From Department of Cell and Molecular Biology
Karolinska Institutet, Stockholm, Sweden

A GENOME-WIDE SCREEN FOR ESSENTIAL GENES THAT CONTROLS THE FORMATION OF HUMAN HEART PROGENITORS

Jiejia Xu

Karolinska
Institutet

Stockholm 2018

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by AJ E-print AB
© Jiejia Xu, 2018
ISBN 978-91-7831-234-4

A Genome-Wide Screen For Essential Genes That Controls The Formation Of Human Heart Progenitors

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Jiejia Xu

Principal Supervisor:
Professor Kenneth Chien
Karolinska Institutet
Department of Cell and Molecular Biology
Department of Medicine

Co-supervisor(s):
Professor Urban Lendahl
Karolinska Institutet
Department of Cell and Molecular Biology

Assistant Professor Xiaojun Lian
The Pennsylvania State University
Department of Biomedical Engineering
Department of Biology
The Huck Institutes of the Life Sciences

Opponent:
Professor Mark Mercola
Stanford University
Department of Medicine
Cardiovascular Institute

Examination Board:
Associate Professor Anna Falk
Karolinska Institutet
Department of Neuroscience

Professor Thomas Eschenhagen
University Medical Center Hamburg-Eppendorf
Department of Experimental Pharmacology and
Toxicology

Associate Professor Lars Jakobsson
Karolinska Institutet
Department of Medical Biochemistry and
Biophysics

ABSTRACT

The heart is a complex organ system composed of multiple types of tissues. These tissues are produced by a diverse set of muscle and non-muscle cells, originated from a few pools of progenitors. During the heart development, these progenitors are able to expand and differentiate in a tightly controlled manner, generating diversified heart cell lineages. The progenies from these progenitors interact with each other and ultimately integrate into distinct heart tissues. The foundation of a healthy and functional heart stems from the state of its progenitor pools. Any errors that occurred during the formation, proliferation, differentiation, and assembly of these progenitors are the potential causes of many congenital heart diseases. To investigate the cellular mechanisms of human heart development and their implications in congenital heart diseases, we face many challenges, two of them are: 1) generation of progenitor cells that can self-assemble into mature cardiac tissue that faithfully resembles native mature adult cardiac tissue; 2) identification of regulators that controls the formation, proliferation and differentiation of these progenitors.

In paper I, we reported the large-scale generation of an enriched pool of human pluripotent stem cells (hPSCs) derived human ventricular progenitors (HVPs). These HVPs can build a function ventricular heart muscle *in vivo* via a cell autonomous pathway, including controlled proliferation followed by normal growth, maturation, and self-assembly. This tissue generation process of HVPs recapitulates one of the earliest and most essential steps of organogenesis. With these properties, HVPs highly likely resembles the progenitors that contribute to the ventricular cardiac muscle tissue during human cardiogenesis. In the study, we also explore the therapeutic potential of HVPs in heart failure. As a resource for further analyzing the genetic and molecular pathways of HVPs, we also documented the transcriptomic transitions of the progenitor formation and subsequent differentiation via sequential RNA-Seq.

With the success of generating HVPs, we next try to identify regulators, specifically, the ones that control the formation of HVPs. **In paper II**, we used CRIPSR-Cas9 system to target β-catenin (encoded by *CTNNB1*), a central component of the canonical WNT signaling pathway. The WNT signaling is a major player in cardiogenesis. By temporal modulating the WNT/β-catenin signaling pathway with small molecules, high differentiation efficiency (>90%) can be achieved. With *CTNNB1* mutated hPSCs, we found that Wnt/β-catenin signaling is neither required for hPSC self-renewal, nor for neuroectoderm formation. However, Wnt/β-catenin signaling is absolutely essential for mesendoderm lineage, including cardiac progenitors and cardiomyocytes. This study pinpoints the β-catenin as the master switch of the human cardiogenesis.

Another set of important signaling pathways in cardiogenesis are the TGFβ superfamily signaling pathways. Due to the complicate interaction between WNT/β-catenin and the TGFβ superfamily signaling pathways, it is difficult to define the roles of TGFβ superfamily signaling pathways from chemical inhibition studies. **In paper III**, we used CRIPSR-Cas9

system to target *SMAD4*, a central component in the whole superfamily. With *SMAD4* mutated hPSCs, we confirmed the dispensable role of *SMAD4* for hPSC self-renewal *in vitro*. Furthermore, we demonstrated the essential requirement of *SMAD4* in the formation of human cardiac mesodermal precursor cell. By transcriptome analysis, we identified that *SMAD4* mutants failed to differentiate into cardiac mesoderm and, after 6 days, switched to neuroectoderm. Primitive steak (PS) genes were expressed in both the wild type and the mutant cells on day 1. And interestingly, on day 1, the only active ligand in the TGFβ superfamily signaling pathways is NODAL, which specifies the pathway in the family as NODAL/SMAD4 pathway. Together, these data suggest that during human mesoderm induction, the WNT/β-catenin is responsible for triggering the expression of PS genes, while NODAL/SMAD4 is responsible for the feedback enhancement for PS gene expression. This study highlights the essential roles of NODAL/SMAD4 signaling pathway in human cardiac mesodermal induction.

In order to unbiasedly uncover the regulators that control the formation of HVPs, **in paper IV**, we developed a genome-wide CRISPR screen based on cardiac differentiation from hPSCs. From the screen output, we compiled a list of 15 candidate genes. After validating 7 of these, we identified *ZIC2* as an essential gene for cardiac progenitor formation. *ZIC2* is known as a master regulator of neurogenesis. hPSCs with *ZIC2* mutated still express pluripotency markers. However, their ability to differentiate into cardiomyocytes has greatly reduced. Transcriptome profiling reveals that they have switched to an alternative mesodermal cell fate. Our results provide a new link between *ZIC2* and human cardiogenesis and document the potential power of genome-wide unbiased CRISPR screens to identify key steps in heart progenitor fate determination during human cardiogenesis with hPSC model systems.

In summary, we have generated HVPs, which can self-assemble into human ventricular muscle tissue and further identified *CTNNB1*, *SMAD4*, and *ZIC2* as the essential regulators that controlled the formation of HVPs.

LIST OF SCIENTIFIC PAPERS

1. Kylie S. Foo[+], Miia L. Lehtinen[+], Chuen Yan Leung[+], Xiaojun Lian, **Jiejia Xu**, Wendy Keung, Lin Geng, Terje RS. Kolstad, Sebastian Thams, Andy On-tik Wong, Nicodemus Wong, Kristine Bylund, Chikai Zhou, Xiaobing He, Shao-Bo Jin, Jonathan Clarke, Urban Lendahl, Ronald A. Li, William E. Louch, Kenneth R. Chien[*] (2018). Human ISL1+ Ventricular Progenitors Self-Assemble into an In Vivo Functional Heart Patch and Preserve Cardiac Function after Infarction.
   Molecular Therapy 6, 1644-1659

2. Xiaojun Lian[*], **Jiejia Xu**, Xiaoping Bao, Lauren N. Randolph (2016). Interrogating Canonical Wnt Signaling Pathway in Human Pluripotent Stem Cell Fate Decisions Using CRISPR-Cas9.
   Cellular and Molecular Bioengineering 9, 325-334.

3. **Jiejia Xu**, Peter J. Gruber, Kenneth R. Chien[*] (2018). SMAD4 Is Essential For Human Cardiac Mesodermal Precursor Cell Formation
   [manuscript | Stem Cells, accepted]

4. **Jiejia Xu**, Chikai Zhou, Kylie S. Foo, Ran Yang, Yao Xiao, Kristine Bylund, Kenneth R. Chien[*] (2018). Genome-Wide CRISPR Screen Identifies *ZIC2* As An Essential Gene That Controls The Fate of Mesodermal Precursors To Human Heart Progenitors.
   [manuscript]

Publications not included in this thesis:

1. Boon-Seng Soh, Shi-Yan Ng, Hao Wu, Kristina Buac, Joo-Hye C. Park, Xiaojun Lian, **Jiejia Xu**, Kylie S. Foo, Ulrika Felldin, Xiaobing He, Massimo Nichane, Henry Yang, Lei Bu, Ronald A. Li, Bing Lim, Kenneth R. Chien[*] (2016). Endothelin-1 Supports Clonal Derivation And Expansion Of Cardiovascular Progenitors Derived From Human Embryonic Stem Cells.
   Nature Communication 7: 10774.

2. Xiaojun Lian, **Jiejia Xu**, Jinsong Li, Kenneth R. Chien[*] (2014). Next-Generation Models of Human Cardiogenesis via Genome Editing. (Review)
   Cold Spring Harbor Perspectives in Medicine 4:a013920

[+] These authors contributed equally

[*] Correspondent author

CONTENTS

1 Introduction 1
2 Congenital Heart Disease 3
2.1 Syndromic Congenital Heart Disease 4
2.2 Non-syndromic Congenital Heart Disease 4
2.3 Others mechanism 6
3 Cardiogenesis in Murine 7
3.1 Heart development with cardiac progenitors 7
3.2 The gene language of heart development 8
3.3 Markers and Lineages 11
3.4 Differences between mice and humans 13
4 Cardiomyocyte From hPSC 15
4.1 Cardiomyocyte Differentiation 15
4.2 Cardiomyocyte Maturation 17
4.3 Cardiomyocyte Subtypes 18
4.4 Disease Modeling with human PSC-CMs 19
4.5 Transplantations 20
5 Genomic Technology 23
5.1 Massive Parallel Sequencing 23
5.2 Transgenic tools for hPSC 24
5.3 Genome Editing in hPSC 25
5.4 CRISPR Screen 27
6 Aims 29
7 Result Summary 31
7.1 Paper I: HVP 31
7.2 Paper II: β-catenin 33
7.3 Paper III: SMAD4 34
7.4 Paper IV: Screen-ZIC2 36
8 Discussion 39
8.1 The Unique Engraftment Property of HVP 39
8.2 A Peek into The Molecular Complexity of Cardiogenesis 40
9 Conclusion 43
Acknowledgements 45
References 47

LIST OF ABBREVIATIONS

hESC	human embryonic stem cell
hiPSC	human induce pluripotent stem cell
hPSC	human pluripotent stem cell, hESC or hiPSC
HVP	human ventricular progenitor
CM	cardiomyocyte
hPSC-CM	hPSC derived cardiomyocyte
NSG	next generation sequencing
MI	moycardiac infaction
GiWi	GSK3 inhibition and Wnt inhibition, a cardiomyocyte differentiation protocol
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
PS	primitive streak
NuRD	Nucleosome Remodeling and Deacetylase
DSB	double strand break
NHEJ	Non-Homologous End Joining
ZFN	Zinc-finger nuclearase
TALE	Transcription activator-like effector
HDR	Homology Directed Repair
ORF	Open Reading Frame
PAM	protospacer adjacent motif
sgRNA	single chimeric guide RNA
RNAi	RNA interference
CHD	Congenital heart disease
SHF	Second heart field
FHF	First heart field
SAN	Sinoatrial node

1 INTRODUCTION

The heart is the very first organ to form during embryogenesis. This complex organ composes of a highly diverse set of cells. Understanding the molecular mechanisms that govern the formation, migration, proliferation, differentiation and assembly of these cells into different tissues of the heart is the central question in heart development. Since laboratory mice rarely have heart problems, the genetic clues usually first came from the congenital heart disease patients. About 1% of the live births carry heart defects, which can affect any part of the heart. By studying congenital heart diseases (CHDs), critical genes have been identified for heart development.

Once a CHD gene is identified, the mice model becomes very useful. By using transgenic mice model, the mechanism of the CHD gene can be explored in depth. Besides the loss-of-function approach, genetic lineage analysis can be performed in mice model. Lineage is one of the fundamental concepts in development, tracing the origins, namely progenitor pools, of every cells in a given heart tissue. Any genetic mutations occurred in these progenitor pools might cause defects in the progenitor function and further result in structural or functional defects in the heart. Insights from studying mice cardiogenesis have brought us from the phenotype-genotype association study to the age of cardiac progenitor biology.

Despite of the prosperity of the transgenic mice model, the complexity of a human heart verses a mouse heart is not at the same scale. The mice model study might not truly reflect the human conditions. Advances in genome editing in human pluripotent stem cell (hPSC) and cardiac differentiation make it possible to use hPSC to model the congenital heart disease and study the human cardiogenesis via genome editing. Compared to the mice model, hPSC models have one obvious advantage: they have the human genome.

Here, we will first review and organize the knowledge of heart development from disease model, mice model and hPSC model. In addition, we also go over some genomic technologies used in our studies. Then, we summarized the present investigations included in this thesis.

2 CONGENITAL HEART DISEASE

Heart development requires a complex interplay of the cells in an embryo, including migration, proliferation, differentiation, and assembly. Any perturbation in these programs can cause defects in heart development or congenital heart disease (CHD). CHD is the most common human congenital defect, and the leading cause of death in infants. Worldwide, 1.35 million infants are born with CHD each year (van der Linde et al. 2011). CHD is also identified in 10% of stillbirths (Fahed et al. 2013). Different countries and continents have different prevalence of CHD. In North America, the prevalence is 6.9 per 1000 live births; in Europe, the incidence is 8.2 per 1,000 live births; in Asia, the prevalence is 9.3 per 1000 live births; in Africa, the incidence is 1.9 per 1,000 live births (van der Linde et al. 2011).

CHD arises during cardiogenesis, manifest itself with various structural and functional defects in the heart. CHD affects most parts of the heart, including valves, septa, inflow tracts, and outflow tracts. Here are some common heart defects: aortic regurgitation (AR), aortic stenosis (AS), atrial septal defect (ASD), atrioventricular septal defect (AVSD), bicuspid aortic valve (BAV), coarctation of the aorta (CoA), double outlet right ventricle (DORV), hypoplastic left heart syndrome (HLHS), hypoplastic right heart syndrome (HRHS), interrupted aortic arch (IAA), mitral regurgitation (MR), mitral valve prolapse (MVP), pulmonary atresia (PA), patent ductus arteriosus (PDA), pulmonary stenosis (PS), single ventricle (SV), tricuspid atresia (TA), total anomalous pulmonary venous return (TAPVR), partial anomalous pulmonary venous return (PAPVR), transposition of great arteries (TGA), tetralogy of Fallot (TOF), ventricular septal defect (VSD). Three broad categories are used to classified CHDs: cyanotic heart disease, left-sided obstruction defects and septation defects (Bruneau 2008). Other types of congenital defects, such as BAV and PDA, do not fit neatly into the three main categories.

The causes of CHD can be genetic or/and environmental. The majority of CHD is thought to be associated with gene mutations. However, it turned out to be much more complicated to understand the molecular mechanism how these gene mutations cause CHD. The complicated picture begins with that not all CHD manifests Mendelian inheritance. Secondly, it is difficult to explain with genetic factors that discordant clinical phenotypes exist within the same family. From the environmental factors aspect, non-genetic causes of CHD include environmental teratogens, maternal exposures, and infectious agents (Fahed et al. 2013).

Simple genotype-phenotype correlations are hard to establish in CHDs. Take familial CHD for example. The same single gene mutation can cause different cardiac defects, even within the same family. On the other hand, different types of structural malformations can be associated with similar gene mutations. CHD arises from the abnormal heart development during embryogenesis, a good knowledge of the genetic and cellular mechanism of heart development is important for explaining the causes of CHDs.

2.1 SYNDROMIC CONGENITAL HEART DISEASE

Cardiac malformations are the most common malformations in congenital syndromes. Congenital heart disease is a primary manifestation is some congenital syndromes. CHD occurs in approximately 40% to 50% in Down Syndrome (Antonarakis et al. 2004), 20% to 50% in Turner syndrome (Phillip et al. 2009), and in almost all in trisomy 18 and 13 (Pont et al. 2006).

DiGeorge Syndrome (DGS) is inherited in an autosomal dominant fashion, but the overwhelming majority of deletions occur *de novo* (Emanuel 2008). The common cause of DGS is a 1.5 to 3.0-Mb deletion on chromosome 22q11 (Scambler 2000), which results in haploinsufficiency of *TBX1* (Jerome and Papaioannou 2001) and other genes. The most common diseases are outflow tract anomalies, which include TOF, TOF with PA, truncus arteriosus, and IAA (Momma 2010).

Holt-Oram Syndrome (HOS) manifests as congenital heart disease and upper limb dysplasia. HOS is an autosomal dominant disease and caused by mutations in *TBX5* gene (Basson et al. 1997; Li et al. 1997b). 85% of HOS patients have congenital heart malformations (Basson et al. 1994), including secundum ASD, VSD, PDA, and conduction system abnormalities.

Alagille Syndrome is inherited in an autosomal dominant fashion and is commonly caused by mutations in *JAG1* gene (Li et al. 1997a; Oda et al. 1997). More than 90% of individuals with a *JAG1* mutation or Alagille Syndrome have cardiovascular anomalies, with branch PA stenosis the most common abnormality and other common defects include TOF, PS and ASD (McElhinney et al. 2002).

Noonan Syndrome (NS) is a relatively common multiple congenital anomaly syndrome, inherited mostly in an autosomal dominant fashion, with some cases occurring sporadically. The affected individuals have characteristic facial features, pterygium colli, short stature, and congenital heart abnormality (Noonan 1994). Congenital Heart Disease is found in about half of Noonan Syndrome patients, with pulmonary stenosis and hypertrophic cardiomyopathy (HCM) are the two most common cardiac manifestations (Noonan 1994; Marino et al. 1999). Other symptoms include ASD, TOF, aortic coarctation, mitral valve anomalies and AVC (Marino et al. 1999). At least 8 genes were found to be associated with NS, including *PTPN11*, *SOS1*, *RAF1*, *KRAS*, *BRAF*, *MEK1*, *MEK2*, and *HRAS* (Tidyman and Rauen 2009). All the genes are part of the Ras/Raf/MEK/ERK signaling pathway, which involves cell proliferation, differentiation, and survival.

2.2 NON-SYNDROMIC CONGENITAL HEART DISEASE

Non-syndromic congenital heart disease is the most prevalent form of CHD. The most common group of genes implicated in CHD are the transcription factor genes. Other genes, such as components signaling transduction pathways, structural components of cardiomyocytes, are also common CHD genes.

GATA4 mutations are a well-established cause of CHD in humans. The most common phenotypes were ASD, VSD, TOF, and AVSD (Garg et al. 2003; Nemer et al. 2006). Heterozygous *GATA6* mutations usually cause CHD characteristic of a severe form of cardiac outflow tract (OFT) defect, namely persistent truncus arteriosus (PTA). (Kodo et al. 2009), and sometimes cause ASD and TOF (Lin et al. 2010).

Many different heterozygous *NKX2-5* mutations have been found in patients with CHDs, and they were transmitted in an autosomal dominant fashion. (Kasahara et al. 2000). The most common phenotype is ASD with or without Atrioventricular (AV) block and other common phenotypes are VSD, TOF (Reamon-Buettner and Borlak 2010).

Apart from Holt-Oram Syndrome, mutations in *TBX5* has not been implicated in non-syndromic CHD, and apart from a single report, findings of mutations in DiGeorge Syndrome gene *TBX1* have not been duplicated in non-syndromic CHD patients (Fahed and Nemer 2012). Mutations in *TBX20* are associated with a family history of CHD and display a complex spectrum of developmental anomalies, including defects in septation, chamber growth, and valvulogenesis. (Kirk et al. 2007; Posch et al. 2010).

As the advances of genomic technology, more and more transcription factors have been found to associated with CHDs. X-linked *ZIC3* mutations have been identified in cohorts and families of cardiac situs abnormalities and TGA (Gebbia et al. 1997; Megarbane et al. 2000). Heterozygous missense *ZFPM2* mutations have been reported to contribute to some sporadic cases TOF patients (Pizzuti et al. 2003). *CITED2* mutations are potential risk factors for CHD and account for about 2% of sporadic cases. Common symptoms include ASD, VSD, PS, and TOF (Sperling et al. 2005). *ANKRD1* mutations have been reported as a possible cause of TAPVR (Cinquetti et al. 2008).

Lots of signaling pathways are crucial for the heart development. Mutations in the components of these signaling pathways are also the common cause of non-syndromic CHD. *NOTCH1* mutations can cause a wild range of non-syndromic cardiac defects, including BAV, AS, CoA, and HLHS (Garg et al. 2005; Mohamed et al. 2006; McBride et al. 2008) Apart from Alagille syndrome, *NOTCH2* (McDaniell et al. 2006) and *JAG1* (Bauer et al. 2010) both cause non-syndromic CHD. Heterozygous *CFC1* mutations in heterotaxy syndrome patients have been associated with outflow tract defects such as TGA and DORV (Goldmuntz et al. 2002). Loss-of-function mutations in *GDF1* have been associated with congenital heart defects, contributing to cardiac defects ranging from TOF to TGA (Karkera et al. 2007). *NODAL* mutations have been identified and characterized in patients with heterotaxy, and isolated cardiovascular malformations (Mohapatra et al. 2009).

