of the protein has a crucial role in maintaining normal levels of *GATA6* activity and that increased *GATA6* activity may stimulate disease progression. RNA-seq analysis identified dysregulation of several pathways in the *GATA6* mutant line related to metabolic pathways, and later lipo-protein handling and ECM production. These changes may represent mechanisms through which the *GATA6*^{c.928_1010delInsA} mutation can cause pathology and will require further investigation.

7.6.1 How does deletion of residues 311-337 affect *GATA6* protein stability and localisation?

Although some work was done to determine how the mutation affects protein activity, further repeats and controls are needed. Firstly, further repeats are required to determine *GATA6*^{c.928_1010delInsA} stability levels in hCMs, and in the *Xenopus* transcriptional activity assay (section 6.2.3). This analysis could then be extended to include examination of *GATA6* localisation within the cell as a mechanism through which the mutation affects *GATA6* activity. The deletion region is histidine rich. Poly-histidine tracts have previously been associated with the localisation of other TFs to nuclear speckles, which may regulate TF activity levels by sequestering or concentrating them in certain areas of the nucleus. Attempts were made to image *GATA6* localisation in the nucleus, with no evidence of speckles (see supplementary data, figure S5). However, this may be due to the steps taken to process the samples and the resolution of these images which can be further improved. Further attempts should ideally be combined with staining for a speckle positive marker such as SRSF2 (SC35), to determine if there is overlap between these factors (Salichs et al. 2009). Furthermore, Bromo-UTP (Br-UTP) could be used to identify active regions of transcription as was carried out by Cmarko *et al.* 1999. This would require cellular injection so this would be better achieved by co-injection of *GATA6*^{c.928_1010delInsA} mRNA and Bromo-UTP into *Xenopus laevis* embryos that are more amenable to injection, or alternative RNA analogues could be used in the *GATA6*^{c.928_1010delInsA} CMs such as 5-ethynyluridine (Jao and

Salic 2008). Then the effect on nuclear localisation and transcriptional activity in these regions may be properly determined.

In addition to the above, the mutation created here is relatively large in comparison to those associated with human CHD cases, which involve discrete deletions, duplications, or replacements of 1-2 amino acid residues within the deletion region. Therefore, the results presented here likely represent an extreme version of events. Thus, it may be more useful to understand how these smaller mutations effect protein localisation and activity, to gain results that are more directly translatable to patients.

7.6.2 Sensitivity to hypertrophy in the GATA6 mutant line

Given the propensity for the *GATA6*^{c.928_1010delInsA} line to produce hypertrophic CMs it would be of interest to see how sensitive this mutant, and any patient mutation lines are to hypertrophic stimuli. This could be easily achieved using established pharmacological treatments known to induce hypertrophy such as, phenylephrine. Treatment with phenylephrine also stimulates CMs to switch from fatty acid metabolism to glycolysis. Based on the potential for the mutation to disrupt the interaction between GATA6 and PPARA it is proposed the mutants may already have a deficiency in fatty acid handling that could predispose them to hypertrophy. There are established platforms and protocols available to test the ability of these cells to utilise fatty-acid metabolism or glycolytic pathways (Readnower et al. 2012).

Furthermore, restoration of fatty acid metabolism through the inhibition of malonyl-CoA production, which can inhibit fatty acid entry into mitochondria, has been shown to attenuate pressure overload induced hypertrophy in mice by maintaining fatty acid oxidation levels (Kolwicz et al. 2012; Ritterhoff et al. 2020). If interaction between GATA6 and PPARA is important for CM metabolism and hypertrophy, then interventions like this may be less effective in attenuating hypertrophy in *GATA6*^{c.928_1010delInsA} CMs. Exploring these avenues of research may provide further insights into how mutations in GATA6 may act as a catalyst for cardiac disease.

7.6.3 Examining the activity of GATA6 in other developmental contexts

Looking beyond heart development, GATA6 has well established roles in the development of endoderm and its derivatives (Fisher et al. 2017; Shi et al. 2017; Sharma et al. 2020). Increased expression of early endoderm markers such as SOX17 and GSC in the

GATA6 mutant created here indicates that endodermal development may also be affected in the line. However, it should be noted that this result needs to be confirmed in further repeats, the addition of further endodermal markers would also increase confidence in these results.

7.7 Concluding remarks

GATA4 is a TF that is broadly expressed throughout cardiac development and in the adult heart. Roles for GATA4 throughout the process of cardiac development have been demonstrated using multiple model organisms. The result presented in this thesis have added to this by assessing the requirement for GATA4 in a human model of cardiogenesis, demonstrating that GATA4 is essential for CM formation in this model. Furthermore, this thesis has provided evidence that *TBX2* and *PRDM1* are conserved targets of GATA4 in this model and demonstrated they both have functions in iPSC-CM development. *TBX2* was shown to be required for the formation of WT-like CMs, revealing a previously unappreciated role for *TBX2* in modulation of the early cardiac GRN. In contrast, *PRDM1* does not seem to be required for the formation of WT-like CMs but does seem to have a modulatory role early in development in the repression of alternative gene programs, and later in modulating the onset of cardiac gene expression. In the final results chapter, the role of GATA4 relative GATA6 in iPSC-CM differentiation was investigated. These investigations revealed amino acid residues 311-337 attenuate GATA6 function and are required for the production of normal CMs. Together the findings presented in these chapters expand upon what is known about the cardiac GRN and establish a robust basis for further exploration of GATA4, *TBX2*, *PRDM1*, and GATA6 in iPSC-CM differentiation.

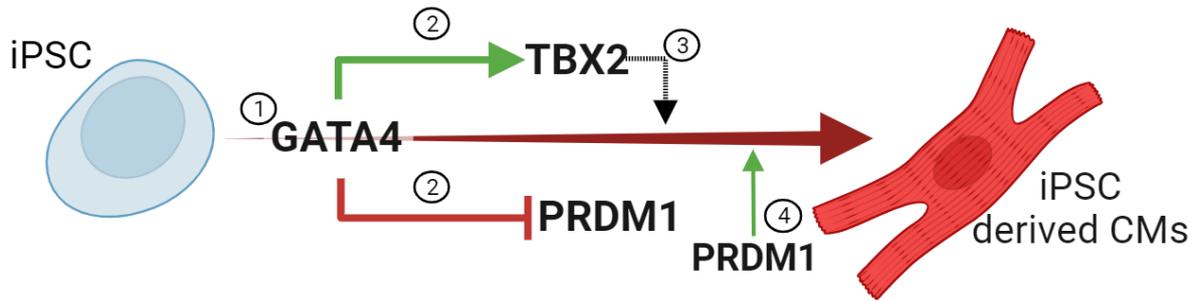


Figure 7.1 A summary of the conclusions produced during this thesis

(1) GATA4 is required for the formation of functional CMs from iPSCs. (2) GATA4 likely regulates the expression of *TBX2* and *PRDM1* in this model (3) Loss of WT *TBX2* expression leads to the formation of hypertrophic CMs that display signs of pathology, thus WT *TBX2* is needed for the production of normal CMs. (4) Loss of *PRDM1* expression results in a small but significant acceleration in iPSC-CM differentiation. *PRDM1* may contribute to cardiac development by repressing alternative gene programs, and by modulating the onset of cardiac gene expression. The final conclusion, which is not depicted is that iPSC-CM differentiation is sensitive to GATA6 activity levels, and that amino acids 311-337 are important for regulating GATA6 activity.

References

- Aanhaanen, W.T.J. et al. 2009. The Tbx2+ primary myocardium of the atrioventricular canal forms the atrioventricular node and the base of the left ventricle. *Circulation Research* 104(11). doi: 10.1161/CIRCRESAHA.108.192450.
- Aanhaanen, W.T.J. et al. 2011. Defective Tbx2-dependent patterning of the atrioventricular canal myocardium causes accessory pathway formation in mice. *Journal of Clinical Investigation* 121(2), pp. 534–544. doi: 10.1172/JCI44350.
- Abrahams, A., Mowla, S., Parker, M.I., Goding, C.R. and Prince, S. 2008. UV-mediated regulation of the anti-senescence factor Tbx2. *Journal of Biological Chemistry* 283(4). doi: 10.1074/jbc.M705651200.
- Abu-Issa, R., Waldo, K. and Kirby, M.L. (2004). Heart fields: One, two or more? *Developmental Biology* 272(2).
- Adams, E., McCloy, R., Jordan, A., Falconer, K. and Dykes, I.M. 2021. Direct reprogramming of cardiac fibroblasts to repair the injured heart. *Journal of Cardiovascular Development and Disease* 8(7). doi: 10.3390/jcdd8070072.
- Ahmad, F.B. and Anderson, R.N. 2021. The Leading Causes of Death in the US for 2020. *JAMA - Journal of the American Medical Association* 325(18). doi: 10.1001/jama.2021.5469.
- Ahmed, M.I., Elias, S., Mould, A.W., Bikoff, E.K. and Robertson, E.J. 2016. The transcriptional repressor Blimp1 is expressed in rare luminal progenitors and is essential for mammary gland development. *Development (Cambridge)* 143(10). doi: 10.1242/dev.136358.
- Aikawa, M. et al. 1993. Human smooth muscle myosin heavy chain isoforms as molecular markers for vascular development and atherosclerosis. *Circulation Research* 73(6). doi: 10.1161/01.RES.73.6.1000.
- Akerberg, B.N. et al. 2019. A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. *Nature Communications* 10(1). doi: 10.1038/s41467-019-12812-3.
- Akhtar, A.A. et al. 2018. Inducible Expression of GDNF in Transplanted iPSC-Derived Neural Progenitor Cells. *Stem Cell Reports* 10(6). doi: 10.1016/j.stemcr.2018.03.024.
- Anderson, R.H., Spicer, D.E., Brown, N.A. and Mohun, T.J. (2014). The development of septation in the four-chambered heart. *Anatomical Record* 297(8). doi: <https://doi.org/10.1002/ar.22949>.
- Ang, Y.S. et al. 2016a. Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell*. doi: 10.1016/j.cell.2016.11.033.
- Ang, Y.S. et al. 2016b. Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell*. doi: 10.1016/j.cell.2016.11.033.

- Angst, B.D. et al. 1997. Dissociated spatial patterning of gap junctions and cell adhesion junctions during postnatal differentiation of ventricular myocardium. *Circulation Research*. doi: 10.1161/01.RES.80.1.88.
- Arguello, C., De La Cruz, M. V. and Gomez, C.S. (1975). Experimental study of the formation of the heart tube in the chick embryo. *Journal of Embryology and Experimental Morphology* 33(1). doi: <https://doi.org/10.1242/dev.33.1.1>.
- Barrangou, R. et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819). doi: 10.1126/science.1138140.
- Barrangou, R. and Doudna, J.A. 2016. Applications of CRISPR technologies in research and beyond. *Nature Biotechnology* 34(9). doi: 10.1038/nbt.3659.
- Basson, C.T. et al. 1997. Mutations in human cause limb and cardiac malformation in Holt-Oram syndrome. *Nature Genetics*. doi: 10.1038/ng0197-30.
- Beck, S., Le Good, J.A., Guzman, M., Haim, N. Ben, Roy, K., Beermann, F. and Constam, D.B. 2002. Extraembryonic proteases regulate Nodal signalling during gastrulation. *Nature Cell Biology* 4(12). doi: 10.1038/ncb890.
- Ben-Haim, N. et al. 2006. The Nodal Precursor Acting via Activin Receptors Induces Mesoderm by Maintaining a Source of Its Convertases and BMP4. *Developmental Cell* 11(3). doi: 10.1016/j.devcel.2006.07.005.
- Van Den Berg, C.W. et al. 2015. Transcriptome of human foetal heart compared with cardiomyocytes from pluripotent stem cells. *Development (Cambridge)* 142(18). doi: 10.1242/dev.123810.
- Bergmann, O. et al. 2009. Evidence for cardiomyocyte renewal in humans. *Science*. doi: 10.1126/science.1164680.
- Van Berlo, J.H., Elrod, J.W., Van Den Hoogenhof, M.M.G., York, A.J., Aronow, B.J., Duncan, S.A. and Molkenkin, J.D. 2010. The transcription factor GATA-6 regulates pathological cardiac hypertrophy. *Circulation Research* 107(8). doi: 10.1161/CIRCRESAHA.110.220764.
- Berridge, M.J. 1997. Elementary and global aspects of calcium signalling. *Journal of Experimental Biology* 200(2). doi: 10.1242/jeb.200.2.315.
- Bhate, P. et al. 2023. 3D genome organization around nuclear speckles drives mRNA splicing efficiency. *Biorxiv Preprint*.
- Bhatnagar, P., Wickramasinghe, K., Wilkins, E. and Townsend, N. 2016. Trends in the epidemiology of cardiovascular disease in the UK. *Heart*. doi: 10.1136/heartjnl-2016-309573.
- Bikoff, E.K., Morgan, M.A. and Robertson, E.J. 2009. An expanding job description for Blimp-1/PRDM1. *Current Opinion in Genetics and Development*. doi: 10.1016/j.gde.2009.05.005.

- Bird, S.D., Doevendans, P.A., Van Rooijen, M.A., Brutel De La Riviere, A., Hassink, R.J., Passier, R. and Mummery, C.L. 2003. The human adult cardiomyocyte phenotype. *Cardiovascular Research* 58(2). doi: 10.1016/S0008-6363(03)00253-0.
- Bisping, E. et al. 2006. Gata4 is required for maintenance of postnatal cardiac function and protection from pressure overload-induced heart failure. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0602543103.
- Blue, G.M. et al. 2014. Targeted next-generation sequencing identifies pathogenic variants in familial congenital heart disease. *Journal of the American College of Cardiology*. doi: 10.1016/j.jacc.2014.09.048.
- Boateng, S.Y. and Goldspink, P.H. 2008. Assembly and maintenance of the sarcomere night and day. *Cardiovascular Research* 77(4). doi: 10.1093/cvr/cvm048.
- Bock, C. et al. 2011. Reference maps of human es and ips cell variation enable high-throughput characterization of pluripotent cell lines. *Cell*. doi: 10.1016/j.cell.2010.12.032.
- Borok, M.J., Papaioannou, V.E. and Sussel, L. 2016. Unique functions of Gata4 in mouse liver induction and heart development. *Developmental Biology*. doi: 10.1016/j.ydbio.2015.12.007.
- von Both, I. et al. 2004. Foxh1 is essential for development of the anterior heart field. *Developmental Cell* 7(3). doi: 10.1016/j.devcel.2004.07.023.
- Bouma, M.J., van Iterson, M., Janssen, B., Mummery, C.L., Salvatori, D.C.F. and Freund, C. 2017. Differentiation-Defective Human Induced Pluripotent Stem Cells Reveal Strengths and Limitations of the Teratoma Assay and In Vitro Pluripotency Assays. *Stem Cell Reports* 8(5). doi: 10.1016/j.stemcr.2017.03.009.
- Bourajjaj, M. et al. 2008. NFATc2 is a necessary mediator of calcineurin-dependent cardiac hypertrophy and heart failure. *Journal of Biological Chemistry* 283(32). doi: 10.1074/jbc.M801296200.
- Brauburger, K., Akyildiz, S., Ruppert, J.G., Graeb, M., Bernkopf, D.B., Hadjihannas, M. V. and Behrens, J. 2014. Adenomatous polyposis coli (APC) membrane recruitment 3, a member of the APC membrane recruitment family of APC-binding proteins, is a positive regulator of Wnt- β -catenin signalling. *FEBS Journal* 281(3). doi: 10.1111/febs.12624.
- Bray, M.A., Sheehy, S.P. and Parker, K.K. 2008. Sarcomere alignment is regulated by myocyte shape. *Cell Motility and the Cytoskeleton* 65(8). doi: 10.1002/cm.20290.
- Brennan, J., Lu, C.C., Norris, D.P., Rodriguez, T.A., Beddington, R.S.P. and Robertson, E.J. 2001. Nodal signalling in the epiblast patterns the early mouse embryo. *Nature* 411(6840). doi: 10.1038/35082103.
- Briganti, F. et al. 2020. iPSC Modeling of RBM20-Deficient DCM Identifies Upregulation of RBM20 as a Therapeutic Strategy. *Cell Reports* 32(10). doi: 10.1016/j.celrep.2020.108117.
- Briggs, R. and King, T.J. (1952). Transplantation of living nuclei from blastula cells into

enucleated frogs' eggs. *Proceedings of the National Academy of Sciences* 38(5). doi: <https://doi.org/10.1073/pnas.38.5.455>.

Broughton, K.M. et al. 2016. A myosin activator improves actin assembly and sarcomere function of human-induced pluripotent stem cell-derived cardiomyocytes with a troponin T point mutation. *American Journal of Physiology - Heart and Circulatory Physiology* 311(1). doi: 10.1152/ajpheart.00162.2016.

Brown, C.O., Chi, X., Garcia-Gras, E., Shirai, M., Feng, X.H. and Schwartz, R.J. 2004. The Cardiac Determination Factor, Nkx2-5, Is Activated by Mutual Cofactors GATA-4 and Smad1/4 via a Novel Upstream Enhancer. *Journal of Biological Chemistry* 279(11). doi: 10.1074/jbc.M301648200.

Bruneau, B.G. et al. 2001. A murine model of Holt-Oram syndrome defines roles of the T-Box transcription factor Tbx5 in cardiogenesis and disease. *Cell*. doi: 10.1016/S0092-8674(01)00493-7.

Bruneau, B.G. 2013. Signaling and transcriptional networks in heart development and regeneration. *Cold Spring Harbor Perspectives in Biology*. doi: 10.1101/cshperspect.a008292.

Bruneau, B.G., Logan, M., Davis, N., Levi, T., Tabin, C.J., Seidman, J.G. and Seidman, C.E. 1999. Chamber-specific cardiac expression of Tbx5 and heart defects in Holt- Oram syndrome. *Developmental Biology*. doi: 10.1006/dbio.1999.9298.

Burridge, P.W. et al. 2014a. Chemically defined generation of human cardiomyocytes. *Nature Methods*. doi: 10.1038/nmeth.2999.

Burridge, P.W. et al. 2014b. Chemically defined generation of human cardiomyocytes. *Nature Methods*. doi: 10.1038/nMeth.2999.

Burridge, P.W., Holmström, A. and Wu, J.C. 2015. Chemically Defined Culture and Cardiomyocyte Differentiation of Human Pluripotent Stem Cells. *Current protocols in human genetics*. doi: 10.1002/0471142905.hg2103s87.

Butcher, J.T. and Markwald, R.R. (2007). Valvulogenesis: The moving target. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362(1484).

Cai, C.L. et al. 2005. T-box genes coordinate regional rates of proliferation and regional specification during cardiogenesis. *Development* 132(10). doi: 10.1242/dev.01832.

Cai, C.L., Liang, X., Shi, Y., Chu, P.H., Pfaff, S.L., Chen, J. and Evans, S. 2003. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Developmental Cell* 5(6). doi: 10.1016/S1534-5807(03)00363-0.

Cai, K.Q., Capo-Chichi, C.D., Rula, M.E., Yang, D.H. and Xu, X.X.M. 2008. Dynamic GATA6 expression in primitive endoderm formation and maturation in early mouse embryogenesis. *Developmental Dynamics* 237(10). doi: 10.1002/dvdy.21703.

