REGULATION OF CARDIAC PROGENITORS BY COMBINATION OF MESP1 AND ETS TRANSCRIPTION FACTORS

A Dissertation

by

KUO-CHAN WENG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Fen Wang
Co-Chair of Committee,	Robert J. Schwartz
Committee Members,	James F. Martin
	Jiang Chang
Head of Department,	Fen Wang

May 2014

Major Subject: Medical Science

ABSTRACT

Heart disease remains the leading cause of death worldwide. By understanding the regulating networks during cardiac development we can exploit those networks to manipulate adult cells into cardiac progenitors and provide an alternative for repairing diseased hearts. Mesp1 is considered to have critical roles during cardiac development but the molecular mechanisms need to be further studied. The roles of ETS transcription factors have been primarily limited to hematopoietic differentiation and cancer progression. The ETS transcription factors are known to have proliferating roles and were hypothesized to also be involved in cardiac differentiation and may potentially be used for cell reprogramming.

The first part of this study characterizes the expression pattern of Mesp1 protein in early mouse embryo from E6.5 to E9.5 and provides a full expression profile in differentiating embryoid bodies *in vitro* from the undifferentiated stage to Day10. Our work showed Mesp1 expresses in the posterior region of E6.5 embryo then starts migrating through the primitive streak toward anterior mesoderm and endoderm in E7.5. A Mesp1 linage tracing ES cell line was established, and it allowed us to trace the Mesp1 derived cell population. The lineage tracing system confirmed Mesp1 expressing cells give rise to a major part of the heart and also contributes to some endodermal derived organs such as the pancreas. The direct DNA binding targets of Mesp1 were determined using a Mesp1 specific antibody to perform ChIP of bound DNA that could then be used in next generation

sequencing. The resulting sequence data included cardiac genes such as *Gata4*, *Hand2*, and *Myocd*. Endoderm correlated genes *Foxa2*, *Pitx2*, and *Gata6* were also shown to be Mesp1 targets. The targeted genes were validated as transcriptional targets using an ES cell line with inducible Mesp1 followed by qPCR of target gene transcript levels with and without Mesp1 expression. Secondly, the complete gene expression profiles of over 20 ETS transcription factors were generated. By comparing the ETS factor expression patterns and identifying which showed cardiac gene activation, ETV4 and ETS2 were chosen for further study of their roles in cardiac differentiation. ETS2 was used in combination with Mesp1 to reprogram Normal Human Dermal Fibroblast into cardiac progenitors. The reprogrammed cells were then characterized for gene expression patterns, surface marker, and structural protein presentation. This work provides thorough insights into the roles of Mesp1 and ETS transcription factors during germ layer development and led to the development of a method to reprogram adult cells into cardiac progenitors that could be applied for clinical use in the future.

DEDICATION

To my parents, who raise me up and support me.

To my mentor, Dr Huan-Yao Lei, for leading me into the scientific world.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Schwartz for his guidance and support. Thanks for his great mentorship; always reminding me to think in higher levels while keeping the quality of research. He is always encouraging us in a very comfortable and humorous way. It is my pleasure and honor to have worked with you for the past few years. The experience learning from you will heighten my future carrier.

I would also like to thank all my committee members: Dr Fen Wang, Dr James Martin, and Dr Jiang Chang. They gave me the freedom to develop the project and provided many valuable and specific suggestions. Thank you for the great support. Special thanks to Dr Anthony Firulli in Indianapolis for keeping me motivated on the project.

Great thanks to my graduate program IBT. Thanks to all faculties that provided a friendly environment to make the research go smoothly. Thanks to all administrative personnel, especially program coordinator Cynthia Lewis and former coordinator Janis Bender. Also thanks to my classmates and friends at IBT: Yan Bai, Yang Xiao, Xiangsheng, and Xi Lin for your company and helpful information.

I would like to express my gratitude to Texas Heart Institute (THI). They provided a great environment in clinical and basic research. It is my pleasure to work with such a great team here. Thanks Dinah, Verna and Ajay to take care all administrative details. Thanks

Dr. Potaman for your guidance and help. Thanks to Schwartz Lab: Jose, Ming, Matt, Dasha, and the Jun Wang Lab: Qiang Ling, Eun Yung, Illembek; Edward Yeh Lab: Dr Long-Sheng Lu, Jeremy and Rui; Dixson Lab: Dean, Amy and Qi; Medel Chen Lab: Jonathan. Thanks to Alon from Flow Cytometry Core for providing technical support and helpful information. Also thanks all other lab to make THI a great place to work.

I would like to say thank you to all Schwartz Lab at University of Houston. Thanks Cindy to take care every administrative work. Thanks Dr Yu Liu for your generosity and help. Thanks Dr Li Chen and your group for the guidance, especially Weijia and Wei Yu. Thanks to lncRNA group: Ashley, for your hand-by-hand guidance; to Ben and Jong for helping finishing the project.

I want to thanks all my friend in Houston area, especially Taiwanese Student and Scholar Association from Baylor, UT-Houston, Rice and UH, to give me lots of useful information in the US and made my life in Houston very colorful. Thanks my housemate, James and my friends Joanne, Elaine, Tina, Rose, Judy, Fu-Jung, Yen-Tun and Shih-Shin for all your support and company.