The cardiomyocyte are the most important cell types in the heart, and contractile proteins take up a large part of cardiomyocyte. Problems in the contractile proteins usually results in the dysfunction of the cardiomyocyte. Mutations in contractile protein genes, including *MYH6*, *MYH7*, *MYH11*, *MYBPC3*, *ACTC1*, usually cause cardiomyopathies, such as hypertophic cardiomyopathy (HCM). However, some of these contractile genes sometimes

also associate with CHDs. For example, mutations in MYH6 is reported to associate with ASD (Ching et al. 2005); MYH7 mutations can cause septal defects and Ebstein's Anomaly (Budde et al. 2007); ASD and PDA have been reported in severe HCM patients with homozygous mutations in MYBPC3 (Xin et al. 2007).

2.3 OTHERS MECHANISM

With the development of new detection technologies, other mechanisms for CHD, such as miRNAs, lncRNAs, histone modification, chromatin remodeling, copy number variations, have also being taken into account nowadays. A small fraction of CHDs can be explained by CNV, and it has been reported that the frequency of causal CNVs in non-syndromic CHD populations is lower than in syndromic CNV populations (3.6 vs 19%) (Breckpot et al. 2011). miRNAs, such as miR-1, miR-133, and miR-208, are reported to be important for heart development and cardiomyocyte differentiation (Chen et al. 2006; van Rooij et al. 2007). Baf60c, encoded by SMARCD3, is an essential component of Swi/Snf-like Brg1/Brm-associated (BAF) chromatin remodeling complexes. It has been shown that this Baf60c containing chromatin remodeling complex plays a crucial role in heart developments (Lickert et al. 2004; Cai et al. 2013). However, most of these researches were based on mice and cell modeling, most of examination for real CHD patients is still focusing on the canonical genetic mutations.

3 CARDIOGENESIS IN MURINE

The heart is a complex system composed of muscle and non-muscle cells. The main questions in heart development are, 1) what are the cells in the heart; 2) where do they come from; 3) how is their relationship with each other. The ventricles take up the major part of the heart. But itself has complex cell type compositions. According to a previous study (Nag 1980), the ventricle of an adult rat heart contains approximately 65%~70% of non-cardiomyocyte cells, while 30%~35% of cardiomyocytes. Assuming similar composition, cardiomyocytes take up about 30% of an adult heart. Through extensive lineage studies, three major heart progenitor populations have been found: cardiogenic mesoderm cells, the pro-epicardium and cardiac neural crest cells (Brade et al. 2013). The cardiac mesoderm cells are of particular interest, as they are the main contributors to the cardiomyocytes.

3.1 HEART DEVELOPMENT WITH CARDIAC PROGENITORS

Cardiac mesoderm precursors are among the first cells that ingress through the primitive streak during gastrulation. After the ingression, these cardiac mesoderm cells migrate away from the primitive streak toward the anterior, and form the cardiac crescent on E7.5.

On E8 the cardiac crescent folds inwards and fuses at the midline and gives rise to the FHF-derived linear heart tube. At this stage, the linear heart tube starts to beat. This linear heart tube is constructed with two layers of cells: outside is a cardiomyocyte layer called the myocardium and inside is an endothelial layer called the endocardium layer. Before the embryo turns, the linear heart tube loops to the right and start to bulge. This linear heart tube expands by self-proliferation and the integration of the SHF progenitors that are moving in the arterial and venous poles of the linear heart.

On E9.5, the preliminary left and right ventricles are visible and keep ballooning. The atrioventricular canal, the septum between atria and ventricle, starts to form at this stage. At this stage, the freely floating pro-epicardial cell vesicles attach to the naked myocardium, starting at the atrioventricular canal region.

On E10.5, the primitive shape of the four chambers has formed.

On E11, the ventricular septum begins to form. The cardiac neural crest cells differentiate into aortic smooth muscle cells and the aorticopulmonary septum cells. These cells reshape the outflow tract into aortic and pulmonary trunks by providing signals as well as cells for these cardiac structures.

By E11.5, the epithelial cell sheet derived from pro-epicardial cells will have covered the whole heart. On E11.5-13.5, some of epicardial cells undergo an epithelial-mesenchymal transition and differentiate toward various cell types, such as coronary vasculature and cardiac fibroblast.

On E14, the four chambers and the septum between them have formed.

The movement of these progenitors are recently possible to be track in-real time in an ex-vivo culture (McDole et al.). However, the genetic programs govern their activity remains elusive. To identify the cardiogenesis genetic program in mice, the lots of loss-of-function genetic experiments have been performed and have provided important clues.

3.2 THE GENE LANGUAGE OF HEART DEVELOPMENT

Through decades of research, the studies of some specific genes have provided us lots of insights about heart development. Some of these genes are identified as markers for important progenitor pools that are crucial for cardiogenesis. As these mark genes become the foundation of our current understand of cardiogenesis, they themselves form a language describing the molecular process of cardiogenesis.

Mesp1&2

The basic helix-loop-helix transcription factor Mesp1 marks the cardiac mesoderm progenitors. Disruption of the *Mesp1* gene in mice leads to aberrant heart morphogenesis, resulting in cardiac bifida (Saga 1998). *Mesp1* and *Mesp2* double-knockout embryos died around 9.5 days postcoitum without developing any posterior structures such as heart, somite or gut (Kitajima et al. 2000). β-gal/lacZ lineage tracing shows the Mesp1-expressing cells migrate out from the primitive streak and are incorporated into the head mesenchyme and form the heart field (Saga et al. 1999). Clonal analysis of Mesp1-expressing cells demonstrates that two temporally distinct pools of Mesp1 progenitors sequentially give rise to the FHF and then the SHF progenitors (Lescroart et al. 2014). It has been estimated that ~244 ± 26 Mesp1 progenitors contributed to heart development (Chabab et al. 2016).

Isl1

LIM homeodomain transcription factor Isl1 marks the second heart field (SHF) progenitors. Hearts of mice lacking Isl1 are completely missing the outflow tract, right ventricle, and much of the atria (Cai et al. 2003). Further studies showed that these Isl1+ progenitors could be found in postnatal rat, mouse and human myocardium (Laugwitz et al. 2005). During the heart development, these progenitors contributed to various cell lineages within the heart, such as cardiomyocytes of the right ventricle and atria as well as vascular smooth muscle and endothelial cells (Moretti et al. 2006). Later studies showed that these progenitors could ultimately generate a series of rare, transient Isl1+ ventricular heart muscle progenitors that are completely committed to the ventricular muscle fate (Domian et al. 2009).

Nkx2-5

NK2 homebox 5, Nkx2-5 is expressed in both the first heart field (FHF) and SHF. *Nkx2-5*$^{-/-}$ mice embryos showed arrest of cardiac development after looping and poor development of blood vessels (Tanaka et al. 1999). In the heart of adult chimeric mice generated from Nkx2-5 null mESCs, there were almost no mESCs-derived cardiomyocytes (Tanaka et al. 1999). The single ventricle-like chamber of Nkx2-5 null embryos is correlated to the normal

left ventricle and mainly derived from the first heart field (Prall et al. 2007). In Nkx2-5-null embryos, Isl1 persisted in cardiomyocytes of the late cardiac crescent and heart tube, suggesting a role of Nkx2-5 as a SHF suppressor (Prall et al. 2007).

Gata4

The zinc finger transcription factor Gata4 is expressed in precardiogeneic splanchnic mesoderm at early heart development stage. Gata4 null mice arrests in development between E7.0 and E9.5. Mutant embryos generated two independent heart tubes. But no random looping was observed in contract to *Mesp1* knock out mice. (Molkentin et al. 1997)

Mef2c

The myocyte enhancer factor Mef2c is expressed in cardiac precursor cells before the formation of the linear heart tube. Mice homozygous for a null mutation of Mef2c have cardiac morphogenesis defects, the heart tube fails to loop, and the future right ventricle fails to form. At the molecular level, a subset of cardiac genes, such as *Hand2*, was not expressed normally. (Lin et al. 1997)

Tbx1/Tbx5/Tbx20

Tbx1$^{+/-}$ mice have a high incidence of cardiac outflow tract anomalies. *Tbx1*$^{-/-}$ mice displayed a wide range of developmental anomalies. As a mice model for the human DiGeorge syndromic CHD, the mutant mice showed almost all of the common DiGeorge syndrome features, including hypoplasia of the thymus and parathyroid glands, cardiac outflow tract abnormalities, abnormal facial structures, abnormal vertebrae and cleft palate.(Jerome and Papaioannou 2001)

Tbx5$^{-/+}$ mice were used to model Holt-Oram syndrome and have cardiac and forelimb abnormalities. Enlarged hearts were found in 8-week-old *Tbx5*$^{-/+}$ mice. Both atria were dilated and the ventricles had a bulbous appearance. *Tbx5*$^{-/-}$ mice have severe hypoplasia of posterior domains in the developing heart. (Bruneau et al. 2001)

Tbx20-null mice exhibit severely hypoplastic hearts and Tbx2 expresses throughout the heart, which is normally restricted to outflow tract and atrioventricular canal. The hearts of Tbx20 mutant mice were unlooped and severely hypoplastic (Cai et al. 2005). In another report, the Tbx20 mutant hearts had a distinct additional compartment in the outflow region became progressively obvious in mutant hearts, resulting by E9.5 in a heart tube with two small chamber-like swellings separated by a circumferential sulcus (Stennard et al. 2005).

Hand1&2

During mouse heart development, Hand1 and Hand2 are expressed in a complementary fashion and are restricted to segments of the heart fated to form the right and left ventricles. Targeted gene deletion of Hand2 in mouse embryos resulted in embryonic lethality at E10.5. Hand2 is required for formation of the right ventricle of the heart and the aortic arch arteries

(Srivastava et al. 1997). Hand1 null mice died between E8.5 and E9.5, and the heart development did not progress beyond the cardiac-looping stage (Thomas et al. 1998).

Foxp4/ Foxh1/ Foxc1&2

In *Foxp4* mutant embryonic mice, each bilateral heart-forming region is capable of developing into a highly differentiated four-chambered mammalian heart in the absence of midline fusion (Li et al. 2004).

Foxh1$^{-/-}$ mutant mouse embryos from a primitive heart tube, but fail to form out flow tract and right ventricle and display loss of outer curvature markers of future working myocardium (von Both et al. 2004).

Single *Foxc1* or *Foxc2* mutants, compound *Foxc1*$^{+/-}$; *Foxc2*$^{+/-}$ and *Foxc1*$^{-/-}$; *Foxc2*$^{+/-}$ mutants exhibited no obvious abnormalities in the formation of the OFT and RV at E9.0. Other compound Foxc1; Foxc2 mutants have a wide spectrum of cardiac abnormalities, including hypoplasia or lack of the outflow tract (OFT) and right ventricle as well as the inflow tract, dysplasia of the OFT and atrioventricular cushions, and abnormal formation of the epicardium, in a dose-dependent manner. Compound *Foxc1*$^{-/-}$; *Foxc2*$^{-/-}$ mutants die around E9.5. (Seo and Kume 2006)

T

The classical mouse mutation brachyury (‘‘short tail’’) was first described in 1927. Heterozygotes for null alleles of Brachyury have short tails and mild skeletal defects due to haploinsufficiency. Homozygous T-null mice have severe developmental disorders, such as defective mesoderm formation and regression of the notochord, and they usually die before E10 (Meisler 1997). The hearts of all T-null mice have some morphological abnormality. The heart looping inversion occurs in T-null mice at a chance of 50%. Aberrant expression of both the lefty genes and nodal is also observed. (King et al. 1998)

Eomes

The T-box gene Eomesodermin (*Eomes*) is essential for both trophoblast development and gastrulation. In chimaeras embryos, where the mutant cells intermingle with the wild type, mesoderm can be partially formed. Further examination showed the wild type cells had migrated efficiently through the primitive streak and become enriched in the mesoderm wings, whereas most mutant cells persist in the ectoderm. When injected into syngeneic hosts, mutant cells give rise to teratomas containing muscle, cartilage and haematopoietic tissue. These experiments showed that Eomes is not required for the differentiation of some mesodermal cell types, and indicates that the mutation may specifically affect the morphogenetic movement of cells from the epiblast into the primitive streak (Russ et al. 2000). In another study, Eomes has been shown as the upstream of *Mesp1* to specify cardiac mesoderm during mouse gastrulation (Costello et al. 2011).

Smad4/Nodal

Smad4 null mice is embryonic lethal on E6.5-E8.5, without gastrulation (Yang et al. 1998). But wild-type tetraploid morale rescues the gastrulation defect (Sirard et al. 1998). To characterize further the requirement of Smad4 during early gastrulation, Sox2Cre was used to specifically inactivate *Smad4* in the epiblast. Mutant embryos form abundant extra-embryonic mesoderm, including allantois, failure to pattern derivatives of the anterior primitive streak, a rudimentary heart, middle primitive streak derivatives such as somites and lateral plate mesoderm. The somites were fused across the midline. No definitive endoderm was derived from Smad4-deficient epiblast cells (Chu et al. 2004). Furthermore, the requirements of Smad4 in later stage of heart development is also found (Wang et al. 2005; Qi et al. 2007; Song et al. 2007).

Mouse embryos deficient for the activity of Nodal fail to form both the mesoderm and the definitive endoderm. They also fail to specify the anterior visceral endoderm (AVE). Their study reveals that Nodal$^{-/-}$ epiblast cells express prematurely and ectopically molecular markers specific of anterior fate. (Camus et al. 2006)

The loss-of-function approach to study the function of a specific gene in heart development is kind of like the stop reaction approach used in the Sanger sequencing. Every time a new defect shows up, we learn another critical step of heart development. When the mutant cells were building the heart, they would stop at the step required the function they have lost, relating the function with the corresponding gene. Take Isl1 for instant, the mutant embryo failed to form OFT and other structures, therefore OFT formation is critical step and Isl1 plays crucial role in it. Although the detail molecular mechanism is usually missing, the relationship between the function, OFT formation, and the gene, Isl1, is established. By examine these development defects, there are three major patterns: 1) failure in the early stage, like gastrulation; 2) the heart tube fails to loop or fuse; 3) the second heart field relative structures fail to form or are dysplastic. The first pattern is a differentiation issue, which is mainly modeled with ES differentiation. The second pattern is mainly a migration issue, which could sometimes be examined by marker, such Mesp1. But the third defect pattern currently are beyond the modeling capacity of ES system, since it could only be examined by formation real structures.

3.3 MARKERS AND LINEAGES

During embryo development, the pluripotent stem cells in the inner cell mass give rise to different precursor cells and increasingly restrict their lineage potential. Understanding the lineage through which the tissue and organs are formed is a fundamental question in developmental biology. To do that, we need two pieces of information, the cell identity and the lineage relationship between cells.

The cell identity problem brings us to one of the most important concepts in developmental genetics, markers. Since the discovery of DNA, we learn every individual has its unique DNA codes. Similarly, we hope there is unique transcription profile for each

individual cell, or at least a subset of cells. Therefore, one marker for one cell type is a very attracted concept. Generally speaking, we can put markers into three categories. The first one is perfect marker, which can truly define a cell type. These markers would not express in any other cell types, such as MLC2v, which usually expresses only in the ventricular cardiomyocytes. However, even within the cells marked by the marker gene, depending the definition of cell type, there might be subtypes. Actually, perfect marker rarely exits, especially when it comes to the intermediates, like progenitors. The second type of marker is tissue restricted marker. Within the specified tissue, the marker gene expresses in some type of cell, but not in other cells. For instant, in the heart, ISL1 marks the second heart field progenitors (Cai et al. 2003). On the other hand, ISL1 also marks islet cells in the pancreas (Ahlgren et al. 1997), and motor neuron in the brain (Pfaff et al. 1996). The third type is temporal restricted marker. This type of marker labels the cells only at the specific time window. Outside of that window, it might label totally different cells. One of such marker is HCN4, which markers the first heart filed progenitor around E7.0, while, at later stages, it labels primarily the SAN node (Spater et al. 2013). In addition, the functional attribute of a maker can be classified as strong marker and weak marker. The strong marker plays a crucial role in the defined cells. Genetically disrupt such marker gene will result in the loss of the labeled cell population. On the hand, the loss function of a weak marker won't affect the labeled cell populations, which usually happens to a surface marker. Taken together, the genetic markers allow us to define certain cells with some spatial-temporal restrictions, but they might not be functionally associated with the cell population.

From the experiment aspect, to define a marker gene, the golden rule is the well-known Cre-*loxP* system. The Cre recombinase is a 38kD protein from P1 bacteriophage that catalyzes the recombination between two *loxP* sites (Hamilton and Abremski 1984). A typical marker lineage tracing experiment will knock-in Cre in the marker gene locus. When the gene activated during development, the Cre would faithfully express and modify the *loxP* containing reporter. As mentioned previous, a maker usually requires some spatial-temporal restriction. The tissue specificity is accomplished by knock-in the Cre into the maker gene loci, while the temporal control is achieved via modified Cre, such as CreERT2 (Feil et al. 1997) and MerCerMer (Zhang et al. 1996), under the control of Tamoxifen at the chosen injection time point.

A Cre reporter usually contains a STOP sequence flanked by two *loxP* sites, which would be excited when Cre is activated, allowing the downstream transcription. On the Cre reporter size, *LacZ* is one of the most comment reporter for Cre lines. The product of *LacZ* is β-galactosidase, which turns X-gal, an artificial substrate, into galactose and 4-Cl,3-Br indigo thus producing a deep blue color. The advantages of *LacZ* reporter is the lineage of marked cells are directly visible after X-gal treated. Another common Cre reporter is to replace the β-galactosidase with fluorescent proteins, such EGFP, or tdTomato. But it is more common that combing both colors, the so-called mTmG reporter, which the *loxP* sites flank the tdTomato and STOP cascade, followed by the EGFP. This reporter gives a red background when the Cre is inactivated, and the marked lineage would have a distinct green fluoresce.

To get some information about dynamics of the marked lineage, stochastic labeling is very useful. The first and most well-known example is the brainbow (Livet et al. 2007). The *loxP* sites or/and its equivalents, and fluorescent proteins were arranged in a specific way that can generate diverse random colors upon Cre activation. The color difference between the progenies/marked cells allows the tracking of the induvial cells that give rise to difference tissues and structures. This type of *in vivo* clonal analyses provide lots of insights about how an organ is developed and regenerated (Gupta and Poss 2012), as well as how the tissue homeostatic is maintained (Snippert et al. 2010).

Beside Cre-lox system, other similar tools have also been developed, such as FLP-FRT (Dymecki and Tomasiewicz 1998), and Dre-rox (Anastassiadis et al. 2009). However, these lineage tracing systems all require some prior knowledge about the maker gene of the lineage. To discover markers for sub-populations and study multiple lineages, newly developing tools such single-cell RNA-Seq (Lescroart et al. 2018) and CRISPR barcoding (Kalhor et al. 2018) go beyond the Cre-lox system and open a new chapter for the old subject of lineage study. Single-cell RNA-Seq profiles individual cells and project them onto a 2- or 3-dimension representation space with visual distant as a metric of the transcriptome similarity. With this approach, new sub types and bifurcation points can be found based on the transcription profile not just pre-existed markers. Pseudo-time and computational lineage can be extract from such datasets, and by examining the expression in the lineage branch new markers for that specific branch can be identified (Kester and van Oudenaarden 2018). CRISRP barcoding construct the lineage tree of the tissue of interested by analysis the mutations induced by Cas9 during development. These new technologies are still in their early development stage, the confidence level of their lineage conclusions is less than the canonical Cre-lox system, but they bring the hope to resolve the lineage mystery at large scale.

Another fruit of thought comes from the long noncoding RNA (lncRNA). LncRNAs have been shown to be more tissue specific comparing to protein markers (Liu et al. 2016). The argument is the diversity of the tissue are more of the result of the diversity in the regulatory elements, while the protein machinery has not changed much in the mammalians evolution trees. Anyway, the importance of marker genes, coding or noncoding, in developmental biology and other branches will continues to growth and providing more and more vocabularies for the language of genes, hopefully one day the language of cells.

