University of Southampton

Faculty of Medicine

Clinical and Experimental Sciences

Investigating the role of NR2F2 in fibroblast during human heart development

by

Maria Gabriela de Oliveira Barbeta

ORCID ID: https://orcid.org/0000-0001-6848-8581

Thesis for the degree of Doctor of Philosophy

February 2023

University of Southampton

Abstract

Faculty of Medicine

Clinical and Experimental Sciences

Doctor of Philosophy

Investigating the role of NR2F2 in fibroblast during human heart development

by

Maria Gabriela de Oliveira Barbeta

Congenital heart disease (CHD) is responsible for significant morbidity and mortality, but the aetiologies are poorly understood despite the impact of cardiac anomalies. Previous evidence shows that genetic mechanisms play a role in CHD, and recent data shows that a key gene involved in embryonic development, NR2F2, has a role in human disease. It was hypothesised that NR2F2 is part of a molecular network with a role in cardiogenesis, and its investigation could elucidate NR2F2 networks associated with heart development. Previous unpublished data showed that fibroblasts and endothelial cells from developing human hearts express NR2F2 mRNA. Transcriptomic analysis in genome editing models enables the investigation of gene expression patterns and transcription factors as regulators of pathways and genomic networks. We aimed to investigate the role of NR2F2 by knocking out its transcription in primary cardiac fibroblasts and further analyse the consequent global transcriptomic changes by single-cell RNA sequencing. We isolated primary migrating cells from human foetal heart samples and characterized them by flow cytometry and single-cell RNA sequencing. We also assessed the expression of both NR2F2 transcript and protein in those cells, and then protocols were optimized for CRISPR/Cas9. We knocked down NR2F2 in fibroblasts using the ribonucleoprotein (RNP) system and then a droplet-based single-cell RNA-sequencing was performed in knocked-down and control samples to compare the perturbations caused by the absence of NR2F2. Bioinformatic pipelines were applied to similar publicly available datasets of cardiac cells for comparison. These pipelines comprised known bulk and single-cell RNA sequencing tools for gene expression analysis, including *limma* and Scanpy. Thus, we could conclude that reduced activity of NR2F2 in human foetal fibroblasts interfere with several cardiac gene expression important for cardiogenesis, e.g. *GATA6* and *HEY1*, and may also influence the activity of other transcription factors, e.g. NKX2-5, that were previously associated with CHD phenotypes.

Keywords: cardiogenesis, cardiac fibroblasts, congenital heart diseases, CRISPR Cas9, gene editing, NR2F2, scRNA-sequencing.

Table of Contents

Table of Tables

Table of Figures

Research Thesis: Declaration of Authorship

Print name: Maria Gabriela de Oliveira Barbeta

Title of thesis: Investigating the role of NR2F2 in fibroblasts during human heart development

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signature: ..... Date: .....

Acknowledgements

During those challenging four years, I found many helpful people. I have grown and learned beyond my expectations, all thanks to the wonderful individuals I pleased to meet during this time.

Firstly, I would like to thank the University of Southampton and my funding agency in Brazil (CNPq) for giving me the opportunity to develop this project.

I am immensely grateful for my incredible supervisors. Having four different “bosses” was, at times, a challenge for my people skills. However, I could not be thankful enough for being guided and supported by them. They have pushed me to evolve personal and professionally. I would like to thank Dr Tilman Sanchez-Elsner for being more than a great supervisor, but also a professional mentor, giving me all the intellectual and emotional support, pushing me to think further than the PhD. I would also like to thank Professor David Ian Wilson which was the primely responsible for me being here in the UK, and in this project today. He was an incredible support, not only during the execution of this work, but even before that, with my university application and the year of preparation before coming to this country. Last, but not least, I am thankful for Dr Marta Polak and Dr Andres Vallejo Pulido for guiding me through the tortuous path of bioinformatics with patience and encouragement. They inspired me to go the extra mile to achieve my best in this project and not give up even when my own data failed me. Because of them, bioinformatics has a special place in my heart! I am, with no doubt, a lucky candidate.

Since the first moment I stepped on level E, I was warmly welcomed. Every single person in those offices is friendly and willing to help. They make a great professional and collaborative place in which I was never denied an advice or a helpful hand. For them, I am also grateful.

Lastly, I would like to thank my incredible family. My parents always went above and beyond to provide me with all I needed to succeed academically and as a person. I owe them my life! And, my biggest supporter, my lovely husband. He endured, with patience and no demand, all the time I had to abdicate with him, so I could be able to dedicate myself to this work. He was by my side, in every difficult day to make me laugh, and celebrated every successful experiment with me. He made Southampton feel like home and provided me with love and encouragement I needed to finish this journey. Finally, I must thank my dog, Rio, that refrain me from a frail mental health by leaving me physically exhausted but with a joyful heart.

Definitions and Abbreviations

Acronyms	Definition
5'UTR	Untranslated region at 5' end
AF-1	Activation function 1
AF-2	Activation function 2
αMyHC	alpha Myosin heavy chain
Ao	Aorta
AoVS	Aortic valve stenosis
APC	Allophycocyanin fluorophore
ASD	Atrial septal defect
AVSD	Atrial ventricular septal defect
BAV	Bicuspid aortic valve
BMI	Body mass index
BSA	Bovine serum albumin protein
CD31	Cluster of differentiation 31
CD90	Cluster of differentiation 90
cDNA	complementary deoxyribonucleic acid
CHAPS	3-((3-cholamidopropyl) dimethylammonium)-1-propanesulfonate
CHD	Congenital heart disease
CNTRL	Control

CNV	Copy number variation
CoA	Coarctation of aorta
COG	Clusters of Orthologous Group
CPM	Counts per million
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CS	Carnegie Staging
CT	Cycle Threshold
DBD	DNA-binding domain
DEG	Differently expressed genes
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DORV	double-outlet right ventricle
DPBS	Dulbecco's phosphate-buffered saline
DSB	Double-strand break
ECM	Extracellular matrix
EDTA	Ethylenediamine tetraacetic acid
EMT	Epithelial-to-Mesenchymal Transition
EndoMT	Endothelial-to-Mesenchymal Transition
ERGO	Ethics and Research Governance Online (University of Southampton)

FACS	Fluorescence-Activated Cell Sorting
FB	Fibroblast
FBS	Foetal bovine serum
FC	Fold-change
FDR	False discovery rate
FSC	Forward scatter
GEM	Gel bead-In EMulsions
GFP	Green fluorescent protein
GO	Gene Ontology
gRNA	single-guide RNA
HDBR	Human Developmental Biology Resource
HDR	Homology directed repair
HEK293	Human embryonic kidney 293 cells
HLHS	Hypoplastic left heart syndrome
HMV	Hypoplastic mitral valve
HTO	Hashtag oligos
IAA	Interrupted aortic arch
iPS-CM	Cardiomyocytes derived from induced pluripotent stem cells
IQR	Interquartile range
KD	Knock-down
KO	Knockout

LBD	Ligand-bind domain
logFC	logarithmic 10 of fold-change
LPAS	Left pulmonary artery stenosis
LSVC	Left superior vena cava
LVH	Left ventricular hypertrophy
LVNC	Left ventricular noncompaction
MEM	Eagle's Minimum Essential Medium
MRC	Medical Research Council
MRC-5	Foetal lung fibroblast cell line
MS RT	MultiScribe Reverse Transcriptase
MVP	Mitral valve prolapse
NCBI	National Centre for Biotechnology Information
NHEJ	Non-homologous end joint
Opti-MEM	Reduced Serum Media. It is a modification of Eagle's Minimum Essential Medium (MEM)
P/S	Penicillin and Streptomycin solution
PA	Pulmonary atresia
PAM	Protospacer Adjacent Motif
PCA	Principal Component Analysis
PCB	Polychlorinated biphenyls
PCR	Polymerase Chain Reaction

pcw	Post-conception weeks
PDA	patent ductus arteriosus
PFA	Paraformaldehyde
PFB	Primary fibroblasts
PFO	Patent foramen ovale
PNK	Polynucleotide Kinase
PPS	Peripheral pulmonary stenosis
PS	Pulmonary valve stenosis
PVDF	Polyvinylidene difluoride
PVR	Pulmonary vascular resistance
PVS	Pulmonary valve stenosis
qPCR	Quantitative polymerase chain reaction
REC	Research Ethics Committee
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RPMI	Roswell Park Memorial Institute medium
rSAP	Shrimp Alkaline Phosphatase
RT	Reverse transcription reaction
scRNA-seq	single-cell RNA sequencing
SOC	Super Optimal broth with Catabolite repression media
SRA	Sequence Read Archive

SSC	Side scatter
STRING	Search Tool for Retrieval of Interacting Gene/Proteins database
SVAS	Supravalvular aortic stenosis
TA	Tricuspid valve atresia
TAE	Tris-acetate-EDTA
TALEN	Transcription Activator-Like Effector Nucleases
TBE	Tris-borate-EDTA
TBST	Tris-buffered saline with Tween 20
TF	Transcription factor
ToF	Tetralogy of Fallot
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique molecule identifier
UNG	Uracil-N-glycosylase
VSD	Ventricular septal defect
ZFN	Zinc Finger Nuclease

Chapter 1 Introduction

1.1 Congenital Heart Diseases

Congenital Heart Disease (CHD) is one of the most frequent congenital disorders and the leading cause of death in newborn babies (1,2). It has been reported that CHD afflicts 0.8-1.2% of live births (1), reaching up to 10% if stillbirths are included (3,4). Cardiac malformations affect the structure of the heart, intra thoracic great vessels, and cause a loss of function in the cardiovascular system. The clinical severity varies depending on the defect but can result in intrauterine, childhood or adulthood death (5).

1.1.1 Aetiologies of Congenital Heart Disease

Despite the epidemiological relevance of CHDs, their aetiologies are poorly understood. Most congenital heart defects do not have a determined cause, and only 44% of the cases are attributed to a known aetiology (4). Genetic alterations are the main cause attributed to CHD (4,6). Of genetic causes, 3% are genetic disorders transmitted from parents, 8% are correlated to *de novo* mutations, 13% belong to aneuploidies, and 10% comprise copy number variation (CNV) causes (4,6). Environmental causes are only attributed to about 10% of CHD cases (4,6).

1.1.1.1 Environmental causes of CHD

Environmental causes encompass non-genetical factors, e.g. exposure to teratogens, pollution, and maternal illnesses. Although there are studies associating environmental factors with CHD, the mechanisms involved in this association are not well understood (7,8).

Among the maternal factors implicated as causes of CHD, pregestational and gestational diabetes mellitus have a significant association with a range of CHD (9). Obesity was also shown to increase the risk of CHD. Women with body mass index (BMI) equal to or higher than 30 kg/m^2 have an increased chance of having offspring with CHD, with an odds ratio increasing with BMI (10,11). Metabolic disorders in pregnant women are also associated with abnormal cardiac development. Maternal phenylketonuria increases the frequency of CHD 15-fold in their progeny (12).

Studies in several vertebrate models showed that exposure to environmental pollutants, such as dioxins, polychlorinated biphenyls (PCBs) and pesticides, are cardio-teratogenic, increasing cardiovascular apoptosis and decreasing cardiomyocyte proliferation (13).

In addition, maternal risk factors such as alcohol, rubella infection and teratogenic drugs (e.g. thalidomide, valproate, and retinoic acid) are also known to increase the risk of offspring CHD (14).

1.1.1.2 Genetic causes of CHD

The understanding of CHD is increasing with the advances in genomic technologies such as next-generation sequencing (15). Genetic causes consist mainly, but are not limited to, chromosomal abnormalities, *de novo,* or inherited single gene variants.

1.1.1.2.1 Chromosomal abnormalities causing CHD

Chromosomal abnormalities are the most frequent genetic cause of CHD and include chromosomal aneuploidies, microdeletions, and duplications. Aneuploidies are defined by an abnormal number of chromosomes (trisomy or monosomy), usually caused by a defect of meiosis in a parental gonadal cell (16). An epidemiologic investigation of chromosomal aneuploidies showed a 100-fold frequency increase in CHD patients compared to control, found in 12.9% of CHD patients (17), an incidence which increases with maternal age (15). The most aneuploidies associated with CHD are trisomy 21 (Down Syndrome) and the complete or partial loss of chromosome X (X0, Turner Syndrome). 40-50% of children with trisomy 21 have cardiac malformation (18), and 20-50% of foetuses with Turner's syndrome have significant heart and aorta defects (19).

Table 1.1 list examples of chromosomal abnormalities that cause CHD, syndromic and non-syndromic, the candidate genes that could be related to the CHD associated and the frequency of CHD in each alteration.

Table 1.1 Chromosomal abnormalities that cause CHD

Chromosomal Abnormality	Syndrome	Candidate Gene for CHD	CHD associated	Frequency of CHD (%)	Reference
Trisomy 21	Down	Unknown	AVSD, VSD, ASD, PDA, ToF	40-50	(20–22)
Trisomy 13	Patau	Unknown	SVAS, PPs, MVP, VSD,	57-85	(23,24)
Trisomy 18	Edward	Unknown	VSD, ASD, AVSD, ToF, DORV, HLHS, CoA	80-95	(24)
Monosomy X (45, X0)	Turner	Unknown	BAV, HLHS, dilated Ao, CoA, ASD, VSD	20-40	(15,25)
Tetrasomy 22p	Cat eye	N/A	PVR, ToF, TA, VSD, ASD	50	(26)
Tetrasomy 12p mosaic	Pallister-Killian	N/A	PDA, ASD, VSD, BAV	39	(27)
22q11.2 deletion/duplication	DiGeorge	*TBX1*	ToF, PA, VSD, TA, IAA	75	(28–30)
7q11.23 deletion	Williams-Beuren	*ELN*	SVAS, PPs, MVP, VSD	75	(31,32)
11q terminal deletion	Jacobsen	*ETS-1*	HLHS, CoA, VSD, MVS, AoVS, BAV	56	(32,33)
1q21.1 deletion/duplication	N/A	*GJA5*	CoA, BAV, ASD, VSD, PVS, ToF, PDA, TA, IAA	30	(34)

Chromosomal Abnormality	Syndrome	Candidate Gene for CHD	CHD associated	Frequency of CHD (%)	Reference
8p23.1 deletion	N/A	*GATA4*	AVC, ASD, VSD, DORV, PDA, LSVC, dextrocardia, PS, HLHS	95	(35,36)
15q terminal deletion	N/A	*IGFR1, NR2F2, MEF2A*	ASD, VSD, CoA, PDA, HMV, LPAS	N/A	(37,38)
1p36 deletion	N/A	*SKI, RERE, PDPN, CASZ1, SPEN, ECE1, HSPG2, LUZP1*	ASD, VSD, PDA, CoA, ToF, valvular abnormalities	71-75	(39,40)
9q34.3 deletion	Kleefstra	*EHMT1*	ASD, VSD, ToF, HLHS, CoA, DORV, Shone's complex	41-50	(41)

Abbreviations: Ao aorta, AoVS aortic valve stenosis, ASD atrial septal defect, AVSD atrioventricular septal defect, BAV bicuspid aortic valve, CoA coarctation of the aorta, DORV double-outlet right ventricle, HLHS hypoplastic left heart syndrome, HMV hypoplastic mitral valve, IAA interrupted aortic arch, LPAS left pulmonary artery stenosis, LSVC left superior vena cava, MVP mitral valve prolapse, MVS mitral valve stenosis, PA pulmonary atresia, PDA patent ductus arteriosus, PPS peripheral pulmonary stenosis, PS pulmonary valve stenosis, PVR pulmonary vascular resistance, TA tricuspid valve atresia, TOF tetralogy of Fallot, SVAS supravalvular aortic stenosis, VSD ventricular septal defect; N/A: not informed.

1.1.1.2.2 Monogenic causes of CHD

Because of technological advances in mutation detection, smaller genomic alterations could be identified within single genes as causes of CHD. Causative mutations in a single gene have been found in ≈12% of patients with CHD (42), and may be responsible for either isolated CHD or the heart malformations as part of a syndrome.

Although some syndromes are caused by mutations in one gene, they can present variable phenotypes of different features or severity (43). One example is the Adams-Oliver syndrome (AOS), which can be caused by mutations in the *NOTCH1* gene. This syndrome is characterised by a range of abnormal features in cranial, limb and cardiac development (44). However, studies have shown that only 20% of AOS present CHD (45). On the other hand, there are syndromic CHD that can present similar phenotypical characteristics but mutation in different genes (43). For example, Alagille syndrome, a disorder caused by a disruption in Notch signalling, is most frequently associated with mutations in the *JAG1* gene, but also mutation in *NOTCH2* in a small proportion of cases (46).

Table 1.2 lists syndromic CHD that could be associated with mutations in a single gene or in one gene of multiple possibilities.

Table 1.2 Syndromic CHD caused by single genes

Mutated gene(s)	Syndrome	CHD associated	Frequency of CHD (%)	Reference
PTPN11 ***SOS1*** ***RAF1*** ***KRAS*** ***NRAS*** ***RIT1*** ***SHOC2*** ***SOS2*** ***BRAF***	Noonan	PVS, ASD, PS, VSD, PDA	76	(47–49)
JAG1 ***NOTCH2***	Alagille	PA, ASD, VSD, ToF, PS, AoVS, CoA, HLHS	90	(46)
TBX5	Holt–Oram	ASD, VSD, PDA	75	(50,51)
NOTCH1	Adams-Oliver	AoVS, AoC, ASD, BAV, CoA, LVH, PA, ToF, VSD	20	(44,45,52)
FOXC1	Axenfeld–Rieger	ASD, AoVS, PS, ToF, BAV, TA	N/A	(53)
CHD7	CHARGE	ToF, IAA, DORV, PDA, AVSD	75-80	(54)
FBN1	Marfan	MVP	80	(55)
TFAP2B	Char	PDA	68	(56,57)
MEF2C	MEF2C haploinsufficiency	VSD, PDA, PS	N/A	(58,59)

Abbreviations: AoVS aortic valve stenosis, ASD atrial septal defect, AVSD atrioventricular septal defect, BAV bicuspid aortic valve, CoA coarctation of the aorta, DORV double-outlet right ventricle, HLHS hypoplastic left heart syndrome, IAA interrupted aortic arch, LVH left ventricular hypertrophy, MVP mitral valve prolapse, PA pulmonary atresia, PDA patent ductus arteriosus, PS pulmonary valve stenosis, PVS pulmonary valve stenosis, TA tricuspid valve atresia, ToF tetralogy of Fallot, VSD ventricular septal defect; N/A: not informed.

In addition to the syndromic cases of CHD, mutations in some genes can cause isolated CHD. About 70% of CHD cases are non-syndromic (60), and the associated genes can be grouped into broad categories, such as transcription factors, signalling molecules and structural proteins (42,61,62). Table 1.3 provides examples of selected genes with associated CHD phenotypes.

Table 1.3 Identified genes that cause non-syndromic CHD

Gene	CHD Phenotype	Reference
	Transcription Factors	
CITED2	ASD, ToF, VSD	(63)
GATA4	ASD, VSD, AVSD, PS, TOF	(64,65)
GATA5	AoVS, PS, DORV, ToF, BAV, ASD, VSD	(66)
GATA6	DORV, VSD, PDA, PS, PTA, ToF	(67,68)
HAND1	ASD, AVSD, VSD, HLHS	(69,70)
HAND2	VSD, PS, DORV, ToF, PDA, PA, ASD	(71,72)
NR2F2	ToF, AVSD, AS, VSD, HLHS, CoA	(73)
NKX2-5	ASD, ToF, VSD, HLHS	(74)
NKX2-6	TA, PAS, VSD, CoA	(75,76)
TBX20	ASD, VSD, MVD	(77)
NFATC1	VSD	(78)
ZFPM2/FOG2	ToF, DORV, TA, MVP	(79,80)
	Signalling molecules	
CRELD1	AVSD	(81)
FOXH1	DORV, ToF, PDA, ASD, VSD	(82)
HEY2	ASD, VSD, TA, MVP, HLHS	(83,84)
NODAL	Heterotaxy, AVSD	(85)
SMAD2	Heterotaxy, DORV, ASD, VSD, PDA, valve anomalies	(86)
SMAD6	BAV	(87,88)
VEGFA	ToF	(89)
	Structural proteins	
ACTC1	ASD, AS, PS, VSD, LVNC	(90,91)
DCHS1	MVP	(92)

Gene	CHD Phenotype	Reference
MYH6	ASD	(93)
MYH7	Ebstein's anomaly, HLHS, LVNC, DORV, VSD	(94,95)
MYH11	PDA	(96,97)

Abbreviations: AoVS aortic valve stenosis, ASD atrial septal defect, AVSD atrioventricular septal defect, BAV bicuspid aortic valve, CoA coarctation of the aorta, DORV double-outlet right ventricle, HLHS hypoplastic left heart syndrome, LVNC left ventricular noncompaction, MVD mitral valve disease, MVP mitral valve prolapse, PA pulmonary atresia, PDA patent ductus arteriosus, PS pulmonary valve stenosis, TA truncus arteriosus, ToF tetralogy of Fallot, VSD ventricular septal defect.

1.2 NR2F2 and its contribution to CHD

NR2F2 is a gene encoding Nuclear Receptor subfamily 2, group F, member 2 (NR2F2), and it is implicated in heart development through several lines of evidence. In 1999, a study in mice showed that *Nr2f2* homozygous knockout (*Nr2f2*$^{-/-}$) resulted in lethality around 10 days of gestation, presenting secondary to neural and cardiac haemorrhage. One-third of *Nr2f2*$^{+/-}$ mice survived until after weaning and were smaller than wild-type littermates. Histological analysis revealed atria and vascular formation defects, indicating that Nr2f2 is involved with angiogenesis and heart development (98). In another study, *Nr2f2* mutants developed cardiovascular defects that included atrioventricular septal defect, thin-walled myocardium, and poor angiogenesis (99). Endothelial-specific *Nr2f2* deficient embryos presented endocardial cushion hypoplasia due to a failure in epithelial-mesenchymal transition (EMT) (99).

Adding to rodent studies, studies have also identified *NR2F2* mutations as a cause of human heart abnormalities. In 2011, a group reported a patient with ventricular and atrial septal defects and a 15q terminal deletion that included *NR2F2*. A comparison made with other reports of CHD associated with chromosomal deletions in the *NR2F2* region suggested deletion of this gene was causative of heart malformations (100). Later, collaborative groups in the UK (University of Southampton and Sanger Institute, Cambridge) and Canada (Hospital for Sick Children, Toronto) carried out exome sequencing in 13 parent-offspring trios and 112 unrelated individuals with atrioventricular septal defects. Their findings showed five rare missense variants in *NR2F2* that were not identified in control groups. In addition, three families with CHD were found to have additional types of mutation in *NR2F2*, including a balanced de novo chromosomal translocation, a *de novo* substitution affecting a splice donor site and a 3-bp duplication, coming to a total of six exon variants of *NR2F2* screened (101). In the same study, the NR2F2 activity was measured by a luciferase assay comparing a wild-type NR2F2 expression with all six variants found by the group, concluding that the mutations impact the transcriptional activity of the gene on target promoters.

More recently, after sequencing 168 unrelated CHD-affected individuals, a group detected a *de novo* nonsense mutation in *NR2F2* in one individual. Functional assays were carried out,

showing NR2F2 mutant protein losing transcriptional activity compared to unaltered protein (102). These findings strongly suggest that NR2F2 is a genetic cause of CHD and has an essential role in heart development.

1.2.1 NR2F2 roles in organogenesis

Several studies have sought to determine the mechanisms and targets related to NR2F2. In reproductive system development, *SOX9* is known to be a pro-testis gene, and *WNT4* is a pro-ovary gene (103). Independent animal studies showed NR2F2 negatively regulates *Sox9* in osteogenic mesenchyme in mice (104) and positively regulates *WNT4* in human endometrium (105), suggesting that NR2F2 participates in a female phenotype of reproductive organs (103,106).

Skeletal muscle formation also depends on a dose-balanced expression of NR2F2 (107). It is required for myoblast proliferation in early mice development, but its expression level decreases along with the myoblast differentiation as Nr2f2 can inhibit the fusion of myoblasts into myotubes, a process required for skeletal muscle formation (107). The expression of Nr2f2 is also required for renal cell maturation. Lack of Nr2f2 in mice impaired metanephric mesenchyme formation, and deletion at later stages reduced the expression of several metanephric mesenchymal genes, e.g. *Eya1*, *Pax2*, *Six2* and *Wt1* (108).

Cell fate decisions are critical during development and are controlled by transcription factor networks (109). In the cardiovascular system, NR2F2 is known to participate in angiogenesis (110). It is predicted to interact with *PROX1* (111,112). *PROX1* expression and maintenance are required to induce the endothelium towards a lymphatic vascular phenotype (111), and it was shown that NR2F2/PROX1 heterodimers would trigger lymphatic endothelial cell lineage, and NR2F2 homodimer would induce venous endothelial cell lineage and inhibit an arterial phenotype by binding to *HEY1/HEY2* promoter regions (113). In the heart, NR2F2 also participates in atrial/ventricular cardiomyocyte identity decisions (109). In mice, immunoprecipitation experiments showed that Nr2f2 binds to the Tbx5 promotor stimulating its expression, which suggests that *Tbx5* is a target of Nr2f2. This modulation could happen through interaction with Sp1, as its binding site is required for Nr2f2 to promote the expression of *Tbx5*. Interestingly, Nr2f2 could also suppress the expression of two pro-

ventricular genes, Hey2 and Irx4, through direct biding to their genomic loci, which supports the proposal that NR2F2 is required for decisions between atrial or ventricular cardiomyocyte cell fate (109,114).

1.2.2 *NR2F2* Mutations

NR2F2 is essential during embryonic development for the heart and other organs. A mice model showed homozygous deletion of *Nf2f2* in the mesentery caused diaphragmatic hernia, herniation of the liver and stomach in the thoracic cavity and asplenia. The findings suggested that *Nr2f2* could have a role in the formation of the diaphragm and attachment of the post-hepatic mesenchymal plate to the body wall (115). Asplenia and congenital diaphragmatic hernia are phenotypes that can also be found in humans with *NR2F2* variants (116,117). Other features associated with *NR2F2* in humans include testicular or ovotesticular differences in sex development, virilization of the external genitalia, dysmorphic features, developmental delay, and underdeveloped limbs (103,117–119).

The most clinically severe disorders associated with NR2F2 mutations relate to the heart. *NR2F2* variants can lead to several isolated heart defects (118). Table 1.4 and Figure 1.1 summarizes the main cardiac phenotypes associated with *NR2F2* variants, with the effect of the mutation.

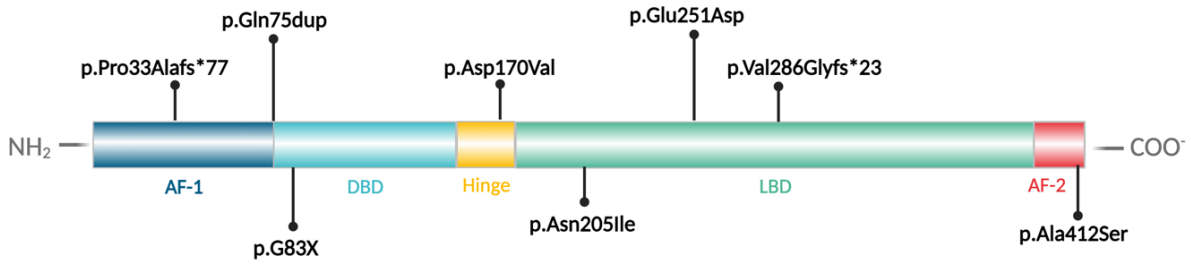


Figure 1.1 NR2F2 mutations and their positions in the canonical isoform

In this schematic representation of the NR2F2 protein, it is indicated the location of each described mutation in Table 1.4. Each coloured square represents a protein domain: AF-1 (activation function 1; blue), DBD (DNA-binding domain; light blue), hinge (yellow), LBD (ligand-binding domain; green) and AF-2 (activation function 2; red). NH_2 is the amino terminus and the COO^- is the carboxyl terminus of the protein. Created with BioRender.com.

Table 1.4 Cardiac phenotypes associated with NR2F2 variants

Phenotype	DNA Mutation	Protein Alteration	Mutation Type	Reference
ToF				
AVSD	c.220_222dup	p.Gln75dup	in-frame duplication	
AS, VSD				
AVSD	c.1022C>A	p.Ser341Tyr	missense	
AVSD	c.614A>T	p.Asn205Ile	missense	(73)
AVSD	c.753G>C	p.Glu251Asp	missense	
AVSD	c.1234G>T	p.Ala412Ser	missense	
AVSD	c.509A>T	p.Asp170Val	missense	
HLHS	c.970þ1G>A	N/A	splice donor	
CoA	t(14;15)(q23;q26.3)	N/A	balanced translocation	
Coa, PFO, PDA	c.856dupG	p.Val286Glyfs*23	frameshift	(118)
ASV	c.92_98delGCCCGC	p.Pro33Alafs*77	frameshift	(116)
CoA	del(15)(q26.1)	N/A	unbalanced translocation	(120)
DORV, VSD	c.247G>T	p.G83X	nonsense	(102)

Abbreviations: ASD atrial septal defect, AVSD atrioventricular septal defect, CoA coarctation of the aorta, DORV double-outlet right ventricle, HLHS hypoplastic left heart syndrome, PDA patent ductus arteriosus, PFO patent foramen ovale, ToF tetralogy of Fallot, VSD ventricular septal defect. N/A not applicable

1.2.3 Genomic structure of *NR2F2*

The *NR2F2* gene (previously known as Chicken Ovalbumin Upstream Promoter Transcription Factor II, COUP-TFII) is located at chromosome 15 (15q26.2: 96,326,046-96,340,263) (Figure 1.2**Error! Reference source not found.**) (121,122). The genomic structure of *NR2F2* was described in 1995, and it is composed of 6 exon regions (123). Each transcript of *NR2F2* is formed by 3 exons, and 3 different protein isoforms (a, b and c) of NR2F2 were identified as splicing products of those 4 transcripts (NCBI, Gene ID: 7026) (122,123). According to the NCBI database, the transcript variant NM_021005.4 is canonical and encodes the longest isoform

(a) of NR2F2. Both transcript variants NM_001145156.1 and NM_001145157.2 encode the same protein isoform (c) (122).

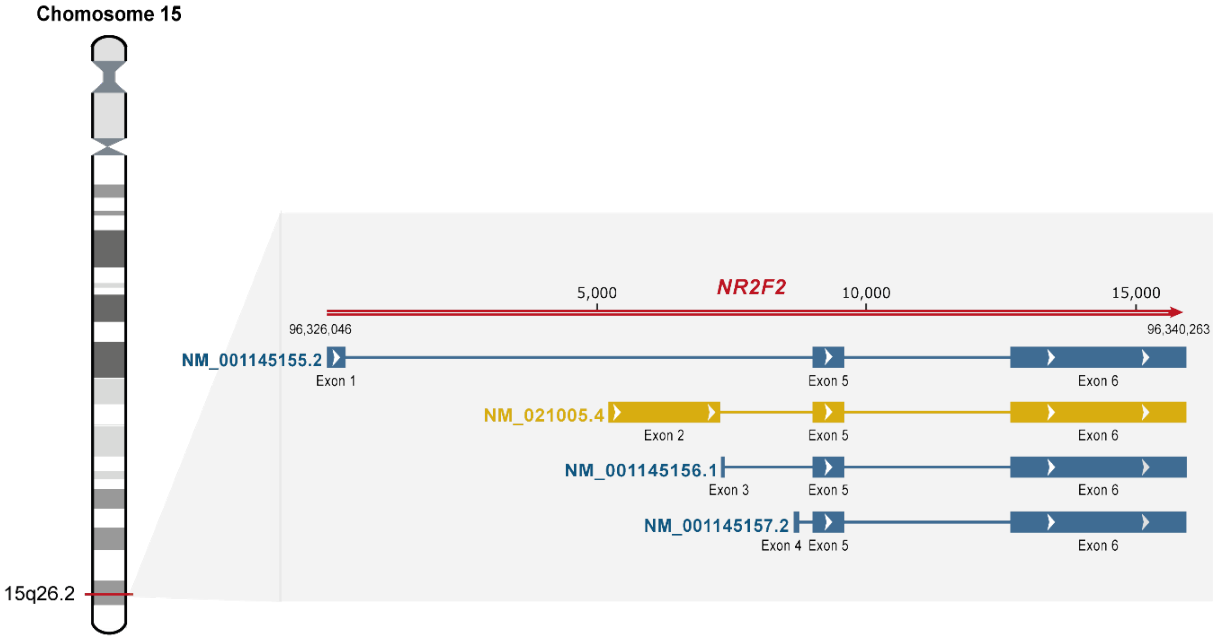


Figure 1.2 *NR2F2* genomic location and structure of transcripts

NR2F2 is located at Chromosome 15 (15q26.2). There are 6 exon regions in the gene that initiate transcription of 4 transcripts: NM_021005.4, NM_001145155.2, NM_001145156.1 and NM_001145157.2 (nomenclature from NCBI ID 7026) formed by 3 exons each (represented by rectangles). The canonical transcript is the NM_021005.4 (golden) and it is formed by the exons 2, 5 and 6. The other three transcript variants (blue) share the exons 5 and 6 with the canonical variant but differ in the first exon of each (exons 1, 3 and 4 compared to exon 2). The *NR2F2* DNA structure is represented in red, and it is contained between the 96,326,046 and 96,340,263 base positions in the genome. The white arrows show the direction of transcription.

1.2.4 NR2F2 protein structure and function

The canonical isoform (a) of NR2F2 protein is formed by 414 amino acids and has an N-terminal DNA-binding domain (DBD), a C-terminal ligand-binding domain (LBD), and two activation function domains: AF1 and AF2 (124) (Figure 1.3). The NR2F2-DBD is composed of two zinc fingers that are similar to the structure of other members of the protein family. The DNA binding function of NR2F2 is considered promiscuous due to its ability to recognize direct repeats and palindromes of the GGTCA motif with different spacings (1-5bp) (125). This property allows NR2F2 to interact with many targets. For functional DNA-binding, two GGTCA half-sites accommodate the activated dimer form of NR2F2 that undergoes structural adaptation to recognise the interspaced GGTCA motifs (125). The NR2F2-LBD domain comprises 10 α-helices that fold into a 3-layered helical sandwich, a known conformation for

NRs. The N-terminal portion of helix α10 is primarily responsible for dimer formation by hydrophobic interaction, which is crucial for the whole activity of the protein (126). As a transcription factor, NR2F2 acts by modulating gene expression, and it has been suggested that it requires the presence of hydrophobic molecules, which have yet to be identified to bind into the LBD ligand site. In the absence of a ligand, the C-terminal portion of helix α10 collapses into the ligand-binding site showing an auto-repressed conformation of NR2F2. Alteration in the AF-2 portion at the C-terminal end of the NR2F2-LDB showed a significant reduction in NR2F2 activity, demonstrating that an intact structure of this region is also required for NR2F2 activation. Considering the importance of NR2F2-LDB for its activation, all the NR2F2 isoforms contain this portion and, therefore, can play a role in NR2F2 function regulation. Compared to the longest isoform (a), isoforms b and c have shorter N-terminus because of the differences in the 5'UTR and coding sequences of their first exon (122).

The transcriptional activity of NR2F2 can occur through several mechanisms. It can form a homodimer or heterodimer with other transcription factors to activate or repress its target (127,128). To active target genes, NR2F2 binds directly into hormone response elements, indirectly as an accessory factor or via protein-protein interaction. It can repress gene expression by binding directly into the target's promoter region, indirectly by competing for occupancy of other transcription factors binding sites or competing with other nuclear receptors for heterodimerization with retinoid X receptors (RXR), and by transrepression (127,128). This last mechanism of repression is defined by a transcription factor binding directly to the LBD of other nuclear receptors preventing their activation (127,128). Transrepression could be the predominant mechanism used by NR2F2 to suppress gene expression (127,129).

Although NR2F2 is still considered an orphan nuclear hormonal receptor, some molecules can modulate its transcription activity *in vitro*. Retinoic acids were described as activators of NR2F2 in a higher concentration than physiological levels (126), whereas a synthetic naphthol compound was described as an inhibitor of it (130). Recently, a screening of 516 metabolic enzymes alongside their substrates and products found that non-canonical alanine-based sphingolipids (1-deoxysphingosines) bind to NR2F2 and NR2F1 LBDs and modulate by increasing their transcriptional activity *in vitro* at physiological concentrations, but not NR2F6

– the third member of the NR2F family (131). Interestingly, the group that described 1-deoxysphingosines as ligands of NR2F2 could not identify any enzymes or retinoids that belong to the retinoic acid pathway able to modulate transcriptional activity of NR2F2/1, conflicting with previous data showing unphysiological concentrations of retinoic acids could put NR2F2 in an activation conformation, capable of recruiting co-activators (126).

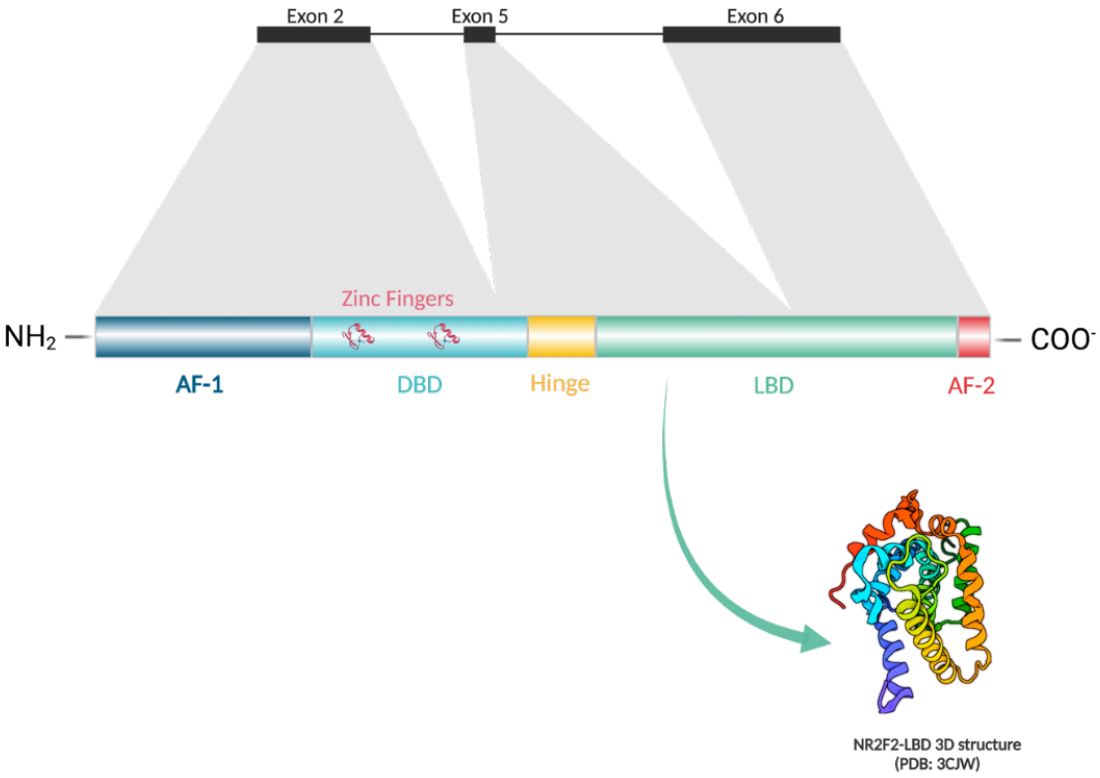


Figure 1.3 Protein structure of NR2F2

NR2F2 protein is composed by 5 domains: AF-1 (blue), DBD (light blue), hinge (yellow), LBD (green) and AF-2 (red). AF-1/2: activation function 1/2; DBD: DNA-binding domain; LBD: ligand-binding domain. NH_2: amino terminus. COO^-: carboxyl terminus. Created with BioRender.com.

1.2.4.1 Homology between NR2F-family members

The molecular structure of human NR2F family members (NR2F2, NR2F1 and NR2F6) is highly conserved. NR2F2 and NR2F1 share the highest degree of homology among the members of the NR2F-family (132). When aligned using Clustal Omega, the protein sequence of NR2F2 (ID Uniprot: P24468) shows ≈89% of similarity with NR2F1 sequence (ID Uniprot: P10589), in comparison to NR2F6 (ID Uniprot: P10588) with only 66% similarity with NR2F2 sequence. Looking at binding domains and comparing NR2F2 and NR2F1, DBD and LBD have 97% and 99%

homology, respectively (Figure 1.4) (123,133). However, the homology between them in the AF1 portion is lower (45%), indicating that they can bind in similar elements on the DNA but may provide distinct functions between them (134). Between NR2F2 and NR2F6, the amino acids that match in DBD and LBD sites correspond to 86% and 76%, respectively (Figure 1.4).

NR2F2 is also highly conserved between species. Its homologous proteins can be found in several members of Metazoa (135). NR2F2 orthologues are found in mice (123,136), chicken (137), *Xenopus* frog (138), and zebrafish (139,140). They form a group in the phylogenic tree of the NR2F-family with a high percentage of similarity (>95%) (124).

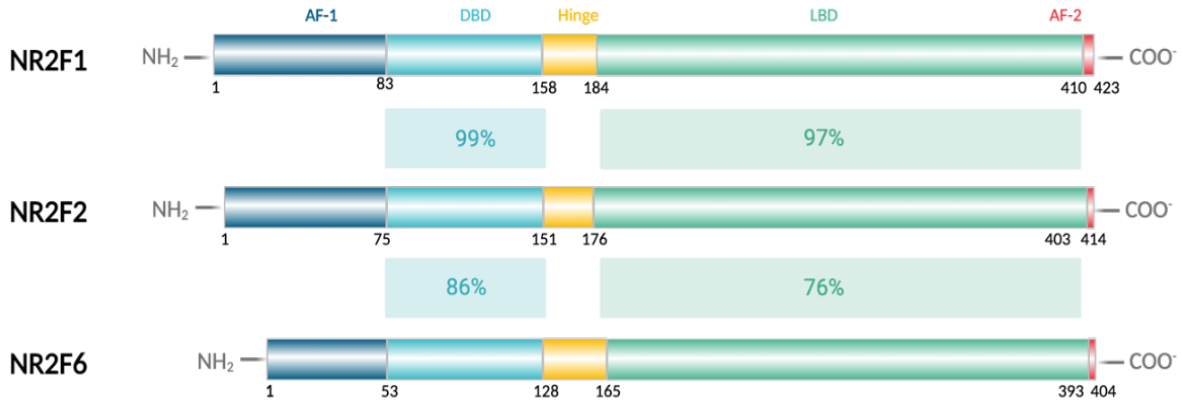


Figure 1.4 NR2F protein homology in humans

The figure shows protein structure of human NR2F1 (top), NR2F2 (middle) and NR2F6 (bottom). The numbers represent the amino acid positions. The percentages in the blue rectangle indicate the similarity of DBDs in relation to DBD-NR2F2. The percentage in green rectangle indicates the similarity of LBDs in reference to LBD-NR2F2. Created with BioRender.com.

1.3 Cellular composition of the heart

The mammalian heart is composed mainly of cardiomyocytes, fibroblasts, endothelial cells, and smooth muscle cells (141). Cardiomyocytes are responsible for the contraction of the heart, but other non-cardiomyocyte cells play other important roles in the heart. They are responsible for vascularisation, extracellular matrix organisation, cell-to-cell communication, and other tasks related to cardiac homeostasis and development (142).

Several studies show that the cellular composition of the mammalian heart can differ between species and ages of development. Rodent studies demonstrated that the murine adult heart is composed of 56% cardiomyocytes and 44% non-cardiomyocyte cells (143), contrasting with

previous reports of the rat adult heart presenting only 30% of cardiomyocytes and 70% of non-cardiomyocyte cells (144). Banerjee et al. also analysed the cellular constituents in different developmental stages of rat and mouse hearts. Similarly, they observed differences in the proportion of cardiomyocytes and non-cardiomyocyte cells in different stages of development, increasing postnatally until adulthood (from ≈10% to ≈30% in mice and from ≈30% to ≈60% in rats) (143,145).

Recently, single-cell transcriptomic analysis has been unveiling the cellular diversity in the human heart. It has been revealed that cardiomyocytes comprise ≈36% of the cells, and regarding nonmyocyte cells, ≈30% are related to fibroblasts (146,147). In human foetal heart studies, this figure changes to 45% and 12% related to cardiomyocytes and fibroblasts, respectively (148).

1.3.1 Noncellular composition of the heart: the extracellular matrix

The extracellular matrix (ECM) is formed by fibrillar and nonfibrillar proteins secreted by the cardiac tissue's resident cells and organised in a tridimensional network (149,150). Although it is known for its scaffolding function, the ECM is also a dynamic component of the heart. A cell-mediated turnover determines its composition, which leads to biomechanical, electrical, and chemical responses between the cardiac cells and the microenvironment (149,151).

The cardiac ECM is composed mainly of collagen fibres (I, III and IV), fibronectin, laminin, proteoglycans, glycosaminoglycans, and elastins (150) distributed in two major regions: interstitial space and basal membrane (152). The interstitial space is between the myocardium and the endocardium, separated from the cellular layer by the basal membrane (153). It is formed 80% by collagen I and 10% of collagen III (150); the remaining portion is formed by an amorphous substance made by proteoglycans and glycoproteins (154,155). The basal membrane is placed in the pericellular region of the cardiac tissue (153), and it is mainly composed by fibronectins, laminins, perlecans, nidogens, and collagen IV (156).

1.4 Role of Fibroblasts in Cardiogenesis

Fibroblasts are essential for maintaining the structural integrity of the heart. They are responsible for remodelling the ECM, communicating with immune cells, and participating in

electrical conductivity and rhythmicity of the cardiac muscle (143,157). Fibroblasts are also responsible for the fibrotic and inflammatory first response during injury of an adult heart (158). However, because of their heterogeneity and lack of specific markers, the study of the origin and role of fibroblasts during human embryonic development remains challenging.

Studies showed that after heart looping, cells from the primitive epicardium undergo epithelial-to-mesenchymal transition (EMT) and migrate to the myocardium (159) and later give rise to most interstitial and adventitial cardiac fibroblasts and vascular smooth muscle cells (160). This process is controlled by a specific set of proteins at specific time points in the EMT. Before the EMT, the transcription factor 21 (Tcf21) expression is important for cell migration, and this is evident in *Tcf21*-null mice with epicardial cells unable to undergo EMT (161). After the transition, interactions between growth factors and transcription factors control later processes. For example, interruptions in fibroblast growth factor 10 (Fgf10) cascade signalling reduce heart size by the reduced migration rate of cardiac fibroblasts to the myocardium (162).

Although most cardiac fibroblasts are of epicardial origin, a small percentage is derived from the endothelium that coats the inner lumen of the heart (i.e. endocardium). The endothelial cells that form the endocardium undergo an endothelium-to-mesenchymal transition (EndoMT) (155,156). The fibroblasts generated by this process can be found in the atrioventricular septum and left ventricle. Lastly, a smaller subset of cardiac fibroblasts is derived from the neural crest. Fibroblasts derived from the neural crest are necessary for outflow tract development and valve formation (163,164). It has been estimated that about 80% of fibroblasts from the adult heart are Wnt1-positive cells, indicating they are of epicardial origin, around 18% come from the endocardium, and the remaining percentage arise from the neural crest (163). Studying the origin of fibroblasts and the processes of EMT and EndoMT is crucial not only to understanding normal cardiogenesis but also to abnormal development and CHD (155).

It is clear that fibroblasts contribute to the three-dimensional organization of the four-chambered heart during embryonic development, but their roles are not limited to the structural framework of the heart. They can interact with cardiomyocytes stimulating

proliferation via ß-integrin signalling (165). Also, although non-contractile cells, they contribute to electrical coupling by contacting with cardiomyocytes via connexins (166,167). Moreover, the importance of fibroblasts remains after birth. Murine neonatal hearts increase the thickness of ventricular walls due to higher systolic pressure. This enlargement of heart wall size is attributed to a double number of fibroblasts postnatally which remodels the extracellular matrix providing distributions of the mechanical distress (166).

1.5 The combined strategy: gene perturbation and transcriptomic analysis

Gene editing strategies associated with transcriptomics analysis can be used to investigate gene networks. Several groups have combined gene perturbations via CRISPR/Cas9 with single-cell RNA sequencing and developed methods to elucidate gene function and interactions. In December 2016, three groups released studies describing similar techniques using this approach: CRISPR-seq and Perturb-seq. CRISPR-seq was applied in mice samples to trace different functions of two transcription factors in immune cells in response to pathogens (167). Perturb-seq, on the other hand, was used by two groups to knock out multiple transcription factors regulated by cell immune response to lipopolysaccharide and investigate aftereffects (168), and to suppress genes involved in cell response to endoplasmic reticulum stress (169). More recently, a study using primary human Langerhans cells combined CRISPR knockout with single-cell Drop-seq RNA sequencing. The group aimed to elucidate the genomic network responsible for context-dependent immune responses of those cells (170). Although these methods share numerous similarities, their distinct experimental design and analysis pipelines demonstrate flexibility in dealing with a wide range of biological interrogations.

In this manner, considering the importance of NR2F2 and fibroblasts in the process of cardiogenesis, the present work will betake this combining methodology of causing gene perturbation in foetal cardiac fibroblasts through CRISPR/Cas9 technology and evaluate these gene expression changes using single-cell RNA sequencing to investigate the role of *NR2F2* in those cells during the heart development.

1.6 Hypothesis and Aims

1.6.1 Hypothesis

NR2F2 interacts with genes related to cardiogenesis pathways in cardiac fibroblasts.

1.6.2 Aims

We aim to investigate the molecular pathway involving NR2F2 during heart development using a gene editing technique in primary foetal fibroblasts combined with transcriptomic analysis.

1.6.2.1 Specific Aims

i. Establish a protocol to isolate and characterize primary foetal cardiac fibroblasts
ii. Perform NR2F2 knockout in primary foetal cardiac fibroblasts using CRISPR/Cas9 technology
iii. Perform single-cell RNA sequencing in *NR2F2* knocked-out primary foetal cardiac fibroblasts
iv. Identify the specific targets downstream of NR2F2 involved in heart development

Chapter 2 Material and Methods

2.1 Tissue Handling and Cell Culture

2.1.1 Heart Sample Collection

Ethics approval for using of foetal tissues was obtained from NRES Committee North East - Newcastle & North Tyneside 1 (REC reference: n08/H0906/21+5 by the Newcastle HDBR. With informed consent and signed permission, human embryonic/foetal material from the developmental period of 6-13 weeks post-conception was provided by the Joint MRC/Wellcome Trust (grant #099175/2/12/2) Human Developmental Biology Resource (www.hdbr.org). All protocols follow the University of Southampton health and safety policy and governance (ERGO #9962).

2.1.2 Foetal Heart Dissection

In total, fourteen hearts were processed. The foetal hearts were embedded in Dulbecco's Phosphate Buffered Saline (DPBS, Sigma). Five hearts were processed without dissection, and nine hearts were dissected in the atria, outflow tract and ventricles using micro-dissection scissors before processing. The atria and outflow track were dissected from the heart, and the three parts (ventricles, atria, and outflow tract) were processed independently. Either the whole heart or each part was chopped into cubicle pieces of 1 mm^3 (McIlwain Tissue Chopper, Campden Instruments) and used to isolate migrating cardiac cells.

2.1.3 Isolation of Migrating Cells

6-well plates were treated with 1 mg/ml of gelatin (Sigma) for at least 30 minutes at 37 °C. Then, the gelatin was removed, and the wells were washed with sterile DPBS. Gelatin coating was used only during the isolation step. The cubicle pieces of tissues were plated 3-4 in each of the gelatin-treated wells and cultivated with 3 ml of Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% of Foetal Bovine Serum (FBS, Sigma) and 1% of Penicillin/Streptomycin (P/S, Gibco). The media was changed after five days to ensure adherence to the explants. The migrating cells were harvested in culture after 7-10 days using a filtering method.

2.1.4 Purification of Migrating Cells

Once the cells migrated, the media was removed, and the cells were washed with sterile DPBS. 700 µl of 0.25% Trypsin-EDTA (Sigma) was added to each well to detach the cells and tissue explants adhered to the plate. After 5 minutes at 37°C, the process was halted by adding 1 ml of culturing media to each well, and the content was collected in a 15 ml tube. A filtering method was used to separate the migrating cells from the tissue explant. The media containing the cells and explants was filtered using a 70 µm filter, separating single cells from the tissue explant retained on the filter. The cellular filtrate was then centrifuged for 5 minutes at 300 g. The resulting pellet was re-suspended in DMEM supplemented with 10% FBS and 1% P/S and plated in T-75 flasks to expand. The media was changed every 3 days, and the cells were expanded until the first passage before being frozen.