Finally I would like to thanks my family, especially my sister, to take care of my Mom and Dad in Taiwan and make me feel carefree studying abroad.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1 Overview of Cardiovascular Diseases

1.1.1 Cardiovascular Diseases Remain Leading Cause of Mortality

Cardiovascular disease remains the leading cause of death in women and men across the world. Even with the exponential advances over the past decade in medicine and clinical treatment options for treating heart disease, the cardiac related deaths still has the highest mortality rate. There is also an increasing number of heart disease related cases in both developmental and low-income countries[1]. In 2010, people in the USA spent $315 billion on medical treatments for cardiovascular diseases alone. However, the rate of cardiac related deaths has not decreased and the majority of current treatment options merely manage heart disease and are unable to cure it[2].

The major cause of cardiovascular diseases is ischemic heart disease (IHD) or coronary artery disease (CAD)[3]. It is triggered by plaque accumulation in the inner wall of the coronary artery, which leads to a loss of oxygen and nutrients that are necessary for proper cardiomyocyte function[4]. Through the progression of CAD, the patient may suffer from angina and eventually suffer a heart attack. The acute heart disease causes the irreversible loss of cardiomyocytes that could lead to permanent heart failure and immediate death if not instantly treated[5].

There are different treatments for myocardial infarctions including surgical repairs, physical therapy, and heart transplants[6, 7]. Those patients, where a transplant is the only remaining option, face major challenges after receiving the new heart. The patients need to take immunosuppressive drug to reduce the transplant rejection by their own immune system and the long term use of these drugs may lead to developing a fatal infection[8, 9]. The long term prognoses are not ideal even with modern technologies and patient care. The 5-year survival rate of patients with a heart transplant is less than 75% according to American Heart Association[10]. The shortage of heart donors is also a major obstacle along with the chance an available heart is suitable for the recipient in need. Therefore, alternative methods are needed to treat patients who are waiting for a heart transplant

1.1.2 Limited Ability of Regeneration of Cardiomyocytes in Adult Heart

The regeneration of cardiomyocytes only happens in vertebrate such as *Zebrafish* and amphibians[11, 12]. Mammalian heart are terminally differentiated and are considered to have poor regenerative ability[13, 14]. However, a recent report showed the heart in neonatal mice could regenerate after heart surgical resection[15]. The ventricular apex of neonatal mice was resected and the full restoration was observed after 21 days. The regenerated regions were revealed to originate from pre-existing cardiomyocytes. This study means the mammalian heart may have regeneration capacity during younger ages. Another group demonstrated the turnover of cardiomyocytes in adult mice might occur at very low rates[16]. This means the mature myocardium of the heart can be replaced

either by an existing cardiac stem cell pool or by recruiting stem cells through circulation. Moreover, the studies showed the population of cardiomyocytes was altered after injury. This newly generated population was from the proliferation of an existing pool of cardiac progenitors around the injured region[17].

The regeneration of cardiomyocytes from a pre-existing population revealed that the signaling pathways involved in this process are important for identifying new therapies. The pathways involved in cell proliferation and differentiation are usually highly conserved throughout the developmental stages. They include BMP, FGF, Wnt and Nodal signaling pathways. To better understanding how cardiomyocytes could be repaired, researchers look to cardiac development during the embryonic stage to help us to focus on the critical factors involved throughout the developmental and repair process.

1.2 Heart Development

1.2.1 Embryo Development

Embryonic development starts from a single, fertilized cell embryo that undergoes cleavage division to become a blastomere that progresses to an egg cylinder before becoming a gastrula that continues to go through asymmetrical proliferation and distinctive germ layer differentiation that blossoms into a complex, three dimensional organism[18]. The gastrulation stage is one of the most important processes in early mouse embryogenesis. The cells divide into three germ layers: ectoderm, mesoderm, and endoderm. The three germ layers will later contribute to tissue development. The

majority of the ectoderm layer will develop into the neural system, skin, and tooth enamel. Mesoderm derives cardiac muscle, skeletal muscle, and blood cells[19]. Endoderm becomes the gastrointestinal tract, respiratory tract, and endocrine organs[20]. The work presented here, focuses on mesoderm and endoderm development.

The endoderm layer will give rise to organs such as the digestive tract and respiratory tube. This includes the mouth, pharynx, trachea, lungs, urinary bladder, thyroid gland, and thymus. There are three components of the mesoderm germ layer: the paraxial mesoderm, the intermediate mesoderm, and the lateral plate mesoderm. The paraxial mesoderm will form somitomeres and later give rise to muscle tissue, cartilage, bone, and subcutaneous, skin tissue. The intermediate mesoderm in the upper thoracic and cervical regions forms the nephrotome, and in caudally regions it forms the nephrogenic cord. The lateral plate mesoderm splits into somatic and splanchnic layers. The somatic layer will form the lateral body wall that folds with the ectoderm and the splanchnic layer will form the walls of gut tube.