3.4 DIFFERENCES BETWEEN MICE AND HUMANS

The study based on mice model have gave us significant insight about the molecular mechanism of heart development. However, there are lots of difference between the human heart and mice's. A mouse heart is created in just a couple of days, while the human heart is continuously built over several months. The physiology of human heart is quite different from mouse heart, including heart rate and blood presure. For example, in mouse the heart rate is usually 450-750bpm, while the human heart rate is usually 60-100bpm. The size of mice heart is about $1\times0.5\times0.3$ cm^3, while the human heart is about $13\times8\times5$ cm^3, which is more than 3000 times bigger (Hansson et al. 2009). The bigger the size is, the more cells are

allocated to build the heart, which also means the complexity of human cardiogenesis is much higher than mice's. This implies a higher degree of progenitor proliferation in human cardiogenesis. To learn the authentic human cardiogenesis, we need to study the cardiogenesis using the human model. But our options are limited, one of them is the human pluripotency stem cells model.

4 CARDIOMYOCYTE FROM HPSC

4.1 CARDIOMYOCYTE DIFFERENTIATION

Human pluripotent stem cell (hPSC) includes human embryonic stem cell (hESC) and induced pluripotent stem cell (iPSC). The first hESC line was derived from human blastocysts in 1998 (Thomson et al. 1998). Human iPSC lines were first derived in 2007 by overexpression of key stem cell related transcriptions factors in somatic cells (Takahashi et al. 2007; Yu et al. 2007). hPSC can self-renew indefinitely in culture while maintaining the ability to differentiate into the 3 primary germ layers, hence potentially any cell types in the human body. In principle, these cells are the ideal sources for many cell-based applications, such as, cell therapy, drug safety test, disease modeling, and other basic researches. However, to make full use of these cell sources, robust and efficient differentiation protocols for generating the cell types of interest are essential.

Embryoid bodies (EB) are the three-dimensional aggregates formed in suspension by pluripotent stem cells. Cells in EBs differentiated to derivatives of the 3 primary germ layers. Different types of cells can be obtained by providing different development cues to these cells. The first successful EB-mediated differentiation of hESCs to cardiomyocytes was reported in 2001 using H9 hESC line. By plating EBs to 0.1% gelatin-coated culture dish, beating areas can be observed after 4 days after (Kehat et al. 2001). Comparable EB-based differentiation protocols were reported to generate cardiomyocytes from a variety of hESC lines: H1, H7, H9, H14 and hiPSC lines: IMR90 C1/4, iPS(Foreskin) C1/2 (Xu et al. 2002; He et al. 2003; Zhang et al. 2009). These EB-based differentiation protocols were difficult to reproduce and exhibited line-to-line variability in efficiency (Osafune et al. 2008), and have two addition issues that limits their applications: 1) the differentiation media and culture conditions were not chemically defined, using feeder cell layer for the culture, containing serum and other animal original components; 2) the EBs formed during the processes were highly variable in size, structure and composition. After a few years' research, serum and feeder free culture conditions were established (Ludwig et al. 2006a; Ludwig et al. 2006b; Chen et al. 2011). And lots progresses have been made in homogenizing the EB formation, such as spin EBs (Ng et al. 2005; Burridge et al. 2007), micro well EBs (Khademhosseini et al. 2006; Mohr et al. 2006) and micro patterned EBs (Peerani et al. 2007; Bauwens et al. 2008).

The second phase of cardiac differentiation begins with the replacement of the serum with growth factors in the differentiation media. Three major serum-free defined basic media for cardiac differentiation were developed, including APEL from Stem Cell Technologies, StemPro-34 and RPMI from Thermo Fisher Scientific/life technology. A range of growth factors implicated in normal cardiac development have been tested, including BMP4, Activin A, FGF2, Wnt agonists and antagonists, and vascular endothelial growth factor. Growth factors based protocols still exhibit line-to-line variability. Different cell lines require different concentrations of Activin A and BMP4. However, by monitoring of KDR^{low}/KIT^{-}

(Yang et al. 2008) or $KDR^+/PDGFRA^+$ (Kattman et al. 2011), the optimal concentrations of Activin A and BMP4 for each cell lines can be found, and high efficient cardiomyocytes differentiation to be achieved, about 50%~65%.

EB methods have lots of drawbacks, such as low cardiomyocytes yield and heterogeneous in the differentiation product. An improved approach is monolayer-based method. hPSCs were first singularized and seeded on Matrigel or defined matrix-coated Petri dishes to produce a homogenized monolayer. When the cells were confluent, Activin A and BMP4 was sequentially added with defined RPMI/B27 medium. This method generated more than 30% cardiomyocytes in the H7 hESC line before optimization Activin A and BMP4 concentration and cardiomyocytes can be enriched to around 80% with Percoll gradient centrifugation purification (Laflamme et al. 2007). However, this method still requires optimizations for different cell lines. One improvement for this protocol is to overlay Matrigel to the cells to produce thicken the mono cell layer, which is subsequently treated with Activin A, BMP4, and FGF2 in RPMI/B27-minus insulin medium to induce robust cardiomyocyte differentiation. This is known as the matrix-sandwich method (Zhang et al. 2012). The matrix-sandwich method generated CMs with high purity (up to 98%) and high yield (up to 11 CMs/input hPSC) from multiple hPSC lines, including H1, H9, IMR90 C4, DF6-9-9T, DF19-9-7T, DF19-9-11T.

The next break-through in cardiomyocyte differentiation comes from the replacement of growth factors stimulation with chemical stimulation. By adding a Gsk3 inhibitor to activate canonical WNT/β-catenin signaling pathway and later a PORCN inhibitor to inhibit WNT signaling pathway, cardiomyocytes can be robustly generated when these chemicals are added at the right time with the right concentration. This temporal modulation of WNT signaling pathway with small chemicals is also known as the GiWi method (Lian et al. 2013). Combing the GiWi method and some chemically defined hPSC culture media, a chemically defined differentiation protocol was developed (Burridge et al. 2014). From media, coating matrix to differentiation stimulus, all the reagents are chemically defined, opening the doors of producing clinical grade hPSC derived cardiomyocytes (hPSC-CMs). Later, the albumin-free GiWi, named GiWi2, was developed, which further simplified the differentiation media with only RPMI and ascorbic acid (VC) (Lian et al. 2015). In another method, the basal differentiation medium is DMEM/F-12 instead of RPMI, and albumin can be removed in the present of heparin (Lin et al. 2016).

From the "spontaneous" EB protocol in 2001 to the total chemically defined protocol in 2014, the differentiation efficiency increased from less than 1% to 95%. After a decade and half's effort, the *in vitro* cardiogenesis model is finally established. The unlimited supply of hPSCs and robust and high efficient differentiation protocols enable a wide range of applications of cardiomyocyte differentiation, such as disease modeling, cardiotoxicity screening, drug discovery, human cardiogenesis research and cell therapy relative research.

4.2 CARDIOMYOCYTE MATURATION

There are a number of reasons why producing mature cardiomyocytes are important. But the primary reason is, compared to mice and the model organism, the access of adult human cardiomyocytes for research is very limited. The mature hPSC-CMs are better proxy for the human cardiomyocytes than mice cardiomyocytes. Tests performed on the more mature cardiomyocytes are more relevant and reliable for developing new therapy, testing drug safety, and modeling heart diseases. However, the status quo of the hPSC-CM is they resemble the immature cardiomyocytes and methods for improving the maturation of hPSC-CMs are underdeveloped. Comparing to the adult cardiomyocytes, the hPSC-CMs are different in many ways, such as morphology, electrophysiology, calcium handling, ultrastructure, metabolism et.al.

In terms of morphology, an adult cardiomyocyte is much larger and cylindrical shape, about 150μm × 10μm, and most of them are bi- or multinucleated. hPSC-CM is smaller and spherical, about 5~10μm in early hPSC-CM and 30μm × 10μm in late hPSC-CM, and most of them are mono-nucleated (Robertson et al. 2013). In adult CM, the contractile machinery, mainly myofibril, takes up 40%~52% of the cytoplasm and mitochondria take up 15%~25% (Tashiro et al. 1990; Barth et al. 1992). hPSC-CM has much lower cytoplasm-nuclear volume ratio and the machinery and mitochondrial take up much less space of the cell (Robertson et al. 2013).

The differences between hPSC-CM and adult CM in electrophysiology are well documented. The adult CM has *maximum diastolic potential* about -85mV~ -90mV (Robertson et al. 2013; Yang et al. 2014), while hPSC-CM is about -30mV at early stage and advances to -60mV to -75mV at late stage (Robertson et al. 2013). The *action potential duration at 90% of repolarization* (APD90) is 250~400ms in hPSC-CM, which is longer than ~213ms in adult CMs (Hoekstra et al. 2012). The biggest difference is the *maximum rate of depolarization* (max dV/dt, Vmax), adult CM is 215~300V/s, while hPSC-CM has only 10~40V/s at late stage (Hoekstra et al. 2012; Robertson et al. 2013). And the propagation of the electrical signal is 20cm/s in hPSC-CM (Lee et al. 2012) and 60cm/s in adult CM (Veerman et al. 2015).

Calcium handling is important for cardiomyocytes, connecting the electrical signal with the contractility. In the hPSC-CMs, a crucial structure is usually missing, transverse tubules or T-tubules. This structure enables special and functional coupling of the L-type calcium channel (CACNA1S) and cardiac ryanodine receptor (RYR2). Lack of this structure results in U-shape Calcium wave fronts, in contrast to uniform calcium wave front in the adult CMs (Lieu et al. 2009). In addition, the expression of some sarcoplasmic reticulum proteins, such as Caveolin-3 (CAV3), Calsequestrin (CSQ), and amphiphysin-2 (BIN1), are lower in hPSC-CMs (Dolnikov et al. 2006; Synnergren et al. 2012).

The force generation unit in cardiomyocyte are the sarcomeres, a structure repeating itself between the two Z-lines. In adult CMs, all the prominent myofilament regions, such as

Z-line, I-band, A-band, H-zone and M-line, are present, suggesting the formation of overlapping and non-overlapping myofilaments. In contrast, M-lines are usually missing in hPSC-CM, as in vivo they do not develop before the neonatal period (Scuderi and Butcher 2017), indicating the immaturity of the hPSC-CM.

Another difference between the adult CMs and hPSC-CMs is the metabolic pathways. The adult CMs use the fatty acid oxidation metabolism, while hPSC-CMs use mostly glycolysis. However, the metabolism switch has already begun in hPSC-CMs, and has been used for large-scale cardiomyocyte purification (Tohyama et al. 2013). Other differences, such contractile force, gap junction distribution, responses to β-adrenergic stimulation, have been systematic reviewed by Yang et al. 2014.

To improve the maturation status of hPSC-CMs, lots approaches have employed, including electrical stimulation, mechanical stimulation, biochemical stimulation, different extracellular matrix, co-culture with different cells, force gene expression. However, so far, these methods can only improve the maturation of hPSC-CM from certain aspects. One example of such studies is reported recently. Applying both electrical and mechanical stimulation at the early stage of hPSC-CM can induce the formation of T-tube, but the electromechanical properties did not achieve the stage of maturity seen in adult human myocardium under such stimulation (Ronaldson-Bouchard et al. 2018). To achieve a satisfied maturation state, lots of works are still needed to be done.

4.3 CARDIOMYOCYTE SUBTYPES

Most of the differentiation protocols yield three types of cardiomyocytes according their action potential characters: ventricular-like, atrial-like and pacemaker-like cardiomyocytes. The variables are the composition of the three types cardiomyocytes. A very nature question how to enrich different types of cardiomyocytes. Since the loss of ventricular cardiomyocytes during heart attack or nature aging can barely regenerate (Bergmann et al. 2009), ventricular cardiomyocytes are in great demands for the cell replacement therapy. And pacemaker cardiomyocytes have the potential of being engineered into a biological artificial pacemaker.

Besides the electrophysiology features, atrial and ventricular CMs have their distinct genetic markers. MLC2v (MYL2) specifically marks the ventricular CMs, while MLC2a (MYL7) marks both atrial and ventricular CMs at early stage, but restrict to the atrial CMs at later stage. With these markers, protocols have been developed to enrich both the ventricular and atrial CMs. Previous study has shown the activation of retinoic acid (RA) signaling increase the atrial CM population, while inhibition of this pathway increases the ventricular CM population (Zhang et al. 2011). And further study shown, the atrial and ventricular CMs are from different mesoderm sub-populations: ALDH+ mesoderm give rise to atrial CMs; and CD235a (GYPA) marked mesoderm becomes ventricular CMs (Lee et al. 2017).

The advance in producing pacemaker CMs is the conversion approach. Inspired by the study in neonatal rat ventricular myocytes (NRVM) that Tbx18 is able to convert NRVMs

into pacemaker like cells both *in vitro* and *in vivo* (Kapoor et al. 2013; Hu et al. 2014), ventricular cardiomyocyte can be first derived from hPSC, then convert to pacemaker cardiomyocyte by overexpressing TBX18 (Gorabi et al. 2018). Similar trans-differentiation by overexpression approaches have also been tried out, such as, overexpressing SHOX2 (Ionta et al. 2015), overexpressing engineered HCN1 channel (Chan et al. 2017).

4.4 DISEASE MODELING WITH HUMAN PSC-CMS

Disease modeling is one of the many applications of hPSC-CMs. Given the challenges of genome editing in hESCs before CRISPR technology, most disease modeling at the early days were using iPSCs from the patients and their relatives. The mutations occurred in the ion channels are among the first diseases that being model by iPSCs, such as KCNQ1 R190Q causes long-QT syndrome (Moretti et al. 2010), a missense mutation in the L-type calcium channel CaV1.2 leads to long-QT with Timothy syndrome (Yazawa et al. 2011), KCNH2 A614V results abnormal action potential in type 2 LQTS (Itzhaki et al. 2011), SCN5A 1975insD/+ results sodium current problem (Davis et al. 2012), PLN R14del mutation induces Ca++ handling abnormalities (Karakikes et al. 2015). The effects of these mutations are relatively simple to measure since once the mutated channels were expressed in the hPSC-CM, the phenotype can be directly assay through electrophysiology studies or calcium measurement.

The second type of diseases being modeled by iPSC-CM are mutations that causes tissue abnormally. One of such examples is TBX20 Y317* T262M mutation cause left ventricular non-compaction cardiomyopathy (Kodo et al. 2016). And the phenotype reveal by this study is the mutation cause ectopic activation of TGF-β signaling, which causes a cardiomyocyte proliferation defect in vivo. This phenotype is at the signaling pathway level, only indirectly suggesting a possible cause of the symptom in patient. Another similar case is the MYH7 A663H mutation results in contractile arrhythmia at the cellular level, while in patient, this mutation associated with familial hypertrophic cardiomyopathy (Lan et al. 2013).

Another interesting disease modeling cases is the arrhythmogenic right ventricular dysplasia (ARVD). ARVD patients are usually carrying mutation in the desmosome proteins: PKP2, DSC2, DSG2, DSP, JUP. The advance stage symptoms are progressive replacement of right ventricular myocytes with fatty or fibrofatty tissue. To model this disease with hPSC-CM is to demonstrate the mutants hPSC-CMs are more susceptible to turn into adipocytes. iPSC-CMs from patients with PKP2 c.2484C>T mutation are indeed easier to become adipocytes, but only under the adipogenic condition (Kim et al. 2013). This study suggests that to reveal the phenotype in cell models, stress challenges are also often required.

The next level of disease modeling with hPSC-CM would require the formation of simple tissue. By ‘simple’, I mean the *in-vitro* tissues that are not vascularized, such as engineered heart tissues (EHT) (Breckwoldt et al. 2017). Barth syndrome (BTHS), a mitochondrial disorder caused by mutation in TAZ. BTHS is characterized by depletion of mature cardiolipin and accumulation of an immature form, monolysocardiolipin. These

phenotypes can be directly measured in cellular assay. The challenge is the myopathy part. By engineering a muscular thin film (MTF) from both wild type and mutant hPSC-CMs, diastolic and peak systolic stresses and Twitch stress can be calculated from the movies of contracting MTFs (Wang et al. 2014a). And these two parameters are useful for scoring the BTHS myopathy.

4.5 TRANSPLANTATIONS

The promise of cell therapy is one of main driven forces of the hPSC-CM research. The first tie of repairing the damage from myocardial infarction with cells were happen during the 90's last century (Marelli et al. 1992; Koh et al. 1993; Chiu et al. 1995). Since then, lots different cells have been used for transplantation in the hope to repaired an infarcted heart, such as skeletal myoblasts (Murry et al. 1996; Jain et al. 2001; Leobon et al. 2003), fetal/neonatal cardiomyocytes (Leor et al. 1996; Li et al. 1996; Scorsin et al. 2000), fibroblasts (Hutcheson et al. 2000), smooth muscle cells (Yasuda et al. 2005), hematopoietic stem cells (Penn et al. 2002; Nygren et al. 2004), mesenchymal stem cells (Toma et al. 2002), endothelial cells (Kocher et al. 2001), adult cardiac progenitor cells (Beltrami et al. 2003; Oh et al. 2003),and hPSC-CMs (Klug et al. 1996; Kehat et al. 2001; Xu et al. 2002; Laflamme et al. 2005; Laflamme et al. 2007; Robey et al. 2008; Fernandes et al. 2010; Shiba et al. 2012; Chong et al. 2014; Ye et al. 2014; Liu et al. 2018). Among them, hPSC-CM transplantation has the most impact so far, and has been systematically studied across several species, and goes all the way to non-human primary studies.

Engraftment is the critical step for cell therapy. In the scenario of cell transplantation into myocardium, engraftment involves cell survival, migration, proliferation, differentiation, maturation, integration, electrical coupling, and vascularization. Once the cells are buried in the heart, the possible controls over the cells are very limited. These cells have to complete a wild range of complicate tasks pretty much on their own. Survival is the first obstacle, only a fraction of the cells transplanted can survival for more than a few weeks (Robey et al. 2008). For hPSC-CMs, there are three principal pathways that reduce the survival of the transplanted cells: ischemia, anoikis, and inflammation-related factors. And pro-survival strategies have been deployed, such as, heat-shocked, incubating with IGF1 and cyclosporine A, supplying the suspension buffer with Matrigel, ZVAD, Bcl-XL BH4, IGF1, CsA, pinacidil (Laflamme et al. 2007; Shiba et al. 2012). Compared to hPSC-CM transplantation, the progenitors should have better survival rate and their proliferation ability is much better than hPSC-CMs. Unfortunately, the identities of the progenitors are varied, and the engraftment outcome depends heavily on the types of progenitors input in.

The next phase of the challenges for the progenitors are migration, proliferation and differentiation. In the moycardiac infaction (MI) heart, the infracted area needs to be invaded by these progenitors. They need to know where to go and where to stop. And at the same time, these progenitor is facing the choice between proliferation and differentiation. If the progenitors differentiate too early, there might not be enough cardiomyocyte to improve the heart wall function. On the other hand, if the progenitors proliferate too much, they might

lose the ability to differentiate into cardiomyocytes. The proliferation-differentiation balance will add another factor to determining the among of progenitors required for the transplantation.

The hPSC-CMs do not have much proliferation ability and no need to worry about differentiation. But both hPSC-CM and progenitor transplantations would have the cardiomyocyte maturation step. The immature cardiomyocytes inside the ventricular myocyte wall can cause problems, such arrhythmia, insufficient force generations. Recent study has shown the stimulations at the early stage helps the hPSC-CM maturation more than that at the later stage (Ronaldson-Bouchard et al. 2018). The moving ventricular myocyte wall provides the stimulations for the hPSC-CMs, either from the direct transplantation or the freshly differentiated from progenitors. In this context, progenitors might yield a better result in terms maturation. But a detail comparison study is still missing till this day.

The integration challenge is how the transplanted cells connect to the native cells, especially the cardiomyocytes. The connection is usually demonstrated by the expression of connexins, such as CX40 and CX43. These membrane proteins form gap-junction channels between cardiomyocytes allowing the transmission of signaling molecules and ions to diffuse between cells. Another integration challenge is the dissolution of fibrosis. The extracellular matrix (ECM) from the fibrosis makes it difficult for the hPSC-CM functionally connect to the native tissue. The transplanted cell would have to find a way to remove these ECMs, restoring the electrical and mechanical property of the tissue. And yet, such study remains difficult to perform.

The long-term survival of a thick graft depends on the oxygen and nutrition supply, which make vascularization a critical requirement. Mature cardiomyocytes use the fatty acid oxidation metabolism. They are very sensitive to oxygen level. The physical diffusion of oxygen is around 150μm, which means any graft more than 10-cells thick require supply from capillary. One strategy is to mix endothelial cell with hPSC-CM to form a network hoping to connect with the vessel network from the host. Another strategy is to recruit the vessel network into the graft either by the cues secreted by the transplanted cells or external chemicals, such as VEGF, added to the cells when transplanted.