- Caprioli, A., Koyano-Nakagawa, N., Iacovino, M., Shi, X., Ferdous, A., Harvey, R.P., ... Garry, D.J. (2011). Nkx2-5 represses gata1 gene expression and modulates the cellular fate of cardiac progenitors during embryogenesis. *Circulation* 123(15). doi: <https://doi.org/10.1161/CIRCULATIONAHA.110.008185>.
- Carlson, H., Ota, S., Campbell, C.E. and Hurlin, P.J. 2001. A dominant repression domain in Tbx3 mediates transcriptional repression and cell immortalization: Relevance to mutations in Tbx3 that cause ulnar-mammary syndrome. *Human Molecular Genetics* 10(21). doi: 10.1093/hmg/10.21.2403.
- Carrasco, M., Delgado, I., Soria, B., Martín, F. and Rojas, A. 2012. GATA4 and GATA6 control mouse pancreas organogenesis. *Journal of Clinical Investigation* 122(10). doi: 10.1172/JCI63240.
- Cartledge, J.E. et al. 2015. Functional crosstalk between cardiac fibroblasts and adult cardiomyocytes by soluble mediators. *Cardiovascular Research* 105(3). doi: 10.1093/cvr/cvu264.
- Castaño, J., Aranda, S., Bueno, C., Calero-Nieto, F.J., Mejia-Ramirez, E., Mosquera, J.L., ... Giorgetti, A. (2019). GATA2 Promotes Hematopoietic Development and Represses Cardiac Differentiation of Human Mesoderm. *Stem Cell Reports* 13(3). doi: <https://doi.org/10.1016/j.stemcr.2019.07.009>.
- Cerrone, M., Remme, C.A., Tadros, R., Bezzina, C.R. and Delmar, M. 2019. Beyond the One Gene–One Disease Paradigm. *Circulation* 140(7). doi: 10.1161/circulationaha.118.035954.
- Chan, S.S.K., Shi, X., Toyama, A., Arpke, R.W., Dandapat, A., Iacovino, M., ... Kyba, M. (2013). Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. *Cell Stem Cell* 12(5). doi: <https://doi.org/10.1016/j.stem.2013.03.004>.
- Charron, F., Paradis, P., Bronchain, O., Nemer, G. and Nemer, M. 1999. Cooperative Interaction between GATA-4 and GATA-6 Regulates Myocardial Gene Expression. *Molecular and Cellular Biology*. doi: 10.1128/mcb.19.6.4355.
- Chazaud, C., Yamanaka, Y., Pawson, T. and Rossant, J. 2006. Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. *Developmental Cell* 10(5). doi: 10.1016/j.devcel.2006.02.020.
- Chen, F., De Diego, C., Chang, M.G., McHarg, J.L., John, S., Klitzner, T.S. and Weiss, J.N. 2010. Atrioventricular conduction and arrhythmias at the initiation of beating in embryonic mouse hearts. *Developmental Dynamics* 239(7). doi: 10.1002/dvdy.22319.
- Cheng, G., Hagen, T.P., Dawson, M.L., Barnes, K. V. and Menick, D.R. 1999. The role of GATA, CARG, E-box, and a novel element in the regulation of cardiac expression of the Na⁺-Ca²⁺ exchanger gene. *Journal of Biological Chemistry* 274(18). doi: 10.1074/jbc.274.18.12819.
- Cheng, H., Lederer, W.J. and Cannell, M.B. 1993. Calcium sparks: Elementary events underlying excitation-contraction coupling in heart muscle. *Science* 262(5134). doi: 10.1126/science.8235594.

Chia, C.Y. et al. 2019. GATA6 Cooperates with EOMES/SMAD2/3 to Deploy the Gene Regulatory Network Governing Human Definitive Endoderm and Pancreas Formation. *Stem Cell Reports*. doi: 10.1016/j.stemcr.2018.12.003.

Cho, G.S., Park, D.S., Choi, S.C. and Han, J.K. 2017. Tbx2 regulates anterior neural specification by repressing FGF signaling pathway. *Developmental Biology* 421(2). doi: 10.1016/j.ydbio.2016.11.020.

Chomczynski, P. and Sacchi, N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*. doi: 10.1016/0003-2697(87)90021-2.

Chong, C.R., Clarke, K. and Levelt, E. 2017. Metabolic remodelling in diabetic cardiomyopathy. *Cardiovascular Research*. doi: 10.1093/cvr/cvx018.

Christoffels, V.M., Habets, P.E.M.H., Franco, D., Campione, M., De Jong, F., Lamers, W.H., ... Moorman, A.F.M. (2000). Chamber formation and morphogenesis in the developing mammalian heart. *Developmental Biology* 223(2). doi: <https://doi.org/10.1006/dbio.2000.9753>.

Christoffels, V. and Jensen, B. (2020). Cardiac morphogenesis: Specification of the four-chambered heart. *Cold Spring Harbor Perspectives in Biology* 12(10). doi: <https://doi.org/10.1101/cshperspect.a037143>.

Christoffels, V.M., Hoogaars, W.M.H., Tessari, A., Clout, D.E.W., Moorman, A.F.M. and Campione, M. 2004. T-Box Transcription Factor Tbx2 Represses Differentiation and Formation of the Cardiac Chambers. *Developmental Dynamics*. doi: 10.1002/dvdy.10487.

Churko, J.M. et al. 2018. Defining human cardiac transcription factor hierarchies using integrated single-cell heterogeneity analysis. *Nature Communications*. doi: 10.1038/s41467-018-07333-4.

Cmarko, D. et al. 1999. Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection. *Molecular Biology of the Cell* 10(1). doi: 10.1091/mbc.10.1.211.

Coll, M., Seidman, J.G. and Müller, C.W. 2002. Structure of the DNA-bound T-box domain of human TBX3, a transcription factor responsible for ulnar-mammary syndrome. *Structure*. doi: 10.1016/S0969-2126(02)00722-0.

Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., Kispert, A., Herrmann, B. and Robertson, E.J. 1994. A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* 120(7). doi: 10.1242/dev.120.7.1919.

Correia, C. et al. 2018. 3D aggregate culture improves metabolic maturation of human pluripotent stem cell derived cardiomyocytes. *Biotechnology and Bioengineering* 115(3). doi: 10.1002/bit.26504.

- Cowling, R.T., Kupsy, D., Kahn, A.M., Daniels, L.B. and Greenberg, B.H. 2019. Mechanisms of cardiac collagen deposition in experimental models and human disease. *Translational Research* 209. doi: 10.1016/j.trsl.2019.03.004.
- Crawford, N.T., McIntyre, A.J., McCormick, A., D'Costa, Z.C., Buckley, N.E. and Mullan, P.B. 2019. TBX2 interacts with heterochromatin protein 1 to recruit a novel repression complex to EGR1-targeted promoters to drive the proliferation of breast cancer cells. *Oncogene*. doi: 10.1038/s41388-019-0853-z.
- Craven, S.E., Lim, K.C., Ye, W., Engel, J.D., de Sauvage, F. and Rosenthal, A. (2004). Gata2 specifies serotonergic neurons downstream of sonic hedgehog. *Development* 131(5). doi: <https://doi.org/10.1242/dev.01024>.
- Crocini, C. and Gotthardt, M. 2021. Cardiac sarcomere mechanics in health and disease. *Biophysical Reviews* 13(5). doi: 10.1007/s12551-021-00840-7.
- Cui, Y. et al. 2019. Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart. *Cell Reports* 26(7). doi: 10.1016/j.celrep.2019.01.079.
- Cuomo, A.S.E. et al. 2020. Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nature Communications* 11(1). doi: 10.1038/s41467-020-14457-z.
- Cyganek, L. et al. 2018. Deep phenotyping of human induced pluripotent stem cell-derived atrial and ventricular cardiomyocytes. *JCI insight* 3(12). doi: 10.1172/jci.insight.99941.
- D'Amato, G. et al. 2016. Sequential Notch activation regulates ventricular chamber development. *Nature Cell Biology* 18(1). doi: 10.1038/ncb3280.
- David, R., Jarsch, V.B., Schwarz, F., Nathan, P., Gegg, M., Lickert, H. and Franz, W.M. 2011. Induction of MesP1 by Brachyury(T) generates the common multipotent cardiovascular stem cell. *Cardiovascular Research* 92(1). doi: 10.1093/cvr/cvr158.
- D.E., S., N., H., P., P. and D.D., D.L. 2015. A novel mutation in GATA6 causes pancreatic agenesis. *Pediatric Diabetes*.
- Decker, K., Goldman, D.C., L. Grasch, C. and Sussel, L. 2006. Gata6 is an important regulator of mouse pancreas development. *Developmental Biology* 298(2). doi: 10.1016/j.ydbio.2006.06.046.
- Devine, W.P., Wythe, J.D., George, M., Koshiba-Takeuchi, K. and Bruneau, B.G. 2014. Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *eLife*. doi: 10.7554/eLife.03848.
- Dixon, I.M.C., Landry, N.M. and Rattan, S.G. 2019. Periostin Reexpression in Heart Disease Contributes to Cardiac Interstitial Remodeling by Supporting the Cardiac Myofibroblast Phenotype. In: *Advances in Experimental Medicine and Biology*. doi: 10.1007/978-981-13-6657-4_4.

- Doench, J.G. et al. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature biotechnology*. doi: 10.1038/nbt.3437.
- Dodou, E., Verzi, M.P., Anderson, J.P., Xu, S.M. and Black, B.L. (2004). Mef2c is a direct transcriptional target of ISL1 and GATA factors in the anterior heart field during mouse embryonic development. *Development*. doi: <https://doi.org/10.1242/dev.01256>.
- Doppler, S.A. et al. 2017. Cardiac fibroblasts: More than mechanical support. *Journal of Thoracic Disease*. doi: 10.21037/jtd.2017.03.122.
- Drakhlis, L. et al. 2021. Human heart-forming organoids recapitulate early heart and foregut development. *Nature Biotechnology*. doi: 10.1038/s41587-021-00815-9.
- Dupays, L., Kotecha, S., Angst, B. and Mohun, T.J. 2009. Tbx2 misexpression impairs deployment of second heart field derived progenitor cells to the arterial pole of the embryonic heart. *Developmental Biology* 333(1). doi: 10.1016/j.ydbio.2009.06.025.
- Durocher, D., Charron, F., Warren, R., Schwartz, R.J. and Nemer, M. 1997. The cardiac transcription factors nkx2-5 and GATA-4 are mutual cofactors. *EMBO Journal*. doi: 10.1093/emboj/16.18.5687.
- Ellesøe, S.G., Jensen, A.B., Ängquist, L.H., Hjortdal, V.E., Larsen, L.A. and Brunak, S. 2016. How Suitable Are Registry Data for Recurrence Risk Calculations? Validation of Diagnoses on 1,593 Families With Congenital Heart Disease. *World journal for pediatric & congenital heart surgery*. doi: 10.1177/2150135115615786.
- Elorbany, R. et al. 2022. Single-cell sequencing reveals lineage-specific dynamic genetic regulation of gene expression during human cardiomyocyte differentiation. *PLoS Genetics* 18(1). doi: 10.1371/journal.pgen.1009666.
- Epstein, J.A., Aghajanian, H. and Singh, M.K. 2015. Semaphorin signaling in cardiovascular development. *Cell Metabolism* 21(2). doi: 10.1016/j.cmet.2014.12.015.
- Eschenhagen, T., Bolli, R., Braun, T., Field, L.J., Fleischmann, B.K., Frisén, J., ... Hill, J.A. (2017). Cardiomyocyte regeneration: A consensus statement. *Circulation*. doi: <https://doi.org/10.1161/CIRCULATIONAHA.117.029343>.
- Fahed, A.C., Roberts, A.E., Mital, S. and Lakdawala, N.K. 2014. Heart failure in congenital heart disease. A confluence of acquired and congenital. *Heart Failure Clinics* 10(1). doi: 10.1016/j.hfc.2013.09.017.
- Fesslova, V., Brankovic, J., Lalatta, F., Villa, L., Meli, V., Piazza, L. and Ricci, C. 2011. Recurrence of congenital heart disease in cases with familial risk screened prenatally by echocardiography. *Journal of pregnancy*. doi: 10.1155/2011/368067.
- Feyen, D.A.M. et al. 2020. Metabolic Maturation Media Improve Physiological Function of Human iPSC-Derived Cardiomyocytes. *Cell Reports* 32(3). doi: 10.1016/j.celrep.2020.107925.
- Fisher, J.B., Pulakanti, K., Rao, S. and Duncan, S.A. 2017. GATA6 is essential for endoderm formation from human pluripotent stem cells. *Biology Open* 6(7). doi: 10.1242/bio.026120.

- Fong, S.H., Emelyanov, A., Teh, C. and Korzh, V. 2005. Wnt signalling mediated by Tbx2b regulates cell migration during formation of the neural plate. *Development*. doi: 10.1242/dev.01933.
- Frangoul, H. et al. 2021. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β -Thalassemia. *New England Journal of Medicine* 384(3). doi: 10.1056/nejmoa2031054.
- Fujii, M. et al. 2017. Sfrp5 identifies murine cardiac progenitors for all myocardial structures except for the right ventricle. *Nature Communications* 8. doi: 10.1038/ncomms14664.
- Fumasoni, I., Meani, N., Rambaldi, D., Scafetta, G., Alcalay, M. and Ciccarelli, F.D. 2007. Family expansion and gene rearrangements contributed to the functional specialization of PRDM genes in vertebrates. *BMC Evolutionary Biology*. doi: 10.1186/1471-2148-7-187.
- Firulli, A.B., McFadden, D.G., Lin, Q., Srivastava, D. and Olson, E.N. (1998). Heart and extra-embryonic mesodermal defects in mouse embryos lacking the bHLH transcription factor Hand1. *Nature Genetics* 18(3). doi: <https://doi.org/10.1038/ng0398-266>.
- Gaio, U. et al. 1999. A role of the cryptic gene in the correct establishment of the left-right axis. *Current Biology* 9(22). doi: 10.1016/S0960-9822(00)80059-7.
- Gajewski, K., Fossett, N., Molkentin, J.D. and Schulz, R.A. 1999. The zinc finger proteins Pannier and GATA4 function as cardiogenic factors in *Drosophila*. *Development (Cambridge, England)*.
- Galdos, F.X. et al. 2022. Combined Lineage Tracing and scRNA-seq Reveals Unexpected First Heart Field Predominance of Human iPSC Differentiation. *bioRxiv*.
- Gallagher, J.M., Komati, H., Roy, E., Nemer, M. and Latinkic, B. V. 2012. Dissociation of Cardiogenic and Postnatal Myocardial Activities of GATA4. *Molecular and Cellular Biology*. doi: 10.1128/mcb.00218-12.
- Garg, V. et al. 2003. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature*. doi: 10.1038/nature01827.
- Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109(39). doi: 10.1073/pnas.1208507109.
- Ge, S.X., Jung, D., Jung, D. and Yao, R. 2020. ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36(8). doi: 10.1093/bioinformatics/btz931.
- Germain, P.L. and Testa, G. 2017. Taming Human Genetic Variability: Transcriptomic Meta-Analysis Guides the Experimental Design and Interpretation of iPSC-Based Disease Modeling. *Stem Cell Reports* 8(6). doi: 10.1016/j.stemcr.2017.05.012.
- Germanguz, I. et al. 2011. Molecular characterization and functional properties of cardiomyocytes derived from human inducible pluripotent stem cells. *Journal of Cellular and Molecular Medicine* 15(1). doi: 10.1111/j.1582-4934.2009.00996.x.

Ghatnekar, A. and Trojanowska, M. 2008. GATA-6 is a novel transcriptional repressor of the human Tenascin-C gene expression in fibroblasts. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1779(3). doi: 10.1016/j.bbagr.2007.11.012.

Gibson-Brown, J.J., I. Agulnik, S., Silver, L.M. and Papaioannou, V.E. 1998. Expression of T-box genes Tbx2-Tbx5 during chick organogenesis. *Mechanisms of Development* 74(1–2). doi: 10.1016/S0925-4773(98)00056-2.

Gillis, W.Q., St John, J., Bowerman, B. and Schneider, S.Q. 2009. Whole genome duplications and expansion of the vertebrate GATA transcription factor gene family. *BMC Evolutionary Biology*. doi: 10.1186/1471-2148-9-207.

Gonzalez-Teran, B. et al. 2022. Transcription factor protein interactomes reveal genetic determinants in heart disease. *Cell* 185(5). doi: 10.1016/j.cell.2022.01.021.

Grncharova, T. et al. 2021. A comprehensive analysis of gene expression changes in a high replicate and open-source dataset of differentiating hiPSC-derived cardiomyocytes. *Scientific Reports* 11(1). doi: 10.1038/s41598-021-94732-1.

Grépin, C., Nemer, G. and Nemer, M. 1997. Enhanced cardiogenesis in embryonic stem cells overexpressing the GATA-4 transcription factor. *Development*.

Grépin, C., Robitaille, L., Antakly, T. and Nemer, M. 1995. Inhibition of transcription factor GATA-4 expression blocks in vitro cardiac muscle differentiation. *Molecular and Cellular Biology*. doi: 10.1128/mcb.15.8.4095.

Greulich, F., Rudat, C. and Kispert, A. 2011. Mechanisms of T-box gene function in the developing heart. *Cardiovascular Research*. doi: 10.1093/cvr/cvr112.

Groenewegen, A., Rutten, F.H., Mosterd, A. and Hoes, A.W. (2020). Epidemiology of heart failure. *European Journal of Heart Failure* 22(8).

Guo, X., Ramirez, A., Waddell, D.S., Li, Z., Liu, X. and Wang, X.F. 2008. Axin and GSK3- β control Smad3 protein stability and modulate TGF- β signaling. *Genes and Development* 22(1). doi: 10.1101/gad.1590908.

Guo, Y. and Pu, W.T. 2020. Cardiomyocyte maturation: New phase in development. *Circulation Research*. doi: 10.1161/CIRCRESAHA.119.315862.

Gurdon, J.B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *Journal of embryology and experimental morphology* 10. doi: <https://doi.org/10.1242/dev.10.4.622>.

Gurdon, J.B., Elsdale, T.R. and Fischberg, M. (1958). Sexually mature individuals of *Xenopus laevis* from the transplantation of single somatic nuclei. *Nature* 182(4627). doi: <https://doi.org/10.1038/182064a0>.

Gurdon, J.B., Laskey, R.A. and Reeves, O.R. (1975). The developmental capacity of nuclei transplanted from keratinized skin cells of adult frogs. *Journal of Embryology and Experimental Morphology* 34(1). doi: <https://doi.org/10.1242/dev.34.1.93>.

Gurdon, J.B. and Uehlinger, V. (1966). 'Fertile' intestine nuclei. *Nature* 210(5042). doi: <https://doi.org/10.1038/2101240a0>.

Guye, P. et al. 2016. Genetically engineering self-organization of human pluripotent stem cells into a liver bud-like tissue using Gata6. *Nature Communications* 7. doi: 10.1038/ncomms10243.

Györy, I., Fejér, G., Ghosh, N., Seto, E. and Wright, K.L. 2003. Identification of a Functionally Impaired Positive Regulatory Domain I Binding Factor 1 Transcription Repressor in Myeloma Cell Lines. *The Journal of Immunology* 170(6). doi: 10.4049/jimmunol.170.6.3125.

Habets, P.E.M.H. et al. 2002a. Cooperative action of Tbx2 and Nkx2.5 inhibits ANF expression in the atrioventricular canal: Implications for cardiac chamber formation. *Genes and Development* 16(10). doi: 10.1101/gad.222902.

Habets, P.E.M.H. et al. 2002b. Cooperative action of Tbx2 and Nkx2.5 inhibits ANF expression in the atrioventricular canal: Implications for cardiac chamber formation. *Genes and Development* 16(10). doi: 10.1101/gad.222902.

Haegel, H., Larue, L., Ohsugi, M., Fedorov, L., Herrenknecht, K. and Kemler, R. 1995. Lack of β -catenin affects mouse development at gastrulation. *Development* 121(11). doi: 10.1242/dev.121.11.3529.

Haftbaradaran Esfahani, P. and Knöll, R. 2020. Cell shape: effects on gene expression and signaling. *Biophysical Reviews* 12(4). doi: 10.1007/s12551-020-00722-4.

Han, L., Shen, W.J., Bittner, S., Kraemer, F.B. and Azhar, S. 2017. PPARs: Regulators of metabolism and as therapeutic targets in cardiovascular disease. Part I: PPAR- α . *Future Cardiology* 13(3). doi: 10.2217/fca-2016-0059.

Harper, J., Mould, A., Andrews, R.M., Bikoff, E.K. and Robertson, E.J. 2011. The transcriptional repressor Blimp1/Prdm1 regulates postnatal reprogramming of intestinal enterocytes. *Proceedings of the National Academy of Sciences of the United States of America* 108(26). doi: 10.1073/pnas.1105852108.

Harrelson, Z., Kelly, R.G., Goldin, S.N., Gibson-Brown, J.J., Bollag, R.J., Silver, L.M. and Papaioannou, V.E. 2004. Tbx2 is essential for patterning the atrioventricular canal and for morphogenesis of the outflow tract during heart development. *Development*. doi: 10.1242/dev.01378.

Harris, S.P. et al. 2002. Hypertrophic cardiomyopathy in cardiac myosin binding protein-C knockout mice. *Circulation Research* 90(5). doi: 10.1161/01.RES.0000012222.70819.64.

Hasegawa, K., Lee, S.J., Jobe, S.M., Markham, B.E. and Kitsis, R.N. 1997. cis-Acting sequences that mediate induction of β -myosin heavy chain gene expression during left ventricular hypertrophy due to aortic constriction. *Circulation* 96(11). doi: 10.1161/01.CIR.96.11.3943.

Haworth, K.E., Kotecha, S., Mohun, T.J. and Latinkic, B. V. 2008. GATA4 and GATA5 are essential for heart and liver development in *Xenopus* embryos. *BMC Developmental Biology* 8. doi: 10.1186/1471-213X-8-74.