2.1.5 Primary Cells Culture

Primary cardiac foetal cells were plated in T-25 flasks at a cell density of $1x10^5$ cells per flask. Once totally confluent, the cells were transferred to T-75 flasks to continue to expand. The media used throughout this process was DMEM supplemented with 10% FBS, which was changed every 2 days.

2.1.6 MRC-5 Cells Culture

MRC-5 (ATCC CLL-171) is a primary lung fibroblast cell line from foetal origin. MRC-5 cells were plated in T-75 flasks at a cell density of $3x10^5$ cells per flask. No gelatin coating was used to treat flasks for MRC-5 culturing at any time. The media used was DMEM supplemented with 10% FBS and changed every 2 days until total confluency.

2.2 Migrating Cells Characterisation by Flow Cytometry

The characterisation of cells isolated from the heart was performed through flow cytometry by using the measure of fibroblast markers (CD90 and Vimentin), endothelial (CD31) and cardiomyocytes (αMyHC) markers. The cells were harvested using 0.25% Trypsin-EDTA and centrifuged at 500 g at 4 °C for 5 minutes. The pellet was re-suspended in DPBS. Cells were washed twice more with DPBS using the last centrifugation setup to remove all residual protein in the solution. Viable cells were stained using Zombie Violet Fixable Viability Kit (BioLegend) in the proportion of 1:100 in DPBS for 30 minutes at room temperature in the

dark. After viability staining, 1 ml of DPBS containing 0.5% FBS was added to chelate the remaining amine-reactive fluorescent dye in the solution, and one more centrifugation step was performed. Cell surface markers staining was performed using 1:20 of CD90-PercP/Cy5.5 (BD Pharmingen) and 1:200 of CD31-APC (ImmunoTools) at room temperature for 45 minutes. Paraformaldehyde (PFA) was added to a final concentration of 1% to fixate the cells for 30 minutes at 4 °C. The cells were washed and re-suspended in 5% FBS in DPBS containing 0.1% saponin for 20 minutes at room temperature, aiming to permeabilise and permit intracellular staining. For intracellular staining, Vimentin-Alexa Fluor 488 (BD Pharmingen) and Myosin4-eFluor 660 (αMyHC, eBioscience) were used at a proportion of 1:20 and 1:200, respectively, and the staining process occurred at room temperature for 45 minutes. The fluorescence measurement was carried out on BD FACSAria II (BD Biosciences).

2.3 Identification of NR2F2 in Cardiac Cells

2.3.1 NR2F2 Transcripts Detection by Quantitative Polymerase Reaction Chain

2.3.1.1 RNA Extraction

Extraction of total ribonucleic acid (RNA) from cells was performed using a phenol-chloroform protocol. For RNA extraction, cells were harvested once confluent using 0.25% Trypsin-EDTA and pelleted at 300 g for 5 minutes. The pellet was then washed in DPBS, transferred to 1.5 ml microcentrifuge tubes and centrifugated at 600 g for 5 minutes. After removal of the supernatant, the cells were re-suspended in 750 μl of TRIzol LS, mixed by vortexing and left at room temperature for 5 minutes. 200 μl of chloroform was added to induce phase separation, in which proteins are in the organic phase, DNA resolves in the interface, and RNA remains in the aqueous phase. The sample was mixed vigorously by vortexing and left at room temperature for 5 minutes. The aqueous phase was carefully collected after centrifuging the samples 9,600 g at 4 °C for 20 minutes. For RNA precipitation, half of the solution volume of isopropanol was added along with 1 μl of glycogen, mixed by vortexing and left at -80 °C for 20 minutes.

The samples were centrifuged at 17,000 g at 4 °C for 30 minutes, and the supernatant containing isopropanol was carefully removed and discarded. 500 μl of 75% ethanol was added, followed by centrifugation of 17,000 g at 4°C for 5 minutes. The ethanol was removed,

and the pellet was air-dried. 20 µl of RNAse free water was then added, and the RNA was measured by wavelength absorbance (A260) using spectrophotometry (Nanodrop, Thermo Scientific).

2.3.1.2 cDNA Synthesis

The extracted RNA was used as a template to synthesise corresponding complementary deoxyribonucleic acid (cDNA) by reverse transcription reaction. For this experiment, 100ng of RNA from each sample was used. A mix of components of the High-Capacity cDNA Reverse Transcription Kit (4368814, Applied Biosystems) was added according to Table 2.1. The final volume per reaction was 10µl. The samples were loaded at the thermocycler to activate the reaction following the setup in Table 2.2.

Table 2.1 Content of Reverse Transcription Master Mix

Reagent	Volume	Concentration
RT Buffer (10x)	1 µl	1x
Random Hexamers (10x)	1 µl	1x
dNTP (100 mM)	0.4 µl	4 mM
RNAse inhibitor (40 U/µl, Promega)	0.5 µl	2 U/µl
MS Reverse Transcriptase (50 U/µl)	0.5 µl	2.5 U/µl
RNA + H_2O	6.6 µl	100 ng of RNA

Table 2.2 Thermocycler set-up for cDNA synthesis

	Primer annealing	cDNA synthesis	Enzyme deactivation	Hold
Temperature	25 °C	37 °C	85 °C	4 °C
Time	10 min	120 min	5 min	∞

2.3.1.3 Quantitative PCR

Quantitative Polymerase Chain Reaction (qPCR) was carried out to identify the expression of *NR2F2* in each sample. TaqMan technology was used to quantify the expression of the transcripts. TaqMan technology delivers a fluorescent probe together with the primer that

accumulates during the PCR cycles. The accumulated fluorescence was detected by the 7900HT Fast Real-Time PCR System (Applied Biosystems). The cDNA was diluted in a proportion of 1:10 using RNAse and DNAse-free distilled H_2O (ddH_2O, Gibco). A commercial primer for *NR2F2* was used (Hs00819630_m1, Thermo Fisher) to identify the canonical *NR2F2* transcript (NM_021005.3). 2.5 µl of TaqMan Universal Master Mix II (Applied Biosystems) with 0.25 µl *NR2F2* TaqMan Gene Expression Assay (Applied Biosystems) were prepared for each reaction, completed with 1.25 µl of ddH2O. 1 µl of diluted cDNA was added to the designed well along with 4 µl of the mix. The total reaction volume was 5 µl per well. *GAPDH* (Hs04420632_g1, Thermo Fisher) and *ACTB* (Hs01060665, Thermo Fisher) were the housekeeping genes used as an endogenous control. The program used to perform the PCR is demonstrated in Table 2.3.

Table 2.3 Thermocycler setup for qPCR

	Initial Steps		**PCR – 40 cycles**	
	UNG Activation	Polymerase Activation	Denaturation	Annealing/Extension
Temperature	50 °C	95 °C	95 °C	60 °C
Time	2 min	10 min	15 s	1 min

2.3.2 NR2F2 Protein Detection by Flow Cytometry

For analysis of NR2F2 at the protein level, cells were harvested using 1x TrypLE Express (Gibco), centrifuged at 300 g at 4 °C for 5 minutes, and the pellet was re-suspended in DPBS. Two washed with DPBS were made using the previous centrifugation setup. Viable cells were stained using Zombie Violet™ Fixable Viability Kit (Bio Legend) in the proportion 1:100 for 30 minutes at room temperature in the dark. 3 ml of DPBS containing 0.5% Bovine Serum Albumin (BSA, Sigma) was added, and cells were centrifuged at 300 g at 4 °C for 5 minutes. The pellet was re-suspended in 1% PFA to fix the cells for 30 minutes at 4 °C. The cells were washed and re-suspended in 5% FBS in DPBS containing 0.1% saponin for 20 minutes at room temperature for permeabilisation. For primary staining, the cells were incubated for 1 hour with 1:100 recombinant Anti-NR2F2 antibody [EPR18443] (Abcam) at room temperature, followed by two washes with 5% FBS in DPBS containing 0.1% saponin. For secondary staining,

1:10,000 Chicken anti-Rabbit IgG (H+L) Cross-Adsorbed- Alexa Fluor 647 antibody (Invitrogen) was added to the cells for 30 minutes. The sample was washed twice using 5% FBS in DPBS with 0.1% saponin. The cells were re-suspended in 5% FBS in DPBS and analysed on BD FACSAria II (BD Biosciences).

2.4 Insertion of CRISPR guide RNAs into Cas9 Vectors

2.4.1 Plasmids Digestion and Dephosphorylation

Two different vectors were used in the gene editing assays, depending on the delivery method of those vectors.

pSpCas9(BB)-2A-Puro (PX459) V2.0 was a gift from Feng Zhang (Addgene plasmid #62988; http://n2t.net/addgene:62988; RRID:Addgene_62988) (171). This plasmid was digested using FastDigest BpiI (Thermo Scientific™) restriction enzyme. This restriction enzyme is an isoschizomer of BbsI and cleavage at the GAAGAC(2/6)^ site of the vector, the same as BbsI. 1 µg of plasmid was digested adding 1 µl of restriction enzyme in a 1.5 ml microtube containing 5 µl of 10x Reaction Buffer (Thermo Scientific) diluted in up to 20 µl of ddH_2O. The digestion reaction was carried out for 30 minutes at 37 °C. After digestion, the linearised plasmid was dephosphorylated to avoid re-ligation. 0.6 units of Shrimp Alkaline Phosphatase (rSAP, New England Biolabs) were added to the same microtube. The dephosphorylation reaction was carried out for 30 minutes at 37 °C. All the enzymes were deactivated at 65 °C for 5 minutes after the reactions. The linearised plasmid yield was purified using column separation steps (from step 6 – Handbook) from the QIAquick Gel Extraction Kit (Qiagen). The pure linearised DNA yield was measured by wavelength absorbance (A260) using spectrophotometry (Nanodrop, Thermo Scientific).

The second vector, lentiCRISPR v2, was a gift from Feng Zhang (Addgene plasmid #52961; http://n2t.net/addgene:52961; RRID:Addgene_52961) (172). This plasmid was digested using FastDigest BsmBI (Thermo Scientific) restriction enzyme. After the digestion, the same oligos can be used as before for ligation because both restriction enzymes leave the same sticky end in both vectors. For 5 µg of plasmid were added 3 µl of restriction enzyme (BsmBI), 3 units of Shrimp Alkaline Phosphatase (rSAP, New England Biolabs), 6 µl of 10x FastDigest Buffer, and ddH_20 up to 60 µl. The reaction was carried at 37 °C for 30 minutes, and the restriction enzyme

was deactivated at 65 °C for 5 minutes. The linearised plasmid ran in 1% agarose gel diluted in Tris-acetate-EDTA (TAE) for 60 minutes at 100 V. Before gel polymerisation, 0.25 µg/ml of Nancy-520 (Sigma-Aldrich) was added. A 1 kb ladder was used as a DNA size reference. The linearised plasmid band was purified using QIAquick Gel Extraction Kit (Qiagen). The purified band had approximately 12 kb in size, as expected. The pure linearised DNA yield was measured by wavelength absorbance (A260) using spectrophotometry (Nanodrop, Thermo Scientific). After this, all steps were the same for both plasmids unless said otherwise.

2.4.2 gRNA Oligos Annealing

Four pairs (top and bottom strands) of oligos were designed to be cloned into Cas9 vectors, each pair corresponding to one single-guide RNA (gRNA) (Table 2.4). The top and bottom strands were designed separately, and an annealing step was necessary before ligating the vectors. A phosphate group was added onto the 5'-end of each strand prior ligation step. The phosphorylation and annealing were performed in the same tube. 10 µM of top and corresponding bottom strands were added to a 1.5 ml microtube containing 1x T4 DNA Ligase Buffer (New England Biolabs) and 1000 units/ml of T4 Polynucleotide Kinase (PNK, New England Biolabs) diluted in ddH_2O in a final volume of 10 µl. The mixture was incubated at 37 °C for 30 minutes. For annealing, the oligos were incubated at 95 °C for 5 minutes and cooled down 0.1 °C/s to 25 °C.

Table 2.4 Sequences of guide RNAs cloned into CRISPR vectors

Name	Oligo Strands	DNA Oligo Sequence
gRNA 1	top	CACCGGACATTTGCGAACTGGCC
	bottom	AAACGGCCAGTTCGCAAATGTCC
gRNA 2	top	CACCGGTATATCCGGACAGGTAC
	bottom	AAACGTACCTGTCCGGATATACC
gRNA 3	top	CACCGGCTCCAGGTGAGGCGAA
	bottom	AAACTTCGCCTCACCTGGAGCC
gRNA 4	top	CACCGGCCTGCCCCTCTGCACCG
	bottom	AAACCGGTGCAGAGGGGCAGGCC

2.4.3 Ligation of CRISPR guides into CRISPR vectors

The ligation reaction was performed by adding 50 ng of the digested plasmids (Methods 2.4.1) in a microtube containing 1x T4 Ligase Buffer (New England Biolabs). 376.1 pg of annealed

guide for the pSpCas9(BB)-2A-Puro (PX459) and 270 pg of annealed guide for LentiCRISPR v2 were added to the microtubes in a final volume of 10 µl diluted using ddH_2O. The mixture was incubated for 10 minutes at room temperature. After incubation, 4000 U of T4 DNA Ligase (New England Biolabs) were added to the mixture, and the reaction was carried out for 30 minutes at room temperature. The enzyme was heat-inactivated at 65 °C for 10 minutes.

2.4.4 Bacterial Transformation

E. coli XL1-Blue Subcloning-Grade Competent Cells (Agilent) were transformed with cloned CRISPR vectors (Methods 2.4.3). During the process, the bacteria were maintained in ice to avoid loss of competency. 25 ng of plasmid was added to 50 µl of bacterial suspension and left on ice for 20 minutes. The bacteria were heat-shocked at 42 °C for 45 seconds to create a pored membrane and allow plasmid transformation into the cell. After, the cells cooled down in ice for 20 minutes. 150 µl of Super Optimal broth with Catabolite repression media (SOC media, Sigma) were added, and the cells were shaken at 800 rpm at 37 °C for 2 hours on Thermomixer (Eppendorf).

2.4.5 Bacterial Culture

Following transformation, the competent cells were plated in Lysogeny broth (LB) agar containing 100 µg/ml of Carbenicillin (Fisher Scientific) and incubated at 37 °C for 16 hours. 10 random colonies were collected and expanded in LB broth containing 100 µg/ml of Carbenicillin and shaking at 250 rpm at 37 °C for 16 hours. The expanded colonies underwent Colony PCR. The colonies that presented the expected band size were purified by Miniprep Kit (Qiagen) following the manufacturer's instructions to extract their plasmid. Purified plasmids from each guide RNA were sent for Sanger Sequencing (Source Bioscience) analysis to verify correct ligation. Colonies presenting correct ligation were expanded using 1 ml of bacterial suspension in 500 ml of LB broth containing 100 µg/ml of Carbenicillin, and their plasmids were purified by Maxiprep Kit (Qiagen) to further cell transfection.

2.4.6 Colony PCR

Colony PCR was performed on colonies of bacteria after the transformation of CRISPR-cloned plasmids. The primer pairs were designed to anneal forward primer (3'-GTGGAAAGGACGAAACACCG-5') on the U6 promoter and reverse primer on guides (bottom

sequence, Table 2.4). 1 µl of bacterial suspension was added into a microtube containing 10 µl of Taq 2x Master Mix (New England Biolabs), 3 µM of the forward primer and 3 µM of the reverse primer diluted in 5 µl of ddH_2O. The samples were incubated at 94 °C for 1 minute, followed by 35 cycles of 94 °C for 20 seconds, 58 °C (annealing temperature) for 10 seconds and 72 °C for 1 kb/minute. Electrophoresis was performed to confirm the presence of 95 bp product for the pSpCas9(BB)-2A-Puro (PX459) and 260 bp for LentiCRISPR v2. The samples ran through a gel made of 1% agarose diluted in Tris-borate-EDTA (TBE) for 60 minutes at 100 V. Before gel polymerisation, 0.25 µg/ml of Nancy-520 (Sigma-Aldrich) was added to it. A 100 bp ladder was used as a DNA size reference. Positive samples related to each guide RNA were sent to Sanger Sequencing (Source Bioscience) using hU6-F (3'-GAGGGCCTATTTCCCATGATT-5)' as a sequencing primer. The traces and quality of sequencing were analysed and aligned using Geneious Prime 2020.0.5 software (Geneious). A reference sequence was created for each guide RNA to analyse the position of the insert using SnapGene software. The vector, guide RNA sequences, and the restriction enzyme used for cloning were input for the software, which simulated the correct cut position and direction of the insert to use as a reference for the sequencing analysis.

2.5 Transfection of GFP and CRISPR Vectors

Five different reagents were assessed to optimise the CRISPR vector transfection. The cells were used at 70-80% of confluency. TransIT-LT1 (Mirus), TransIT-X2 (Mirus), VIROMER RED (Mirus), VIROMER PLASMID (Mirus) and Lipofectamine 3000 (Invitrogen) were used following the respective manufacturer's protocol for each reagent. We measured the efficiency of transfection through GFP expression of pVmTHL (Addgene) and pmaxGFP (Lonza) plasmids. For electroporation, the 4D-Nucleofector System was applied using P3 Primary Cell 4D-Nucleofector X Kit (Lonza) and following the manufacturer's protocol. The program used for electroporation was the DC-100, as suggested by the company. The proportion used for transfecting cells was 3 µl of transfection reagent to 1 µg of DNA, and GFP was measured after 48 hours by fluorescence microscopy and flow cytometry in all assessments. Viable cells were stained using SYTOX AADvanced Dead Cell Stain Kit (Invitrogen) following the manufacturer's protocol.

2.6 Transduction of CRISPR Lentiviral particles to Fibroblasts

2.6.1 HEK293 transfection for Lentiviral packaging

HEK293 cells were plated in T-75 flasks and expanded in DMEM supplemented with 10% FBS. The media was changed every 2 days. When confluent, the cells were plated in two 100 mm^2 plates at a cell density of $1x10^6$ cells per plate 24 hours before transfection. For transfection, two tubes with 500 µl of Opti-MEM each were used to dilute the reagents. In the first tube, 45 µl de Lipofectamine 3000 (Invitrogen) was diluted. In the second tube, 15 µg total of vector DNA was diluted (Table 2.5). In addition to the LentiCRISPR plasmid, lentivirus packaging vectors were added to the mixture (pVSVg and pPAX2). The control group was transfected with only LentiCRISPR v2 undigested, and the knockout group was transfected with four plasmids in equal parts, one for each single-guide RNA. Then, 30 µl of P3000 Reagent was added to the diluted plasmid DNA and mixed well. The mixture was added to the diluted Lipofectamine 3000 and incubated for 15 minutes at room temperature, forming a transfection master mix. Lastly, 1 ml transfection master mix was carefully added by drops and cells were incubated for 24 hours. After 24 hours, 5 ml of DMEM supplemented with 10% FBS was added to the plate without removing the media to complete 6 ml, and the cells were incubated for another 24 hours.

Table 2.5 Amount of plasmid DNA added to transfection

Plasmid DNA	Amount
LentiCRISPR	6.6 µg
pVSVg	3.4 µg
pPAX2	5 µg

2.6.2 Fibroblast transduction with Lentiviral particles

2 ml, 1 ml, and 0.5 ml of HEK293 supernatant were added to fibroblasts to optimise the volume of supernatant-containing lentivirus particles used for transduction (data not shown). No multiplicity of infection was calculated, and the transduction efficiency was qualitatively measured by comparing the rate of survival between the transduced cells (surviving cells) and

untreated cells (no surviving cells) after Puromycin selection. Considering the results, the volume of 2 ml of supernatant containing lentivirus particles was set for transduction.

Fibroblasts were plated in 6-well plates at a cell density of $5x10^4$ cells per well 24 hours before transduction. 3 wells were used to be transduced with the CRISPR vector (single-guide RNA cloned vector), and 3 wells were used as negative controls being transduced with empty vector (no single-guide RNA cloned vector). On the day of transduction, the cell media was changed to DMEM with 10% FBS supplemented with 10 µg/ml polybrene and placed back in the incubator. In the meantime, the 6 ml supernatant from transfected HEK293 was collected into 15 ml centrifuge tubes and centrifuged at 800 g for 15 minutes to pellet cell debris from the supernatant. The supernatant was carefully collected and supplemented with 10 µg/ml polybrene. The media in the 6-well plate was removed, and 2 ml per well was added to the corresponding 3-well set of experiment (treatment or negative control). The 6-well plate containing the fibroblasts in the supernatant was centrifuged at 975 g for 90 minutes at 37 °C. Once the centrifugation finished, the 6-well plate was incubated for 2 hours at 37 °C. Finally, the supernatant was changed to fresh DMEM with 10% FBS. The process was repeated 24 hours later. Next, transduced cells were selected using 1 µg/ml Puromycin added to the media. The duration of the selection was determined by adding the same concentration of Puromycin to untreated cells, and when all untreated cells were dead, the selection media (DMEM with 10% FBS and 1 µg/ml Puromycin) was replaced by fresh growth media (DMEM with 10% FBS). Once the transduced cells were selected, they were cultured in growth media for 1 week to recover and expand. At the end of the week, cells were collected for downstream analysis.

2.7 Nucleofection of CRISPR Ribonucleoprotein to Fibroblasts

Primary cardiac foetal cells from one donor (14597) and MRC-5 cells were harvested and counted. Each nucleofection reaction requires $1x10^5$ cells, and it was performed in duplicates. Before nucleofection, 200µl of RPMI-1640 medium (Sigma) for each well to be used were pre-warmed to 37°C in two duplicated 48-well plates. Cells on the first plate were used for CRISPR validation assays. The cells of the second plate were used for single-cell transcriptomic analysis and banking. For this assay, Gene Knockout Kit v2 (Synthego) was used following the manufacturer's protocol. Briefly, ribonucleoprotein (RNP) complexes were assembled using 120pmol of 3 gRNA mix (Table 2.6) and 20pmol of Cas9 in a gRNA:Cas9 ratio of 6:1. The RNP

complex was made in P3 Primary Cell Nucleofector Solution (Lonza) in a total volume of 15µl per reaction. The cells were resuspended in 5µl per reaction at a concentration of $2x10^4$ cells/µl, added to the RNP complex mix and incubated for 10 minutes at room temperature. For negative control, only Cas9 was delivered to cells. The pmaxGFP vector (Lonza) was used instead of the RNP complex to analyse the efficiency of transfection. The program used for electroporation was EN-150 in primary cardiac fibroblasts and DC-100 in MRC-5 cells. After nucleofection, cells were recovered, and plates in pre-warmed RPMI-1640 supplemented with 10% FBS. Media was changed every two days, and cells were passaged to 24, 12 and 6-well plates when 90% confluence was reached.

Table 2.6 Sequences of guide RNA used in CRISPR RNP system

gRNA Sequences
CCGUGGGUCGGCUGGGUCGG
UAUUUCCCUGCUGUUGCGCG
UCGAUACCCAUGAUGUUGUU

2.8 CRISPR Validation Assays

Cells were collected in 500 µl of Tryzol for nucleic acid extraction. The extraction of DNA and RNA from the same samples was done by using the Direct-zol DNA/RNA Miniprep kit (Zymo Research) following the manufacturer's instructions. The nucleic acid concentrations were measured by wavelength absorbance (A260) using spectrophotometry (Nanodrop, Thermo Scientific). RNA was stored at -80 °C, and the DNA was stored at -20 °C. For protein extraction, cells were lysed using 8 M Urea with 1% CHAPS Detergent (Thermo Scientific) and scraped from the plate after 20 minutes at room temperature. The lysate was then vortexed and centrifuged at 14,000 g for 15 minutes. For measuring protein concentration, the supernatant was taken and analysed using Pierce 660 nm Protein Assay Reagent (Thermo Scientific) following the manufacturer's instructions and the plate was read at Multiskan FC Microplate Photometer (Thermo Scientific).

2.8.1 *NR2F2* Knockdown Detection

2.8.1.1 PCR amplification of the targeted region

The targeted region aimed by the designed guide RNAs is exon 5 of *NR2F2*. A pair of primers was designed to amplify the region in transfected cells with cloned LentiCRISPR v2. The forward (3' - TGGTCATTAACTGTGGAG - 5') annealed 39 bp before the beginning of the exon, and the reverse (3'- ACTTTAAGTTGCACTCAA - 5') annealed 113 bp after its end. The product for this amplification was 698 bp. For cells electroporated with RNP complexes, the *NR2F2* region was amplified using 3' – GACTCCAGCTCTGTGGCAG – 5' as forward and 3' – GCGCCGCATTCAACACAAA – 5' as reverse primers, with an amplification product of 479 bp. The input of DNA was 100 ng, and PCR was performed using the Taq 2X Master Mix (New England Biolabs) following the manufacturer's protocol (Table 2.7).

Table 2.7 PCR protocol for genomic DNA amplification

Reagent	Volume	Concentration
Taq 2x Master Mix	12.5 µl	1x
Primer Forward (10µM)	0.5 µl	0.2 µM
Primer Reverse (10µM)	0.5 µl	0.2 µM
DNA input	---	100 ng
H_2O	Up to 25 µl	---

The samples were loaded at the thermocycler to activate the reaction following the setup in Table 2.8. After the DNA amplification, the products were purified using the AMPure XP system by adding 0.6x beads to the DNA amplification product and pipetting up and down 20 times. The mixture of DNA and beads was incubated for 5 minutes at room temperature, allowing the beads to bind the DNA. Next, the mixture was exposed to a strong magnet for 2 minutes, causing the DNA-beads complexes to be pushed towards the wall of the microtube. The supernatant was then removed, and the DNA-beads complexes were washed twice with 200 µl of 80% ethanol solution. The tubes were taken from the magnet, and the DNA was eluted using 20 µl of DNA Elution Buffer (10 mM Tris, pH 8.5, 0.1 mM EDTA). Following the elution, the microtubes were exposed again to the magnet causing the DNA solution to

separate from the free beads. The supernatant containing the DNA was then transferred to a clean microtube and measured by wavelength absorbance (A260) using spectrophotometry (Nanodrop, Thermo Scientific).

Table 2.8 Thermocycler set-up for CRISPR PCR

	Initial Denaturation	PCR – 30 cycles			Final Extension
		Denaturation	Annealing	Extension	
Temperature	95 °C	95 °C	42 °C	68 °C	68 °C
Time	30 s	15 s	15 s	1 min	10 min

2.8.1.2 Detection of mutation in amplified DNA targeted region

Two assays were performed to detect mutation at the DNA level of samples. The first assay uses a T7 Endonuclease that cleaves double-strand DNA mismatches. For this, 200 ng of DNA input is added in 19 µl of 1x NEBuffer 2 (New England Biolabs). The DNA was submitted to a denaturation and hybridization cycle in a thermocycler, following the setup in Table 2.9. 1 µl of T7 Endonuclease I (New England Biolabs) was added to the mix after the hybridization step and incubated for 15 minutes at 37 °C. The reaction was stopped using 1.5 µl of 0.25 M EDTA. The analysis of the fragmented PCR products was performed on an Agilent Bioanalyzer using the High Sensitivity DNA kit (Agilent).

Table 2.9 Hybridization Condition

STEP	TEMPERATURE	RAMP RATE	TIME
Initial Denaturation	95 °C		5 minutes
Annealing	95-85 °C	-2 °C/second	
	85-25 °C	-0.1 °C/second	
Hold	4 °C		∞

The second method to detect mutation was through Sanger sequencing. 50 ng of PCR product from each sample was sent to sequencing (Source Bioscience) using 3'-TGGTCATTAACTGTGGAG-5' for LentiCRISPR transfected samples and 3'-

GCGCCGCATTCAACACAAAC-5' for RNP electroporated samples. The traces and quality of sequencing were analysed and aligned using Geneious Prime 2020.0.5 software (Geneious) and ICE Analysis Software (Synthego).

2.8.2 NR2F2 mRNA Detection

qPCR for NR2F2 was performed to detect knockdown at the transcriptomic level, using the same protocol as stated before in Methods 2.3.1.2 and 2.3.1.3. However, only GAPDH was used as the housekeeping gene. The analysis of these results was made by calculating the relative expression ($2^{-\Delta\Delta CT}$) of *NR2F2*.

2.8.3 NR2F2 Protein Detection

The detection of NR2F2 proteins was made by Western Blot analysis. 12 µg of total lysate proteins were separated by electrophoresis in Bolt 10% Bis-Tris gels (Invitrogen) using 1x Bolt MOPS SDS Running Buffer (Invitrogen), and the run was set up for 1 hour and 30 minutes on 100 V for the first hour and 150 V for the last 30 minutes. Proteins were transferred to a Polyvinylidene difluoride (PVDF) membrane using 1x Bolt Transfer Buffer (Invitrogen) at 20 V for 1 hour and 30 minutes. The quality of transfer was checked by staining the membrane with Ponceau solution. After transfer, the membrane was blocked using 5% non-fat milk in 1x Tris-buffered saline with Tween 20 (TBST, 25 mM Tris, 0.15 M NaCl, 0.05% Tween-20, pH 7.5) at 4 °C overnight. For staining, the primary antibodies diluted 1:1000 in blocking buffer were incubated for 1 hour and 30 minutes shaking at room temperature, and the secondary antibodies diluted 3:7000 in blocking buffer were incubated for 1 hour shaking at room temperature. The antibodies used are listed in Table 2.10. For revelation, the membrane was embedded in HRP Substrate Solution for 3 minutes and exposed to chemiluminescence using ChemiDoc XRS+ Gel Imaging System (Bio-Rad).

Table 2.10 List of antibodies used in Wester Blotting

Antibody	Host	Brand
Recombinant Anti-NR2F2 [EPR18443]	Rabbit	Abcam (ab211777)
GAPDH Monoclonal [1E6D9]	Mouse	Proteintech (60004-1-IG)
Polyclonal Goat Anti-Rabbit Immunoglobulins/HRP (affinity isolated)	Goat	Agilent Dako (P044801-2)
Polyclonal Goat Anti-Mouse Immunoglobulins/HRP (affinity isolated)	Goat	Agilent Dako (P044701-2)

2.9 Single-Cell RNA Sequencing

2.9.1 Cell Preparation and Encapsulation

Cells were harvested using 1x TrypLE Express (Gibco) and counted and checked for viability using the TC20 automated cell counter (Bio-Rad). After counting, cells were added to a FACS tube and washed in PBS 1%BSA. The cells were resuspended in 50 µl of Cell Staining Buffer (Biolegend) and added 5 µl Human TruStain FcX (Fc Receptor Blocking Solution, Biolegend). After an incubation of 10 minutes on ice, 5 µl of hashtag antibodies (TotalSeq-A Antibodies and Cell Hashing, Biolegend) were added to each sample, followed by a 45 µl Cell Staining Buffer (Biolegend). The hashtag antibodies allow combination and subsequently deconvolute samples in the downstream bioinformatic analysis. Following the antibody staining, the cells were washed twice with 3 ml of Cell Staining Buffer (Biolegend). On the last wash, cells were resuspended in 1 ml of PBS 1% BSA. Cells were again counted and pooled together in equal numbers. This pooled sample was filtered in a 40 µm Flowmi Cell Strainer (SP Bel-Art) and counted again. After the last counting, cells were resuspended in a concentration of ≈1.0-1.3x10^6 cells/ml. Chromium Next GEM Single Cell 3' v3.1 with Feature Barcoding technology for Cell Surface Protein User Guide (Document number CG000206 Rev D, 10x Genomics) was followed as protocol from the generation of Gel Bead-in Emulsions (GEMs) to library construction. Briefly, a Reverse Transcription (RT) Master Mix was prepared on ice containing 18.8 µl of RT Reagent B (10x Genomics), 2.4 µl of Template Switch Oligo (10x Genomics), 2µl of Reducing Agent B (10x Genomics) and 8.7µl of RT Enzyme C (10x Genomics). 31.8 µl Master Mix was added into each tube of a PCR 8-tube strip on ice with 43.2 µl of the cell suspension

to achieve ≈20.000 cells recovered in sequencing. Assemble of Chromium Next GEM Chip G (10x Genomics) was followed by attaching it to a Chromium Next GEM Secondary Holder and loading unused wells with 50% Glycerol. One well from row 1 was loaded with the combination of the Master Mix with single cell suspension previously prepared. At row 2, one well was filled with Chromium Single Cell 3' v3.1 Gel Beads (10x Genomics) previously vortex for 30 seconds, along with Partitioning Oil (10x Genomics) in one well of row 3. During this process, all reagents were carefully pipetted to avoid the presence of bubbles in any well of the chip. The chip was covered by a 10x Gasket and loaded into the Chromium Controller (10x Genomics) for GEMs formation. At the end of 18 minutes run, the chip was removed, and the GEMs were carefully transferred to a PCR tube strip on ice. A reverse transcription reaction was then performed in a thermocycler at 53 °C for 45 minutes, followed by 85 °C for 5 minutes, holding at 4 °C.

2.9.2 Post GEM-RT clean-up and cDNA Amplification

Following the reverse transcription reaction, a step of cDNA clean-up was performed using 125 µl of Recovery Agent (10x Genomics) and retaining the aqueous phase. Barcoding reagents were then removed by using Dynabead MyOne SILANE beads (10x Genomics) clean-up protocol, and cDNA was eluted in Elution Buffer I (DNA Elution Buffer, 0.1% Tweet 20 and 1% Reducing Agent B, 10X Genomics). cDNA amplification was performed by adding 65 µl of Amplification Master Mix (Table 2.11) into 35 µl of sample and incubating in a thermal cycler following the program: 98 °C for 3 minutes, and 10 cycles of 98 °C for 15 seconds, 63 °C for 20 seconds, 72 °C for 1 minute, then a last step of 72 °C for 1 minute, holding at 4 °C. SPRIselect reagent size-selection using 0.6x bead:sample ratio (60 µl) was performed to separate the 3' Gene Expression library from the Cell Surface Protein library. A magnet was used to separate the beads bonded to cDNA retained at the pellet from the supernatant containing the Cell Surface Protein library. The pellet was treated with 2 washes of 80% ethanol, and cDNA was eluted in 40 µl of DNA Elution Buffer. The supernatant was cleaned using 70 µl (2.0x) of SPRIselect reagent, followed by 2 washes of 80% ethanol in a magnet and eluted in 40 µl of DNA Elution Buffer. cDNA quality control and quantification were performed using a Bioanalyzer High Sensitivity chip (Agilent).

Table 2.11 cDNA Amplification Master Mix

cDNA Amplification Reagents (10x Genomics)	Volume
Amp Mix	50 µl
Feature cDNA Primers 2	14 µl
HTO Additive Primer v2 (0.2 µM)	1 µl

2.9.3 3' Gene Expression and Cell Surface Protein Library Construction

The library construction begins with the fragmentation of the cDNA from 3' Gene Expression followed by an A-tailing. For the Fragmentation Mix, 10 µl of Fragmentation Enzyme (10x Genomics) was added to 5 µl of Fragmentation Buffer (10x Genomics). 25% of the cDNA was mixed with 15 µl of Fragmentation Mix and DNA Elution Buffer up to 50 µl and incubated in a 4 °C pre-cooled thermal cycler with the following protocol: 32 °C for 5 minutes for fragmentation, 65 °C for 30 minutes for DNA end repair and A-tailing. Post fragmentation, a double-sided sized selection clean-up was carried out. 30 µl (0.6x) SPRIselect reagent was added to cDNA for the first step. The supernatant was transferred to a new PCR tube, and a second selection was carried in the supernatant using 10 µl (0.8x) of SPRIselect reagent. The beads containing the bound DNA were eluted in 50 µl of DNA Elution Buffer. Following clean-up, the Adaptor Ligation Mix was prepared using 20 µl of Ligation Buffer (10x Genomics), 10 µl of DNA Ligase (10x Genomics) and 20 µl of Adaptor Oligos (10x Genomics). 50 µl of Adaptor Ligation Mix was added to the sample and incubated at 20 °C for 15 minutes. After ligation, another clean-up step using 80 µl (0.8x) SPRIselect reagent was done, and DNA was eluted in 30 µl DNA Elution Buffer at the end of this process. The final step of library construction is Indexing which allows multiplexed sequencing to run with both Cell Surface Protein and 3' Gene Expression libraries. For this step, 60 µl of Sample Index PCR mix was added to 30 µl of each library. Then, 10 µl of an individual Single Index from Single Index Plate T Set A well ID (10x Genomics). After, the samples were incubated in a thermocycler with the protocol stated in Table 2.12. Steps 2, 3 and 4 were done in cycles. For the primary fibroblasts' characterization, it was used 12 cycles, and for the edited fibroblasts' analysis, it was used 11 cycles. Lastly, a double-sided sized selection clean-up was carried out. 60 µl (0.6x) SPRIselect reagent was added to each library after PCR. The supernatant was transferred to a new PCR tube, and a

second selection was carried in the supernatant using 20 µl (0.8x) of SPRIselect reagent. The beads containing the bound DNA were eluted in 35 µl of DNA Elution Buffer. At the end of the library preparation, their quality was assessed by the Bioanalyzer High Sensitivity chip (Agilent).

Table 2.12 Thermal cycler protocol for Sample Index PCR

Temperature	Time
98 °C	45 seconds
98 °C	20 seconds
54 °C	30 seconds
72 °C	20 seconds
72 °C	1 minute

2.9.4 Data Sequencing

Cell Surface Protein and 3' Gene Expression libraries were diluted to 2 nM in ddH_2O. 15 µl of Cell Surface Protein diluted library was pooled together with 85 µl of 3' Gene Expression library. DNA from the pooled library was denatured by hydrogen bond break using 0.2 N of NaOH (Sigma) for 5 minutes at room temperature. The reaction was neutralised using 200 mM of Tris-HCl (pH 7). The denatured library was diluted to 20 pM using prechilled HT1 solution (Illumina) and, again, until final dilution to 2 pM in 1.3 ml of solution. The library was then sequenced using NextSeq 500 platform (Illumina) set up to paired-end read with 28 cycles for Read 1, 8 cycles for Index 1 and 52 cycles for Read 2. No cycles were run in Index 2.

2.9.5 Data Processing

Codes are described in Appendices A.1 and C.1 for primary migrating cells and CRISPR/Cas9 edited fibroblasts samples, respectively.

2.9.5.1 Read Alignment and Demultiplexing

The raw base calls (BCL) files obtained from the sequencer were demultiplexed into FASTAQ files using bcl2fastaq v2.20 software (Illumina). For the gene expression matrix readout, the reads were aligned to the genome using the Cell Ranger 6.1.2 (10x Genomics) pipeline

generating a “cell x genes” count matrix. For the hashtag’s matrix readout, the hashtag reads were aligned to a reference file containing the antibodies identification and the nucleotide sequence for each, using the kallisto -bustools scRNA-seq workflow (173). Briefly, the ‘kallisto bus’ command was used to pseudoalign the reads against a reference index. The output is a BUS file entry for every unique feature barcode, UMI and set combination. Finally, the BUS file is converted to a “cells × hashtag” count matrix by collapsing counts from each feature barcode with those corresponding to its Hamming distance 1 mismatch.

To demultiplex the samples pooled together during cell preparation, Hashsolo (174) was used to label each cell to the hashtag assigned to its sample. Table 2.13 shows the assigned hashtag to the samples sequenced to determine the phenotype of migrating cells from heart explants. Table 2.14 shows the assigned hashtags to the samples sequenced to analyse perturbations after NR2F2 knockout via CRISPR-Cas9.

Table 2.13 Hashtags assigned to each sample sequenced to determine cell types in migrating cell population of heart explants

Sample	Hashtag
Atria	TotalSeq™-A0259 anti-human Hashtag 9 (BioLegend)
Ventricle	TotalSeq™-A0260 anti-human Hashtag 10 (BioLegend)
Outflow Tract	TotalSeq™-A0262 anti-human Hashtag 12 (BioLegend)
Whole heart	TotalSeq™-A0263 anti-human Hashtag 13 (BioLegend)

Table 2.14 Hashtags assigned to each sample sequenced for CRISPR-Cas9 knockout perturbation analysis

Sample	Hashtag
MRC-5 CNTRL	TotalSeq™-A0259 anti-human Hashtag 9 (BioLegend)
MRC-5 KD	TotalSeq™-A0260 anti-human Hashtag 10 (BioLegend)
PFB CNTRL	TotalSeq™-A0262 anti-human Hashtag 12 (BioLegend)
PFB KD	TotalSeq™-A0263 anti-human Hashtag 13 (BioLegend)

PFB: primary fibroblasts from foetal hearts; MRC-5: foetal lung fibroblast cell line; CNTRL: control; KD: NR2F2-knockeddown.

2.9.5.2 Data Filtering and Normalisation

The pre-processing and data visualisation was done using Scanpy (175) toolkit. Firstly, cell doublets were identified by Scrublet (176) and subsequently filtered out. After hashing, "ambient" RNA was taken out using EmptyDrops (177), and the cell count matrix was filtered to remove low-quality cells. Genes expressed in less than 5 cells were also excluded and stressed or dying cells were filtered out based on fraction counts of mitochondrial genes (10% or more). The normalisation of counts and highly variable genes selection were carried out using analytic Pearson residuals (178).

2.9.5.3 Clustering and Visualisation

Graph-based clustering was performed using the Leiden-based method (179) on a neighbourhood graph (n_neighbours=10) projected from a number of principal components (n_PCA) of a Principal Component Analysis (PCA) on the data log-transformed normalised counts. The data was dimensionally reduced and visualised by Uniform Manifold Approximation and Projection (UMAP) (180). Table 2.15 summarises the settings used in the data processing steps.

Table 2.15 Summary of settings used in scRNA-sequencing data processing

Sample	Read count filtering	Gene count filtering	n_PCA	Leiden resolution
Primary migrating cells	Min: 500 Max: 25,000	Min: 500	20	0.1
CRISPR/Cas9 edited fibroblasts	Min: 500 Max: 50,000	Min: 500	50	0.2

Min: minimum of read or gene counts to consider good quality cells for downstream analysis; Max: minimum of read or gene counts to consider good quality cells for downstream analysis; n_PCA: number of principal components selected for clustering the data.

2.10 Public Dataset Analysis

2.10.1 Bulk RNA-sequencing

A public dataset (Accession Number: GSE81585) was obtained from GEO as a series matrix. This dataset was generated in a study using cardiomyocytes derived from induced pluripotent stem cells (iPS-CMs)(181). The expression matrix was analysed using BioJupies (182). For data normalisation, raw counts were normalised to log10-Counts Per Million (logCPM) by dividing

each column by the total sum of its counts, multiplying it by 10^6, followed by the application of a log10-transform. The differentially expressed genes (DEGs) analysis was carried out using the *limma* R package (183), and the genes were selected based on log fold-change (logFC) of the group of perturbed samples in contrast with the control group (|logFC|≥ 1.5). The gene ontology of the DEGs was assessed by ShinyGO (v0.76.1)(184). Proteins related to DEGs were used as input to the Search Tool for Retrieval of Interacting Gene/Proteins database (STRING v11) to evaluate protein-protein interactions. STRING database gives the protein-protein interaction network of the input list by combining information from publicly available sources with computational predictions (185). All interactions were set at a medium level of confidence (all interaction sources active; combined score ≥ 0.40). According to STRING, a score of 0.5 indicates that every second interaction is a false positive. For Gene Ontology Enrichment analysis, STRING uses the defined functional categories of the Clusters of Orthologous Groups of proteins (COGs) database, and the terms are ranked by the false discovery rate (FDR).

2.10.2 Single-cell RNA-sequencing

A public dataset (BioProject ID: PRJNA576243) was obtained from Sequence Read Archive (SRA). This dataset comprised transcriptomes of 3 healthy foetal human hearts of ages 19 and 22 gestational weeks sequenced at the single-cell level (148). Cell barcode error correction was performed using the 10x barcode whitelist (186). Subsequently, sequenced reads were processed with kallisto-bustools' scRNA-seq workflow (173). The sequencing quality control and analysis were made using the Scanpy toolkit (175). 5861 cells were obtained after filtration. Genes expressed in less than 10 cells were excluded. Stressed or dying cells were filtered based on fraction counts of mitochondrial genes (10% or more). The batch effects were removed by BBKNN (187). Counts were normalised using scran (188). UMAP (n_neighbours=20; n_pca=20) was run for visualisation, and the clustering of each cell type was performed using the Leiden-based method (189) at a resolution = 0.5. Clusters 4, 5, 7 and 9 from Leiden clustering had their resolution increased by 0.1. Clusters 0, 10 and 14 were excluded due high expression of ambient mRNA of erythroblast genes, as was done in the original paper (148). Cell types were identified by known marker genes from the original study (148).

Chapter 3 Results

Isolation and Characterisation of Migrating Cells from Foetal Hearts

3.1 Introduction

3.1.1 Identifying cells of the heart

The heart is formed by a complex heterogeneous population of cells. Its tissues are composed of four main cardiac cell types: endothelial cells, fibroblasts, smooth muscle cells and cardiomyocytes (142,190). The identification of this diverse group of cells is made through morphology, location in the tissue, genetic blueprint, and protein expression.

Cardiomyocytes are contracting cells responsible for the rhythmical beating of the heart. Their morphology in culture is marked by a polygonal flatten shape, with cytoplasmatic protrusions (191). Ultra-structurally, cardiomyocytes are characterised by sarcomeres and express the sarcomeric proteins Myosin-6 (α MyHC), and Troponins T (TnT) and I (TnI) which are frequently used to identify these cells (148,192). Regarding transcriptomic signature, studies in the adult human heart define the cardiomyocyte population from other cell types by enrichment of myosin and troponin genes (*MYL2, MYL3, MYL4, MYL7, MYL9, MYH7, MYH7B, MYH6*), and Ca^{+2}, Na^{+} and K^{+} channels genes (*SLC8A1, CACNA1C, KCNA6*) (146,147,157). These ion channels participate in the polarization and depolarisation of the cell membrane that triggers cell contraction (193). Transcription factors involved in the regulation of cardiogenesis and cardiomyocyte differentiation from cardiac progenitors such as Nkx2-5, Tbx5 and Gata6 are also frequently detected in these cells, although they are not specific to them (194).

In addition to cardiomyocytes, the non-myocyte cell types play important roles in the heart like vascularisation, extracellular matrix organisation, and cell-to-cell communication, among other tasks related to heart homeostasis and development (195).

Endothelial cells are responsible for the inner lining of vessels and heart chambers. Although their morphology and markers can vary, endothelial cells share common features. Pan-markers for endothelial cells include platelet endothelial cell adhesion molecule (*PECAM-1*, CD31) (196), a transmembrane glycoprotein located at the intercellular clefts that participate

in the establishment and maintenance of the vascular barrier, and cadherin-5 (*CDH5*), an adhesion molecule involved in e cytoskeleton organisation in cells (197–200). Another common element in endothelial cells is Weibel–Palade bodies, organelles that can also be used to identify these cells in culture and are responsible for the storage of von Willebrand factor (*VWF*) that is secreted by endothelial cells (201). In the human developing heart, other endothelial cell markers are related to vasculogenesis such as TEK receptor tyrosine kinase (*TEK*), Vascular endothelial growth factor receptor 2 (*VEGFR2)* and E-selectin (*SELE*) (202).

Smooth muscle cells are in the tunica media of vessel walls and are responsible for vasculature structure, smooth muscle contraction and extracellular matrix production (203). Because of their contractility, smooth muscle cells express muscle-contracting proteins such as smooth muscle myosin (*MYH11*), caldesmon (*CALD1*), myocardin (*MYOCD*) and vascular smooth muscle actin (αSMA, *ACTA2*) (203,204). Other markers are related to cytoskeleton organisation: Transgelin (*TGLN*) is ubiquitously expressed in smooth muscle cells and an early marker of smooth muscle differentiation (205).

Fibroblasts play an important role in cardiac tissue integrity by organizing the extracellular matrix. However, their identification is challenging due to their phenotype heterogeneity and lack of a specific marker. Morphologically, cardiac fibroblasts present an oval nucleus, a spindle-shaped form, and a flat granular cytoplasm with multiple sheet-like extensions (206,207). Vimentin (*VIM*) has been commonly used to identify fibroblast populations (208–210). Despite its wide use, Vimentin is vastly expressed in endothelial cells and smooth muscle cells (201,211). Like vimentin, thymocyte differentiation antigen 1 (*THY1*) is broadly used as a marker but is also present in other cell types, such as endothelial cells and leukocytes (212–214). Therefore, fibroblast identification is made based on morphology and a combination of unspecific markers and the exclusion of markers specific to other cell types.

3.1.2 NR2F2 expression in the developing heart

NR2F2 is a transcription factor that plays a key role during organogenesis, including the heart. In E8 mice, Nr2f2 is detected in early embryonic structures as the cardiac crescent and the lateral mesoderm (99). Later, it is expressed in other heart structures, including the sinus venosus and common atria, endocardium and myocardium of the atrium, endocardium of the ventricle and atrioventricular canal (99). The expression in the atria and ventricles is seen

throughout cardiac development (99). In human foetal hearts, areas of the cardiovascular system expressing NR2F2 comprise the atria, aorta, and coronary and lymphatic vessels (101). These data suggest that NR2F2 has a role from early cardiogenesis until late septation, valvulogenesis and coronary vasculature formation.

At the cellular level, NR2F2 is co-localized with markers of endothelial cells, smooth muscle cells and cardiomyocytes (101). Also, unpublished transcriptomic data from our group showed mRNA for *NR2F2* being highly expressed in fibroblasts and endothelial cells compared to other cardiac residing cell types (Wilson DI, 2018, personal communication).

Thus, this chapter aims to isolate migrating cells from the heart and assess their phenotypes and NR2F2 levels of expression to investigate further the NR2F2 role in fibroblasts during heart development and disease.

3.2 Cardiac migrating cells are isolated from tissue explant.

Human foetal tissue was provided by the Newcastle HDBR (Methods 2.1.1). Tissue was obtained from 14 hearts. In 9 hearts, dissection was performed that enabled separate chambers to be cultured, and in 5 hearts, culture was performed using the whole heart (Methods 2.1.2). The foetal ages of these hearts are shown in Table 3.1.

Cardiac migrating cells were isolated from foetal hearts using a protocol that combines the expansion of migrating cells from tissue explant and isolation by a filtering method (215). The tissue explants were cultured for 7 to 10 days in media to allow the cells to migrate out of it (Methods 2.1.3). After this, the tissue explants were filtered out and migrating cells were maintained in culture for expanding and downstream analysis (Methods 2.1.4 and 2.1.5). Migrating cells were visible after 3 days of explant culture (Figure 3.1A). The cells had an elongated flattened form when in monolayer (Figure 3.2B). Interestingly, cells isolated from atria, ventricle and outflow tract presented different morphology between them (Figure 3.2).

Table 3.1 Foetal heart summary information

Sample ID	Foetal Age	Outflow Tract	Atria	Ventricles	Whole Heart
14552	L8pcw	+	+	+	-
14553	9pcw	+	+	+	-
14555	9pcw	+	+	+	-
14562	9pcw	+	+	+	-
14576	L8pcw	+	+	+	-
14579	CS23	-	-	-	+
14578	10pcw	-	-	-	+
14577	11pcw	-	-	-	+
14580	9pcw	-	-	-	+
14590	CS23	-	-	-	+
14591	L8pcw	+	+	+	-
14592	CS21	+	+	+	-
14596	L8pcw	+	+	+	-
14597	11pcw	+	+	+	-

-/+ indicates which area of the heart or if the entire heart was used to isolate migrating cells. CS: Carnegie Staging; L: late; pcw: post conception weeks.