Transplanting hPSC-CM or progenitors as cell therapy is still in their early research stage. Most of the time, all we can do is to put some cells into the animal heart hoping something nice would happen. When it happens or dose not, we learn something. By accumulating these information piece by piece, there is still hope to truly pre-program the cells to accomplish the complicated heart regeneration tasks.

5 GENOMIC TECHNOLOGY

During my study two big technology evolutions happen in biology: the rapid development of mass parallel sequencing and CRISPR-Cas9 genome editing. Both technologies are focusing on the genome: one is about reading the genome and the other is about writing it. In this section, we briefly introduce the tools used in our studies.

5.1 MASSIVE PARALLEL SEQUENCING

Illumina sequencer is the main sequencing platform in our studies. A basic knowledge of how the sequencer works can help to understand of library preparations procedure. In an illumina sequencer, the reactions take place on the surface of a flow cell. From the DNA fragments to digital letters, there are two main steps: template amplification and sequencing by synthesis.

The template amplification strategy in illumina machine is called solid-phase bridge amplification. First, the templates are constructed by adding the adapters to both ends of the DNA fragments. Then, these templates are load into the illumina flow cell, on the surface of which were pre-coated with the two types of single strand DNAs that are able to bind to the adapters on the template. Third, during the amplification process, the single strand DNAs on the surface serves as the primers for the PCR reaction. Finally, the elongated DNAs carries the information of the template and can bind to the other primer on the surface, forming a bridge like structure. After dozens of reactions, the spot where a template has bound to will form a cluster containing the DNAs with the same sequence. This spot amplifies the signal from sequencing-by-synthesis to the detectable level.

The second step is to read out the DNA sequence in the spot. This is done by sequencing by synthesis. The key component of this process is the fluorophore-labelled, terminally blocked nucleotide. The fluorophore enables the identification of the incorporated nucleotide through imaging. The terminally block only allow adding one nucleotide at each reaction. Moreover, after the readout, the fluorophore can be cleaved exposing the 3'-OH group. Thus, the blockage is removed, and new nucleotide can be appended at the end. Combing these smart designs, the DNA sequence can be read out one nucleotide by one nucleotide. Though the process is slow, the throughput is high, as millions of spots can be imaged at one reaction cycle.

There is trip worth to mention is how to add adaptors to the DNA fragments to construct the sequencing compatible template. For the DNA fragments whose sequence are partially known, adaptors can be simply added by a simple PCR. For the unknown fragments, random primers or T-A ligations can be used to add the adaptors. But the efficiencies of these methods are not so good and often large amount of input sample material is required. One of the solutions is Tn5 transposase. Like other transposase, Tn5 operates in cut-and-paste manner. After Tn5 is loaded with adaptors, it can just randomly cut the long DNA chain into

small fragments and paste the adaptors at the cut site (Picelli et al. 2014). The combination of the two procedure is called Tagmentation. Because this technique the efficiency of adding adaptors are much high. Materials from even a single cell is enough to produce the sequence library.

With all these technical innovations, breakthroughs in DNA sequencing significantly lower the cost of genome-scale sequencing and make many new applications possible, such as RNA-Seq, ChIP-seq, DNase-Seq, FAIRE-Seq, ATAC-Seq, RRBS. These new applications add layers after layers new information on top of the human genome. RNA-Seq provides the transcriptome map. ChIP-Seq provides histone modification map and identify some transcription binding sites. DNase-Seq and FAIRE-Seq provide information about the open chromatins. ATAC-Seq shows the interactions of the chromatins. RRBS shows the methylation status of the CpG islands. Beside these standardized application, knowing the basic sequencing principle also allow us to develop new applications. Barcode back tracking used in the CRISPR screen is one of such examples.

5.2 TRANSGENIC TOOLS FOR HPSC

Transgenic tools are very useful when we just want to deliver some functional DNA fragment in the genome without targeting specific locus. Lentivirus and *PiggyBac* transposon system are two of such tools. When generating the mutant cell library, low titer Lentivirus was used to deliver the sgRNAs into the hPSCs avoiding multiple integration. In terms of the deliver efficiency in hPSCs, lentivirus is lower than PiggyBac, which is more suitable for library generation. On the hand, for overexpression purpose, *PiggyBac* system is able deliver large cargo at high efficiency and very simple to used. Below, we have included some basic information about these two transgenic tools.

Lentiviruses are a subclass of retroviruses. Retroviruses are useful for *ex-vivo* delivery of somatic cells because they can be linearly insert into host cell genome. However, they have a few problems, such as the low *in-vivo* efficiency, immunogenic problems, the inability to transduce the non-dividing cells. Lentiviruses distinct itself from other retroviruses with their ability to naturally integrate to the genome of non-dividing cells. It has been used as tool for engineering transgenic cell lines due to the high-efficiency infection of dividing and non-dividing cells, long-term stable expression of a transgene, low immunogenicity and the ability to accommodate medium size transgenes with 8kb shuttle capacity.

The drawback of this tool is the safety issues. To transfect the cells, virus must be first produced. To increase the safety, the components necessary for virus production are split across multiple plasmids: transfer plasmid, packaging plasmid and envelope plasmid. Without the packaging and envelope components in genetic material inside the virus, it cannot replicate itself making it safer to work with. But there is still some small chance that the packaging and envelope can rejoin the transfer DNA sequence. The third generation of lentiviruses tool further separate the components, increasing from three-plasmid system to four-plasmid system.

DNA transposons are recently used as a molecular tool for inserting foreign DNA to the host cells. DNA transposons are genetic elements that can relocate between genomic sites by a “cut and paste” mechanism. The system function in pairs. It requires the transposon sequence and the corresponding transposase, the enzyme that moves the sequences around the genome. *Sleeping Beauty* was the first DNA transposon system shown to be functional in mammalian cells. Its transposition efficiency is relatively low and has strong “local hopping” tendency. The system widely used for engineering hPSCs is *PiggyBac* transposon system, which was isolated from insect. One important feature of the *PiggyBac* transposon is that it nearly always excises itself precisely and leaves no footprint behind. *PiggyBac* has significantly higher transposition activity in mammalian cell lines than *Sleeping Beauty* and Tol2.

PiggyBac transposon system has serval advantages over the traditional lentivirus method in engineering new cell lines. First, the inserted DNA fragment can reach 100kb or more (Li et al. 2011), much higher than lentivirus. Second, the whole operation is DNA delivery and there is no need for virus production, purification and transfection. It is a much safer and easier system to work with. Third, the efficiency is much higher in hPSCs than lentivirus. Finally, there is an option that we can remove the inserted DNA fragment. By expressing a mutated form of transposase, which the insertion function is impaired, the inserted DNA fragments can be remove and more importantly sequence at insertion site can be perfectly restored. Because of these advantages, our main validation tool and rescue construct is *PiggyBac* based.

5.3 GENOME EDITING IN HPSC

Four genome editing technologies have been developed, including homology recombination (Gordon et al. 1980), zinc-finger nucleases (Kim et al. 1996), transcription activator-like effector nucleases (Hockemeyer et al. 2011), and the RNA-guided CRISPR/Cas9 system (Cong et al. 2013; Mali et al. 2013).

Since genome editing via homology recombination (HR) is low efficiency and time consuming, only a few cases have been reported. The first genome editing case in hPSCs was reported in 2003, which generated *HPRT1* knock-out and POU5F1-GFP knock-in hESC lines (Zwaka and Thomson 2003). To study cardiogenesis using hESC model, GFP was knocked in to the early mesoderm marker gene *MIXL1* and pan cardiac marker gene *NKX2-5* in hESCs. The *MIXL1*-GFP knock-in hESC line was used to optimize mesoderm induction from hESCs (Davis et al. 2008). The *NKX2-5*-GFP knock-in hESC line was used to optimize cardiomyocyte differentiation from hESCs (Elliott et al. 2011). Based on the pre-existed *NKX2-5*-GFP hESC line, another fluoresce protein mCherry was inserted into the *MESP1* locus creating the first dual-reported hESC line (Den Hartogh et al. 2015). To determine the lineage derivatives from the second heart field progenitors, Cre was knocked in to the locus of *ISL1*, a marker gene for second heart field progenitor (Bu et al. 2009).

DAN double strand breaks (DSBs) are the key for genome editing. There are two general DSB repairing pathways: the Non-Homologous End Joining (NHEJ) DNA repair pathway or the Homology Directed Repair (HDR) pathway. In the present of suitable repair template, usual from an externally introduced donor vector, the DSB is repaired by the HDR pathways. In this repairing pathway, the sequence in the template will be transferred precisely to targeted region. By providing DNA templates with designed features, the feature sequence can be insert precisely into the gene loci in the genome. The NHEJ DNA repair pathway is activated when DSB occurs in the absence of a suitable repair template. With NHEJ repair pathway, a small number, sometimes a couple hundreds, of nucleotides are either randomly inserted or deleted at the DSB site. These indels may cause the shifting of open reading frame (ORF) of the gene, or introducing a premature stop codon. Any of these outcomes from the NHEJ repair pathway will disrupt the normal expression target gene, achieving the loss-of-function of the targeted gene. For decades, the major efforts in genome editing have been focusing on making DSB at specific locus. ZFNs, TALENs and CRISPR-Cas9 systems were developed as tools for genome editing.

Zinc-finger (ZF) and Transcription activator-like effector (TALE) are a class of versatile and programmable DNA-binding proteins. ZFs are stackable motifs of ~30 amino acids that recognize approximately three base pairs of DNA. However, simply stacking the ZFs cannot reliably recognize longer sequences (Ramirez et al. 2008), which makes using ZFN to target arbitrate sequences difficult.

TAL effector can recognize DNA sequence without the ZFNs' problem. The difficulty lays in the assembly such plasmids, which contain lots of repeats. It takes 34 amino acids TAL motif to recognize a single base pair through contracts with amino acids 12 and 13, known as the repeat variable di-residue (VDR) (Boch et al. 2009).

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR Associated (Cas) system was first discovered in bacteria and archaea (Bhaya et al. 2011; Terns and Terns 2011; Wiedenheft et al. 2012). This system has been reduced down to three core components: a specificity-determining crRNA, an auxiliary trans-activating crRNA (tracrRNA) and the CRISPR-associated nuclease Cas9. The crRNA contains the protospacer sequence, which is the most important sequence because this sequence guides the Cas9 complex to complementary DNA locus in the genome. Additional sequences in the crRNA, allows the hybridization with the tracrRNA to form a crRNA-tracrRNA duplex. A fully functional crRNA-tracrRNA-Cas9 complex is formed when the RNA duplex joins together with nuclease Cas9, which contains domains homologous to both HNH and RuvC endonucleases (Jinek et al. 2012; Nishimasu et al. 2014). Once the complex position in the right loci, a DNA-RNA-protein complex is formed, and the Cas9 begins to cut the double strand DNA. The right loci are defined by the complementary of 20 nucleotides guide sequence in crRNAs RNA and the DNA, as well as the presence of an appropriate protospacer adjacent motif (PAM) sequence immediately downstream of the guide sequence (Jinek et al. 2012). Different Cas9 may have different PAM sequences (Mojica et al. 2009).

The Cas9 protein from Streptococcus pyogenes (SpCas9) requires 5'-NGG-3' PAM sequence (Jinek et al. 2012). To make the three-component CRISPR-Cas9 system easier to use, an artificial linker sequence is introduce to fuse crRNA and tracrRNA into a single chimeric guide RNA (sgRNA) that mimics the natural crRNA-tracrRNA hybrid RNA (Jinek et al. 2012). In this tow-component system, the SpCas9 protein along with the sgRNA could be programmed to cleave any sequence preceding a 5'-NGG-3' PAM sequence in mammalian cells, including sequence after 5'-CCN-3' on the complementary strand (if you using only one strand information). Across the whole human genome, there are about 190,000 specific gRNA-targetable sequences, and around 40.5% exons of genes can be targeted with SpCas9 (Mali et al. 2013).

5.4 CRISPR SCREEN

In the history of biology, genome-wide loss-of-function screenings have provided a wealth of information in diverse cell model, human cells, Drosophila cells, and mouse cells (Berns et al. 2004; Boutros et al. 2004; Carette et al. 2009; Jiang et al. 2009; Rad et al. 2010). Recent advances in CRISPR-Cas9 genome editing technology makes genome-wide loss-of-function screens in human cells possible.

Large-scale genetic screen was previously dominated by RNA interference (RNAi), but recently CRISPR-Cas9 technology has make its way into the field of genetic screen (Shalem et al. 2014; Wang et al. 2014b; Zhou et al. 2014). CRISPR screen usually contains five steps. 1) Design sgRNAs libraries at large scale according the demand of the applications. 2) Generate a pool of sgRNA-expressing lentivirus or other genome integration constructs by array synthesis and library cloning. 3) Generate mutant cell libraries by transfect or transfer these sgRNA-expressing libraries into the cells of interest. 4) Apply positive or negative selections to the mutant library, such as FACS for markers, and drug resistance. 5) Back-tracking the mutations by reading the sgRNAs sequence integrated in the genome, usually with deep-sequencing. The cells with selected mutations will enrich in the destined population thus the sgRNAs carried by the cells will ranks on the top comparing the other by-chance sgRNAs in the population. The genes targeted by the top-ranked sgRNAs are the hits for the screen.

Compared to traditional genetic screen with RNAi, there are several improvements with CRISPR-Cas9 technology. First, CRISPR-Cas9 inactivates genes at the DNA level, while RNAi acts on the mRNA, which might result in incompleteness of protein depletion. Second, CRISPR-Cas9 should enable the screens for non-transcribed or regulatory elements, such as promoters, enhancers, silencers, as well as the intergenic regions, which are normally inaccessible by means of RNAi. Third, mutations caused by the CRISPR-Cas9 are some permanent scars fixed in the genome independent of the state of the cells, while the effects of RNAi are undetectable when the gene complete its mission and stop transcribed. Therefore, with CRISPR-Cas9 technology, it is possible to verify mutations at endpoint of a transition process, such as differentiation, while RNAi effect is difficult to verified at the end point or the beginning. However, there are also some limitations with CRISPR-Cas9 technology, such

as the off-target problem, the different efficiency of knocking out genes. But overall, CRISPR screen is a powerful tool for genome-wild functional genetic interrogation.

The primary assay is the key for a screen. Since the publication of CRISPR screen, numbers of assays have been used in such screens. Survival assay is the simplest, but strictest, assay for CRISPR screen. There are lots of such studies. Shalem et al screened for genes whose loss in involved in resistance to vemurafenib (a therapeutic RAF inhibitor) in melanoma (Shalem et al. 2014). Wang et al screened for resistance to the nucleotide analog 6-thioguanine (6-TG) identified all expected members of the DNA mismatch repair (MMR) pathway (Wang et al. 2014b). Ma et al identified genes essential for West-Nile-Virus-induced cell death (Ma et al. 2015). Birsoy et al Identifies metabolic genes whose loss sensitizes human jurkat cells to phenformin (a mitochondrial complex I inhibitor) (Birsoy et al. 2015). CRISPR screens can also be used in identifying genes for signaling transduction (Parnas et al. 2015; DeJesus et al. 2016; Wang et al. 2017; Pusapati et al. 2018), tumor metastasis (Chen et al. 2015), drug resistant in primary tumor cells (Ruiz et al. 2016), muscle formation (Bi et al. 2017) and cellular metabolism (Arroyo et al. 2016).

Approaching to the cardiogenesis problem, genome-wide screen is an unmet needed. For large-scale screens, *in-vitro* cardiac differentiation is currently the best choice. For any screens, variable control plays a central role during the process. Influence from external factors should be reduced to the minimal. Compared to *in-vivo* system, the *in-vitro* system has much less uncontrollable factors and gives the most lucid readout. Cardiac differentiation from hPSCs closely mimics cardiac development in the embryo and is the ideal model for cardiogenesis *in vitro* (Laflamme and Murry 2011). With hPSCs cardiac differentiation models, not only can we verify the established cardiac disease model, but also genetically explore new processes of cardiogenesis. CRISPR-Cas9 loss-of-function screen technology together with the robust, high efficient GiWi cardiac differentiation protocol make it feasible to carry out genome-wide genetic screen for the important genes for cardiogenesis. Differentiation of hPSCs into the cardiovascular lineages is a multi-step process that involves initial epithelial to mesenchyme transition, mesoderm induction and specification, cardiac specification and differentiation, and functional maturation (Laflamme and Murry 2011). Each intermediate cell can be characterized by the expression of different markers, although the lineage marker tree is in its preliminary state.

In contrast to RNA-seq profiling method which gives a genome-wild expression profile of the specific intermediate cell types, CRISPR screen is aim to discover genes that contributes to the formation of the specific intermediate cell type or directs the fate decision of the progenitors. CRISPR screen is more based on the gene function, making use of the cause-and-effect relationship between the genes and direct by the function of the genes. CRISPR-Cas9 genome-wide screen would be able to provided more wealthy information than just RNA-seq described profiling.

6 AIMS

In order to provide the molecular and cellular mechanism of human cardiogenesis and congenital heart disease, we aim to generate progenitors from hPSCs that can self-assemble into the ventricular muscle tissue and to identify essential genes that control the formation of such progenitors.

Paper I – Aim 1: to generate human ventricular progenitors that can produce the ventricular muscle tissue *in vivo*. Aim 2: to functionally and molecularly characterized these progenitors. Aim 3: explore the therapeutic potential of these progenitors.

Paper II – Aim 1: to establish CRISPR-Cas9 as a tool for investigating cell fate decision in hPSCs. Aim 2: to investigate the roles of β-catenin in early cell fate decision of hPSCs.

Paper III – Aim 1: to examine the role of SMAD4 in human cardiomyocyte differentiation. Aim 2: to identify the earlies signaling pathways in TGFβ superfamily and their essential roles in human cardiogenesis.

Paper IV – Aim 1: to develop genome-wide CRISPR screen for human heart progenitor formation. Aim 2: To investigate role of the identified candidate gene in cell fate decision of early precursor cells.

7 RESULT SUMMARY

7.1 PAPER I: HVP

Human ISL1+ Ventricular Progenitors Self-Assemble into an In Vivo Functional Heart Patch and Preserve Cardiac Function after Infarction

The generation of human ventricular tissue from human pluripotent stem cells not only holds great promise to regenerate a failing heart, but also will serve as valuable model for studying human cardiogenesis. *In-vivo* tissue formation requires a niche that allows the cells to trigger the machinery that produce vascularization and matrix formation cues. Therefore, additional components are usually required, such as synthetic matrices, scaffolds and vascular lineage cells. Moreover, these components, together with the cells, are demanded to assembly themselves correctly into a functional cardiac muscle tissue. In paper I, we applied developmental principles using ventricular progenitors instead of beating cardiomyocytes to generate *in-vivo* human ventricular muscle patch without the assistant of additional matrices and scaffolds.

By using more potent inhibitors of GSK3 and PORCN, we optimized an existing two-step protocol to efficient generation of human ventricular progenitor (HVP) over 6 days. And the stage conversion from pluripotent stem cells to ISL1+ HVPs and subsequent *in-vitro* differentiation were extensively characterized by immunostaining and RNA-sequencing. These analyses demonstrated the distinct markers for each stage: pluripotency stage, *POU5F1*, *NANOG*, and *SOX2*; primitive streak or early mesoderm stage, *TBXT*, *EOMES* and *MIXL1*; cardiac mesoderm stage, *MESP1*, *KDR*, *PDGFRA*; cardiac progenitor or early cardiomyocyte stage, *ISL1*, *NKX2-5*, *MEF2C*; and cardiomyocyte stage, *TNNT2*, *MYH7*, *PLN*, *RYR2*, *MYL7*, *MYL2*. By examining the marker expression, we also define the corresponding time points during the differentiation for these stages. The early mesoderm stage happens on day 1, and quickly transitive to cardiac mesoderm on day 2-3. The cardiac mesoderm cells enter the cardiac progenitor stage on day 5-7. The beating cardiomyocytes can be observed usually on day 7-8 or slight later day 9-10. With these results, we focus on the day 6 progenitor cells, to study and character their properties.

In order to further purified cardiac progenitors from the day 6 cells, surface markers are needed. By analyzing the gene expression patterns in the RNA-Seq dataset, we applied 4 criteria to select candidate marker genes: (1) high day 6 to day 0 ratio; (2) correlation with ISL1 expression pattern; (3) cell surface proteins; and (4) highly expressed at day 6. Five candidates fulfill all the criteria: JAG1, FZD4, FGFR3, LIFR, and TNFSF9. Further live-cell flow cytometry analysis identifies, in these 5 candidates, LIFR is the best surface protein mark for cardiac progenitor on day 6. The LIFR marks more than 87.4% of the cell population on day 6.