- He, A. et al. 2014. Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nature Communications* 5. doi: 10.1038/ncomms5907.
- He, A., Kong, S.W., Ma, Q. and Pu, W.T. 2011. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1016959108.
- He, C., Cheng, H. and Zhou, R. (2007). GATA family of transcription factors of vertebrates: Phylogenetics and chromosomal synteny. *Journal of Biosciences* 32(3). doi: <https://doi.org/10.1007/s12038-007-0136-7>.
- He, L. et al. 2022. Zebrafish Foxc1a controls ventricular chamber maturation by directly regulating *wwtr1* and *nkx2.5* expression. *Journal of Genetics and Genomics* 49(6). doi: 10.1016/j.jgg.2021.12.002.
- Heikinheimo, M., Scandrett, J.M. and Wilson, D.B. 1994. Localization of Transcription Factor GATA-4 to Regions of the Mouse Embryo Involved in Cardiac Development. *Developmental Biology*. doi: 10.1006/dbio.1994.1206.
- Heras-Bautista, C.O. et al. 2019. Cardiomyocytes facing fibrotic conditions re-express extracellular matrix transcripts. *Acta Biomaterialia* 89. doi: 10.1016/j.actbio.2019.03.017.
- Heslop, J.A., Pournasr, B., Liu, J.T. and Duncan, S.A. 2021. GATA6 defines endoderm fate by controlling chromatin accessibility during differentiation of human-induced pluripotent stem cells. *Cell Reports* 35(7). doi: 10.1016/j.celrep.2021.109145.
- Hinger, S.A., Wei, J., Dorn, L.E., Whitson, B.A., Janssen, P.M.L., He, C. and Accornero, F. 2021. Remodeling of the m6A landscape in the heart reveals few conserved post-transcriptional events underlying cardiomyocyte hypertrophy. *Journal of Molecular and Cellular Cardiology* 151. doi: 10.1016/j.yjmcc.2020.11.002.
- Hinz, B., Phan, S.H., Thannickal, V.J., Galli, A., Bochaton-Piallat, M.L. and Gabbiani, G. 2007. The myofibroblast: One function, multiple origins. *American Journal of Pathology* 170(6). doi: 10.2353/ajpath.2007.070112.
- Hiroi, Y., Kudoh, S., Monzen, K., Ikeda, Y., Yazaki, Y., Nagai, R. and Komuro, I. 2001. Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nature Genetics*. doi: 10.1038/90123.
- Ho, D., Yan, L., Iwatsubo, K., Vatner, D.E. and Vatner, S.F. 2010. Modulation of β -adrenergic receptor signaling in heart failure and longevity: Targeting adenylyl cyclase type 5. *Heart Failure Reviews* 15(5). doi: 10.1007/s10741-010-9183-5.
- Holtzinger, A. and Evans, T. 2007. Gata5 and Gata6 are functionally redundant in zebrafish for specification of cardiomyocytes. *Developmental Biology* 312(2). doi: 10.1016/j.ydbio.2007.09.018.
- Home, P., Kumar, R.P., Ganguly, A., Saha, B., Milano-Foster, J., Bhattacharya, B., ... Paul, S. (2017). Genetic redundancy of GATA factors in the extraembryonic trophoblast lineage

ensures the progression of preimplantation and postimplantation mammalian development. *Development (Cambridge)* 144(5). doi: <https://doi.org/10.1242/dev.145318>.

Homsy, J. et al. 2015. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science*. doi: 10.1126/science.aac9396.

Hoogaars, W.M.H., Barnett, P., Rodriguez, M., Clout, D.E., Moorman, A.F.M., Goding, C.R. and Christoffels, V.M. 2008. TBX3 and its splice variant TBX3 + exon 2a are functionally similar. *Pigment Cell and Melanoma Research* 21(3). doi: 10.1111/j.1755-148X.2008.00461.x.

Horb, M.E. and Thomsen, G.H. 1999. Tbx5 is essential for heart development. *Development*.

Horikoshi, Y. et al. 2019. Fatty Acid-Treated Induced Pluripotent Stem Cell-Derived Human Cardiomyocytes Exhibit Adult Cardiomyocyte-Like Energy Metabolism Phenotypes. *Cells* 8(9). doi: 10.3390/cells8091095.

Horsley, V. et al. 2006. Blimp1 Defines a Progenitor Population that Governs Cellular Input to the Sebaceous Gland. *Cell*. doi: 10.1016/j.cell.2006.06.048.

Hoshino, H., Shioi, G. and Aizawa, S. 2015. AVE protein expression and visceral endoderm cell behavior during anterior-posterior axis formation in mouse embryos: Asymmetry in OTX2 and DKK1 expression. *Developmental Biology* 402(2). doi: 10.1016/j.ydbio.2015.03.023.

Hsiao, T., Maures, T., Waite, K., Yang, J., Kelso, R., Holden, K. and Stoner, R. 2018. Inference of CRISPR Edits from Sanger Trace Data. *bioRxiv*. doi: 10.1101/251082.

Hsu, P.D. et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*. doi: 10.1038/nbt.2647.

Huang, C.Y., Peres Moreno Maia-Joca, R., Ong, C.S., Wilson, I., DiSilvestre, D., Tomaselli, G.F. and Reich, D.H. 2020. Enhancement of human iPSC-derived cardiomyocyte maturation by chemical conditioning in a 3D environment. *Journal of Molecular and Cellular Cardiology*. doi: 10.1016/j.yjmcc.2019.10.001.

Hug, N., Longman, D. and Cáceres, J.F. 2015. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research* 44(4). doi: 10.1093/nar/gkw010.

Hung, J., Teng, T.H.K., Finn, J., Knuiman, M., Briffa, T., Stewart, S., ... Hobbs, M. (2013). Trends from 1996 to 2007 in incidence and mortality outcomes of heart failure after acute myocardial infarction: A population-based study of 20 812 patients with first acute myocardial infarction in Western Australia. *Journal of the American Heart Association* 2(5). doi: <https://doi.org/10.1161/JAHA.113.000172>.

Ieda, M., Fu, J.D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G. and Srivastava, D. 2010. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell*. doi: 10.1016/j.cell.2010.07.002.

- Ilik, İ.A. and Aktaş, T. 2022. Nuclear speckles: dynamic hubs of gene expression regulation. *FEBS Journal* 289(22). doi: 10.1111/febs.16117.
- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakamura, A. 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isoenzyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology* 169(12). doi: 10.1128/jb.169.12.5429-5433.1987.
- Ivanovitch, K. et al. 2021. Ventricular, atrial, and outflow tract heart progenitors arise from spatially and molecularly distinct regions of the primitive streak. *PLoS Biology* 19(5). doi: 10.1371/journal.pbio.3001200.
- Ivey, M.J., Kuwabara, J.T., Riggsbee, K.L. and Tallquist, M.D. 2019. Platelet-derived growth factor receptor- α is essential for cardiac fibroblast survival. *American Journal of Physiology - Heart and Circulatory Physiology* 317(2). doi: 10.1152/ajpheart.00054.2019.
- Iyer, D. et al. 2015. Robust derivation of epicardium and its differentiated smooth muscle cell progeny from human pluripotent stem cells. *Development (Cambridge)* 142(8). doi: 10.1242/dev.119271.
- Jain, R. and Epstein, J.A. 2018. Competent for commitment: You've got to have heart! *Genes and Development* 32(1). doi: 10.1101/gad.308353.117.
- Jao, C.Y. and Salic, A. 2008. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proceedings of the National Academy of Sciences of the United States of America* 105(41). doi: 10.1073/pnas.0808480105.
- Jiang, C.L. et al. 2021. Cell type determination for cardiac differentiation occurs soon after seeding of human induced pluripotent stem cells. *bioRxiv*.
- Jiang, F. and Doudna, J.A. 2017. CRISPR-Cas9 Structures and Mechanisms. *Annual Review of Biophysics* 46. doi: 10.1146/annurev-biophys-062215-010822.
- Jiang, Y.Y., Hou, H.T., Yang, Q., Liu, X.C. and He, G.W. 2017. Chloride Channels are Involved in the Development of Atrial Fibrillation - A Transcriptomic and proteomic Study. *Scientific Reports* 7(1). doi: 10.1038/s41598-017-10590-w.
- Jianli Tao, Daniel E. Bauer and Roberto Chiarle. 2023. Assessing and advancing the safety of CRISPR-Cas tools: from DNA to RNA editing. *Nature Communications*.
- Jin, S.C. et al. 2017. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nature Genetics*. doi: 10.1038/ng.3970.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096). doi: 10.1126/science.1225829.
- Kala, K., Haugas, M., Lillväli, K., Guimera, J., Wurst, W., Salminen, M. and Partanen, J. (2009). *Gata2* is a tissue-specific post-mitotic selector gene for midbrain GABAergic neurons. *Development* 136(2). doi: <https://doi.org/10.1242/dev.029900>.

- Kallies, A. et al. 2007. Initiation of Plasma-Cell Differentiation Is Independent of the Transcription Factor Blimp-1. *Immunity*. doi: 10.1016/j.immuni.2007.04.007.
- Kane, C. and Terracciano, C.M.N. 2017. Concise Review: Criteria for Chamber-Specific Categorization of Human Cardiac Myocytes Derived from Pluripotent Stem Cells. *Stem Cells*. doi: 10.1002/stem.2649.
- Karbassi, E., Fenix, A., Marchiano, S., Muraoka, N., Nakamura, K., Yang, X. and Murry, C.E. 2020. Cardiomyocyte maturation: advances in knowledge and implications for regenerative medicine. *Nature Reviews Cardiology* 17(6). doi: 10.1038/s41569-019-0331-x.
- Kattman, S.J., Huber, T.L. and Keller, G.M.M. 2006. Multipotent Flk-1+ Cardiovascular Progenitor Cells Give Rise to the Cardiomyocyte, Endothelial, and Vascular Smooth Muscle Lineages. *Developmental Cell* 11(5). doi: 10.1016/j.devcel.2006.10.002.
- Kaul-Strehlow, S. and Stach, T. 2013. A detailed description of the development of the hemichordate *Saccoglossus kowalevskii* using SEM, TEM, Histology and 3D-reconstructions. *Frontiers in Zoology*. doi: 10.1186/1742-9994-10-53.
- Kelly, R.G., Brown, N.A. and Buckingham, M.E. (2001). The Arterial Pole of the Mouse Heart Forms from Fgf10-Expressing Cells in Pharyngeal Mesoderm. *Developmental Cell* 1(3). doi: [https://doi.org/10.1016/S1534-5807\(01\)00040-5](https://doi.org/10.1016/S1534-5807(01)00040-5).
- Kelly, R.G. and Buckingham, M.E. (2002). The anterior heart-forming field: Voyage to the arterial pole of the heart. *Trends in Genetics* 18(4).
- Kelly, R.G., Buckingham, M.E. and Moorman, A.F. (2014). Heart fields and cardiac morphogenesis. *Cold Spring Harbor Perspectives in Medicine* 4(10). doi: <https://doi.org/10.1101/cshperspect.a015750>.
- Kilpinen, H. et al. 2017. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature*. doi: 10.1038/nature22403.
- Kim, K., Zhao, R., Doi, A., Ng, K., Unternaehrer, J., Cahan, P., ... Daley, G.Q. (2011). Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nature Biotechnology* 29(12). doi: <https://doi.org/10.1038/nbt.2052>.
- Kimura-Yoshida, C. et al. 2005. Canonical Wnt signaling and its antagonist regulate anterior-posterior axis polarization by guiding cell migration in mouse visceral endoderm. *Developmental Cell* 9(5). doi: 10.1016/j.devcel.2005.09.011.
- King, T.J. and Briggs, R. (1954). Transplantation of Living Nuclei of Late Gastrulae into Enucleated Eggs of *Rana pipiens*. *Development* 2(1). doi: <https://doi.org/10.1242/dev.2.1.73>.
- Kinnunen, S.M. et al. 2018. Cardiac Actions of a Small Molecule Inhibitor Targeting GATA4-NKX2-5 Interaction. *Scientific Reports* 8(1). doi: 10.1038/s41598-018-22830-8.
- Kispert, A. and Herrmann, B.G. 1994. Immunohistochemical analysis of the Brachyury protein in wild-type and mutant mouse embryos. *Developmental Biology* 161(1). doi: 10.1006/dbio.1994.1019.

Koban, M.U., Brugh, S.A., Riordon, D.R., Dellow, K.A., Yang, H.T., Tweedie, D. and Boheler, K.R. 2001. A distant upstream region of the rat multipartite Na⁺-Ca²⁺ exchanger NCX1 gene promoter is sufficient to confer cardiac-specific expression. *Mechanisms of Development* 109(2). doi: 10.1016/S0925-4773(01)00548-2.

Kobayashi, S., Lackey, T., Huang, Y., Bisping, E., Pu, W.T., Boxer, L.M. and Liang, Q. 2006. Transcription factor gata4 regulates cardiac BCL2 gene expression in vitro and in vivo. *FASEB Journal*. doi: 10.1096/fj.05-5426fje.

Kodo, K. et al. 2009. GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proceedings of the National Academy of Sciences of the United States of America* 106(33). doi: 10.1073/pnas.0904744106.

Kojima, Y., Tam, O.H. and Tam, P.P.L. 2014. Timing of developmental events in the early mouse embryo. *Seminars in Cell and Developmental Biology* 34. doi: 10.1016/j.semcdb.2014.06.010.

Kolwicz, S.C., Olson, D.P., Marney, L.C., Garcia-Menendez, L., Synovec, R.E. and Tian, R. 2012. Cardiac-specific deletion of acetyl CoA carboxylase 2 prevents metabolic remodeling during pressure-overload hypertrophy. *Circulation Research* 111(6). doi: 10.1161/CIRCRESAHA.112.268128.

Konhilas, J.P., Irving, T.C. and De Tombe, P.P. 2002. Frank-Starling law of the heart and the cellular mechanisms of length-dependent activation. *Pflugers Archiv European Journal of Physiology*. doi: 10.1007/s00424-002-0902-1.

Kotlikoff, M.I. 2003. Calcium-induced calcium release in smooth muscle: The case for loose coupling. *Progress in Biophysics and Molecular Biology* 83(3). doi: 10.1016/S0079-6107(03)00056-7.

Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R. and Grosveld, F. 1999. The transcription factor GATA6 is essential for early extraembryonic development. *Development* 126(4). doi: 10.1242/dev.126.9.723.

Kuo, C.T. et al. 1997. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes and Development* 11(8). doi: 10.1101/gad.11.8.1048.

Kuo, T.C. and Calame, K.L. 2004. B Lymphocyte-Induced Maturation Protein (Blimp)-1, IFN Regulatory Factor (IRF)-1, and IRF-2 Can Bind to the Same Regulatory Sites. *The Journal of Immunology*. doi: 10.4049/jimmunol.173.9.5556.

Krishnan, A., Samtani, R., Dhanantwari, P., Lee, E., Yamada, S., Shiota, K., ... Lo, C.W. (2014). A detailed comparison of mouse and human cardiac development. *Pediatric Research* 76(6). doi: <https://doi.org/10.1038/pr.2014.128>.

Laforest, B. 2011. *Essential role of GATA5 in the mammalian heart*.

Lambert, S.A. et al. 2018. The Human Transcription Factors. *Cell*. doi: 10.1016/j.cell.2018.01.029.

- Landrum, M.J. et al. 2018. ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research* 46(D1). doi: 10.1093/nar/gkx1153.
- Lansford, R. and Rugonyi, S. (2020). Follow me! a tale of avian heart development with comparisons to mammal heart development. *Journal of Cardiovascular Development and Disease* 7(1).
- Laskey, R.A. and Gurdon, J.B. (1970). Genetic content of adult somatic cells tested by nuclear transplantation from Cultured Cells. *Nature* 228(5278). doi: <https://doi.org/10.1038/2281332a0>.
- Latinkic, B. V. 2003. Induction of cardiomyocytes by GATA4 in *Xenopus* ectodermal explants. *Development*. doi: 10.1242/dev.00599.
- Latinkic, B. V. 2004. Transcriptional regulation of the cardiac-specific MLC2 gene during *Xenopus* embryonic development. *Development*. doi: 10.1242/dev.00953.
- Laverriere, A.C., MacNeill, C., Mueller, C., Poelmann, R.E., Burch, J.B.E. and Evans, T. 1994. GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut. *Journal of Biological Chemistry* 269(37). doi: 10.1016/s0021-9258(17)31636-8.
- Lee, W. and van den Veyver, I.B. 2017. 22q11.2 deletion syndrome. In: *Obstetric Imaging: Fetal Diagnosis and Care, 2nd Edition*. doi: 10.1016/B978-0-323-44548-1.00154-6.
- Lee, Y., Shioi, T., Kasahara, H., Jobe, S.M., Wiese, R.J., Markham, B.E. and Izumo, S. 2015. The Cardiac Tissue-Restricted Homeobox Protein Csx/Nkx2.5 Physically Associates with the Zinc Finger Protein GATA4 and Cooperatively Activates Atrial Natriuretic Factor Gene Expression. *Molecular and Cellular Biology*. doi: 10.1128/mcb.18.6.3120.
- Lentjes, M.H., Niessen, H.E., Akiyama, Y., de Bruïne, A.P., Melotte, V. and van Engeland, M. 2016a. The emerging role of GATA transcription factors in development and disease. *Expert Reviews in Molecular Medicine*. doi: 10.1017/erm.2016.2.
- Lentjes, M.H.F.M., Niessen, H.E.C., Akiyama, Y., Bruïne, D.A.P., Melotte, V. and Engeland, M.V.A.N. 2016b. The emerging role of GATA transcription factors in development and disease. *Expert Reviews in Molecular Medicine*. doi: 10.1017/erm.2016.2.
- Lepore, J.J., Mericko, P.A., Cheng, L., Lu, M.M., Morrissey, E.E. and Parmacek, M.S. 2006. GATA-6 regulates semaphorin 3C and is required in cardiac neural crest for cardiovascular morphogenesis. *Journal of Clinical Investigation* 116(4). doi: 10.1172/JCI27363.
- Lescroart, F., Chabab, S., Lin, X., Rulands, S., Paulissen, C., Rodolosse, A., ... Blanpain, C. (2014). Early lineage restriction and regional segregation during mammalian heart development. *Nature cell biology* 16(9).
- Li, J. et al. 2021. EphA4 is highly expressed in the atria of heart and its deletion leads to atrial hypertrophy and electrocardiographic abnormalities in rats. *Life Sciences* 278. doi: 10.1016/j.lfs.2021.119595.
- Li, T., Yang, Y., Qi, H., Cui, W., Zhang, L., Fu, X., ... Yu, T. (2023). CRISPR/Cas9 therapeutics: progress and prospects. *Signal Transduction and Targeted Therapy* 8(1).

Lian, X. et al. 2012. Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proceedings of the National Academy of Sciences of the United States of America* 109(27). doi: 10.1073/pnas.1200250109.

Lian, X. et al. 2013. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/ β -catenin signaling under fully defined conditions. *Nature Protocols* 8(1). doi: 10.1038/nprot.2012.150.

Liang, P. et al. 2016. Patient-Specific and Genome-Edited Induced Pluripotent Stem Cell-Derived Cardiomyocytes Elucidate Single-Cell Phenotype of Brugada Syndrome. *Journal of the American College of Cardiology* 68(19). doi: 10.1016/j.jacc.2016.07.779.

Liang, Q., De Windt, L.J., Witt, S.A., Kimball, T.R., Markham, B.E. and Molkentin, J.D. 2001. The Transcription Factors GATA4 and GATA6 Regulate Cardiomyocyte Hypertrophy in Vitro and in Vivo. *Journal of Biological Chemistry*. doi: 10.1074/jbc.M102174200.

Lickert, H., Kutsch, S., Kanzler, B., Tamai, Y., Taketo, M.M. and Kemler, R. 2002. Formation of multiple hearts in mice following deletion of β -catenin in the embryonic endoderm. *Developmental Cell* 3(2). doi: 10.1016/S1534-5807(02)00206-X.

Lili Zhu, K.C.B.G.-T.Y.-S.A.R.T.N.R.S.L.L.P.Z.C.Z.H.R.Y.H.S.J.A.P.F.K.A.P., Nandhini Sadagopan, A.H.M.W.C.C.A.G.J.G. van B.R.H.V.V.B.R.C.B.L.B.G.B.L.S.N.J.K. and Katherine S Pollard. 2022. Transcription Factor GATA4 Regulates Cell Type-Specific Splicing Through Direct Interaction With RNA in Human Induced Pluripotent Stem Cell-Derived Cardiac Progenitors. *Circulation* 146(10), pp. 770–787.

Lim, K.C., Lakshmanan, G., Crawford, S.E., Gu, Y., Grosveld, F. and Engel, J.D. (2000). Gata3 loss leads to embryonic lethality due to noradrenaline deficiency of the sympathetic nervous system. *Nature Genetics* 25(2). doi: <https://doi.org/10.1038/76080>.

Lin, Q., Schwarz, J., Bucana, C. and Olson, E.N. (1997). Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science* 276(5317). doi: <https://doi.org/10.1126/science.276.5317.1404>.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., ... Ecker, J.R. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 471(7336). doi: <https://doi.org/10.1038/nature09798>.