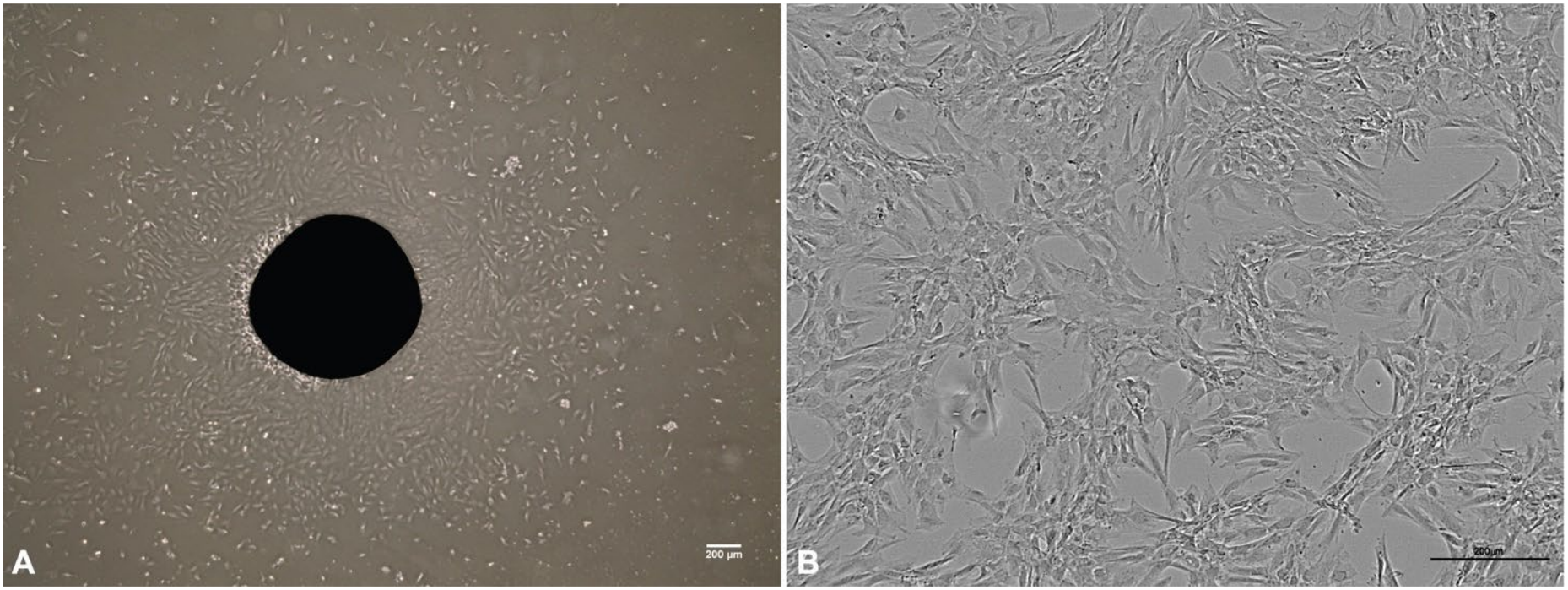


Figure 3.1 Isolation of Migrating Cells from Cardiac Tissue

(A) A ring of cells migrating out of the tissue explant after 3 days in culture using a gelatin-covered plate. (B) After cardiac tissue explant removal, cells have a spindle-shaped morphology with protrusions and flatted cytoplasm at day 5. Plates were not treated with gelatin during expansion. N=14. Scale bars: 200µm

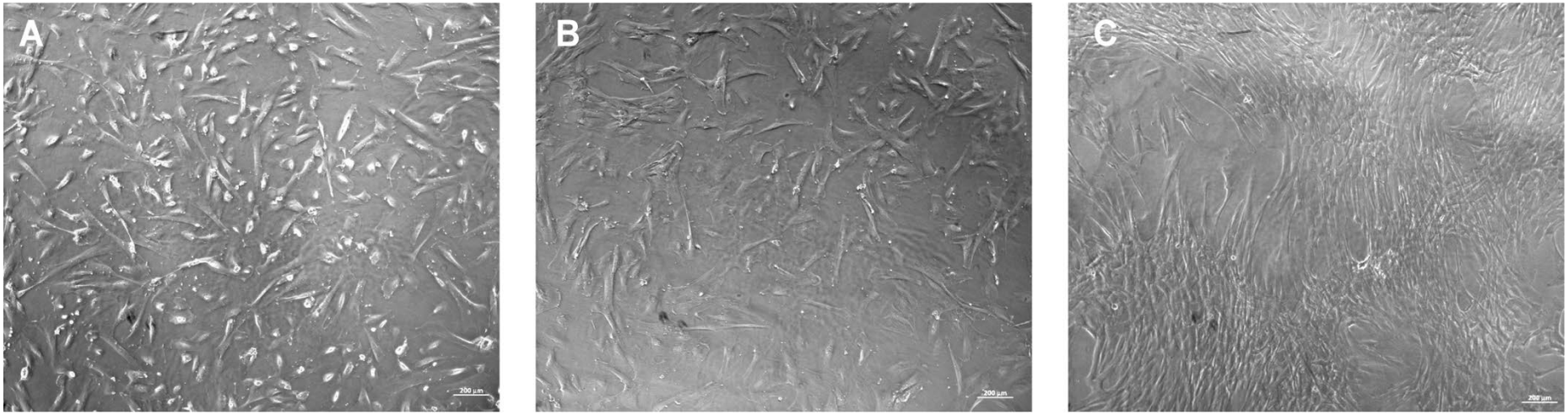


Figure 3.2 Migrating cells from cardiac explant derived from atria, ventricles, and outflow tract

(A) Atrial cells (n=1) (B) Ventricular (n=1), and (C) Outflow tract cells (n=1) cultivated in T-25 plates with no gelatin-coating treatment. Scale bars: 200µm.

3.3 Determining the cellular phenotype of the migrating cells

To determine the cellular composition of migrating cells - in particular, to assess the fraction of fibroblasts, cardiomyocytes, and endothelial cells - flow cytometry analysis was performed (Methods 2.2). Cells from 5 heart samples and a lung fibroblast cell line (MRC-5) were analysed. Figure 3.3A show the gating strategy used to identify the population of viable cells. Singlets were identified based on the height (FSC-H) and area (FSC-A) of the signal produced by the event size. From the singlets gate, cells were identified based on event size (Forward Scatter, FSC) and granularity (Side Scatter, SSC). Viable cells were selected based on the negative signal of Zombie Violet™ Fixable Viability staining (Live/Dead). Figure 3.3B demonstrate the gating strategy to assess the expression of cardiac cell markers Thy-1 (CD90), Vimentin, Myosin-6 (αMyHC) and PECAM-1 (CD31).

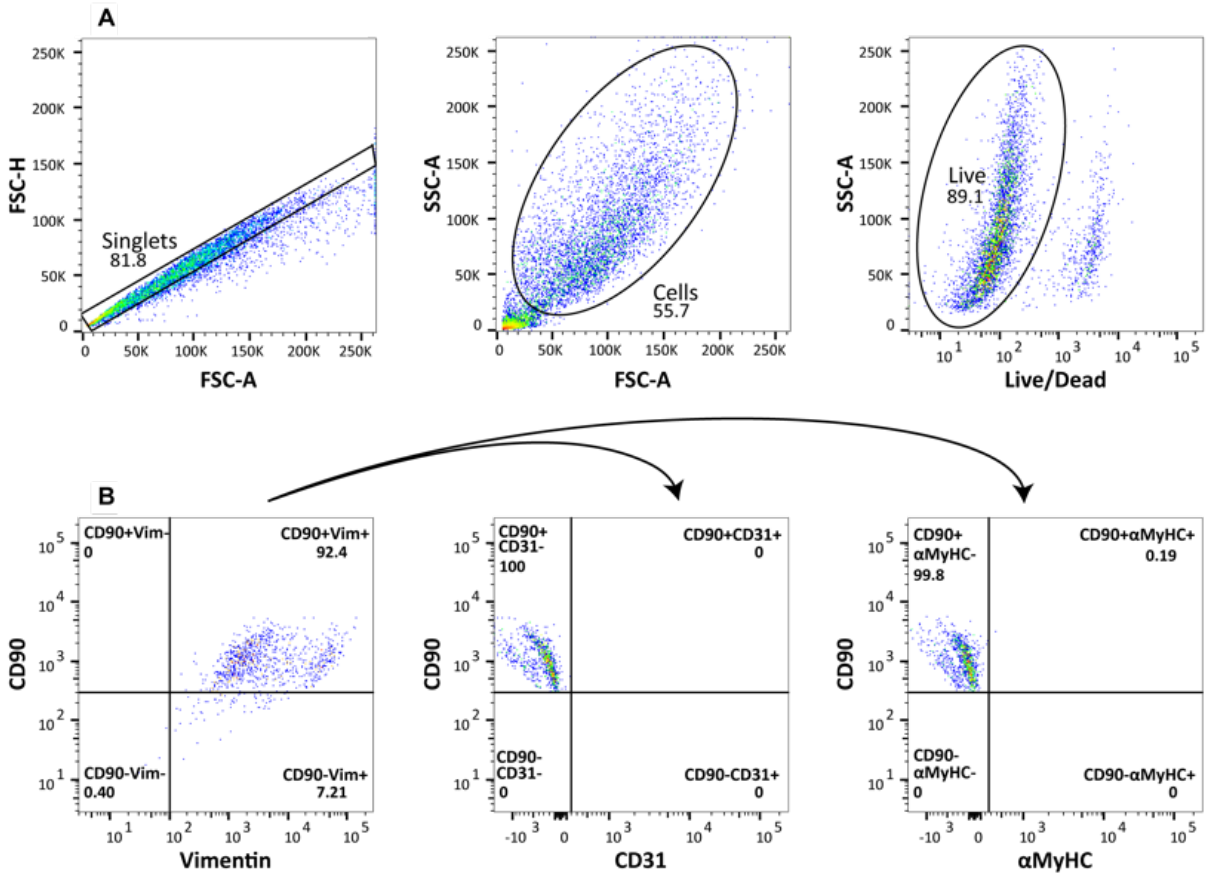


Figure 3.3 Gating strategy to determine cardiac cell populations by flow cytometry

(A) FSC-A x FSC-H was used to define singlets, FSC-A x SSC -A was used to identify the cellular population and SSC-A x Live/Dead to select viable cells – negative population. (B) Cell type populations were determined by co-expression of Thy-1(CD90) and Vimentin, and from gating, the co-expression of Thy-1 (CD90) and endothelial (CD31) or cardiomyocyte (αMyHC) markers.

To determine the fibroblast population, the expression of CD90 and Vimentin was assessed by flow cytometry in each of the 5 migrating cell samples tested, and MRC-5 cells. Fibroblasts are determined by the populations $CD90^+$ Vim^+ and $CD90^-Vim^+$. Figure 3.4 and Table 3.2 show the results of this assessment. Vimentin was highly expressed in all samples, and CD90 was present in most samples, including MRC-5. However, CD90 was virtually absent in FB14590 and expressed in lower frequency in FB14580.

Table 3.2 Percentage of $CD90^+$ Vim^+ and $CD90^-$ Vim^+ populations in migrating cells

Sample ID	$CD90^+$ Vim^+	$CD90^-Vim^+$
FB14577	78.7	18.4
FB14579	73.9	25.9
FB14580	25.9	68.8
FB14590	1.22	97
FB14562	99.1	0.38
MRC-5	92.8	6.82

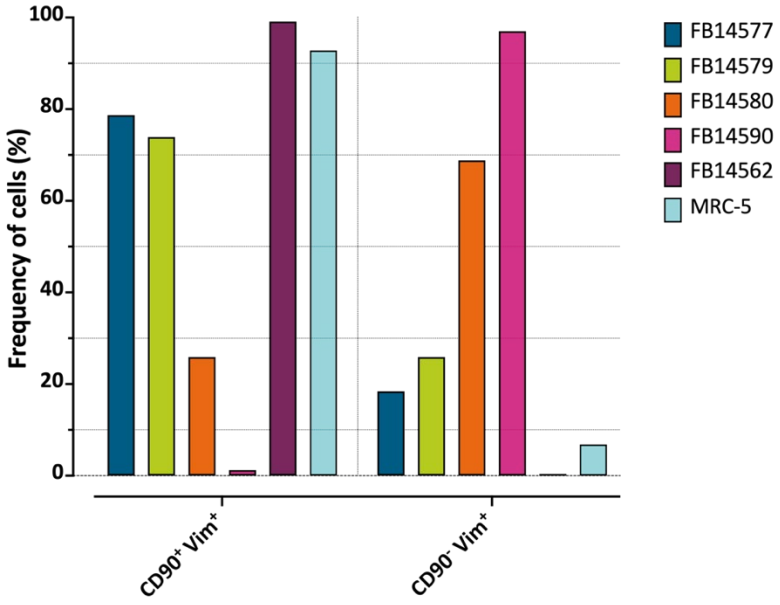


Figure 3.4 Migrating cells and MRC-5 expression of CD90 and Vimentin proteins

The bars represent the percentage of live cells (y-axis) in each population (x-axis). The legend shows the colours of the bar that are related to each sample analysed. Vim: Vimentin.

Since fibroblasts do not have specific protein markers, other cell types' specific protein markers are used to rule out their presence. To assess the presence of endothelial cells and cardiomyocytes, the expression of CD31 and αMyHC, respectively, was measured in the

previous populations ($CD90^+Vim^+CD31^+$, $CD90^+Vim^+\alpha MyHC^+$ and $CD90^-Vim^+\alpha MyHC^+$). Table 3.3 and Figure 3.5 show the results of this assessment. The lung fibroblast cell line (MRC-5) and 2 of the primary cardiac samples (FB14577 and FB14580) had fibroblasts as the predominant cell type identified by the cell markers. FB14579 and FB14590 had cells presenting cardiomyocyte markers, in contrast to FB14562 having cells with endothelial markers.

Table 3.3 Percentage of CD31 and αMyHC expressed in $CD90^+Vim^+$ and $CD90^-Vim^+$ populations

	From $CD90^+Vim^+$ population		From $CD90^-Vim^+$ population
Sample ID	**$CD31^+$**	**$\alpha MyHC^+$**	**$\alpha MyHC^+$**
FB14577	0	0.29	0.028
FB14579	16.1	56.9	85.8
FB14580	0.32	2.02	4.65
FB14590	6.45	34.1	35.6
FB14562	58.4	0	10.8
MRC-5	0	0.19	0

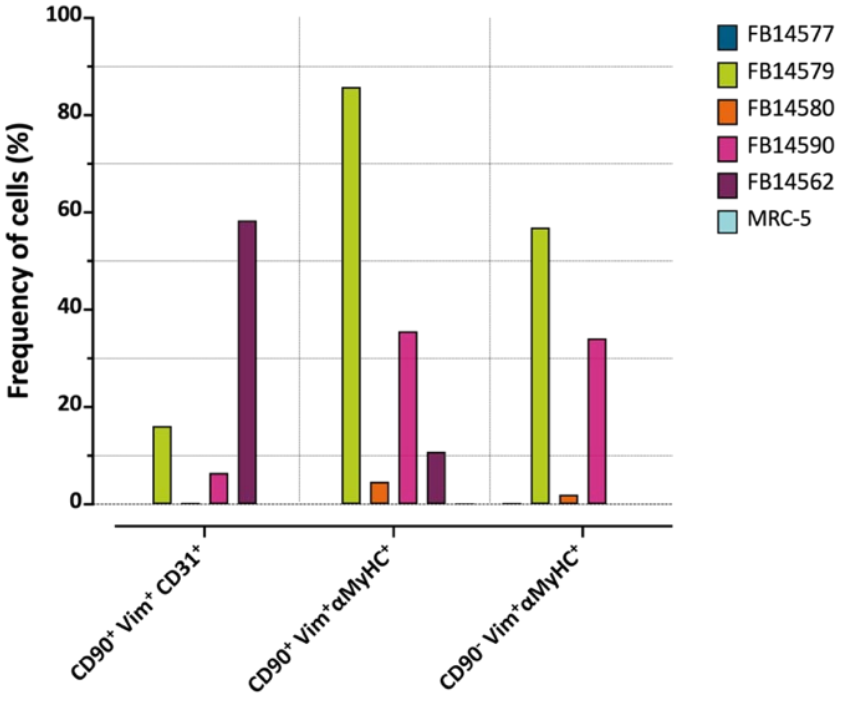


Figure 3.5 Expression of CD31 and αMyHC proteins in the $CD90^+Vim^+$ and $CD90^-Vim^+$ populations of migrating cells, and MRC-5

The bars represent the percentage of cells (y-axis) expressing CD31 or αMyHC in $CD90^+Vim^+$ or $CD90^-Vim^+$ populations (x-axis). The legend shows the colours of the bar that are related to each sample analysed. Vim: Vimentin.

To assess the abundance of different cell types by transcript profile, single-cell RNA sequencing was performed. Figure 3.6 demonstrates the schematic workflow analysis for this assessment (Methods 2.9.5.1). The binary base call (BCL) files generated by the sequencer were converted into FASTAQ files to be used for count matrices writing after transcriptome alignment. Sequentially, the cells with read count lower than 500 and higher than 25,000, and gene detection lower than 500 were filtered out due to low quality. Genes expressed in less than 5 cells were also filtered out. After normalization, the data was dimensionally reduced and clustered (Methods 2.9.5.2). Graph-based clustering was performed using the Leiden-based method (179) projected from the first 20 PCA on this dataset of log-transformed normalised counts. The reduced data was plotted in Uniform Manifold Approximation and Projection (UMAP) graphs (Methods 2.9.5.3). Figure 3.7 shows the results of these steps that precede the cell type assessment in the migrating cell populations. Cells from two foetal hearts were used (14562 and 14597). From 14562, cells from atria, ventricles and outflow tract were cultured and handled separately; from 14597 the entire heart was used to isolate cells for single-cell RNA sequencing.

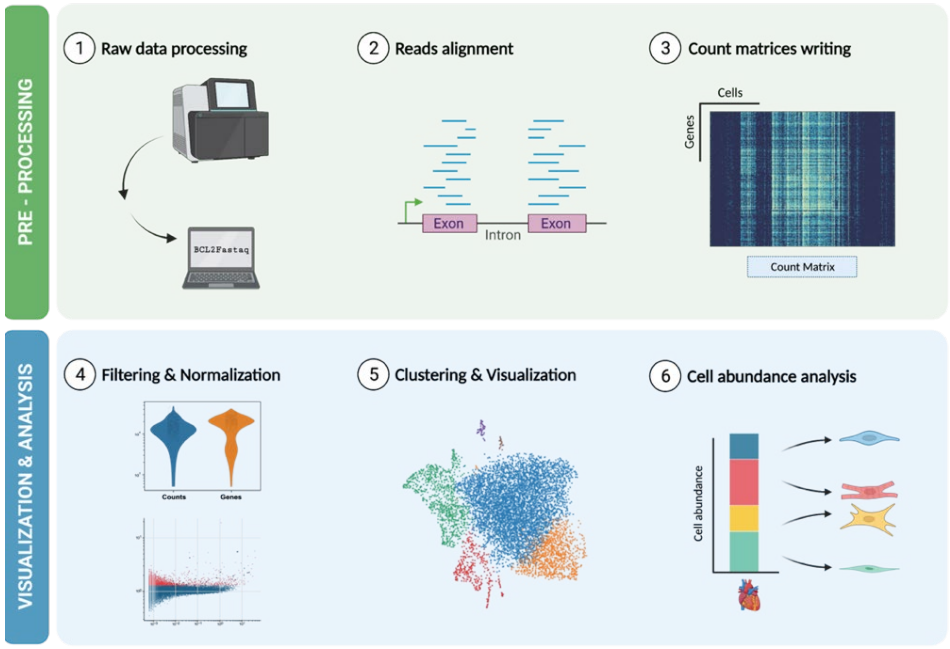


Figure 3.6 Workflow of the scRNA-sequencing analysis in cardiac migrating cells

In the pre-processing step, the raw data (BLC files) is converted into Fastaq files (1). After the file conversion, the Fastaq files are used to align the reads into the human transcriptome reference (2) generating a count matrix with gene ID in the row and cell ID in the columns (3). To visualise the data, the filtering and normalization were carried out to remove low-quality cells and sequencing depth bias, respectively (4), followed by clustering and UMAP plotting (5). The final step is the analysis of the data by assessing the abundance of each cell type in the dataset (6). Created with BioRender.com.

In total, 7700 cells and 20528 genes passed the filtering criteria and were used in the downstream pipeline. The cells and the genes were evenly distributed among the four samples - Atria, Outflow tract, Ventricle and Whole heart (Figure 3.7A-C, Table 3.4). PanglaoDB (216) was used to annotate cell clusters. PanglaoDB is a prior knowledge database of cell gene markers accessed via the Decoupler library (217). To determine the annotation of each cell, Decoupler runs Fisher's exact tests that measure the enrichment of gene markers from a cell type in each cell from the dataset. This analysis selects the top expressed genes for each cell and tests if a collection of marker genes from cell types are enriched resulting in a score based on the *p*-value.

The automatic annotation of PanglaoDB showed that the cells were mostly fibroblasts when compared to the other main cardiac cells, in all samples analysed (whole heart, outflow tract, atria, and ventricles) (Figure 3.8). Leiden r=0.1 was used as a clustering method and the clusters could be visualised in UMAP plots (Figure 3.7F).

As a proof-of-concept of the automatic annotation, individual gene markers for cardiomyocytes, endothelial cells, fibroblasts, and smooth muscle cells were assessed. The gene markers were determined by a study in human foetal hearts (age 19-22 pcw) (148), and they are listed in Table 3.5. It observed an even level of expression of the gene markers through the dataset. Also, all the presented fibroblast gene markers were expressed. One gene marker for Cardiomyocytes (*MYL9)* and two gene markers for smooth muscle cells (*TGLN* and *ACTA2*) were also expressed in the cardiac migrating cells. No gene marker of endothelial cells was expressed in these cells (Figure 3.9).

Table 3.4 Age and number of cells per sample

Sample ID	ID (age)	Number of cells
Ventricle	14562 (9pcw)	2996
Outflow Tract	14562 (9pcw)	2141
Atria	14562 (9pcw)	1425
Whole Heart	14597 (11pcw)	1138

Pcw: post conception weeks

Table 3.5 Gene markers used to identify cardiac cell types

Cell type	Marker genes
Cardiomyocytes	*MYL3, MYL4, MYL7, MYL9, MYH7B, TNNT2, TNNC1, RYR2, TTN, PLN*
Endothelial cells	*CDH5, PECAM-1*
Fibroblasts	*COL1A1, COL1A2, COL3A1, COL1A2, TGFBI, DCN, LUM, FN1, BGN*
Smooth muscle cells	*ACTA2, TAGLN, RGS5*

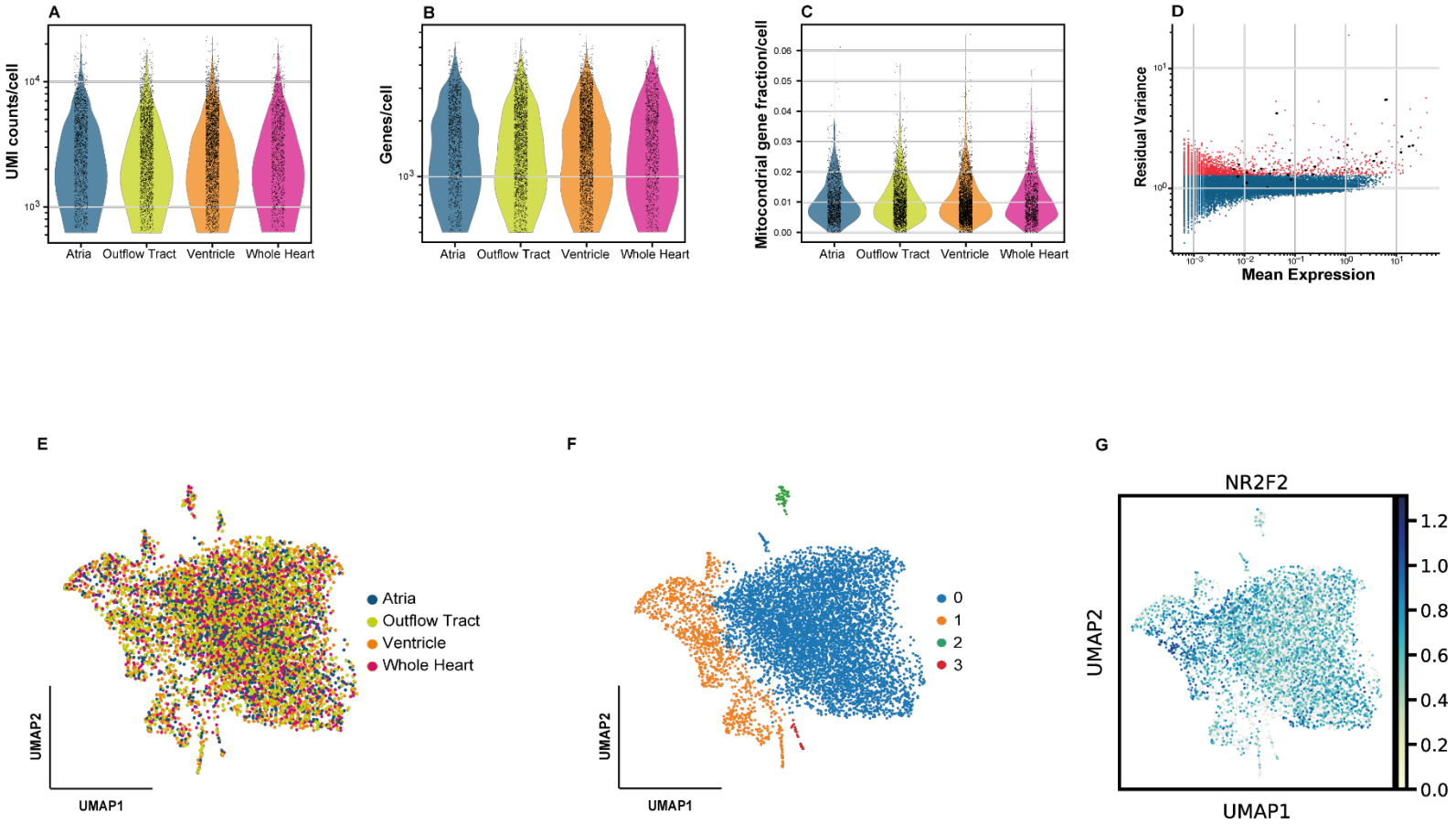


Figure 3.7 Quality control and visualisation of cardiac migrating cell single-cell RNA-sequencing dataset.

(A-C) Violin plots show the distribution of UMI counts (A), genes (B) and mitochondrial gene fraction (C) in each cell. Each black dot is a cell and the values for these parameters are plotted on the y-axis. The sample ID is on the x-axis. (D) After normalisation, most of the marker genes (black dots) were selected as highly variable genes (red dots). (E-F) To visualise the data, UMAP was used to reduce it and plot it in two dimensions. Leiden (r=0.1) was used to calculate the clusters (F) and the deconvolution of the hashtag staining during cell preparation was applied to identify the cells from each sample (E). (G) UMAP visualisation of *NR2F2* expression across each cell (dots). The expression abundance is coloured according to the colour bar and the expression values show the normalised counts of *NR2F2*.

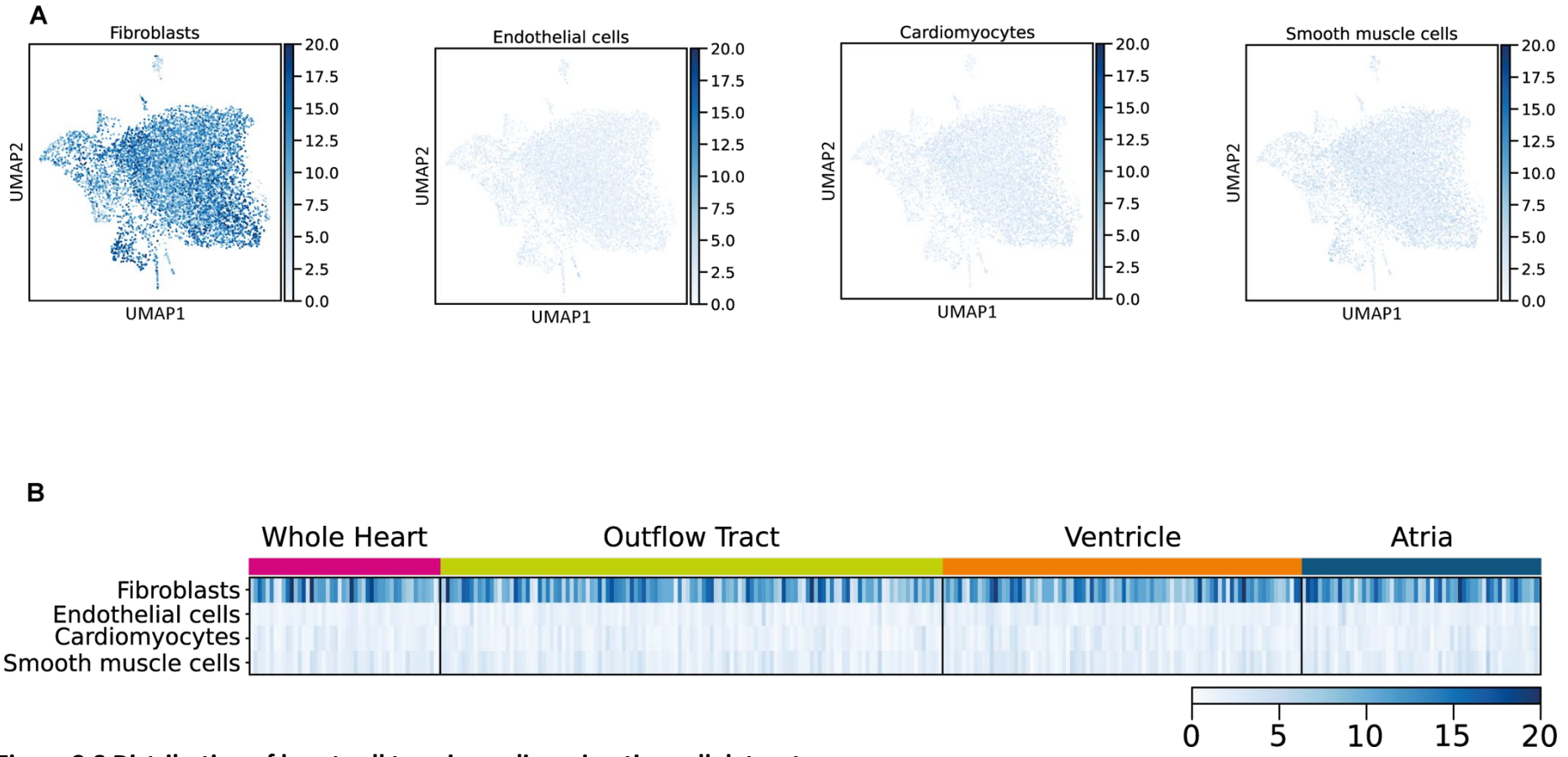


Figure 3.8 Distribution of heart cell type in cardiac migrating cell dataset.

(A) UMAP plots of selected cardiac cell types. Each cell (dots) is coloured according to Fisher's exact test score (-log10[*p*-value]) for fibroblasts, endothelial cells, cardiomyocytes, and smooth muscle cells. (B) Heatmap plot colours representing Fisher's exact test scores for each cell clustered by regions of the foetal heart from where the cells were isolated.

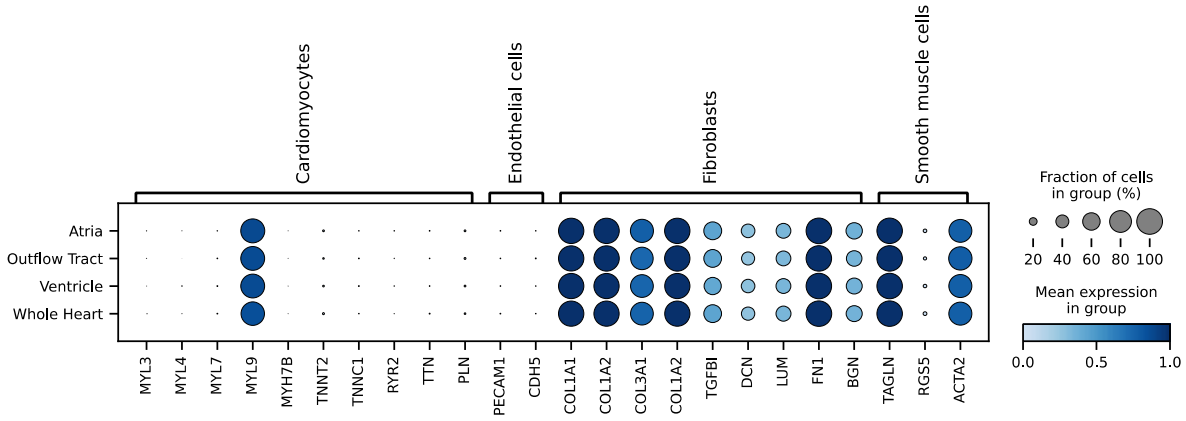


Figure 3.9 Expression of gene markers of cardiac cell types in migrating cardiac cells

Dot plot showing the expression of gene markers to identify cardiomyocytes, endothelial cells, fibroblasts, and smooth muscle cells. Each dot represents the percentage of cells (size of the dot) expressing the gene (x-axis) and the mean expression (intensity of the colour) within each cluster (y-axis).

3.4 *NR2F2* mRNA transcripts are expressed in migrating cells from foetal heart

Two housekeeping genes, *GAPDH* and *ACTB*, were assessed as endogenous control and the data is shown in Figure 3.10. *ACTB* expression was variable between samples, but GAPDH had consistent levels and was therefore used as the control housekeeping gene for NR2F2 qPCR assays. The CT values of this assessment are shown in Table 3.6.

Table 3.6 CT values from qPCR to compare housekeeping genes (*GAPDH* and *ACTB*) used to measure *NR2F2* mRNA relative expression

	NR2F2			***GAPDH***					***ACTB***				
Sample	**CT_1**	**CT_2**	**CT_{Mean}**	**CT_1**	**CT_2**	**CT_{Mean}**	**ΔCT**	**$2^{-\Delta CT}$**	**CT_1**	**CT_2**	**CT_{Mean}**	**ΔCT**	**$2^{-\Delta CT}$**
MRC-5	29.0	28.3	28.7	22.8	22.1	22.4	**6.2**	**0.013**	25.5	24.2	24.8	**3.8**	**0.069**
FB14578	28.5	28.7	28.6	22.2	21.7	22.0	**6.7**	**0.010**	24.4	23.9	24.2	**4.4**	**0.046**
FB14579	28.4	28.2	28.3	21.8	21.4	21.6	**6.7**	**0.010**	22.6	22.2	22.4	**5.9**	**0.017**
FB14580	28.3	28.4	28.3	22.0	22.1	22.0	**6.3**	**0.013**	21.9	22.1	22.0	**6.3**	**0.013**
FB14590	37.5	36.6	37.1	30.0	30.6	30.3	**6.8**	**0.009**	29.8	29.8	29.8	**7.3**	**0.007**

CT: Number of cycles required for the florescent signal of the amplified gene to exceed baseline; Cycle Threshold; CT_{Mean} = (CT_1+CT_2) ÷ 2; CT_1: CT value of the duplicate 1; CT_2: CT value of duplicate 2; ΔCT = *NR2F2* CT_{Mean} - Housekeeping gene (*GAPDH/ACTB*) CT_{Mean}; $2^{-\Delta CT}$: Relative Expression of *NR2F2* mRNA for each housekeeping gene. These are the values represented in Figure 3.10.

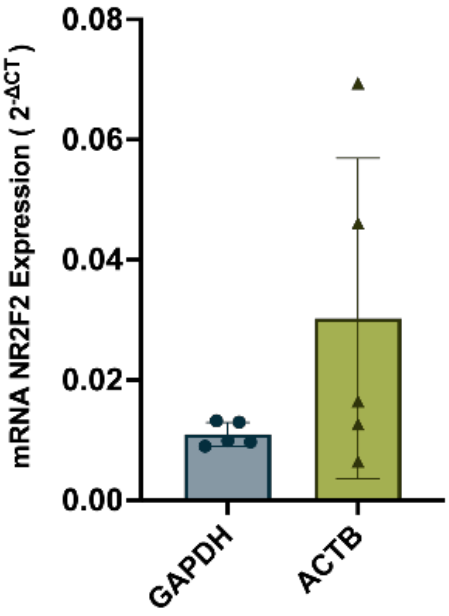


Figure 3.10 Comparison of two Housekeeping Genes for measuring *NR2F2* expression analysis by qPCR

NR2F2, in comparison with GAPDH, presented a lower, but more homogenous expression rate (0.009-0.013) compared to NR2F2 expression normalised by ACTB (0.006-0.069). Bars and error bars represent Median-IQR. N = 5.

qPCR was performed to determine the presence of *NR2F2* transcripts (RefSeq NM_021005.3) in the migrating cells and MRC-5 cells, the data is shown in Figure 3.11. *NR2F2* mRNA was detected in cardiac cells from 6 migrating cell samples tested, and MRC-5 cells. To calculate the relative mRNA expression, the $2^{-\Delta Ct}$ method was used (Methods 2.3.1). Compared with MRC-5, the cardiac cells had similar levels of *NR2F2* mRNA expression. MRC-5 cells had an *NR2F2* mRNA a median expression of 0.011 (IQR 0.006), in contrast with cardiac cells showing median expression of 0.010 (IQR 0.003) (Figure 3.11). The CT values of this assessment are shown in Table 3.7.

Table 3.7 CT values from qPCR to measure *NR2F2* mRNA relative expression

	NR2F2			*GAPDH*				
Sample	**CT_1**	**CT_2**	**CT_{Mean}**	**CT_1**	**CT_2**	**CT_{Mean}**	**ΔCT**	**$2^{-\Delta CT}$**
FB14578	28.5	28.7	28.6	22.2	21.7	22.0	**6.7**	**0.010**
FB14579	28.4	28.2	28.3	21.8	21.4	21.6	**6.7**	**0.010**
FB14580	28.3	28.4	28.3	22.0	22.1	22.0	**6.3**	**0.013**
FB14562	28.7	28.8	28.7	22.4	22.3	22.3	**6.4**	**0.012**
FB14597	27.9	27.9	27.9	21.4	21.4	21.4	**6.5**	**0.011**
FB14590	37.5	36.6	37.1	30.0	30.6	30.3	**6.8**	**0.009**
MRC-5 (1)	29.0	28.3	28.7	22.8	22.1	22.4	**6.2**	**0.013**
MRC-5 (2)	29.5	29.3	29.4	22.3	22.5	22.4	**7.1**	**0.007**
MRC-5 (3)	29.3	29.4	29.3	22.8	22.9	22.8	**6.5**	**0.011**

CT: Number of cycles required for the florescent signal of the amplified gene to exceed baseline; Cycle Threshold; CT_{Mean} = (CT_1+CT_2) ÷ 2; CT_1: CT value of the duplicate 1; CT_2: CT value of duplicate 2; ΔCT = *NR2F2* CT_{Mean} - Housekeeping gene (*GAPDH*) CT_{Mean}; $2^{-\Delta CT}$: Relative Expression of *NR2F2* mRNA normalised by the housekeeping gene (GAPDH). These are the values represented in Figure 3.11.

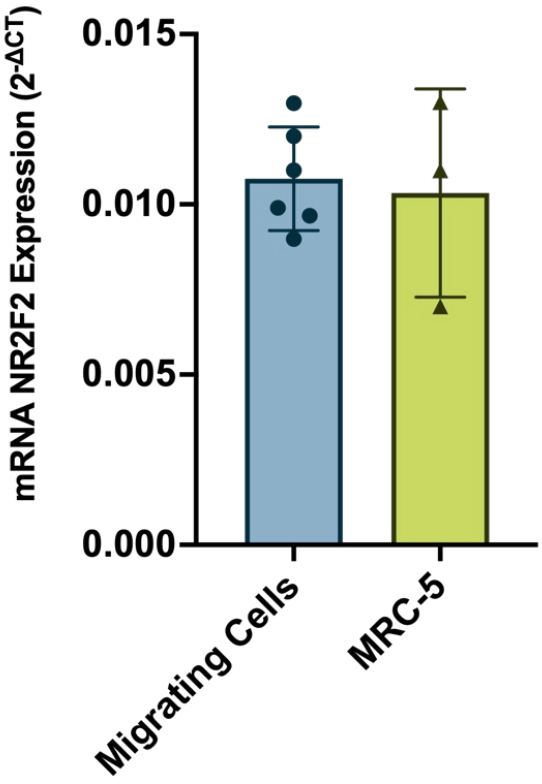


Figure 3.11 *NR2F2* mRNA expression in cells migrating from the foetal heart

Migrating cells showed similar expression of *NR2F2* mRNA compared to MRC-5. For this analysis, GAPDH was used as housekeeping gene. For MRC-5, N=3. For cardiac primary cells, N=6. Each dot represents one measurement. Bars and error bars represent Median-IQR.

3.5 NR2F2 is detected at protein level in foetal heart's migrating cells

Flow cytometry was carried out to identify NR2F2 at the protein level in the foetal heart's migrating cells and MRC-5 cells (Methods 2.3.2). The gating strategy is shown in Figure 3.12. Singlets were identified based on the FSC-H vs FSC-A. From the singlets gate, cells were identified based on FSC vs SSC. Viable cells were selected based on the negative signal of Zombie Violet™ Fixable Viability staining (Live/Dead).

The MRC-5 assessment showed 74% of singlets events of which 68.9% were cells. From the identified cells, 92.6% were viable. 90.5% of the live/dead events were positive for NR2F2 (Figure 3.12). In the assessment of cardiac cells, the same gating strategy to select viable cells was adopted. The MRC-5 presented a median of 49.95% of cells positive for NR2F2 (IQR 55.7%). The primary cells had a median of 23.9% of NR2F2^{+} cells (IQR 32.95%) (Figure 3.13).

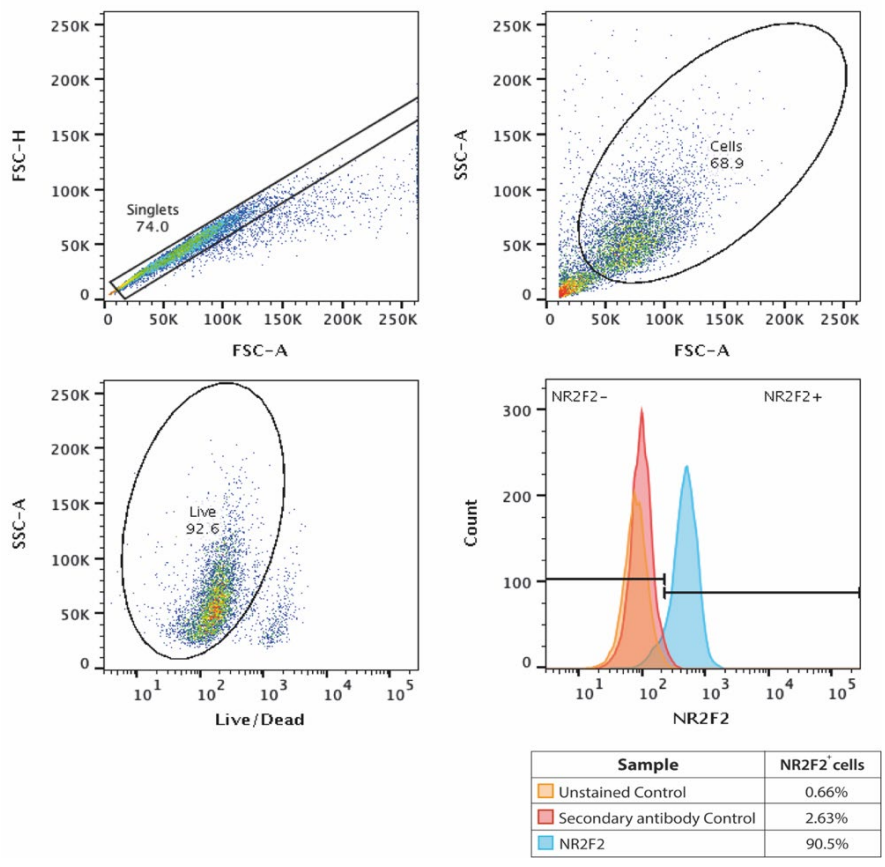


Figure 3.12 MRC-5 expressed NR2F2 at protein level

Gating strategy used to define singlets (FSC-A x FSC-H), cell population (FSC x SSC) and live cells (SSC x DAPI). The histogram shows the fluorescence intensity of the unstained (orange) cells, i.e., no antibody addition, secondary control sample (red), i.e., cells only stained with secondary antibody, and cells stained with anti-NR2F2 (blue) and secondary antibodies.

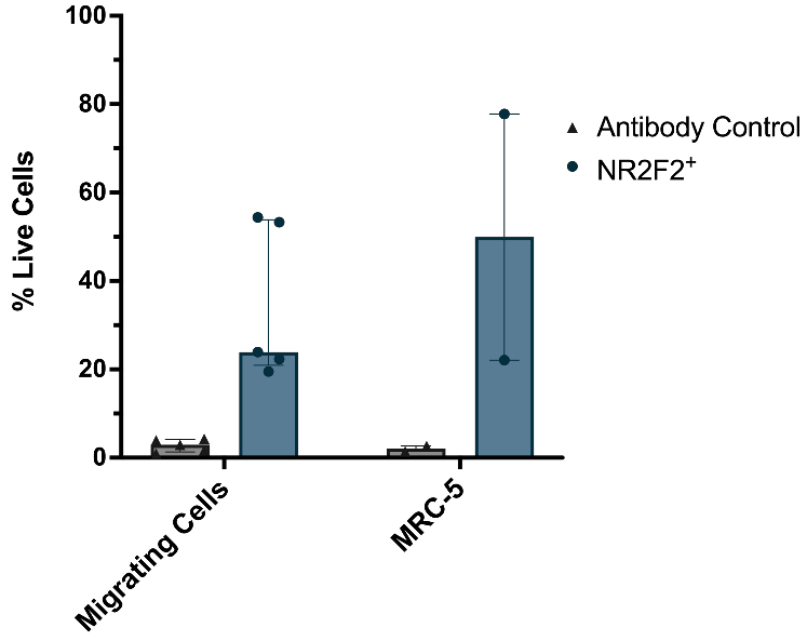


Figure 3.13 Percentage of positive cells for NR2F2 in foetal cardiac migrating cells.

The bar chart shows the percentage median of positive cells expressing NR2F2 protein. The MRC-5 (righthand side) has a median of 49.95% NR2F2+ (IQR 55.7%), N=2. The migrating cardiac cells (lefthand side) present a median of 23.9% NR2F2+ cells (IQR 32.95%), N=5. Bars and error bars indicate Median-IQR.

3.6 Discussion

Previous unpublished data from our group showed expression of *NR2F2* in fibroblasts and endothelial cells isolated from human foetal hearts, suggesting that altering the *NR2F2* expression in these cell types has the potential to be related to heart malformations. In support of this, we isolated migrating cells from the heart in the early stages of development to further assess the role of NR2F2 in fibroblasts.

Fibroblasts *in vitro* present a selective adhesion preference to culture plate surface (218), and they can migrate out of tissue explants. However, using an isolation protocol based on differential migration rates, it is possible to have the presence of other cell types (215). Therefore, further characterisation of the phenotypes of the cells that migrated out of heart explants was necessary. A panel of biomarkers was assessed by flow cytometry, and single-cell transcriptomic profiling was used to determine the identity of the migrating cells. In flow cytometry, 2 out of the 5 samples from foetal hearts had a similar pattern as MRC-5 – a foetal lung fibroblast cell line. The presence of a population $CD90^{-}VIM^{+}\alpha MHC^{-}$ could suggest the presence of vascular smooth muscle cells (219,220) or that fibroblasts may have changed their phenotype to myofibroblasts losing the Thy-1 expression during cell culture (212,221). Thy-1

(CD90) can also be expressed in endothelial cells (213), so a specific marker for endothelial cells – Pecam-1 – was also assessed to exclude this cell type. 3 samples showed CD31 (Pecam-1) expression, indicating the presence of endothelial cells. In addition, 4 samples had cells positive for αMyHC indicating the presence of cardiomyocytes. Unexpectedly, these cells were also positive for Thy-1 (CD90) or Vimentin. Previous studies in cardiac progenitor cells extracted from adult and foetal human hearts reported that the cells that express Thy-1, although present lower cardiomyogenic potential, could co-express sarcomeric proteins, such as αMyHC (222,223), which is consistent with population showing a $CD90^{+}\alpha MyHC^{+}$ population. This could indicate that this $CD90^{+}\alpha MyHC^{+}$ population are representing immature cardiomyocytes. Moreover, a study from 2007 in rats showed that cardiac fibroblasts can influence the phenotype of cardiomyocytes *in vitro*. They observed that cardiomyocytes cultured for 6 days using conditioned media from fibroblasts increased expression of Vimentin over time (224), which may explain the presence of the $VIM^{+}\alpha MyHC^{+}$ population. These data confirmed the importance of using exclusion criteria using cell markers specific for other cell types to determine fibroblast presence in the samples due to the lack of fibroblast-specific markers.

In the study that developed the migration-based and filtration protocol to isolate and purify fibroblasts from heart explants, Ieda et al. isolated cells from transgenic mice hearts. These mice would express GFP driven by the αMyHC promoter, which means only cardiomyocytes would express GFP (215). In their population of migrating cells, they could confirm no expression of GFP or other proteins specific for cardiomyocytes by flow cytometry. However, they did not assess cell markers for smooth muscle cells or endothelial cells. Recently, a group compared the three most used isolation techniques: (1) selective adhesion (same principle used in this thesis) following enzymatic digestion, (2) cell sorting using fluorescence, and (3) magnetic bead separation to isolate adult fibroblasts and myofibroblasts from rodent hearts (218). Looking into the purity of each isolated cell population, they concluded that the only method with high purity was the cell sorting fluorescence. In this technique, they used a transgenic mouse with the GFP expression driven by the α-smooth muscle actin promoter, as their goal was to isolate myofibroblasts. Although could be an efficient alternative to obtain a pure cell population, it presents an obvious limitation for human samples: recombination technique is necessary. Based on these several pieces of evidence, it is possible to obtain an

enriched population of fibroblasts from foetal hearts, however, it is important to assess the phenotype of each sample as there are varying degrees of heterogeneity of the isolated cells and a highly purified population of fibroblasts may not be possible with the protocols used in human samples.

Regarding the single-cell transcriptomic data, it was observed that fibroblasts specific genes were enriched throughout the entire dataset. However, when the expression of gene markers was looked at individually, two genes of smooth muscle cells presented significant expression: *TAGLN* and *ACTA2*, in contrast to *RGS5*. Transgelin (*TAGLN*) and α-smooth muscle actin (*ACTA2*) are known to be genes involved in the contractility of myofibroblasts (225). When the diseased heart starts its fibrotic process, TGF-β activates resident fibroblasts into myofibroblasts, which express those genes (145). It is also known that, in normal culture conditions, primary fibroblasts can still differentiate into myofibroblasts, even in the absence of TGF-β or serum (226). This combined knowledge could lead one to believe that the high z-score for smooth muscle cell genes presented in this data can be a consequence of the activation of fibroblasts into myofibroblasts due to the limitations of those cell cultures *in vitro* apart from the presence of smooth muscle cells itself. In addition, those markers were not localised into a specific cluster, they were co-expressed with fibroblast markers in most cells, which contribute to the hypothesis of the presence of activated fibroblasts rather than smooth muscle cell contamination. In corroboration, only one gene from cardiomyocytes had its expression observed in the migrating cells: *MYL9.* Although it is present in cardiomyocytes as part of their contractile machinery, this myosin light chain gene is also known to be a contractile factor in activated fibroblasts (227,228).

The data presented in this chapter also shows that NR2F2 can be detected by flow cytometry and qPCR in all migrating cell samples isolated from foetal hearts, similarly to the fibroblast cell line – MRC-5. However, the expression of NR2F2 presented a wide intra sample variability in MRC-5 in both mRNA and protein levels (Figure 3.11 and Figure 3.13). This could be attributed to the lack of serum starvation of those cells prior to analysis. Serum starvation is shown to synchronise cell cycle in fibroblasts, which would result in a metabolic homogeneous cell population (229).

Thus, although the isolation of a purified population of fibroblasts from human foetal hearts is challenging, it is possible to obtain a cell population enriched for fibroblasts and NR2F2 can be detected, at both transcript and protein levels in them, which makes these samples a coherent model of study.

Chapter 4 Results

Transcriptomic analysis of public datasets from foetal cardiac cells

4.1 Introduction

4.1.1 Methods of transcriptomic analysis

Over the past 20 years, approaches to investigate gene expression in a genome-wide manner became available with the evolution of microarray techniques that permitted high-throughput transcriptomic analysis. Initially, microarrays were designed to measure gene expression levels, but soon became possible to do differential expression analysis between samples (230). However, microarrays present several limitations, e.g. a limited dynamic range of detection due to high backgrounds and saturation of the signals (231).

RNA sequencing (RNA-seq) is a sequencing-based technology (versus the hybridization-based method of microarrays) that allowed high sensitivity, low to no background as the reads map unambiguously to the genome, and high dynamic range of detection on the gene expression levels (231). Several studies comparing microarrays and RNA-seq showed that, while microarrays had results with a degree of confidence, RNA-seq presented as being highly reproducible, better estimates of transcript expression, and better detections of low abundance transcripts (232–234).