Next, we characterized the lineage potency of these day 6 cardiac progenitors and the electrophysiology (EP) properties of the cardiomyocytes derived from them. In the clonal assay, day 6 cells were re-plated and after 3 weeks of differentiation, the cells became cTnI+SMA+, which indicates immature cardiomyocytes, but no VE-cadherin (CD144, CDH5) positive endothelial cells. And on differentiation day 15 and onwards, the majority of the cells were positive for ventricular marker MLC2v (MYL2), cTnT(TNNT2), and NKX2-5, but negative for MLC2a (MYL7), and HCN4. These results demonstrated these day 6 cardiac progenitors are ventricular muscle progenitors, hence named human ventricular progenitors (HVPs). To further characterized the maturation potential of HVPs, we performed patch clamping, optical mapping and calcium imaging of the cardiomyocytes derived from HVPs. Via patch clamp, we showed over 90% of the NKX2-5+ cells had a polarized diastolic membrane potentials and a ventricular-like action potential. And by using optical mapping and calcium imaging, we demonstrated the propagation of both action potential and calcium transient *in-vitro*. This result indicates the network activity and electrical coupling.

By examining the parameters from the EP-studies, there are a couple of EP-parameters indicated the cardiomyocyte derived from HVP are mature. Firstly, the APD90/APD50 ratio is small than 1.2, which recapitulates the AP configuration of native adult human ventricular myocytes. Secondly, comparing with the cardiomyocytes derived from a EB-based protocol and HVPs, the funny current is reduced by half. These features suggestion the maturation of the HVP-derived cardiomyocytes.

To analysis HVPs *in-vivo*, we transplanted the HVPs under the renal capsule in a highly vascularized site. The choice of renal capsule is based on these advantages:1) it is separated from other organs. The transplanted cells would be contained within the capsule. 2) the capsule is highly vascularized, it is easier for the transplanted cells to survive. 3) the retroperitoneal location of the kidney is easy accessed for both the surgery and imaging. In order to further purify the day 6 cells, we negatively sorted out the cells expressing TRA-1-60 (PODXL), a pluripotent stem cell surface marker. Three million TRA-1-60 negative day 6 cells, mainly HVPs, were transplanted under the kidney capsule of immunocompromised NOD scid gamma (NSG) mice (n =30). A patch of xeno tissue was engrafted on the surface of the murine kidney after 2 months. The graft resembled a human ventricular muscle wall, as the immunostaining reveals the graft consisted cTnT and MLC2v positive ventricular myocytes. Furthermore, electron microscopy (EM) showed aligned sarcomeres, with distinct A-band, I-band, M-line and Z-line. The desmosomes were also presence in the EM image of the graft.

To assay the engraftment of these cells, the physical size is measured, which is about .6 cm in length and .2 cm in thickness. More importantly, the graft was vascularized by the host circulation network, as red-lectin, which is injected through the tail of the host, can be found in the graft's vessel. The graft also produced its own extracellular matrix, containing human fibronectin and laminin. In addition, the beating graft can be detected by ultrasound imaging

in vivo. Taken together, the results demonstrated a fully function ventricular muscle tissue is assembled through a cell-autonomous pathway of the HVPs.

Engraftment would require the cells to commit the ventricular myocyte fate and create necessary cues for assembly, and proliferate for scaling-up after the transplantation. If the cells were transplanted at earlier stage, such as day 3, MESP1+ cardiac mesoderm cells, the cells might become other cell types other than ventricular myocytes, which might overwhelm the tissue and turn into tumor-ish graft. When transplanted cells at later stage, such as day 10, NKX2-5+ beating cardiomyocytes, the cells had very limited proliferation capacity and lose the capacity to interface with other tissue to build a necessary niche for the formation of the muscle tissue. The day 3 and 10 cells transplantation experiment demonstrated the unique engraftment property of the HVPs, and a developmental time window for tissue formation.

To further explore the therapeutic potential of the HVPs, we transplanted the HVPs into the hearts of NSG mice, with and without myocardial infarction (MI). Ventricular myocytes can be found in the heart without MI in both short term (2 months) and long term (8 months). The ventricular graft patch was vascularized and integrated with the host myocardium via connexin 43. In the MI hearts, at the 2-month time point, MRI revealed improved ejection fraction, augmented wall thickening, and improved contractility.

In summary, these results demonstrate the HVPs generated from hPSC reassemble the ventricular myogenic progenitors that builds the ventricular muscle tissue during human cardiogenesis. It will be of interest to study the genetic control of these cells, such as their formation, proliferation, and differentiation. In the next phase of the study, we try to identify the essential genes for the formation of HVPs, from the absolute basic to unbiased genome-wide screen.

7.2 PAPER II: B-CATENIN

Interrogating Canonical Wnt Signaling Pathway in Human Pluripotent Stem Cell Fate Decisions Using CRISPR-Cas9

Canonical Wnt/β-catenin signaling pathway is one of the most important signaling pathways in developmental biology. And β-catenin sits in the center of this pathway. In the absent of Wnt, a destruction complex consisting AXIN, APC, and GSK3 binds to and phosphorylate and degraded then β-catenin, preventing the accumulation of β-catenin in the nucleus. When Wnt binds to the Frizzled receptors, the signaling cascade inhibits the destruction complex, allowing the β-catenin gets inside the nucleus and activate targeted genes.

The Wnt signaling has been documented of importance for both the self-renewal and differentiation of hPSCs. The role of Wnt signaling in the hPSC self-renewal is controversial. Mouse embryonic stem cells (mESCs) are common cultured in medium contains GSK3β & Mek1/2 inhibitors, sometimes, but not required to include leukemia inhibitory factor (LIF) (Ying et al. 2008). This demonstrated the self-renewal of ES can be driven by Wnt signaling.

Similar results for human ES were observed in early culture conditions (Sato et al. 2004). On the other hand, β-catenin knock down experiments showed the Wnt signaling is not essential for the self-renewal of hPSCs (Davidson et al. 2012; Lian et al. 2012). Therefore, the essential role of β-catenin in the self-renewal of hPSCs is yet to be determined.

In the GiWi cardiomyocyte differentiation protocol, GSK3 inhibitor were used to inhibit the destruction complex and activate the Wnt targeted genes. However, the GSK3 has nearly 100 proteins proposed to be its substrates, including c-MYC, HIF-1α and β-catenin. The inhibition of GSK3 might have a broader impact on the cells than merely Wnt activation. To find out the essential role of β-catenin in human cardiac differentiation, direct genetic evidence is required.

In this study, we used CRISPR-Cas9 system to generate β-catenin mutated hPSC lines and further utilized these cell lines to study the role of canonical Wnt signaling in hPSC self-renewal and differentiation. We used an engineered nuclease Cas9-GFP and an sgRNA targeting exon 5 of CTNNB1 to mutate β-catenin. We chose one of the β-catenin negative clones for further study. By sequencing the targeted region, we found the first allele had 8bp deletion and the second allele had more than 100bp deletion. This compound heterozygous clone can be expanded for long-term without differentiation in E8 medium and express pluripotency marker Oct4. The success of generating β-catenin mutated hPSC clone demonstrates β-catenin is not required for hPSC maintenance.

Next, we moved on the question that whether β-catenin is required cardiomyocyte differentiation. We induced hPSC differentiation with 6μM CHIR for 24 hours. The mesendoderm marker brachyury (T or TBXT) was strongly expressed in the wild types, while it was undetectable in the mutants. This result demonstrated the CHIR induced mesendoderm differentiation is triggered through β-catenin dependent pathways and the requirement of β-catenin for initiating the mesendoderm differentiation. By using GiWi cardiomyocyte differentiation protocol, the wild type hPSCs yielded more than 90% cTnT+ cells by day 15, while no cTnT+ cells were detected from the mutant hPSCs.

With CRISPR-Cas9 genome editing technology, we provide the first genetic evidence that β-catenin is essential for the mesendoderm formation and subsequence cardiomyocyte differentiation, but is dispensable for human pluripotent stem self-renewal.

7.3 PAPER III: SMAD4

SMAD4 is Essential for Human Cardiac Mesodermal Precursor Cell Formation

Using CIRSPR-Cas9 to target CTNNB1 clarifies the chemical stimulation in the GiWi protocol is indeed through β-catenin. But it wouldn't allow us to see downstream of the chain reaction caused by β-catenin and its targets. TGF-β superfamily signaling pathways, including BMPs, Activin, and Nodal, have been used in growth factor based cardiomyocyte differentiation. In small molecular based differentiation protocol, such as GiWi, these pathways have been left untouched. Pharmacology studies have been showed that these

pathways are required for the differentiation (Lian et al. 2012; Burridge et al. 2014). However, the complex interactions between the WNT/ β-catenin and the TGF-β superfamily signaling pathways make these studies a bit difficult to interpret in the contest. A genetic study would clarify the essential role of the TGF-β superfamily signaling pathways in cardiac differentiation.

There are more than 30 ligands, 4 type-II and 7 type-I receptors in TGF-β superfamily signaling pathways. The complexity of singling pathways required a better way to target them. The Ligands of the TGF-β superfamily binds to their receptors and caused the phosphorylation of the R-Smads (SMAD1/5/8 and SMAD2/3). The phosphorylated R-SMAD(s) then binds the common SMAD, SMAD4, and transported into the nucleus to active or suppress the targeted genes. Therefore, by targeting SMAD4, the whole signaling pathway family would be impaired.

When targeting a gene, especially the ones expressed in hPSCs, is hard to predict whether or not it will affect the self-renewal and pluripotency of the hPSCs. The status of the hPSCs must be checked first. In our cases, the hPSCs were cultured in E8 medium, which contains TGF-β1 as one of the essential components. And numbers of studies have suggested that Nodal signaling pathway can enhance the expression of NANOG, which supports the pluripotency of the hPSCs. However, there is a study showed that the knockdown of SMAD4 in hESCs would not induce rapid differentiation, just has some trouble in long term culture (Avery et al. 2010). Thus, whether SMAD4 is essential for the self-renewal and pluripotency remains a question.

In this study, we used *PiggyBac* CRISPR-Cas9 system to generate *SMAD4* mutated hPSC lines. The successful of generating single-cell derived cell lines demonstrated the SMAD4 is not required for the self-renewal of hPSCs. By immunostaining, we demonstrated these *SMAD4* mutated cell lines express pluripotency marker POU5F1, NANOG, SOX2. Furthermore, RNA-Seq analysis revealed these pluripotency markers were not differentially expressed between the wild type and mutant hPSCs, while the TGF-β signaling and target genes including *LEFTY1*, *LEFTY2* and *CER1* were down regulated in mutant hPSCs. Taken together, we demonstrated that SMAD4 is not required for hPSC self-renewal.

Since the *SMAD4* mutated hPSCs maintains their pluripotent identify, we next test whether SMAD4 is essential for cardiomyocyte differentiation. We differentiated the wild type and mutant hPSCs, using both the small molecule based and growth factor based differentiation methods. In both cases, the *SMAD4* mutants yield zero detectable TNNT2+ cardiomyocytes, while the wild type cells yielded more than 80% of cardiomyocyte with GiWi method, and 40% with Activin A/BMP4 method. We further re-expressed SMAD4 in the mutant cells, and these cells restored their cardiomyocyte differentiation capacity. These results suggest SMAD4 is absolutely essential for cardiomyocyte differentiation.

To find out how the SMAD4 mutant hPSCs fail to generate cardiomyocytes, we performed transcriptome profiling at different stages during the differentiation: day 0

pluripotent stem cells stage, day 1 primitive streak and early mesoderm stages, day 3 cardiac mesoderm stage, and day 6 progenitor and early cardiomyocyte stages. From the transcriptome profile, we found that the major difference begins at the cardiac mesoderm, suggesting the failure of generating cardiac mesodermal precursor cells. And further immunostaining showed MESP1+ cardiac mesoderm precursors were missing from the day 3 mutant cells. And we further examined the ligands of TGFβ superfamily. It turned out that NODAL was the only active signal in the TGFβ superfamily, instead of Activin A or BMPs, which we used to induce cardiomyocytes differentiation. On the other hand, from the developmental view, NODAL and WNT are the major signals that establish the axis of embryos (Tam and Loebel 2007; Arnold and Robertson 2009). By detail analysis, the differential expression of the wild type and mutant cells on day 1, we found the expression of primitive streak genes were present but much weaker than the wild type cells, while the expression of pro-neuroectoderm genes were higher in the mutants.

Taken together, these results suggested the distinct the role of the WNT/β-catenin signaling and the NODAL/SMAD4 signaling in the earliest step of human cardiogenesis. The WNT/β-catenin signaling can active the cardiogenesis relative genes, which largely in the primitive streak (PS) genes set, and this process can happen in both wild type and the mutants. The activation process is independent of NODAL/SMAD4 signaling. However, the role NODAL/SMAD4 signaling play in the process is to enhance the expression of the PS genes to the level that allowing the formation of mesendoderm lineage, otherwise the cells, even with the activated PS genes, would fall back the default neuroectoderm fate.

7.4 PAPER IV: SCREEN-ZIC2

Genome-Wide CRISPR Screen Identifies ZIC2 As an Essential Gene That Controls the Fate of Mesodermal Precursors to Human Heart Progenitors

After validating the β-catenin and SMAD4 as two of the essential genes for the formation of human heart progenitors, we sought out to development a new technology to unbiasedly screen for the essential genes across the whole genome. In this study, we developed a two-stage genome-wide CRISPR screen using hPSC system.

First, we created three hPSC libraries carrying gene mutations covering more than 6000 genes. Then, we combined these three libraries to ensure the diversity of the libraries and differentiate the libraries into cardiac progenitors. By monitoring two key cardiac markers: MESP1 and ISL1, we were able to assay the formation of cardiac mesodermal precursors and cardiac progenitors respectively. Cells with essential gene mutated would fail to express these markers and enrich in the MESP1 or ISL1 negative population. Therefore, by comparing the sgRNA enrichment in the marker negative and positive population, we can identify the corresponding gene as the potential candidates.

From the screen output, we compiled a list of 15 candidate genes. After validating 7 of these, we identified *ZIC2* as a new essential gene for cardiac progenitor formation. All three *ZIC2* mutated hPSC lines produced less than 10% of cardiomyocytes. Via immunostaining,

we found the *ZIC2* mutated cells failed to express MESP1 at the cardiac mesoderm formation stage. And surprisingly, the *ZIC2* mutated cells also express ISL1 on day 6, but there were few cardiomyocytes produced in the end. These results suggest ISL1 can be active by a MESP1 independent pathway, but the myogenic ISL1 progenitors are from the MESP1 lineage.

To further study the essential role of *ZIC2* in human cardiogenesis, we used RNA-Seq to compare the transcriptome difference at different stages: day 0, the epiblast stage, day 1, primitive streak, day 2, early mesoderm, day 3, cardiac mesoderm and day 6 cardiac progenitor. The mutant and wild cells follow a similar differentiation trajectory, with further apart at the later stage. And the differential expression (DE) analysis reveals the DE genes are increasing from 50 on day 1, to 293 on day 6. By using the classical cardiac differentiation marker genes, we found the mutant cells were able to become mesoderm precursors, but failed to turn into cardiac progenitors. And the up-regulated genes in the mutant cells formed a network regulating the skeleton development. Among these up-regulated genes, there are a number of osteoblast marker genes, such as *RUNX2*, *DCN*, and *OGN*, while the cardiac genes, such as *NKX2-5*, *MEF2C*, *PLN*, and *MYH6*, were down-regulated. These results suggest the essential role of ZIC2 in hPSC cardiac differentiation is to regulate the cell fate decision of mesodermal precursor to cardiac progenitor. Under the GiWi differentiation condition, the *ZIC2* mutated hPSCs switch to osteoblast fate from the cardiac progenitor fate.

Two functions of ZIC2 have been reported. First, ZIC2 can bind directly to TCF4 and inhibit the transcriptional activity of the β-catenin/TCF4 complex (Pourebrahim et al. 2011). Second, ZIC2 can also function with Mbd3/NuRD in regulating the chromatin state and transcriptional output of genes linked to differentiation (Luo et al. 2015). Based on these previous studies, the mechanisms we proposed here is the inhibitory function of ZIC2, which allows the more specific differentiation molecular program correctly happen. During gastrulation, cells were migrating through the primitive streak and forms all types of mesodermal precursor cells. The activation signal is overwhelming. Without a strong inhibitory system, different differentiation processes were all activated and compete to take control of the cells. The inhibitory system can dampen lots of spiking signaling and allowing the primary signals to convert the cells into right progenitors.

8 DISCUSSION

8.1 THE UNIQUE ENGRAFTMENT PROPERTY OF HVP

Besides the committing to ventricular myocyte fate, engraftment is one of the distinct features the HVPs. As mention previous, engraftment is a complicate process, involving cell survival, migration, proliferation, differentiation, maturation, integration, electrical coupling, and vascularization. The HVPs resemble the *in-vivo* ventricular myocyte wall builder. Their genetically build-in program can accomplish lots of these engraftment tasks naturally, such as proliferation, differentiation, matrix formation, vessel recruitment, and self-assembly. The proliferation capacity of HVPs enables the cells to expand and differentiate from a few million HVPs to, perhaps, a billion of ventricular myocytes. The strong bias towards ventricular myocyte in differentiation generates specifically ventricular muscular tissue instead of a mixture of other types cardiomyocytes. The self-secreting matrix protein provides the mechanism support and signaling niche for the maturation and integration of the ventricular myocytes. And the recruitment of host vessel allows the tissue to survive and expand into a 3D functional, force-generation ventricular myocyte tissue. And the kidney capsule HVP transplantation clearly demonstrated a cell-autonomous pathway, through which the HVPs can generate a 3D vascularized, functional ventricular wall without the assistant of external cues.

This unique engraftment property of HVPs resides in the cell autonomous pathway that empower them. With RNA-Seq data, we have the transcriptome profile of the HVPs, which offers the chance to identify the molecular modules for this engraftment pathway, such as the cell fate restriction, matrix secretion, and vessel recruitment. By comparing the profiles of the engraftable and un-engraftable cell, the modules should stand out for themselves. And the time window, we have defined in the HVP paper, have provide a comparison guide line. However, due to the dramatically change of cell identify, lots of genes were differentially expressed. As all these molecular modules pool together, all we can conclude is that the progenitor genes are important the engraftment. Though a few interesting genes are turn up, such as, *LIFR*, *FZD4*, *LAMMA1*, *VEGFA*, further studies are required to investigate their roles in engraftment. Meanwhile, the surface markers we identified according to their expression patterns, have the potential to further enrich HVPs from the day 6 cell population, which might improve the engraftment.

Other aspects of engraftment, such as mechanical and electrical integration, might not be a build property of HVP. For that, other interventions should be employed. These challenges are the obstacles have to be overcome, before the use of HVP for regenerating the heart. By studying the details molecular pathways of the HVPs, we should one day be able to pre-program these HVPs to taken on new property, like resolving scar tissue, turning fully mature.

8.2 A PEEK INTO THE MOLECULAR COMPLEXITY OF CARDIOGENESIS

For a decade and half, efforts have been focus on improving the cardiomyocyte differentiation protocols, not so much on the molecular mechanism how the cells become cardiomyocytes. Since the development of robust differentiation protocols, the mechanistic studies of the differentiation program have become more of interest. One of such example is Rao et al. 2016. In Rao's study, EOMES is proposed to be the driving force to suppress SOX2, which intends to direct the hPSC towards neuronal fate and later study (Pfeiffer et al. 2018) showed by controlling the expression of EOMES can indeed induce cardiac differentiation.

In our first study, the HVP project, we have documented a detail transcriptome profiles of the differentiation process. There are around 4K genes, 20% of the protein coding genes, are differentially expressed during the whole cardiac differentiation process. Most genes required for initiating the differentiation program were activated at very early stage. Lots of genes were turned on in day 2 samples. Another thing we learn from the expression pattern is that not only the activation is important, the gene inactivation play equally important roles, as lots of genes were shut down on day 4 and 6. Another interesting in the profiles are the distinct stages can be obtain automatically via clustering the expression, and these distinct stages corresponding to the developmental stages. These genome-wide description gives a very rough picture how the program is executed, and similar dataset can also be found in other studies (Piccini et al. 2016; Tompkins et al. 2016).