Litviňuková, M. et al. 2020. Cells of the adult human heart. *Nature* 588(7838). doi: 10.1038/s41586-020-2797-4.

Liu, C., Banister, C.E., Weige, C.C., Altomare, D., Richardson, J.H., Contreras, C.M. and Buckhaults, P.J. 2018a. PRDM1 silences stem cell-related genes and inhibits proliferation of human colon tumor organoids. *Proceedings of the National Academy of Sciences of the United States of America* 115(22). doi: 10.1073/pnas.1802902115.

Liu, Y., Chen, S., Zühlke, L., Black, G.C., Choy, M.K., Li, N. and Keavney, B.D. (2019). Global birth prevalence of congenital heart defects 1970-2017: Updated systematic review and meta-analysis of 260 studies. *International Journal of Epidemiology*. doi: <https://doi.org/10.1093/ije/dyz009>.

- Liu, N. et al. 2018b. Functional variants in TBX2 are associated with a syndromic cardiovascular and skeletal developmental disorder. *Human Molecular Genetics*. doi: 10.1093/hmg/ddy146.
- Liu, Y., Chen, S., Zühlke, L., Black, G.C., Choy, M.K., Li, N. and Keavney, B.D. 2019. Global birth prevalence of congenital heart defects 1970-2017: Updated systematic review and meta-analysis of 260 studies. *International Journal of Epidemiology*. doi: 10.1093/ije/dyz009.
- Lopaschuk, G.D., Karwi, Q.G., Tian, R., Wende, A.R. and Abel, E.D. 2021. Cardiac Energy Metabolism in Heart Failure. *Circulation Research* 128(10). doi: 10.1161/CIRCRESAHA.121.318241.
- Lowry, J.A. and Atchley, W.R. (2000). Molecular evolution of the GATA family of transcription factors: Conservation within the DNA-binding domain. *Journal of Molecular Evolution* 50(2). doi: <https://doi.org/10.1007/s002399910012>.
- Lüdtke, T.H., Wojahn, I., Kleppa, M.J., Schierstaedt, J., Christoffels, V.M., Künzler, P. and Kispert, A. 2021. Combined genomic and proteomic approaches reveal DNA binding sites and interaction partners of TBX2 in the developing lung. *Respiratory Research* 22(1). doi: 10.1186/s12931-021-01679-y.
- Lukyanenko, V., Viatchenko-Karpinski, S., Smirnov, A., Wiesner, T.F. and Györke, S. 2001. Dynamic regulation of sarcoplasmic reticulum Ca²⁺ content and release by luminal Ca²⁺-sensitive leak in rat ventricular myocytes. *Biophysical Journal* 81(2). doi: 10.1016/S0006-3495(01)75741-4.
- Luna-Zurita, L. et al. 2016. Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. *Cell*. doi: 10.1016/j.cell.2016.01.004.
- Lyra-Leite, D.M., Gutiérrez-Gutiérrez, Ó., Wang, M., Zhou, Y., Cyganek, L. and BurrIDGE, P.W. 2022. A review of protocols for human iPSC culture, cardiac differentiation, subtype-specification, maturation, and direct reprogramming. *STAR Protocols* 3(3). doi: 10.1016/j.xpro.2022.101560.
- Ma, Q., Zhou, B. and Pu, W.T. (2008). Reassessment of Isl1 and Nkx2-5 cardiac fate maps using a Gata4-based reporter of Cre activity. *Developmental Biology* 323(1). doi: <https://doi.org/10.1016/j.ydbio.2008.08.013>.
- Ma, L.Y. et al. 2020. China cardiovascular diseases report 2018: An updated summary. *Journal of Geriatric Cardiology* 17(1). doi: 10.11909/j.issn.1671-5411.2020.01.001.
- Maitra, M., Koenig, S.N., Srivastava, D. and Garg, V. 2010. Identification of GATA6 sequence variants in patients with congenital heart defects. *Pediatric Research* 68(4). doi: 10.1203/PDR.0b013e3181ed17e4.
- Maitra, M., Schluterman, M.K., Nichols, H.A., Richardson, J.A., Lo, C.W., Srivastava, D. and Garg, V. 2009. Interaction of Gata4 and Gata6 with Tbx5 is critical for normal cardiac development. *Developmental Biology*. doi: 10.1016/j.ydbio.2008.11.004.

Makarova, K.S., Grishin, N. v., Shabalina, S.A., Wolf, Y.I. and Koonin, E. v. 2006. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* 1. doi: 10.1186/1745-6150-1-7.

Makinde, A.O., Kantor, P.F. and Lopaschuk, G.D. 1998a. Maturation of fatty acid and carbohydrate metabolism in the newborn heart. In: *Molecular and Cellular Biochemistry*. doi: 10.1023/A:1006860104840.

Manuylov, N.L. and Tevosian, S.G. (2009). Cardiac expression of Tnnt1 requires the GATA4-FOG2 transcription complex. *TheScientificWorldJournal* 9. doi: <https://doi.org/10.1100/tsw.2009.75>.

Marchetto, M.C.N., Yeo, G.W., Kainohana, O., Marsala, M., Gage, F.H. and Muotri, A.R. (2009). Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PLoS ONE* 4(9). doi: <https://doi.org/10.1371/journal.pone.0007076>.

Materna, S.C., Sinha, T., Barnes, R.M., Lammerts van Bueren, K. and Black, B.L. 2019. Cardiovascular development and survival require Mef2c function in the myocardial but not the endothelial lineage. *Developmental Biology* 445(2). doi: 10.1016/j.ydbio.2018.12.002.

Mcdermott, D.A. et al. 2005. TBX5 genetic testing validates strict clinical criteria for Holt-Oram syndrome. *Pediatric Research*. doi: 10.1203/01.PDR.0000182593.95441.64.

Mesnard, D., Guzman-Ayala, M. and Constam, D.B. 2006. Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning. *Development* 133(13). doi: 10.1242/dev.02413.

Min, Y.L., Bassel-Duby, R. and Olson, E.N. 2019. CRISPR correction of duchenne muscular dystrophy. *Annual Review of Medicine* 70. doi: 10.1146/annurev-med-081117-010451.

Misra, C. et al. 2012. Congenital heart disease-causing Gata4 mutation displays functional deficits in vivo. *PLoS Genetics*. doi: 10.1371/journal.pgen.1002690.

Misra, C., Chang, S.W., Basu, M., Huang, N. and Garg, V. 2014. Disruption of myocardial Gata4 and Tbx5 results in defects in cardiomyocyte proliferation and atrioventricular septation. *Human Molecular Genetics* 23(19). doi: 10.1093/hmg/ddu215.

Mjaatvedt, C.H., Nakaoka, T., Moreno-Rodriguez, R., Norris, R.A., Kern, M.J., Eisenberg, C.A., ... Markwald, R.R. (2001). The outflow tract of the heart is recruited from a novel heart-forming field. *Developmental Biology* 238(1). doi: <https://doi.org/10.1006/dbio.2001.0409>.

Mohiuddin, M., Kooy, R.F. and Pearson, C.E. (2022). De novo mutations, genetic mosaicism and human disease. *Frontiers in Genetics* 13.

Mojica, F.J.M., Díez-Villaseñor, C., Soria, E. and Juez, G. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology* 36(1). doi: 10.1046/j.1365-2958.2000.01838.x.

Molkentin, J.D. et al. 1998. A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* 93(2). doi: 10.1016/S0092-8674(00)81573-1.

- Molkentin, J.D. 2013. Parsing good versus bad signaling pathways in the heart role of calcineurin-nuclear factor of activated t-cells. *Circulation Research* 113(1). doi: 10.1161/CIRCRESAHA.113.301667.
- Molkentin, J.D., Kalvakolanu, D. V and Markham, B.E. 1994. Transcription factor GATA-4 regulates cardiac muscle-specific expression of the alpha-myosin heavy-chain gene. *Molecular and cellular biology*.
- Molkentin, J.D., Lin, Q., Duncan, S.A. and Olson, E.N. 1997. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes and Development*. doi: 10.1101/gad.11.8.1061.
- Molkentin, J.D., Tymitz, K.M., Richardson, J.A. and Olson, E.N. 2002. Abnormalities of the Genitourinary Tract in Female Mice Lacking GATA5. *Molecular and Cellular Biology*. doi: 10.1128/mcb.20.14.5256-5260.2000.
- Moller, P.C. and Philpott, C.W. 1973. The circulatory system of Amphioxus (*Branchiostoma floridae*) I. Morphology of the major vessels of the pharyngeal area. *Journal of Morphology*. doi: 10.1002/jmor.1051390403.
- Moorman, A.F.M., Christoffels, V.M., Anderson, R.H. and Van Den Hoff, M.J.B. (2007). The heart-forming fields: One or multiple? *Philosophical Transactions of the Royal Society B*:
- Moretti, A. et al. 2006. Multipotent Embryonic Isl1+ Progenitor Cells Lead to Cardiac, Smooth Muscle, and Endothelial Cell Diversification. *Cell* 127(6). doi: 10.1016/j.cell.2006.10.029.
- Morin, S. 2000. GATA-dependent recruitment of MEF2 proteins to target promoters. *The EMBO Journal*. doi: 10.1093/emboj/19.9.2046. *Biological Sciences* 362(1484).
- Moriguchi, T., Takako, N., Hamada, M., Maeda, A., Fujioka, Y., Kuroha, T., ... Engel, J.D. (2006). Gata3 participates in a complex transcriptional feedback network to regulate sympathoadrenal differentiation. *Development* 133(19). doi: <https://doi.org/10.1242/dev.02553>.
- Morrissey, E.E., Ip, H.S., Lu, M.M. and Parmacek, M.S. 1996. GATA-6: A zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Developmental Biology* 177(1). doi: 10.1006/dbio.1996.0165.
- Morrissey, E.E., Ip, H.S., Tang, Z., Lu, M.M. and Parmacek, M.S. 1997a. GATA-5: A transcriptional activator expressed in a novel temporally and spatially-restricted pattern during embryonic development. *Developmental Biology* 183(1). doi: 10.1006/dbio.1996.8485.
- Morrissey, E.E., Ip, H.S., Tang, Z. and Parmacek, M.S. 1997. GATA-4 activates transcription via two novel domains that are conserved within the GATA-4/5/6 subfamily. *Journal of Biological Chemistry* 272(13). doi: 10.1074/jbc.272.13.8515.

- Morrisey, E.E., Tang, Z., Sigrist, K., Lu, M.M., Jiang, F., Ip, H.S. and Parmacek, M.S. 1998. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes and Development* 12(22). doi: 10.1101/gad.12.22.3579.
- Mosqueira, D. et al. 2018. CRISPR/Cas9 editing in human pluripotent stemcell-cardiomyocytes highlights arrhythmias, hypocontractility, and energy depletion as potential therapeutic targets for hypertrophic cardiomyopathy. *European Heart Journal* 39(43). doi: 10.1093/eurheartj/ehy249.
- Mould, A.W., Morgan, M.A.J., Nelson, A.C., Bikoff, E.K. and Robertson, E.J. 2015. Blimp1/Prdm1 Functions in Opposition to Irf1 to Maintain Neonatal Tolerance during Postnatal Intestinal Maturation. *PLoS Genetics*. doi: 10.1371/journal.pgen.1005375.
- Mummery, C.L., Zhang, J., Ng, E.S., Elliott, D.A., Elefanty, A.G. and Kamp, T.J. 2012. Differentiation of human embryonic stem cells and induced pluripotent stem cells to cardiomyocytes: a methods overview. *Circulation research*. doi: 10.1161/CIRCRESAHA.110.227512.
- Naito, A.T., Shiojima, I., Akazawa, H., Hidaka, K., Morisaki, T., Kikuchi, A. and Komuro, I. 2006. Developmental stage-specific biphasic roles of Wnt/ β -catenin signaling in cardiomyogenesis and hematopoiesis. *Proceedings of the National Academy of Sciences of the United States of America* 103(52). doi: 10.1073/pnas.0605768103.
- Nakamura, M. and Sadoshima, J. 2018. Mechanisms of physiological and pathological cardiac hypertrophy. *Nature Reviews Cardiology* 15(7). doi: 10.1038/s41569-018-0007-y.
- Nakanishi, M. et al. 2019. Human Pluripotency Is Initiated and Preserved by a Unique Subset of Founder Cells. *Cell* 177(4). doi: 10.1016/j.cell.2019.03.013.
- Nakano, H. et al. 2017. Glucose inhibits cardiac muscle maturation through nucleotide biosynthesis. *eLife*. doi: 10.7554/eLife.29330.
- Nam, Y.J. et al. 2013. Reprogramming of human fibroblasts toward a cardiac fate. *Proceedings of the National Academy of Sciences of the United States of America* 110(14). doi: 10.1073/pnas.1301019110.
- Nemer, G. and Nemer, M. 2003. Transcriptional activation of BMP-4 and regulation of mammalian organogenesis by GATA-4 and -6. *Developmental Biology*. doi: 10.1016/S0012-1606(02)00026-X.
- Ng, R.K. and Gurdon, J.B. (2008). Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nature Cell Biology* 10(1). doi: <https://doi.org/10.1038/ncb1674>.
- Nishii, K. et al. 2008. Targeted disruption of the cardiac troponin T gene causes sarcomere disassembly and defects in heartbeat within the early mouse embryo. *Developmental Biology* 322(1). doi: 10.1016/j.ydbio.2008.07.007.
- Niu, Z., Yu, W., Zhang, S.X., Barron, M., Belaguli, N.S., Schneider, M.D., ... Schwartz, R.J. (2005). Conditional mutagenesis of the murine serum response factor gene blocks

cardiogenesis and the transcription of downstream gene targets. *Journal of Biological Chemistry* 280(37). doi: <https://doi.org/10.1074/jbc.M501372200>.

Office for National Statistics. 2020. National life tables – life expectancy in the UK. *Office for National Statistics*.

Ohno, Y. et al. 2013. Distinct iPS cells show different cardiac differentiation efficiency. *Stem Cells International*. doi: 10.1155/2013/659739.

Ohi, Y., Qin, H., Hong, C., Blouin, L., Polo, J.M., Guo, T., ... Ramalho-Santos, M. (2011). Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nature Cell Biology* 13(5). doi: <https://doi.org/10.1038/ncb2239>.

Oka, T., Maillet, M., Watt, A.J., Schwartz, R.J., Aronow, B.J., Duncan, S.A. and Molkenkin, J.D. 2006. Cardiac-specific deletion of *gata4* reveals its requirement for hypertrophy, compensation, and myocyte viability. *Circulation Research* 98(6). doi: 10.1161/01.RES.0000215985.18538.c4.

Ohneda, K. and Yamamoto, M. (2002). Roles of hematopoietic transcription factors GATA-1 and GATA-2 in the development of red blood cell lineage. *Acta Haematologica* 108(4). doi: <https://doi.org/10.1159/000065660>.

Oliveros, J.C. 2021. Venny 2.1. *BioinfoGP, CNB-CSIC*.

Olson, E.N. 2006. Gene regulatory networks in the evolution and development of the heart. *Science*. doi: 10.1126/science.1132292.

Osorno, R. and Chambers, I. 2011. Transcription factor heterogeneity and epiblast pluripotency. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366(1575). doi: 10.1098/rstb.2011.0043.

Packham, E.A. 2003. T-box genes in human disorders. *Human Molecular Genetics*. doi: 10.1093/hmg/ddg077.

Paige, S.L., Plonowska, K., Xu, A. and Wu, S.M. 2015. Molecular regulation of cardiomyocyte differentiation. *Circulation Research* 116(2). doi: 10.1161/CIRCRESAHA.116.302752.

Pandolfi, P.P., Roth, M.E., Karis, A., Leonard, M.W., Dzierzak, E., Grosveld, F.G., ... Lindenbaum, M.H. (1995). Targeted disruption of the *GATA3* gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nature Genetics* 11(1). doi: <https://doi.org/10.1038/ng0995-40>.

Papayioannou, V.E. 2001. T-box genes in development: From hydra to humans. *International Review of Cytology*. doi: 10.1016/S0074-7696(01)07002-4.

Papayioannou, V.E. 2014a. The T-box gene family: emerging roles in development, stem cells and cancer. *Development*. doi: 10.1242/dev.104471.

Papayioannou, V.E. 2014b. The t-box gene family: Emerging roles in development, Stem cells and cancer. *Development (Cambridge)*. doi: 10.1242/dev.104471.

- Papanayotou, C. et al. 2014. A Novel Nodal Enhancer Dependent on Pluripotency Factors and Smad2/3 Signaling Conditions a Regulatory Switch During Epiblast Maturation. *PLoS Biology* 12(6). doi: 10.1371/journal.pbio.1001890.
- Paul, M.H., Harvey, R.P., Wegner, M. and Sock, E. 2014. Cardiac outflow tract development relies on the complex function of Sox4 and Sox11 in multiple cell types. *Cellular and Molecular Life Sciences* 71(15). doi: 10.1007/s00018-013-1523-x.
- Perea-Gomez, A. et al. 2002. Nodal antagonists in the anterior visceral endoderm prevent the formation of multiple primitive streaks. *Developmental Cell* 3(5). doi: 10.1016/S1534-5807(02)00321-0.
- Peterkin, T., Gibson, A., Loose, M. and Patient, R. 2005. The roles of GATA-4, -5 and -6 in vertebrate heart development. *Seminars in Cell and Developmental Biology*. doi: 10.1016/j.semcdb.2004.10.003.
- Peterkin, T., Gibson, A. and Patient, R. 2007. Redundancy and evolution of GATA factor requirements in development of the myocardium. *Developmental Biology* 311(2). doi: 10.1016/j.ydbio.2007.08.018.
- Peters, N.S., Severs, N.J., Rothery, S.M., Lincoln, C., Yacoub, M.H. and Green, C.R. (1994). Spatiotemporal relation between gap junctions and fascia adherens junctions during postnatal development of human ventricular myocardium. *Circulation*. doi: <https://doi.org/10.1161/01.CIR.90.2.713>.
- Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H. and Costantini, F. (1995). Development of hematopoietic cells lacking transcription factor GATA-1. *Development* 121(1). doi: <https://doi.org/10.1242/dev.121.1.163>.
- Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., ... Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* 349(6306). doi: <https://doi.org/10.1038/349257a0>.
- Pikkarainen, S. et al. 2003. GATA-4 is a nuclear mediator of mechanical stretch-activated hypertrophic program. *Journal of Biological Chemistry* 278(26). doi: 10.1074/jbc.M302719200.
- Pilon, N., Raiwet, D., Viger, R.S. and Silversides, D.W. 2008. Novel pre- and post-gastrulation expression of Gata4 within cells of the inner cell mass and migratory neural crest cells. *Developmental Dynamics* 237(4). doi: 10.1002/dvdy.21496.
- Ponnalagu, D. et al. 2016. Molecular identity of cardiac mitochondrial chloride intracellular channel proteins. *Mitochondrion* 27. doi: 10.1016/j.mito.2016.01.001.
- Prajapati, R.S., Hintze, M. and Streit, A. 2019. PRDM1 controls the sequential activation of neural, neural crest and sensory progenitor determinants. *Development (Cambridge)* 146(24). doi: 10.1242/dev.181107.
- Prall, O.W., Menon, M.K., Solloway, M.J., Yusuke Watanabe, S.Z., Bajolle, F., Biben, C., ... Harvey, R.P. (2007). An Nkx2-5/Bmp2/Smad1 negative feedback loop controls second heart

field progenitor specification and proliferation. *Ce* 46(1).

Prendiville, T., Jay, P.Y. and Pu, W.T. 2014. Insights into the genetic structure of congenital heart disease from human and murine studies on monogenic disorders. *Cold Spring Harbor Perspectives in Medicine* 4(10). doi: 10.1101/cshperspect.a013946.

Prodan, N. et al. 2022. Direct reprogramming of cardiomyocytes into cardiac Purkinje-like cells. *iScience* 25(11). doi: 10.1016/j.isci.2022.105402.

Quijada, P., Trembley, M.A. and Small, E.M. (2020). The Role of the Epicardium during Heart Development and Repair. *Circulation Research*.

Qureshi, W.T., Zhang, Z.M., Chang, P.P., Rosamond, W.D., Kitzman, D.W., Wagenknecht, L.E. and Soliman, E.Z. (2018). Silent Myocardial Infarction and Long-Term Risk of Heart Failure: The ARIC Study. *Journal of the American College of Cardiology* 71(1). doi: <https://doi.org/10.1016/j.jacc.2017.10.071>.

Rahbari, R., Wuster, A., Lindsay, S.J., Hardwick, R.J., Alexandrov, L.B., Al Turki, S., ... Hurles, M.E. (2016). Timing, rates and spectra of human germline mutation. *Nature Genetics* 48(2). doi: <https://doi.org/10.1038/ng.3469>.