1.2.2 Heart Formation

Heart is derived from the mesoderm germ layer during embryonic development[21]. At E5.0, Nodal signaling induces the mesoderm in proximal epiblast and allows Bmp4 expression in the extraembryonic ectoderm. The expression of Bmp4 keeps Wnt3 expression in the proximal epiblast. At E5.5, the Wnt inhibitors Dkk1, Lefty1, and Cer1 are expressed in the anterior region so that they restrict the Wnt and Nodal signaling in

the posterior epiblast. Wnt signaling induces mesodermal marker T (Brachyury) and Eomes expression in E5.75 that will later induce Mesp1 expression to drive cardiac, mesoderm specification through Dkk1-mediated inhibition of Wnt pathway[22].

The cardiac mesoderm will give rise to endocardium, the first heart field (FHF) and the second heart field (SHF). The first heart field will form the left ventricle, atria, and conduction system whereas the second heart field will give rise to the right ventricle, part of atria, and outflow tract[23]. The transcription networks coordinate the formation of the heart during development. This includes the transcription factors, Nkx2-5, Gata4, Tbx20, found in both heart fields and Hand2, Isl1, Mef2c, and Fgf10 found in the second heart field[24]. Knocking out these genes usually leads to embryonic lethality and/or a defect of cardiac morphogenesis[25]. At E9.5, the cardiac crescent forms a looping heart tube through rightward looping. The cardiac neural crest migrates into the outflow tract and separates and configures the aortic arch arteries[26]. Later, the heart forms four chambers at E10.5[27, 28].

1.2.3 Signal Pathways in Cardiac Development

1.2.3.1 TGF Beta Signaling Pathway

Transforming Growth Factor beta (TGFB) super family signaling pathway plays critical roles during embryo development that includes cell proliferation and cell differentiation. The general molecular mechanism of the TGFB pathway is through ligand-receptor interactions. The ligands bind on type II TGFB receptors and type II receptors recruit

type I receptors. Then the Type I receptors phosphorylate receptor-regulated SMADs (R-SMAD) and R-SMADs interact with the other coSMAD SMAD4. The R-SMAD/coSMAD complex then translocates to the nucleus to regulate downstream target genes.

The ligands include Bone Morphogenetic Proteins (BMPs), Activin, Nodal, and TGFβ. BMPs bind to type II receptor BMPR2 and activate SAMD1, 5, and 8 through type I receptor BMPR1A (ALK3) and BMPR1B (ALK6) and the ligands can be inhibited by Noggin and Chordin. Nodal binds to ACVR2B and activates SMADs through ACVR1B (ALK4) or ACVR1C (ALK7) and can be inhibited by Lefty. TGFβs bind to TGFβRII and activate SMADs through TGFβRI (ALK5). Activin A binds to ACVR2A and activates SMADs through ACVR1B (ALK4) and can be inhibited by Follistatin.

Among BMP lignads, BMP2 and BMP4 can induce cardiomyocyte differentiation[29-31] and both BMP2 and BMP4 null mice are embryonic lethal[32, 33]. The phenotype of BMP4 null mice showed plasticity. It suggested the multiple roles of BMP4 such as gastrulation or mesoderm specification. Mice lacking Activin did not show significant heart defects while the roles of mesoderm formation had been confirmed in lower vertebrates[34]. Only ACVR2B knockout mice showed heart defects on left-right patterning[35].

1.2.3.2 FGF Signaling Pathway

Fibroblast growth factors (FGF) are growth factors that play roles in the process of cell proliferation, differentiation, and migration[36, 37]. The mitogenic functions include mesoderm formation, anterior-posterior patterning, neural development, and wound healing[38, 39]. Twenty-two FGF family members in vertebrates have been identified and some of them show important roles in cardiac development[40, 41]. The FGF signal pathway undergoes downstream signaling by interacting with tyrosine kinase receptors. There are four major members of FGF receptors (FGFR1-4). The heterodimerization of FGF and FGFR results in phosphorylation of intracellular residue of FGFR and activates the signal transduction pathways. Two major signaling cascades include MAPK and PI3K cascades[42, 43].

FGF signaling pathway showed multiple roles in cardiac development. FGF2 (basic FGF) plays a critical role by inducing the expression of important core cardiac transcription factors[44]. Fgf4 regulates valve precursor cell proliferation through inducing the expansion of cushion mesenchyme[45]. Fgf8 is shown to regulate the out flow tract and right ventricle development. Fgf9 is required for myocardial proliferation and coronary vascular development. The ablation of FGFR also shows abnormality during development. The Fgfr1 knockout mice are lethal during gastrulation showing defective mesoderm patterning[46].

1.2.3.3 Notch Signaling Pathway

Notch signaling pathway is a conserved signaling pathway in multiple cellular organisms that plays a role in cell fate specification, differentiation, and patterning. There are four types of Notch receptors in mammals: NOTCH1, NOTCH2, NOTCH3, and NOTCH4. The Notch receptor consists of a large extracellular domain (NECD), a single-pass transmembrane and a shorter intracellular domain (NICD). The extracellular domain can interact with the membrane-bound ligands expressed by neighbor cells. The interaction depends on the E3 ubiquitination and leads to the endocytosis of ligands and the Notch receptors will be cleaved at the S2 site. The remaining Notch receptors were cleaved at S3 site by γ-secretase and release the NICD. The NICD then translocates to the nucleus and interacts with transcription factor RBPJ. In the absent of Notch signaling, the RBPJ recruit the transcription co-activator MAML and activate the bHLH repressor Hes and Hey transcription family.