Although RNA-seq changes the way that genomic biology is investigated, it cannot be applied in heterogenous cell populations as it measures the average of expression across a pool of cells of mixed phenotypes. Single-cell RNA sequencing (scRNA-seq) brought the resolution necessary to investigate complex multicellular systems and discover new cell types or subpopulations. There are several methods to obtain RNA from a unique cell in a high-throughput approach. The first paper sequencing transcriptome from a unique mouse blastomere isolated cells using a glass capillary. As result, they identified 5,270 more genes compared with DNA microarrays from hundreds of blastomeres. They also identified, previously unknown splice junctions and expression patterns of transcript variants in the blastomere (235).

The landmark paper described a droplet-based single-cell RNA sequencing (Drop-seq) reporting the analysis of the human retina (236). Retinal tissue was disaggregated, transcriptome-wide single-cell data was obtained, and the group was able to identify 39 distinct clusters of cells using two-dimensional representation (tSNE). *De novo* clustering of 44,808 single-cell transcriptomes demonstrated for the first time that the human retina was comprised of 39 different cell types (236).

In this chapter, *NR2F2* expression was investigated using two public datasets: bulk RNA-sequence (181) and single cell RNA sequence data (148). The bulk RNA-sequence dataset was generated from cardiomyocytes derived from induced pluripotent cells (iPS-CM) that were NR2F2 knocked-out by CRISPR/Cas9 technology. The single-cell RNA sequencing was generated from human foetal hearts, and we used this dataset to evaluate the expression of NR2F2 in the cell types.

4.2 Absence of NR2F2 causes perturbation in other cardiogenic genes

A public dataset of bulk RNA sequencing from cardiomyocytes derived from iPS-CM was analysed (Methods 2.10.1)(181). This dataset comprised transcriptome of iPS-CM sequenced after 30 days of cardiac differentiation protocol followed by *NR2F2* edited from these cells using CRISPR/Cas9 system. DEG analysis was carried out using the *limma* R package. Differential gene expression analysis identified 114 upregulated genes (Appendix B.2), and 45 downregulated genes in *NR2F2* knockout (Appendix B.3). Among the upregulated genes ($\log_{10}(FC) \geq 1.5$), it is possible to see cardiogenic genes as *SOX9* (237,238) and *BMP4* (239). In the downregulated gene set ($\log_{10}(FC) \leq -1.5$), genes correlated to heart development were also listed, as *NKX2-5* (240,241), which contributed for the biological processes enriched by the downregulated genes are liked with cardiac cell development. The top 10 biological processes enriched in the upregulated and downregulated gene sets are demonstrated in Table 4.1 and Table 4.2, respectively. Looking into normalised counts of each sample, the *NR2F2* presented $\log_{10}(FC)=0.06$ when its expression in the knocked-out iPS-CM (Mean of normalised counts = 2.61, n=6) was compared with the control group (Mean of normalised counts = 2.27, n= 4).

Table 4.1 Top 10 Biological Processes enriched in the upregulated gene set of *NR2F2* knocked-out iPS-CM (public dataset)

Enrichment FDR	Fold Enrichment	Biological Process	Genes
2.21E-05	4.82	Positive regulation of locomotion	*SERPINE1 BMP4 SOX9 TUBB2B THBS1 SPHK1 FBLN1 FN1 TGFB2 AMOT ANXA1 EDN1 NEDD9 CD99 CEMIP CYR61 SEMA6B*
1.42E-05	4.02	Circulatory system development	*SERPINE1 CCL2 BMP4 SOX9 APOE THBS1 FREM2 SPHK1 NKX2-6 FN1 TGFB2 CTGF ANKRD1 AMOT ID1 EDN1 COL2A1 CPE CYR61 ACTA2 ROBO1 ADM*
1.25E-05	3.97	Cell adhesion	*EPDR1 CCL2 SOX9 THBS1 PCDH10 ADD2 ITGA9 FREM2 SLC7A11 IGFBP7 CTNND2 FN1 TGFB2 CTGF ID1 ANXA1 NEDD9 CD99 CYR61 CLSTN2 ROBO1 TENM2 CDH8*
1.25E-05	3.94	Biological adhesion	*EPDR1 CCL2 SOX9 THBS1 PCDH10 ADD2 ITGA9 FREM2 SLC7A11 IGFBP7 CTNND2 FN1 TGFB2 CTGF ID1 ANXA1 NEDD9 CD99 CYR61 CLSTN2 ROBO1 TENM2 CDH8*
2.21E-05	3.87	Regulation of cell migration	*SERPINE1 CCL2 BMP4 SOX9 APOE SULF1 THBS1 SPHK1 FBLN1 DPYSL3 FN1 TGFB2 AMOT ANXA1 EDN1 NEDD9 CD99 CEMIP CYR61 ROBO1 SEMA6B*
7.73E-06	3.79	Regulation of anatomical structure morphogenesis	*SERPINE1 CCL2 BMP4 SOX9 APOE TUBB2B SULF1 THBS1 ETV5 MAP6 SPHK1 FBLN1 FN1 APCDD1 TGFB2 AMOT ID1 ANXA1 EDN1 NEDD9 LZTS1 ROBO1 ADM KIF1A SEMA6B MYC*

Enrichment FDR	Fold Enrichment	Biological Process	Genes
2.21E-05	3.64	Regulation of cellular component movement	*SERPINE1 CCL2 BMP4 SOX9 APOE TUBB2B SULF1 THBS1 SPHK1 FBLN1 DPYSL3 FN1 ATP1A2 TGFB2 AMOT ANXA1 EDN1 NEDD9 CD99 CEMIP CYR61 ROBO1 SEMA6B*
2.21E-05	3.31	Locomotion	*CCL2 BMP4 DCLK1 TUBB2B THBS1 ADD2 ITGA9 SLC7A11 GPC2 TUBB3 CRMP1 FN1 APCDD1 ATP1A2 TGFB2 CTGF AMOT ID1 ANXA1 NEDD9 CD99 LIMA1 CYR61 ROBO1 TENM2 SEMA6B*
2.21E-05	2.89	Neurogenesis	*NNAT CCL2 BMP4 SOX9 APOE DCLK1 TUBB2B UCHL1 GPC2 MAP1A ETV5 MAP6 CTNND2 TUBB3 CRMP1 DPYSL3 FN1 APCDD1 TGFB2 ANKRD1 ID1 ANXA1 ID4 LZTS1 ROBO1 TENM2 ADM KIF1A SEMA6B ENC1*
2.21E-05	2.81	Tissue development	*BMP4 SOX9 SULF1 STC2 FREM2 SLC7A11 MYOZ2 NKX2-6 TUBB3 ENO3 FN1 APCDD1 ACTA1 TGFB2 CTGF ANKRD1 ID1 ANXA1 ID4 EDN1 COL2A1 TNFRSF19 ODAM ACTG2 CYR61 ACTA2 ROBO1 TAGLN ADM SEMA6B MYC*

FDR: False discovery rate. FDR is calculated based on the nominal *p*-value, which states the likelihood of enrichment by chance. The fold enrichment is the percentage of genes in the list of differentially expressed genes belongs to the biological process, indicating the overrepresentation. The biological processes are ranked by Fold Enrichment.

Table 4.2 Top 10 Biological Processes enriched in the downregulated gene set of bulk RNA-sequencing

Enrichment FDR	Fold Enrichment	Biological Process	Genes
2.06E-06	41.21	Cardiac cell development	*TGFBR3 MYL2 CSRP3 ALPK3 NKX2-5 LRRC10*
7.11E-07	25.01	Muscle cell development	*TNNT1 MYL2 CSRP3 ALPK3 LMOD2 MYOZ2 NKX2-5 LRRC10*
7.68E-07	17.15	Striated muscle cell differentiation	*TNNT1 MYL2 SMYD1 CSRP3 ALPK3 LMOD2 MYOZ2 NKX2-5 LRRC10*
7.11E-07	14.34	Muscle cell differentiation	*TNNT1 MYL2 SMYD1 CSRP3 ALPK3 NPNT LMOD2 MYOZ2 NKX2-5 LRRC10*
3.63E-06	13.01	Muscle contraction	*TNNT1 MYL2 CSRP3 TRIM63 NPNT LMOD2 DES NKX2-5 FXYD1*
4.51E-06	12.50	Striated muscle tissue development	*TGFBR3 MYL2 BMP5 SMYD1 CSRP3 ALPK3 MYOZ2 NKX2-5 LRRC10*
5.98E-06	11.92	Muscle tissue development	*TGFBR3 MYL2 BMP5 SMYD1 CSRP3 ALPK3 MYOZ2 NKX2-5 LRRC10*
2.08E-06	11.65	Muscle system proc.	*TNNT1 MYL2 CSRP3 TRIM63 NPNT LMOD2 MYOZ2 DES NKX2-5 FXYD1*
3.28E-06	9.16	Muscle structure development	*TGFBR3 TNNT1 MYL2 SMYD1 CSRP3 ALPK3 NPNT LMOD2 MYOZ2 NKX2-5 LRRC10*
1.25E-05	8.95	Heart development	*TGFBR3 MYL2 BMP5 SMYD1 CSRP3 ALPK3 SFRP2 DHRS3 NKX2-5 LRRC10*

FDR: False discovery rate. FDR is calculated based on the nominal *p*-value, which states the likelihood of enrichment by chance. The fold enrichment is the percentage of genes in the list of differentially expressed genes belongs to the biological process, indicating the overrepresentation. The biological processes are ranked by Fold Enrichment.

Protein-protein interaction analysis of DEG-related set of proteins was performed using STRING. The network had significantly more interactions than expected for a random set of proteins of similar size, indicating that the proteins are at least partially biologically related ($p < 10^{-16}$). *NR2F2* was added to the list of genes at STRING analysis, and this manual input modification did not drastically alter the results (data not shown). From the 159 DEGs, the network presented 148 nodes (i.e., proteins that have at least 1 association) and 288 protein-protein associations (Figure 4.1). GO evaluation of this network (Table 4.3) highlighted developmental process (FDR = 1.09×10^{-16}) as biological process. Protein binding (FDR = 1.78×10^{-6}) and extracellular region part (FDR = 1.24×10^{-6}) were enriched as molecular function and cellular component, respectively. Two of the 114 upregulated genes interact with *NR2F2*: *SOX9* ($\log_{10}$FC = 4.89) and *BMP4* ($\log_{10}$FC = 1.61). From downregulated genes, 2 genes out of 45 also interacted to *NR2F2*: *NKX-2.5* ($\log_{10}$FC = -2.01) and *NPNT* ($\log_{10}$FC = -1.55).

Considering genes associated with *NR2F2*, we created a second network with *SOX9*, *BMP4*, *NKX2-5*, *NPNT* and *NR2F2* using the same parameters as before, and this network also had significantly more interactions than expected ($p = 5.59 \times 10^{-8}$). Biological processes highlighted positive regulation of stem cell differentiation, muscle structure development and tube morphogenesis. Cellular components enriched only nuclear transcription factor complex (Table 4.4).

Table 4.3 Top 5 enriched terms of biased GO analysis for DEGs using STRING

Biological Process		
GO-term	***Description***	***FDR***
GO:0032502	developmental process	1.09E-16
GO:0007275	multicellular organism development	1.74E-16
GO:0048856	anatomical structure development	2.02E-16
GO:0048731	system development	6.00E-16
GO:0032501	multicellular organismal process	8.70E-16
Molecular Function		
GO-term	***Description***	***FDR***
GO:0005515	protein binding	1.78E-06
GO:0008092	cytoskeletal protein binding	1.69E-05
GO:0005102	signalling receptor binding	0.00018
GO:0005488	binding	0.0011
GO:0003779	actin binding	0.002
Cellular Component		
GO-term	***Description***	***FDR***
GO:0005576	extracellular region	1.24E-06
GO:0005615	extracellular space	1.24E-06
GO:0044421	extracellular region part	1.24E-06
GO:0044449	contractile fibre part	1.25E-06
GO:0099512	supramolecular fibre	1.25E-06

FDR: false discovery rate. GO-term: gene ontology term.

Table 4.4 GO enrichment analysis of up and down regulated transcription factors that interact with NR2F2

Biological Process		
GO-term	***Description***	***FDR***
GO:0007507	heart development	2.86E-06
GO:0014706	striated muscle tissue development	2.86E-06
GO:0055002	striated muscle cell development	2.86E-06
GO:0055006	cardiac cell development	2.86E-06
GO:0061061	muscle structure development	2.86E-06
Cellular Component		
GO-term	***Description***	***FDR***
GO:0044798	nuclear transcription factor complex	0.0425

FDR: false discovery rate. GO-term: gene ontology term.

Figure 4.1 Protein-protein interaction network of up and down regulated transcription factors that interact with NR2F2

The lines indicate association between proteins. The red lines are the interactions between *NR2F2* (yellow node) and *SOX9*, *NPNT*, *BMP4* and *NKX2-5* (yellow nodes). p = 5.59 x 10^{-8}. Red nodes are indicating the upregulated genes (*SOX9* and *BMP4*). The blue nodes show the downregulated genes (*NPNT* and *NKX2-5*).

4.3 NR2F2 expression pattern in human foetus heart

A second public dataset of single-cell RNA sequencing of 3 healthy foetal hearts (ages 19 and 22 gestational weeks) was analysed to understand the expression trend of *NR2F2* in the developing heart (148) (Methods 2.10.2). 5861 cells were obtained after filtration.

Following pre-processing and QC steps, UMAP calculation with Leiden clustering (r=0.5), 12 clusters were annotated by the expression of the cell-type specific genes cited in the original study (Figure 4.3). Cardiomyocytes were identified by co-expression of myosin light (*MYL2*, *MYL3*, and *MYL4*) and heavy chains (*MYH7*, *MYH7B*, and *MYH6*) along to troponins (*TNNT2*, *TNNI1*, *TNNI3*, and *TNNC1*), encoding forming the contractile apparatus of those cells. A variety of immune cells populations were also captured. From the lymphoid lineage, T cells were identified by the expression of *LTB* and *IL7R*, natural killer (NK) cells presented *NKG7*, *GNLY* expression, and B cells were differentiated by *MS4A1* expression. Myeloid lineage cell included Mast cells with expression of *TPSB2* and *GATA2.* Monocytes expressed *CD68*, *LYZ* and *S100A8*, and macrophages were identified by presence of *MS4A4* transcripts. $CD1C^{+}$ Dendric cells were also identified as an individual cluster. Endothelial cells were clustered by the expression of *PECAM1* and *CDH5*. Some genes in the endothelial cells cluster were shared with fibroblasts such as *COL3A1*, *COL1A2*, and *FN1,* indicating that those cells are undergoing endothelial-mesenchymal transition (EndoMT) (148), crucial for endocardial cushion and valves formation (242). Individually, fibroblasts expressed *DCN*, *DLK1*, and *LUM*, and smooth muscle cells expressed uniquely *TAGLN*, *RGS5* and *ACTA2* (Figure 4.2). Table 4.5 shows the number of cell counts presented in each cluster.

Table 4.5. Number of cells per annotated cluster

Cluster	Cell count	Cluster	Cell count	Cluster	Cell count
Cardiomyocytes	987	**Mesothelial cells**	91	**NK cells**	233
Endothelial cells	1198	**B cells**	269	**Mast cells**	181
Fibroblasts	1049	**T cells**	476	**Macrophages**	408
Smooth muscle cells	357	**Monocytes**	353	**Dendritic cells**	259

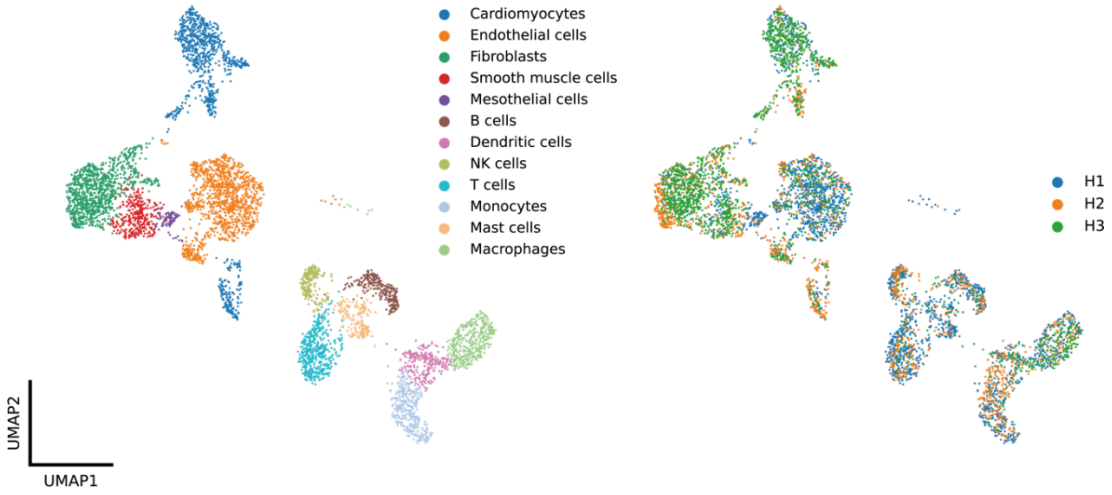


Figure 4.2 Cell populations of human foetal heart

UMAP visualization of 5861 cells from3 healthy foetal heart samples coloured by cell identity (left): cardiomyocytes fibroblasts, endothelial cells, smooth muscle cells, natural killer cells, T cells, B cells, mast cells, macrophages, monocytes and CD1C+ dendritic cells. On the righ-hand side, UMAP clustered by each foetal heart (H1, H2, H3). Scanpy, UMAP plot: Leiden r = 0.5, n_pcs=20, n_neighbours =20.

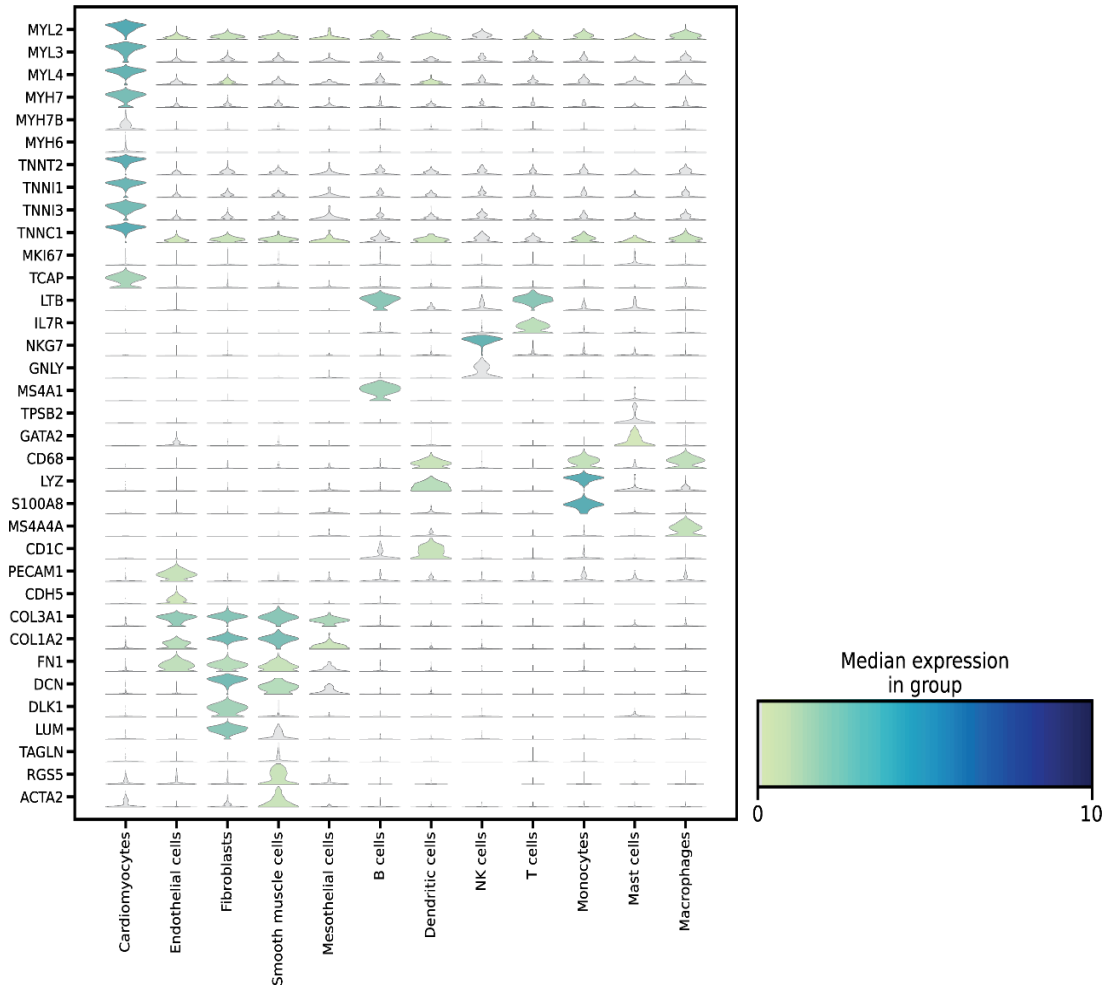


Figure 4.3 Cardiac cells expressed specific gene markers for each cell type.

Violin-plot of expression levels of cell-type specific gene markers in cells dissociated from foetal hearts. The x-axis contains the cell-type clusters, the left y-axis contains the gene markers, and colour bar represents the median of normalized counts for each gene.

Looking at *NR2F2* mRNA expression at single cell level, it was observed that *NR2F2* was mainly expressed in fibroblasts, smooth muscle cells and endothelial cells, which the least expressing higher levels (Figure 4.4). It was also observed lower expression in the cluster of cardiomyocytes and mesothelial cells (Figure 4.4).

Since the data analysis of *NR2F2* knockout iPS-CM (Results 4.2) identified 4 DEGs interacting with *NR2F2* (*SOX9*, *BMP4*, *NKX2-5* and *NPNT*), we investigated if their expressions co-localised in the clusters with *NR2F2* (Figure 4.5). From the four genes, only *SOX9* and *BMP4* had similar expression pattern as *NR2F2* mainly localised in fibroblasts population. *BMP4* was also expressed in endothelial cells.

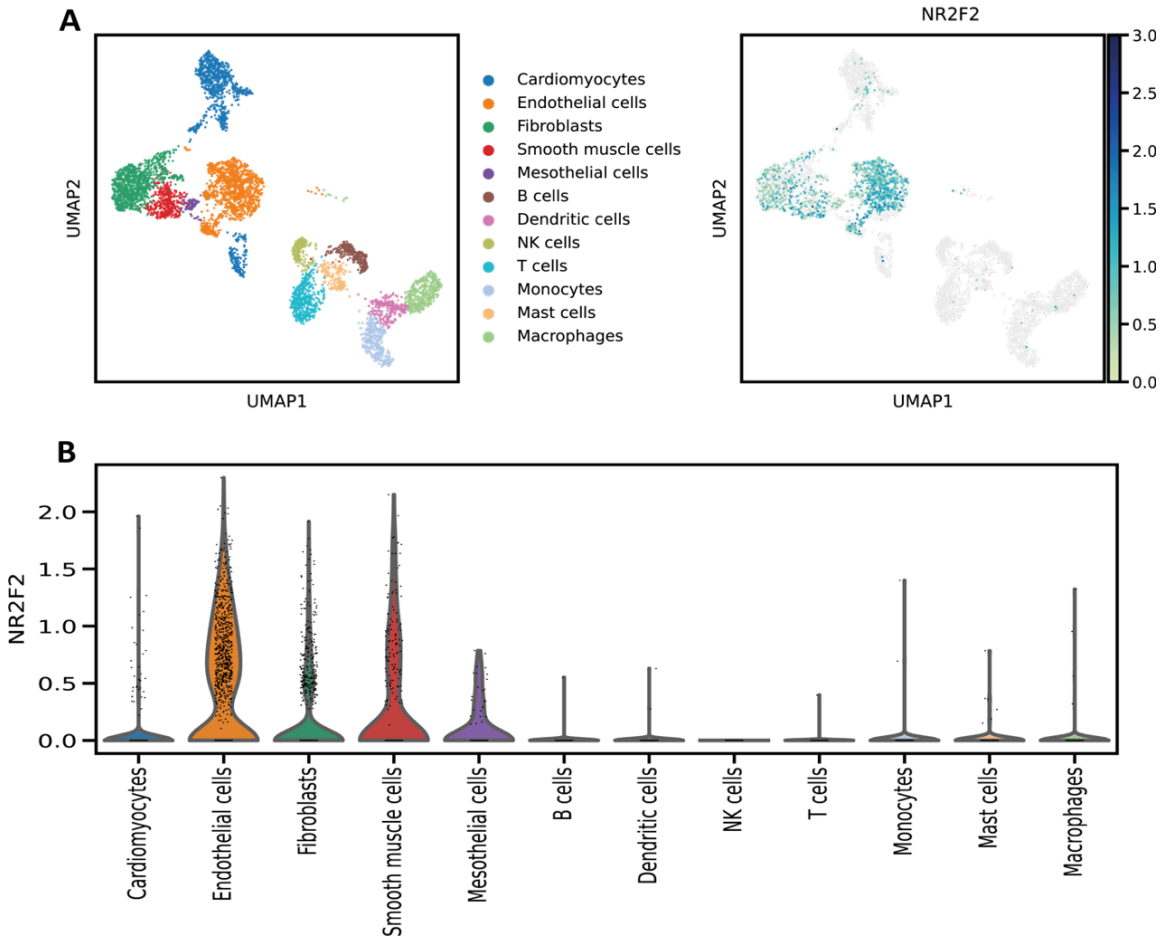


Figure 4.4 Expression of *NR2F2* across the cell types in human foetal heart

(A) UMAP plots of coloured cell-type clusters (left) and *NR2F2* expression across the clusters (right). (B) Violin plot of *NR2F2* expression across the cell-type clusters. The y-axis indicates the normalised counts of *NR2F2* in each cell (black dots). In the x-axis are plotted the clusters identified at the dataset.

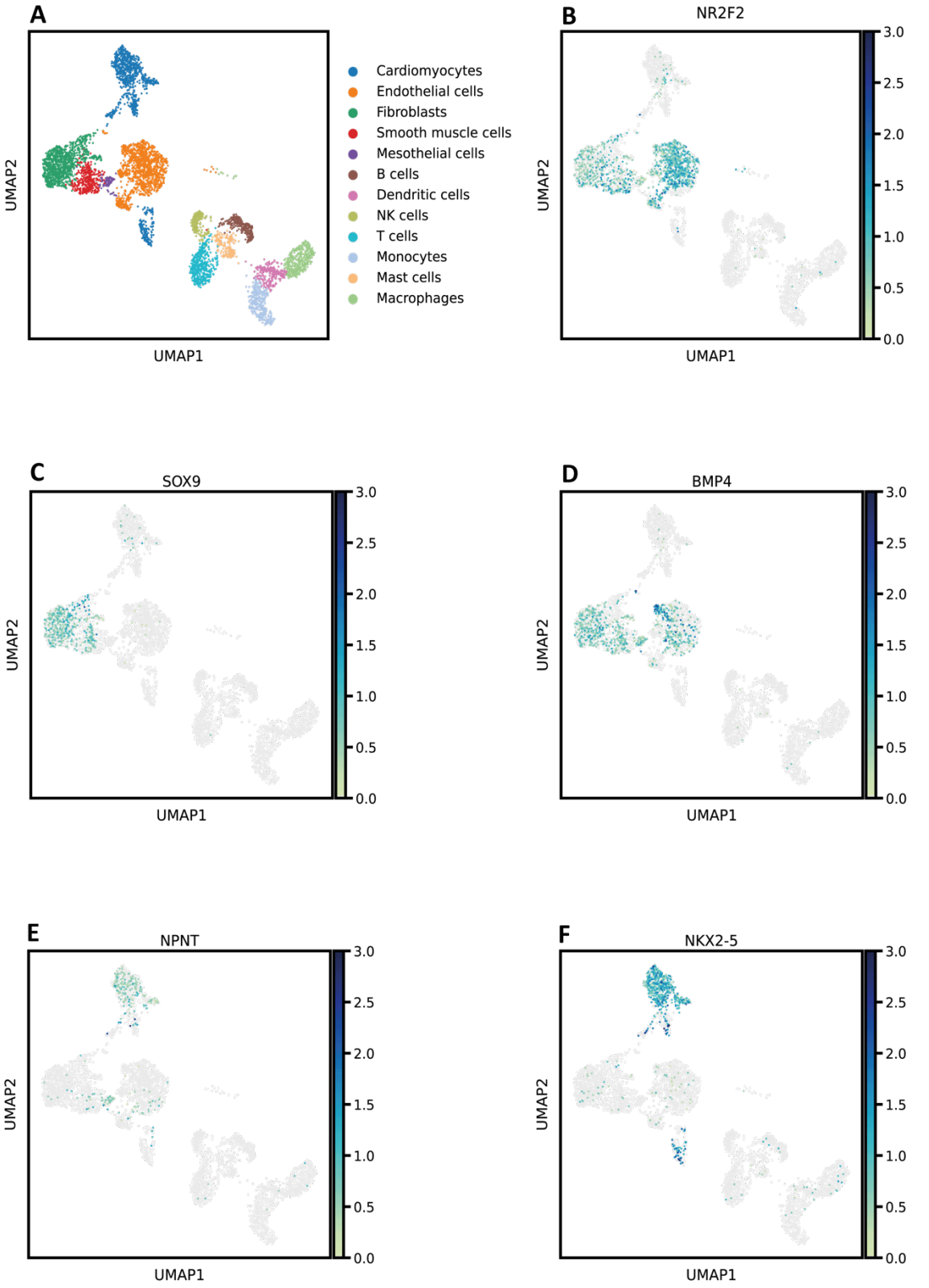


Figure 4.5 Expression pattern of NR2F2, SOX9, BMP4, NPNT and NKX2-5 in the developing heart

(A) UMAP visualisation of 5861 cells in three foetal hearts coloured by cell type. Expression abundance (coloured bar in the right) of (B) *NR2F2*, (C) *SOX9*, (D) *BMP4*, (E) *NKX2-5* and (F) *NPNT* in the heart. NR2F2 is mostly expressed by endothelial cells, fibroblasts, smooth muscle cells and mesothelial cells. *SOX9* and *BMP4* are expressed in fibroblasts and smooth muscle cells. The latter is also expressed in endothelial cells. *NKX2-5* is mainly present in cardiomyocytes. *NPNT* is lowly present in cardiomyocytes.

4.4 Discussion

To understand the importance of NR2F2 in the developing heart, we used published datasets and applied transcriptomic analysis. Firstly, a dataset generated from a bulk RNA sequencing of *NR2F2* knocked out iPS-CM via CRISPR/Cas9 was analysed (181).

NR2F2 did not emerge as a differentially expressed gene in the comparison between *NR2F2* knockout and wild-type cells ($log_{10}FC=0.06$), although the original study has shown *NR2F2* depletion in these cells through genomic DNA analysis – Sanger sequencing, and genomic DNA PCR (181). However, no mRNA analysis was carried out by original paper's group, which makes difficult to determine the baseline levels of *NR2F2* mRNA in the samples. Also, some 4 out of 6 knocked out samples had the editing in the *NR2F2* loci, but only produced heterozygous mutation, which could be a limitation for the sensitivity of the bulk RNA sequencing not being high enough to detect a greater decrease of a lowly expressed gene in a heterozygous population. Moreover, looking into the normalised counts of the samples from the bulk RNA sequencing, the average count in the knockout group was slightly higher than the control, but not significantly.

Bulk RNA sequencing is a high-throughput sequencing method commonly used to compare transcriptomes of different conditions, e.g. gene perturbation vs control. In this approach, the sequenced RNA is extracted from a pool of cells and used to assess the average levels of expression of each gene (243). However, the output of this method does not identify the transcriptome in individual cells which could mean loss of sensitivity and biological information. Thus, single-cell RNA sequencing emergence enabled to investigate at high resolution the expression levels of each gene across different cell types (244). By assessing the expression levels of this gene by single-cell RNA sequencing across the cardiac cell types, it was evident that *NR2F2* is not highly expressed overall, especially in cardiomyocytes. This could indicate that, even in a knockout condition, the fold-change between samples was not abrupt enough to meet the set requirements for DEGs, although it was enough to cause perturbations in the transcriptome. (243) (244)

To have a better understand of the expression pattern of *NR2F2* in the foetal heart, we used a public dataset derived from single-cell RNA sequencing from 3 human foetal hearts (148). We used their cited specific gene markers to identify and annotate the cell clusters. It was observed that *NR2F2* was expressed in fibroblasts, endothelial cells, smooth muscle

cells, cardiomyocytes, and mesothelial cells, with the higher expression in endothelial cells followed by smooth muscle cells and fibroblasts. Although the higher expression of *NR2F2* is in endothelial cells, most of cells expressing *NR2F2* in endothelial cells cluster are co-expressing genes related to EndoMT (Appendix B.4)(148), which is a process to form fibroblasts from endothelial sources (245). Defects in EndoMT event can be associated with a variety of CHDs (245), reenforcing the importance of understand the role of *NR2F2* in fibroblasts.

The knockout of *NR2F2* in iPS-CMs caused perturbations in genes related to cardiogenesis and associated with *NR2F2*, i.e. *SOX9* (237,238), *BMP4* (239), *NKX2-5* (240,241), and *NPNT* (237,246). However, when observed the expression of these genes at single-cell resolution, only *SOX9* and *BMP4* was co-expressed in the same cell types as *NR2F2*. *SOX9* is expressed in the endocardial cushions and known to be involved in septovalvulogenesis (237). *BMP4*, is also involved in heart development. *BMP4*$^{-/-}$ mice die at early stages without mesoderm formation, the embryonic layer that gives rise to the heart (247). By Cre-recombination, mice with BMP4 ablation in different stages of heart development, showed interruption of the processes responsible for outflow track and septum formation.

The present analysis helped us to start understanding some the changes that could take place in the absence of *NR2F2*. Nevertheless, to have a complete overview of this gene function in the fibroblasts during heart development, it is crucial to combine the knockout in fibroblasts with single-cell transcriptomic approach.

Chapter 5 Results

NR2F2 knockout in Fibroblasts using CRISPR/Cas9 technology

5.1 Introduction

Knowing the importance of NR2F2 for the development of CHD, its regulatory network is investigated in this chapter. For this, after isolating fibroblasts from human foetal hearts, a genomic knockout in NR2F2 loci was performed on those cells and their transcriptome was further analysed. Fibroblasts are known to be difficult-to-transfect cells; therefore, it is pivotal to apply a cost-efficient method that it is easier to implement and execute. In comparison with other approaches, CRISPR/Cas9 presents as an alternative to those types of cells.

5.1.1 Methods commonly used to generate gene perturbations

From 1970's, gene expression modulation is possible through homologous recombination (248). Although recombinant DNA was widely used, especially in murine models, this methodology could be expensive and time-consuming, so RNA interference technology was developed to provide a more straightforward inexpensive alternative. However, the use of RNA molecules to modify gene expression is transient creating the need to alter the genome directly to create long-term perturbations (249,250). The use of nucleases associated with DNA-binding molecules increased over the decades to generate permanent modifications on DNA. Nucleases are delivered to specify sites of the DNA and cause a double-strand break (DSB). After introducing DSB in DNA, two cellular mechanisms are applied to repair the break: non-homologous end-joining causing small deletions or insertions (indels), or homology-directed repair, leading to gene correction during cellular mitosis or insertion of exogenous homologous DNA template (251).

In 1985, the discovery of zinc finger proteins and its ability of specifically binding DNA (252) provided the biochemical knowledge to engineer the fusion of a chain of zinc finger proteins to the catalytic portion of a bacterial nuclease called FokI. This system, known as Zinc Finger Nuclease (ZFN), was pioneer in targeted nucleases to cleave DNA strand at a specific gene location (253). Each amino acid present on the α-helix portion of the zinc finger proteins would be able to bind to 3bp of the targeted DNA. Despite the range of applications unlocked by ZFN,

challenges like high-level expertise to engineer ZFN, pushed the scientific community to continue pursue for new technologies (249).

Like ZFN, transcription activator-like effector nucleases (TALENs) are DNA-binding proteins fused to FokI endonuclease in the C-terminal domain (254). The advantage against ZFN is that ability of TALENs to recognise 1bp which make this method more efficient and easier to engineer and implement in non-expert laboratories. The biggest disadvantage around TALENs is the large protein size making it limits the methods to deliver to cells, and its expensive design (249,254).

The most recently developed genome engineering approach is CRISPR/Cas9, which has been widely used for genome editing. CRISPR/Cas9 technology is based on two major components: a Cas9 endonuclease, responsible for generating DNA strand breaks and the single-guide RNA (gRNA) that is capable of binding to a specific region of the genome using Watson-Crick base-pairing principles. Additionally, the complex Cas9-gRNA can recognise a motif upstream of the targeted sequence termed the Protospacer Adjacent Motif (PAM), which reduce off-target editing (251). Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) is a known immunological system in prokaryotes (255,256). In 2012, the mechanism behind CRISPR guided targeting and isolation of its components *in vitro* (256,257) represented a landmark which led to successful DNA editing in human cells (171,258–260).

5.1.2 CRISPR/Cas9 and its applications

Like most nuclease-based techniques, CRISPR/Cas9 usually cause a DSB to provoke permanent alterations in the genome. This rupture of DNA strands triggers mechanisms of repair that could be a non-homologous end joining (NHEJ) or a homology-direct repair (HDR) depending on the conditions. The NHEJ pathway occurs in the absence of a DNA template and frequently introduce base insertions or deletions (indels) that would disturb the target gene by knocking it out. The HDR pathway is high-fidelity way of repairing the DNA. It uses a DNA template to reproduce the base sequence on the targeted gene. This method can be used to introduce specific exogenous mutations at the DBS site (Figure 5.1) (261).

CRISPR/Cas9 system goes beyond gene editing. It is also possible to modulate gene expression profile using a modification of Cas9 having both nucleases' domains inactivated (dCas9). The

difference in dCas9 is the retained capacity of bind the DNA via sgRNA but not cause DNA cleavage (262). For a robust modulation response in mammalians, the dCas9 needs to be fused with transcription activator or repressor molecules, making possible specific gene expression modulation (Figure 5.2) (263).

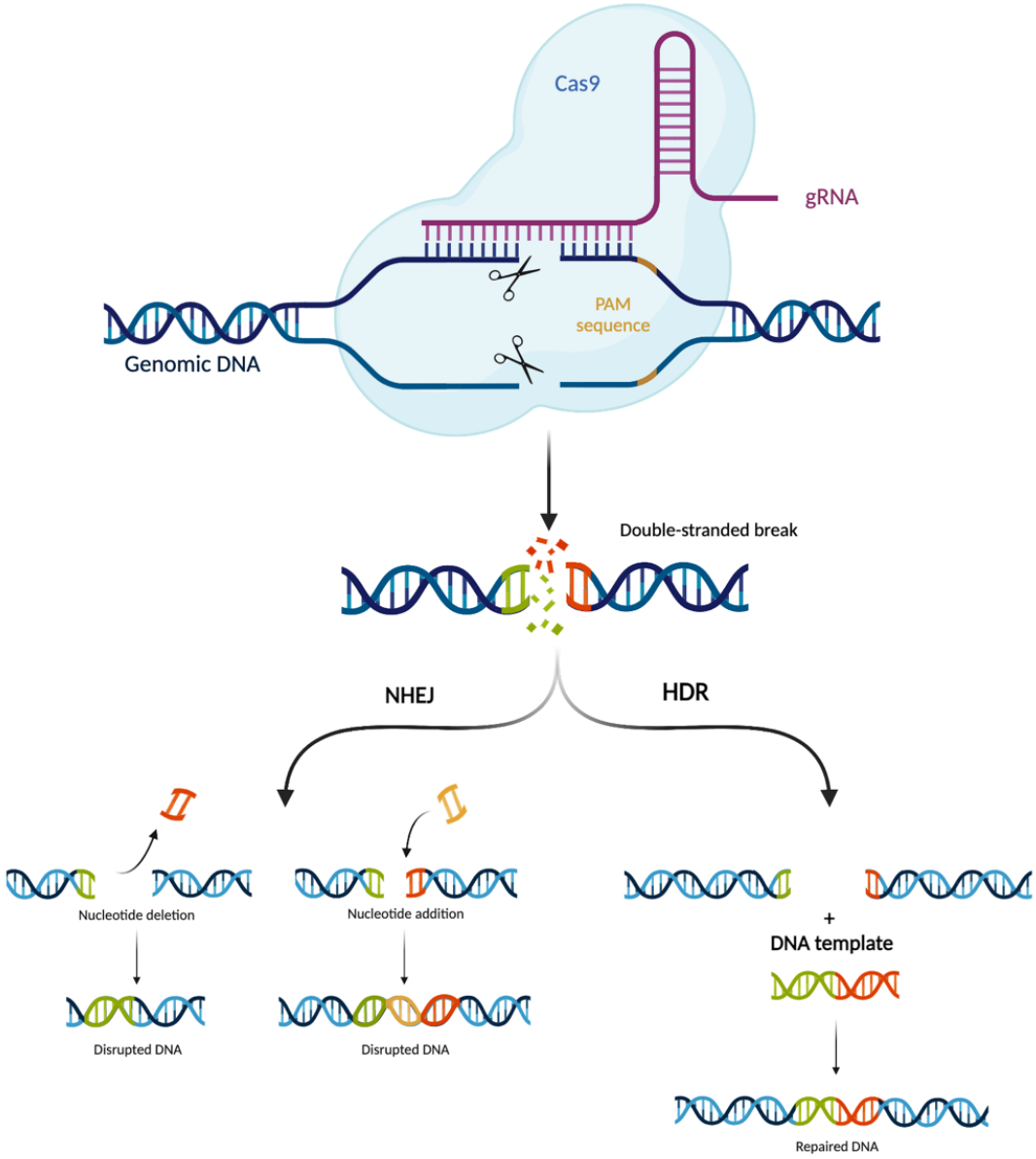


Figure 5.1 Scheme of CRISPR-Cas9 gene edition system

The endonuclease Cas9 depends on the guide-RNA (gRNA) to be delivered to the targeted gene. Apart from the complementary base pairing of the gRNA, Cas9 also recognise a protospacer adjacent motif (PAM) downstream the targeted gene. Cas9 causes a blunt DNA DSB through its two nuclease domains (scissors). The DSB can be repaired by two mechanisms: non-homologous end joint (NHEJ) and homologous direct repair (HDR). The NHEJ can delete or insert nucleotides in the location of the break and the HDR uses a DNA template to repair the break conserving the original sequence. Created with BioRender.com.

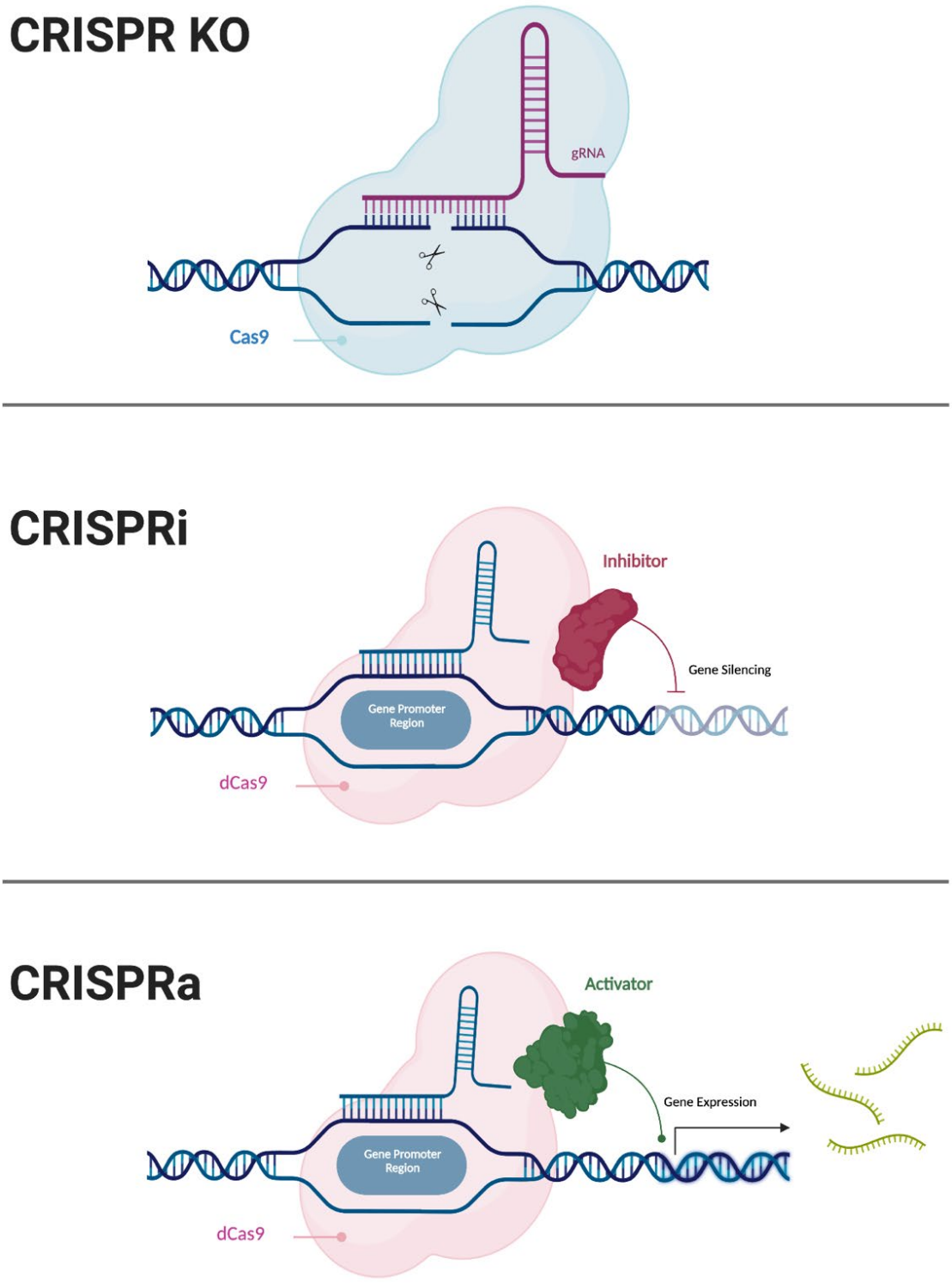


Figure 5.2 Applications beyond gene edition of CRISPR-Cas9 system

CRISPR/Cas9 technology can be used to inhibit (CRISPRi) or activate (CRISPRa) gene expression, in addition to the most common knockout application. In CRISPRi and CRISPRa, the nuclease domains of Cas9 are deactivated and an inhibitor or activator molecule is fused to it. Created with BioRender.com.

5.1.3 Comparison between CRISPR/Cas-9 methods

There are several methods to deliver CRISPR/Cas9 components to mammalian cells, and to achieve the desired outcome it is pivotal to choose the correct CRISPR/Cas9 tools. One of the main steps to establish a knockout through CRISPR/Cas9 technology is to decide in which format CRISPR components to be delivered: 1) DNA plasmids (pDNA) expressing both the Cas9 endonuclease and the guide RNA, 2) the mRNA for Cas9 translation alongside with the small guide RNA or 3) the Cas9 ribonucleoprotein (RNP) with the guide RNA directly to the cell.(264) Once the suitable CRISPR component format for the experiment workflow is chosen, it is necessary to decide the vehicle in which those components to be delivered. Lipid-mediated transfection is a standard method used to overcome the lipophobic structure of DNA molecules. The association of the exogenous DNA with the positively charged lipids form liposomes capable of fusing to the cell membrane realising the DNA into the cytoplasm.(265) Once inside the cell, the pDNA will express the components of CRISPR/Cas9 machinery responsible to find and edit the targeted gene.

As an alternative to lipid-based transfections, polymers with polycationic core, called Viromers, can be used. Viromers bind to nucleic acids through its positively charged core, but its neutral surface allows the complex to be taken by the cell through endocytosis. Once the complex is up taken in endosomes, its hydrophobicity increases due acidification of the endosome permitting the complex to be released into the cytosol and escape lysosomal degradation (266).

Viral transduction is also commonly use to carry CRISPR elements into cells. Although non-viral systems offer advantages in terms of safety and immunogenicity, virus-based approaches allow higher genomic edition efficiency.(267) One of the virus-based approaches used are lentiviruses. Lentiviruses belongs to *Retroviridae* family known for its ability to integrate the transgene into the targeted cell's genome which can guarantee long-term expression of CRISPR components (268). The main reason for using Lentivirus over other retroviruses is their competence in transduce not only dividing cells but also nondividing ones, which amplify the range of possible targeted cells.

Another widely used delivery method is electroporation. As the name suggests, electroporation uses an electrical pulse to create pores enabling exogenous DNA to enter the

cell. The formation of the pores is temporally, and the diameter can limit the DNA molecule size(269). In this work, both lipofection and electroporation were tested to deliver DNA plasmids containing the CRISPR component sequences and the RNP system to fibroblasts.

The aim of this chapter was to test several methods to deliver CRISPR/Cas9 components to create a *NR2F2* knockout mutation in fibroblasts for further transcriptome analysis of its role in this cell type.

5.2 Nonviral transfection was not able to carry CRISPR plasmids onto primary fibroblasts

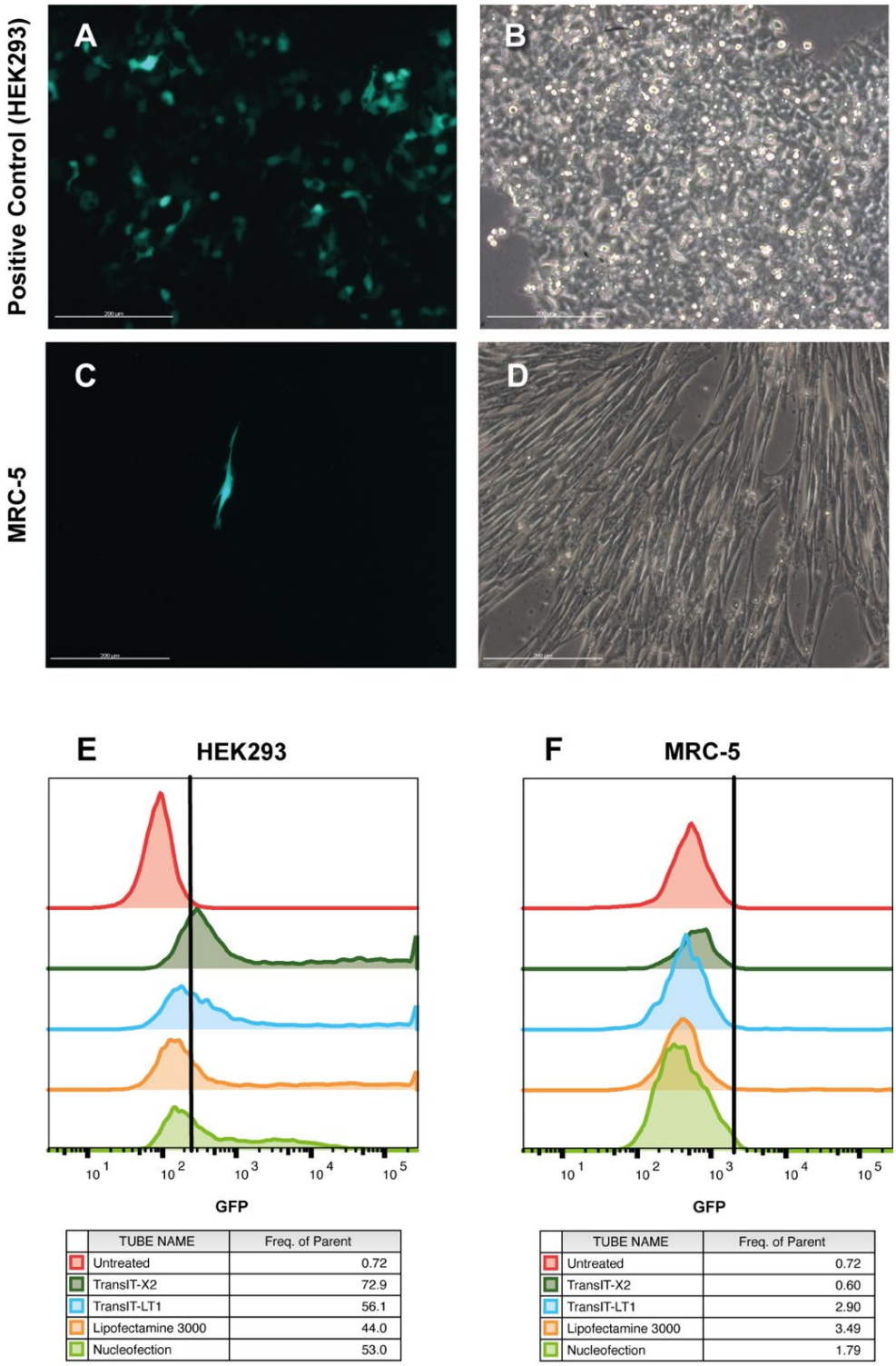
Several methods were tested to transfect CRISPR vector into fibroblasts. Five transfection reagents with different properties - lipid-based (Lipofectamine 3000 and TransIT-LT1), nonlipid polymer-based (TrasnIT-X2, Viromer Red and Viromer Plasmid) and an electroporation technology designed to delivery molecules directly to the nucleus, called Nucleofection - were used. A GFP vector (pLVTHM – 11Kb) of a size similar to CRISPR vector (9.1Kb) was used to track transfection efficiency by fluorescence microscopy and flow cytometry. HEK293 cells were used as a positive control of the transfection in both assays.