Raid, R., Krinka, D., Bakhoff, L., Abdelwahid, E., Jokinen, E., Kärner, M., ... Karis, A. (2009). Lack of Gata3 results in conotruncal heart anomalies in mouse. *Mechanisms of Development* 126(1–2). doi: <https://doi.org/10.1016/j.mod.2008.10.001>.

Rana, M.S., Christoffels, V.M. and Moorman, A.F.M. (2013). A molecular and genetic outline of cardiac morphogenesis. *Acta Physiologica* 207(4).

Rajagopal, S.K. et al. 2007. Spectrum of heart disease associated with murine and human GATA4 mutation. *Journal of Molecular and Cellular Cardiology*. doi: 10.1016/j.yjmcc.2007.06.004.

Readnower, R.D., Brainard, R.E., Hill, B.G. and Jones, S.P. 2012. Standardized bioenergetic profiling of adult mouse cardiomyocytes. *Physiological Genomics* 44(24). doi: 10.1152/physiolgenomics.00129.2012.

Ren, B., Chee, K.J., Kim, T.H. and Maniatis, T. 1999. PRDI-BF1/Blimp-1 repression is mediated by corepressors of the Groucho family of proteins. *Genes and Development* 13(1). doi: 10.1101/gad.13.1.125.

Ren, Y. et al. 2011. Small molecule Wnt inhibitors enhance the efficiency of BMP-4-directed cardiac differentiation of human pluripotent stem cells. *Journal of Molecular and Cellular Cardiology* 51(3). doi: 10.1016/j.yjmcc.2011.04.012.

Ribeiro, A.J.S. et al. 2015. Contractility of Single cardiomyocytes differentiated from pluripotent stem cells depends on physiological shape and substrate stiffness. *Proceedings of the National Academy of Sciences of the United States of America* 112(41). doi: 10.1073/pnas.1508073112.

Riley, P., Anson-Cartwright, L. and Cross, J.C. (1998). The Hand1 bHLH transcription factor is essential for placentation and cardiac morphogenesis. *Nature Genetics* 18(3). doi:

<https://doi.org/10.1038/ng0398-271>.

Ribeiro, I. et al. 2007. Tbx2 and Tbx3 regulate the dynamics of cell proliferation during heart remodeling. *PLoS ONE*. doi: 10.1371/journal.pone.0000398.

Risebro, C.A. et al. 2009. Prox1 maintains muscle structure and growth in the developing heart. *Development* 136(3). doi: 10.1242/dev.030007.

Ritterhoff, J. et al. 2020. Metabolic remodeling promotes cardiac hypertrophy by directing glucose to aspartate biosynthesis. *Circulation Research*. doi: 10.1161/CIRCRESAHA.119.315483.

Robbe, Z.L. et al. 2022. CHD4 is recruited by GATA4 and NKX2-5 to repress noncardiac gene programs in the developing heart. *Genes & Development* 36(7–8), pp. 468–482. doi: 10.1101/gad.349154.121.

Robertson, E.J. et al. 2007. Blimp1 regulates development of the posterior forelimb, caudal pharyngeal arches, heart and sensory vibrissae in mice. *Development*. doi: 10.1242/dev.012047.

Rodriguez, T.A., Srinivas, S., Clements, M.P., Smith, J.C. and Beddington, R.S.P. 2005. Induction and migration of the anterior visceral endoderm is regulated by the extra-embryonic ectoderm. *Development* 132(11). doi: 10.1242/dev.01847.

Rojas, A., Kong, S.W., Agarwal, P., Gilliss, B., Pu, W.T. and Black, B.L. 2008. GATA4 Is a Direct Transcriptional Activator of Cyclin D2 and Cdk4 and Is Required for Cardiomyocyte Proliferation in Anterior Heart Field-Derived Myocardium. *Molecular and Cellular Biology*. doi: 10.1128/mcb.00717-08.

Romero-García, R. et al. 2020. Differential epigenetic regulation between the alternative promoters, PRDM1 α and PRDM1 β , of the tumour suppressor gene PRDM1 in human multiple myeloma cells. *Scientific Reports* 10(1). doi: 10.1038/s41598-020-72946-z.

Rony, I.K., Baten, A., Bloomfield, J.A., Islam, M.E., Billah, M.M. and Islam, K.D. (2015). Inducing pluripotency in vitro: Recent advances and highlights in induced pluripotent stem cells generation and pluripotency reprogramming. *Cell Proliferation* 48(2).

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölpfer, S., Krüger, M. and Stainier, D.Y.R. 2015. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature*. doi: 10.1038/nature14580.

Sadahiro, T. et al. 2018. Tbx6 Induces Nascent Mesoderm from Pluripotent Stem Cells and Temporally Controls Cardiac versus Somite Lineage Diversification. *Cell Stem Cell* 23(3). doi: 10.1016/j.stem.2018.07.001.

Saga, Y., Miyagawa-Tomita, S., Takagi, A., Kitajima, S., Miyazaki, J.I. and Inoue, T. (1999). MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development*.

- Salichs, E., Ledda, A., Mularoni, L., Albà, M.M. and De La Luna, S. 2009. Genome-Wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. *PLoS Genetics* 5(3). doi: 10.1371/journal.pgen.1000397.
- Saijoh, Y. and Hamada, H. (2020). Making the Right Loop for the heart. *Developmental Cell* 55(4).
- Sam, J., Mercer, E.J., Torregroza, I., Banks, K.M. and Evans, T. 2020. Specificity, redundancy and dosage thresholds among gata4/5/6 genes during zebrafish cardiogenesis. *Biology Open* 9(6). doi: 10.1242/bio.053611.
- Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Research* 39(21). doi: 10.1093/nar/gkr606.
- Schlesinger, J. et al. 2011. The cardiac transcription network modulated by gata4, mef2a, nkx2.5, srf, histone modifications, and microRNAs. *PLoS Genetics*. doi: 10.1371/journal.pgen.1001313.
- Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*. doi: 10.1038/nmeth.2089.
- Schram, K., de Girolamo, S., Madani, S., Munoz, D., Thong, F. and Sweeney, G. 2010. Leptin regulates MMP-2, TIMP-1 and collagen synthesis via p38 MAPK in HL-1 murine cardiomyocytes. *Cellular and Molecular Biology Letters* 15(4). doi: 10.2478/s11658-010-0027-z.
- Sebé-Pedrós, A. et al. 2013. Early evolution of the T-box transcription factor family. *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1309748110.
- Sedletcaia, A. and Evans, T. 2011. Heart chamber size in zebrafish is regulated redundantly by duplicated *tbx2* genes. *Developmental Dynamics* 240(6). doi: 10.1002/dvdy.22622.
- Seeger, T. et al. 2019. A premature termination codon mutation in MYBPC3 causes hypertrophic cardiomyopathy via chronic activation of nonsense-mediated decay. *Circulation* 139(6). doi: 10.1161/CIRCULATIONAHA.118.034624.
- Sehnert, A.J., Huq, A., Weinstein, B.M., Walker, C., Fishman, M. and Stainier, D.Y.R. 2002. Cardiac troponin T is essential in sarcomere assembly and cardiac contractility. *Nature Genetics* 31(1). doi: 10.1038/ng875.
- Seya, D., Ihara, D., Shirai, M., Kawamura, T., Watanabe, Y. and Nakagawa, O. (2021). A role of Hey2 transcription factor for right ventricle development through regulation of Tbx2-Mycn pathway during cardiac morphogenesis. *Development Growth and Differentiation* 63(1). doi: <https://doi.org/10.1111/dgd.12707>.
- Seo, B.R. et al. 2020. Collagen microarchitecture mechanically controls myofibroblast differentiation. *Proceedings of the National Academy of Sciences of the United States of America* 117(21). doi: 10.1073/pnas.1919394117.

- Sevim Bayrak, C., Zhang, P., Tristani-Firouzi, M., Gelb, B.D. and Itan, Y. 2020. De novo variants in exomes of congenital heart disease patients identify risk genes and pathways. *Genome Medicine*. doi: 10.1186/s13073-019-0709-8.
- Shaheen, R. et al. 2015. Positional mapping of PRKD1, NRP1 and PRDM1 as novel candidate disease genes in truncus arteriosus. *Journal of Medical Genetics* 52(5). doi: 10.1136/jmedgenet-2015-102992.
- Sharma, A. et al. 2020. GATA6 mutations in hiPSCs inform mechanisms for maldevelopment of the heart, pancreas, and diaphragm. *eLife* 9. doi: 10.7554/eLife.53278.
- Shi, Z.D. et al. 2017. Genome Editing in hPSCs Reveals GATA6 Haploinsufficiency and a Genetic Interaction with GATA4 in Human Pancreatic Development. *Cell Stem Cell*. doi: 10.1016/j.stem.2017.01.001.
- Shirai, M., Imanaka-Yoshida, K., Schneider, M.D., Schwartz, R.J. and Morisaki, T. 2009. T-box 2, a mediator of Bmp-Smad signaling, induced hyaluronan synthase 2 and Tgf β 2 expression and endocardial cushion formation. *Proceedings of the National Academy of Sciences of the United States of America* 106(44). doi: 10.1073/pnas.0900635106.
- Sim, X., Cardenas-Diaz, F.L., French, D.L. and Gadue, P. 2014. A doxycycline-inducible system for genetic correction of iPSC disease models. In: *Methods in Molecular Biology*. doi: 10.1007/7651_2014_179.
- Singh, M.K. et al. 2010. Gata4 and Gata5 cooperatively regulate cardiac myocyte proliferation in mice. *Journal of Biological Chemistry*. doi: 10.1074/jbc.M109.038539.
- Singh, R. et al. 2012. Tbx2 and Tbx3 induce atrioventricular myocardial development and endocardial cushion formation. *Cellular and Molecular Life Sciences* 69(8). doi: 10.1007/s00018-011-0884-2.
- Sissman, N.J. (1970). Developmental landmarks in cardiac morphogenesis: Comparative chronology. *The American Journal of Cardiology* 25(2). doi: [https://doi.org/10.1016/0002-9149\(70\)90575-8](https://doi.org/10.1016/0002-9149(70)90575-8).
- Smith, C.L., Baek, S.T., Sung, C.Y. and Tallquist, M.D. 2011. Epicardial-derived cell epithelial-to-mesenchymal transition and fate specification require PDGF receptor signaling. *Circulation Research* 108(12). doi: 10.1161/CIRCRESAHA.110.235531.
- Song, K. et al. 2012. Heart repair by reprogramming non-myocytes with cardiac transcription factors. *Nature* 485(7400). doi: 10.1038/nature11139.
- Später, D., Hansson, E.M., Zangi, L. and Chien, K.R. 2014. How to make a cardiomyocyte. *Development (Cambridge)*. doi: 10.1242/dev.091538.
- Srivastava, D., Thomas, T., Kirby, M.L., Brown, D. and Olson, E.N. (1997). Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nature Genetics*. doi: <https://doi.org/10.1038/ng0697-154>.

Stansfield, W.E., Andersen, N.M., Tang, R.H. and Selzman, C.H. 2009. Periostin Is a Novel Factor in Cardiac Remodeling After Experimental and Clinical Unloading of the Failing Heart. *Annals of Thoracic Surgery* 88(6). doi: 10.1016/j.athoracsur.2009.07.038.

Stansfield, W.E., Ranek, M., Pendse, A., Schisler, J.C., Wang, S., Pulinilkunnil, T. and Willis, M.S. 2014. The Pathophysiology of Cardiac Hypertrophy and Heart Failure. In: *Cellular and Molecular Pathobiology of Cardiovascular Disease*. doi: 10.1016/B978-0-12-405206-2.00004-1.

Steimle, J.D. and Moskowitz, I.P. 2017. TBX5: A Key Regulator of Heart Development. In: *Current Topics in Developmental Biology*. doi: 10.1016/bs.ctdb.2016.08.008.

Stennard, F.A., Costa, M.W., Lai, D., Biben, C., Furtado, M.B., Solloway, M.J., ... Harvey, R.P. (2005). 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(10). doi: <https://doi.org/10.1242/dev.01799>.

Stemmer, M., Thumberger, T., del Sol Keyer, M., Wittbrodt, J. and Mateo, J.L. 2015. CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS ONE*. doi: 10.1371/journal.pone.0124633.

Stephen, L.J., Fawkes, A.L., Verhoeve, A., Lemke, G. and Brown, A. 2007. A critical role for the EphA3 receptor tyrosine kinase in heart development. *Developmental Biology* 302(1). doi: 10.1016/j.ydbio.2006.08.058.

Stirnimann, C.U., Ptchelkine, D., Grimm, C. and Müller, C.W. 2010. Structural basis of TBX5-DNA recognition: The T-box domain in its DNA-bound and -unbound form. *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2010.04.052.

Su, S.-T., Ying, H.-Y., Chiu, Y.-K., Lin, F.-R., Chen, M.-Y. and Lin, K.-I. 2009. Involvement of Histone Demethylase LSD1 in Blimp-1-Mediated Gene Repression during Plasma Cell Differentiation. *Molecular and Cellular Biology* 29(6). doi: 10.1128/mcb.01158-08.

Sulo, G., Iglund, J., Vollset, S.E., Nygård, O., Ebbing, M., Sulo, E., ... Tell, G.S. (2016). Heart failure complicating acute myocardial infarction; burden and timing of occurrence: A nationwide analysis including 86 771 patients from the cardiovascular disease in norway (CVDNOR) project. *Journal of the American Heart Association* 5(1). doi: <https://doi.org/10.1161/JAHA.115.002667>.

Sun, N. et al. 2012. Patient-specific induced pluripotent stem cells as a model for familial dilated cardiomyopathy. *Science Translational Medicine* 4(130). doi: 10.1126/scitranslmed.3003552.

Sun, Q., Liu, S., Liu, K. and Jiao, K. 2018. Role of semaphorin signaling during cardiovascular development. *Journal of the American Heart Association* 7(11). doi: 10.1161/JAHA.118.008853.

Sun, Y., Liang, X. and Evans, S.M. 2007. Islet1 Progenitors in Developing and Postnatal Heart. *Advances in Developmental Biology* 18. doi: 10.1016/S1574-3349(07)18006-6.

- Suzuki, E., Evans, T., Lowry, J., Truong, L., Bell, D.W., Testa, J.R. and Walsh, K. 1996. The human GATA-6 gene: Structure, chromosomal location, and regulation of expression by tissue-specific and mitogen-responsive signals. *Genomics* 38(3). doi: 10.1006/geno.1996.0630.
- Suzuki, S. et al. 2014. A case of pancreatic agenesis and congenital heart defects with a novel GATA6 nonsense mutation: Evidence of haploinsufficiency due to nonsense-mediated mRNA decay. *American Journal of Medical Genetics, Part A*. doi: 10.1002/ajmg.a.36275.
- Szklarczyk, D. et al. 2019. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*. doi: 10.1093/nar/gky1131.
- Takaoka, K. and Hamada, H. 2012. Cell fate decisions and axis determination in the early mouse embryo. *Development* 139(1). doi: 10.1242/dev.060095.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* 131(5). doi: <https://doi.org/10.1016/j.cell.2007.11.019>.
- Takahashi, K. and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* 126(4). doi: <https://doi.org/10.1016/j.cell.2006.07.024>.
- Takeuchi, J.K. and Bruneau, B.G. 2009. Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. *Nature*. doi: 10.1038/nature08039.
- Tan, F.L., Moravec, C.S., Li, J., Apperson-Hansen, C., McCarthy, P.M., Young, J.B. and Bond, M. 2002. The gene expression fingerprint of human heart failure. *Proceedings of the National Academy of Sciences of the United States of America* 99(17). doi: 10.1073/pnas.162370099.
- Tanaka, A. et al. 2014. Endothelin-1 induces myofibrillar disarray and contractile vector variability in hypertrophic cardiomyopathy-induced pluripotent stem cell-derived cardiomyocytes. *Journal of the American Heart Association* 3(6). doi: 10.1161/JAHA.114.001263.
- Tanaka, M., Chen, Z., Bartunkova, S., Yamasaki, N. and Izumo, S. (1999). The cardiac homeobox gene *Csx/Nkx2.5* lies genetically upstream of multiple genes essential for heart development. *Development* 126(6). doi: <https://doi.org/10.1242/dev.126.6.1269>.
- Taylor, C.J., Ordóñez-Mena, J.M., Roalfe, A.K., Lay-Flurrie, S., Jones, N.R., Marshall, T. and Hobbs, F.D.R. 2019. Trends in survival after a diagnosis of heart failure in the United Kingdom 2000-2017: population based cohort study. *The BMJ*. doi: 10.1136/bmj.l223.
- The British Heart Foundation (2023). *British Heart Foundation (2023) Heart & Circulatory Disease Statistics, 2023 Compendium*.
- Ting, C.N., Olson, M.C., Barton, K.P. and Leiden, J.M. (1996). Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature* 384(6608). doi: <https://doi.org/10.1038/384474a0>.

Tomoya Sakamoto, K.B.S.W.Y.G.L.L.R.B.V.D.P.K. 2022. The nuclear receptor ERR cooperates with the cardiogenic factor GATA4 to orchestrate cardiomyocyte maturation. *Nature Communications*.

Towbin, H., Staehelin, T. and Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America* 76(9). doi: 10.1073/pnas.76.9.4350.

Towbin, J.A. and Jefferies, J.L. 2017. Cardiomyopathies due to left ventricular noncompaction, mitochondrial and storage diseases, and inborn errors of metabolism. *Circulation Research* 121(7). doi: 10.1161/CIRCRESAHA.117.310987.

Tran, D.Q., Andersson, J., Wang, R., Ramsey, H., Unutmaz, D. and Shevach, E.M. 2009. GARP (LRRC32) is essential for the surface expression of latent TGF- β on platelets and activated FOXP3+ regulatory T cells. *Proceedings of the National Academy of Sciences of the United States of America* 106(32). doi: 10.1073/pnas.0901944106.

Triedman, J.K. and Newburger, J.W. 2016. Trends in Congenital Heart Disease. *Circulation*. doi: 10.1161/CIRCULATIONAHA.116.023544.

Troncoso, M.F. et al. 2021. VCAM-1 as a predictor biomarker in cardiovascular disease. *Biochimica et Biophysica Acta - Molecular Basis of Disease* 1867(9). doi: 10.1016/j.bbadis.2021.166170.

Tsai, F.Y., Keller, G., Kuo, F.C., Weiss, M., Chen, J., Rosenblatt, M., ... Orkin, S.H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* 371(6494). doi: <https://doi.org/10.1038/371221a0>.

Tsai, F.Y. and Orkin, S.H. (1997). Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood* 89(10). doi: https://doi.org/10.1182/blood.v89.10.3636.3636_3636_3643.

Tuladhar, R. et al. 2019. CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nature Communications* 10(1). doi: 10.1038/s41467-019-12028-5.

di Tullio, F., Schwarz, M., Zorgati, H., Mzoughi, S. and Guccione, E. 2022. The duality of PRDM proteins: epigenetic and structural perspectives. *FEBS Journal* 289(5). doi: 10.1111/febs.15844.

Tyser, R.C.V., Miranda, A.M.A., Chen, C.M., Davidson, S.M., Srinivas, S. and Riley, P.R. 2016. Calcium handling precedes cardiac differentiation to initiate the first heartbeat. *eLife* 5(OCTOBER2016). doi: 10.7554/eLife.17113.

Tyser, R.C.V. and Srinivas, S. 2020. The first heartbeat—origin of cardiac contractile activity. *Cold Spring Harbor Perspectives in Biology* 12(7). doi: 10.1101/cshperspect.a037135.

Tyser, R.C.V., Ibarra-Soria, X., McDole, K., Jayaram, S.A., Godwin, J., Brand, T.A.H.V. Den, ... Srinivas, S. (2021). Characterization of a common progenitor pool of the epicardium and

myocardium. *Science* 371(6533). doi: <https://doi.org/10.1126/science.abb2986>.

Tzahor, E. and Poss, K.D. (2017). Cardiac regeneration strategies: Staying young at heart. *Science* 356(6342).

Van Berlo, J.H., Elrod, J.W., Van Den Hoogenhof, M.M.G., York, A.J., Aronow, B.J., Duncan, S.A. and Molkenkin, J.D. (2010). The transcription factor GATA-6 regulates pathological cardiac hypertrophy. *Circulation Research* 107(8). doi: <https://doi.org/10.1161/CIRCRESAHA.110.220764>.

Vance, K.W., Carreira, S., Brosch, G. and Goding, C.M. 2005. Tbx2 is overexpressed and plays an important role in maintaining proliferation and suppression of senescence in melanomas. *Cancer Research* 65(6). doi: 10.1158/0008-5472.CAN-04-3045.

Van Der Linde, D., Konings, E.E.M., Slager, M.A., Witsenburg, M., Helbing, W.A., Takkenberg, J.J.M. and Roos-Hesselink, J.W. (2011). Birth prevalence of congenital heart disease worldwide: A systematic review and meta-analysis. *Journal of the American College of Cardiology*. doi: <https://doi.org/10.1016/j.jacc.2011.08.025>.