The expression of Notch pathway starts from E7.5 endocardium. N1ICD and ligand Dll4 are detectable during this stage[47]. At E8.0, the Notch ligand Jag1 is expressed in myocardium and Notch2, Notch4, Dll4, and N1ICD are in endocardium[48-50]. As the progression of embryo development continues, the expression of Notch pathway after E9.5 showed complex expression to coordinate the looped heart and left-right asymmetry. The mutation of factors in Notch pathway usually leads to a congenital heart disease such as Alagille syndrome, bicuspid aortic valve disease, and calcification of heart valves. The ablation of *Dll4* is embryonic lethal at E9.5 and showed reduced atrial

and ventricular chambers. Notch1 knockout mice is also embryonic lethal at E10.5 and showed hypoplastic cardiac cushions. Overexpression of Notch pathway also showed abnormal heart development. Overexpressing *N1CID* by *Mesp1-Cre* showed impaired ventricular myocardial differentiation and activation of *Hey1* by *Mesp1-Cre* showed reduction of atrioventricular canal[51].

1.2.3.4 Wnt Signaling Pathway

The Wnt signaling pathways play roles in cell communication, passing signals from outside the cellular membrane through surface receptors into cells and can be either paracrine or autocrine[52, 53]. Wnt proteins can interact with their signal mediator, β-catenin. In the absent of Wnt, β-catenin forms complexes with Axin, APC, and GSK and is phosphorylated to be subsequently degraded by proteasomes. With the presence of Wnt, β-catenin is released from complex and enters the nucleus to regulate downstream genes through TCF/LEF transcription factors.

There are three major Wnt pathways: canonical Wnt pathways, non-canonical Wnt/JNK planar cell polarity pathways, and non-canonical Wnt/calcium pathways. Canonical Wnt pathways mediate cell proliferation through regulating genes such as *c-myc* and *cyclinD1* [54] and this can be inhibited by Dkk-1[55]. The non-canonical pathways are mediated by β-catenin independent signaling cascade. In Wnt/JNK pathways, Wnt signaling is carried by Rho-JNK kinase cascade while Wnt/Ca pathways are regulated by calmodulin-depend kinase (CaMKII) cascades[56]. The recent reports showed non-

canonical Wnt pathways could crosstalk to canonical Wnt pathways through different mechanisms[57].

The inhibition of Wnt pathways is essential during myocardial specification. The tissue specific ablation of β-catenin leads to the formation of multiple hearts[58]. However, there are also reports that suggest the positive roles of Wnt pathways during P19 teratocarcinoma stem cell differentiation[59]. These controversial results demonstrate the complex and multiple functions of Wnt pathways during developmental processes.

1.3 Stem Cell and Research

1.3.1 Stem Cell Research

The signaling pathway in mouse heart development showed the complexity of signaling pathways. Some genetic ablation mouse model even showed controversial roles. It made the research on single factor become impractical. The in vitro stem cell research could simplify the roles of individual factor and make manipulating cells easier. It also reduced time and costs comparing using animal models. The manipulated stem cells may also benefit to experiments such as performing cell therapy in injury model.

The stem cells can be isolated from inner cell mass (ICM) of blastocyst and cultured in vivo while maintaining their pluripotency[60, 61]. Stem cells have the differentiate potential to cardiomyocytes, neurons, islet cells, and blood cells and are considered very high valuable source for basic research and clinical application. Mouse embryonic stem

cells (mESC) can be differentiated in vitro by force aggregating. The most established way is hanging drop method, a method that aggregate mESC by gravity in a droplet. The mESC will form enbryoid body (EB) and EB will beat spontaneously about one week after aggregating. Recently report showed expression of Gata4, Tbx5, and Baf60c was sufficient to convert noncardiac mesoderm into beating cardiomyocyte[62]. Keller Lab also demosrate the optimized treatment of signals such as BMP and Activin could promote cardiac differentiation[63]. The stem cell research provides an insight method to better understand molecular mechanisms during cardiac development.

1.3.2 Reprogramming Somatic and IPS Cell Types

The terminally differentiated cells were considered lost the proliferating ability. By introducing stem cell associated factors *Oct4*, *Sox2*, *cMyc*, and *Klf4*, Yamanaka and colleagues showed the converting of somatic cell back to pluripotent cells (induced pluripotent cells, iPS cells)[64]. This strategy bypasses the potential ethical issue using human embryonic stem cells. The iPS cells could provide the source of patient specific cells for cell therapies. It could also establish a genetic disease model from patients for basic research.

The concept of converting cell fates from one to another (transdifferentiation) had been proven successful. People further test whether adult cells can be transdifferentiated toward cardiomyocyte or not. Recent report suggested by introducing transcription factors *Gata4*, *Mef2c*, *Tbx5*, and *Hand2* could transdifferentiate fibroblast into

cardiomyocyte[65, 66]. To sum up, manipulating adult cells by introducing exogenous genes or essential signaling molecules provide a practical ways to perform stem cell therapies.