Three of five reagents (Lipofectamine 3000, TransIT-X2, TransIT-LT1) and Nucleofection were tested in the lung fibroblast cell line (MRC-5) (Figure 5.3). In this comparison, HEK293 cells median of transfected between the tested reagents was 54.55% (IQR 22.45%). The highest percentage of HEK293-GFP^{+} cells was obtained using TrasnIT-X2 (72.9%), and the lowest was using Lipofectamine 3000 (44%). In contrast, the median transfection efficiency in MRC-5 was 2.35% (IQR 2.44%) with the highest using Lipofection 3000 (3.5%) and the lowest, TransIt-X2 (0.6%). For nucleofection, the HEK293 presented 53% of positive cells for GFP, but MRC-5 had only 1.8% of transfected cells.

In primary cardiac fibroblasts, the reagents tested were Lipofectamine 3000, Viromir Plasmid, Viromir Red and TransIT-LT1 (Figure 5.4). At this assay, HEK293 cells presented a median of GFP expression of 30.9% (IQR 38.72%). The highest transfection efficiency percentage was obtained using Lipofectamine 3000 (75%), while the lower percentage of cells positive for GFP were gained using Viromir Plasmid (23.7%). Therefore, these data demonstrated that the reagents were working and could successfully transfect the plasmid into MRC-5 and HEK293

cells. However, in primary cardiac fibroblasts assessment, no transfected cells could be detected using any of the reagents.

Data have shown that the size of the plasmid could be inversely proportional to efficiency in transfections (270). Speculations for this correlation include poor membrane diffusion, lower motility of large plasmids into the cytoplasm and effects in transgene expression (270,271). Facing this information and the unsuccessful transfection rates using several reagents, we investigated if the size of the vector could influence the efficiency of the delivery. For that, a smaller GFP plasmid was used (pmaxGFP, 3Kb). The smaller GFP-expressing plasmid and the pVmTHL GFP-expressing (11Kb) were transfected into MRC-5 cells using TransIT-LT1 on the same conditions as before. HEK293 cells were, again, used as a positive control for this assessment (not shown). The efficiency of the transfections was measured by flow cytometry (data not shown). For the 11Kb plasmid, the efficiency was 0.26% GFP^+ cells, in comparison with the 3Kb plasmid efficiency of 3.25% GFP^+ cells. Fluorescence microscopy images also showed a higher number of cells transfected using a smaller plasmid compared to the ones transfected with the 11Kb GFP plasmid (Figure 5.5).

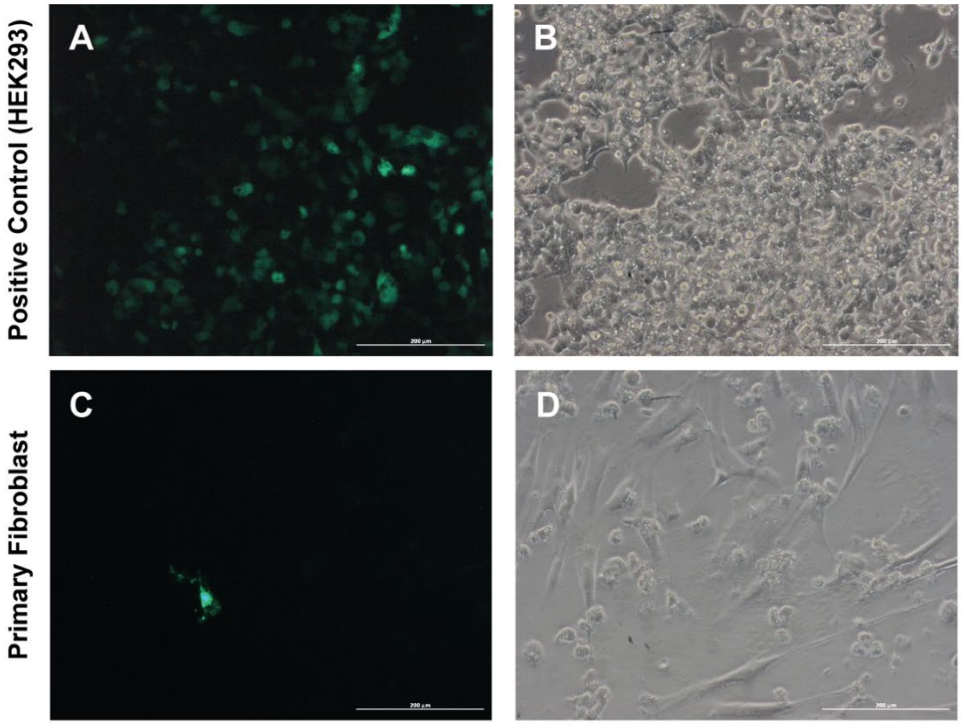


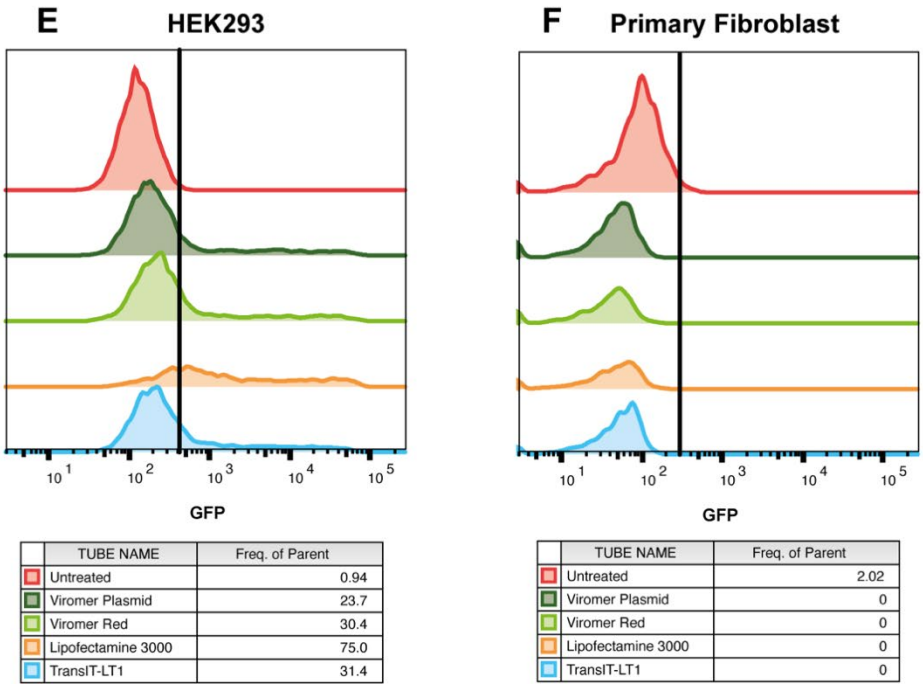
	TUBE NAME	Freq. of Parent
	Untreated	0.72
	TransIT-X2	72.9
	TransIT-LT1	56.1
	Lipofectamine 3000	44.0
	Nucleofection	53.0

	TUBE NAME	Freq. of Parent
	Untreated	0.72
	TransIT-X2	0.60
	TransIT-LT1	2.90
	Lipofectamine 3000	3.49
	Nucleofection	1.79

Figure 5.3 MRC-5 presented low rates of transfection

(A) Fluorescence microscopy of HEK293 cells after 48hrs of transfection; scale bar: 200µm. (B) Phase-contrast microscopy of HEK293 cells after 48hrs of transfection (same field as "A"); scale bar: 200µm. (C) Fluorescence microscopy of MRC-5 cells after 48hrs of transfection; scale bar: 200µm. (D) Phase-contrast microscopy of MRC-5 cells after 48hrs of transfection (same field as "C"); scale bar: 200µm. (E) Histograms showing the peaks of fluorescence in GFP^+ HEK293. (F) Histograms showing the peaks of fluorescence in GFP^+ MRC-5. N=1 for each reagent.





	TUBE NAME	Freq. of Parent
	Untreated	0.94
	Viromer Plasmid	23.7
	Viromer Red	30.4
	Lipofectamine 3000	75.0
	TransIT-LT1	31.4

	TUBE NAME	Freq. of Parent
	Untreated	2.02
	Viromer Plasmid	0
	Viromer Red	0
	Lipofectamine 3000	0
	TransIT-LT1	0

Figure 5.4 Primary cells could not be transfected with any tested condition

(A) Fluorescence microscopy of HEK293 cells after 48hrs of transfection; scale bar: 200µm. (B) Phase-contrast microscopy of HEK293 after 48hrs of transfection (same field as "A"); scale bar: 200µm. (C) Fluorescence microscopy of primary cells after 48hrs of transfection; scale bar: 200µm. (D) Phase-contrast microscopy of primary cells after 48hrs of transfection (same field as "C"); scale bar: 200µm. (E) Histograms showing the peaks of fluorescence in GFP^+ HEK293. (F) Histograms showing the peaks of fluorescence in GFP^+ primary cells. Each reagent had N=1.

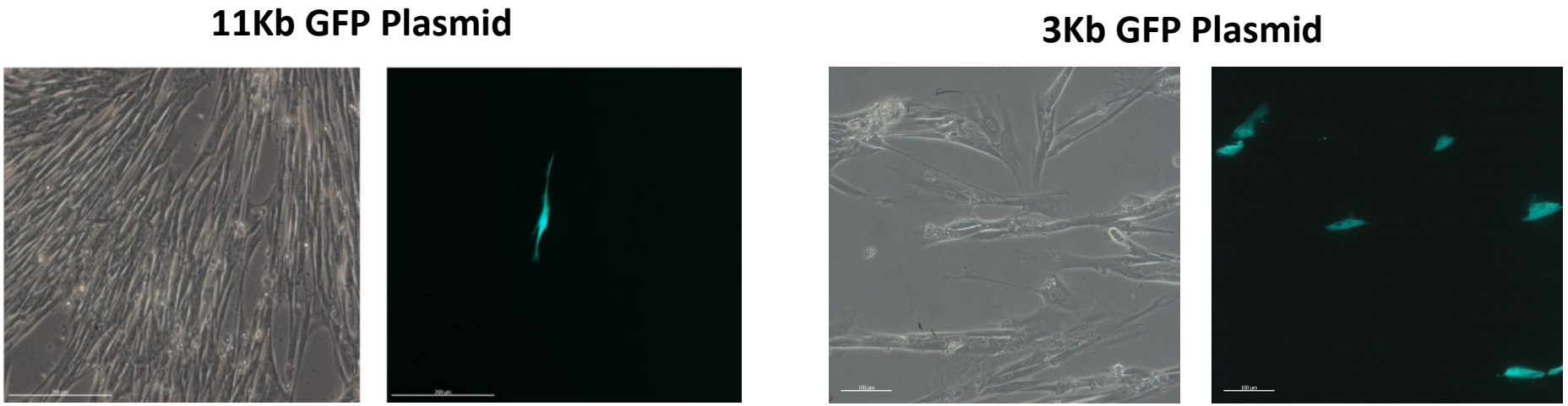


Figure 5.5 Transfection using a smaller (3Kb) plasmid showed transfection improvement

MRC-5 presented higher rates of transfection in comparison with transfection using a bigger (11Kb) plasmid. On the left-hand side, MRC-5 cells transfected with 11Kb GFP plasmid. On the right-hand side, MRC-5 cells transfected with 3Kb GFP plasmid. Both transfection assays were performed using same conditions. Scale bars: 200µm for the 11Kb GFP plasmid image, and 100µm for the 3Kb GFP plasmid image. N=1.

5.3 Lentiviral CRISPR delivery system did not create NR2F2 knockout in Fibroblasts

Nonviral delivery systems have advantages like lower cost, less optimization steps and lower immunogenicity compared to viral delivery systems. However, in terms of transgene delivery and expression, lentivirus-based methods are more efficient due its tropism to a broad range of cell types and its capacity to continuously express CRISPR/Cas9 elements after integration into targeted cells' genome (272,273).

In response to the previous complications to deliver CRISPR components to fibroblasts via direct transfection and the demand of increase transfection rate, lentivirus was further used. Plasmids containing Cas9 and NR2F2-targeted guide-RNA sequences as well as puromycin resistance gene, were transduced to fibroblasts through lentivirus to attempt knockout *NR2F2*. The viral particles were produced into packaging cells (HEK293) and the supernatant containing those particles was used directly onto fibroblasts.

Transduced cells were selected with a 5 to 7-day course of puromycin treatment. Both MRC-5 and primary fibroblasts transduced populations had majority of cells surviving the antibiotic selection (data not shown), indicating that the puromycin resistance gene was expressed in those cells. Genomic DNA was extracted from surviving cells to assess *NR2F2* mutation via T7 endonuclease (T7E1) mismatch detection assay. Briefly, T7E1 is an endonuclease which cleaves double-stranded DNA that contains mismatches formed via cross-annealing of mutated and wild-type strands. To assess the presence of mutations through this assay, extracted DNA from samples undergo hybridization and annealing cycles, followed by treatment with T7E1 and electrophoresis. Through this technique, the occurrence of mutations is confirmed by the presence of two smaller DNA fragments in edited samples due to cleavage of DNA heteroduplex after hybridization process, in contrast to one single (uncut) larger DNA fragment in wild-type samples (274,275).

Although most cells survived the antibiotic treatment, no cleaved fragments were detected by the T7 endonuclease mismatch detection assay, indicating the absence of DNA heteroduplex on those samples (Figure 5.6). The PCR amplicon had an expected size of 648bp which corresponded to the detected band size in the samples.

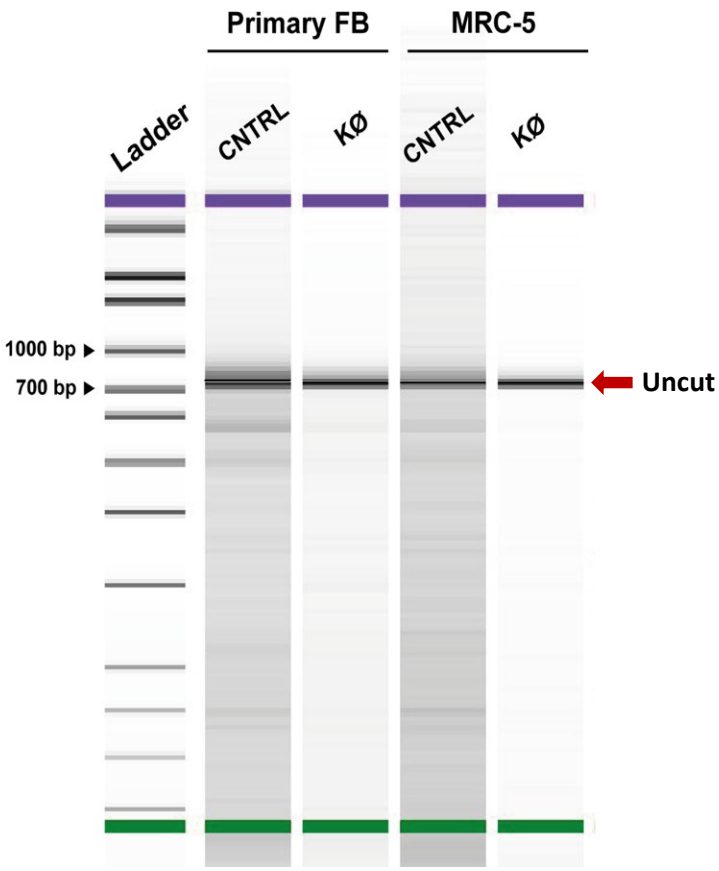


Figure 5.6 T7E1 mismatch detection assay implied no NR2F2 genomic mutation in fibroblasts

NR2F2 site in genomic DNA of each sample was amplified by. PCR and treated with T7E1 after temperature-induced hybridization. The expected PCR product was 698bp corresponding to the band detected in the samples. No mutation was detected as indicated by the undigested fragment (red arrow) and, consequently, the lack of smaller fragments. N=1.

5.4 NR2F2 was knocked down in Fibroblasts using CRISPR RNP system

Lentivirus is widely used in difficult-to-transfect cells, as primary fibroblasts, due its high rate of success. The data portrayed in Results 5.3 showed that the presence surviving cells after antibiotic implied successful transduction as expected with this application. However, the data failed to evidence any mutation into the genome of the cells by the lack of fragmented DNA after T7E1 assay.

Facing the challenges of using plasmid-delivering systems in hard-to-transfect cell such as fibroblasts, a more straightforward method had to be applied. RNPs improve the odds of getting edition into genome because no transcription and translation are required (276). To use RNP system, Cas9 protein and 3 small-guide RNA molecules were directly administrated to the nuclei of cells via nucleofection. In this method, no antibiotic selection was carried out

and the heterogenic pool of cells was assessed in the DNA, mRNA and protein levels to validate the gene mutation. Cells electroporated with only Cas9 were considered the negative control for this experiment.

To evaluate the occurrence of *NR2F2* mutation in genomic DNA, both mismatch detection using T7E1 and Sanger sequencing were performed. Differently from what was observed in the lentiviral system, both MRC-5 and primary fibroblast presented cleaved DNA when treated with T7E1 after hybridization cycles, indicating presence of indels due to non-homologous end-joint DNA repair after DNA double-strand break (Figure 5.7). Based on the traces generated on Sanger sequencing (Figure 5.8) and considering the position of the breaking point present with higher frequency, fragments sizes of 316bp and 163 bp were expected and subsequently observed in the gel. The wild-type PCR amplicon size was 498bp. To analyse the indels, PCR amplified DNA from heterogeneous population of edited primary fibroblasts and their respective negative control were sequenced. The traces were demultiplexed by ICE software (Figure 5.8), showing a proportion of frameshifts or indels ≥21bp (Knockout-Score) of 96%. A deletion of 144bp was detected in the majority (67%) of the trace mixture (Figure 5.9). The location of this deletion size in the base call sequences matched the cut point observed in the brightest fragment bands in T7E1 assay gel. At transcript level, the knockout of *NR2F2* was assessed by qPCR and the knockout percentage was calculated using negative control expression levels as reference. MRC-5 showed 53% reduction on NR2F2 mRNA expression compared to negative control. Primary fibroblasts had 51% of knockdown on NR2F2 transcript levels in comparison with their paired negative control (Figure 5.10). NR2F2 protein presence was evaluated by Western blotting. HEK293 lysate was used as positive control for NR2F2 content, and three primary fibroblasts protein lysates (wild-type, negative control and edited) were analysed. The wild-type sample had the same protein size band for NR2F2 protein (46kDa) as the negative and positive controls. On the other hand, the expected size band for NR2F2 (46kDa) was absent in the edited protein lysate. Instead, a 40kDa band was observed. Taking in consideration the most frequent deletion (-144bp) present on this population of cells, this size band matches the predicted truncated protein that this mutation would translate (Figure 5.11).

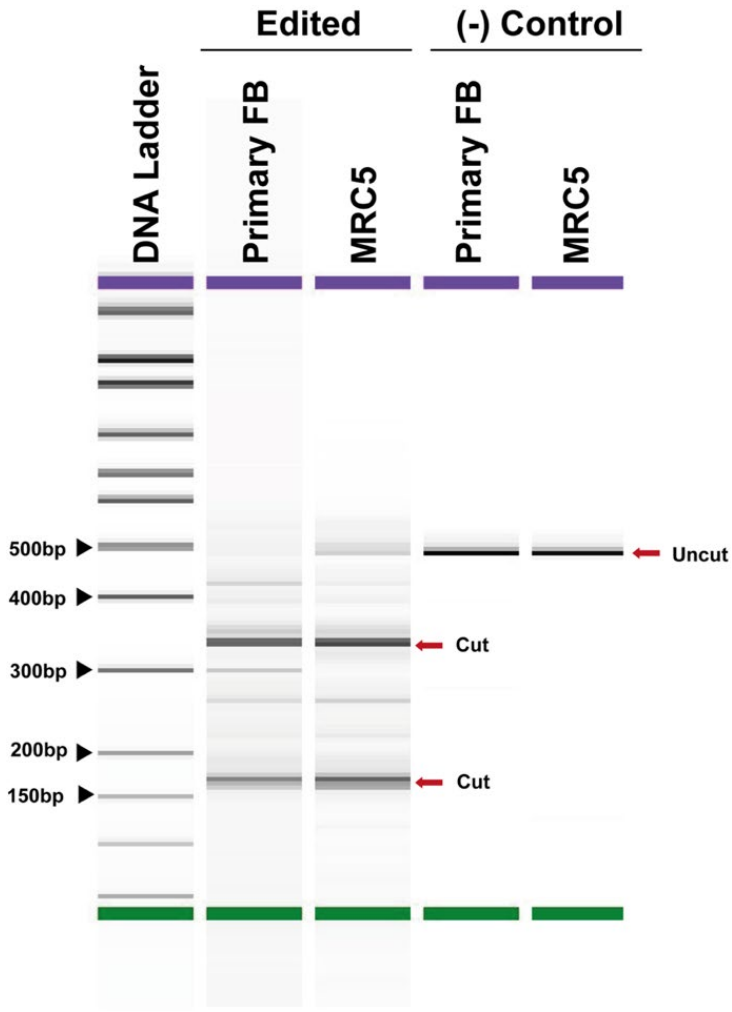


Figure 5.7 T7E1 detected DNA mismatches in fibroblasts due to CRISPR/Cas9 genomic editing

NR2F2 exon in which the 3 guides anneal to was amplified by. PCR and treated with T7E1 after temperature-induced hybridization. The expected wild-type PCR product was 479bp. Mutation was detected as indicated by the cut fragments. The expected sizes of the fragments are 316bp and 163bp. N=1.

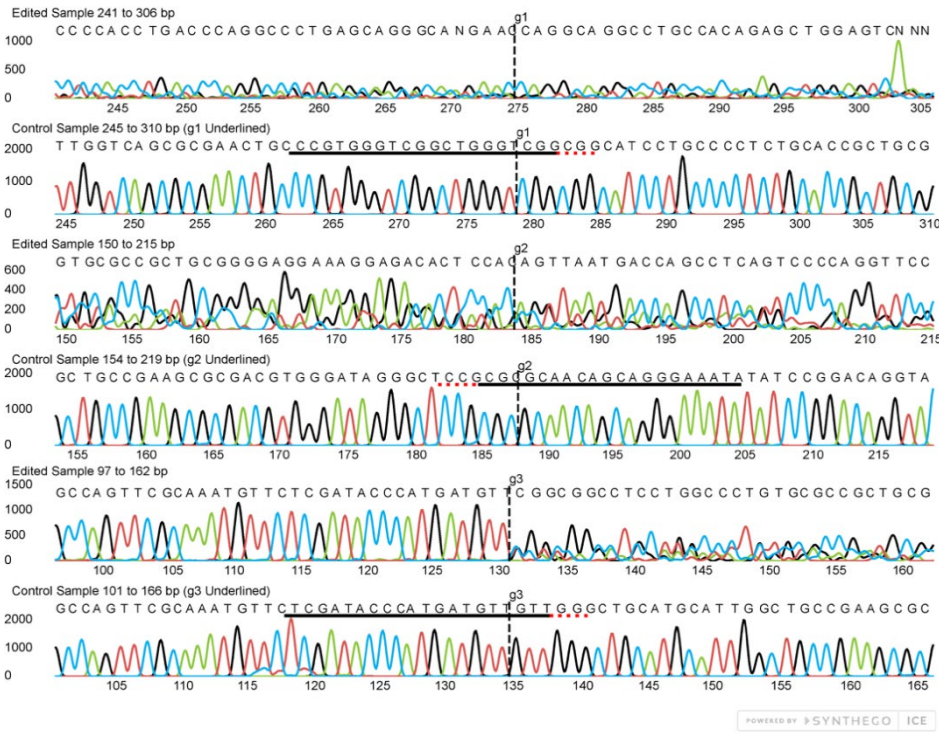


Figure 5.8 Sequence base calls from edited primary fibroblasts

Sanger sequence traces of edited and wild-type (control) samples. The horizontal black underline sequences represent the align location of the 3 guides. The horizontal red dotted underline is marking the PAM sequence of each guide. The vertical black dotted line shows the DNA double-strand break point. The mixed traces are result of error-prone repair that happens in non-homologous end-joint DNA repair. N=1.

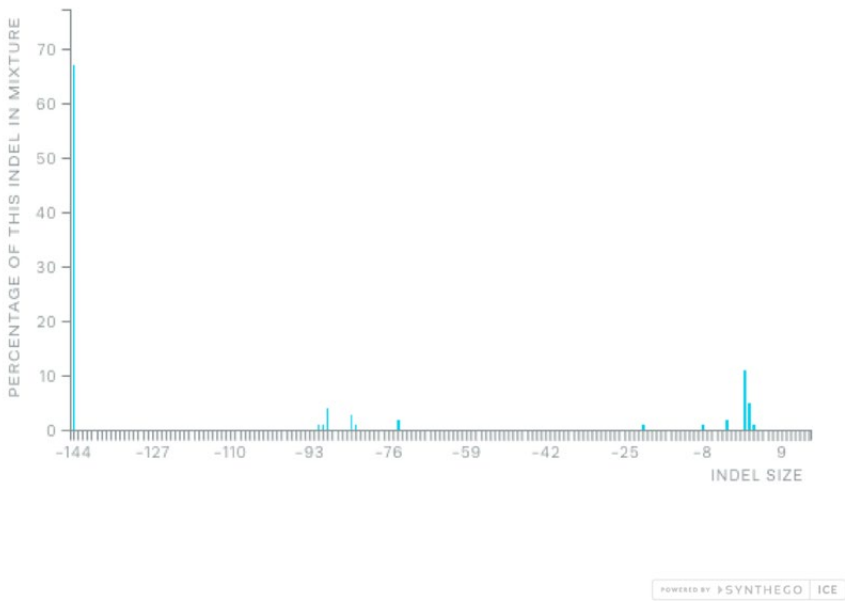


Figure 5.9 Percentage of Indels in edited primary fibroblasts mixed population

Percentage of Indels in edited primary fibroblasts mixed population. The x axis shows the size of insertion of deletion. The y axis shows the percentage of genomes that contains the indels. N=1.

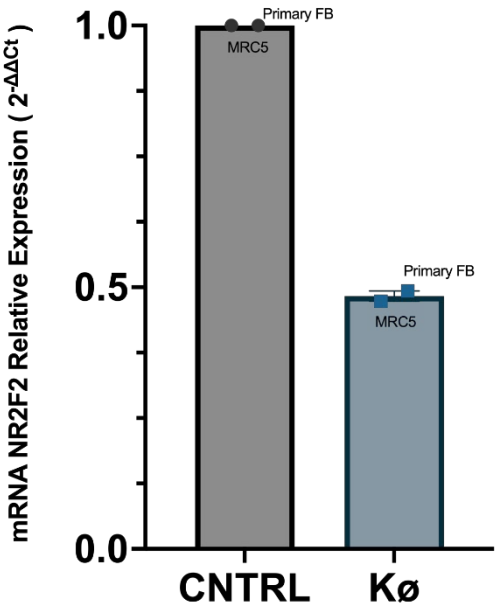


Figure 5.10 Fibroblasts have shown knockdown on NR2F2 expression

MRC-5 showed 53% reduction on NR2F2 mRNA expression ($2^{-\Delta\Delta CT}$ = 0.47) in knockdown samples compared to their paired negative control expression. Primary fibroblasts had a decrease of 51% on NR2F2 transcript levels ($2^{-\Delta\Delta CT}$ = 0.49) in knockdown population compared to their paired negative control.

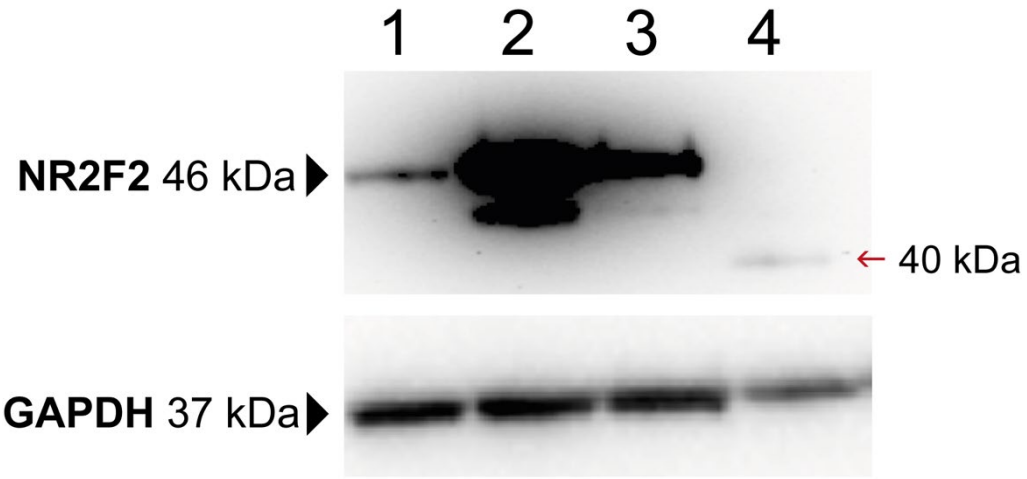


Figure 5.11 Absence of NR2F2 protein in edited primary fibroblasts

The expected size of NR2F2 protein is 46kDa. The 40kDa band shows a predicted truncated NR2F2 protein. GAPDH was used as loading reference. Lane 1: HEK293 (positive control). Lane 2: Wild-type primary fibroblasts. Lane 3: Primary fibroblast Cas9-only (negative control). Lane 4: Edited primary fibroblasts.

5.5 Discussion

NR2F2 has been shown to participate in heart development since variations on its structure often goes along with phenotypic anomalies in the cardiac development. However, the molecular mechanisms involving NR2F2 in this developmental process is not fully understood. The combination of genetic perturbations and omics analyses is a powerful tool to unveil the function of genes and their role in biological processes. Therefore, to unveil the molecular network behind NR2F2 in the heart, it is pivotal to interrupt the gene and assess the transcriptomic implications trigged by its disruption.

To obtain a successful knockout assay in a challenging cell type such as primary fibroblasts (277), it is necessary to use a simple and effective method. CRISPR/Cas9 is straightforward and offers several options of cargo and delivery applications, making it easily adaptable to all experimental needs. Plasmid DNA is the most used cargo due to its low cost and easier manipulation, but it demands optimisation for each cell type. In this study, five different transfection reagents and nucleofection were tested to carry exogen pDNA into this cell type. Due to the large size of Cas9 sequence, the final plasmid size (>7kb) can be problematic to an efficient delivery (265). For all reagents, the positive control (HEK293 cells) presented good transfection efficiency – around 50% of transfected cells, contrasting to low efficiency rates in fibroblasts. Although similar protocols were used in this assessment, studies showing efficient transfection protocol in primary fibroblasts used a smaller plasmid. This evidence led to speculation that large plasmids could be causing the low efficiency. To test this hypothesis, a smaller pDNA was transfected using one of the reagents in the same experimental condition as the larger pDNA. A robust improvement in the transfection efficiency was visually noticed by fluorescence microscopy, suggesting that the large size pDNA initially used would be a reason for the low efficiency.

To address this problem, other delivery systems were further tested. Studies demonstrate CRISPR/Cas9 system being applied using lentivector system (268,278,279). Although, virus-based systems present higher efficiency compared to non-viral methods in primary fibroblasts (277), in here, it did not lead to *NR2F2* interruption. After lentiviral transfection with the CRISPR/Cas9 components, there were cells surviving the antibiotic selection, suggesting that they had been successfully transduced. However, when the genome of these cells was subjected to the T7E1 assay, there were no extra fragments pointing towards the presence of a CRISPR/Cas9 mutation in the DNA. A possible explanation for

these results is that the pDNA was delivered into the cell, but one or all elements of CRISPR/Cas9 failed to be transcribed or they were not sufficiently expressed to then integrate into the genomic DNA.

Considering these unsuccessful attempts to disrupt NR2F2, the RNP approach is an alternative to resolve the issues previously described. This method using a complex formed by the Cas9 protein and guide RNAs is transient and does not demand transcription or translation, which reduces off-target effects and makes the gene editing process quicker and efficient. In this result chapter, the data showed reduction on NR2F2 expression on both transcript and protein level, as well as the identification of significant deletions on the DNA caused by the RNP complex. The complex was carried onto the cell through nucleofection. While no expression of NR2F2 protein was observed, qPCR showed up to 53% knockdown on mRNA relative expression. This discrepancy could be justified by the presence of heterogeneous mutation in the mixed population being detected by qPCR due its higher sensitivity in comparison to Western blotting assay. Another hypothesis is that the nonsense-mediated mRNA decay does not happen as fast as the protein degradation, which could lead to the detection of mRNA – although reduced – in the qPCR, but no protein detection in the Western blotting assay, explaining the contradicting results in Figure 5.10 compared to Figure 5.11. Moreover, a band of 40kDa was observed which matches the weight of a predicted protein generated *in silico* from the traces with 144bp deletion (not shown). The deleted bases were located in the exon 5 of NR2F2 gene, precisely in the region that codifies the transcription factor DNA-binding domain. This could indicate the presence of a truncated dysfunctional NR2F2 protein. Regarding MRC-5 cells' protein assessment, the amount of cells collected after CRISPR/Cas 9 knockdown assay was lower in comparison to primary fibroblasts, which led to a undetectable amount of protein extracted from this group for Western blotting. This limitation averted the assessment and comparison of the results presented in Figure 5.11 with MRC-5 cells.

The collective data presented in this chapter leads to believe that NR2F2 was successfully knocked out in foetal primary fibroblasts and MRC-5 cell lines which directed to transcriptomic investigations crucial to further understand the role of NR2F2 in those cells.

Chapter 6 Results

Transcriptomic analysis of fibroblasts with NR2F2 edited by CRISPR/Cas9 system

6.1 Introduction

Several approaches are used to unveil the role of key transcription factors (TFs) in the heart development. The studies can focus on the identification of expression pattern of those transcription factors (280), morphology changes due to depletion of the gene (281) or identification of DNA-binding site in targeted genes (282), among others (283).

Another approach is the combination of gene perturbations and transcriptomic analysis has been used to identify the role of TFs in development. A group investigated the role of *OCT4* in preimplantation events by knocking out it via CRISPR/Cas9 and preforming single-cell RNA sequencing. They found that the loss of *OCT4* caused mis-expression of trophectoderm markers (284). Another study, in cardiogenesis field, *NKX2-5*, an important cardiogenic TF, was deleted from cardiomyocytes derived from human embryonic stem cells. An RNA sequencing from those cells revealed that *NKX2-5* function is mediated by *HEY2* in ventricular development (285).

The function of NR2F2 was also analysed using the combined approach to investigate its role in myocardial cells. NR2F2 was ablated by a Cre-recombinase technique in mice. The myocardium transcriptome was explored by microarray assay concluding that NR2F2 is important for atrial identity of cardiomyocytes (286).

In this chapter, we use the same combined approach to understand the role of NR2F2 by analysing the global transcriptomic changes that the reduction of NR2F2 expression cause in fibroblasts.

6.2 NR2F2 transcriptional activity inference is reduced in NR2F2-knockeddown fibroblasts

Once lung fibroblast cell-line (MRC-5) and cardiac primary fibroblasts had *NR2F2* gene knocked down via CRISPR/Cas9 method, single-cell RNA sequencing was performed on the control and knocked down samples from these both sources of fibroblast cells. The method

for sequencing, data quality control assessment and visualisation were described in Methods 2.9. In the quality control assessment step, cells presenting read counts below 500 and above 50000 were filtered out. Cell expressing less than 2000 genes were also taken out of the analysis. The number of cells that passed the filtering criteria in each sample is informed in Table 6.1. Genes expressed in less than 5 cells were also excluded. The cell quality assessment of the dataset is shown in Appendix C.2. Cell clusters were annotated using PanglaoDB (216), a prior knowledge database of cell markers accessed via Decoupler library (217). The annotation data is shown in Figure 6.1. It was observed that cells were annotated as fibroblasts by this method. Individual gene markers for cardiomyocytes, endothelial cells, fibroblasts, and smooth muscle cells were also assessed (Figure 6.2). The gene markers used were determined by a study in human foetal hearts (age 19-22 pcw) (148), and they are listed in Table 6.2. It was observed that cells from both control and knockdown samples from primary fibroblasts expressed all gene markers from fibroblasts, consistent to the automatic annotation from PanglaoDB. Regarding the other cell type gene markers, only one gene marker from cardiomyocytes (*MYL9*) and two gene markers from smooth muscle cells (*TAGLN* and *ACTA2*) were expressed. No expression of endothelial cell gene markers was detected. MRC-5 cells (control and knockdown) presented similar expression pattern and the data is shown in Appendix C.4.

Table 6.1 Number of cells per sample after filtering sequencing data

Sample	#Cells
PFB CNTRL	1632
PFB KO	952
TOTAL	**2584**
MRC-5 CNTRL	1392
MRC-5 KO	1303
TOTAL	**2695**

PFB: cardiac primary fibroblasts; MRC-5: foetal lung fibroblast cell line; CNTRL: control samples; KO: NR2F2 knocked down samples

Table 6.2 Gene markers used to assess cell types in CRISPR/Cas9 edited fibroblasts

Cell type	Marker genes
Cardiomyocytes	*MYL3, MYL4, MYL9, MYH7B, RYR2, TTN, PLN*
Endothelial cells	*CDH5, PECAM-1*
Fibroblasts	*COL1A1, COL1A2, COL3A1, TGFBI, DCN, LUM, FN1, BGN*
Smooth muscle cells	*ACTA2, TAGLN, RGS5*

Comparing *NR2F2* mRNA counts in each sample, it is observed in the collection of MRC-5 cells that the median of counts has no difference between control and edited cells, and in primary fibroblasts, the edited cell sampling presented higher number of counts for NR2F2 mRNA compared to control sample (Figure 6.3 – left column). However, to understand the effects caused by the lack of NR2F2 function in fibroblasts, Decoupler library also uses an integrated resource that scores transcription factor-target interaction to infer transcriptional activities. Decoupler creates a reference matrix using a curated collection (i.e., Dorothea (287)) containing the transcription factors, their regulons and the weight of their interaction based on the confidence level of the evidence and the sign of its mode of regulation (activation or inhibition). After, Decoupler runs a Multivariate Linear Model method that models the data gene expression by using the regulatory adjacency matrix created using Dorothea prior knowledge collection as covariates of a linear model. The obtained t-values of the fitted model correlates with the activity scores. It is observed that the edited samples from both MRC-5 and cardiac primary fibroblasts had reduced NR2F2 transcriptional activity (Figure 6.3 – right column).

After establishing that NR2F2 transcriptional activity inference was reduced, a differential expression gene (DEG) analysis was performed between the conditions (control and knockdown). If DEG is performed treating each cell as a sample, the *p*-value could be inflated and the variation across the population of cells be lost. To mitigate it, a pseudo-bulk of the cell populations was performed using the Decoupler library. Control and knockdown were compared, and a ranked t-test (Wilcoxon) was used to assess the differently expressed genes (DEG). To select the DEG, the genes were filtered out by a *p*-value ≤ 0.05 and a $-1.5 \geq \log_2(FC) \geq 1.5$. The list of down and up-regulated genes are listed in Appendix C.6. When applied these cut-offs, for MRC-5, 37 genes appeared upregulated,

and one gene (*THRB*) was downregulated. For primary fibroblasts, 154 genes were upregulated, and fifty-four genes were downregulated. Two upregulated genes were common between MRC-5 and primary fibroblasts (*NFIB* and *PLD5*).

Figure 6.1 Cell annotation using PanglaoDB in CRISPR/Cas9 edited primary fibroblasts

UMAP plots of primary fibroblasts samples – CNTLR and KO – distribution and PanglaoDB annotation (first row). Bellow, UMAP plots showing each cell (dots) coloured according to the Fisher's exact test score (-log10[*p*-value]) for fibroblasts, endothelial cells, cardiomyocytes, and smooth muscle cells. FB: fibroblasts; CNTRL: control; KO: CRISPR/Cas9 edited cells.

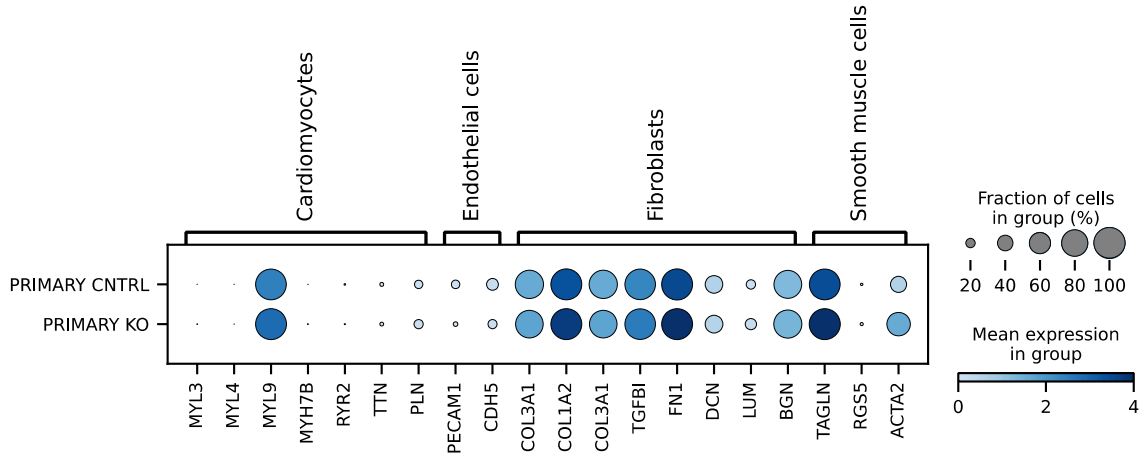


Figure 6.2 Individual expression of cardiac cell type's gene markers in CRISPR/Cas9 primary cardiac fibroblasts

Dot plot showing the expression of gene markers to identify cardiomyocytes, endothelial cells, fibroblasts, and smooth muscle cells in CRISPR/Cas9 edited primary fibroblasts from foetal hearts. Each dot represents the percentage of cells (size of the dot) expressing the gene (x axis) and the mean expression (intensity of the colour) within each cluster (y axis).

Following DEG analysis, the gene ontology (GO) of biological processes enriched in each set of up and downregulated genes were identified by ShinyGo v0.76 (184). The top 20 biological processes enriched were selected by the lowest values of False Discovery Rate (FDR), and then ranked by Fold Enrichment. The Fold Enrichment is defined by the ratio of a percentage of the gene set that belongs to a gene ontology term per the percentage in the list background. At the MRC-5's set of upregulated genes the most enriched term in downregulated genes was the "Cardiac right ventricle morphogenesis" (GO:0003215) with *HAND2*, *TBX20* and *GATA4* genes related to it (Appendix C.7). Following this, most of the other enriched terms were related to nervous system ad neuron cell's development. For the upregulated genes in the cardiac primary fibroblasts, the most enriched biological processes were identified as "Regulation of cardioblast differentiation" (GO:0051890), "Regulation of cardiocyte differentiation" (GO:1905207) and "Cardiac chamber development" (GO:0003205). For the downregulated gene set, the biological processes enriched were "Cardiac epithelial to mesenchymal transition" (GO:0060317) and "Ventricular septum morphogenesis" (GO:0060412), in addition to terms related to tube development and angiogenesis (Figure 6.4). The summary of the GO analysis of biology processes for cardiac primary fibroblasts samples are shown in Table 6.3 and Table 6.4.

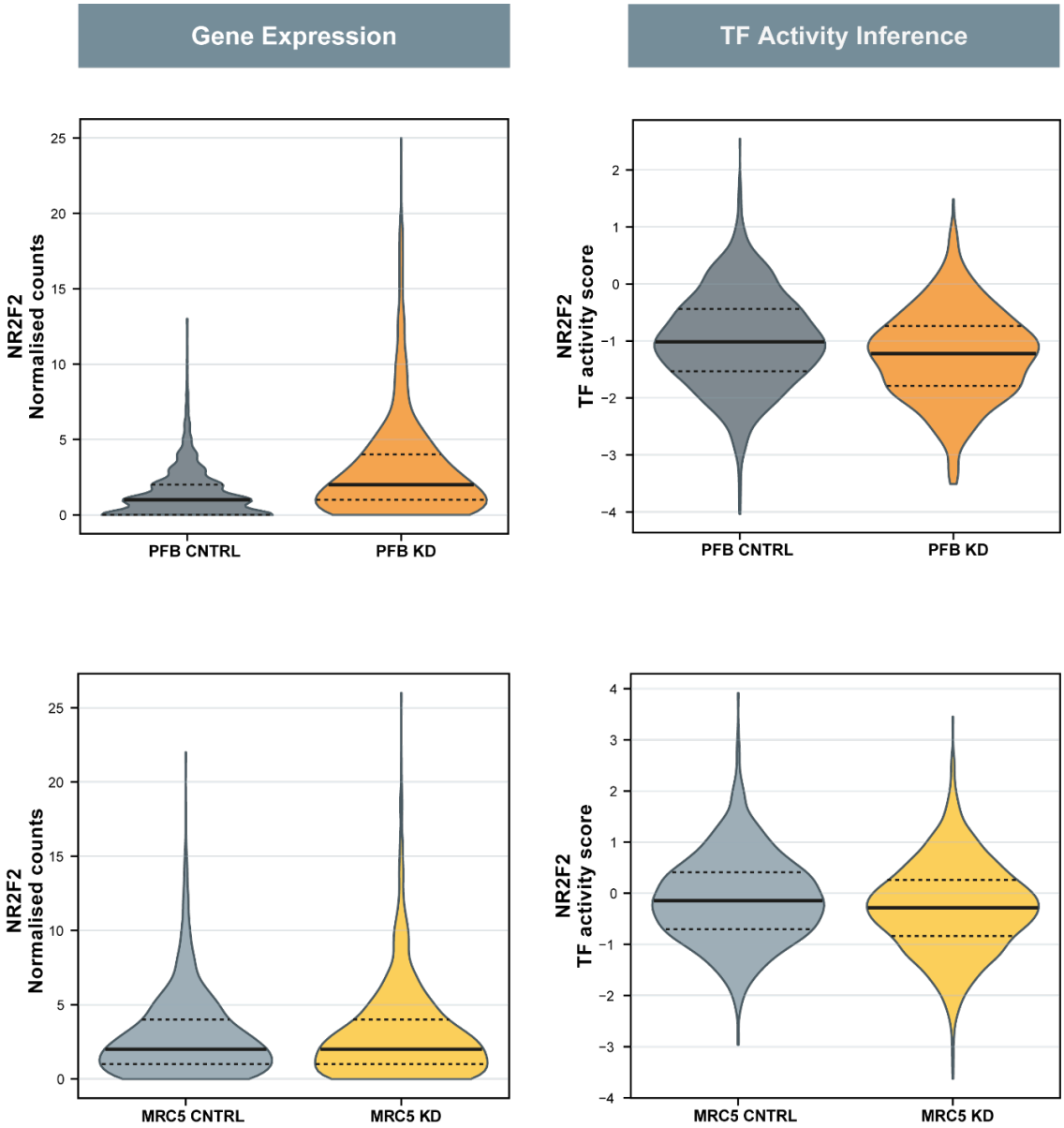


Figure 6.3 Expression of mRNA versus transcriptional activity of NR2F2 in Fibroblasts.

The left column shows the number of counts of NR2F2 mRNA in each sample. For primary fibroblasts (top), it is shown that edited fibroblasts presented higher median (median=2) compared to control (median=1). For MRC-5 (bottom), the median of control and edited samples shows no difference (median=2). The right column shows the transcriptional activity of NR2F2. For primary fibroblasts (top), in the knockdown sample, NR2F2 presented lower transcriptional activity (median=-1.22) compared to control sample (median=-1.01). For MRC-5 (bottom), the same pattern is observed. The edited fibroblasts presented reduced activity for NR2F2 (median=-0.28) compared to unedited sample (median=-0.15). The dotted black lines represent the 25th and the 75th quartiles. The full black lines represent the median values. TF: transcription factor; PFB: foetal cardiac primary fibroblasts; MRC-5: foetal lung fibroblast cell line; KD: knockdown; CNTRL: control.

Table 6.3 Gene Ontology of Biological Processes enriched in the upregulated gene set of cardiac primary fibroblasts

Enrichment FDR	Fold Enrichment	Biological Process	Genes
1.92E-03	63.3	Regulation of cardioblast differentiation	*TGFB2 PRICKLE1 GATA6*
2.63E-03	25.3	Regulation of cardiocyte differentiation	*TGFB2 PRICKLE1 GATA6 ABI3BP*
6.43E-05	9.1	Cardiac chamber development	*KCNK2 TGFB2 DSP CPE GATA6 ANK2 ADAMTS1 SLIT3*
1.38E-05	8.0	Cell-cell adhesion via plasma-membrane adhesion molecules	*CLDN11 CDH10 ICAM1 TGFB2 MYPN PCDH10 HMCN1 SDK1 LRRC4C CLDN1 NLGN1 PCDH7*
7.69E-04	7.4	Potassium ion transmembrane transport	*NEDD4L KCNK2 KCNK6 NALCN RGS4 ANK2 KCNE4 KCNT2 KCNQ5*
2.44E-03	7.2	Response to ketone	*ICAM1 TGFB2 GHR FOXP2 POSTN CLDN1 OXTR SLIT3*
8.77E-06	6.3	Extracellular matrix organization	*NTN4 ICAM1 CCDC80 TGFB2 LOX POSTN CYP1B1 COL8A1 HAPLN1 ABI3BP ADAMTS1 PTX3 MFAP4 ADAMTSL1 MFAP5*
8.77E-06	6.3	Extracellular structure organization	*NTN4 ICAM1 CCDC80 TGFB2 LOX POSTN CYP1B1 COL8A1 HAPLN1 ABI3BP ADAMTS1 PTX3 MFAP4 ADAMTSL1 MFAP5*
8.77E-06	6.3	External encapsulating structure organization	*NTN4 ICAM1 CCDC80 TGFB2 LOX POSTN CYP1B1 COL8A1 HAPLN1 ABI3BP ADAMTS1 PTX3 MFAP4 ADAMTSL1 MFAP5*
1.05E-03	5.5	Cell-substrate adhesion	*LIMCH1 NTN4 EPDR1 CCDC80 BST1 POSTN COL8A1 ABI3BP ANGPT1 EDIL3 ITGBL1*

Enrichment FDR	Fold Enrichment	Biological Process	Genes
2.25E-04	5.5	Regulation of membrane potential	*NEDD4L KCNK2 DSP KCNK6 NALCN FGF14 RGS4 IFI6 ANK2 KCNE4 GRIK2 SCN9A NLGN1*
1.44E-04	4.9	Heart development	*KCNK2 TGFB2 DSP CPE LOX RGS4 PRICKLE1 GATA6 ANK2 ANKRD1 ABI3BP ADAMTS1 OXTR SLIT3 CRIP1*
3.28E-03	4.7	Muscle system proc.	*HSPB6 NEDD4L DSP ACTA2 RGS4 GATA6 STAC ANK2 KCNE4 LMOD1 OXTR*
3.05E-09	3.9	Cell adhesion	*CLDN11 CDH10 LIMCH1 NTN4 EPDR1 ICAM1 CCDC80 TGFB2 DSP VSIR BST1 SELPLG POSTN GPNMB CYP1B1 MYPN PCDH10 HMCN1 COL8A1 HAPLN1 SDK1 LRRC4C ABI3BP ANGPT1 CNTNAP3B PKD1L1 CLDN1 EDIL3 MFAP4 STXBP6 NLGN1 PCDH7 ADAMTSL1 ITGBL1*
3.05E-09	3.9	Biological adhesion	*CLDN11 CDH10 LIMCH1 NTN4 EPDR1 ICAM1 CCDC80 TGFB2 DSP VSIR BST1 SELPLG POSTN GPNMB CYP1B1 MYPN PCDH10 HMCN1 COL8A1 HAPLN1 SDK1 LRRC4C ABI3BP ANGPT1 CNTNAP3B PKD1L1 CLDN1 EDIL3 MFAP4 STXBP6 NLGN1 PCDH7 ADAMTSL1 ITGBL1*
3.28E-03	3.9	Circulatory system proc.	*SLC38A5 NEDD4L ICAM1 TGFB2 DSP KCNK6 ACTA2 RGS4 POSTN ANK2 KCNE4 ANGPT1 OXTR*
8.77E-06	3.9	Circulatory system development	*HSPB6 KCNK2 TGFB2 DSP ACTA2 CPE LOX CTH RGS4 GPNMB CYP1B1 PRICKLE1 GATA6 COL8A1 ANK2 RSPO3 ANKRD1 ABI3BP ANGPT1 ADAMTS1 OXTR SLIT3 CRIP1*
3.09E-04	3.6	Cell-cell adhesion	*CLDN11 CDH10 ICAM1 TGFB2 DSP VSIR SELPLG GPNMB CYP1B1 MYPN PCDH10 HMCN1 SDK1 LRRC4C PKD1L1 CLDN1 STXBP6 NLGN1 PCDH7*

Enrichment FDR	Fold Enrichment	Biological Process	Genes
4.41E-04	3.3	Animal organ morphogenesis	*NTN4 TGFB2 DSP ACTA2 CPE VDR GHR PAPPA2 PRICKLE1 GATA6 COL8A1 RSPO3 SDK1 NFIB ANKRD1 ADAMTS1 SLIT3 ROR1 MFAP5 CRIP1*

FDR: False discovery rate. FDR is calculated based on the nominal *p*-value, which states the likelihood of enrichment by chance. The fold enrichment is the percentage of genes in the list of differentially expressed genes belongs to the biological process, indicating the overrepresentation. The biological processes are ranked by Fold Enrichment.