Vervoort, M., Meulemeester, D., Béhague, J. and Kerner, P. 2016. Evolution of Prdm genes in animals: Insights from comparative genomics. *Molecular Biology and Evolution*. doi: 10.1093/molbev/msv260.

Vigil-Garcia, M. et al. 2021. Gene expression profiling of hypertrophic cardiomyocytes identifies new players in pathological remodelling. *Cardiovascular Research* 117(6). doi: 10.1093/cvr/cvaa233.

Vincent, S.D. and Buckingham, M.E. 2010. How to make a heart. The origin and regulation of cardiac progenitor cells. In: *Current Topics in Developmental Biology*. doi: 10.1016/S0070-2153(10)90001-X.

Vincent, S.D., Mayeuf-Louchart, A., Watanabe, Y., Brzezinski, J.A., Miyagawa-Tomita, S., Kelly, R.G. and Buckingham, M. 2014. Prdm1 functions in the mesoderm of the second heart field, where it interacts genetically with Tbx1, during outflow tract morphogenesis in the mouse embryo. *Human molecular genetics*. doi: 10.1093/hmg/ddu232.

Vincentz, J.W., Barnes, R.M., Firulli, B.A., Conway, S.J. and Firulli, A.B. 2008. Cooperative interaction of Nkx2.5 and Mef2c transcription factors during heart development. *Developmental Dynamics* 237(12). doi: 10.1002/dvdy.21803.

Van Vliet, P., Wu, S.M., Zaffran, S. and Pucéat, M. 2012. Early cardiac development: A view from stem cells to embryos. *Cardiovascular Research* 96(3). doi: 10.1093/cvr/cvs270.

Virágh, S. and Challice, C.E. (1973). Origin and differentiation of cardiac muscle cells in the mouse. *Journal of Ultrastructure Research* 42(1–2). doi: [https://doi.org/10.1016/S0022-5320\(73\)80002-4](https://doi.org/10.1016/S0022-5320(73)80002-4).

Volpato, V. and Webber, C. 2020. Addressing variability in iPSC-derived models of human disease: Guidelines to promote reproducibility. *DMM Disease Models and Mechanisms* 13(1). doi: 10.1242/dmm.042317.

- Vrzalikova, K. et al. 2012. Hypomethylation and over-expression of the beta isoform of BLIMP1 is induced by Epstein-Barr virus infection of B cells; potential implications for the pathogenesis of EBV-associated lymphomas. *Pathogens* 1(2). doi: 10.3390/pathogens1020083.
- Waldo, K.L., Kumiski, D.H., Wallis, K.T., Stadt, H.A., Hutson, M.R., Platt, D.H. and Kirby, M.L. (2001). Conotruncal myocardium arises from a secondary heart field. *Development* 128(16). doi: <https://doi.org/10.1242/dev.128.16.3179>.
- Wang, E. et al. 2013. Identification of Functional Mutations in GATA4 in Patients with Congenital Heart Disease. *PLoS ONE*. doi: 10.1371/journal.pone.0062138.
- Wang, J. et al. 2012. Novel GATA6 mutations associated with congenital ventricular septal defect or tetralogy of Fallot. *DNA and Cell Biology* 31(11). doi: 10.1089/dna.2012.1814.
- Wang, T., Tian, J. and Jin, Y. 2021a. VCAM1 expression in the myocardium is associated with the risk of heart failure and immune cell infiltration in myocardium. *Scientific Reports* 11(1). doi: 10.1038/s41598-021-98998-3.
- Wang, W. et al. 2021b. A histidine cluster determines YY1-compartmentalized coactivators and chromatin elements in phase-separated super-enhancers. *bioRxiv*.
- Watanabe, M. and Whitman, M. 1999. FAST-1 is a key maternal effector of mesoderm inducers in the early *Xenopus* embryo. *Development* 126(24). doi: 10.1242/dev.126.24.5621.
- Watt, A.J., Battle, M.A., Li, J. and Duncan, S.A. 2004. GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 101(34). doi: 10.1073/pnas.0400752101.
- Weidgang, C.E. et al. 2013. TBX3 directs cell-fate decision toward mesendoderm. *Stem Cell Reports*. doi: 10.1016/j.stemcr.2013.08.002.
- Whitcomb, J., Gharibeh, L. and Nemer, M. 2020. From embryogenesis to adulthood: Critical role for GATA factors in heart development and function. *IUBMB Life* 72(1). doi: 10.1002/iub.2163.
- Wiesinger, A. et al. 2022. A single cell transcriptional roadmap of human pacemaker cell differentiation. *eLife* 11. doi: 10.7554/eLife.76781.
- Wilczewski, C.M. et al. 2018. CHD4 and the NuRD complex directly control cardiac sarcomere formation. *Proceedings of the National Academy of Sciences of the United States of America* 115(26). doi: 10.1073/pnas.1722219115.
- Wilkins, B.J., De Windt, L.J., Bueno, O.F., Braz, J.C., Glascock, B.J., Kimball, T.F. and Molkentin, J.D. 2002. Targeted Disruption of NFATc3, but Not NFATc4, Reveals an Intrinsic Defect in Calcineurin-Mediated Cardiac Hypertrophic Growth. *Molecular and Cellular Biology* 22(21). doi: 10.1128/mcb.22.21.7603-7613.2002.
- Wilkinson, D.G., Bhatt, S. and Herrmann, B.G. 1990. Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature* 343(6259). doi: 10.1038/343657a0.

- Willett, R.T. and Greene, L.A. (2011). Gata2 is required for migration and differentiation of retinorecipient neurons in the superior colliculus. *Journal of Neuroscience* 31(12). doi: <https://doi.org/10.1523/JNEUROSCI.4616-10.2011>.
- Wilm, T.P. and Solnica-Krezel, L. 2005. Essential roles of a zebrafish prdm1/blimp1 homolog in embryo patterning and organogenesis. *Development* 132(2). doi: 10.1242/dev.01572.
- De Windt, L.J. et al. 2001. Targeted inhibition of calcineurin attenuates cardiac hypertrophy in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 98(6). doi: 10.1073/pnas.031371998.
- Wunderlich, Z. and Mirny, L.A. 2009. Different gene regulation strategies revealed by analysis of binding motifs. *Trends in Genetics*. doi: 10.1016/j.tig.2009.08.003.
- Wyles, S.P., Hrstka, S.C., Reyes, S., Terzic, A., Olson, T.M. and Nelson, T.J. 2016. Pharmacological Modulation of Calcium Homeostasis in Familial Dilated Cardiomyopathy: An In Vitro Analysis From an RBM20 Patient-Derived iPSC Model. *Clinical and Translational Science* 9(3). doi: 10.1111/cts.12393.
- Xin, M., Davis, C.A., Molkenin, J.D., Lien, C.L., Duncan, S.A., Richardson, J.A. and Olson, E.N. 2006a. A threshold of GATA4 and GATA6 expression is required for cardiovascular development. *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0604604103.
- Xin, M., Davis, C.A., Molkenin, J.D., Lien, C.-L., Duncan, S.A., Richardson, J.A. and Olson, E.N. 2006b. A threshold of GATA4 and GATA6 expression is required for cardiovascular development. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.0604604103.
- Xu, H., Morishima, M., Wylie, J.N., Schwartz, R.J., Bruneau, B.G., Lindsay, E.A. and Baldini, A. 2004. Tbx1 has a dual role in the morphogenesis of the cardiac outflow tract. *Development*. doi: 10.1242/dev.01174.
- Xu, Y.J. et al. 2018. GATA6 loss-of-function mutation contributes to congenital bicuspid aortic valve. *Gene*. doi: 10.1016/j.gene.2018.04.018.
- Xuan, S. et al. 2012. Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis. *Journal of Clinical Investigation*. doi: 10.1172/JCI63352.
- Yamada, M., Revelli, J.P., Eichele, G., Barron, M. and Schwartz, R.J. 2000. Expression of chick Tbx-2, Tbx-3, and Tbx-5 genes during early heart development: Evidence for BMP2 induction of Tbx2. *Developmental Biology*. doi: 10.1006/dbio.2000.9927.
- Yamak, A., Latinkic, B. V., Dali, R., Temsah, R. and Nemer, M. 2014. Cyclin D2 is a GATA4 cofactor in cardiogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 111(4). doi: 10.1073/pnas.1312993111.
- Yamashiro, C. et al. 2016. Persistent requirement and alteration of the key targets of PRDM1 during primordial germ cell development in mice. *Biology of Reproduction* 94(1). doi: 10.1095/biolreprod.115.133256.

- Yang, H. et al. 2016. Dynamic Myofibrillar Remodeling in Live Cardiomyocytes under Static Stretch. *Scientific Reports* 6. doi: 10.1038/srep20674.
- Yang, X. et al. 2019. Fatty Acids Enhance the Maturation of Cardiomyocytes Derived from Human Pluripotent Stem Cells. *Stem Cell Reports*. doi: 10.1016/j.stemcr.2019.08.013.
- Yang, X., Pabon, L. and Murry, C.E. 2014. Engineering adolescence: Maturation of human pluripotent stem cell-derived cardiomyocytes. *Circulation Research* 114(3). doi: 10.1161/CIRCRESAHA.114.300558.
- Yang, Y.Q. et al. 2013. GATA4 Loss-of-Function Mutations Underlie Familial Tetralogy of Fallot. *Human Mutation*. doi: 10.1002/humu.22434.
- Yao, C.X. et al. 2012. Transcription factor GATA-6 recruits PPAR α to cooperatively activate Glut4 gene expression. *Journal of Molecular Biology* 415(1). doi: 10.1016/j.jmb.2011.11.011.
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T.L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*. doi: 10.1186/1471-2105-13-134.
- Yechikov, S., Copaciu, R., Gluck, J.M., Deng, W., Chiamvimonvat, N., Chan, J.W. and Lieu, D.K. 2016. Same-Single-Cell Analysis of Pacemaker-Specific Markers in Human Induced Pluripotent Stem Cell-Derived Cardiomyocyte Subtypes Classified by Electrophysiology. *Stem Cells* 34(11). doi: 10.1002/stem.2466.
- Yin, F. and Herring, B.P. 2005. GATA-6 can act as a positive or negative regulator of smooth muscle-specific gene expression. *Journal of Biological Chemistry* 280(6). doi: 10.1074/jbc.M411585200.
- Yokouchi-Konishi, T. et al. 2019. Recurrent Congenital Heart Diseases Among Neonates Born to Mothers with Congenital Heart Diseases. *Pediatric Cardiology*. doi: 10.1007/s00246-019-02083-6.
- Yoon, C.H. et al. 2018. Gata6 in pluripotent stem cells enhance the potential to differentiate into cardiomyocytes. *BMB Reports* 51(2). doi: 10.5483/BMBRep.2018.51.2.176.
- Yoshida, T., Vivatbutsiri, P., Morriss-Kay, G., Saga, Y. and Iseki, S. (2008). Cell lineage in mammalian craniofacial mesenchyme. *Mechanisms of Development*. doi: <https://doi.org/10.1016/j.mod.2008.06.007>.
- Yu, J., Angelin-Duclos, C., Greenwood, J., Liao, J. and Calame, K. 2000. Transcriptional Repression by Blimp-1 (PRDI-BF1) Involves Recruitment of Histone Deacetylase. *Molecular and Cellular Biology* 20(7). doi: 10.1128/mcb.20.7.2592-2603.2000.
- Zaidi, S. et al. 2013. De novo mutations in histone-modifying genes in congenital heart disease. *Nature*. doi: 10.1038/nature12141.
- Zeisberg, E.M., Ma, Q., Juraszek, A.L., Moses, K., Schwartz, R.J., Izumo, S. and Pu, W.T. 2005. Morphogenesis of the right ventricle requires myocardial expression of Gata4. *Journal of Clinical Investigation*. doi: 10.1172/JCI23769.

- Zhai, J., Xiao, Z., Wang, Y. and Wang, H. 2022. Human embryonic development: from peri-implantation to gastrulation. *Trends in Cell Biology* 32(1). doi: 10.1016/j.tcb.2021.07.008.
- Zhang, G.Q., Wei, H., Lu, J., Wong, P. and Shim, W. 2013. Identification and Characterization of Calcium Sparks in Cardiomyocytes Derived from Human Induced Pluripotent Stem Cells. *PLoS ONE* 8(2). doi: 10.1371/journal.pone.0055266.
- Zhang, T.-N. et al. 2021. Environmental Risk Factors and Congenital Heart Disease: An Umbrella Review of 165 Systematic Reviews and Meta-Analyses With More Than 120 Million Participants. *Frontiers in Cardiovascular Medicine* 8. doi: 10.3389/fcvm.2021.640729.
- Zhang, Q., Carlin, D., Zhu, F., Cattaneo, P., Ideker, T., Evans, S.M., ... Chi, N.C. (2021). Unveiling Complexity and Multipotentiality of Early Heart Fields. *Circulation Research* 129(4). doi: <https://doi.org/10.1161/CIRCRESAHA.121.318943>.
- Zhang, Z., Huynh, T. and Baldini, A. 2006. Mesodermal expression of Tbx1 is necessary and sufficient for pharyngeal arch and cardiac outflow tract development. *Development*. doi: 10.1242/dev.02539.
- Zhao, R., Watt, A.J., Battle, M.A., Li, J., Bondow, B.J. and Duncan, S.A. 2008. Loss of both GATA4 and GATA6 blocks cardiac myocyte differentiation and results in acardia in mice. *Developmental Biology*. doi: 10.1016/j.ydbio.2008.03.013.
- Zhao, R., Watt, A.J., Li, J., Luebke-Wheeler, J., Morrissey, E.E. and Duncan, S.A. 2005. GATA6 Is Essential for Embryonic Development of the Liver but Dispensable for Early Heart Formation. *Molecular and Cellular Biology* 25(7). doi: 10.1128/mcb.25.7.2622-2631.2005.
- Zheng, G.F., Wei, D., Zhao, H., Zhou, N., Yang, Y.Q. and Liu, X.Y. 2012. A novel GATA6 mutation associated with congenital ventricular septal defect. *International Journal of Molecular Medicine* 29(6). doi: 10.3892/ijmm.2012.930.
- Zhou, W. et al. 2019. Induced Pluripotent Stem Cell–Derived Cardiomyocytes from a Patient with MYL2-R58Q-Mediated Apical Hypertrophic Cardiomyopathy Show Hypertrophy, Myofibrillar Disarray, and Calcium Perturbations. *Journal of Cardiovascular Translational Research* 12(5). doi: 10.1007/s12265-019-09873-6.
- Zhu, B., Zhang, M., Byrum, S.D., Tackett, A.J. and Davie, J.K. 2014. TBX2 blocks myogenesis and promotes proliferation in rhabdomyosarcoma cells. *International Journal of Cancer*. doi: 10.1002/ijc.28721.
- Zhu, J., Min, B., Hu-Li, J., Watson, C.J., Grinberg, A., Wang, Q., ... Paul, W.E. (2004). Conditional deletion of Gata3 shows its essential function in TH1-TH2 responses. *Nature Immunology* 5(11). doi: <https://doi.org/10.1038/ni1128>.
- Zhu, T. xiang et al. 2013. ECM-related gene expression profile in vascular smooth muscle cells from human saphenous vein and internal thoracic artery. *Journal of Cardiothoracic Surgery* 8(1). doi: 10.1186/1749-8090-8-155.
- Zimmer, A., Bagchi, A.K., Vinayak, K., Bello-Klein, A. and Singal, P.K. (2019). Innate immune response in the pathogenesis of heart failure in survivors of myocardial infarction. *American*

Journal of Physiology - Heart and Circulatory Physiology 316(3).

Zhao, R., Watt, A.J., Li, J., Luebke-Wheeler, J., Morrisey, E.E. and Duncan, S.A. (2005). GATA6 Is Essential for Embryonic Development of the Liver but Dispensable for Early Heart Formation. *Molecular and Cellular Biology* 25(7). doi: <https://doi.org/10.1128/mcb.25.7.2622-2631.2005>.

Zhou, Y., Lim, K.C., Onodera, K., Takahashi, S., Ohta, J., Minegishi, N., ... Engel, J.D. (1998). Rescue of the embryonic lethal hematopoietic defect reveals a critical role for GATA-2 in urogenital development. *EMBO Journal* 17(22). doi: <https://doi.org/10.1093/emboj/17.22.6689>.

Zhou, Y., Yamamoto, M. and Engel, J.D. (2000). GATA2 is required for the generation of V2 interneurons. *Development* 127(17). doi: <https://doi.org/10.1242/dev.127.17.3829>.

Supplementary Data

S1. Generating gain of function lines

Another aim of this PhD was to create gain of function (GoF) lines to allow control over GATA4, GATA6, PRDM1 and TBX2 expression in iPSC-CM differentiation. In otherwise normal WT cells with endogenous expression of these genes this was to be used to determine if over-expression of these genes can negatively or positively affect cardiogenesis. In null lines this could be used to define when and for how long the expression of these genes is necessary, and to conduct rescue experiments if necessary. This was attempted in the WT Rebl Pat line first, following the protocol published by (Sim et al. 2014). In brief, WT cells were lipofected with four plasmids, two containing zinc finger nucleases designed to induce a double strand break in the AAVS1, to stimulate homology direct repair using the integration constructs contained in the other two plasmids. The gene to be inserted into the locus was put under control of a tetracycline response element (TRE), this construct also contained a puromycin resistance gene, and was surrounded by homology arms specific to the AAVS1 locus. The final plasmid contained the TET activator sequence and a neomycin resistance gene (rtTA). All plasmids used are detailed in table S1 below. Following lipofection cells positive for both constructs were selected using puromycin followed by neomycin treatment.

PCR was used to screen the cells and determine if integration had occurred at the AAVS1 locus, the primers used for this are shown in table S2. A number of cells were identified that were negative for the WT amplicon of 500bp (figure S1a) indicating that the AAVS1 locus has been disrupted and that integration had likely taken place. These likely positive clones also showed a positive band for the integration of the TRE-gene construct with an amplicon of ~1,380bp (figure S1b). However, integration of the TET activator construct was not detected despite cell growth in media supplemented with Neomycin. This may be due to integration of the plasmid at another site, or the selection method used. Neomycin selection takes 8-10 days and requires that cell numbers are low. This is not that compatible with iPSC culture, where cells can grow rapidly.

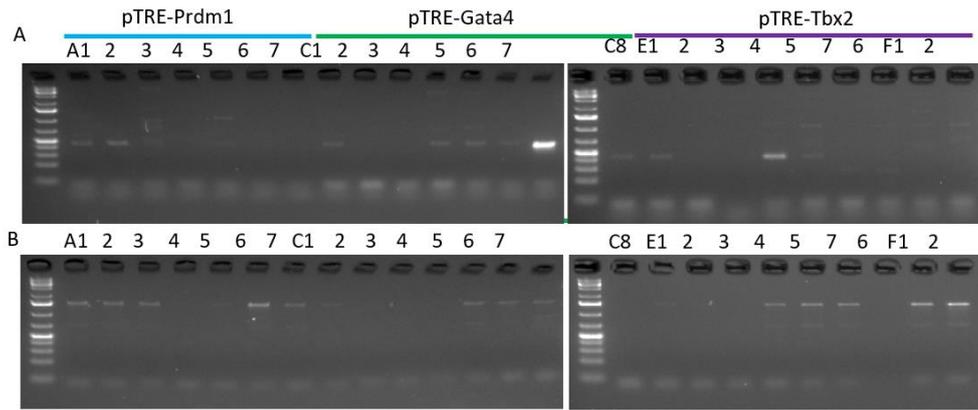


Figure S1. PCR based screening identifies A4, A6, and A7 as potentially double positive clones for pTRE-*PRDM1* integration

(A) 500bp amplicons indicate the presence of a WT AAVS1 amplification, therefore in clones with integration in both alleles this band should be absent. (B) 1,380bp amplicons indicate that the pTRE-gene of interest construct has been integrated into the AAVS1 locus.

The ladders two brightest bands show a 1,500bp marker, and 500bp marker (descending).

Table S1. A summary of the plasmids used to generate inducible cell lines.

Plasmids 60915 and 60916 encode zinc finger nucleases that are designed to induce a double strand break in the AAVS1 locus. Plasmid 22074, is the vector plasmid into which our genes of interest have been inserted, in place of EGFP under the control of the pTRE promoter, and these are also listed below. Plasmid 60431 encodes the tet activator sequence, which when bound to doxycycline can activate the pTRE promoter allowing controlled transgene expression. The donor plasmids containing the genes of interest are shown in bold, these genes were cloned into the pTRE vector by Dr. Pavel Kirilenko (PK).