The cardiac lineage arises from the cardiac mesoderm and the cardiac mesoderm comes from the mesoderm cells migrated though the primitive streak. However, the diversity of the mesoderm cells paints a complicate developmental landscape (Loh et al. 2016). These cells can give rise to multiple lineages, including, paraxial mesoderm and lateral mesoderm. Subsequently, the paraxial mesoderm give rise to early somite and then differentiated into dermomyotome and sclerotome et al. And the lateral mesoderm can become cardiac mesoderm and limb bud mesoderm. Therefore, the mesoderm cells have lots of different cell fate choices, and their decisions are the combination of the signaling cue from the surrounding and the genetic programs build-in these cells. By varying the signaling cues, the mesoderm cells can be directed to a given lineage. On the other side of the question, given an optimized signaling cue, what is the genetic program that locks the fate of these mesodermal cells is equally interesting.

To define this genetic program, the first step is to identify the essential component in this program. In the β-catenin study, we showed β-catenin is the key switch for initializing the differentiation program. Upon stimulation, the *CTNNB1* mutant cells didn't express even the first stage, primitive streak (PS) stage, maker *TBXT*. And these mutants didn't produce any cardiomyocyte in the end. In the *SMAD4* mutants, primitive streak stage can be reached. PS markers, such as *TBXT*, *MIXL1*, *EOMES*, did expressed in the day 1 *SMAD4* mutant cells, but just not as high as the wild type cells. And previous study show β-catenin can activate both primitive streak and neural crest genes (Funa et al. 2015) during the GSK3 inhibition in

hPSCs. And our RNA-Seq result had similar result in the mutant cells, but not in the wild type cells. This suggests the β-catenin is able to trigger the expression of the PS genes, and is independent of NODAL/SMAD4, but the NODAL/SMAD4 pathway can enhance the expression of PS genes. Without this enhancement, the cell just cannot enter the next phase of differentiation, namely cardiac mesoderm formation, and eventually drop back to the neuroectoderm fate. In the *ZIC2* study, with both WNT/β-catenin and NODAL/SMAD4 signaling pathways intact, the *ZIC2* mutant cells were able to enter the mesoderm cells, but most of them fail to commit the cardiac lineage. Unlike β-catenin and SMAD4, ZIC2 is not known to link to a specific signaling pathway. Two studies reported ZIC2 interacts with β-catenin/TCF4 and Mbd3/NuRD. Both interactions attenuate the transcriptional activity. And our RNA-Seq data supports this notion, as there are more active gene in the *ZIC2* mutant cells comparing to the wild type cells. We hypothesis that ZIC2 functions as a key component of a gene suppressor complex, attenuating the transcriptional activity after the β-catenin activation and NODAL/SMAD4 enhancement. Most of the *ZIC2* mutant cells were misled by other signaling into another mesodermal fate, such as osteoblast. Taken together, the basic molecular picture of the cardiac differentiation is: 1) β-catenin enters the nuclear and activates both PS and neuroectoderm genes, including NODAL. 2) the NODAL/SMAD4 cascade, allows the SMAD complex translocate into the nuclear and enhance the PS gene and direct the cells at PS stage to the mesoderm fate. 3) the ZIC2 repressor complex attenuate the transcription activity after the mesoderm precursors are formed, leaving the cell fate decision to the external signaling cues.

In hPSC cardiomyocyte differentiation, the first stage stimulation plays a crucial role. The strength and duration of stimulation often determines the efficient of the differentiation. Our studies have provided some information about the genetic program of this crucial step. This step is corresponding to the gastrulation and mesoderm formation steps *in-vivo*, when the mesodermal precursor begins for emerge. Some of these mesodermal precursors later commit to cardiac mesoderm, which give rise cardiac progenitors. To form a healthy heart, these progenitors are demanded to correctly completely complicate tasks, such as the migration, proliferation, differentiation, and assembly et al. Mistakes in these progenitors are the potential causes of many congenital heart diseases.

9 CONCLUSION

More and more advance techniques have been adapted to study congenital heart disease. Our knowledge about human genome is growing faster than ever before. More information from congenital heart disease will soon be available in the databases. In the research field, the framework we choose to integrate these clinical results would be the key to advance the whole field. The progenitor theory framework operates above the molecular interaction at the cellular level, focusing on the cells instead individual gene. However, the difficulty is how we identify the progenitors. Single gene marker is not always faithful. Through evolution, many key genes have been used and re-used multiple times, therefore a unique marker gene rarely exists. Single-cell biology provides whole-genome profile, but it is always difficult to used comparing to a marker gene. To get a clear picture of the cardiac progenitors, a better approach is needed to define and assay these heart-builders.

CHD, mice, hPSCs, three models, each has its own advantages and limitations. Observations from the clinical are valuable but we can do only little more than just keeping records, which limits our options in testing hypothesis. The mice model provides us great insights. However, the difference between mice and human mean that some process in human can never be observed in mice. The hPSC model is an *in-vitro* system. We are unsure the faithfulness of the system and how human cardiogenesis really happen *in vivo*. A new cardiogenesis system is needed to propel the next break-through in human cardiogenesis. The HVP project has put forward a new organ-on-organ system to study the human ventricular muscle tissue formation, at both cellular and molecular level. Powerful genome editing tools allows us to directly interrogate the critical role the gene plays in the progenitor. Our study has advanced the use of hPSC model to understanding the process of cardiogenesis in human.

ACKNOWLEDGEMENTS

Making things happen is our way to contribute to the society and the planet. To make thing under your own control happen, all you need is just some self-discipline. Sometimes is easy, sometimes is hard. But the truly valuable thing is always the work of many. It cannot be imagined to be accomplished by any single individual. Here, I gratefully thank everyone who have directly or indirectly helped me in the past few years.

Special thanks to **Kenneth Chien**, my supervisor. You are such a legend. The things you have accomplished in the past few years amazed me. After more than 38 years in the science community, I guess you have reached another level. A good lab is not only about fund raising and paper publishing, but also, and more importantly, making things happen. To realize the value of the project demands to take it beyond just words on the paper. And you did it, taking numbers of projects to the real word. It is always my honor to be one of your students. I have truly learned a lot form you. The most important one may be criticism, which has enlightened me to critically think about the work done by myself and others. To understand something, criticizing it is a quick way to do it. And thank you for giving me the opportunity few years ago. It has been such a fun journey.

Thank you, my co-supervisor, **Xiaojun (Lance) Lian**. Thanks for teaching me every basic little thing I need to know to start the hPSC research and showing the career path of becoming an independent assistant professor. Your ability to dig up the details is very impressive. You know every chemical component in the medium. You can read a paper in a few minutes and adopt their new techniques in our projects. Finally, your cardiac differentiation protocol is destined to change the whole field.

Thank you, **Urban Lendahl**, my co-supervisor. From our first hand-shake on my first day in KI, I know you will always be there when I need help. You have made me feel proud of being an CMBer.

Thanks to all the past and present members of Chien Lab. Thank you, **Xiaobing He** for taking such a good care of the daily laboratory operation. Setting up the old lab, and moving to the new, your effort can be found in every tangible item in the lab. We are very lucky to have you as our lab manager. **Kristine Bylund**, thank you for letting us mess up your house for so many times. It is really fun to party in your home. And thanks to the secretary group, **Katarina Drakenberg**, **Jay Donovan**, and **Raffaella Giugliano**, your excellent work has made so many things possible. Thanks, too, to the hard-working technicians, **Jesper Sohlmér**, **Kalaiselvan Krishnan**, **Tamara Szattler**, **Hasina Nasser**, **Chimezie Harrison Umeano**, **Pedro Sousa**.

Thank you to my fellow students. **Chikai Zhou**, your excellent skill in creating genetic engineered mice has empower lots of our projects. **Yat Long (Sunny) Tsoi**, your remarkable skill in designing complex plasmids has impressed me a lot. **Mimmi Mononen**, thanks for

helping me with the single cell sequencing. **Eduarde Rohner** and **Christopher Yen**, thanks for being such wonderful companions around. And thank you to our guest students, **Whitney Ching**, **Edvin Porovic**, **Minghui Li**, **Katrin Mangold,** and **Jingcheng Zhao**. We have lots of good time together.

Big thanks to the post-doc fellows, the key players in the lab. **Kylie Foo**, thanks for taking good care of the HVP project after Lance left, helping me with those transplantation experiments, and editing the ZIC2 manuscript. **Federica Santoro**, **Nevin Witman**, thank you for creating such a joyful lab atmosphere. You have made the Chien lab a very happy place to do research. **Ran Yang**, and **Yao Xiao**, my old PKU alumnus, it felt good to whine about the past and thank you for helping me with lots of experiments. **Makoto Sahara**, **Alexander Goedel**, **Elif Eroglu**, and **Chuen Yan Leung**, thank you for proof reading those manuscripts and giving many good suggestions. And to **Miia Lehtinen**, **Jonathan Clarke**, **Gianluigi Pironti**, **Erwin De Genst**, **Wei Fu**, and **Alimujiang Fulati**, you have made the Chien lab a wonderful place to work in.

Thank you to the collaborators of Chien lab. Thanks **Peter Gruber** for the intense editing the SMAD4 manuscript and insightful inputs for the ZIC2 manuscript. **Shao-Bo Jin**, thank you for sharing your experience with me. That finance thing you taught me, may be one of the most rewarding things I learn in the past few years. **Javier Avila-Cariño**, thank you for helping me with FACS experiments. Without you, the CMB FACS facility is never the same again. And also thanks to the collaborators in William Louch’s lab, Oslo and Ronald Li’s, HKU, who have made HVP project a better work.

Thanks to the CMB staff, specially **Matti Nikkola**, **Linda Lindel**, and **Lina Pettersson**, thank you for making the department such a wonderful place to work.

Finally, my big thanks to my parents and my sister. 我要衷心感谢我父母在一个改革与发展的时代，面对生活中的变数，含辛茹苦将我们抚养大，送我们上好学校。感谢你们给予我的鼓励和支持。姐姐，感谢你照顾爸妈和多年来对我的关心与帮助。

REFERENCES

Ahlgren U, Pfaff SL, Jessell TM, Edlund T, Edlund H. 1997. Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells. *Nature* **385**: 257-260.

Anastassiadis K, Fu J, Patsch C, Hu S, Weidlich S, Duerschke K, Buchholz F, Edenhofer F, Stewart AF. 2009. Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E. coli, mammalian cells and mice. *Dis Model Mech* **2**: 508-515.

Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S. 2004. Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nat Rev Genet* **5**: 725-738.

Arnold SJ, Robertson EJ. 2009. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat Rev Mol Cell Biol* **10**: 91-103.

Arroyo JD, Jourdain AA, Calvo SE, Ballarano CA, Doench JG, Root DE, Mootha VK. 2016. A Genome-wide CRISPR Death Screen Identifies Genes Essential for Oxidative Phosphorylation. *Cell Metab* **24**: 875-885.

Avery S, Zafarana G, Gokhale PJ, Andrews PW. 2010. The role of SMAD4 in human embryonic stem cell self-renewal and stem cell fate. *Stem Cells* **28**: 863-873.

Barth E, Stammler G, Speiser B, Schaper J. 1992. Ultrastructural quantitation of mitochondria and myofilaments in cardiac muscle from 10 different animal species including man. *J Mol Cell Cardiol* **24**: 669-681.

Basson CT, Bachinsky DR, Lin RC, Levi T, Elkins JA, Soults J, Grayzel D, Kroumpouzou E, Traill TA, Leblanc-Straceski J et al. 1997. Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet* **15**: 30-35.

Basson CT, Cowley GS, Solomon SD, Weissman B, Poznanski AK, Traill TA, Seidman JG, Seidman CE. 1994. The clinical and genetic spectrum of the Holt-Oram syndrome (heart-hand syndrome). *N Engl J Med* **330**: 885-891.

Bauer RC, Laney AO, Smith R, Gerfen J, Morrissette JJ, Woyciechowski S, Garbarini J, Loomes KM, Krantz ID, Urban Z et al. 2010. Jagged1 (JAG1) mutations in patients with tetralogy of Fallot or pulmonic stenosis. *Hum Mutat* **31**: 594-601.

Bauwens CL, Peerani R, Niebruegge S, Woodhouse KA, Kumacheva E, Husain M, Zandstra PW. 2008. Control of human embryonic stem cell colony and aggregate size heterogeneity influences differentiation trajectories. *Stem Cells* **26**: 2300-2310.

Beltrami AP, Barlucchi L, Torella D, Baker M, Limana F, Chimenti S, Kasahara H, Rota M, Musso E, Urbanek K et al. 2003. Adult cardiac stem cells are multipotent and support myocardial regeneration. *Cell* **114**: 763-776.

Bergmann O, Bhardwaj RD, Bernard S, Zdunek S, Barnabe-Heider F, Walsh S, Zupicich J, Alkass K, Buchholz BA, Druid H et al. 2009. Evidence for cardiomyocyte renewal in humans. *Science* **324**: 98-102.

Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B et al. 2004. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**: 431-437.

Bhaya D, Davison M, Barrangou R. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* **45**: 273-297.

Bi P, Ramirez-Martinez A, Li H, Cannavino J, McAnally JR, Shelton JM, Sanchez-Ortiz E, Bassel-Duby R, Olson EN. 2017. Control of muscle formation by the fusogenic micropeptide myomixer. *Science* **356**: 323-327.

Birsoy K, Wang T, Chen WW, Freinkman E, Abu-Remaileh M, Sabatini DM. 2015. An Essential Role of the Mitochondrial Electron Transport Chain in Cell Proliferation Is to Enable Aspartate Synthesis. *Cell* **162**: 540-551.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**: 1509-1512.

Boutros M, Kiger AA, Armknecht S, Kerr K, Hild M, Koch B, Haas SA, Paro R, Perrimon N, Heidelberg Fly Array C. 2004. Genome-wide RNAi analysis of growth and viability in Drosophila cells. *Science* **303**: 832-835.

Brade T, Pane LS, Moretti A, Chien KR, Laugwitz KL. 2013. Embryonic heart progenitors and cardiogenesis. *Cold Spring Harb Perspect Med* **3**: a013847.

Breckpot J, Thienpont B, Arens Y, Tranchevent LC, Vermeesch JR, Moreau Y, Gewillig M, Devriendt K. 2011. Challenges of interpreting copy number variation in syndromic and non-syndromic congenital heart defects. *Cytogenet Genome Res* **135**: 251-259.

Breckwoldt K, Letuffe-Breniere D, Mannhardt I, Schulze T, Ulmer B, Werner T, Benzin A, Klampe B, Reinsch MC, Laufer S et al. 2017. Differentiation of cardiomyocytes and generation of human engineered heart tissue. *Nature protocols* **12**: 1177-1197.

Bruneau BG. 2008. The developmental genetics of congenital heart disease. *Nature* **451**: 943-948.

Bruneau BG, Nemer G, Schmitt JP, Charron F, Robitaille L, Caron S, Conner DA, Gessler M, Nemer M, Seidman CE et al. 2001. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* **106**: 709-721.

Bu L, Jiang X, Martin-Puig S, Caron L, Zhu S, Shao Y, Roberts DJ, Huang PL, Domian IJ, Chien KR. 2009. Human ISL1 heart progenitors generate diverse multipotent cardiovascular cell lineages. *Nature* **460**: 113-117.

Budde BS, Binner P, Waldmuller S, Hohne W, Blankenfeldt W, Hassfeld S, Bromsen J, Dermintzoglou A, Wieczorek M, May E et al. 2007. Noncompaction of the ventricular myocardium is associated with a de novo mutation in the beta-myosin heavy chain gene. *PloS one* **2**: e1362.

Burridge PW, Anderson D, Priddle H, Barbadillo Munoz MD, Chamberlain S, Allegrucci C, Young LE, Denning C. 2007. Improved human embryonic stem cell embryoid body homogeneity and cardiomyocyte differentiation from a novel V-96 plate aggregation system highlights interline variability. *Stem Cells* **25**: 929-938.

Burridge PW, Matsa E, Shukla P, Lin ZC, Churko JM, Ebert AD, Lan F, Diecke S, Huber B, Mordwinkin NM et al. 2014. Chemically defined generation of human cardiomyocytes. *Nat Methods* **11**: 855-860.

Cai CL, Liang X, Shi Y, Chu PH, Pfaff SL, Chen J, Evans S. 2003. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev Cell* **5**: 877-889.

Cai CL, Zhou W, Yang L, Bu L, Qyang Y, Zhang X, Li X, Rosenfeld MG, Chen J, Evans S. 2005. T-box genes coordinate regional rates of proliferation and regional specification during cardiogenesis. *Development* **132**: 2475-2487.

Cai W, Albini S, Wei K, Willems E, Guzzo RM, Tsuda M, Giordani L, Spiering S, Kurian L, Yeo GW et al. 2013. Coordinate Nodal and BMP inhibition directs Baf60c-dependent cardiomyocyte commitment. *Genes Dev* **27**: 2332-2344.

Camus A, Perea-Gomez A, Moreau A, Collignon J. 2006. Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. *Dev Biol* **295**: 743-755.

Carette JE, Guimaraes CP, Varadarajan M, Park AS, Wuethrich I, Godarova A, Kotecki M, Cochran BH, Spooner E, Ploegh HL et al. 2009. Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**: 1231-1235.

Chabab S, Lescroart F, Rulands S, Mathiah N, Simons BD, Blanpain C. 2016. Uncovering the Number and Clonal Dynamics of Mesp1 Progenitors during Heart Morphogenesis. *Cell Rep* **14**: 1-10.

Chan PKW, Geng L, Gao Y, Keung W, Li RA. 2017. AAV-mediated conversion of human pluripotent stem cell-derived pacemaker. *Biochem Biophys Res Commun* **494**: 346-351.

Chen G, Gulbranson DR, Hou Z, Bolin JM, Ruotti V, Probasco MD, Smuga-Otto K, Howden SE, Diol NR, Propson NE et al. 2011. Chemically defined conditions for human iPSC derivation and culture. *Nat Methods* **8**: 424-429.

Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, Conlon FL, Wang DZ. 2006. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet* **38**: 228-233.

Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R et al. 2015. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**: 1246-1260.

Ching YH, Ghosh TK, Cross SJ, Packham EA, Honeyman L, Loughna S, Robinson TE, Dearlove AM, Ribas G, Bonser AJ et al. 2005. Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat Genet* **37**: 423-428.

Chiu RC, Zibaitis A, Kao RL. 1995. Cellular cardiomyoplasty: myocardial regeneration with satellite cell implantation. *Ann Thorac Surg* **60**: 12-18.

Chong JJ, Yang X, Don CW, Minami E, Liu YW, Weyers JJ, Mahoney WM, Van Biber B, Cook SM, Palpant NJ et al. 2014. Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts. *Nature* **510**: 273-277.

Chu GC, Dunn NR, Anderson DC, Oxburgh L, Robertson EJ. 2004. Differential requirements for Smad4 in TGFbeta-dependent patterning of the early mouse embryo. *Development* **131**: 3501-3512.

Cinquetti R, Badi I, Campione M, Bortoletto E, Chiesa G, Parolini C, Camesasca C, Russo A, Taramelli R, Acquati F. 2008. Transcriptional deregulation and a missense mutation define ANKRD1 as a candidate gene for total anomalous pulmonary venous return. *Hum Mutat* **29**: 468-474.

Cong L, Ran FA, Cox D, Lin SL, Barretto R, Habib N, Hsu PD, Wu XB, Jiang WY, Marraffini LA et al. 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**: 819-823.

Costello I, Pimeisl IM, Drager S, Bikoff EK, Robertson EJ, Arnold SJ. 2011. The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nature cell biology* **13**: 1084-1091.

Davidson KC, Adams AM, Goodson JM, McDonald CE, Potter JC, Berndt JD, Biechele TL, Taylor RJ, Moon RT. 2012. Wnt/beta-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4. *Proc Natl Acad Sci U S A* **109**: 4485-4490.

Davis RP, Casini S, van den Berg CW, Hoekstra M, Remme CA, Dambrot C, Salvatori D, Oostwaard DW, Wilde AA, Bezzina CR et al. 2012. Cardiomyocytes derived from pluripotent stem cells recapitulate electrophysiological characteristics of an overlap syndrome of cardiac sodium channel disease. *Circulation* **125**: 3079-3091.

Davis RP, Ng ES, Costa M, Mossman AK, Sourris K, Elefanty AG, Stanley EG. 2008. Targeting a GFP reporter gene to the MIXL1 locus of human embryonic stem cells identifies human primitive streak-like cells and enables isolation of primitive hematopoietic precursors. *Blood* **111**: 1876-1884.

DeJesus R, Moretti F, McAllister G, Wang Z, Bergman P, Liu S, Frias E, Alford J, Reece-Hoyes JS, Lindeman A et al. 2016. Functional CRISPR screening identifies the ufmylation pathway as a regulator of SQSTM1/p62. *Elife* **5**.