1.4 Gene Regulations and Transcription Factors

1.4.1 Gene Regulation

In multiple cellular organisms, developing cells are surrounded by a constantly changing environment. They can receive and be affected by extracellular signals through the physical interaction with neighboring cells by secreted molecules from remote cells. Developing cells are guided by differing signals to respond to distinct developmental cues such as cell proliferation, lineage specification, differentiation, left-right patterning, and apoptosis. The progression through development requires that signals quickly find and communicate their message to its desired target(s) so that the developmental process can be achieved in the right place and right time. This involves the tight and orchestrated regulation at the molecular level for cells to be propelled into the stages on development and differentiation.

There are several ways to regulate gene expressions including transcriptional regulation, post-transcriptional regulation, translational regulation, post-translational regulation, and epigenetic regulation. Transcription is usually initiated by the interaction of DNA and transcription factors. The transcription factors bind on DNA by recognizing specific region and recruit the RNA polymerase near the transcription complex. The RNA

polymerase then starts transcription and synthesizes mRNA for protein synthesis. In some cases, however, the transcription factors may also function as negative regulators by competing for the DNA binding sites with other transcription factors and subsequently blocks gene transcription. Post-transcriptional regulation occurs after the gene transcript has been made and is blocked from undergoing translation either by degradation or by regulatory miRNAs that bind to the 3'UTR of mRNA to prevent the ribosomal complex formation. Post-translational regulation is regulating the protein activities or half-life by modifying the protein by means of phosphorylation and tagging the protein for ubiquitination. Epigenetic regulation means regulating the genes activity without changing or directly binding to the DNA sequences but by modifying or interacting with the histone protein complexes the DNA is coiled around. By manipulating the scaffold in which the DNA strands are coiled around, we can promote or prevent a gene's expression. In this study, we will focus on the transcriptional level of gene expression regulation during the developmental process.

1.4.2 Transcription Regulation

Transcription factors are proteins with DNA binding domains that bind to genomic DNA by recognizing a specific sequence motif or genomic structural feature. The DNA binding domains recognize specific DNA sequences and the trans-activating domain can interact with other protein such as co-activator or co-repressor. The transcription factors have been primarily grouped into different TF families based on what their DNA binding

domain specifically interacts with such as basic helix-loop-helix, basic-leucine zipper, zinc fingers, homeodomain, and helix-turn-helix.

Transcription factors can bind genomic DNA in diverse regions of the genome including promoter, enhancer, and intergenic regions. They can also be found within the intron regions of other genes. Usually, the promoters are regions that are near the transcription start site (TSS) (within 5kb upstream of the start site). Enhancers are genomic regions that can promote transcription of a gene that may not be located near the gene's promoter and seem too far from the transcription start site but enhancers can be geometrically close and form the transcription complex. The promoters and enhancers of genes are usually evolutional conserved and have been intensively studied in developmental biology. The transgenic mice consisted of a specific promoter driven reporter that was widely used to study the expression pattern of genes. For example, Nkx2-5 had been shown to regulate cardiac development. By dissecting the different Nkx2-5 upstream regions, several regions were identified to regulate different regions of heart including the right atria, right ventricle, cardiac crescent and inter-ventricular septum.

1.4.3 The Roles of Mesp1 during Cardiac Development

Mesp1 is a bHLH (basic helix-loop-helix) transcription factor that was first identified by screening enriched transcripts in the posterior region of mouse embryo at E7.5 stage and named "mesoderm posterior 1"[67]. The earliest expression in mouse embryo can be

detected in the initiation stage of gastrulation around E6.25-6.5 at the junction of epiblast and extraembryonic ectoderm. It reaches the highest expression level in nascent mesoderm and primitive streak at E6.5-6.75. Mesp1 expression is soon down regulated after E7.5 and localized only in the base of allantois. The expression moves from the base of allantois to the edge of the tailbud and almost disappears after E8.5. The only adult tissue that expresses Mesp1 is mature testis. The Mesp1 lineage tracing was designed by using Cre-loxP system. It showed Mesp1 expressing cells ingresses through primitive streak and cooperates with the mesodermal part of amnion. It contributed to the myocardium but not the endocardium of the later heart. The ablation of Mesp1 is embryonic lethal due to cardia bifida. This suggested the essential role of Mesp1 during cardiac development. Flk1 expression was accumulated in the posterior region in Mesp1 deficient mice. Genes that regulated cardiac development such as Hand2 and Gata4, however, were not significantly altered[68, 69]. This means the roles of Mesp1 during cardiac development is in the specification of cardiac progenitors. However knocking out both Mesp1 and Mesp2 leads to the absence of the heart[70]. This result suggested the functional role of Mesp1 during cardiac mesoderm development and might be compensated by Mesp2 when Mesp1 is lost.

Recent studies of Mesp1 from molecular approaches suggest Mesp1 regulates downstream targets through DNA binding on the promoter region of Dkk1, a Wnt signaling inhibitor that plays critical roles in cardiovascular differentiation. It is suggested that Mesp1 regulates cardiovascular restriction through regulating core cardiac

developmental genes such as Gata4, Mef2c, Tbx5, and Hand2. It also regulates Epithelial-Mesenchymal Transition (EMT) through induction of Snai1, which agrees the roles of Mesp1 during gastrulation of early mouse embryos[71]. A similar study suggested that Mesp1 regulates the core cardiac transcription factors also through binding to their promoter regions[72]. These molecular studies define Mesp1 as the most critical factor for cardiovascular development.