Plasmid Name	Gene	Supplier	Catalogue #	Reference
RGS-6xHis- <i>BLIMP1/PRDM1</i> -pcDNA3.1	<i>PRDM1</i>	Addgene	52518	RGS-6xHis-BLIMP1-pcDNA3.1- was a gift from Adam Antebi (Addgene plasmid # 52518)
pBabe 3xFLAG- wt <i>GATA6</i> -3xAU1 puro	<i>GATA6</i>	Addgene	72607	pBabe 3XFLAG-wt <i>GATA6</i> -3XAU1 puro was a gift from Kevin Janes (Addgene plasmid # 72607)
<i>GATA4</i> in pCR4- TOPO Vector	<i>GATA4</i>	Source Bioscience	IRATp970E02121D	-
<i>TBX2</i> (human) in pOTB7 Vector	<i>TBX2</i>	Source Bioscience	IRAU p969B09104D	-
PGK-AAVS1ZFNR	ZFNR	Addgene	60915	PGK-AAVS1ZFNR was a gift from Paul Gadue (Addgene plasmid # 60915 ; http://n2t.net/addgene:60915 ; RRID:Addgene_60915)
PGK-AAVS1ZFNL	ZFNL	Addgene	60916	PGK-AAVS1ZFNL was a gift from Paul Gadue (Addgene plasmid # 60916 ; http://n2t.net/addgene:60916 ; RRID:Addgene_60916)
pTRE-TIGHT- EGFP-donor fw copy	EGFP	Addgene	22074	pTRE-TIGHT-EGFP-donor fw copy was a gift from Rudolf Jaenisch (Addgene plasmid # 22074 ; http://n2t.net/addgene:22074 ; RRID:Addgene_22074)
AAVS1-SA-2A- Neo-CAG-RTTA	rtTA3	Addgene	60431	AAVS1-SA-2A-NEO-CAG-RTTA3 was a gift from Paul Gadue (Addgene plasmid # 60431 ; http://n2t.net/addgene:60431 ; RRID:Addgene_60431)

pTRE-TIGHT- 3xFLAG-GATA6- 3xAU1-donor fw copy	<i>GATA6</i>	-	-	Dr. Pavel Kirilenko, Cardiff University
pTRE-TIGHT- <i>TBX2</i> HA-donor fw copy	<i>TBX2</i>	-	-	Dr. Pavel Kirilenko, Cardiff University
pTRE-TIGHT-- 6xHis- <i>BLIMP1/PRDM1</i> - donor fw copy	<i>BLIMP1/PRDM1</i>	-	-	Dr. Pavel Kirilenko, Cardiff University
pTRE-TIGHT- GATA4 HA- donor fw copy	<i>GATA4</i>	-	-	Dr. Pavel Kirilenko, Cardiff University

Table S2. A list of primers used to assess plasmid integration into the AAVS1 locus.

These primer sequences are taken from Sim *et al.* 2014, where the expected results are explained in further detail.

Primer ID	Forward Primer	Primer ID	Reverse Primer	Amplicon Length (bp)
AAVS1 WT F	CCCCTATGTCCACTTCAGGA	AAVS1 WT R	CAGCTCAGGTTCTGGGAGAG	441
AAVS1 CAG F	GAGCATCTGACTTCTGGCTAATA	AAVS1 CAG R	GAAGGATGCAGGACGAGAAA	500 for WT
AAVS1 TRE F	GCAATAGCATCACAAATTTAC	AAVS1 TRE R	GAAGGATGCAGGACGAGAAA	-
Neo F	GAAGGCGATAGAAGGCGATG	Neo R	GCTTGCCGAATATCATGGTG	204
rtTA3 F	GCTGTTTCTCCAGGCCAC	rtTA3 R	CAAGACTTTCTGCGGAACAACG	376

S2 – Maturation media testing: Glucose is essential for survival in a high fatty acid environment

With the view of creating more mature CMs that may provide more physiological results the elements of a number of published maturation medias were trialled. It has been noted that the inclusion of fatty acids as an energy source encourages maturation (Horikoshi et al. 2019; Yang et al. 2019). Others have shown that inclusion of T_3 , and a cortisol mimic dexamethasone in cell culture media can encourage maturation (Huang et al. 2020). Relevant literature also suggests that the inclusion of glucose can inhibit maturation (Nakano et al. 2017). Therefore, the initial maturation media combined these aspects with glucose free media, supplemented with fatty acids, T_3 , dexamethasone, and insulin, henceforth referred to as high fatty acid-no glucose media. The other medias formulations trialled are shown in table S3.

High fatty acid-no glucose media proved to be quickly toxic to cells, after just 24 hours cells can be seen to accumulate lipid droplets, and the CMs begins to dissociate from the cell surface cells (figure S2). Within 5 days of culture the number of cell present was vastly reduced such that the culture was no longer viable. To examine if this was due to the high concentration of fatty acids used this was reduced, but this still proved to be highly toxic to the cells. However, addition of a reduced amount of glucose to the cultures seemed to prevent cell death, and cultures can be viably maintained in high fatty acid-low glucose media long term. These results demonstrate that glucose seems to be necessary for cell survival in a high fatty acid environment.

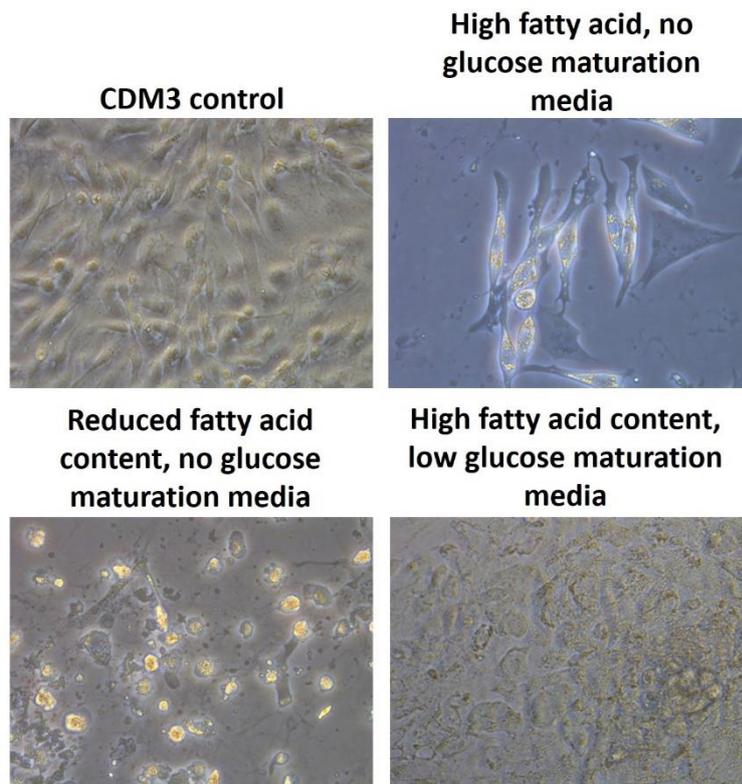


Figure S2. Glucose is necessary for cell survival in a high fatty acid environment.

The panel above shows cells maintained in different medias for ~6 days. Details of the media content are given in tableS3. Cells maintained in CDM3 are shown for comparison. Pictures are taken at 5x.

Table S3. A selection of the maturation media formulations tested

All maturation medias were made up in a base of RPMI 1640 with or without glucose.

Maturation Media 1 High fatty acid content, no glucose - fatty acid media		Reduced fatty acid content, no glucose - maturation media		High fatty acid, low glucose maturation media	
Sugars		Sugars		Sugars	
Glucose	0 mM	Glucose	0 mM	Glucose	2.75 mM
Lactate	5 mM	Lactate	5 mM	Lactate	5 mM
Fatty Acids		Fatty Acids		Fatty Acids	
Oleic Acid-BSA	33.28 μ M	Oleic Acid-BSA	16.64 μ M	Oleic Acid-BSA	33.28 μ M
Linoleic Acid-BSA	33.52 μ M	Linoleic Acid-BSA	16.76 μ M	Linoleic Acid-BSA	33.52 μ M
Palmitic Acid-BSA	0 μ M	Palmitic Acid-BSA	0 μ M	Palmitic Acid-BSA	0 μ M
Hormones/Growth Factors		Hormones/Growth Factors		Hormones/Growth Factors	
T3	100 nM	T3	100nM	T3	100nM
Dexamethsone	1 μ M	Dexamethasone	1 μ M	Dexamethasone	1 μ M
Insulin	1.72 μ M	Insulin	1.72 μ M	Insulin	1.72 μ M
Other Factors		Other Factors		Other Factors	
Transferrin	5.5 μ g/mL	Transferrin	5.5 μ g/ml	Transferrin	5.5 μ g/ml
Selenite	5 μ g/mL	Selenite	5 μ g/ml	Selenite	5 μ g/ml
BSA-FA free	500 μ g/mL	BSA-FA free	500 μ g/ml	BSA-FA free	500 μ g/ml
L-carnitine	120 μ M	L-carnitine	120 μ M	L-carnitine	120 μ M

S2. Further optimisation of maturation media

To further dissect the effect of each component in the medias tested, another experiment was carried out using the high fatty acid-low glucose maturation media (High-FA, low glucose), or CDM3 containing just thyroid hormone (T_3), or T_3 and dexamethasone. The cells were maintained in this media from day 25 to day 32 before the cells were fixed and stained for TNNT2 to identify CMs. Aspects of a more mature CM phenotype such as increased cell area, a larger aspect ratio, and better alignment of sarcomeres were assessed. This experiment was only conducted once, thus this data is preliminary. No significant difference in aspect ratio was observed. However, an increase in cell area was observed for cells treated with T_3 alone and high fatty acid – low glucose media. This may perhaps also be true for the cells treated with T_3 and Dexamethasone but fewer cells were analysed for the sample. CM organisation was analysed using the classification system used throughout (see section 2.5.8). Treatment with T_3 , T_3 and dexamethasone, or high fatty acid – low glucose media lead to an increase in the number of cells belonging to class IV the most disorganised class. From these limited results it is not possible to draw definitive conclusions, but these results give a hint that these maturation medias may have positive effects on some aspects of cell condition, but negative effects on others.

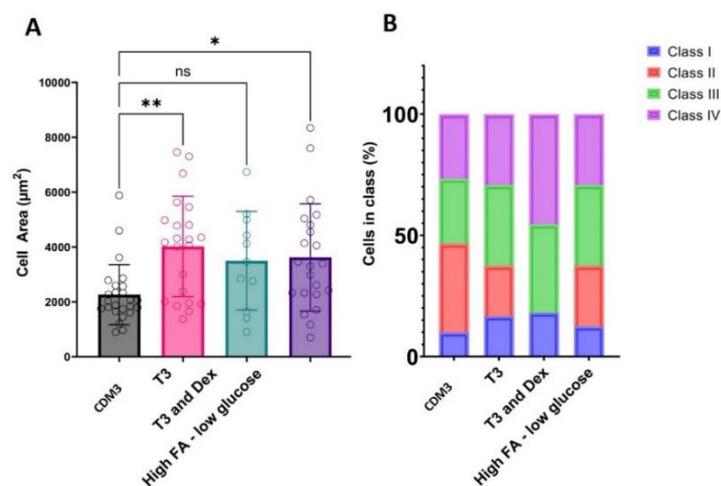


Figure S3. Maturation media has mixed effects on CM phenotype

(A) Measurement of CM area in day 32 cells maintained in different medias from days 25-32 of differentiation. (CDM3 n = 30, T_3 = 24, T_3 and Dexamethasone n = 12, and High fatty acid – low glucose media n = 24). (B) Classification of those cells based on the criteria described previously in figure S1. Only 1 biological repeat was conducted. Statistical test used = one-way ANOVA. P-value = * \leq 0.05, ** \leq 0.01.

S4. Optimisation of 3D culture conditions: A comparison of beating and non-beating heart forming organoids (HFOs)

Differentiation of iPSCs to CMs in a 3D format was considered as another method for analysing the iPSC lines generated in this thesis, as this differentiation format has been reported to generate cell cultures with a higher level of organisation and diversity of cells (Drakhlis et al., 2021). Before this could be implemented for any mutant lines the application of these protocols in our lab required optimisation as did the analysis of these cells by whole mount immunofluorescence. These methods are described in the main thesis methods and materials section. Some of the optimisation results are shown below. Firstly, using WT cells only, differentiation efficiency was assessed by recording beating versus non-beating organoids. Following 15 days of differentiation 50% (n =22) of the organoids created were beating. A selection of beating, and non-beating HFO's were subject to whole mount immunofluorescence and stained for TNNT2 and nuclear stain Hoechst 33342, followed by light sheet imaging to generate a 3D picture of the cells (figure S4). As expected, beating HFOs contained more TNNT2 positive cells, and there is more connectivity between these cells. Although, there were a small number of CMs present in the non-beating HFOs.

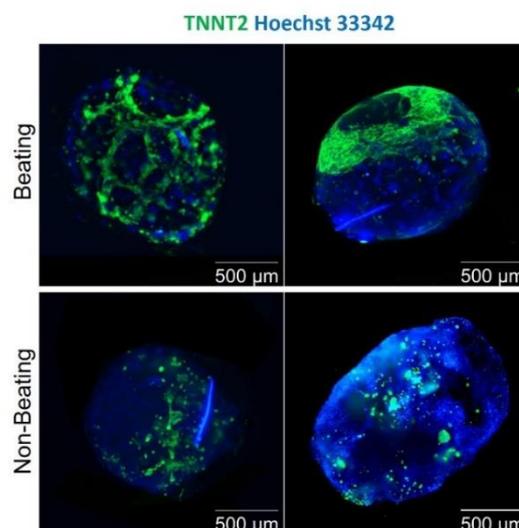


Figure S4. A comparison of the morphology between beating and non-beating HFOs

WT Rebl Pat iPSCs were differentiated until day 15. They were then processed for whole mount immunofluorescence for detection of TNNT2, and then counterstained with Hoechst 33342 to identify nuclei. The upper panel shows HFO's that had clearly detectable beating. The bottom panel shows two HFO's which contain a small amount of CMs, but no beating was detectable.

S5. Preliminary experiments suggest an essential cell autonomous role for GATA4

When combined with WT cells GATA4^{-/-} mESCs can contribute to functional myocardium in mice (Narita et al. 1997; Watt et al. 2004). This suggests that the WT cells surrounding the GATA4 null cells supply diffusible factors that are sufficient to drive cardiomyogenesis in the null cells. In this model the potential for a cell non-autonomous effect has been assessed by combining WT and GATA4^{-/-} iPS cells in a 1:1 ratio prior to differentiation using the CDM3 protocol. The cells were then dual stained for GATA4 and a CM marker, TNNT2 or MYBPC3, on day 12.

Firstly, it should be noted that GATA4 staining is present in non-CM cells in the WT with an average of 53.5% of cells present positive for GATA4 across the 2 experiments. However, higher GATA4 expression is generally seen in the CMs (figure S5d-eiii). In the GATA4^{-/-} mutant no GATA4 signal was seen in any of the cells present. In the WT differentiations (n = 2) just 11-33% of cells were TNNT2 positive so these happened to be low efficiency differentiations, staining for MYBPC3 was similar with 13-34% of cells showing positive staining. Cells positive for TNNT2 and MYBPC3 were all also GATA4 positive. In the GATA4^{-/-} mutant no TNNT2 or MYBPC3 positive cells were noted, and no GATA4 staining was observed (figure S5a-b).

In the WT:Gata4^{-/-} mixed wells 23.5-53% of cells stained positively for GATA4. This decrease suggests that some GATA4^{-/-} cells are present in the end population. If there was a non-cell autonomous effect it would be expected that an increase in TNNT2+/MYBPC3+ and GATA4- cells would be seen, however, none were detected.

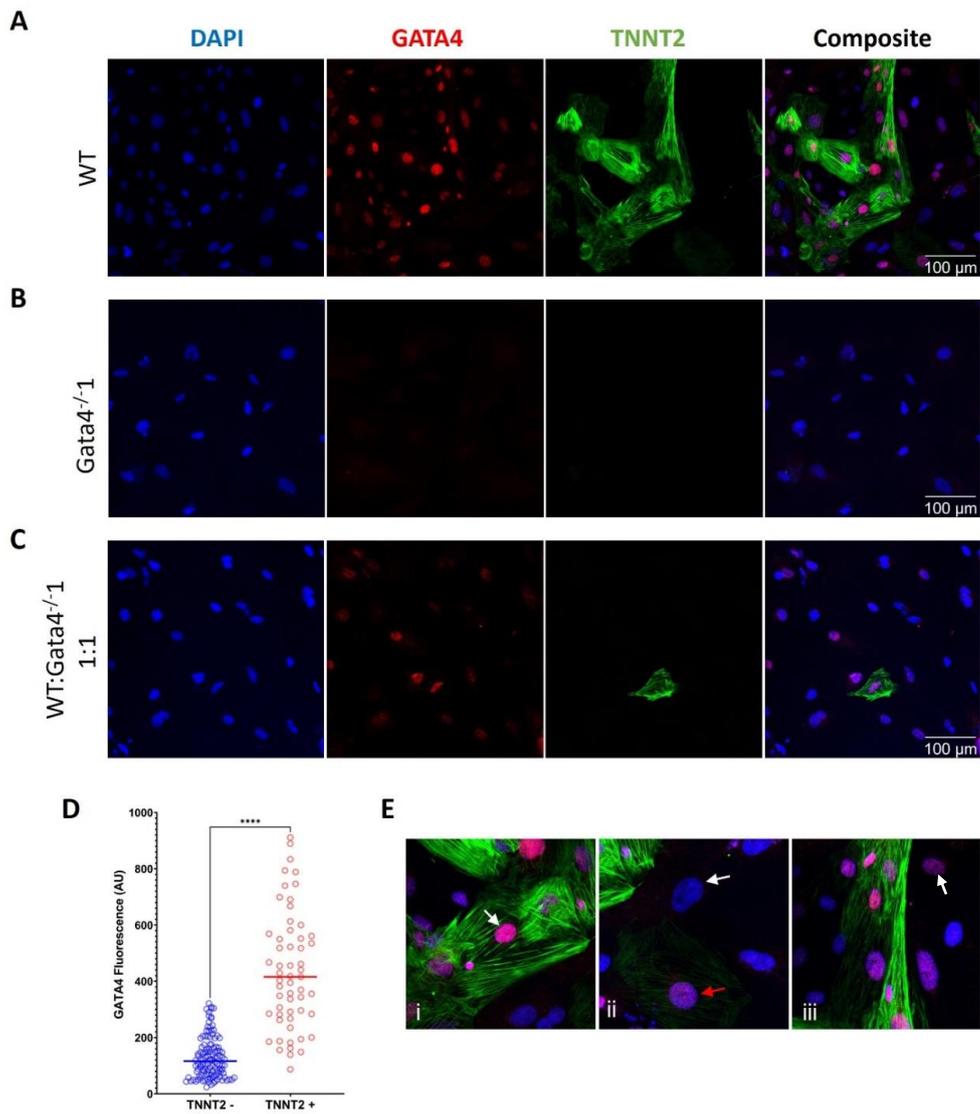


Figure S5. The addition of WT cells to GATA null iPS cell cultures does not improve their cardiomyogenic differentiation potential

(A) WT, (B) $GATA4^{-/-}$, and a 1:1 combination of (C) WT and $GATA4^{-/-}$ iPS cells were differentiated using the CDM3 protocol to day 12, and then stained for GATA4 (red) and TNNT2 (green). DAPI (blue) was used to counterstain and identify nuclei. (D) A plot of the values for GATA4 signal per nuclei for TNNT2 positive WT cells vs TNNT2 negative WT cells. For the TNNT2 negative category $n=127$ cells, across 2 biological repeats, and for TNNT2 positive category $n=57$ across, across 2 biological repeats. A student's t-test was used for statistical comparison. P-value $**** \leq 0.0001$. (E) Examples of the variation of GATA4 positivity in a WT culture; (i) a TNNT2 positive cell with high GATA4 expression (white arrow), (ii) a TNNT2 negative cell that is also GATA4 negative (white arrow) and a cell that is both lowly positive for TNNT2 and GATA4 (red arrow), and (iii) a TNNT2 negative cell with low Gata4 expression.