Den Hartogh SC, Schreurs C, Monshouwer-Kloots JJ, Davis RP, Elliott DA, Mummery CL, Passier R. 2015. Dual reporter MESP1 mCherry/w-NKX2-5 eGFP/w hESCs enable studying early human cardiac differentiation. *Stem Cells* **33**: 56-67.

Dolnikov K, Shilkrut M, Zeevi-Levin N, Gerecht-Nir S, Amit M, Danon A, Itskovitz-Eldor J, Binah O. 2006. Functional properties of human embryonic stem cell-derived cardiomyocytes: intracellular Ca2+ handling and the role of sarcoplasmic reticulum in the contraction. *Stem Cells* **24**: 236-245.

Domian IJ, Chiravuri M, van der Meer P, Feinberg AW, Shi X, Shao Y, Wu SM, Parker KK, Chien KR. 2009. Generation of functional ventricular heart muscle from mouse ventricular progenitor cells. *Science* **326**: 426-429.

Dymecki SM, Tomasiewicz H. 1998. Using Flp-recombinase to characterize expansion of Wnt1-expressing neural progenitors in the mouse. *Dev Biol* **201**: 57-65.

Elliott DA, Braam SR, Koutsis K, Ng ES, Jenny R, Lagerqvist EL, Biben C, Hatzistavrou T, Hirst CE, Yu QC et al. 2011. NKX2-5(eGFP/w) hESCs for isolation of human cardiac progenitors and cardiomyocytes. *Nat Methods* **8**: 1037-1040.

Emanuel BS. 2008. Molecular mechanisms and diagnosis of chromosome 22q11.2 rearrangements. *Dev Disabil Res Rev* **14**: 11-18.

Fahed AC, Gelb BD, Seidman JG, Seidman CE. 2013. Genetics of congenital heart disease: the glass half empty. *Circ Res* **112**: 707-720.

Fahed AC, Nemer GM. 2012. Genetic Causes of Syndromic and Non-Syndromic Congenital Heart Disease. In *Mutations in Human Genetic Disease*, doi:10.5772/48477.

Feil R, Wagner J, Metzger D, Chambon P. 1997. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun* **237**: 752-757.

Fernandes S, Naumova AV, Zhu WZ, Laflamme MA, Gold J, Murry CE. 2010. Human embryonic stem cell-derived cardiomyocytes engraft but do not alter cardiac remodeling after chronic infarction in rats. *J Mol Cell Cardiol* **49**: 941-949.

Funa NS, Schachter KA, Lerdrup M, Ekberg J, Hess K, Dietrich N, Honore C, Hansen K, Semb H. 2015. beta-Catenin Regulates Primitive Streak Induction through Collaborative Interactions with SMAD2/SMAD3 and OCT4. *Cell Stem Cell* **16**: 639-652.

Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, Rothrock CR, Eapen RS, Hirayama-Yamada K, Joo K et al. 2003. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* **424**: 443-447.

Garg V, Muth AN, Ransom JF, Schluterman MK, Barnes R, King IN, Grossfeld PD, Srivastava D. 2005. Mutations in NOTCH1 cause aortic valve disease. *Nature* **437**: 270-274.

Gebbia M, Ferrero GB, Pilia G, Bassi MT, Aylsworth A, Penman-Splitt M, Bird LM, Bamforth JS, Burn J, Schlessinger D et al. 1997. X-linked situs abnormalities result from mutations in ZIC3. *Nat Genet* **17**: 305-308.

Goldmuntz E, Bamford R, Karkera JD, dela Cruz J, Roessler E, Muenke M. 2002. CFC1 mutations in patients with transposition of the great arteries and double-outlet right ventricle. *American journal of human genetics* **70**: 776-780.

Gorabi AM, Hajighasemi S, Tafti HA, Atashi A, Soleimani M, Aghdami N, Saeid AK, Khori V, Panahi Y, Sahebkar A. 2018. TBX18 transcription factor overexpression in human-induced pluripotent stem cells increases their differentiation into pacemaker-like cells. *J Cell Physiol* doi:10.1002/jcp.27018.

Gordon JW, Scangos GA, Plotkin DJ, Barbosa JA, Ruddle FH. 1980. Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc Natl Acad Sci U S A* **77**: 7380-7384.

Gupta V, Poss KD. 2012. Clonally dominant cardiomyocytes direct heart morphogenesis. *Nature* **484**: 479-484.

Hamilton DL, Abremski K. 1984. Site-specific recombination by the bacteriophage P1 lox-Cre system. Cre-mediated synapsis of two lox sites. *J Mol Biol* **178**: 481-486.

Hansson EM, Lindsay ME, Chien KR. 2009. Regeneration next: toward heart stem cell therapeutics. *Cell Stem Cell* **5**: 364-377.

He JQ, Ma Y, Lee Y, Thomson JA, Kamp TJ. 2003. Human embryonic stem cells develop into multiple types of cardiac myocytes: action potential characterization. *Circ Res* **93**: 32-39.

Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, Cost GJ, Zhang L, Santiago Y, Miller JC et al. 2011. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* **29**: 731-734.

Hoekstra M, Mummery CL, Wilde AA, Bezzina CR, Verkerk AO. 2012. Induced pluripotent stem cell derived cardiomyocytes as models for cardiac arrhythmias. *Front Physiol* **3**: 346.

Hu YF, Dawkins JF, Cho HC, Marban E, Cingolani E. 2014. Biological pacemaker created by minimally invasive somatic reprogramming in pigs with complete heart block. *Sci Transl Med* **6**: 245ra294.

Hutcheson KA, Atkins BZ, Hueman MT, Hopkins MB, Glower DD, Taylor DA. 2000. Comparison of benefits on myocardial performance of cellular cardiomyoplasty with skeletal myoblasts and fibroblasts. *Cell Transplant* **9**: 359-368.

Ionta V, Liang W, Kim EH, Rafie R, Giacomello A, Marban E, Cho HC. 2015. SHOX2 overexpression favors differentiation of embryonic stem cells into cardiac pacemaker cells, improving biological pacing ability. *Stem Cell Reports* **4**: 129-142.

Itzhaki I, Maizels L, Huber I, Zwi-Dantsis L, Caspi O, Winterstern A, Feldman O, Gepstein A, Arbel G, Hammerman H et al. 2011. Modelling the long QT syndrome with induced pluripotent stem cells. *Nature* **471**: 225-229.

Jain M, DerSimonian H, Brenner DA, Ngoy S, Teller P, Edge AS, Zawadzka A, Wetzel K, Sawyer DB, Colucci WS et al. 2001. Cell therapy attenuates deleterious ventricular remodeling and improves cardiac performance after myocardial infarction. *Circulation* **103**: 1920-1927.

Jerome LA, Papaioannou VE. 2001. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nat Genet* **27**: 286-291.

Jiang D, Zhao L, Clapham DE. 2009. Genome-wide RNAi screen identifies Letm1 as a mitochondrial Ca2+/H+ antiporter. *Science* **326**: 144-147.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816-821.

Kalhor R, Kalhor K, Mejia L, Leeper K, Graveline A, Mali P, Church GM. 2018. Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**.

Kapoor N, Liang W, Marban E, Cho HC. 2013. Direct conversion of quiescent cardiomyocytes to pacemaker cells by expression of Tbx18. *Nat Biotechnol* **31**: 54-62.

Karakikes I, Stillitano F, Nonnenmacher M, Tzimas C, Sanoudou D, Termglinchan V, Kong CW, Rushing S, Hansen J, Ceholski D et al. 2015. Correction of human phospholamban R14del mutation associated with cardiomyopathy using targeted nucleases and combination therapy. *Nat Commun* **6**: 6955.

Karkera JD, Lee JS, Roessler E, Banerjee-Basu S, Ouspenskaia MV, Mez J, Goldmuntz E, Bowers P, Towbin J, Belmont JW et al. 2007. Loss-of-function mutations in growth differentiation factor-1 (GDF1) are associated with congenital heart defects in humans. *American journal of human genetics* **81**: 987-994.

Kasahara H, Lee B, Schott JJ, Benson DW, Seidman JG, Seidman CE, Izumo S. 2000. Loss of function and inhibitory effects of human CSX/NKX2.5 homeoprotein mutations associated with congenital heart disease. *J Clin Invest* **106**: 299-308.

Kattman SJ, Witty AD, Gagliardi M, Dubois NC, Niapour M, Hotta A, Ellis J, Keller G. 2011. Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell* **8**: 228-240.

Kehat I, Kenyagin-Karsenti D, Snir M, Segev H, Amit M, Gepstein A, Livne E, Binah O, Itskovitz-Eldor J, Gepstein L. 2001. Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. *J Clin Invest* **108**: 407-414.

Kester L, van Oudenaarden A. 2018. Single-Cell Transcriptomics Meets Lineage Tracing. *Cell Stem Cell* doi:10.1016/j.stem.2018.04.014.

Khademhosseini A, Ferreira L, Blumling J, 3rd, Yeh J, Karp JM, Fukuda J, Langer R. 2006. Co-culture of human embryonic stem cells with murine embryonic fibroblasts on microwell-patterned substrates. *Biomaterials* **27**: 5968-5977.

Kim C, Wong J, Wen J, Wang S, Wang C, Spiering S, Kan NG, Forcales S, Puri PL, Leone TC et al. 2013. Studying arrhythmogenic right ventricular dysplasia with patient-specific iPSCs. *Nature* **494**: 105-110.

Kim YG, Cha J, Chandrasegaran S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* **93**: 1156-1160.

King T, Beddington RS, Brown NA. 1998. The role of the brachyury gene in heart development and left-right specification in the mouse. *Mechanisms of development* **79**: 29-37.

Kirk EP, Sunde M, Costa MW, Rankin SA, Wolstein O, Castro ML, Butler TL, Hyun C, Guo G, Otway R et al. 2007. Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. *American journal of human genetics* **81**: 280-291.

Kitajima S, Takagi A, Inoue T, Saga Y. 2000. MesP1 and MesP2 are essential for the development of cardiac mesoderm. *Development* **127**: 3215-3226.

Klug MG, Soonpaa MH, Koh GY, Field LJ. 1996. Genetically selected cardiomyocytes from differentiating embronic stem cells form stable intracardiac grafts. *J Clin Invest* **98**: 216-224.

Kocher AA, Schuster MD, Szabolcs MJ, Takuma S, Burkhoff D, Wang J, Homma S, Edwards NM, Itescu S. 2001. Neovascularization of ischemic myocardium by human bone-marrow-derived angioblasts prevents cardiomyocyte apoptosis, reduces remodeling and improves cardiac function. *Nature medicine* **7**: 430-436.

Kodo K, Nishizawa T, Furutani M, Arai S, Yamamura E, Joo K, Takahashi T, Matsuoka R, Yamagishi H. 2009. GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc Natl Acad Sci U S A* **106**: 13933-13938.

Kodo K, Ong SG, Jahanbani F, Termglinchan V, Hirono K, InanlooRahatloo K, Ebert AD, Shukla P, Abilez OJ, Churko JM et al. 2016. iPSC-derived cardiomyocytes reveal abnormal TGF-beta signalling in left ventricular non-compaction cardiomyopathy. *Nature cell biology* doi:10.1038/ncb3411.

Koh GY, Klug MG, Soonpaa MH, Field LJ. 1993. Differentiation and long-term survival of C2C12 myoblast grafts in heart. *J Clin Invest* **92**: 1548-1554.

Laflamme MA, Chen KY, Naumova AV, Muskheli V, Fugate JA, Dupras SK, Reinecke H, Xu C, Hassanipour M, Police S et al. 2007. Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nat Biotechnol* **25**: 1015-1024.

Laflamme MA, Gold J, Xu C, Hassanipour M, Rosler E, Police S, Muskheli V, Murry CE. 2005. Formation of human myocardium in the rat heart from human embryonic stem cells. *Am J Pathol* **167**: 663-671.

Laflamme MA, Murry CE. 2011. Heart regeneration. *Nature* **473**: 326-335.

Lan F, Lee AS, Liang P, Sanchez-Freire V, Nguyen PK, Wang L, Han L, Yen M, Wang Y, Sun N et al. 2013. Abnormal calcium handling properties underlie familial hypertrophic cardiomyopathy pathology in patient-specific induced pluripotent stem cells. *Cell Stem Cell* **12**: 101-113.

Laugwitz KL, Moretti A, Lam J, Gruber P, Chen Y, Woodard S, Lin LZ, Cai CL, Lu MM, Reth M et al. 2005. Postnatal isl1+ cardioblasts enter fully differentiated cardiomyocyte lineages. *Nature* **433**: 647-653.

Lee JH, Protze SI, Laksman Z, Backx PH, Keller GM. 2017. Human Pluripotent Stem Cell-Derived Atrial and Ventricular Cardiomyocytes Develop from Distinct Mesoderm Populations. *Cell Stem Cell* **21**: 179-194 e174.

Lee P, Klos M, Bollensdorff C, Hou L, Ewart P, Kamp TJ, Zhang J, Bizy A, Guerrero-Serna G, Kohl P et al. 2012. Simultaneous voltage and calcium mapping of genetically purified human induced pluripotent stem cell-derived cardiac myocyte monolayers. *Circ Res* **110**: 1556-1563.

Leobon B, Garcin I, Menasche P, Vilquin JT, Audinat E, Charpak S. 2003. Myoblasts transplanted into rat infarcted myocardium are functionally isolated from their host. *Proc Natl Acad Sci U S A* **100**: 7808-7811.

Leor J, Patterson M, Quinones MJ, Kedes LH, Kloner RA. 1996. Transplantation of fetal myocardial tissue into the infarcted myocardium of rat. A potential method for repair of infarcted myocardium? *Circulation* **94**: II332-336.

Lescroart F, Chabab S, Lin X, Rulands S, Paulissen C, Rodolosse A, Auer H, Achouri Y, Dubois C, Bondue A et al. 2014. Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nature cell biology* **16**: 829-840.

Lescroart F, Wang X, Lin X, Swedlund B, Gargouri S, Sanchez-Danes A, Moignard V, Dubois C, Paulissen C, Kinston S et al. 2018. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science* doi:10.1126/science.aao4174.

Li L, Krantz ID, Deng Y, Genin A, Banta AB, Collins CC, Qi M, Trask BJ, Kuo WL, Cochran J et al. 1997a. Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nat Genet* **16**: 243-251.

Li MA, Turner DJ, Ning Z, Yusa K, Liang Q, Eckert S, Rad L, Fitzgerald TW, Craig NL, Bradley A. 2011. Mobilization of giant piggyBac transposons in the mouse genome. *Nucleic Acids Res* **39**: e148.

Li QY, Newbury-Ecob RA, Terrett JA, Wilson DI, Curtis AR, Yi CH, Gebuhr T, Bullen PJ, Robson SC, Strachan T et al. 1997b. Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet* **15**: 21-29.

Li RK, Jia ZQ, Weisel RD, Mickle DA, Zhang J, Mohabeer MK, Rao V, Ivanov J. 1996. Cardiomyocyte transplantation improves heart function. *Ann Thorac Surg* **62**: 654-660; discussion 660-651.

Li S, Zhou D, Lu MM, Morrisey EE. 2004. Advanced cardiac morphogenesis does not require heart tube fusion. *Science* **305**: 1619-1622.

Lian X, Bao X, Zilberter M, Westman M, Fisahn A, Hsiao C, Hazeltine LB, Dunn KK, Kamp TJ, Palecek SP. 2015. Chemically defined, albumin-free human cardiomyocyte generation. *Nat Methods* **12**: 595-596.

Lian X, Hsiao C, Wilson G, Zhu K, Hazeltine LB, Azarin SM, Raval KK, Zhang J, Kamp TJ, Palecek SP. 2012. Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc Natl Acad Sci U S A* **109**: E1848-1857.

Lian X, Zhang J, Azarin SM, Zhu K, Hazeltine LB, Bao X, Hsiao C, Kamp TJ, Palecek SP. 2013. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions. *Nature protocols* **8**: 162-175.

Lickert H, Takeuchi JK, Von Both I, Walls JR, McAuliffe F, Adamson SL, Henkelman RM, Wrana JL, Rossant J, Bruneau BG. 2004. Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* **432**: 107-112.

Lieu DK, Liu J, Siu CW, McNerney GP, Tse HF, Abu-Khalil A, Huser T, Li RA. 2009. Absence of transverse tubules contributes to non-uniform Ca(2+) wavefronts in mouse and human embryonic stem cell-derived cardiomyocytes. *Stem Cells Dev* **18**: 1493-1500.

Lin Q, Schwarz J, Bucana C, Olson EN. 1997. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science* **276**: 1404-1407.

Lin X, Huo Z, Liu X, Zhang Y, Li L, Zhao H, Yan B, Liu Y, Yang Y, Chen YH. 2010. A novel GATA6 mutation in patients with tetralogy of Fallot or atrial septal defect. *J Hum Genet* **55**: 662-667.

Lin Y, Linask KL, Mallon B, Johnson K, Klein M, Beers J, Xie W, Du Y, Liu C, Lai Y et al. 2016. Heparin Promotes Cardiac Differentiation of Human Pluripotent Stem Cells in Chemically Defined Albumin-Free Medium, Enabling Consistent Manufacture of Cardiomyocytes. *Stem Cells Transl Med* doi:10.5966/sctm.2015-0428.

Liu SJ, Nowakowski TJ, Pollen AA, Lui JH, Horlbeck MA, Attenello FJ, He D, Weissman JS, Kriegstein AR, Diaz AA et al. 2016. Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol* **17**: 67.

Liu YW, Chen B, Yang X, Fugate JA, Kalucki FA, Futakuchi-Tsuchida A, Couture L, Vogel KW, Astley CA, Baldessari A et al. 2018. Human embryonic stem cell-derived cardiomyocytes restore function in infarcted hearts of non-human primates. *Nat Biotechnol* **36**: 597-605.

Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, Lichtman JW. 2007. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**: 56-62.

Loh KM, Chen A, Koh PW, Deng TZ, Sinha R, Tsai JM, Barkal AA, Shen KY, Jain R, Morganti RM et al. 2016. Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. *Cell* **166**: 451-467.

Ludwig TE, Bergendahl V, Levenstein ME, Yu J, Probasco MD, Thomson JA. 2006a. Feeder-independent culture of human embryonic stem cells. *Nat Methods* **3**: 637-646.

Ludwig TE, Levenstein ME, Jones JM, Berggren WT, Mitchen ER, Frane JL, Crandall LJ, Daigh CA, Conard KR, Piekarczyk MS et al. 2006b. Derivation of human embryonic stem cells in defined conditions. *Nat Biotechnol* **24**: 185-187.

Luo Z, Gao X, Lin C, Smith ER, Marshall SA, Swanson SK, Florens L, Washburn MP, Shilatifard A. 2015. Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Mol Cell* **57**: 685-694.

Ma H, Dang Y, Wu Y, Jia G, Anaya E, Zhang J, Abraham S, Choi JG, Shi G, Qi L et al. 2015. A CRISPR-Based Screen Identifies Genes Essential for West-Nile-Virus-Induced Cell Death. *Cell Rep* **12**: 673-683.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science* **339**: 823-826.

Marelli D, Desrosiers C, el-Alfy M, Kao RL, Chiu RC. 1992. Cell transplantation for myocardial repair: an experimental approach. *Cell Transplant* **1**: 383-390.

Marino B, Digilio MC, Toscano A, Giannotti A, Dallapiccola B. 1999. Congenital heart diseases in children with Noonan syndrome: An expanded cardiac spectrum with high prevalence of atrioventricular canal. *J Pediatr* **135**: 703-706.

McBride KL, Riley MF, Zender GA, Fitzgerald-Butt SM, Towbin JA, Belmont JW, Cole SE. 2008. NOTCH1 mutations in individuals with left ventricular outflow tract malformations reduce ligand-induced signaling. *Hum Mol Genet* **17**: 2886-2893.

McDaniell R, Warthen DM, Sanchez-Lara PA, Pai A, Krantz ID, Piccoli DA, Spinner NB. 2006. NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *American journal of human genetics* **79**: 169-173.

McDole K, Guignard L, Amat F, Berger A, Malandain G, Royer LA, Turaga SC, Branson K, Keller PJ. 2018. In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level. *Cell* doi:10.1016/j.cell.2018.09.031.

McElhinney DB, Krantz ID, Bason L, Piccoli DA, Emerick KM, Spinner NB, Goldmuntz E. 2002. Analysis of cardiovascular phenotype and genotype-phenotype correlation in individuals with a JAG1 mutation and/or Alagille syndrome. *Circulation* **106**: 2567-2574.