Table 6.4 Gene Ontology of Biological Processes enriched in the downregulated gene set of cardiac primary fibroblasts

Enrichment FDR	Fold Enrichment	Biological Process	Genes
4.05E-03	46.0	Cardiac epithelial to mesenchymal transition	*SPRY1 HEY1 NOG*
6.34E-03	37.5	Ventricular septum morphogenesis	*HES1 HEY1 NOG*
2.12E-03	31.1	Smooth muscle cell differentiation	*NFATC2 OLFM2 HES1 HEY1*
2.23E-03	9.1	Muscle cell differentiation	*NFATC2 TNNT1 OLFM2 HES1 IGFBP5 SORBS2 HEY1*
3.62E-03	6.4	Positive regulation of cellular component biogenesis	*NOX4 RGCC CSF3 ID1 MMP3 CSF2 SLITRK4 MMP1*
1.18E-03	6.4	Vasculature development	*CALCRL NFATC2 RGCC LAMA4 HES1 ID1 ROBO4 HEY1 ANGPTL4 NOG*
4.51E-03	6.1	Muscle structure development	*NFATC2 TNNT1 OLFM2 HES1 IGFBP5 SORBS2 HEY1 NOG*
4.67E-03	6.0	Blood vessel morphogenesis	*CALCRL RGCC HES1 ID1 ROBO4 HEY1 ANGPTL4 NOG*

Enrichment FDR	Fold Enrichment	Biological Process	Genes
2.48E-03	6.0	Blood vessel development	*CALCRL RGCC LAMA4 HES1 ID1 ROBO4 HEY1 ANGPTL4 NOG*
5.50E-04	5.6	Negative regulation of multicellular organismal proc.	*CALCRL RGCC TNNT1 LAMA4 HES1 IGFBP5 SOCS2 ID1 ADAMTS18 STC1 SPRY1 HEY1 NOG*
2.23E-03	5.4	Tube morphogenesis	*CALCRL RGCC HES1 ID1 ROBO4 EDA SPRY1 HEY1 ANGPTL4 NOG*
5.61E-04	5.2	Anatomical structure formation involved in morphogenesis	*CALCRL NFATC2 RGCC TNNT1 HES1 ID1 ROBO4 EDA SPRY1 HEY1 FAT3 ANGPTL4 NOG*
1.18E-03	5.1	Circulatory system development	*CALCRL NFATC2 RGCC LAMA4 HES1 ID1 ROBO4 SORBS2 SPRY1 HEY1 ANGPTL4 NOG*
3.42E-03	4.9	Negative regulation of developmental proc.	*NFATC2 RGCC HES1 IGFBP5 SOCS2 ID1 SPRY1 HEY1 FAT3 NOG*
2.23E-03	4.8	Tube development	*CALCRL RGCC HES1 IGFBP5 ID1 ROBO4 EDA SPRY1 HEY1 ANGPTL4 NOG*
1.18E-03	4.5	Positive regulation of developmental proc.	*NFATC2 RGCC MEDAG OLFM2 CSF3 HES1 SOCS2 SPRY1 CSF2 ANGPTL4 PDE3A SLITRK4 NOG*
3.62E-03	4.0	Regulation of multicellular organismal development	*NFATC2 RGCC CSF3 LAMA4 HES1 ID1 SPRY1 CSF2 HEY1 ANGPTL4 SLITRK4 NOG*
2.12E-03	3.8	Regulation of cell differentiation	*NFATC2 RGCC MEDAG OLFM2 CSF3 HES1 IGFBP5 SOCS2 ID1 SPRY1 CSF2 HEY1 PDE3A NOG*

Enrichment FDR	Fold Enrichment	Biological Process	Genes
2.48E-03	3.6	Regulation of cell death	*NOX4 RGCC CSF3 SOCS2 ID1 MMP3 SPRY1 CSF2 ANGPTL4 BDKRB2 PDE3A TCIM NOG TNFRSF6B*
7.19E-03	3.6	Cell migration	*NOX4 NFATC2 RGCC LAMA4 HES1 IGFBP5 ID1 ROBO4 STC1 FAT3 NOG MMP1*

FDR: False discovery rate. FDR is calculated based on the nominal *p*-value, which states the likelihood of enrichment by chance. The fold enrichment is the percentage of genes in the list of differentially expressed genes belongs to the biological process, indicating the overrepresentation. The biological processes are ranked by Fold Enrichment.

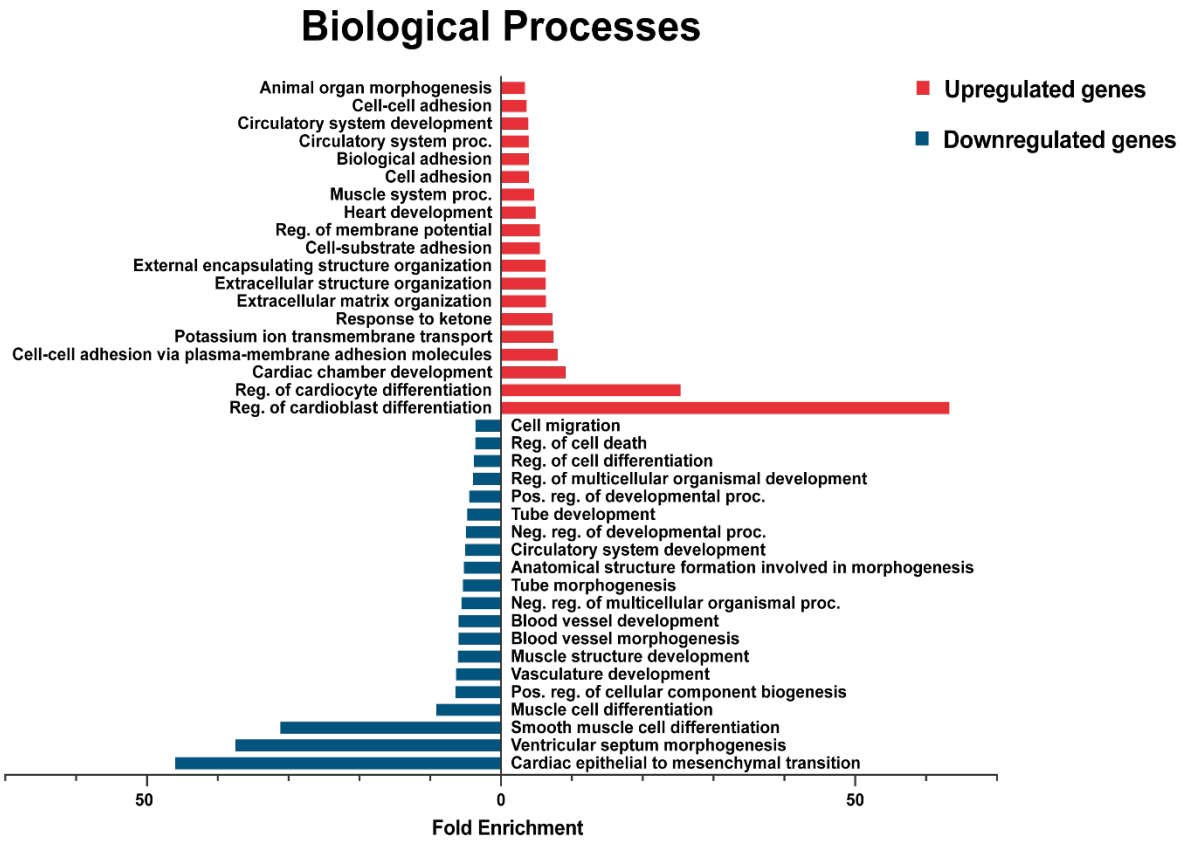


Figure 6.4 Biological processes enriched in the DEG set from cardiac primary fibroblasts

The biological processes were selected by the top 10 significant FDR values (in a list of FDR ≤ 0.05) and ranked by enriched fold enrichment. The upregulated genes in the NR2F2-knockeddown fibroblasts were related to heart development and other cardiogenic processes, focusing on "regulation of cardiocyte differentiation" and "cardiac chamber development" that are mostly enriched in this gene set. The downregulated genes in NR2F2-knockeddown cells are also mainly related to cardiogenic processes, with "Cardiac epithelial to mesenchymal transition", "Ventricular septum morphogenesis" and "Smooth muscle cell differentiation" being the most enriched in this gene set.

6.3 Lack of NR2F2 activity perturbed other important cardiogenic TFs inferred activities

In addition to NR2F2, a global inference analysis of transcription factor (TF) activities was carried out using Decoupler. Scanpy library (175) was used to compare the activity of the transcription factors between control and knockdown populations using Wilcoxon method and ranked by z-score.

In both MRC-5 and cardiac primary fibroblasts a difference in transcription factors activity between control and NR2F2-knocked down samples was observed (Figure 6.5). Contrasting with MRC-5, the knockdown cardiac primary fibroblasts showed reduction of activity in important cardiogenic transcription factors, as NKX2-5 and SOX4 (288,289). Other TFs are known for forming dimers or being targets of NR2F2, as PROX1 and HEY1 (113). Among the

TFs with increased activity, it was observed a gene from NR family, NR2C2. Also, a known activator of NR2F2, ETS1 (290), presented higher activity in samples knockdown for NR2F2. MRC-5 and cardiac primary fibroblasts shared 2 TFs with increased activity, ETS1 and MYC, known being part of Neural Crest Cell differentiation pathway(291,292).

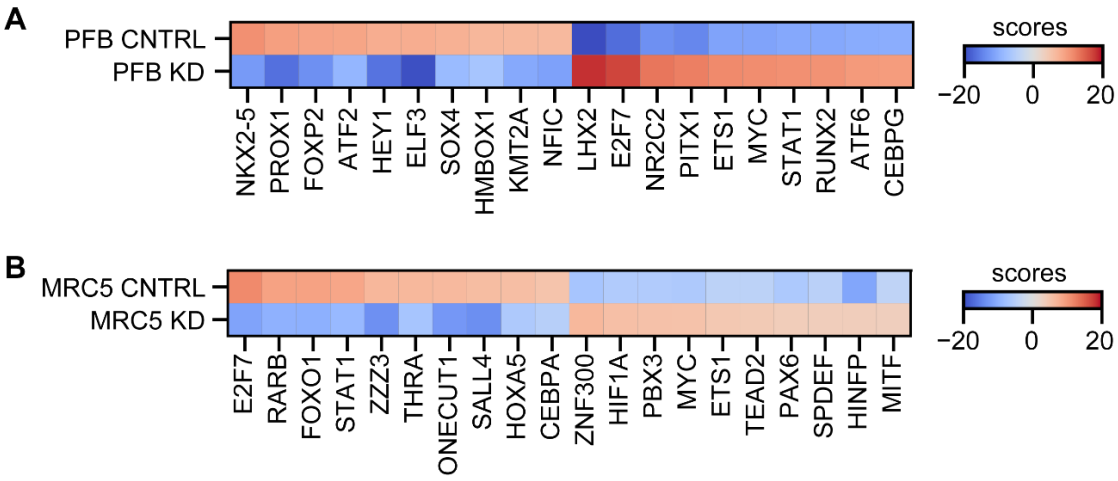


Figure 6.5 Top 10 transcription factor activity inferences in fibroblasts after NR2F2 knockdown

Transcription factors presenting increased (positive scores - red) and reduced activity (negative scores - blue) in the primary fibroblasts isolated from foetal hearts. Cells from control and NR2F2-knockeddown showed contrasting differences regarding the inferred transcriptional activities of the top 10 TFs presented. The scores represent the z-score for the inferred transcriptional activity for each TF in each group. PFB: primary fibroblasts from foetal hearts; CNTRL: control; KD: NR2F2-knockeddown; MRC-5: foetal lung fibroblast cell line.

6.4 Discussion

The use of scRNA-sequencing is usually justified to analyse heterogenous populations. Although the use of this technique could be justified in primary fibroblasts as these could be considered to be heterogeneous, the application of this methodology in a cell line (MRC-5) could be called into question. In this chapter results, the use of scRNA- sequencing in MRC-5 cells was justified by the fact that the fibroblasts could not survive and grow as single cells in culture, which prevented clonal selection. The clonal selection is an important step after CRISPR/Cas9 editing as it makes possible to create a homogeneous cell population to assess the type of mutation (homozygous or heterozygous), separately. This cell culturing limitation of the fibroblasts created a population of cells with different transcriptomes, directing the use of single-cell RNA sequencing in the edited MRC-5 fibroblasts as well.

This chapter of results also shown the molecular variations caused by the reduced presence of NR2F2 in fibroblasts. Previous chapter results showed the mRNA levels of NR2F2 about

50% lower in CRISPR/Cas9 edited samples in comparison with the control, which is not reproduced in the scRNA sequencing data exposed here (Results 6.2, Figure 6.3). This could be explained by limitations presented by global transcriptomic performance, which makes difficult to interpret relevant biology by gene expression analysis alone (217). One hypothesis could be the low expression NR2F2 in fibroblasts that could cause high dropout rates and misrepresentation of the TF expression. Also, the higher amount of mRNA counts in the knockdown group in comparison with the control could be explained by the hypothesis of the knockdown group presenting an alteration in the cell or nuclear membrane which could have been made the cells from this group more susceptible to lysis than the control group, leading to a higher capture rate of mRNA by the beads, thus masking the mRNA reduction. To overcome this issue, Decoupler package was used to predict its transcriptional activity (217). Using this analysis strategy, NR2F2 transcriptional activity was predicted to be reduced in CRISPR/Cas9 altered fibroblast cells in comparison with the control sample, in accordance with the lack of NR2F2 protein previously demonstrated in these cells (Results 5.4).

The cardiac primary fibroblasts samples with the decrease in TF activity had expression relevant cardiogenic genes altered. Interestingly, the cardiac primary fibroblasts presented relevant differentially expressed genes to heart development than the lung fibroblast cell line, which could indicate that, although NR2F2 is widely expressed in various tissues, it could present a tissue-specific gene regulation function during the development.

The upregulated genes were mainly involved in biological processes as cardiomyocyte differentiation and processes related to extracellular matrix organization (Table 6.3). *GATA6* is one of the upregulated genes and it is described to cause CHDs (293). A gain-of-function mutation in *GATA6* was found in a patient with Tetralogy of Fallot, together with another variant present in a VSD patient (293). More recently, ablation of *GATA6* expression in human iPS cells during cardiac differentiation demonstrated significant increase in *NR2F2* expression compared to wild type iPS cells (294). This, in addition to our results, could suggest a regulatory mechanism between those two genes. Notably, *GATA6* variants were also associated with congenital diaphragmatic hernia associated to CHD phenotypes, defects that were also associated with variants in *NR2F2* (116,295).

From the downregulated genes, *HEY1*, *HES1* and *NOG* were enriched. The gene ontologies related to them were associated with endothelial-mesenchymal transition, an important

process in cardiogenesis, and ventricular septum morphogenesis (296). Intriguingly, defects in the septal process are phenotypes associated with NR2F2. In addition, endothelial-mesenchymal transition is the process necessary for the epicardial and endothelial-derived fibroblasts to migrate and differentiate, crucial to the atrioventricular septation (166).

In addition to the gene expression perturbations, the reduction of NR2F2 in the fibroblasts also inferred abnormal activity of other important cardiac TFs (Results 6.3). In foetal cardiac fibroblasts it was observed reduction of activity of *NKX2-5*, which has been described to participate in cardiogenesis (241). In humans, several heart defects have been associated with NKX2-5, such as atrial and ventricular septal defects and Tetralogy of Fallot (74,297,298).

Some TFs that are predicted to have their transcriptional activity reduced (Figure 6.5A) interact with NR2F2, i.e. PROX1 (113), FOXP2 (299) and HEY1 (113). The nature of FOXP2 and NR2F2 interaction is protein-protein binding (299). Although Foxp2 is associated mainly with neurodevelopmental disorders, it has been shown to be expressed in the outflow tract region of developing heart from mice but its function in cardiogenesis requires further investigation (300). Interestingly, FOXP1 contain the FOXP2 conserved residue motif for NR2F2 binding (299), and in contrast to its homologue, FOXP1 has been shown to be downregulated in patients with HLHS (301).

PROX1 is known to participate in vascular cell fate, and it was demonstrated to heterodimerize with NR2F2 for this cell-fate determination (113). Aranguren et al. demonstrated that in venous endothelial cells, NR2F2 homodimer binds to HEY1/2 promoter repressing its expression and, consequently, artery-specific gene expression by Notch signalling. In contrast, lymphatic endothelial cells are determined by the heterodimerization of NR2F2 and PROX1 (111,113). The induction of PROX1 expression in lymphatic cells is dependent of SOX18, but the latter is not sufficient to induce PROX1 in arterial endothelial cells, suggesting that NR2F2 could cooperate with SOX18 to determine the lymphatic cell fate (111). In addition, venous cells, ETS-1 binds to *NR2F2* promoter inducing its transcription (128,302). Although the cell types are different from this work – endothelial cells versus fibroblasts – this evidence can suggest that ETS-1 is upregulated in the absence of NR2F2 due to a compensation mechanism.

Thus, it is conclusive that NR2F2 reduced expression can perturb genes and transcription factor activities related to cardiogenesis in cardiac foetal fibroblasts. The range of genetic perturbation could justify the different CHD phenotypes presented in patients with mutations in NR2F2.

Chapter 7 Discussion

7.1 Summary

Congenital heart malformations are one the most common birth anomalies affecting new-borns across the globe (62). NR2F2 is a transcription factor known to be involved in cardiogenesis, but its contribution to CHDs and the regulatory network behind it are unclear.

The hypothesis of this work was that NR2F2 may interact with other important cardiogenic genes in fibroblasts. To support this hypothesis, firstly, primary fibroblasts were isolated from human foetal hearts with the final aim to knockout *NR2F2* via CRISPR/Cas9 in these cells and evaluate the differences in gene expression in comparison with a negative control cell population. *NR2F2* was also edited by CRISPR/Cas9 in a known foetal lung fibroblast cell line (MRC-5) and gene perturbations were compared.

The primary cardiac fibroblasts were isolated through a protocol based on migration of cells from tissue explants derived from human foetal hearts. An initial assessment of specific proteins and gene expression was carried out to determine the cell types present in this possibly heterogenic population of migrating cells. It was shown that, although other cell types were identified in the migrating cell population, there was an enrichment of fibroblasts in most of the samples. In addition, NR2F2 was detected at mRNA and protein levels in fibroblasts both from foetal hearts and the cell line.

To begin to understand the role of NR2F2 in heart development, preliminary transcriptomic analyses in public datasets were performed. Firstly, a dataset of bulk RNA-sequencing from iPS-CM with *NR2F2* knocked-out by CRISPR/Cas9 presented altered expression in important cardiogenic genes, i.e. *SOX9* (237,238), *BMP4* (239), *NKX2-5* (240,241), and *NPNT* (246). They can be related to septovalvulogenesis (74,238,246,298,303) and outflow tract development (304,305).

Next, we evaluated the efficacy of NR2F2 ablation specifically in fibroblasts. Due to its characteristic of being hard-to-transfect cells, the fibroblasts were successfully edited via CRISPR/Cas9 in the *NR2F2* gene loci using a system that transfects Cas9 ribonucleoprotein together with the guide RNAs by nucleofection, but not by plasmid-based methods. Although the RNP system has a higher cost, its efficiency relies on the lack of the

components' transcription and translation requirements, delivering the CRISPR elements directly to the nucleus of the cell (306).

Finally, transcriptomes of *NR2F2* edited fibroblast samples, together with their respective paired negative controls, were analysed at single-cell resolution. In accordance with previous data, *NKX2-5*, although did not surge as a downregulated gene, was inferred to have a decrease in its transcriptional activity in primary foetal fibroblasts, leading to speculate that NR2F2 could in fact interfere with the regulatory network related to NKX2-5.

7.2 Limitations of studying cardiac fibroblasts

Despite its importance, studying fibroblasts is a challenging task (307). Fibroblasts are a heterogeneous population, leading to not having a specific cell marker to truly identify this cell type (308). In the heart, fibroblast can differ by their phenotype, specific functions, spatial location, and developmental origins (245,307). Because of this, single-cell RNA sequencing has been applied to try unveiling the fibroblast heterogeneity and distinguish them from other cells in several tissues (309).

In addition to the difficulties inherent to the cell type, there are also technical issues to overcome regarding the use of fibroblasts in basic research. In cell culture, beyond the intricacies of isolating fibroblasts due to lack of cell markers, there are some limitations in keeping this cell type in 2D culture (310). It's well known that under standard conditions, the stiffness of the plastic surface from the bottom of the culture plate can trigger activation into myofibroblasts within a few hours (145,310). Fibroblasts demand convoluted conditions to avoid transdifferentiating into myofibroblasts, such as reduced amount of serum in the media and coating plates with gelatin to control stiff covering (311). The protocols used here resorted of plate coating with gelatin to reduce the effects of stiffens of the plastic in the activation of fibroblasts. However, the reduction of serum in the media reduced the rate of growth of the cells. This could justify the presence of genes expression related to myofibroblasts in the tested fibroblasts in this thesis, exposing a limitation of the protocol implemented.

We also presented challenges in deliver CRISPR/Cas9 elements to primary fibroblasts by a plasmid-based system (277). Due to limited life span in culture (312,313), primary fibroblasts are known to be difficult-to-transfect cells. Transfection success is directly

related to the cell cycle phase of the cell. During the S- and M-phase, the nuclear envelope dissolve temporarily, which leads to pDNA translocation to the nucleus (277,314). Also, transgene expression is increased because of the higher global transcriptional activity during DNA synthesis (277). Moreover, gene edition in fibroblasts presented another intricated concern. The limited number of passages that fibroblasts tolerate before senescence also hinder the clonal selection in CRISPR/Cas9 edited fibroblasts. Clonal selection is an important step for CRISPR/Cas9 process for it allows expansion and downstream analysis in a homogeneous population of homozygous mutated cells, and it was an issue experienced in this work. However, this limitation was overcome by the multiplexed single-cell RNA sequencing which increased resolution and sensitivity for the heterogenous population of edited cells.

7.3 The importance of NR2F2 in fibroblasts during cardiogenesis

During the heart development, cells from epicardium undergo through EMT that migrate to the interstitial region of the myocardium (214). The EMT process is responsible to provide fibroblasts to myocardium, allowing its compaction and enlargement of the heart during cardiogenesis (315). This process of EMT is controlled by a fine signalling by several factors including, but not limited to, FGF, PDGF, ETS factors, SOX9, TBX5, and TGFβ (166). Interestingly, rodents with impaired βfgf and Pdgf presented hypoplastic left ventricle (316). ETS-1 is also implicated to the development of HLHS (32,317). Moreover, in this study, we found that lack of NR2F2 activity can upregulate *TGF2* (Results 6.2), increase transcriptional activity of ETS-1 and decrease transcriptional activity of TBX5 (Results 6.3), which indicates that NR2F2 in fibroblasts is necessary to cell migration and the normal heart growth during development. This assumption is corroborated by a recent study in the context of human breast cancer, where *NR2F2* knocked down breast cancer epithelial cell lines showed impaired EMT (318), which confirm the role of NR2F2 in cell migration and mesenchymal invasion.

Another prime source of fibroblasts are the endocardial cells that underwent EndoMT (214). This process is crucial to the formation of valves and the impairment of it is associated with congenital valve defects, such as mitral prolapse valves (319). EndoMT is also responsible for the formation of endocardial cushions, structure that rises the atrioventricular septation (320). Because of this, interruptions on this process during development are also associated with atrioventricular septal defects (321). For the

formation of endocardial cushions, endocardial cells would detach from the basal membrane and migrate to the region between the endocardium and myocardium while differentiate into fibroblasts. The fibroblasts would, then, proliferate and modulate the cushions into valves or extend to begin the septation process (320). Four main pathways are related to EndoMT: TGF, Notch, Wnt and BMP (319). In the present work, the lack of NR2F2 physiological activity shown perturbation in genes directly related to TGFβ and Notch pathways, such as *HES1*, *HEY1*, *NOG*, and *TGFB2* (Results 6.2), which could lead to a conclusion of the role of NR2F2 in this process.

The collective evidence demonstrates that NR2F2 has a crucial function in the balanced signalling for EMT and EndoMT in fibroblasts. Intriguingly, these processes are often related to CHD phenotypes also associated with mutations in NR2F2, like atrioventricular septation defects, mitral valve prolapse, hypoplastic left heart syndrome, and outflow tract defects(319,321) .

7.4 Conclusions and Future Work

In this work it was demonstrated that the decrease of transcriptional activity from NR2F2 in fibroblast can interfere in the fine genetic balance required for a normal heart development.

As a conclusion, NR2F2 could participate in many important pathways for the heart and may perturb other important cardiogenic TFs activities which could led to a wide unbalance in the normal regulatory network in foetal cardiac fibroblasts. Some of the genes altered are involved in chamber septation, valvulogenesis, angiogenesis and heart growth.

However important, this approach is limited to an *in vitro* scenario of one cell-type which give us the global gene expression in which NR2F2 participates, but only allow speculation around the morphological changes that variants in this gene could lead *in vivo,* based on the previous knowledge of each cardiac gene function in normal development. Therefore, an expanded investigation to a cell specific knockout *in vivo* could confirm morphological alterations specific to the role of fibroblasts and extrapolate to other cardiac cellular components in the context of CHDs.

Moreover, using the approach executed in this thesis it was not possible to assess if these gene expression modifications were due to direct binding of NR2F2 to genomic targets or

by indirect regulation. Thus, further investigations should focus in elucidate which are the direct downstream targets using techniques that can identify where the NR2F2 could be direct bond in the genome of foetal cardiac cells, e.g. chromatin immunoprecipitation. The identification of those downstream targets specific for cardiac cell-types could help design genetic therapeutic strategies for CHDs targeting the regulatory network in which NR2F2 is involved.

Appendix A

Single-cell RNA sequencing of migrating primary cells

A.1 Pipelines used in this single-cell RNA analysis

A.1.1 Filtering of low-quality cells using Scanpy

```
# Filter cells according to identified QC thresholds:
sc.pp.filter_cells(adata, min_counts = 500)
sc.pp.filter_cells(adata, max_counts = 25000)
adata = adata[adata.obs['mt_frac'] < 0.1]
sc.pp.filter_cells(adata, min_genes = 500)

#Filter genes not expressed in a minimum number of cells:
sc.pp.filter_genes(adata, min_cells=5)
```

A.1.2 Removal of doublets using Scrublet

Doublets are formed when two or more cells are captured in one GEM during the encapsulation step

```
sce.pp.scrublet(adata,
    adata_sim = None,
    sim_doublet_ratio= 2.0,
    expected_doublet_rate = 0.05,
    stdev_doublet_rate = 0.02,
    synthetic_doublet_umi_subsampling= 1.0,
    knn_dist_metric = 'euclidean',
    normalize_variance= True,
    log_transform= False,
    mean_center= True,
    n_prin_comps= 30,
```

```
    use_approx_neighbors= True,
    get_doublet_neighbor_parents= False,
    n_neighbors = None,
    threshold = None,
    verbose = True,
    copy= False,
random_state= 0,)
adata = adata[adata.obs.predicted_doublet == False]
```

A.1.3 Removal of mitochondrial and ribosomal protein genes

```
ribo_genes = adata.var_names.str.startswith(("RPS","RPL"))
mito_genes = adata.var_names.str.startswith('MT-')
remove = np.add(ribo_genes,mito_genes)
kept_genes = np.invert(remove)
adata = adata[:,kept_genes]
```

A.1.4 Normalisation by Pearson Residuals

```
adata.layers["raw"] = adata.X.copy()
adata.layers["sqrt_norm"] = np.sqrt(
    sc.pp.normalize_total(adata, inplace=False)["X"])
sc.experimental.pp.recipe_pearson_residuals(adata,n_top_genes
=2000,n_comps=100,random_state=42,inplace=True,)
```

A.1.5 Clustering and Visualisation

```
sc.pp.neighbors(adata,n_pcs=50,n_neighbors=10,random_state=42)
sc.tl.umap(adata,random_state=42)
sc.tl.leiden(adata,random_state=42, resolution=0.1)
sc.pl.umap(adata, color='leiden_r0.1')
```

A.1.6 Cell annotation using PanglaoDB from DecoupleR

```
markers = dc.get_resource('PanglaoDB')
markers[(markers['human']=='True')&(markers['canonical_marker
']=='True')]
```

```
markers[~markers.duplicated(['cell_type', 'genesymbol'])]

dc.run_ora(mat=adata,     net=markers,     source='cell_type',
target='genesymbol', min_n=3, verbose=True)

dc.get_acts(adata, obsm_key='ora_estimate')

mean_enr   =   dc.summarize_acts(acts,   groupby='leiden_r0.1',
min_std=2)

annotation_dict = dc.assign_groups(mean_enr)

adata.obs['cell_type'] = [annotation_dict[clust] for clust in
adata.obs['leiden_r0.1']]

sc.pl.umap(adata, color='cell_type')
```

A.1.7 Cardiac cell types' individual gene markers plotting

```
marker_genes_dict = {
    'Cardiomyocytes':   ['MYL3',   'MYL4',   'MYL7',   'MYL9',
'MYH7B','TNNT2','TNNC1','RYR2','TTN','PLN'],
    'Endothelial cells':['PECAM1', 'CDH5'],
    'Fibroblasts':['COL1A1',  'COL1A2',  'COL3A1',  'COL1A2',
'TGFBI','DCN', 'LUM', 'FN1','BGN'],
    'Smooth muscle cells':['TAGLN', 'RGS5', 'ACTA2']}
sc.pl.dotplot(adata,
                  var_names=marker_genes_dict,
                  vmax=1,cmap=mymap2,       groupby='sample',
return_fig=False, dendrogram=False)
```

Appendix B

Transcriptomic analysis of public datasets

B.1 Pipeline used in in public dataset single-cell RNA-sequencing analysis

B.1.1 Filtering of low-quality cells from each foetal heart sample (H1, H2 and H3)

- Heart 1 (H1)

```
# Filter cells according to identified QC thresholds:
print('Total number of cells: {:d}'.format(adata.n_obs))
sc.pp.filter_cells(adata, min_counts = 1000)
sc.pp.filter_cells(adata, max_counts = 30000)
adata = adata[adata.obs['mt_frac'] < 0.1]
sc.pp.filter_cells(adata, min_genes = 200)

#Filter genes not expressed in a minimum number of cells:
sc.pp.filter_genes(adata, min_cells=10)
```

- Heart 2 (H2)

```
# Filter cells according to identified QC thresholds:
sc.pp.filter_cells(adata, min_counts = 600)
sc.pp.filter_cells(adata, max_counts = 40000)
adata = adata[adata.obs['mt_frac'] < 0.1]
sc.pp.filter_cells(adata, min_genes = 250)

#Filter genes not expressed in a minimum number of cells:
sc.pp.filter_genes(adata, min_cells=10)
```

- Heart 3 (H3)

```
# Filter cells according to identified QC thresholds:

sc.pp.filter_cells(adata, min_counts = 500)

sc.pp.filter_cells(adata, max_counts = 20000)

adata = adata[adata.obs['mt_frac'] < 0.1]

sc.pp.filter_cells(adata, min_genes = 500)


#Filter genes not expressed in a minimum number of cells:

sc.pp.filter_genes(adata, min_cells=10)
```

B.1.2 Data Concatenation

```
adata1.obs['sample'] = ['H1']*adata1.n_obs

adata2.obs['sample'] = ['H2']*adata2.n_obs

adata3.obs['sample'] = ['H3']*adata3.n_obs

adata = adata1.concatenate(adata2, adata3, join='inner',
index_unique=None)

adata.var_names_make_unique()
```

B.1.3 Batch effect correction using BBKNN

```
sc.tl.pca(adata, random_state=42)

sc.external.pp.bbknn(adata, pynndescent_random_state=42)

sc.pp.neighbors(adata, n_neighbors=20, n_pcs=20,
random_state=42)

sc.tl.umap(adata, random_state=42)
```

B.1.4 Normalisation using *scran*

```
#Perform a clustering for scran normalization in clusters

adata_pp = adata.copy()

sc.pp.normalize_per_cell(adata_pp, counts_per_cell_after=1e6)

sc.pp.log1p(adata_pp)

sc.pp.pca(adata_pp, n_comps=15)

sc.pp.neighbors(adata_pp)
```

```
sc.tl.leiden(adata_pp, key_added='groups', resolution=0.5)


#Preprocess variables for scran normalization

input_groups = adata_pp.obs['groups']

data_mat = adata.X.T


%%R -i data_mat -i input_groups -o size_factors

size_factors                                                    =
sizeFactors(computeSumFactors(SingleCellExperiment(list(count
s=data_mat)), clusters=input_groups, min.mean=0.1))


#Normalize adata

adata.X = adata.obs['size_factors'].values[:,None]

adata.X = sp.sparse.csr_matrix(adata.X)

sc.pp.log1p(adata)
```

B.1.5 Filtering of highly variable genes and Clustering

```
sc.pp.highly_variable_genes(adata,      flavor='cell_ranger',
n_top_genes=4000)


# Perform clustering - using highly variable genes

sc.tl.leiden(adata, resolution=0.5, key_added='leiden')


#Incresing the resolution in specific clusters

sc.tl.leiden(adata,                          resolution=0.1,
restrict_to=('leiden_increase',
['4']),key_added='leiden_increase')

sc.tl.leiden(adata,                          resolution=0.1,
restrict_to=('leiden_increase',
['5']),key_added='leiden_increase')

sc.tl.leiden(adata,                          resolution=0.1,
restrict_to=('leiden_increase',
['7']),key_added='leiden_increase')
```

```
sc.tl.leiden(adata,                                    resolution=0.1,
restrict_to=('leiden_increase',
['9']),key_added='leiden_increase')


# Excluding low-quality cluster

adata[~adata.obs['annotated'].isin(['14','Erythroblasts']),:]
```

B.2 List of upregulated genes of *NR2F2* knocked out cells (public dataset of bulk RNA-sequencing)

Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC
SOX9	4.89	***MICAL1***	2.41	***RGS5***	1.99	***SERINC2***	1.83	***RNU12***	1.70	***ADM***	1.58
APCDD1	3.98	***ANKRD1***	2.40	***EPDR1***	1.99	***ID4***	1.83	***EFR3B***	1.69	***EVA1B***	1.57
EDN1	3.58	***TUBB3***	2.38	***LZTS1***	1.99	***PTX3***	1.83	***PCDH10***	1.68	***KCNK3***	1.57
DUSP6	3.57	***TGFB2***	2.35	***TAGLN***	1.98	***CDH8***	1.83	***NQO1***	1.67	***PIR***	1.56
ACTA1	3.43	***TUBB2B***	2.30	***ACTG2***	1.97	***MSI1***	1.82	***LIMD2***	1.66	***CRMP1***	1.56
CTGF	3.21	***EMID2***	2.29	***SERPINE1***	1.97	***FREM2***	1.80	***STK17A***	1.65	***FBLN1***	1.55
NKX2-6	3.16	***NNAT***	2.21	***ROBO1***	1.94	***REEP2***	1.79	***CNTFR***	1.65	***MDFI***	1.54
KIF1A	2.91	***KIAA1199***	2.20	***ATP1A2***	1.93	***PCYT1B***	1.78	***COTL1***	1.63	***ENO3***	1.54
DCLK1	2.81	***SLC7A11***	2.12	***ADAMTS19***	1.93	***SPHK1***	1.77	***MYOZ2***	1.63	***CD99***	1.53
CCL2	2.75	***LEFTY2***	2.08	***TNFRSF19***	1.88	***THBS1***	1.77	***BMP4***	1.62	***AMOT***	1.53

Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC
TENM2	2.72	***WBSCR17***	2.08	***GPC2***	1.88	***CLSTN2***	1.76	***GLIPR1***	1.61	***EFCAB4A***	1.53
ITGA9	2.66	***DEPDC1***	2.05	***MYC***	1.87	***C4orf48***	1.75	***SLFN5***	1.61	***ETV5***	1.53
TM4SF1	2.62	***ENC1***	2.05	***ADD2***	1.86	***LIMA1***	1.75	***UCHL1***	1.61	***TMEM88***	1.52
A2M	2.50	***IGFBP7***	2.05	***DBNDD1***	1.86	***PRSS23***	1.75	***STC2***	1.60	***MAP6***	1.51
FN1	2.49	***DPYSL3***	2.04	***SERPINA3***	1.86	***CPA4***	1.74	***NEDD9***	1.60	***PPM1H***	1.51
LEFTY2	2.48	***XRCC4***	2.03	***SULF1***	1.86	***NNAT***	1.74	***SEMA6B***	1.60	***CPE***	1.51
CDO1	2.47	***ODAM***	2.02	***MAP1A***	1.85	***CTNND2***	1.72	***COL2A1***	1.59	***RP11-488C13.7***	1.51
LIX1	2.43	***ZFHX2***	2.02	***COL2A1***	1.85	***ACTA2***	1.70	***ID1***	1.59	***DTL***	1.50
MGST1	2.43	***SLC22A23***	2.01	***APOE***	1.84	***CYR61***	1.70	***ANXA1***	1.59	***TLE1***	1.50

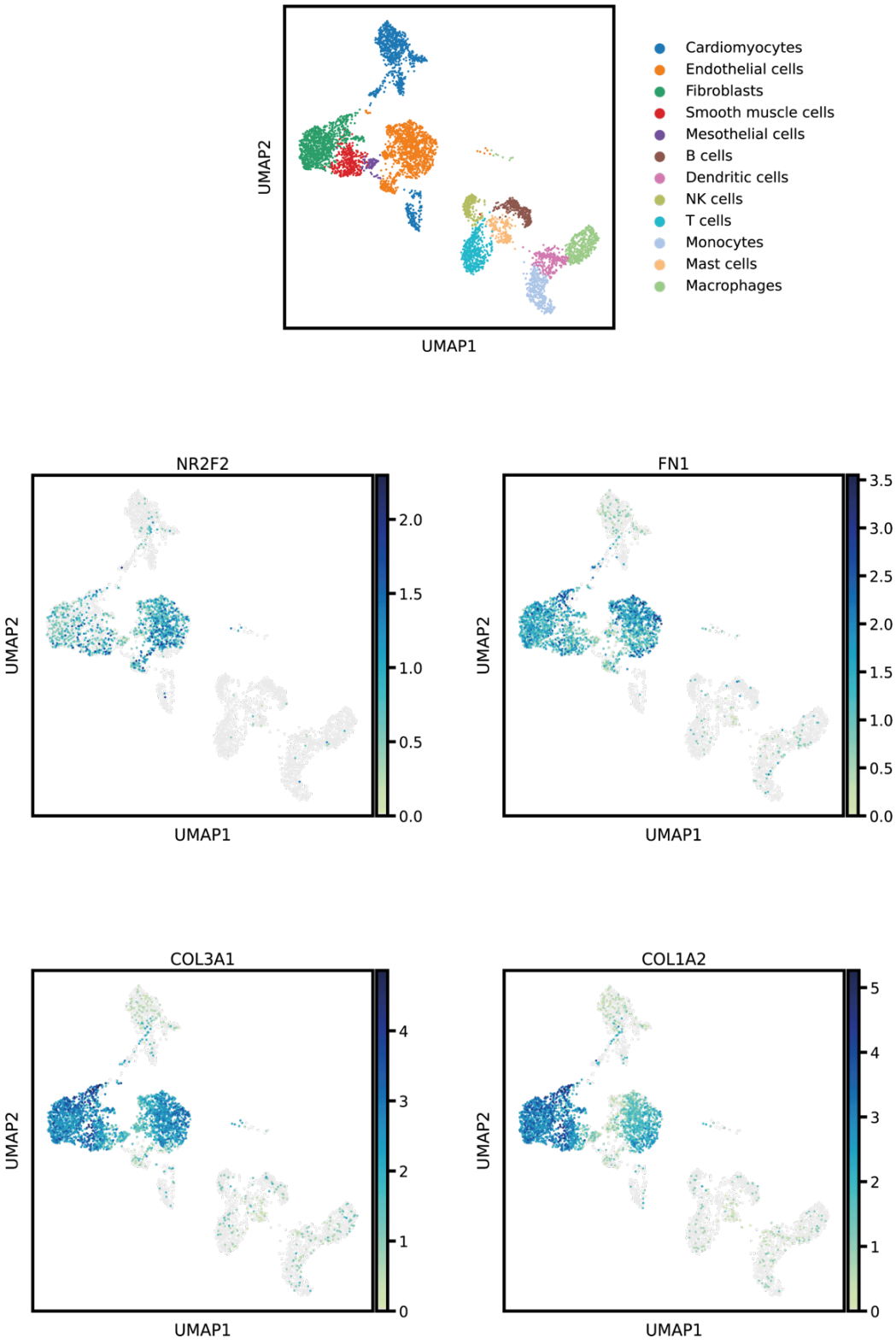
LogFC = Log10 fold-change

B.3 List of downregulated genes of *NR2F2* knocked out cells (public dataset of bulk RNA-sequencing)

Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC
DUSP5	-1.50	***MALAT1***	-1.57	***SEPP1***	-1.74	***FRZB***	-1.87	***TNNT1***	-2.33
ALPK3	-1.50	***SYDE2***	-1.58	***ARHGEF3***	-1.75	***SULT1E1***	-1.91	***A2M***	-2.40
PODXL	-1.52	***SNORA14B***	-1.59	***MAB21L2***	-1.75	***FMOD***	-1.91	***PLAT***	-2.72
TRIM63	-1.53	***ZNF330***	-1.60	***LINC00881***	-1.75	***NKX2-5***	-2.01	***MIR4761***	-2.78
MYOZ2	-1.54	***SFRP2***	-1.63	***TGFBR3***	-1.78	***G0S2***	-2.01	***HPCAL4***	-2.85
ATP7B	-1.55	***LRRC10***	-1.67	***SLC40A1***	-1.81	***DES***	-2.05	***DLK1***	-3.16
NPNT	-1.55	***BMP5***	-1.68	***CRABP1***	-1.83	***NXPH2***	-2.11	***MYL2***	-3.25
SNORA28	-1.55	***LMOD2***	-1.71	***FXYD1***	-1.85	***SLC39A8***	-2.14	***PARM1***	-3.51
NEXN	-1.56	***SMYD1***	-1.72	***DHRS3***	-1.85	***CSRP3***	-2.15	***BANCR***	-4.27

LogFC = Log10 fold-change

B.4 Presence of cells going through endothelial-mesenchymal transition in foetal heart (public dataset)



UMAP plots from single-cell RNA sequencing of three foetal human hearts (public dataset). In the first row, the clusters show the annotation and position of each cell type identified in the dataset. The four UMAPs below demonstrate the expression of *NR2F2, FN1, COL3A1* and *COL1A2* in the dataset. Each cell is colorised according to the level of expression.

Appendix C

Single-cell RNA sequencing of NR2F2 knocked-down fibroblasts

C.1 Pipelines used in this single-cell RNA analysis

C.1.1 Filtering of low-quality cells using Scanpy

```
# Filter cells according to identified QC thresholds:
sc.pp.filter_cells(adata, min_counts = 500)
sc.pp.filter_cells(adata, max_counts = 50000)
adata = adata[adata.obs['mt_frac'] < 0.1]
sc.pp.filter_cells(adata, min_genes = 2000)

#Filter genes not expressed in a minimum number of cells:
sc.pp.filter_genes(adata, min_cells=5)
```

C.1.2 Removal of doublets using Scrublet

Doublets are formed when two or more cells are captured in one GEM during the encapsulation step

```
sce.pp.scrublet(adata,
    adata_sim = None,
    sim_doublet_ratio= 2.0,
    expected_doublet_rate = 0.05,
    stdev_doublet_rate = 0.02,
    synthetic_doublet_umi_subsampling= 1.0,
    knn_dist_metric = 'euclidean',
    normalize_variance= True,
    log_transform= False,
    mean_center= True,
    n_prin_comps= 30,
```

```
    use_approx_neighbors= True,
    get_doublet_neighbor_parents= False,
    n_neighbors = None,
    threshold = None,
    verbose = True,
    copy= False,
random_state= 0,)
adata = adata[adata.obs.predicted_doublet == False]
```

C.1.3 Removal of mitochondrial and ribosomal protein genes

```
ribo_genes = adata.var_names.str.startswith(("RPS","RPL"))
mito_genes = adata.var_names.str.startswith('MT-')
remove = np.add(ribo_genes,mito_genes)
kept_genes = np.invert(remove)
adata = adata[:,kept_genes]
```

C.1.4 Normalisation by Pearson Residuals

```
adata.layers["raw"] = adata.X.copy()
adata.layers["sqrt_norm"] = np.sqrt(
    sc.pp.normalize_total(adata, inplace=False)["X"])
sc.experimental.pp.recipe_pearson_residuals(adata,n_top_genes
=2000,n_comps=100,random_state=42,inplace=True,)
```

C.1.5 Clustering and Visualisation

```
sc.pp.neighbors(adata,n_pcs=50,n_neighbors=10,random_state=42)
sc.tl.umap(adata,random_state=42)
sc.tl.leiden(adata,random_state=42, resolution=0.2)
sc.pl.umap(adata, color='leiden_r0.2')
```

C.1.6 Cell annotation using PanglaoDB from DecoupleR

```
markers = dc.get_resource('PanglaoDB')
markers[(markers['human']=='True')&(markers['canonical_marker
']=='True')]
```

```
markers[~markers.duplicated(['cell_type', 'genesymbol'])]

dc.run_ora(mat=adata,     net=markers,     source='cell_type',
target='genesymbol', min_n=3, verbose=True)

dc.get_acts(adata, obsm_key='ora_estimate')

mean_enr   =   dc.summarize_acts(acts,   groupby='leiden_r0.2',
min_std=0.2)

annotation_dict = dc.assign_groups(mean_enr)

adata.obs['cell_type'] = [annotation_dict[clust] for clust in
adata.obs['leiden_r0.2']]

sc.pl.umap(adata, color='cell_type')
```

C.1.7 Expression checking of known cardiac cell type gene markers

```
markers=['MYL3','MYL4','MYL9','MYH7B','RYR2','TTN','PLN','PEC
AM1','CDH5','COL3A1','COL1A2','COL3A1','TGFBI','FN1','DCN','L
UM','BGN','TAGLN','RGS5','ACTA2']

sc.pl.umap(adata,     color=markers,color_map=mymap2,ncols=4,
vmax=20,    legend_loc='on    data',    legend_fontoutline=3,
legend_fontweight='normal',use_raw=False,layer="sqrt_norm")

marker_genes_dict = {


'Cardiomyocytes:['MYL3','MYL4','MYL9','MYH7B','RYR2','TTN','P
LN'],

    'Endothelial cells: ['PECAM1', 'CDH5'],

    'Fibroblasts':
['COL3A1','COL1A2','COL3A1','TGFBI','FN1','DCN','LUM','BGN'],

    'Smooth muscle cells: ['TAGLN', 'RGS5', 'ACTA2']}

sc.pl.dotplot(adata,   var_names=marker_genes_dict,   vmin=0,
vmax=4, cmap=mymap2, groupby='sample', figsize= (8,1))
```

C.1.8 Pseudo-bulk DEG analysis using DecoupleR

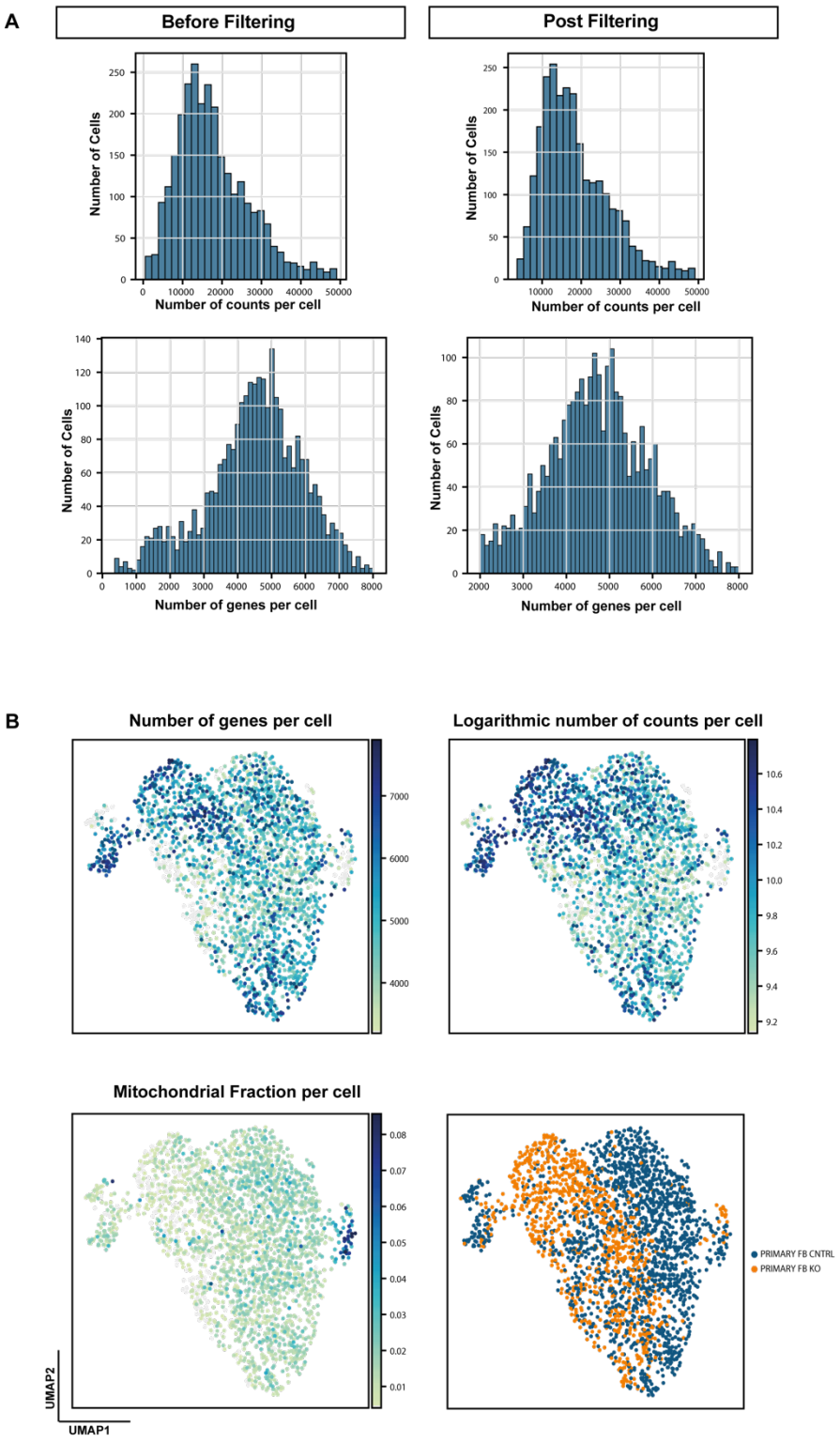
```
sc.pp.log1p(adata)

logFCs, pvals = dc.get_contrast (padata,

                                 group_col='cell_type',

                                 condition_col='sample',

                                 condition='KO',

                                 reference='CNTRL',
```

```
                                    method='wilcoxon'

                                   )

dc.format_contrast_results (logFCs, pvals)
```

C.1.9 TF activity inference using Dorothea from DecoupleR

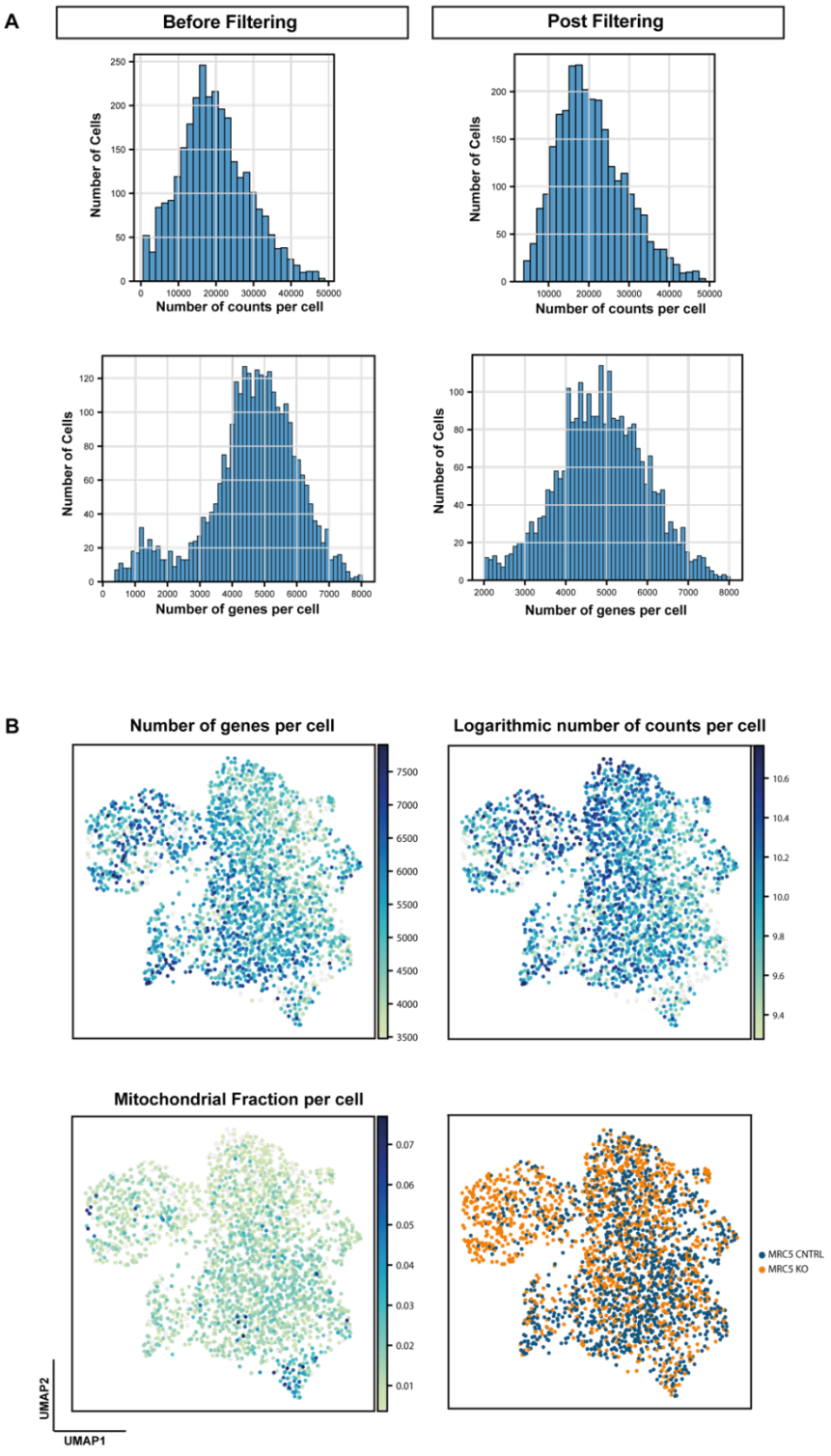
```
dc.get_dorothea (organism='human', levels=['A','B','C','D'])

dc.run_mlm      (mat=adata,      net=net,      source='source',
target='target', weight='weight', verbose=True)

adata.obsm['dorothea_mlm_estimate']                          =
adata.obsm['mlm_estimate'].copy()

adata.               obsm['dorothea_mlm_pvals']              =
adata.obsm['mlm_pvals'].copy()

dc.get_acts (adata, obsm_key='dorothea_mlm_estimate')

sc.pp.log1p(acts)

sc.tl.rank_genes_groups(acts,                groupby='sample',
key_added='dorothea_rankgenes', method='wilcoxon')
```

C.2 Quality control of CRISPR/Cas9 edited primary fibroblasts



(A) Bar charts showing the number of cells (y axis) containing a number of counts/genes (x axis). The left-hand side shows the bar charts before the application of filtering method present in the pipeline and the right-hand side shows the result post-filtering. (B) UMAP plots showing the number of counts, the logarithmic number of counts and the mitochondrial fraction in each cell (dots). It also shows a UMAP with the location of each cell that belongs to control (non-edited) fibroblasts and edited fibroblasts populations. PRIMARY CNTRL: control (non-edited) cardiac foetal fibroblasts; PRIMARY KO: edited cardiac foetal fibroblasts.