S.6 Examination of GATA6 nuclear localisation

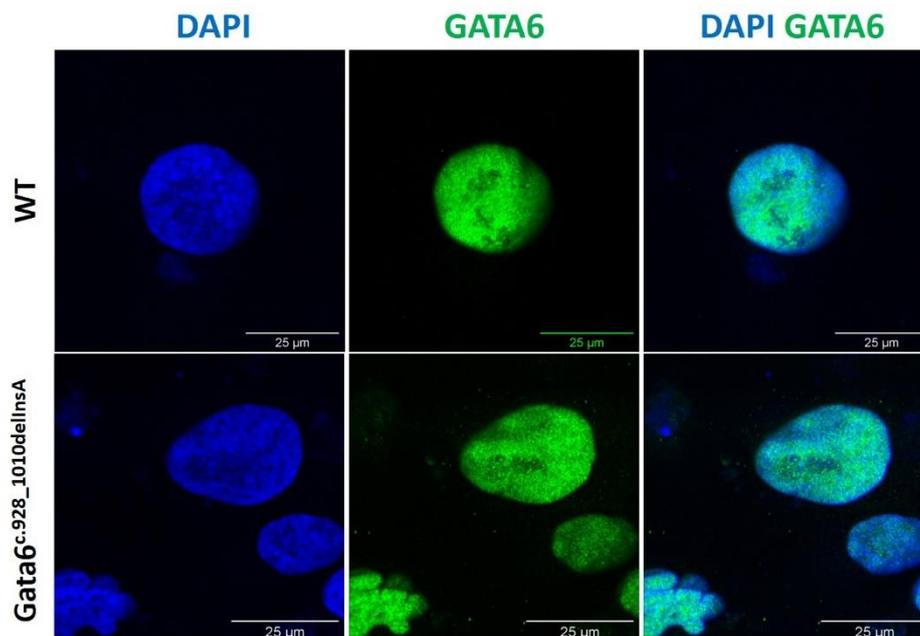


Figure S5. IF staining for GATA6 shows it to be homogenously distributed throughout the nucleus of WT and *GATA6*^{c.928_1010delInsA} CMs

WT and *GATA6*^{c.928_1010delInsA} mutant iPSCs were differentiated using the CDM3 protocol. Then stained for GATA6 (green), and counterstained with DAPI (blue) on day 32 of a CDM3 differentiation. The images shown were taken at 63x with an oil immersion lens, with 2x zoom. N=1.

S.7 Further examples of sarcomere deficiencies in GATA4 null TNNT2+ cells

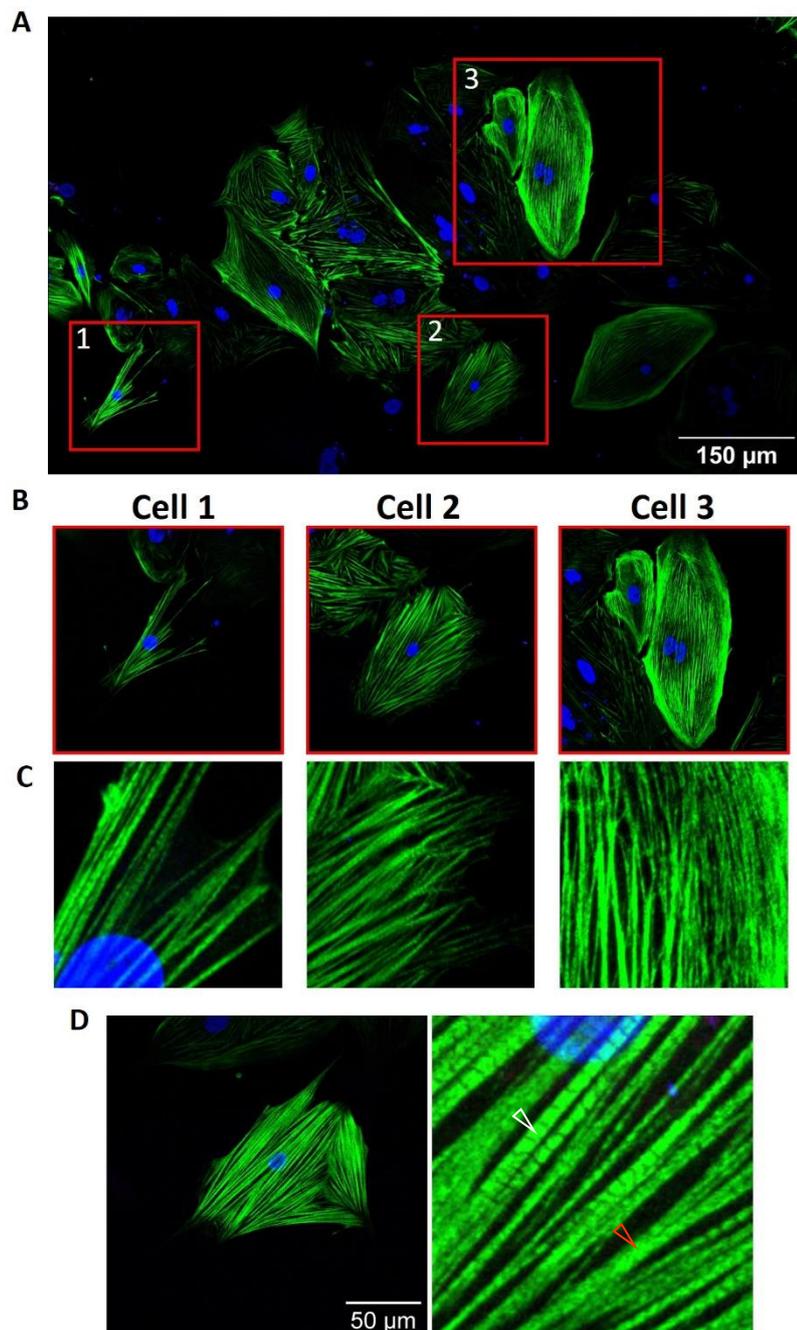


Figure S6. Most GATA4 null cells fail to form any sarcomeres

The pictures shown are of GATA4 null D32 TNNT2+ cells. Produced using the CDM3 differentiation protocol. (A) A tiled scan picture showing many cells, with different levels of TNNT2 expression. (B-C) Closer inspection of some of the brightest cells found in (A) reveals few sarcomeres are present. (D) This cell is the best example of a GATA4 null TNNT2 positive cell in which some fibrils have striations. White arrow = myofibril with striations. Red arrow = myofibril without striations.

S.8 Quality control for the RNA-sequencing data obtained from WT and TBX2 mutant cell lines

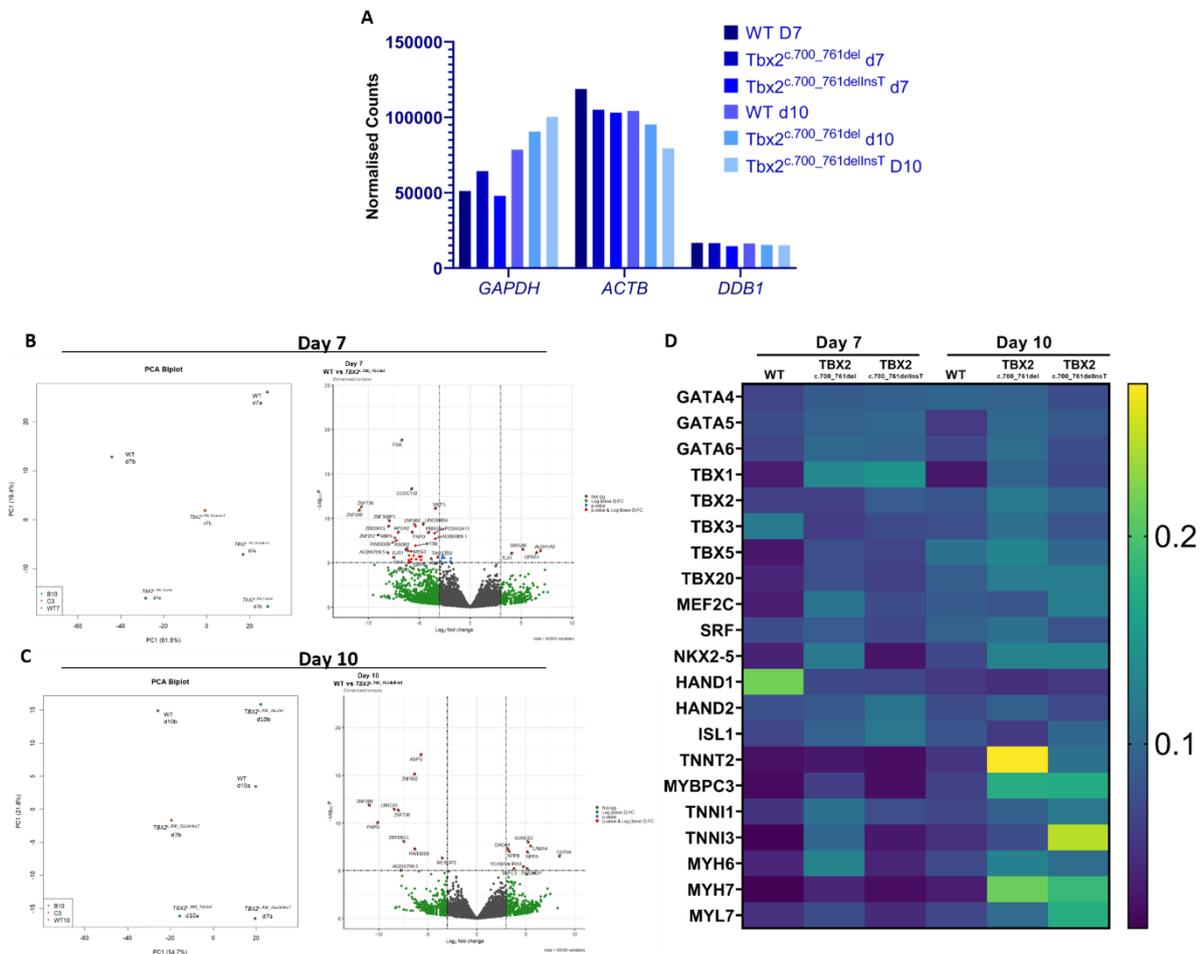


Figure S7. Quality control data for WT and TBX2 mutant iPSC-CM differentiation RNA samples

Cardiomyocytes were generated using the CDM3 differentiation protocol and samples taken at day 7 and day 10 were submitted for RNA-sequencing. (A) The mean average of the normalised counts for each sample have been plotted for comparison of housekeeping genes; *GAPDH*, *ACTB*, and *DDB1*. (B) A panel showing a principal component analysis and volcano plot for the WT, *TBX2*^{c.700_761del}, and *TBX2*^{c.700_761delInsT} produced at day 7. (C) A panel showing a principal component analysis and volcano plot for the WT, *TBX2*^{c.700_761del}, and *TBX2*^{c.700_761delInsT} produced at day 10. (D) A heatmap comparing the normalised expression of cardiac progenitor markers *GATA4-ISL1*, and cardiomyocyte markers (*TNNT2-MYL7*) between the WT and TBX2 mutants. N = 2 biological repeats.

The PCA and Volcano Plots shown were created by Antonios Tselingas, Cardiff University.

S.9 Quality control for the RNA-sequencing data obtained from WT and PRDM1^{-/-} cells

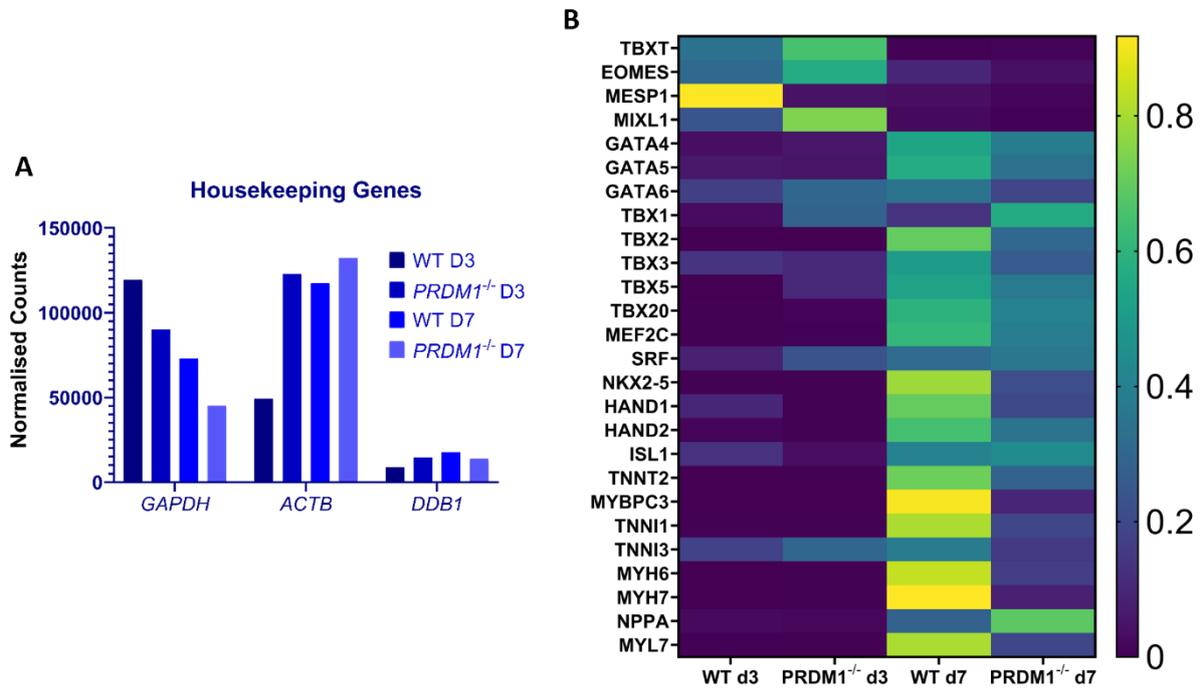


Figure S8. Quality control data for the RNA-seq data generated from WT and PRDM1^{-/-} mutant lines during iPSC-CM differentiation

Cardiomyocytes were generated using the CDM3 differentiation protocol and samples taken at day 3 and day 10 were submitted for RNA-sequencing. (A) The mean average of the normalised counts for each sample have been plotted for comparison of the expression of housekeeping genes; *GAPDH*, *ACTB*, and *DDB1*. (B) A heatmap comparing the normalised expression of mesodermal genes (*TBXT* and *EOMES*), cardiac mesoderm (*MESP1* and *MIXL1*), cardiac progenitor markers *GATA4-ISL1*, and cardiomyocyte markers (*TNNT2-MYL7*) between WT and *PRDM1^{-/-}* cells. N = 1 biological repeat.

These results have been generated using a single repeat for both lines, thus no further analysis has been completed.

S.10 Quality control for the RNA-sequencing data obtained from WT and *GATA6*^{c.928_1010delInsA}

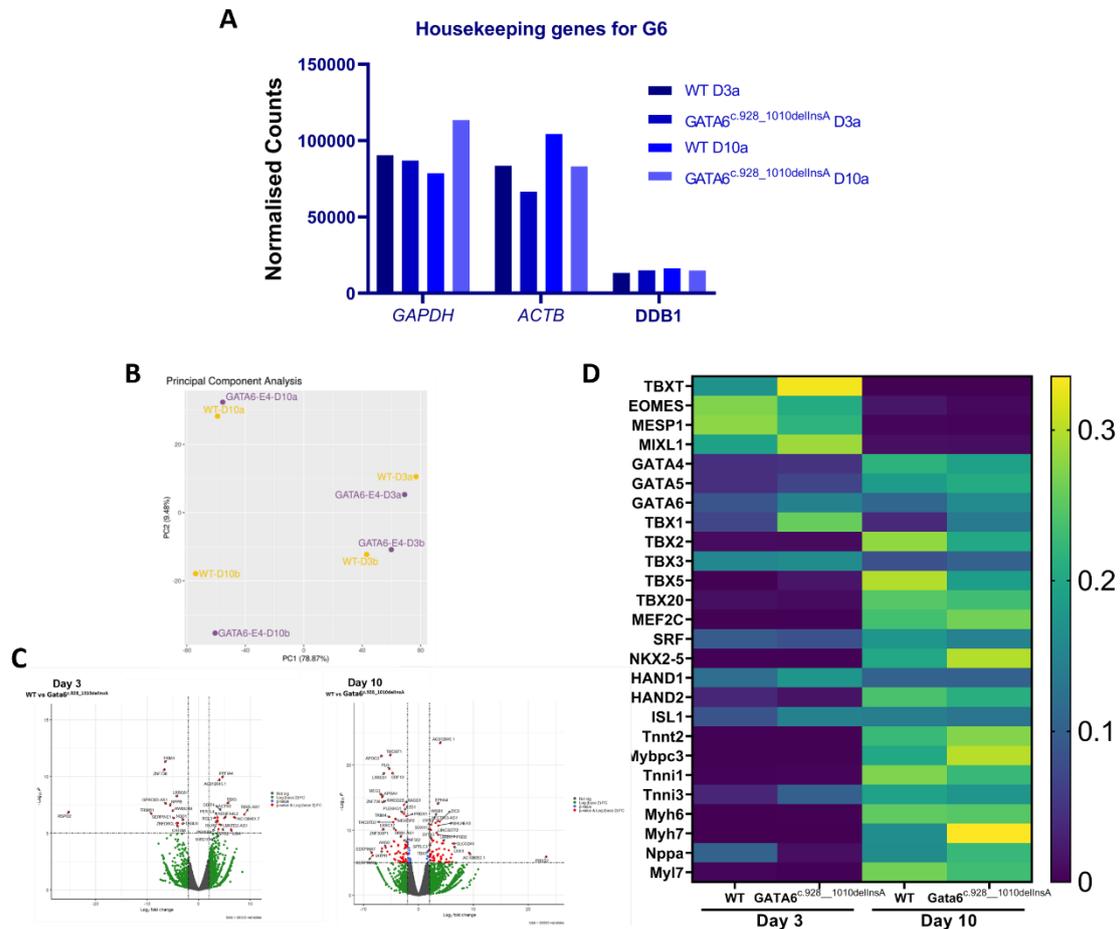


Figure S9. Quality control data for WT and *GATA6*^{c.928_1010delInsA} mutant iPSC-CM differentiation RNA samples

Cardiomyocytes were generated using the CDM3 differentiation protocol and samples taken at day 3 and day 10 were submitted for RNA-sequencing. (A) The mean average of the normalised counts for each sample have been plotted for comparison of housekeeping genes: *GAPDH*, *ACTB*, and *DDB1*. (B) A principal component analysis for the WT and *GATA6*^{c.928_1010delInsA} lines at day 3 and 10 of their differentiation. (C) A panel showing volcano plots comparing the transcriptional profiles of the WT and *GATA6*^{c.928_1010delInsA} cells at day 3 and day 10 of iPSC-CM differentiation. (D) A heatmap comparing the normalised expression of mesodermal genes (*TBXT* and *EOMES*), cardiac mesoderm (*MESP1* and *MIXL1*), cardiac progenitor markers *GATA4-ISL1*, and cardiomyocyte markers (*TNNT2-MYL7*) between WT and *GATA6*^{c.928_1010delInsA/-} cells. N = 2 biological repeat. **The PCA plot shown was created by Emma Moth, Cardiff University, and the Volcano Plots created by Antonios Tselingas, Cardiff University.**

S.11 A comparison between the number of TNNT2 and MYBPC3 positive cells produced from the WT and GATA4 mutant lines

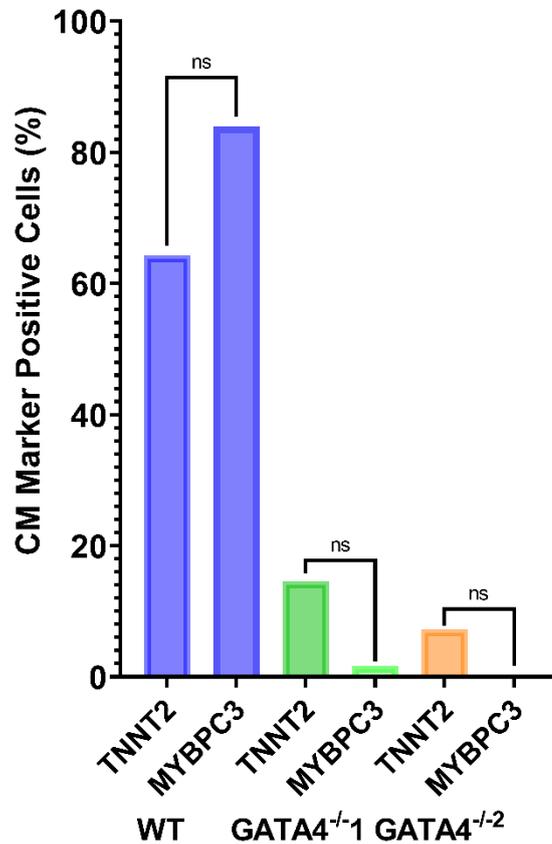


Figure S10. There is no significant difference in the proportion of TNNT2 positive and MYBPC3 positive cells produced within each cell line

(A) Immunofluorescent staining for TNNT2 and MYBPC3 was conducted on D32 WT, GATA4^{-/-1}, and GATA4^{-/-2} cells differentiated using the CDM3 protocol. Here, the proportion of cells staining positively for TNNT2 and MYBPC3 have been compared within each cell line. Statistical comparison was made using a one-way ANOVA. For WT TNNT2 n = 4, WT MYBPC3 n=3, GATA4^{-/-1} TNNT2 n=4, GATA4^{-/-1} MYBPC3 n=4, GATA4^{-/-2} TNNT2 n=4, and for GATA4^{-/-2} TNNT2 n=1.