1.4.4 ETS Transcription Factors

The ETS transcription factors are named after E-twenty-six or E26 transforming specific factors. They belong to ETS transcription factor family that all contain an ETS DNA binding domain. It was first identified from the gene transduced by E26 leukemia virus. The ETS transcription factors were reported to be associated with cancer progression in different tissues. The DNA binding ETS domain is an evolutionally conserved region that recognizes DNA specifically through a common sequence GGA(A/T). Besides the ETS domain, some ETS transcription factors consist of conserved PNT (Pointed) domain, which could associate with potential binding partners to regulate different genes.

The expression patterns of all ETS transcription factors in human tissues and cell lines are ubiquitous. There were 16 ETS transcription factors expressed in all tissues and more than half of all ETS transcription factors were considered as ubiquitous. The mRNA level may not be the best way to identify tissue specific ETS transcription factors.

However, some ETS transcription factors are highly expressed in certain tissues. For example, *ERG* and *FLI1* expression is higher in endothelial cells, which explains their roles in endothelial development[73, 74]. Another example is PEA3 families that are highly expressed in cancer cell lines, which match their roles in tumor progression.

Although ETS transcription factors all bind a similar DNA sequence, GGA(A/T), the genome wide ChIP (chromatin immunoprecipitation) arrays or ChIP-sequencing results provide the detailed binding specificities of different ETS transcription factors in different tissues. One report showed the binding regions of GABPA in Jurkat cells were also occupied by ETS1. The shared regions bound by different ETS transcription factors are often located in the promoters of house-keeping genes. However, the DNA binding sequences were also reported to be specific for only some ETS transcription factors. For example, ERG and FLI1 shared redundant groups of genes while SPI1 was less overlapping with these regions in hematopoietic stem cells.

The ETS transcription factors could also regulate target genes through thr combination of different association partners within transcription complexes. The interactions of Serum Response Factors (SRF) and TCF subfamily member such as ELK1, ELK3, and ELK4 have been well studied. The co-occupancy of genomic regions by ETS1 and RUNX1 was characterized and the target genes regulated by this interaction were *ZAP70*, *CXCR4*, and *CD38* during T cell activation. These studies demonstrated the

multiple roles of ETS transcription factors during the developmental process of different tissues.

The ETS transcription factors may play multiple roles during embryo development. The following are some examples how ETS transcription factors regulate cardiovascular development. ETS-related protein 71 (Etsrp71, human ETV2) is expressed in endothelium and endocardium of the developing heart. It was shown that Nkx2-5 binds to the promoter regions of ETV2 and the ablation of Nkx2-5 causes down regulation of ETV2 and leads to embryonic lethality. The regulation of ETV2 during cardiac development was shown through direct regulation of *Tie2* expression. It was also shown that the cardiac gene *Mef2c* could be regulated by the combination of Etv2 and Foxc2. Combination of Etv2 and Foxc2 activate *Flk1* expression in *Xenopus* embryos and is also required for vascular development in Zebrafish. These studies provide evidence that ETS transcription factors could be involved in regulating cardiac development through the combination of assembling proteins in the nearby regulatory, chromatin regions or in the transcription complexes.

2. MATERIALS AND METHODS

2.1 Embryonic Stem Cell Culture

Mouse embryonic stem (ES) cell line, AB2.2, was purchased from ATCC (American Type Culture Collection) and stored in liquid nitrogen until ready for use. Dr. Yu Liu from the University of Houston generated UH3 cell line used throughout this work. Dr. Yu Liu from the University of Houston also generated Mesp1 doxycycline inducible ES cells. To be brief, AB2.2 cells were infected with the lentivirus carrying inducible Mesp1 construct and the lentivirus carrying rtTA (Tet-on 3G trans-activator) construct. ES cells were put under puromycin selection to obtain a stable cell line with the desired construct.

Cells were grown with M15 ES growth media that is DMEM (SH30022, HyClone) supplemented with 15% ES defined fetal bovine serum (FBS) (SH30070.03, HyClone), L-Glutamine (25030, Gibco), Penicillin-Streptomycin (15140 Gibco), 100 mM of β-mercaptoethanol (M7522, Sigma Aldrich). ES cells were sub-cultured every other day and media changed daily supplemented with 1000 U/ mL of LIF (ESG1107, Millipore) to maintain pluripotency.

ES cells were differentiated to Embryoid Bodies (EB) by hanging drop as previously described. The ES cells were harvested and cell density is determined in 2×10^4 cells/ml and hung in 20 µl volume droplet on the cap of sterile, square dish (400 cells/drop). The

lids containing the droplets were inverted and placed back onto the base of the plate with10 ml of sterile water to maintain moisture/humidity content. The square dishes were maintained at 37°C, 5% CO_2 and at 95% humidity. Resulting EBs were collected at different times based on the experimental design. For single cell analysis, harvested EBs were trypsinized at 37°C for five minutes if necessary.