Megarbane A, Salem N, Stephan E, Ashoush R, Lenoir D, Delague V, Kassab R, Loiselet J, Bouvagnet P. 2000. X-linked transposition of the great arteries and incomplete penetrance among males with a nonsense mutation in ZIC3. *Eur J Hum Genet* **8**: 704-708.

Meisler MH. 1997. Mutation watch: mouse brachyury (T), the T-box gene family, and human disease. *Mamm Genome* **8**: 799-800.

Mohamed SA, Aherrahrou Z, Liptau H, Erasmi AW, Hagemann C, Wrobel S, Borzym K, Schunkert H, Sievers HH, Erdmann J. 2006. Novel missense mutations (p.T596M and p.P1797H) in NOTCH1 in patients with bicuspid aortic valve. *Biochem Biophys Res Commun* **345**: 1460-1465.

Mohapatra B, Casey B, Li H, Ho-Dawson T, Smith L, Fernbach SD, Molinari L, Niesh SR, Jefferies JL, Craigen WJ et al. 2009. Identification and functional characterization of NODAL rare variants in heterotaxy and isolated cardiovascular malformations. *Hum Mol Genet* **18**: 861-871.

Mohr JC, de Pablo JJ, Palecek SP. 2006. 3-D microwell culture of human embryonic stem cells. *Biomaterials* **27**: 6032-6042.

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733-740.

Molkentin JD, Lin Q, Duncan SA, Olson EN. 1997. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev* **11**: 1061-1072.

Momma K. 2010. Cardiovascular anomalies associated with chromosome 22q11.2 deletion syndrome. *Am J Cardiol* **105**: 1617-1624.

Moretti A, Bellin M, Welling A, Jung CB, Lam JT, Bott-Flugel L, Dorn T, Goedel A, Hohnke C, Hofmann F et al. 2010. Patient-specific induced pluripotent stem-cell models for long-QT syndrome. *N Engl J Med* **363**: 1397-1409.

Moretti A, Caron L, Nakano A, Lam JT, Bernshausen A, Chen Y, Qyang Y, Bu L, Sasaki M, Martin-Puig S et al. 2006. Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell* **127**: 1151-1165.

Murry CE, Kay MA, Bartosek T, Hauschka SD, Schwartz SM. 1996. Muscle differentiation during repair of myocardial necrosis in rats via gene transfer with MyoD. *J Clin Invest* **98**: 2209-2217.

Nag AC. 1980. Study of non-muscle cells of the adult mammalian heart: a fine structural analysis and distribution. *Cytobios* **28**: 41-61.

Nemer G, Fadlalah F, Usta J, Nemer M, Dbaibo G, Obeid M, Bitar F. 2006. A novel mutation in the GATA4 gene in patients with Tetralogy of Fallot. *Hum Mutat* **27**: 293-294.

Ng ES, Davis RP, Azzola L, Stanley EG, Elefanty AG. 2005. Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust, reproducible hematopoietic differentiation. *Blood* **106**: 1601-1603.

Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**: 935-949.

Noonan JA. 1994. Noonan syndrome. An update and review for the primary pediatrician. *Clin Pediatr (Phila)* **33**: 548-555.

Nygren JM, Jovinge S, Breitbach M, Sawen P, Roll W, Hescheler J, Taneera J, Fleischmann BK, Jacobsen SE. 2004. Bone marrow-derived hematopoietic cells generate cardiomyocytes at a low frequency through cell fusion, but not transdifferentiation. *Nature medicine* **10**: 494-501.

Oda T, Elkahloun AG, Pike BL, Okajima K, Krantz ID, Genin A, Piccoli DA, Meltzer PS, Spinner NB, Collins FS et al. 1997. Mutations in the human Jagged1 gene are responsible for Alagille syndrome. *Nat Genet* **16**: 235-242.

Oh H, Bradfute SB, Gallardo TD, Nakamura T, Gaussin V, Mishina Y, Pocius J, Michael LH, Behringer RR, Garry DJ et al. 2003. Cardiac progenitor cells from adult myocardium: homing, differentiation, and fusion after infarction. *Proc Natl Acad Sci U S A* **100**: 12313-12318.

Osafune K, Caron L, Borowiak M, Martinez RJ, Fitz-Gerald CS, Sato Y, Cowan CA, Chien KR, Melton DA. 2008. Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat Biotechnol* **26**: 313-315.

Parnas O, Jovanovic M, Eisenhaure TM, Herbst RH, Dixit A, Ye CJ, Przybylski D, Platt RJ, Tirosh I, Sanjana NE et al. 2015. A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162**: 675-686.

Peerani R, Rao BM, Bauwens C, Yin T, Wood GA, Nagy A, Kumacheva E, Zandstra PW. 2007. Niche-mediated control of human embryonic stem cell self-renewal and differentiation. *EMBO J* **26**: 4744-4755.

Penn MS, Francis GS, Ellis SG, Young JB, McCarthy PM, Topol EJ. 2002. Autologous cell transplantation for the treatment of damaged myocardium. *Prog Cardiovasc Dis* **45**: 21-32.

Pfaff SL, Mendelsohn M, Stewart CL, Edlund T, Jessell TM. 1996. Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell* **84**: 309-320.

Pfeiffer MJ, Quaranta R, Piccini I, Fell J, Rao J, Ropke A, Seebohm G, Greber B. 2018. Cardiogenic programming of human pluripotent stem cells by dose-controlled activation of EOMES. *Nat Commun* **9**: 440.

Phillip M, Lebenthal Y, Lebl J, Zuckerman-Levin N, Korpal-Szczyrska M, Marques JS, Steensberg A, Jons K, Kappelgaard AM, Ibanez L et al. 2009. European multicentre study in children born small for gestational age with persistent short stature: comparison of continuous and discontinuous growth hormone treatment regimens. *Horm Res* **71**: 52-59.

Piccini I, Arauzo-Bravo M, Seebohm G, Greber B. 2016. Functional high-resolution time-course expression analysis of human embryonic stem cells undergoing cardiac induction. *Genom Data* **10**: 71-74.

Picelli S, Bjorklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24**: 2033-2040.

Pizzuti A, Sarkozy A, Newton AL, Conti E, Flex E, Digilio MC, Amati F, Gianni D, Tandoi C, Marino B et al. 2003. Mutations of ZFPM2/FOG2 gene in sporadic cases of tetralogy of Fallot. *Hum Mutat* **22**: 372-377.

Pont SJ, Robbins JM, Bird TM, Gibson JB, Cleves MA, Tilford JM, Aitken ME. 2006. Congenital malformations among liveborn infants with trisomies 18 and 13. *Am J Med Genet A* **140**: 1749-1756.

Posch MG, Gramlich M, Sunde M, Schmitt KR, Lee SH, Richter S, Kersten A, Perrot A, Panek AN, Al Khatib IH et al. 2010. A gain-of-function TBX20 mutation causes congenital atrial septal defects, patent foramen ovale and cardiac valve defects. *Journal of medical genetics* **47**: 230-235.

Pourebrahim R, Houtmeyers R, Ghogomu S, Janssens S, Thelie A, Tran HT, Langenberg T, Vleminckx K, Bellefroid E, Cassiman JJ et al. 2011. Transcription factor Zic2 inhibits Wnt/beta-catenin protein signaling. *J Biol Chem* **286**: 37732-37740.

Prall OW, Menon MK, Solloway MJ, Watanabe Y, Zaffran S, Bajolle F, Biben C, McBride JJ, Robertson BR, Chaulet H et al. 2007. An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. *Cell* **128**: 947-959.

Pusapati GV, Kong JH, Patel BB, Krishnan A, Sagner A, Kinnebrew M, Briscoe J, Aravind L, Rohatgi R. 2018. CRISPR Screens Uncover Genes that Regulate Target Cell Sensitivity to the Morphogen Sonic Hedgehog. *Dev Cell* **44**: 271.

Qi X, Yang G, Yang L, Lan Y, Weng T, Wang J, Wu Z, Xu J, Gao X, Yang X. 2007. Essential role of Smad4 in maintaining cardiomyocyte proliferation during murine embryonic heart development. *Dev Biol* **311**: 136-146.

Rad R, Rad L, Wang W, Cadinanos J, Vassiliou G, Rice S, Campos LS, Yusa K, Banerjee R, Li MA et al. 2010. PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. *Science* **330**: 1104-1107.

Ramirez CL, Foley JE, Wright DA, Muller-Lerch F, Rahman SH, Cornu TI, Winfrey RJ, Sander JD, Fu F, Townsend JA et al. 2008. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods* **5**: 374-375.

Reamon-Buettner SM, Borlak J. 2010. NKX2-5: an update on this hypermutable homeodomain protein and its role in human congenital heart disease (CHD). *Hum Mutat* **31**: 1185-1194.

Robertson C, Tran DD, George SC. 2013. Concise review: maturation phases of human pluripotent stem cell-derived cardiomyocytes. *Stem Cells* **31**: 829-837.

Robey TE, Saiget MK, Reinecke H, Murry CE. 2008. Systems approaches to preventing transplanted cell death in cardiac repair. *J Mol Cell Cardiol* **45**: 567-581.

Ronaldson-Bouchard K, Ma SP, Yeager K, Chen T, Song L, Sirabella D, Morikawa K, Teles D, Yazawa M, Vunjak-Novakovic G. 2018. Advanced maturation of human cardiac tissue grown from pluripotent stem cells. *Nature* **556**: 239-243.

Ruiz S, Mayor-Ruiz C, Lafarga V, Murga M, Vega-Sendino M, Ortega S, Fernandez-Capetillo O. 2016. A Genome-wide CRISPR Screen Identifies CDC25A as a Determinant of Sensitivity to ATR Inhibitors. *Mol Cell* **62**: 307-313.

Russ AP, Wattler S, Colledge WH, Aparicio SA, Carlton MB, Pearce JJ, Barton SC, Surani MA, Ryan K, Nehls MC et al. 2000. Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* **404**: 95-99.

Saga Y. 1998. Genetic rescue of segmentation defect in MesP2-deficient mice by MesP1 gene replacement. *Mechanisms of development* **75**: 53-66.

Saga Y, Miyagawa-Tomita S, Takagi A, Kitajima S, Miyazaki J, Inoue T. 1999. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development* **126**: 3437-3447.

Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH. 2004. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature medicine* **10**: 55-63.

Scambler PJ. 2000. The 22q11 deletion syndromes. *Hum Mol Genet* **9**: 2421-2426.

Scorsin M, Hagege A, Vilquin JT, Fiszman M, Marotte F, Samuel JL, Rappaport L, Schwartz K, Menasche P. 2000. Comparison of the effects of fetal cardiomyocyte and skeletal myoblast transplantation on postinfarction left ventricular function. *J Thorac Cardiovasc Surg* **119**: 1169-1175.

Scuderi GJ, Butcher J. 2017. Naturally Engineered Maturation of Cardiomyocytes. *Front Cell Dev Biol* **5**: 50.

Seo S, Kume T. 2006. Forkhead transcription factors, Foxc1 and Foxc2, are required for the morphogenesis of the cardiac outflow tract. *Dev Biol* **296**: 421-436.

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, Heckl D, Ebert BL, Root DE, Doench JG et al. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**: 84-87.

Shiba Y, Fernandes S, Zhu WZ, Filice D, Muskheli V, Kim J, Palpant NJ, Gantz J, Moyes KW, Reinecke H et al. 2012. Human ES-cell-derived cardiomyocytes electrically couple and suppress arrhythmias in injured hearts. *Nature* **489**: 322-325.

Sirard C, de la Pompa JL, Elia A, Itie A, Mirtsos C, Cheung A, Hahn S, Wakeham A, Schwartz L, Kern SE et al. 1998. The tumor suppressor gene Smad4/Dpc4 is required for gastrulation and later for anterior development of the mouse embryo. *Genes Dev* **12**: 107-119.

Snippert HJ, van der Flier LG, Sato T, van Es JH, van den Born M, Kroon-Veenboer C, Barker N, Klein AM, van Rheenen J, Simons BD et al. 2010. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**: 134-144.

Song L, Yan W, Chen X, Deng CX, Wang Q, Jiao K. 2007. Myocardial smad4 is essential for cardiogenesis in mouse embryos. *Circ Res* **101**: 277-285.

Spater D, Abramczuk MK, Buac K, Zangi L, Stachel MW, Clarke J, Sahara M, Ludwig A, Chien KR. 2013. A HCN4+ cardiomyogenic progenitor derived from the first heart field and human pluripotent stem cells. *Nature cell biology* **15**: 1098-1106.

Sperling S, Grimm CH, Dunkel I, Mebus S, Sperling HP, Ebner A, Galli R, Lehrach H, Fusch C, Berger F et al. 2005. Identification and functional analysis of CITED2 mutations in patients with congenital heart defects. *Hum Mutat* **26**: 575-582.

Srivastava D, Thomas T, Lin Q, Kirby ML, Brown D, Olson EN. 1997. Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nat Genet* **16**: 154-160.

Stennard FA, Costa MW, Lai D, Biben C, Furtado MB, Solloway MJ, McCulley DJ, Leimena C, Preis JI, Dunwoodie SL et al. 2005. Murine T-box transcription factor Tbx20 acts as a repressor during heart development, and is essential for adult heart integrity, function and adaptation. *Development* **132**: 2451-2462.

Synnergren J, Ameen C, Jansson A, Sartipy P. 2012. Global transcriptional profiling reveals similarities and differences between human stem cell-derived cardiomyocyte clusters and heart tissue. *Physiol Genomics* **44**: 245-258.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**: 861-872.

Tam PP, Loebel DA. 2007. Gene function in mouse embryogenesis: get set for gastrulation. *Nat Rev Genet* **8**: 368-381.

Tanaka M, Chen Z, Bartunkova S, Yamasaki N, Izumo S. 1999. The cardiac homeobox gene Csx/Nkx2.5 lies genetically upstream of multiple genes essential for heart development. *Development* **126**: 1269-1280.

Tashiro A, Masuda T, Segawa I. 1990. Morphometric comparison of mitochondria and myofibrils of cardiomyocytes between hypertrophic and dilated cardiomyopathies. *Virchows Arch A Pathol Anat Histopathol* **416**: 473-478.

Terns MP, Terns RM. 2011. CRISPR-based adaptive immune systems. *Curr Opin Microbiol* **14**: 321-327.

Thomas T, Yamagishi H, Overbeek PA, Olson EN, Srivastava D. 1998. The bHLH factors, dHAND and eHAND, specify pulmonary and systemic cardiac ventricles independent of left-right sidedness. *Dev Biol* **196**: 228-236.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. 1998. Embryonic stem cell lines derived from human blastocysts. *Science* **282**: 1145-1147.

Tidyman WE, Rauen KA. 2009. The RASopathies: developmental syndromes of Ras/MAPK pathway dysregulation. *Curr Opin Genet Dev* **19**: 230-236.

Tohyama S, Hattori F, Sano M, Hishiki T, Nagahata Y, Matsuura T, Hashimoto H, Suzuki T, Yamashita H, Satoh Y et al. 2013. Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. *Cell Stem Cell* **12**: 127-137.

Toma C, Pittenger MF, Cahill KS, Byrne BJ, Kessler PD. 2002. Human mesenchymal stem cells differentiate to a cardiomyocyte phenotype in the adult murine heart. *Circulation* **105**: 93-98.

Tompkins JD, Jung M, Chen CY, Lin Z, Ye J, Godatha S, Lizhar E, Wu X, Hsu D, Couture LA et al. 2016. Mapping Human Pluripotent-to-Cardiomyocyte Differentiation: Methylomes, Transcriptomes, and Exon DNA Methylation "Memories". *EBioMedicine* **4**: 74-85.

van der Linde D, Konings EE, Slager MA, Witsenburg M, Helbing WA, Takkenberg JJ, Roos-Hesselink JW. 2011. Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. *J Am Coll Cardiol* **58**: 2241-2247.

van Rooij E, Sutherland LB, Qi X, Richardson JA, Hill J, Olson EN. 2007. Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science* **316**: 575-579.

Veerman CC, Kosmidis G, Mummery CL, Casini S, Verkerk AO, Bellin M. 2015. Immaturity of human stem-cell-derived cardiomyocytes in culture: fatal flaw or soluble problem? *Stem Cells Dev* **24**: 1035-1052.

von Both I, Silvestri C, Erdemir T, Lickert H, Walls JR, Henkelman RM, Rossant J, Harvey RP, Attisano L, Wrana JL. 2004. Foxh1 is essential for development of the anterior heart field. *Dev Cell* **7**: 331-345.

Wang G, McCain ML, Yang L, He A, Pasqualini FS, Agarwal A, Yuan H, Jiang D, Zhang D, Zangi L et al. 2014a. Modeling the mitochondrial cardiomyopathy of Barth syndrome with induced pluripotent stem cell and heart-on-chip technologies. *Nature medicine* **20**: 616-623.

Wang J, Xu N, Feng X, Hou N, Zhang J, Cheng X, Chen Y, Zhang Y, Yang X. 2005. Targeted disruption of Smad4 in cardiomyocytes results in cardiac hypertrophy and heart failure. *Circ Res* **97**: 821-828.

Wang T, Wei JJ, Sabatini DM, Lander ES. 2014b. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**: 80-84.

Wang T, Yu H, Hughes NW, Liu B, Kendirli A, Klein K, Chen WW, Lander ES, Sabatini DM. 2017. Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* **168**: 890-903 e815.

Wiedenheft B, Sternberg SH, Doudna JA. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**: 331-338.

Xin B, Puffenberger E, Tumbush J, Bockoven JR, Wang H. 2007. Homozygosity for a novel splice site mutation in the cardiac myosin-binding protein C gene causes severe neonatal hypertrophic cardiomyopathy. *Am J Med Genet A* **143A**: 2662-2667.

Xu C, Police S, Rao N, Carpenter MK. 2002. Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells. *Circ Res* **91**: 501-508.

Yang L, Soonpaa MH, Adler ED, Roepke TK, Kattman SJ, Kennedy M, Henckaerts E, Bonham K, Abbott GW, Linden RM et al. 2008. Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature* **453**: 524-528.

Yang X, Li C, Xu X, Deng C. 1998. The tumor suppressor SMAD4/DPC4 is essential for epiblast proliferation and mesoderm induction in mice. *Proc Natl Acad Sci U S A* **95**: 3667-3672.

Yang X, Pabon L, Murry CE. 2014. Engineering adolescence: maturation of human pluripotent stem cell-derived cardiomyocytes. *Circ Res* **114**: 511-523.

Yasuda T, Weisel RD, Kiani C, Mickle DA, Maganti M, Li RK. 2005. Quantitative analysis of survival of transplanted smooth muscle cells with real-time polymerase chain reaction. *J Thorac Cardiovasc Surg* **129**: 904-911.

Yazawa M, Hsueh B, Jia X, Pasca AM, Bernstein JA, Hallmayer J, Dolmetsch RE. 2011. Using induced pluripotent stem cells to investigate cardiac phenotypes in Timothy syndrome. *Nature* **471**: 230-234.

Ye L, Chang YH, Xiong Q, Zhang P, Zhang L, Somasundaram P, Lepley M, Swingen C, Su L, Wendel JS et al. 2014. Cardiac repair in a porcine model of acute myocardial infarction with human induced pluripotent stem cell-derived cardiovascular cells. *Cell Stem Cell* **15**: 750-761.

Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A. 2008. The ground state of embryonic stem cell self-renewal. *Nature* **453**: 519-523.

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R et al. 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**: 1917-1920.

Zhang J, Klos M, Wilson GF, Herman AM, Lian X, Raval KK, Barron MR, Hou L, Soerens AG, Yu J et al. 2012. Extracellular matrix promotes highly efficient cardiac differentiation of human pluripotent stem cells: the matrix sandwich method. *Circ Res* **111**: 1125-1136.

Zhang J, Wilson GF, Soerens AG, Koonce CH, Yu J, Palecek SP, Thomson JA, Kamp TJ. 2009. Functional cardiomyocytes derived from human induced pluripotent stem cells. *Circ Res* **104**: e30-41.

Zhang Q, Jiang J, Han P, Yuan Q, Zhang J, Zhang X, Xu Y, Cao H, Meng Q, Chen L et al. 2011. Direct differentiation of atrial and ventricular myocytes from human embryonic stem cells by alternating retinoid signals. *Cell Res* **21**: 579-587.

Zhang Y, Riesterer C, Ayrall AM, Sablitzky F, Littlewood TD, Reth M. 1996. Inducible site-directed recombination in mouse embryonic stem cells. *Nucleic Acids Res* **24**: 543-548.

Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y, Wei W. 2014. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**: 487-491.

Zwaka TP, Thomson JA. 2003. Homologous recombination in human embryonic stem cells. *Nat Biotechnol* **21**: 319-321.