C.3 Quality control of CRISPR/Cas9 edited MRC-5 cells

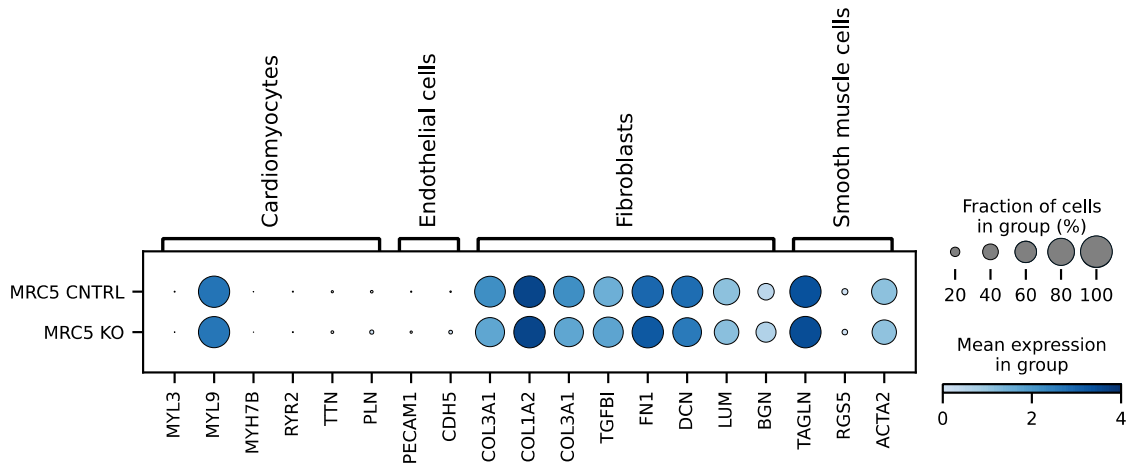


(A) Bar charts showing the number of cells (y axis) containing a number of counts/genes (x axis). The left-hand side shows the bar charts before the application of filtering method present in the pipeline and the right-hand side shows the result post-filtering. (B) UMAP plots showing the number of counts, the logarithmic number of counts and the mitochondrial fraction in each cell (dots). It also shows a UMAP with the location of each cell that belongs to control (non-edited) fibroblasts and edited fibroblasts populations. MRC-5 CNTRL: control (non-edited) foetal lung fibroblasts cell line; MRC-5 KO: edited foetal lung fibroblasts cell line.

C.4 Cell annotation using PanglaoDB in CRISPR/Cas9 edited MRC-5 fibroblasts cell line

UMAP plots of MRC-5 samples – CNTLR and KO – distribution and PanglaoDB annotation (first row). Bellow, UMAP plots showing each cell (dots) coloured according to the Fisher's exact test score (-log10[*p*-value]) for fibroblasts, endothelial cells, cardiomyocytes, and smooth muscle cells. CNTRL: control; KO: CRISPR/Cas9 edited cells.

C.5 Individual expression of cardiac cell type gene markers in CRISPR/Cas9 MRC-5 fibroblast cell line



Dot plot showing the expression of gene markers to identify cardiomyocytes, endothelial cells, fibroblasts, and smooth muscle cells in CRISPR/Cas9 edited MRC-5 cells. Each dot represents the percentage of cells (size of the dot) expressing the gene (x axis) and the mean expression (intensity of the colour) within each cluster (y axis).

C.6 List of differentially expressed genes of fibroblasts with NR2F2 edited by CRISPR/Cas9

C.6.1 List of upregulated genes from primary fibroblasts from foetal heart

Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC
LINC00968	3.90	**ANKRD1**	2.41	**LINC01638**	1.95	**POSTN**	1.66
AC100822.1	3.35	**CLDN11**	2.41	**NUPR1**	1.93	**LMOD1**	1.66
ITGBL1	3.35	**CYP1B1**	2.37	**KCNQ5**	1.93	**ADAMTS1**	1.64
IGFBP3	3.07	**C1orf143**	2.35	**PCDH10**	1.92	**EDIL3**	1.63
CPA4	3.05	**ROR1**	2.34	**MYPN**	1.92	**PKD1L1**	1.63
TMEM178B	3.05	**LOX**	2.34	**KCTD4**	1.90	**LINC00607**	1.63
C5orf17	3.05	**TGFB2**	2.33	**CCDC80**	1.89	**SLIT3**	1.62
PAPPA2	2.96	**KCNE4**	2.32	**FAM155A**	1.89	**GALNT5**	1.62
DOK5	2.94	**MFAP5**	2.31	**PCDH7**	1.88	**AC113383.1**	1.62
CDH10	2.89	**C7orf69**	2.30	**NALCN**	1.87	**AC002460.2**	1.61
TMEM233	2.82	**HAPLN1**	2.29	**LIMCH1**	1.86	**DOK6**	1.60
KRTAP2-3	2.82	**FGF14**	2.28	**KCNT2**	1.86	**KIAA1324**	1.60
AC010624.5	2.78	**SH3BGR**	2.25	**NLGN1**	1.84	**GPNMB**	1.59
GATA6	2.75	**SLC38A5**	2.24	**CDKL4**	1.84	**PSG5**	1.57
AL162493.1	2.74	**ACTA2**	2.19	**ZFYVE28**	1.83	**VSIR**	1.55
LINC01133	2.73	**AC026316.5**	2.16	**HSPB6**	1.82	**GHR**	1.55
PTX3	2.72	**CPE**	2.15	**RGS4**	1.82	**CCDC144A**	1.55
PLD5	2.70	**COL8A1**	2.14	**MAPRE3**	1.81	**LRMDA**	1.54
ABI3BP	2.67	**LINC00973**	2.14	**LINC00960**	1.81	**MGARP**	1.54
CPXM2	2.66	**OCIAD2**	2.12	**CLDN1**	1.80	**OXTR**	1.53
DEPTOR	2.66	**GPR1**	2.11	**ZNF804A**	1.79	**DSP**	1.53

Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC
NEFM	2.64	**CDK15**	2.11	**NR2F1**	1.76	**SELPLG**	1.53
AL591043.2	2.64	**KCTD8**	2.08	**SDSL**	1.76	**AF165147.1**	1.52
FOXP2	2.63	**PSG4**	2.03	**PLAC9**	1.75	**LRRC4C**	1.52
NTN4	2.59	**PPP4R4**	2.03	**ADAMTSL1**	1.74	**STAT4**	1.51
ARRDC4	2.57	**SLC7A14**	2.03	**SERPINB2**	1.74	**C5orf46**	1.50
NGF-AS1	2.55	**STAC**	2.02	**AC073114.1**	1.72	**AC110058.1**	1.50
NGF	2.55	**LINC00452**	2.01	**SCN9A**	1.71	**AC005736.1**	1.50
CNTNAP3B	2.52	**ANK2**	2.01	**CRIP1**	1.71	**NFIB**	1.49
AC069277.1	2.48	**FOXC2-AS1**	2.00	**GRIK2**	1.71	**LINC02755**	1.49
HMCN1	2.47	**AMPH**	2.00	**NEDD4L**	1.71	**PPP1R3C**	1.49
KCNK6	2.44	**RSPO3**	1.99	**NR2F1-AS1**	1.70	**F2RL2**	1.48
STXBP6	2.43	**LINC01340**	1.98	**PSAT1**	1.70	**BST1**	1.48
LINC00882	2.43	**ANGPT1**	1.96	**PLXDC2**	1.68	**AC239799.2**	1.48
CCDC81	2.43	**LINC02615**	1.95	**KCNK2**	1.68	**KIF26B**	1.47
TPD52L1	2.42	**EPDR1**	1.95	**IFI6**	1.67	**ICAM1**	1.47
FGD4	2.42	**LRRN3**	1.95	**HIST1H4H**	1.66	**PRICKLE1**	1.46

LogFC = Log10 fold-change

C.6.2 List of downregulated genes from primary fibroblasts from foetal heart

Gene Symbol	logFC	Gene Symbol	logFC	Gene Symbol	logFC
DCLK3	-3.90	**RNF144A**	-2.33	**SOCS2**	-2.05
AP002989.1	-3.61	**CSF3**	-2.29	**SLITRK4**	-2.04
NOG	-3.40	**NOX4**	-2.22	**TMEM132B**	-2.00
LAMA4	-2.73	**MGAT4C**	-2.21	**TCIM**	-1.92
FAT3	-2.64	**CALB2**	-2.15	**MMP1**	-1.89
TEX26-AS1	-2.54	**AC079949.3**	-2.13	**ROBO4**	-1.87
PCDH1	-2.50	**STC1**	-2.10	**SPRY1**	-1.87
OLFM2	-2.40	**MMP3**	-2.06	**ST8SIA4**	-1.86
EFR3B	-1.84	**DENND11**	-1.63	**MEDAG**	-1.54
HEY1	-1.83	**COL5A3**	-1.61	**CALCRL**	-1.54
NFATC2	-1.75	**SORBS2**	-1.61	**EDA**	-1.52
TFPI2	-1.74	**HES1**	-1.59	**TNNT1**	-1.51
AC068234.2	-1.67	**HS6ST3**	-1.58	**BDKRB2**	-1.51
LINC01411	-1.66	**ADAMTS18**	-1.58	**PDE3A**	-1.49
BTBD11	-1.66	**CSF2**	-1.58	**ID1**	-1.49
TNFRSF6B	-1.64	**GNGT1**	-1.56	**RGCC**	-1.48
AC004918.1	-1.64	**ANGPTL4**	-1.55	**RBP1**	-1.48
DENND2A	-1.64	**IGFBP5**	-1.55	**CFAP58**	-1.46

LogFC = Log10 fold-change

C.6.3 List of differently expressed genes from MRC-5 cells

Gene Symbol	logFC	Gene Symbol	logFC
THRB	-1.55		
TBX20	2.38	**MYCT1**	1.71
LINC02008	2.35	**TMEM132D**	1.68
HAND2	2.29	**EMB**	1.67
NCAM1	2.26	**AL157392.2**	1.66
ESM1	2.22	**FOXO1**	1.65
FOXG1	2.20	**CXCL3**	1.64
SHOX2	2.14	**ANK3**	1.60
SNTG2	2.11	**GATA4**	1.60
NTNG1	2.08	**CXCL5**	1.54
EPHA5	1.97	**CXCL1**	1.54
NCALD	1.97	**SLC1A3**	1.52
NFIB	1.96	**CXCL6**	1.52
CACNG8	1.91	**BCL11A**	1.51
AP000695.2	1.90	**IL1A**	1.50
PLD5	1.84	**COLEC12**	1.50
CCDC85A	1.80	**EBF3**	1.48
DSC3	1.77	**HHIP**	1.47
SIM2	1.73	**SULF1**	1.46

LogFC = Log10 fold-change

C.7 Gene Ontology of Biological Processes enriched in the upregulated gene set of MRC-5 cells

Enrichment FDR	Fold Enrichment	Pathway	Genes
0.00024096	100.57	Cardiac right ventricle morphogenesis	*HAND2 TBX20 GATA4*
2.10E-05	12.31	Axonogenesis	*BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 SHOX2 EMB FOXG1*
2.88E-05	11.32	Axon development	*BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 SHOX2 EMB FOXG1*
2.10E-05	10.81	Cell morphogenesis involved in neuron differentiation	*SLC1A3 BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 SHOX2 EMB FOXG1*
6.54E-05	8.80	Neuron projection morphogenesis	*BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 SHOX2 EMB FOXG1*
6.54E-05	8.63	Chemotaxis	*CXCL6 EPHA5 NFIB NCAM1 CXCL3 CXCL5 CXCL1 EMB FOXG1*
6.54E-05	8.62	Plasma membrane bounded cell projection morphogenesis	*BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 SHOX2 EMB FOXG1*
6.54E-05	8.60	Taxis	*CXCL6 EPHA5 NFIB NCAM1 CXCL3 CXCL5 CXCL1 EMB FOXG1*
6.54E-05	8.57	Cell projection morphogenesis	*BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 SHOX2 EMB FOXG1*

Enrichment FDR	Fold Enrichment	Pathway	Genes
6.56E-05	8.37	Cell part morphogenesis	*BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 SHOX2 EMB FOXG1*
6.54E-05	7.57	Tube morphogenesis	*IL1A PODXL SULF1 NFIB HAND2 HHIP ESM1 TBX20 SHOX2 GATA4*
0.00016891	7.40	Cellular component morphogenesis	*BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 SHOX2 EMB FOXG1*
6.54E-05	6.74	Tube development	*IL1A PODXL SULF1 NFIB SIM2 HAND2 HHIP ESM1 TBX20 SHOX2 GATA4*
0.00019697	6.18	Cell morphogenesis	*SLC1A3 BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 SHOX2 EMB FOXG1*
6.56E-05	6.13	Neuron development	*SLC1A3 BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 HAND2 SHOX2 EMB FOXG1*
6.56E-05	5.46	Neuron differentiation	*SLC1A3 BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 HAND2 TBX20 SHOX2 EMB FOXG1*
6.54E-05	5.37	Generation of neurons	*SLC1A3 BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 HAND2 HHIP TBX20 SHOX2 EMB FOXG1*

Enrichment FDR	Fold Enrichment	Pathway	Genes
6.54E-05	4.96	Neurogenesis	*SLC1A3 BCL11A EPHA5 NFIB NCAM1 ANK3 NTNG1 HAND2 HHIP TBX20 SHOX2 EMB FOXG1*
6.54E-05	4.74	Locomotion	*CXCL6 PODXL SULF1 EPHA5 NFIB NCAM1 NTNG1 CXCL3 CXCL5 CXCL1 HAND2 TBX20 EMB FOXG1*

FDR: False discovery rate. FDR is calculated based on the nominal *p*-value, which states the likelihood of enrichment by chance. The fold enrichment is the percentage of genes in the list of differentially expressed genes belongs to the biological process, indicating the overrepresentation. The biological processes are ranked by Fold Enrichment.

List of References

1. Wu W, He J, Shao X. Incidence and mortality trend of congenital heart disease at the global, regional, and national level, 1990–2017. Medicine [Internet]. 2020 [cited 2022 May 10];99(23). Available from: /pmc/articles/PMC7306355/

2. Gillian M Blue, Edwin P Kirk, Gary F Sholler, Richard P Harvey, David S Winlaw. Congenital heart disease: current knowledge about causes and inheritance. Med J Aust [Internet]. 2012 Aug 6 [cited 2022 May 10];197(3):155–9. Available from: www.mja.com.au

3. Triedman JK, Newburger JW. Trends in congenital heart disease. Circulation. 2016;133(25):2716–33.

4. Diab NS, Barish S, Dong W, Zhao S, Allington G, Kahle KT, et al. Molecular Genetics and Complex Inheritance of Congenital Heart Disease. Genes (Basel). 2021;12(7):1020.

5. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, et al. Heart Disease and Stroke Statistics - 2014 Update: A report from the American Heart Association. Vol. 129, Circulation. 2014.

6. Zaidi S, Brueckner M. Genetics and Genomics of Congenital Heart Disease. Circ Res [Internet]. 2017 [cited 2022 Jun 15];120(6):923–40. Available from: http://circres.ahajournals.org

7. Gillian M Blue, Edwin P Kirk, Gary F Sholler, Richard P Harvey, David S Winlaw. Congenital heart disease: current knowledge about causes and inheritance. MJA [Internet]. 2012 Aug 6 [cited 2020 Feb 24];197(3):155–9. Available from: www.mja.com.au

8. Boyd R, Mcmullen H, Beqaj H, Kalfa D. Environmental Exposures and Congenital Heart Disease. 2022 [cited 2022 May 10]; Available from: www.pediatrics.org/cgi/10.1542/peds.2021-052151

9. Correa A, Gilboa SM, Besser LM, Botto LD, Moore CA, Hobbs CA, et al. Diabetes mellitus and birth defects. Am J Obstet Gynecol [Internet]. 2008 [cited 2022 May 10];199(3):237.e1-237.e9. Available from: www.AJOG.org

10. Madsen NL, Schwartz SM, Lewin MB, Mueller BA. Prepregnancy Body Mass Index and Congenital Heart Defects among Offspring: A Population-based Study. Congenit Heart Dis. 2013;8(2):131–41.

11. Cai GJ, Sun XX, Zhang L, Hong Q. Association between maternal body mass index and congenital heart defects in offspring: A systematic review. Am J Obstet Gynecol [Internet]. 2014 [cited 2022 May 10];211(2):91–117. Available from: http://dx.doi.org/10.1016/j.ajog.2014.03.028

12. Levy HL, Guldberg P, Güttler F, Hanley WB, Matalon R, Rouse BM, et al. Congenital Heart Disease in Maternal Phenylketonuria: Report from the Maternal PKU Collaborative Study. Pediatric Research 2001 49:5 [Internet]. 2001 [cited 2022 Jul 15];49(5):636–42. Available from: https://www.nature.com/articles/pr2001100

13. Kopf PG, Walker MK. Overview of developmental heart defects by dioxins, PCBs, and pesticides. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 2009;27(4):276–85.

14. Bouma BJ, Mulder BJM. Changing Landscape of Congenital Heart Disease. Circ Res [Internet]. 2017;120(6):908–22. Available from: http://circres.ahajournals.org/lookup/doi/10.1161/CIRCRESAHA.116.309302

15. Pierpont ME, Brueckner M, Chung WK, Garg V, Lacro R v., McGuire AL, et al. Genetic Basis for Congenital Heart Disease: Revisited: A Scientific Statement From the American Heart Association. Circulation [Internet]. 2018 Nov 20 [cited 2022 May 10];138(21):e653–711. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000000606

16. Chatfield KC, Deardorff MA, Chatfield KC, Deardorff MA. Chromosomal Anomalies Associated with Congenital Heart Disease. In: Cruz EM da, Ivy D, Jaggers J, editors. Pediatric and Congenital Cardiology, Cardiac Surgery and Intensive Care [Internet].

1st editio. Springer, London; 2014 [cited 2022 Jul 16]. p. 47–71. Available from: https://link.springer.com/referenceworkentry/10.1007/978-1-4471-4619-3_93

17. Ferencz C, Boughman JA, Neill CA, Brenner JI, Perry LW. Congenital cardiovascular malformations: Questions on inheritance. J Am Coll Cardiol. 1989 Sep 1;14(3):756–63.

18. Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S. Chromosome 21 and Down syndrome: From genomics to pathophysiology. Nat Rev Genet. 2004;5(10):725–38.

19. Bondy CA. Turner syndrome. Horm Res. 2009;71(SUPPL. 1):52–6.

20. Sallie B. Freeman, Lisa F. Taft, Kenneth J. Dooley, Katherine Allran, Stephanie L. Sherman, Terry J. Hassold, et al. Population-based study of congenital heart defects in Down syndrome. Am J Med Genet [Internet]. 1998 [cited 2022 Jun 13];80(3):213–7. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1002/%28SICI%291096-8628%2819981116%2980%3A3%3C213%3A%3AAID-AJMG6%3E3.0.CO%3B2-8?saml_referrer

21. Freeman SB, Bean LH, Allen EG, Tinker SW, Locke AE, Druschel C, et al. Ethnicity, sex, and the incidence of congenital heart defects: A report from the National Down Syndrome Project. Genetics in Medicine. 2008 Mar;10(3):173–80.

22. Bull MJ. Down Syndrome. Ropper AH, editor. https://doi.org/101056/NEJMra1706537 [Internet]. 2020 Jun 10 [cited 2022 Jun 13];382(24):2344–52. Available from: https://www.nejm.org/doi/10.1056/NEJMra1706537

23. Polli JB, de P. Groff D, Petry P, Mattos VF, Rosa RCM, Zen PRG, et al. Trisomy 13 (Patau syndrome) and congenital heart defects. Am J Med Genet A. 2014 Jan;164(1):272–5.

24. Springett A, Wellesley D, Greenlees R, Loane M, Addor MC, Arriola L, et al. Congenital anomalies associated with trisomy 18 or trisomy 13: A registry-based study in 16 european countries, 2000-2011. Am J Med Genet A. 2015 Dec 1;167(12):3062–9.

25. Bondy CA. Congenital Cardiovascular Disease in Turner Syndrome. Congenit Heart Dis [Internet]. 2008 Jan 1 [cited 2022 Jun 14];3(1):2–15. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1747-0803.2007.00163.x

26. Freedom RM, Gerald PS. Congenital Cardiac Disease and the Cat Eye Syndrome. American Journal of Diseases of Children [Internet]. 1973 Jul 1 [cited 2022 Jul 19];126(1):16–8. Available from: https://jamanetwork.com/journals/jamapediatrics/fullarticle/505064

27. Wilkens A, Liu H, Park K, Campbell LB, Jackson M, Kostanecka A, et al. Novel clinical manifestations in Pallister–Killian syndrome: Comprehensive evaluation of 59 affected individuals and review of previously reported cases. Am J Med Genet A [Internet]. 2012 Dec 1 [cited 2022 Jul 19];158A(12):3002–17. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/ajmg.a.35722

28. Digilio M, Marino B, Capolino R, Dallapiccola B. Clinical manifestations of Deletion 22q11.2 syndrome (DiGeorge/Velo-Cardio-Facial syndrome). Images Paediatr Cardiol [Internet]. 2005 Apr [cited 2022 Jun 14];7(2):23. Available from: /pmc/articles/PMC3232571/

29. Merscher S, Funke B, Epstein JA, Heyer J, Puech A, Lu MM, et al. TBX1 Is Responsible for Cardiovascular Defects in Velo-Cardio-Facial/DiGeorge Syndrome. Cell. 2001 Feb 23;104(4):619–29.

30. Marino B, Digilio MC, Toscano A, Anaclerio S, Giannotti A, Feltri C, et al. Anatomic patterns of conotruncal defects associated with deletion 22q11. Genetics in Medicine. 2001;3(1):45–8.

31. Pober BR. Williams–Beuren Syndrome. https://doi.org/101056/NEJMra0903074 [Internet]. 2010 Jan 21 [cited 2022 Jun 14];362(3):239–52. Available from: https://www.nejm.org/doi/10.1056/NEJMra0903074

32. Ye M, Coldren C, Liang X, Mattina T, Goldmuntz E, Benson DW, et al. Deletion of ETS-1, a gene in the Jacobsen syndrome critical region, causes ventricular septal defects and abnormal ventricular morphology in mice. Hum Mol Genet [Internet]. 2010 Feb

15 [cited 2022 Jun 14];19(4):648–56. Available from: https://academic.oup.com/hmg/article/19/4/648/609309

33. Favier R, Akshoomoff N, Mattson S, Grossfeld P. Jacobsen syndrome: Advances in our knowledge of phenotype and genotype. Am J Med Genet C Semin Med Genet. 2015 Sep 1;169(3):239–50.

34. Digilio MC, Bernardini L, Consoli F, Lepri FR, Giuffrida MG, Baban A, et al. Congenital heart defects in recurrent reciprocal 1q21.1 deletion and duplication syndromes: Rare association with pulmonary valve stenosis. Eur J Med Genet. 2013 Mar;56(3):144–9.

35. Cicenia M, Alesi V, Orlando V, Magliozzi M, di Tommaso S, Iodice FG, et al. 8p23.1 deletion: Look out for left ventricular hypertrabeculation and not only congenital heart diseases. Single-center experience and literature revision. Am J Med Genet A [Internet]. 2022 Mar 1 [cited 2022 Jul 19];188(3):883–95. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/ajmg.a.62598

36. Wat MJ, Shchelochkov OA, Holder AM, Breman AM, Dagli A, Bacino C, et al. Chromosome 8p23.1 Deletions as a Cause of Complex Congenital Heart Defects and Diaphragmatic Hernia. Am J Med Genet A [Internet]. 2009 Aug [cited 2022 Jul 19];149A(8):1661. Available from: /pmc/articles/PMC2765374/

37. Poot M, Eleveld MJ, van ’t Slot R, van Genderen MM, Verrijn Stuart AA, Hochstenbach R, et al. Proportional growth failure and oculocutaneous albinism in a girl with a 6.87 Mb deletion of region 15q26.2→qter. Eur J Med Genet. 2007 Nov;50(6):432–40.

38. Ono M, Tanaka M, Hiroshima S, Sawano K, Ogawa Y, Nagasaki K, et al. Diagnosis of Chromosome 15q-Terminal Deletion Syndrome through Elevated Fasting Serum Growth Hormone Levels. Endocrines 2022, Vol 3, Pages 92-99 [Internet]. 2022 Feb 23 [cited 2022 Jul 15];3(1):92–9. Available from: https://www.mdpi.com/2673-396X/3/1/8/htm

39. Jordan VK, Zaveri HP, Scott DA. 1p36 deletion syndrome: an update. Appl Clin Genet [Internet]. 2015 Aug 27 [cited 2022 Jul 16];8:189. Available from: /pmc/articles/PMC4555966/

40. Lodato V, Orlando V, Alesi V, di Tommaso S, Bengala M, Parlapiano G, et al. 1p36 Deletion Syndrome and the Aorta: A Report of Three New Patients and a Literature Review. J Cardiovasc Dev Dis [Internet]. 2021 Nov 1 [cited 2022 Jun 15];8(11). Available from: /pmc/articles/PMC8618808/

41. Kohli U. Shone's complex in a patient with chromosome 9q34.3 deletion (Kleefstra syndrome). Cardiol Young [Internet]. 2019 Feb 1 [cited 2022 Jul 19];29(2):249–51. Available from: https://www.cambridge.org/core/journals/cardiology-in-the-young/article/abs/shones-complex-in-a-patient-with-chromosome-9q343-deletion-kleefstra-syndrome/803F989BDCD6EDB1B96E5A46EC0E98E2

42. Yasuhara J, Garg V. Genetics of congenital heart disease: A narrative review of recent advances and clinical implications. Transl Pediatr. 2021;10(9):2366–86.

43. Mercer CL, Wilson DI. Structural Heart Disease: Genetic Influences. In: Mark D. Kilby, Anthony Johnson, Dick Oepkes, editors. Fetal Therapy [Internet]. 2nd editio. Cambridge University Press; 2019 [cited 2022 Apr 5]. p. 123–32. Available from: https://doi.org/10.1017/9781108564434.014

44. Southgate L, Sukalo M, Karountzos ASV, Taylor EJ, Collinson CS, Ruddy D, et al. Haploinsufficiency of the NOTCH1 Receptor as a Cause of Adams-Oliver Syndrome with Variable Cardiac Anomalies. Circ Cardiovasc Genet [Internet]. 2015 Aug 25 [cited 2022 Jul 19];8(4):572–81. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCGENETICS.115.001086

45. Digilio MC, Marino B, Dallapiccola B. Autosomal dominant inheritance of aplasia cutis congenita and congenital heart defect: A possible link to the Adams-Oliver syndrome. Am J Med Genet A. 2008 Nov 1;146(21):2842–4.

46. Turnpenny PD, Ellard S. Alagille syndrome: pathogenesis, diagnosis and management. European Journal of Human Genetics 2012 20:3 [Internet]. 2011 Sep

21 [cited 2022 Jul 19];20(3):251–7. Available from: https://www.nature.com/articles/ejhg2011181

47. Roberts AE, Allanson JE, Tartaglia M, Gelb BD. Noonan syndrome. The Lancet. 2013;381(9863):333–42.

48. Prendiville TW, Gauvreau K, Tworog-Dube E, Patkin L, Kucherlapati RS, Roberts AE, et al. Cardiovascular disease in Noonan syndrome. Arch Dis Child [Internet]. 2014 [cited 2022 Jul 20];99:629–34. Available from: http://adc.bmj.com/

49. Romano AA, Allanson JE, Dahlgren J, Gelb BD, Hall B, Pierpont ME, et al. Noonan Syndrome: Clinical Features, Diagnosis, and Management Guidelines. Pediatrics [Internet]. 2010 Oct 1 [cited 2022 Jun 15];126(4):746–59. Available from: /pediatrics/article/126/4/746/65699/Noonan-Syndrome-Clinical-Features-Diagnosis-and

50. Spiridon MR, Petris AO, Gorduza EV, Petras AS, Popescu R, Caba L. Holt-Oram Syndrome With Multiple Cardiac Abnormalities. Cardiol Res [Internet]. 2018 [cited 2022 Jun 15];9(5):324. Available from: /pmc/articles/PMC6188042/

51. Ashley F. Krauser, Subitchan Ponnarasu, Mark P. Schury. Holt Oram Syndrome [Internet]. StatPearls Publishing ; 2022 [cited 2022 Jul 19]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK513339/

52. Digilio CM, Marino B, Baban A, Dallapiccola B. Cardiovascular malformations in Adams-Oliver syndrome. Am J Med Genet A. 2015 May 1;167(5):1175–7.

53. Gripp KW, Hopkins E, Jenny K, Thacker D, Salvin J. Cardiac anomalies in Axenfeld-Rieger syndrome due to a novel FOXC1 mutation. Am J Med Genet A. 2013 Jan;161(1):114–9.

54. Corsten-Janssen N, Kerstjens-Frederikse WS, Du GJ, Sarvaas M, Baardman ME, Bakker MK, et al. The Cardiac Phenotype in Patients With a CHD7 Mutation. Circ Cardiovasc Genet [Internet]. 2013 [cited 2022 Jul 20];6:248–54. Available from: http://www.rug.nl/research/portal.

55. Thacoor A. Mitral valve prolapse and Marfan syndrome. Congenit Heart Dis. 2017 Jul 1;12(4):430–4.

56. Bruce D Gelb. Char Syndrome. In: Margaret P Adam, Ghayda M Mirzaa, Roberta A Pagon, Stephanie E Wallace, Lora JH Bean, Karen W Gripp, et al., editors. GeneReviews® [Internet] [Internet]. Seattle (WA): University of Washington, Seattle; 2003 [cited 2022 Jul 28]. p. 1993–2022. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1106/#char.Chapter_Notes

57. Satoda M, Zhao F, Diaz GA, Burn J, Goodship J, Davidson HR, et al. Mutations in TFAP2B cause Char syndrome, a familial form of patent ductus arteriosus. Nature Genetics 2000 25:1 [Internet]. 2000 May [cited 2022 Jul 28];25(1):42–6. Available from: https://www.nature.com/articles/ng0500_42

58. Qiao XH, Wang F, Zhang XL, Huang RT, Xue S, Wang J, et al. MEF2c loss-of-function mutation contributes to congenital heart defects. Int J Med Sci. 2017 Sep 8;14(11):1143–53.

59. Rocha H, Sampaio M, Rocha R, Fernandes S, Leão M. MEF2C haploinsufficiency syndrome: Report of a new MEF2C mutation and review. Eur J Med Genet. 2016 Sep 1;59(9):478–82.

60. Calcagni G, Digilio MC, Sarkozy A, Dallapiccola B, Marino B. Familial recurrence of congenital heart disease: An overview and review of the literature. Eur J Pediatr [Internet]. 2007 Feb 8 [cited 2022 Jul 21];166(2):111–6. Available from: https://link.springer.com/article/10.1007/s00431-006-0295-9

61. Nees SN, Chung WK. Genetic basis of human congenital heart disease. Cold Spring Harb Perspect Biol. 2020 Sep 1;12(9):1–40.

62. Fahed AC, Gelb BD, Seidman JGG, Seidman CE. Genetics of Congenital Heart Disease: The Glass Half Empty. Circ Res. 2013;112(4):707–20.

63. Sperling S, Grimm CH, Dunkel I, Mebus S, Sperling HP, Ebner A, et al. Identification and functional analysis of CITED2 mutations in patients with congenital heart defects. Hum Mutat. 2005 Dec;26(6):575–82.

64. Butler TL, Esposito G, Blue GM, Cole AD, Costa MW, Waddell LB, et al. GATA4 mutations in 357 unrelated patients with congenital heart malformation. Genet Test Mol Biomarkers [Internet]. 2010 Dec 1 [cited 2022 Jul 21];14(6):797–802. Available from: https://pubmed.ncbi.nlm.nih.gov/20874241/

65. Zhang Y, Ai F, Zheng J, Peng B. Associations of GATA4 genetic mutations with the risk of congenital heart disease: A meta-analysis. Medicine [Internet]. 2017 May 1 [cited 2022 Jul 21];96(18). Available from: /pmc/articles/PMC5419936/

66. Song ZP, Yan B. Potential roles of GATA binding protein 5 in cardiovascular diseases. Rev Cardiovasc Med [Internet]. 2020 Jun 1 [cited 2022 Jul 21];21(2):253–61. Available from: https://www.imrpress.com/journal/RCM/21/2/10.31083/j.rcm.2020.02.5104/htm

67. Wang J, Luo XJ, Xin YF, Liu Y, Liu ZM, Wang Q, et al. Novel GATA6 Mutations Associated with Congenital Ventricular Septal Defect or Tetralogy of Fallot. DNA Cell Biol [Internet]. 2012 Nov 1 [cited 2022 Jul 21];31(11):1610. Available from: /pmc/articles/PMC3482375/

68. Zhang E, Hong N, Chen S, Fu Q, Li F, Yu Y, et al. Targeted sequencing identifies novel GATA6 variants in a large cohort of patients with conotruncal heart defects. Gene. 2018 Jan 30;641:341–8.

69. Reamon-Buettner SM, Ciribilli Y, Traverso I, Kuhls B, Inga A, Borlak J. A functional genetic study identifies HAND1 mutations in septation defects of the human heart. Hum Mol Genet [Internet]. 2009 Oct 1 [cited 2022 Jul 21];18(19):3567–78. Available from: https://pubmed.ncbi.nlm.nih.gov/19586923/

70. Reamon-Buettner SM, Ciribilli Y, Inga A, Borlak J. A loss-of-function mutation in the binding domain of HAND1 predicts hypoplasia of the human hearts. Hum Mol Genet [Internet]. 2008 May 15 [cited 2022 Jul 21];17(10):1397–405. Available from: https://pubmed.ncbi.nlm.nih.gov/18276607/

71. Cohen ASA, Simotas C, Webb BD, Shi H, Khan WA, Edelmann L, et al. Haploinsufficiency of the basic helix–loop–helix transcription factor HAND2 causes congenital heart defects. Am J Med Genet A. 2020 May 1;182(5):1263–7.

72. Sun YM, Wang J, Qiu XB, Yuan F, Li RG, Xu YJ, et al. A HAND2 Loss-of-Function Mutation Causes Familial Ventricular Septal Defect and Pulmonary Stenosis. G3: Genes|Genomes|Genetics [Internet]. 2016 [cited 2022 Jul 21];6(4):987. Available from: /pmc/articles/PMC4825666/

73. Turki S al, Manickaraj AK, Mercer CL, Gerety SS, Hitz P, Lindsay S, et al. Rare variants in NR2F2 cause congenital heart defects in humans. Am J Hum Genet [Internet]. 2014 [cited 2022 Jun 11];94(4):574–85. Available from: http://dx.doi.org/10.1016/j.ajhg.2014.03.007.

74. Ellesøe SG, Johansen MM, Bjerre JV, Hjortdal VE, Brunak S, Larsen LA. Familial Atrial Septal Defect and Sudden Cardiac Death: Identification of a Novel NKX2-5 Mutation and a Review of the Literature. Congenit Heart Dis [Internet]. 2016 May 1 [cited 2022 Jul 27];11(3):283–90. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/chd.12317

75. Ta-Shma A, El-lahham N, Edvardson S, Stepensky P, Nir A, Perles Z, et al. Conotruncal malformations and absent thymus due to a deleterious NKX2-6 mutation. J Med Genet [Internet]. 2014 [cited 2022 Jul 21];51(4):268–70. Available from: https://pubmed.ncbi.nlm.nih.gov/24421281/

76. Ritter A, Werner P, Latney B, Krock BL, Santani A, Bedoukian E, et al. NKX2-6 related congenital heart disease: Biallelic homeodomain-disrupting variants and truncus arteriosus. Am J Med Genet A [Internet]. 2020 Jun 1 [cited 2022 Jul 21];182(6):1454–9. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/ajmg.a.61550

77. Kirk EP, Sunde M, Costa MW, Rankin SA, Wolstein O, Castro ML, et al. Mutations in Cardiac T-Box Factor Gene TBX20 Are Associated with Diverse Cardiac Pathologies, Including Defects of Septation and Valvulogenesis and Cardiomyopathy. The American Journal of Human Genetics. 2007 Aug 1;81(2):280–91.

78. Shen L, Li ZZ, Shen AD, Liu H, Bai S, Guo J, et al. Association of NFATc1 gene polymorphism with ventricular septal defect in the Chinese Han population. Chin Med J (Engl) [Internet]. 2013 Jan 5 [cited 2022 Jul 28];126(1):78–81. Available from: https://journals.lww.com/cmj/Fulltext/2013/01010/Association_ofNFATc1gene_polymorphism_with.15.aspx

79. de Luca A, Sarkozy A, Ferese R, Consoli F, Lepri F, Dentici M, et al. New mutations in ZFPM2/FOG2 gene in tetralogy of Fallot and double outlet right ventricle. Clin Genet [Internet]. 2011 Aug 1 [cited 2022 Jul 28];80(2):184–90. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1399-0004.2010.01523.x

80. Pizzuti A, Sarkozy A, Newton AL, Conti E, Flex E, Digilio MC, et al. Mutations of ZFPM2/FOG2 gene in sporadic cases of tetralogy of Fallot. Hum Mutat [Internet]. 2003 Nov 1 [cited 2022 Jul 28];22(5):372–7. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/humu.10261

81. Robinson SW, Morris CD, Goldmuntz E, Reller MD, Jones MA, Steiner RD, et al. Missense Mutations in CRELD1 Are Associated with Cardiac Atrioventricular Septal Defects. Am J Hum Genet [Internet]. 2003 Apr 1 [cited 2022 Jul 28];72(4):1047. Available from: /pmc/articles/PMC1180336/

82. Wei W, Li B, Li F, Sun K, Jiang X, Xu R. Identification of FOXH1 mutations in patients with sporadic conotruncal heart defect. Clin Genet [Internet]. 2020 Apr 1 [cited 2022 Jul 28];97(4):576–85. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/cge.13710

83. van Walree ES, Dombrowsky G, Jansen IE, Mirkov MU, Zwart R, Ilgun A, et al. Germline variants in HEY2 functional domains lead to congenital heart defects and thoracic aortic aneurysms. Genetics in Medicine. 2021 Jan 1;23(1):103–10.

84. Fischer A, Klamt B, Schumacher N, Glaeser C, Hansmann I, Fenge H, et al. Phenotypic variability in Hey2 -/- mice and absence of HEY2 mutations in patients with congenital heart defects or Alagille syndrome. Mammalian Genome. 2004 Sep;15(9):711–6.

85. Barnes RM, Black BL. Nodal Signaling and Congenital Heart Defects. In: Toshio Nakanishi, Roger R. Markwald, H.Scott Baldwin, Bradley B. Keller, Deepak Srivastava, Hiroyuki Yamagishi, editors. Etiology and Morphogenesis of Congenital Heart Disease: From Gene Function and Cellular Interaction to Morphology [Internet]. Springer; 2016 [cited 2022 Jul 28]. p. 183–92. Available from: https://www.ncbi.nlm.nih.gov/books/NBK500272/

86. Granadillo JL, Chung WK, Hecht L, Corsten-Janssen N, Wegner D, Nij Bijvank SWA, et al. Variable cardiovascular phenotypes associated with SMAD2 pathogenic variants. Hum Mutat. 2018 Dec 1;39(12):1875–84.

87. Kloth K, Bierhals T, Johannsen J, Harms FL, Juusola J, Johnson MC, et al. Biallelic variants in SMAD6 are associated with a complex cardiovascular phenotype. Hum Genet [Internet]. 2019 Jun 1 [cited 2022 Jul 28];138(6):625–34. Available from: https://link.springer.com/article/10.1007/s00439-019-02011-x

88. Gillis E, Kumar AA, Luyckx I, Preuss C, Cannaerts E, Beek G van de, et al. Candidate gene resequencing in a large bicuspid aortic valve-associated thoracic aortic aneurysm cohort: SMAD6 as an important contributor. Front Physiol. 2017 Jun 13;8(JUN):400.

89. Reuter MS, Jobling R, Chaturvedi RR, Manshaei R, Costain G, Heung T, et al. Haploinsufficiency of vascular endothelial growth factor related signaling genes is associated with tetralogy of Fallot. Genetics in Medicine 2018 21:4 [Internet]. 2018 Sep 20 [cited 2022 Jul 28];21(4):1001–7. Available from: https://www.nature.com/articles/s41436-018-0260-9

90. Augière C, Mégy S, el Malti R, Boland A, el Zein L, Verrier B, et al. A Novel Alpha Cardiac Actin (ACTC1) Mutation Mapping to a Domain in Close Contact with Myosin Heavy Chain Leads to a Variety of Congenital Heart Defects, Arrhythmia and Possibly Midline Defects. PLoS One [Internet]. 2015 Jun 10 [cited 2022 Jul 28];10(6):e0127903. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127903

91. Frank D, Rangrez AY, Friedrich C, Dittmann S, Stallmeyer B, Yadav P, et al. Cardiac α-actin (ACTC1) gene mutation causes atrial-septal defects associated with late-onset dilated cardiomyopathy. Circ Genom Precis Med [Internet]. 2019 Aug 1 [cited 2022 Jul 28];12(8):345–56. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCGEN.119.002491

92. Ronen Durst, Kimberly Sauls, David S. Peal, Annemarieke deVlaming, Katelynn Toomer, Maire Leyne, et al. Mutations in DCHS1 cause mitral valve prolapse. Leticia Fernandez-Friera. 2015;525:109–13.

93. Razmara E, Garshasbi M. Whole-exome sequencing identifies R1279X of MYH6 gene to be associated with congenital heart disease. BMC Cardiovasc Disord [Internet]. 2018 Jul 3 [cited 2022 Jul 28];18(1):1–7. Available from: https://bmccardiovascdisord.biomedcentral.com/articles/10.1186/s12872-018-0867-4

94. Ritter A, Leonard J, Gray C, Izumi K, Levinson K, Nair DR, et al. MYH7 variants cause complex congenital heart disease. Am J Med Genet A [Internet]. 2022 [cited 2022 Jul 28]; Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/ajmg.a.62766

95. van der Linde IHM, Hiemstra YL, Bökenkamp R, van Mil AM, Breuning MH, Ruivenkamp C, et al. A Dutch MYH7 founder mutation, p.(Asn1918Lys), is associated with early onset cardiomyopathy and congenital heart defects. Netherlands Heart Journal [Internet]. 2017 Dec 1 [cited 2022 Jul 28];25(12):675. Available from: /pmc/articles/PMC5691818/

96. Lahaye S, Corsmeier D, Basu M, Bowman JL, Fitzgerald-Butt S, Zender G, et al. Utilization of Whole Exome Sequencing to Identify Causative Mutations in Familial Congenital Heart Disease. Circ Cardiovasc Genet [Internet]. 2016 Aug 1 [cited 2022 Jul 28];9(4):320–9. Available from: https://www.ahajournals.org/doi/abs/10.1161/circgenetics.115.001324

97. Zhu L, Vranckx R, van Kien PK, Lalande A, Boisset N, Mathieu F, et al. Mutations in myosin heavy chain 11 cause a syndrome associating thoracic aortic aneurysm/aortic dissection and patent ductus arteriosus. Nature Genetics 2006 38:3 [Internet]. 2006 Jan 29 [cited 2022 Jul 28];38(3):343–9. Available from: https://www.nature.com/articles/ng1721

98. Pereira FA, Yuhong Q, Zhou G, Tsai MJ, Tsai SY. The orphan nuclear receptor COUP-TFII is required for angiogenesis and heart development. Genes Dev. 1999;13(8):1037–49.

99. Lin FJ, You LR, Yu CT, Hsu WH, Tsai MJ, Tsai SY. Endocardial cushion morphogenesis and coronary vessel development require chicken ovalbumin upstream promoter-transcription factor II. Arterioscler Thromb Vasc Biol. 2012;32(11):135–46.

100. Nakamura E, Makita Y, Okamoto T, Nagaya K, Hayashi T, Sugimoto M, et al. 5.78 Mb terminal deletion of chromosome 15q in a girl, evaluation of NR2F2 as candidate gene for congenital heart defects. Eur J Med Genet [Internet]. 2011;54(3):354–6. Available from: http://dx.doi.org/10.1016/j.ejmg.2010.12.004

101. al Turki S, Manickaraj AK, Mercer CL, Gerety SS, Hitz MP, Lindsay S, et al. Rare variants in NR2F2 cause congenital heart defects in humans. Am J Hum Genet. 2014;94(4):574–85.

102. Qiao XH, Wang Q, Wang J, Liu XY, Xu YJ, Huang RT, et al. A novel NR2F2 loss-of-function mutation predisposes to congenital heart defect. Eur J Med Genet [Internet]. 2018;61(4):197–203. Available from: https://doi.org/10.1016/j.ejmg.2017.12.003

103. Bashamboo A, Eozenou C, Jorgensen A, Bignon-Topalovic J, Siffroi JP, Hyon C, et al. Loss of Function of the Nuclear Receptor NR2F2, Encoding COUP-TF2, Causes Testis Development and Cardiac Defects in 46,XX Children. The American Journal of Human Genetics [Internet]. 2018 Mar 1 [cited 2022 Jul 21];102(3):487–93. Available from: http://www.cell.com/article/S0002929718300430/fulltext

104. Hsu WH, Chen CM, You LR. COUP-TFII is required for morphogenesis of the neural crest-derived tympanic ring. Scientific Reports 2017 7:1 [Internet]. 2017 Sep 28 [cited 2022 Aug 3];7(1):1–13. Available from: https://www.nature.com/articles/s41598-017-12665-0

105. Li X, Large MJ, Creighton CJ, Lanz RB, Jeong JW, Young SL, et al. COUP-TFII Regulates Human Endometrial Stromal Genes Involved in Inflammation. Molecular Endocrinology [Internet]. 2013 Dec 1 [cited 2022 Aug 3];27(12):2041–54. Available from: https://academic.oup.com/mend/article/27/12/2041/2615061

106. Grinspon RP, Rey RA. Molecular characterization of XX maleness. Int J Mol Sci. 2019 Dec 1;20(23).

107. Lee HJ, Kao CY, Lin SC, Xu M, Xie X, Tsai SY, et al. Dysregulation of nuclear receptor COUP-TFII impairs skeletal muscle development. Sci Rep [Internet]. 2017 Jun 9 [cited

2022 Aug 3];7(1):1–10. Available from: https://www.nature.com/articles/s41598-017-03475-5

108. Ishii S, Koibuchi N. COUP-TFII in Kidneys, from Embryos to Sick Adults. Diagnostics 2022, Vol 12, Page 1181 [Internet]. 2022 May 9 [cited 2022 Aug 2];12(5):1181. Available from: https://www.mdpi.com/2075-4418/12/5/1181/htm

109. Wu SP, Yu CT, Tsai SY, Tsai MJ. Choose your destiny: Make a cell fate decision with COUP-TFII. J Steroid Biochem Mol Biol. 2016 Mar 1;157:7–12.