2.2 Normal Human Dermal Fibroblast and Reprogrammed Cell Culture

Normal Human Dermal Fibroblasts (NHDF) were purchase from Lonza (CC-2509, Lonza) and cultured with FBM medium (CC-3131) supplemented with FGM-2 SingleQuots (CC-4126). NHDF cells were cultured following the manufacturer's recommendations and sub-cultured weekly.

Reprogrammed cells were cultured in StemPro® hES SFM (A1000701, Invitrogen) and reprogrammed cardiac progenitors were grown with alpha MEM medium supplemented with 15% FBS, 0.1 mM β-mercaptoethanol, 2 mM glutamine (25030, Gibco) and 0.1 mM nonessential amino acids (11140, Invitrogen).

2.3 Lentivirus Production

Lentiviruses were produced by 293FT cells. 293FT cells were grown under DMEM supplemented with 10% FBS (SH30910), sodium pyruvate (11360, Gibco), L-Glutamine (25030, Gibco), MEM-NEAA (11140, Gibco) and Penicillin-Streptomycin (15140 Gibco). To produce lentivirus, 2.5×10^6 cells were seeded on 60 mm plate one day

before transfection. The next day, 4.5 μg of lentiviral plasmid, 1.87 μg of pMD2.G and 2.8 μg of psPAX2 were transfected to 293FT cell by FuGene HD (4709705001, Roche). Lentiviral particles were collected 24 to 48 hours after transfection and filtered with 0.45 μm filter following 0.22 μm filter then stored at -80°C until used.

2.4 Quantitative Real-Time PCR

RNA was isolated by RNeasy kit following the manual. To avoid genomic DNA contamination, on-column DNA digestion step was added during RNA isolation. RNA concentrations were determined by Spectrophotometer (NanoDrop 2000, Thermo Scientific). 2 μg of RNA were used to perform reverse transcription by M-MLV reverse transcription kit (28025-013, Life Technologies) following the manufacturer's recommended procedure. After the reverse transcription step, the cDNA samples were diluted to the final concentration of 10 ng/μl for the qPCR reactions. Brilliant II SYBR Green QPCR Master Mix (600828, Agilent Technologies) was used to perform qPCR and reactions were run on HT7900 Real-Time PCR System (Life Technologies). The qPCR primers of each gene measured are listed in Table 1 and Table 2.

Table 1 Primer List of Mouse Genes

Gene	Forward	Reverse
Acvr2a	TTCAAATCCTGTTACACCGAAGC	AGGAGGGTAGGCCATCTTGTG
Cxcr4	GAAGTGGGTTCTGGAGACTAT	TTGCCGACTATGCCAGTCAAG
Ehf	TCAAATGGGAAGACCGTTCGG	CCGGCTGAGCTTCTCGTAT
Elf1	TGTCCAACAGAACGACCTAGT	CACACAAGCTAGACCAGCATAA
Elf2	GTTCACAGCAGTAATGCTCACT	TCAAGCAGGTAGGAGATTCCAT
Elf3	GCTGCCACCTGTGAGATCAG	GTGCCAAAGGTAGTCGGAGG
Elf4	CTGGAAGACCCTTCAGTGTTC	GCACAAGGTCCTGTCCCTTATG
Elf5	ATGTTGGACTCCGTAACCCAT	GCAGGGTAGTAGTCTTCATTGCT
Elk1	TCCTGGACCTCACGGGATG	GGGTAGGACACAAACTTGTAGAC
Elk3	TCCTCACGCGGTAGAGATCAG	GTGGAGGTACTCGTTGCGG
Elk4	ATCTAACAATGGGGAGTTCAAGC	GGCTCGGCTGAGTTTATCATAAT
Eomes	CCCTATGGCTCAAATTCCAC	AGAACCACTTCCACGAAAACAT
Erf	GGTTTGCCTTCCCAGATTGG	CCTGGTACTCCTCTTTCCGAA
Erg	ACCTCACCCCTCAGTCCAAA	TGGTCGGTCCCAGGATCTG
Ets1	TCCTATCAGCTCGGAAGAACTC	TCTTGCTTGATGGCAAAGTAGTC
Ets2	CCTGTCGCCAACAGTTTTCG	TGGAGTGTCTGATCTTCACTGA
Etv1	TTAAGTGCAGGCGTCTTCTTC	GGAGGCCATGAAAAGCCAAA
Etv2	CTGGGAGCGGAATTTGGTTTC	GTAAAGCGGGGTTCCAGTCC
Etv3	ATGAAAGCAGGCTGTAGCATC	ATGGCGGAACTCTTCCTTCTG
Etv4	CGGAGGATGAAAGGCGGATAC	TCTTGGAAGTGACTGAGGTCC
Etv5	TCAGTCTGATAACTTGGTGCTTC	GGCTTCCTATCGTAGGCACAA
Etv6	AGCAGGAACGAATTTCATACACG	GGCAGGTGGATCGAGTCTTC
Fev	ACGCCTACCGCTTTGACTTC	AAGCTGCCATCAAGTTGAGTT
Fli1	ATGGACGGGACTATTAAGGAGG	GAAGCAGTCATATCTGCCTTGG
Foxa2	TCCGACTGGAGCAGCTACTAC	GCGCCCACATAGGATGACA
Gabpa	AGCGCATCTCGTTGAAGAAG	TCCTGCTCTTTTCTGTAGCCT
Gapdh	AGGTCGGTGTGAACGGATTTG	TGTAGACCATGTAGTTGAGGTCA
Gata1	GCACTAACTGTCAAACGACCAC	TCTGGATTCCATCTTTCCTCAT
Gata2	CACCCCGCCGTATTGAATG	CCTGCGAGTCGAGATGGTTG
Gata4	CACGCTGTGGCGTCGTAAT	CTGGTTTGAATCCCCTCCTTC
Gata6	CTTGCGGGCTCTATATGAAACT	TAGGTTTTCGTTTCCTGGTTTG
Gsc	CCCCGGTTCTGTACTGGTG	TCTGGGTACTTCGTCTCCTGG
Hand2	CTACATCGCCTACCTCATGGAT	TGCTCACTGTGCTTTTCAAGAT
Hhex	CGGACGGTGAACGACTACAC	CGTTGGAGAACCTCACTTGAC
Isl1	ATGATGGTGGTTTACAGGCTAAC	TCGATGCTACTTCACTGCCAG
Mef2c	ATGCCATCAGTGAATCAAAGGAT	GTGGTACGGTCTCCCAACT
Mesp1	GCTCGGTCCCCGTTTAAGC	ACGATGGGTCCCACGATTCT
Mesp2	GCCTAGGAACAAGACTGGACAC	AGGGTTCTGGAGACACAGAAAG
Mixl1	CATGTACCCAGACATCCACTTG	CTTGAGGATAAGGGCTGAAATG
Myocd	ACCACTGAGCAATACCCCTCTA	CGGATTCGAAGCTGTTGTCTTA
Nkx2-5	GACAAAGCCGAGACGGATGG	CTGTCGCTTGCACTTGTAGC
Pdgfra	GGAGACTCAAGTAACCTTGCAC	TCAGTTCTGACGTTGCTTTCAA
Pitx2	ACCCCGGCTATTCGTACAAC	GAGGACAGGGGATTGACGTTC
Sall4	TCCAACATTTATCCGAGCACAG	TGGCAGACGAGAAGTTCTTTC
Sfpi1	ATGTTACAGGCGTGCAAAATGG	TGATCGCTATGGCTTTCTCCA
Smad1	GCTTCGTGAAGGGTTGGGG	CGGATGAAATAGGATTGTGGGG
Smo	CAATCGCTACCCTGCGGTTAT	CTGCTCGGCAAACAATCTCTC
Sox17	GATGCGGGATACGCCAGTG	CCACCACCTCGCCTTTCAC
Spdef	AAGGCAGCATCAGGAGCAATG	CTGTCAATGACGGGACACTG