110. Pereira FA, Yuhong Q, Zhou G, Tsai MJ, Tsai SY. The orphan nuclear receptor COUP-TFII is required for angiogenesis and heart development. Genes Dev [Internet]. 1999 Apr 15 [cited 2022 Jun 11];13(8):1037–49. Available from: https://pubmed.ncbi.nlm.nih.gov/10215630/

111. Srinivasan RS, Geng X, Yang Y, Wang Y, Mukatira S, Studer M, et al. The nuclear hormone receptor Coup-TFII is required for the initiation and early maintenance of Prox1 expression in lymphatic endothelial cells. Genes Dev. 2010 Apr 1;24(7):696–707.

112. Risebro CA, Searles RG, Melville AAD, Ehler E, Jina N, Shah S, et al. Prox1 maintains muscle structure and growth in the developing heart. Development [Internet]. 2009 [cited 2020 Feb 26];136:699. Available from: http://rsb.info.nih.gov/ij

113. Aranguren XL, Beerens M, Coppiello G, Wiese C, Vandersmissen I, Nigro A lo, et al. COUP-TFII orchestrates venous and lymphatic endothelial identity by homo- or hetero-dimerisation with PROX1. J Cell Sci [Internet]. 2013 Mar 1 [cited 2022 Jul 12];126(5):1164–75. Available from: https://journals.biologists.com/jcs/article/126/5/1164/54245/COUP-TFII-orchestrates-venous-and-lymphatic

114. Wu S pin, Cheng CM, Lanz RB, Wang T, Respress JL, Ather S, et al. Atrial Identity Is Determined by A COUP-TFII Regulatory Network. Dev Cell [Internet]. 2013 May 5 [cited 2022 Aug 4];25(4):417. Available from: /pmc/articles/PMC3687546/

115. You LR, Takamoto N, Yu CT, Tanaka T, Kodama T, Demayo FJ, et al. Mouse lacking COUP-TFII as an animal model of Bochdalek-type congenital diaphragmatic hernia.

Proc Natl Acad Sci U S A [Internet]. 2005 Nov 8 [cited 2022 Aug 2];102(45):16351–6. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.0507832102

116. High FA, Bhayani P, Wilson JM, Bult CJ, Donahoe PK, Longoni M. De novo frameshift mutation in COUP-TFII (NR2F2) in human congenital diaphragmatic hernia. Am J Med Genet A [Internet]. 2016 Sep 1 [cited 2022 Jun 12];170(9):2457–61. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/ajmg.a.37830

117. Arsov T, Kelecic J, Frkovic SH, Sestan M, Kifer N, Andrews D, et al. Expanding the clinical spectrum of pathogenic variation in NR2F2: Asplenia. Eur J Med Genet. 2021 Dec 1;64(12).

118. Upadia J, Gonzales PR, Robin NH. Novel de novo pathogenic variant in the NR2F2 gene in a boy with congenital heart defect and dysmorphic features. Am J Med Genet A. 2018 Jun 1;176(6):1423–6.

119. Carvalheira G, Malinverni AM, Moysés-Oliveira M, Ueta R, Cardili L, Monteagudo P, et al. The Natural History of a Man With Ovotesticular 46,XX DSD Caused by a Novel 3-Mb 15q26.2 Deletion Containing NR2F2 Gene. J Endocr Soc [Internet]. 2019 Nov 1 [cited 2022 Sep 5];3(11):2107–13. Available from: https://academic.oup.com/jes/article/3/11/2107/5555550

120. Biggio JR, Descartes MD, Carroll AJ, Holt RL. Congenital diaphragmatic hernia: Is 15q26.1-26.2 a candidate locus? Am J Med Genet A [Internet]. 2004 Apr 15 [cited 2022 Jun 12];126A(2):183–5. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/ajmg.a.20464

121. Auwerx J, Baulieu E, Beato M, Becker-Andre M, Burbach PH, Camerino G, et al. A unified nomenclature system for the nuclear receptor superfamily. Cell. 1999.

122. NR2F2 nuclear receptor subfamily 2 group F member 2 [Homo sapiens (human)] [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. 2004 [cited 2022 Sep 7]. Available from: https://www.ncbi.nlm.nih.gov/gene/7026

123. Qiu Y, Krishnan V, Zeng Z, Gilbert DJ, Copeland NG, Gibson L, et al. Isolation, Characterization, and Chromosomal Localization of Mouse and Human COUP-TF I and II Genes. Genomics. 1995;29:240–6.

124. Tsai SY, Tsai MJ. Chick Ovalbumin Upstream Promoter-Transcription Factors (COUP-TFs): Coming of Age. Endocr Rev. 1997;18(2):229–40.

125. Cooney AJ, Tsai SY, O'malley BW, Tsai MJ. Chicken Ovalbumin Upstream Promoter Transcription Factor (COUP-TF) Dimers Bind to Different GGTCA Response Elements, Allowing COUP-TF To Repress Hormonal Induction of the Vitamin D3, Thyroid Hormone, and Retinoic Acid Receptors [Internet]. MOLECULAR AND CELLULAR BIOLOGY. 1992 [cited 2020 Jan 8]. Available from: http://mcb.asm.org/

126. Kruse SW, Suino-Powell K, Zhou XE, Kretschman JE, Reynolds R, Vonrhein C, et al. Identification of COUP-TFII orphan nuclear receptor as a retinoic acid-activated receptor. PLoS Biol. 2008;6(9):2002–15.

127. Park JI, Tsai SY, Tsai MJ. Molecular mechanism of chicken ovalbumin upstream promoter-transcription factor (COUP-TF) actions. Keio J Med. 2003;52(3):174–81.

128. Polvani S, Pepe S, Milani S, Galli A. COUP-TFII in Health and Disease. Cells. 2020;9(101):1–30.

129. Achatz G, Hölzl B, Speckmayer R, Hauser C, Sandhofer F, Paulweber B. Functional domains of the human orphan receptor ARP-1/COUP-TFII involved in active repression and transrepression. Mol Cell Biol. 1997;17(9):4914–32.

130. Rémy R, Guével LG, Frédérik ##, Frédérik F, Martinez-Jimenez CP, Bizot M, et al. Inactivation of the Nuclear Orphan Receptor COUP-TFII by Small Chemicals. 2017 [cited 2022 Aug 4]; Available from: https://pubs.acs.org/sharingguidelines

131. Wang T, Wang Z, de Fabritus L, Tao J, Saied EM, Lee HJ, et al. 1-deoxysphingolipids bind to COUP-TF to modulate lymphatic and cardiac cell development. Dev Cell. 2021 Nov 22;56(22):3128-3145.e15.

132. Bertacchi M, Parisot J, Studer M. The pleiotropic transcriptional regulator COUP-TFI plays multiple roles in neural development and disease. Vol. 1705, Brain Research. Elsevier B.V.; 2019. p. 75–94.

133. Alfano C, Magrinelli E, Harb K, Studer M. The nuclear receptors COUP-TF: A long-lasting experience in forebrain assembly. Cellular and Molecular Life Sciences [Internet]. 2014 Jan 23 [cited 2022 Aug 8];71(1):43–62. Available from: https://link.springer.com/article/10.1007/s00018-013-1320-6

134. Pereira FA, Tsai MJ, Tsai SY. COUP-TF orphan nuclear receptors in development and differentiation. CMLS, Cell Mol Life Sci. 2000;57.

135. Coppola U, Waxman JS. Origin and evolutionary landscape of Nr2f transcription factors across Metazoa. PLoS One [Internet]. 2021 Nov 1 [cited 2022 Aug 8];16(11):e0254282. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0254282

136. Jonk LJC, de Jonge MEJ, Pals CEGM, Wissink S, Vervaart JMA, Schoorlemmer J, et al. Cloning and expression during development of three murine members of the COUP family of nuclear orphan receptors. Mech Dev. 1994 Jul 1;47(1):81–97.

137. Lutz B, Kuratani S, Cooney AJ, Wawersik S, Tsai SY, Eichele G, et al. Developmental regulation of the orphan receptor COUP-TF II gene in spinal motor neurons. Development [Internet]. 1994 Jan [cited 2022 Aug 12];120(1):25–36. Available from: https://pubmed.ncbi.nlm.nih.gov/8119130/

138. Matharu PJ, Sweeney GE. Cloning and sequencing of a COUP transcription factor gene expressed in Xenopus embryos. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 1992 Feb 11;1129(3):331–4.

139. Fjose A, Weber U, Mlodzik M. A novel vertebrate svp-related nuclear receptor is expressed as a step gradient in developing rhombomeres and is affected by retinoic acid. Mech Dev [Internet]. 1995 [cited 2022 Aug 12];52(2–3):233–46. Available from: https://pubmed.ncbi.nlm.nih.gov/8541212/

140. Love CE, Prince VE. Expression and retinoic acid regulation of the zebrafish nr2f orphan nuclear receptor genes. Developmental Dynamics [Internet]. 2012 Oct 1

[cited 2022 Aug 12];241(10):1603–15. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/dvdy.23838

141. Lim GB. Complexity and plasticity of cardiac cellular composition. Nature Reviews Cardiology 2020 17:12 [Internet]. 2020 Oct 6 [cited 2022 Jul 22];17(12):759–759. Available from: https://www.nature.com/articles/s41569-020-00464-6

142. Zhou P, Pu WT. Recounting cardiac cellular composition. Circ Res [Internet]. 2016 [cited 2020 Feb 26];118(3):368–70. Available from: http://ahajournals.org

143. Banerjee I, Fuseler JW, Price RL, Borg TK, Baudino TA. Determination of cell types and numbers during cardiac development in the neonatal and adult rat and mouse. Am J Physiol Heart Circ Physiol [Internet]. 2007 Sep [cited 2022 Jul 22];293(3):1883–91. Available from: www.ajpheart.org

144. Nag AC. Study of non-muscle cells of the adult mammalian heart: a fine structural analysis and distribution. Cytobios [Internet]. 1980 Jan 1 [cited 2022 Sep 14];28(109):41–61. Available from: https://europepmc.org/article/med/7428441

145. Hall C, Gehmlich K, Denning C, Pavlovic D. Complex Relationship Between Cardiac Fibroblasts and Cardiomyocytes in Health and Disease. J Am Heart Assoc [Internet]. 2021 Mar 2 [cited 2022 Jun 6];10(5):1–15. Available from: https://www.ahajournals.org/doi/abs/10.1161/JAHA.120.019338

146. Tucker NR, Chaffin M, Fleming SJ, Hall AW, Parsons VA, Bedi KC, et al. Transcriptional and Cellular Diversity of the Human Heart. Circulation [Internet]. 2020 Aug 4 [cited 2022 Mar 24];142(5):466–82. Available from: https://singlecell.broadinstitute.

147. Litviňuková M, Talavera-López C, Maatz H, Reichart D, Worth CL, Lindberg EL, et al. Cells of the adult human heart. Nature 2020 588:7838 [Internet]. 2020 Sep 24 [cited 2021 Dec 2];588(7838):466–72. Available from: https://www.nature.com/articles/s41586-020-2797-4

148. Suryawanshi H, Clancy R, Morozov P, Halushka MK, Buyon JP, Tuschl T. Cell atlas of the fetal human heart and implications for autoimmune-mediated congenital heart block. Cardiovasc Res. 2019;

149. Valiente-Alandi I, Schafer AE, Blaxall BC. Extracellular matrix-mediated cellular communication in the heart. J Mol Cell Cardiol. 2016;91:228–37.

150. Silva AC, Pereira C, Fonseca ACRG, Pinto-do-Ó P, Nascimento DS. Bearing My Heart: The Role of Extracellular Matrix on Cardiac Development, Homeostasis, and Injury Response. Vol. 8, Frontiers in Cell and Developmental Biology. Frontiers Media S.A.; 2021.

151. Chang CW, Dalgliesh AJ, López JE, Griffiths LG. Cardiac extracellular matrix proteomics: Challenges, techniques, and clinical implications. Proteomics Clin Appl. 2016;10(1):39–50.

152. Takawale A, Sakamuri SSVP, Kassiri Z. Extracellular matrix communication and turnover in cardiac physiology and pathology. Compr Physiol. 2015;5(2):687–719.

153. Sarohi V, Chakraborty S, Basak T. Exploring the cardiac ECM during fibrosis: A new era with next-gen proteomics. Front Mol Biosci. 2022;9(November):1–25.

154. Kumar V, Abbas AK, Aster JC, Perkins JA. Robbins and Cotran Pathologic Basis of Disease. Ninth edit. Kumar V, Abbas AK, Aster JC, editors. Elsevier Saunders; 1391 p.

155. Freedman BR, Bade ND, Riggin CN, Zhang S, Haines PG, Ong KL, et al. The (dys)functional extracellular matrix. Biochim Biophys Acta Mol Cell Res [Internet]. 2015;1853(11):3153–64. Available from: http://dx.doi.org/10.1016/j.bbamcr.2015.04.015

156. Boland E, Quondamatteo F, van Agtmael T. The role of basement membranes in cardiac biology and disease. Biosci Rep. 2021;41(8):1–21.

157. Anto Michel N, Ljubojevic-Holzer S, Bugger H, Zirlik A. Cellular Heterogeneity of the Heart. Front Cardiovasc Med. 2022 Apr 25;9.

158. Souders CA, Bowers SLK, Baudino TA. Cardiac Fibroblast: The Renaissance Cell. Circ Res. 2009;105(12):1164–76.

159. Gittenberger-de Groot AC, Vrancken Peeters MPFM, Mentink MMT, Gourdie RG, Poelmann RE. Epicardium-derived cells contribute a novel population to the myocardial wall and the atrioventricular cushions. Circ Res. 1998;82(10):1043–52.

160. Dettman RW, Denetclaw W, Ordahl CP, Bristow J. Common epicardial origin of coronary vascular smooth muscle, perivascular fibroblasts, and intermyocardial fibroblasts in the avian heart. Dev Biol. 1998 Jan 15;193(2):169–81.

161. Acharya A, Baek ST, Huang G, Eskiocak B, Goetsch S, Sung CY, et al. The bHLH transcription factor Tcf21 is required for lineagespecific EMT of cardiac fibroblast progenitors. Development (Cambridge). 2012;139(12):2139–49.

162. Vega-Hernández M, Kovacs A, de Langhe S, Ornitz DM. FGF10/FGFR2b signaling is essential for cardiac fibroblast development and growth of the myocardium. Development. 2011;138(15):3331–40.

163. Moore-Morris T, Guimarães-Camboa N, Banerjee I, Zambon AC, Kisseleva T, Velayoudon A, et al. Resident fibroblast lineages mediate pressure overload-induced cardiac fibrosis. Journal of Clinical Investigation. 2014 Jul 1;124(7):2921–34.

164. Ali SR, Ranjbarvaziri S, Talkhabi M, Zhao P, Subat A, Hojjat A, et al. Developmental heterogeneity of cardiac fibroblasts does not predict pathological proliferation and activation. Circ Res [Internet]. 2014 Sep 12 [cited 2020 Feb 26];115(7):625–35. Available from: https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.115.303794

165. Ieda M, Tsuchihashi T, Ivey KN, Ross RS, Hong TT, Shaw RM, et al. Cardiac Fibroblasts Regulate Myocardial Proliferation through β1 Integrin Signaling. Dev Cell. 2009 Feb 17;16(2):233–44.

166. Lajiness JD, Conway SJ. Origin, development, and differentiation of cardiac fibroblasts. Vol. 70, Journal of Molecular and Cellular Cardiology. Academic Press; 2014. p. 2–8.

167. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell. 2016;

168. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens.

Cell [Internet]. 2016;167(7):1853-1866.e17. Available from: http://dx.doi.org/10.1016/j.cell.2016.11.038

169. Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell. 2016 Dec 15;167(7):1867-1882.e21.

170. Sirvent S, Vallejo AF, Davies J, Clayton K, Wu Z, Woo J, et al. Genomic programming of IRF4-expressing human Langerhans cells. Nature Communications [Internet]. 2020 [cited 2020 Feb 26];11(313). Available from: https://doi.org/10.1038/s41467-019-14125-x

171. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Feng Zhan. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013;8(11):2281–308.

172. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nature Methods 2014 11:8 [Internet]. 2014 Jul 30 [cited 2021 Oct 26];11(8):783–4. Available from: https://www.nature.com/articles/nmeth.3047

173. Melsted P, Booeshaghi AS, Gao F, Beltrame E, Lu L, Eldjárn K, et al. Modular and efficient pre-processing of single-cell RNA-seq. [cited 2020 Feb 26]; Available from: https://doi.org/10.1101/673285

174. Bernstein NJ, Fong NL, Lam I, Roy MA, Hendrickson DG, Kelley DR. Solo: Doublet Identification in Single-Cell RNA-Seq via Semi-Supervised Deep Learning. 2020 [cited 2022 Jun 16]; Available from: https://doi.org/10.1016/j.cels.2020.05.010

175. Wolf FA, Angerer P, Theis FJ. SCANPY: Large-scale single-cell gene expression data analysis. Genome Biol [Internet]. 2018 Feb 6 [cited 2022 Jul 11];19(1):1–5. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1382-0

176. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. 2019 Apr 24;8(4):281-291.e9.

177. Lun ATL, Riesenfeld S, Andrews T, Dao TP, Gomes T, Marioni JC. EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing

data. Genome Biol [Internet]. 2019 Mar 22 [cited 2022 Jun 16];20(1):1–9. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1662-y

178. Lause J, Berens P, Kobak D. Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data. Genome Biol [Internet]. 2021 Dec 1 [cited 2022 Jun 16];22(1):1–20. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02451-7

179. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Scientific Reports 2019 9:1 [Internet]. 2019 Mar 26 [cited 2022 Jun 16];9(1):1–12. Available from: https://www.nature.com/articles/s41598-019-41695-z

180. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. J Open Source Softw [Internet]. 2018 Sep 2 [cited 2022 Jun 16];3(29):861. Available from: https://joss.theoj.org/papers/10.21105/joss.00861

181. Churko JM, Garg P, Treutlein B, Venkatasubramanian M, Wu H, Lee J, et al. Defining human cardiac transcription factor hierarchies using integrated single-cell heterogeneity analysis. Nat Commun. 2018;9(1).

182. Torre D, Lachmann A, Ma'ayan Correspondence A. BioJupies: Automated Generation of Interactive Notebooks for RNA-Seq Data Analysis in the Cloud. 2018 [cited 2020 Jan 31]; Available from: https://doi.org/10.1016/j.cels.2018.10.007

183. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res [Internet]. 2015 [cited 2020 Jan 31];43(7). Available from: https://academic.oup.com/nar/article-abstract/43/7/e47/2414268

184. Ge SX, Jung D, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics [Internet]. 2020 Apr 15 [cited 2022 Jul 8];36(8):2628–9. Available from: https://academic.oup.com/bioinformatics/article/36/8/2628/5688742

185. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res [Internet]. 2018 [cited 2020 Feb 17];47:607–13. Available from: https://string-db.org/.

186. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nature Communications 2017 8:1 [Internet]. 2017 Jan 16 [cited 2022 Sep 15];8(1):1–12. Available from: https://www.nature.com/articles/ncomms14049

187. Polański K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park JE. BBKNN: fast batch alignment of single cell transcriptomes. Bioinformatics [Internet]. 2020 Feb 1 [cited 2022 Jun 17];36(3):964–5. Available from: https://academic.oup.com/bioinformatics/article/36/3/964/5545955

188. Aaron T. L. Lun, Karsten Bach, John C. Marioni. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol. 2016;17(75):1–14.

189. Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre1. Fast unfolding of communities in large networks. J Stat Mech. 2008;

190. Xin M, Olson EN, Bassel-Duby R. Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair. Nat Rev Mol Cell Biol [Internet]. 2013 Aug [cited 2022 Sep 11];14(8):529. Available from: /pmc/articles/PMC3757945/

191. Wang YC, Neckelmann N, Mayne A, Herskowitz A, Srinivasan A, Sell KW, et al. Establishment of a Human Fetal Cardiac Myocyte Cell Line. In Vitro Cellular & Developmental Biology. 1991;27A(1):63–74.

192. Dewing J. Reprogramming of Primary Human Fetal Fibroblasts towards Cardiomyocytes. University of Southampton; 2015.

193. Bers DM. Cardiac excitation–contraction coupling. Nature 2002 415:6868 [Internet]. 2002 Jan 10 [cited 2022 Sep 11];415(6868):198–205. Available from: https://www.nature.com/articles/415198a

194. Furtado MB, Nim HT, Boyd SE, Rosenthal NA. View from the heart: cardiac fibroblasts in development, scarring and regeneration. Development [Internet]. 2016;143(3):387–97. Available from: http://dev.biologists.org/cgi/doi/10.1242/dev.120576

195. Zhou P, Pu WT. Recounting cardiac cellular composition. Circ Res [Internet]. 2016 [cited 2022 May 26];118(3):368–70. Available from: http://ahajournals.org

196. Ribatti D, Tamma R, Ruggieri S, Annese T, Crivellato E. Surface markers: An identity card of endothelial cells. Microcirculation. 2020;27(1):1–8.

197. Brutsaert DL. Cardiac endothelial-myocardial signaling: Its role in cardiac growth, contractile performance, and rhythmicity. Physiol Rev [Internet]. 2003 [cited 2022 May 23];83(1):59–115. Available from: https://journals.physiology.org/doi/full/10.1152/physrev.00017.2002

198. Ferrero E, Ferrero ME, Pardi R, Zocchi MR. The platelet endothelial cell adhesion molecule-1 (PECAM1) contributes to endothelial barrier function. FEBS Lett. 1995 Nov 6;374(3):323–6.

199. Wang L, Yang Y, Ma H, Xie Y, Xu J, Near D, et al. Single-cell dual-omics reveals the transcriptomic and epigenomic diversity of cardiac non-myocytes. [cited 2022 May 26]; Available from: https://academic.oup.com/cardiovascres/article/118/6/1548/6220323

200. Sauteur L, Krudewig A, Herwig L, Ehrenfeuchter N, Lenard A, Affolter M, et al. Cdh5/VE-cadherin Promotes Endothelial Cell Interface Elongation via Cortical Actin Polymerization during Angiogenic Sprouting. Cell Rep. 2014 Oct 23;9(2):504–13.

201. Brutsaert DL, Andries LJ. The endocardial endothelium. Am J Physiol Heart Circ Physiol. 1992;263(4 32-4).

202. Garlanda C, Dejana E. Heterogeneity of endothelial cells: Specific markers. Arterioscler Thromb Vasc Biol [Internet]. 1997 [cited 2022 May 23];17(7):1193–202. Available from: http://ahajournals.org

203. Zhuge Y, Zhang J, Qian F, Wen Z, Niu C, Xu K, et al. Role of smooth muscle cells in Cardiovascular Disease. Int J Biol Sci [Internet]. 2020 [cited 2022 May 26];2020(14):2741–51. Available from: http://www.ijbs.com://creativecommons.org/licenses/by/4.0/

204. Doll S, Dreßen M, Geyer PE, Itzhak DN, Braun C, Doppler SA, et al. Region and cell-type resolved quantitative proteomic map of the human heart. Nature Communications 2017 8:1 [Internet]. 2017 Nov 13 [cited 2022 May 26];8(1):1–13. Available from: https://www.nature.com/articles/s41467-017-01747-2

205. Hilenski LL, Griendling KK. Vascular Smooth Muscle [Internet]. Second Edi. Vascular Medicine: A Companion to Braunwald's Heart Disease: Second Edition. Elsevier Inc.; 2013. 25–42 p. Available from: http://dx.doi.org/10.1016/B978-1-4377-2930-6.00003-3

206. Bhana B, Iyer RK, Chen WLK, Zhao R, Sider KL, Likhitpanichkul M, et al. Influence of substrate stiffness on the phenotype of heart cells. Biotechnol Bioeng [Internet]. 2010 [cited 2020 Jan 27];n/a-n/a. Available from: http://doi.wiley.com/10.1002/bit.22647

207. Sugimoto H, Mundel TM, Kieran MW, Kalluri R. Identification of fibroblast heterogeneity in the tumor microenvironment. Cancer Biol Ther. 2006;5(12):1640–6.

208. Camelliti P, Green CR, LeGrice I, Kohl P. Fibroblast Network in Rabbit Sinoatrial Node: Structural and Functional Identification of Homogeneous and Heterogeneous Cell Coupling. Circ Res [Internet]. 2004 Apr 2 [cited 2020 Feb 26];94(6):828–35. Available from: https://www.ahajournals.org/doi/10.1161/01.RES.0000122382.19400.14

209. Kim HD. Expression of intermediate filament desmin and vimentin in the human fetal heart. Anatomical Record. 1996;246(2):271–8.

210. Zuppinger C, Gibbons G, Dutta-Passecker P, Segiser A, Most H, Suter TM. Characterization of cytoskeleton features and maturation status of cultured human iPSC-derived cardiomyocytes. [cited 2020 Jan 27]; Available from: http://rsb.

211. Wang R, Li Q, Tang DD. Role of vimentin in smooth muscle force development. Am J Physiol Cell Physiol [Internet]. 2006 [cited 2022 May 27];291(3):C483. Available from: /pmc/articles/PMC1538637/

212. Baranyi U, Winter B, Gugerell A, Hegedus B, Brostjan C, Laufer G, et al. Primary Human Fibroblasts in Culture Switch to a Myofibroblast-Like Phenotype Independently of TGF Beta. 2019 [cited 2020 Feb 11]; Available from: www.mdpi.com/journal/cells

213. Rege TA, Hagood JS. Thy-1 as a regulator of cell-cell and cell-matrix interactions in axon regeneration, apoptosis, adhesion, migration, cancer, and fibrosis. The FASEB Journal. 2006;20:1045–54.

214. Doppler SA, Carvalho C, Lahm H, Deutsch MA, Dreßen M, Puluca N, et al. Cardiac fibroblasts: More than mechanical support. J Thorac Dis. 2017;9(Suppl 1):S36–51.

215. Ieda M, Fu JD, Delgado-Olguin P, Vedantham V, Hayashi Y, Bruneau BG, et al. Direct Reprogramming of Fibroblasts into Functional Cardiomyocytes by Defined Factors. Cell. 2010;142:375–86.

216. Franzén O, Gan LM, Björkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database [Internet]. 2019 Jan 1 [cited 2022 Jul 1];2019(1):46. Available from: https://academic.oup.com/database/article/doi/10.1093/database/baz046/5427041

217. Badia-i-Mompel P, elez Santiago J v, Braunger J, Geiss C, Dimitrov D, Mü ller-Dott S, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. Bioinformatics Advances [Internet]. 2022 Jan 10 [cited 2022 Jul 1];2(1). Available from: https://academic.oup.com/bioinformaticsadvances/article/2/1/vbac016/6544613

218. Melzer M, Beier D, Young PP, Saraswati S. Isolation and Characterization of Adult Cardiac Fibroblasts and Myofibroblasts. J Vis Exp [Internet]. 2020 Mar 3 [cited 2022 Sep 15];2020(157). Available from: /pmc/articles/PMC7325628/

219. Holdsworth-Carson SJ, Zaitseva M, Vollenhoven BJ, Rogers PAW. Clonality of smooth muscle and fibroblast cell populations isolated from human fibroid and myometrial tissues. Mol Hum Reprod [Internet]. 2014 Mar 1 [cited 2022 Sep 12];20(3):250–9. Available from: https://academic.oup.com/molehr/article/20/3/250/1114116

220. Clément M, Chappell J, Raffort J, Lareyre F, Vandestienne M, Taylor AL, et al. Vascular Smooth Muscle Cell Plasticity and Autophagy in Dissecting Aortic Aneurysms. Arterioscler Thromb Vasc Biol [Internet]. 2019 Jun 1 [cited 2022 Sep 12];39(6):1149–59. Available from: https://www.ahajournals.org/doi/abs/10.1161/ATVBAHA.118.311727

221. Tarbit E, Singh I, Peart JN, Rose'meyer RB. Biomarkers for the identification of cardiac fibroblast and myofibroblast cells. 2018 [cited 2020 Jan 27]; Available from: https://doi.org/10.1007/s10741-018-9720-1

222. Gago-Lopez N, Awaji O, Zhang Y, Ko C, Nsair A, Liem D, et al. THY-1 receptor expression differentiates cardiosphere-derived cells with divergent cardiogenic differentiation potential. Stem Cell Reports [Internet]. 2014;2(5):576–91. Available from: http://dx.doi.org/10.1016/j.stemcr.2014.03.003

223. van Vliet P, Smits AM, de Boer TP, Korfage TH, Metz CHG, Roccio M, et al. Foetal and adult cardiomyocyte progenitor cells have different developmental potential. J Cell Mol Med [Internet]. 2010 Apr [cited 2022 Sep 13];14(4):861. Available from: /pmc/articles/PMC3823117/

224. LaFramboise WA, Scalise D, Stoodley P, Graner SR, Guthrie RD, Magovern JA, et al. Cardiac fibroblasts influence cardiomyocyte phenotype in vitro. Am J Physiol Cell Physiol [Internet]. 2007 May [cited 2022 Jun 6];292(5):1799–808. Available from: https://journals.physiology.org/doi/full/10.1152/ajpcell.00166.2006

225. Guerrero-Juarez CF, Dedhia PH, Jin S, Ruiz-Vega R, Ma D, Liu Y, et al. Single-cell analysis reveals fibroblast heterogeneity and myeloid-derived adipocyte progenitors in murine skin wounds. Nat Commun [Internet]. 2019 [cited 2022 Jun 6];10(650). Available from: https://doi.org/10.1038/s41467-018-08247-x

226. Baranyi U, Winter B, Gugerell A, Hegedus B, Brostjan C, Laufer G, et al. Primary Human Fibroblasts in Culture Switch to a Myofibroblast-Like Phenotype Independently of TGF Beta. Cells [Internet]. 2019 Jul 13 [cited 2022 Jun 6];8(7):721. Available from: /pmc/articles/PMC6678602/

227. Guerrero-Juarez CF, Dedhia PH, Jin S, Ruiz-Vega R, Ma D, Liu Y, et al. Single-cell analysis reveals fibroblast heterogeneity and myeloid-derived adipocyte progenitors in murine skin wounds. Nature Communications 2019 10:1 [Internet]. 2019 Feb 8 [cited 2022 Sep 25];10(1):1–17. Available from: https://www.nature.com/articles/s41467-018-08247-x

228. Forte E, Ramialison M, Nim HT, Mara M, Li JY, Cohn R, et al. Adult mouse fibroblasts retain organ-specific transcriptomic identity. Elife. 2022 Mar 1;11.

229. Kues WA, Anger M, Carnwath JW, Paul D, Motlik J, Niemann H. Cell Cycle Synchronization of Porcine Fetal Fibroblasts: Effects of Serum Deprivation and Reversible Cell Cycle Inhibitors. Biol Reprod [Internet]. 2000 [cited 2023 Jan 17];62:412–9. Available from: http://www.biolreprod.org

230. Trevino V, Falciani F, Barrera-Saldaña HA. DNA microarrays: A powerful genomic tool for biomedical and clinical research. Molecular Medicine [Internet]. 2007 Sep 1 [cited 2022 Aug 5];13(9–10):527–41. Available from: https://molmed.biomedcentral.com/articles/10.2119/2006-00107.Trevino

231. Zhong Wang, Mark Gerstein, Michael Snyder. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009 Jan;10:57–63.

232. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One [Internet]. 2014 Jan 16 [cited 2022 Aug 5];9(1). Available from: https://pubmed.ncbi.nlm.nih.gov/24454679/

233. Fu X, Fu N, Guo S, Yan Z, Xu Y, Hu H, et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. BMC Genomics [Internet]. 2009 Apr 16 [cited 2022 Aug 5];10(1):1–9. Available from: https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-10-161

234. John H Malone, Brian Oliver. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol [Internet]. 2011 [cited 2022 Aug 5];9(34). Available from: http://www.biomedcentral.com/1741-7007/9/34

235. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nature Methods 2009 6:5 [Internet]. 2009 Apr 6 [cited 2022 Aug 5];6(5):377–82. Available from: https://www.nature.com/articles/nmeth.1315

236. Macosko EZ, Basu A, Regev A, Mccarroll SA, Macosko EZ, Basu A, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell [Internet]. 2015;161(5):1202–14. Available from: http://dx.doi.org/10.1016/j.cell.2015.05.002

237. Garside VC, Cullum R, Alder O, Lu DY, Werff R vander, Bilenky M, et al. SOX9 modulates the expression of key transcription factors required for heart valve development. 2015;

238. Akiyama H, Chaboissier MC, Behringer RR, Rowitch DH, Schedl A, Epstein JA, et al. Essential role of Sox9 in the pathway that controls formation of cardiac valves and septa [Internet]. Vol. 101, PNAS. 2004 [cited 2020 Feb 28]. Available from: www.pnas.orgcgidoi10.1073pnas.0401711101

239. Yuasa S, Fukuda K, Edelberg -Bristol-Myers Squibb Co J. Multiple roles for BMP signaling in cardiac development. Drug Discov Today Ther Strateg [Internet]. 2008 Dec [cited 2022 Apr 6];5(4):209–14. Available from: www.drugdiscoverytoday.com

240. Lints TJ, Parsons LM, Hartley L, Lyons I, Harvey RP. Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. Development. 1993;119:419–31.

241. Zhang L, Nomura-Kitabayashi A, Sultana N, Cai W, Cai X, Moon AM, et al. Mesodermal Nkx2.5 is necessary and sufficient for early second heart field development. Dev Biol. 2014 Jun 1;390(1):68–79.

242. Criem N, Zwijsen A. The epicardium obscures interpretations on endothelial-to-mesenchymal transition in the mouse atrioventricular canal explant assay. Scientific

Reports 2018 8:1 [Internet]. 2018 Mar 16 [cited 2022 Sep 15];8(1):1–9. Available from: https://www.nature.com/articles/s41598-018-22971-w

243. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nature Publishing Group [Internet]. 2011 [cited 2020 Feb 19]; Available from: www.nature.com/reviews/genetics

244. Olsen TK, Baryawno N. Introduction to Single-Cell RNA Sequencing. Curr Protoc Mol Biol. 2018;122(1):1–14.

245. Zeisberg EM, Kalluri R. Origins of Cardiac Fibroblasts. Circ Res [Internet]. 2010 Nov 26 [cited 2022 Jul 27];107(11):1304–12. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCRESAHA.110.231910

246. Patra C, Diehl F, Ferrazzi F, van Amerongen MJ, Novoyatleva T, Schaefer L, et al. Nephronectin regulates atrioventricular canal differentiation via Bmp4-Has2 signaling in zebrafish. Development. 2011;138(20):4499–509.

247. Winnier G, Blessing M, Labosky PA, Hogan BLM. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. Genes Dev. 1995;9(17):2105–16.

248. Jackson DA, Symonst RH, Berg - P, Lobban DP, Kaiser AD. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. Proc Natl Acad Sci U S A [Internet]. 1972 [cited 2022 Apr 21];69(10):2904–9. Available from: https://www.pnas.org

249. Gupta RM, Musunuru K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. Journal of Clinical Investigation. 2014;124(10):4154–61.

250. Gaj T. ZFN, TALEN and CRISPR/Cas based methods for genome engineering. 2013. 2014;31(7):397–405.

251. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science (1979). 2014;346(6213).

252. Miller J, McLachlan AD, Klug A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. 1985;4(6):1609–14.

253. Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A [Internet]. 1996 Feb 6 [cited 2022 Apr 21];93(3):1156–60. Available from: https://www.pnas.org

254. Joung JK, Sander JD. TALENs: A widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol [Internet]. 2013;14(1):49–55. Available from: http://dx.doi.org/10.1038/nrm3486

255. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science (1979) [Internet]. 2007 Mar 23 [cited 2022 Apr 25];315(5819):1709–12. Available from: https://www.science.org/doi/abs/10.1126/science.1138140

256. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A [Internet]. 2012 Sep 25 [cited 2022 Apr 25];109(39). Available from: www.pnas.org/cgi/doi/10.1073/pnas.1208507109

257. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (1979) [Internet]. 2012 Aug 17 [cited 2022 Apr 25];337(6096):816–21. Available from: https://www.science.org

258. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science (1979). 2013 Feb 15;339(6121):819–23.

259. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. Elife. 2013 Jan 29;2013(2).

260. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science (1979). 2013 Feb 15;339(6121):823–6.

261. Chang AY. Genome engineering with CRISPR/Cas9, ZFNs, and TALENs. In: CRISPR Genome Surgery in Stem Cells and Disease Tissues [Internet]. INC; 2022. p. 39–45. Available from: http://dx.doi.org/10.1016/B978-0-12-817876-8.00009-7

262. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell. 2013 Feb 28;152(5):1173–83.

263. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013 Jul 18;154(2):442.

264. Cevher E, Sezer AD, Şefik Çağlar E. Gene Delivery Systems: Recent Progress in Viral and Non-Viral Therapy. In: Sezer AD, editor. Recent Advances in Novel Drug Carrier Systems. InTech; 2012.

265. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, et al. In vivo genome editing using Staphylococcus aureus Cas9.

266. Jansig E, Geissler S, Rieckmann V, Kuenemund A, Hietel B, Schenk M, et al. Viromers as carriers for mRNA-mediated expression of therapeutic molecules under inflammatory conditions. Scientific Reports 2020 10:1 [Internet]. 2020 Sep 15 [cited 2022 May 6];10(1):1–13. Available from: https://www.nature.com/articles/s41598-020-72004-8

267. Kantor B, Bailey RM, Wimberly K, Kalburgi SN, Gray SJ. Methods for Gene Transfer to the Central Nervous System. Adv Genet. 2014 Jan 1;87:125–97.

268. Xu X, Tan X, Tampe B, Wilhelmi T, Hulshoff MS, Saito S, et al. High-fidelity CRISPR/Cas9- based gene-specific hydroxymethylation rescues gene expression and attenuates renal fibrosis. Nat Commun. 2018 Dec 1;9(1):1–15.

269. Li L, Hu S, Chen X. Non-viral delivery systems for CRISPR/Cas9-based genome editing: Challenges and opportunities. Vol. 171, Biomaterials. Elsevier Ltd; 2018. p. 207–18.

270. Kreiss P, Cameron B, Rangara R, Mailhe P, Aguerre-Charriol O, Airiau M, et al. Plasmid DNA size does not affect the physicochemical properties of lipoplexes but modulates

gene transfer efficiency. Nucleic Acids Res [Internet]. 1999 [cited 2022 May 7];27(19):3792–8. Available from: https://academic.oup.com/nar/article/27/19/3792/1054764

271. Hornstein BD, Roman D, Aré valo-Soliz LM, Engevik MA, Zechiedrich L. Effects of Circular DNA Length on Transfection Efficiency by Electroporation into HeLa Cells. 2016;

272. Thomas CE, Ehrhardt A, Kay MA. Progress and problems with the use of viral vectors for gene therapy. Nat Rev Genet. 2003;4(5):346–58.

273. Xu CL, Ruan MZC, Mahajan VB, Tsang SH. Viral delivery systems for crispr. Viruses. 2019;11(1):1–12.

274. Mashal RD, Koontz J, Sklar J. Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. 1995 [cited 2022 Apr 13]; Available from: http://www.nature.com/naturegenetics

275. Guschin DY, Waite AJ, Katibah GE, Miller JC, Holmes MC, Rebar EJ. A Rapid and General Assay for Monitoring Endogenous Gene Modification. Methods in Molecular Biology [Internet]. 2010 [cited 2022 Apr 13];649:247–56. Available from: https://link.springer.com/protocol/10.1007/978-1-60761-753-2_15

276. Yip BH. Recent Advances in CRISPR/Cas9 Delivery Strategies. Biomolecules [Internet]. 2020 [cited 2022 May 10];10. Available from: https://clinicaltrials.gov/

277. Kucharski M, Mrowiec P, Ocło E. Current standards and pitfalls associated with the transfection of primary fibroblast cells. Biotechnol Prog [Internet]. 2021;37(4). Available from: https://doi.org/10.1002/btpr.3152

278. Liu G, Liu K, Wei H, Li L, Zhang S. Generation of porcine fetal fibroblasts expressing the tetracycline-inducible Cas9 gene by somatic cell nuclear transfer. Mol Med Rep. 2016 Sep 1;14(3):2527–33.

279. Black JB, Adler AF, Wang HG, D'Ippolito AM, Hutchinson HA, Reddy TE, et al. Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. Cell Stem Cell. 2016;

280. Hatcher CJ, Goldstein MM, Mah CS, Delia CS, Basson CT. Identification and Localization of TBX5 Transcription Factor During Human Cardiac Morphogenesis. 2000;

281. Bruneau BG, Nemer G, Schmitt JP, Charron F, Robitaille L, Caron S, et al. A murine model of Holt-Oram syndrome defines roles of the T-Box transcription factor Tbx5 in cardiogenesis and disease. Cell [Internet]. 2001 Sep 21 [cited 2022 Aug 12];106(6):709–21. Available from: http://www.cell.com/article/S0092867401004937/fulltext

282. Ghosh TK, Packham EA, Bonser AJ, Robinson TE, Cross SJ, Brook JD. Characterization of the TBX5 binding site and analysis of mutations that cause Holt–Oram syndrome. Hum Mol Genet [Internet]. 2001 Sep 1 [cited 2022 Aug 12];10(18):1983–94. Available from: https://academic.oup.com/hmg/article/10/18/1983/2901498

283. Kathiriya IS, Nora EP, Bruneau BG. Investigating the Transcriptional Control of Cardiovascular Development. Circ Res [Internet]. 2015 Feb 13 [cited 2022 Aug 12];116(4):700–14. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCRESAHA.116.302832

284. Fogarty NME, McCarthy A, Snijders KE, Powell BE, Kubikova N, Blakeley P, et al. Genome editing reveals a role for OCT4 in human embryogenesis. Nature 2017 550:7674 [Internet]. 2017 Sep 20 [cited 2022 Aug 12];550(7674):67–73. Available from: https://www.nature.com/articles/nature24033

285. Anderson DJ, Kaplan DI, Bell KM, Koutsis K, Haynes JM, Mills RJ, et al. NKX2-5 regulates human cardiomyogenesis via a HEY2 dependent transcriptional network. Nature Communications 2018 9:1 [Internet]. 2018 Apr 10 [cited 2022 Aug 12];9(1):1–13. Available from: https://www.nature.com/articles/s41467-018-03714-x

286. Wu S pin, Cheng CM, Lanz RB, Wang T, Respress JL, Ather S, et al. Atrial Identity Is Determined by a COUP-TFII Regulatory Network. Dev Cell. 2013 May 28;25(4):417–26.

287. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities.

Genome Res [Internet]. 2019 Aug 1 [cited 2022 Jul 4];29(8):1363–75. Available from: https://genome.cshlp.org/content/29/8/1363.full

288. Benson DW, Silberbach GM, Kavanaugh-McHugh A, Cottrill C, Zhang Y, Riggs S, et al. Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. Journal of Clinical Investigation [Internet]. 1999 Dec 12 [cited 2022 Jul 12];104(11):1567. Available from: /pmc/articles/PMC409866/

289. Paul MH, Harvey RP, Wegner M, Sock E. Cardiac outflow tract development relies on the complex function of Sox4 and Sox11 in multiple cell types. Cell Mol Life Sci [Internet]. 2014 [cited 2022 Jul 12];71(15):2931–45. Available from: https://pubmed.ncbi.nlm.nih.gov/24310815/

290. Petit FG, Salas R, Tsai MJ, Tsai SY. The regulation of COUP-TFII gene expression by Ets-1 is enhanced by the steroid receptor co-activators. Mech Ageing Dev. 2004 Oct 1;125(10–11):719–32.

291. Betancur P, Bronner-Fraser M, Sauka-Spengler T. Assembling neural crest regulatory circuits into a gene regulatory network. Annu Rev Cell Dev Biol. 2010 Nov 10;26:581–603.

292. Nelms B, Labosky P. bHLH Proteins. In: Transcriptional Control of Neural Crest Development [Internet]. San Rafael (CA): Morgan & Claypool Life Sciences; 2010 [cited 2022 Jul 12]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK53143/

293. Maitra M, Koenig SN, Srivastava D, Garg V. Identification of GATA6 Sequence Variants in Patients With Congenital Heart Defects. Pediatric Research 2010 68:4 [Internet]. 2010 Oct [cited 2022 Sep 2];68(4):281–5. Available from: https://www.nature.com/articles/pr2010175

294. Sharma A, Wasson LK, Willcox JAL, Morton SU, Gorham JM, Delaughter DM, et al. GATA6 mutations in hiPSCs inform mechanisms for maldevelopment of the heart, pancreas, and diaphragm. Elife. 2020 Oct 1;9:1–28.

295. Yu L, Bennett JT, Wynn J, Carvill GL, Cheung YH, Shen Y, et al. Whole exome sequencing identifies de novo mutations in GATA6 associated with congenital

diaphragmatic hernia. J Med Genet [Internet]. 2014 Mar 1 [cited 2022 Sep 2];51(3):197–202. Available from: https://jmg.bmj.com/content/51/3/197

296. Lin CJ, Lin CY, Chen CH, Zhou B, Chang CP. Partitioning the heart: mechanisms of cardiac septation and valve development. Development [Internet]. 2012;139(18):3277–99. Available from: http://dev.biologists.org/cgi/doi/10.1242/dev.063495

297. Stallmeyer B, Fenge H, Nowak-Göttl U, Schulze-Bahr E. Mutational spectrum in the cardiac transcription factor gene NKX2.5 (CSX) associated with congenital heart disease. Clin Genet. 2010;78(6):533–40.

298. McElhinney DB, Geiger E, Blinder J, Benson DW, Goldmuntz E. NKX2.5 mutations in patients with congenital heart disease. J Am Coll Cardiol. 2003 Nov 5;42(9):1650–5.

299. Estruch SB, Graham SA, Quevedo M, Vino A, Dekkers DHW, Deriziotis P, et al. Proteomic analysis of FOXP proteins reveals interactions between cortical transcription factors associated with neurodevelopmental disorders. Hum Mol Genet [Internet]. 2018 Apr 1 [cited 2022 Sep 2];27(7):1212–27. Available from: https://academic.oup.com/hmg/article/27/7/1212/4819278

300. Shu W, Yang H, Zhang L, Lu MM, Morrisey EE. Characterization of a New Subfamily of Winged-helix/Forkhead (Fox) Genes That Are Expressed in the Lung and Act as Transcriptional Repressors. Journal of Biological Chemistry. 2001 Jul 20;276(29):27488–97.

301. Gambetta K, Al-Ahdab MK, Ilbawi MN, Hassaniya N, Gupta M. Transcription repression and blocks in cell cycle progression in hypoplastic left heart syndrome. Am J Physiol Heart Circ Physiol [Internet]. 2008 May [cited 2022 Sep 2];294(5):2268–75. Available from: https://journals.physiology.org/doi/10.1152/ajpheart.91494.2007

302. Petit FG, Salas R, Tsai MJ, Tsai SY. The regulation of COUP-TFII gene expression by Ets-1 is enhanced by the steroid receptor co-activators. Mech Ageing Dev. 2004 Oct 1;125(10–11):719–32.

303. Qian B, Mo R, Da M, Peng W, Hu Y, Mo X. Common Variations in BMP4 Confer Genetic Susceptibility to Sporadic Congenital Heart Disease in a Han Chinese Population. Pediatr Cardiol [Internet]. 2014 Dec 1 [cited 2022 Sep 3];35(8):1442–7. Available from: https://link.springer.com/article/10.1007/s00246-014-0951-1

304. Liu W, Selever J, Wang D, Lu MF, Moses KA, Schwartz RJ, et al. Bmp4 signaling is required for outflow-tract septation and branchial-arch artery remodeling [Internet]. 2004 [cited 2020 Feb 28]. Available from: www.pnas.orgcgidoi10.1073pnas.0308466101

305. McCulley DJ, Kang JO, Martin JF, Black BL. BMP4 is required in the anterior heart field and its derivatives for endocardial cushion remodeling, outflow tract septation, and semilunar valve development. Developmental Dynamics [Internet]. 2008 Nov 1 [cited 2022 Sep 3];237(11):3200–9. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/dvdy.21743

306. Zhang S, Shen J, Li D, Cheng Y. Strategies in the delivery of Cas9 ribonucleoprotein for CRISPR/Cas9 genome editing. Theranostics [Internet]. 2021 [cited 2022 Sep 28];11(2):614. Available from: /pmc/articles/PMC7738854/

307. Tallquist MD. Cardiac Fibroblast Diversity. https://doi.org/101146/annurev-physiol-021119-034527 [Internet]. 2020 Feb 10 [cited 2022 Sep 26];82:63–78. Available from: https://www.annualreviews.org/doi/abs/10.1146/annurev-physiol-021119-034527

308. Tallquist MD. Developmental Pathways of Cardiac Fibroblasts. Cold Spring Harb Perspect Biol [Internet]. 2020 Apr 1 [cited 2022 Jul 27];12(4):a037184. Available from: http://cshperspectives.cshlp.org/content/12/4/a037184.full

309. Muhl L, Genové G, Leptidis S, Liu J, He L, Mocci G, et al. Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. Nature Communications 2020 11:1 [Internet]. 2020 Aug 7 [cited 2022 Sep 26];11(1):1–18. Available from: https://www.nature.com/articles/s41467-020-17740-1

310. Baranyi U, Winter B, Gugerell A, Hegedus B, Brostjan C, Laufer G, et al. Primary Human Fibroblasts in Culture Switch to a Myofibroblast-Like Phenotype

Independently of TGF Beta. Cells [Internet]. 2019 Jul 13 [cited 2020 Feb 11];8(7):721. Available from: /pmc/articles/PMC6678602/

311. Landry NM, Rattan SG, Dixon IMC. An Improved Method of Maintaining Primary Murine Cardiac Fibroblasts in Two-Dimensional Cell Culture. Scientific Reports 2019 9:1 [Internet]. 2019 Sep 9 [cited 2022 Sep 14];9(1):1–13. Available from: https://www.nature.com/articles/s41598-019-49285-9

312. Campisi J. Cellular senescence. Chromosomal Instability and Aging: Basic Science and Clinical Implications [Internet]. 2003 Jan 1 [cited 2022 Sep 28];102(4):29–50. Available from: http://www.cell.com/article/S0092867400000465/fulltext

313. Choi AMK, Olsen DR, Cook KG, Deamond SF, Uitto J, Bruce SA. Differential extracellular matrix gene expression by fibroblasts during their proliferative life span in vitro and at senescence. J Cell Physiol [Internet]. 1992 Apr 1 [cited 2022 Sep 28];151(1):147–55. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/jcp.1041510119

314. Akita H, Ito R, Kamiya H, Kogure K, Harashima H. Cell cycle dependent transcription, a determinant factor of heterogeneity in cationic lipid-mediated transgene expression. J Gene Med [Internet]. 2007 Mar 1 [cited 2022 Sep 28];9(3):197–207. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/jgm.1010

315. Lajiness JD, Conway SJ. The Dynamic Role of Cardiac Fibroblasts in Development and Disease. J Cardiovasc Transl Res. 2012;5:739–48.

316. Guarino N, Shima H, Puri P. The hypoplastic heart in congenital diaphragmatic hernia: reduced expression of basic fibroblast growth factor and platelet-derived growth factor. Pediatr Surg Int. 2000;16:243–6.

317. Grossfeld P. ETS1 and HLHS: Implications for the Role of the Endocardium. J Cardiovasc Dev Dis [Internet]. 2022 Jul 1 [cited 2022 Sep 25];9(7). Available from: /pmc/articles/PMC9319889/

318. Xia B, Hou L, Kang H, Chang W, Liu Y, Zhang Y, et al. NR2F2 plays a major role in insulin-induced epithelial-mesenchymal transition in breast cancer cells. BMC

Cancer [Internet]. 2020 Jul 6 [cited 2022 Sep 26];20(1):1–12. Available from: https://bmccancer.biomedcentral.com/articles/10.1186/s12885-020-07107-6

319. Peng Q, Shan D, Cui K, Li K, Zhu B, Wu H, et al. The Role of Endothelial-to-Mesenchymal Transition in Cardiovascular Disease. Cells 2022, Vol 11, Page 1834 [Internet]. 2022 Jun 3 [cited 2022 Sep 26];11(11):1834. Available from: https://www.mdpi.com/2073-4409/11/11/1834/htm

320. Kovacic JC, Dimmeler S, Harvey RP, Finkel T, Aikawa E, Krenning G, et al. Endothelial to Mesenchymal Transition in Cardiovascular Disease: JACC State-of-the-Art Review. J Am Coll Cardiol. 2019 Jan 22;73(2):190–209.

321. Anbara T, Sharifi M, Aboutaleb N. Endothelial to Mesenchymal Transition in the Cardiogenesis and Cardiovascular Diseases. Curr Cardiol Rev [Internet]. 2020 Aug 8 [cited 2022 Sep 28];16(4):306. Available from: /pmc/articles/PMC7903503/