Table 1 (Continued)

Gene	Forward	Reverse
Spib	AGGAGTCTTCTACGACCTGGA	GAAGGCTTCATAGGGAGCGAT
Spic	AAACATTTCAAGACGCCATTGAC	CTCTGACGTGAGGATAAGGGT
T	GCTTCAAGGAGCTAACTAACGAG	CCAGCAAGAAAGAGTACATGGC
Tal1	CGGCAGCAGAATGTGAATGG	CTCCTGGTCATTGAGTAACTTGG
Tbx20	AAACCCCTGGAACAATTTGTGG	CATCTCTTCGCTGGGGATGAT

Table 2 Primer List of Human Genes

Gene	Forward	Reverse
ACTA2	CGGTGCTGTCTCTCTATGCC	CACGCTCAGTCAGGATCTTCA
ACVR2B	GCCTCTGGCAAATGAGTGAAG	TTCAACTGCTACGATAGGCAGG
CTNNB1	TACCTCCCAAGTCCTGTATGAG	TGAGCAGCATCAAACTGTGTAG
BMP2	ACCTTTATGGAGGGAAACCCA	CCGGATCTGGTTCAAGCATGA
BMP4	TGGTCTTGAGTATCCTGAGCG	GCTGAGGTTAAAGAGGAAACGA
CITED1	ACCAGCGACATCAGCACTTC	GACCCCTGCAAAACTGGGAC
CX43	AGAGGAAGAACTCAAGGTTGCC	AGGCCACCTCAAAGATAGACT
CX45	AAGATTGCCAAAATGGAGCACG	CTCCTCCGTTTCTTCCAGAGC
DKK1	ATTCCAACGCTATCAAGAACC	CCAAGGTGCTATGATCATTACC
FOXH1	GTCGTAGATGCTTTTGTTGGGT	CTGCCCACCTCCTACTTGC
GAPDH	TGTTGCCATCAATGACCCCTT	CTCCACGACGTACTCAGCG
GATA2	CATCTTCATGCTCTCCGTCAG	GGCCCACTCTCTGTGTACC
GATA4	CATGGGCCCTCCATCCACCCTG	CCTCCAAGTCCCAGGTCCGTGC
GATA6	AGAAACGCCGAGGGTGAAC	AGTTGGAGTCATGGGAATGGAA
HAND2	ATGAGTCTGGTAGGTGGTTTTCC	CATACTCGGGGCTGTAGGACA
ISL1	GTGGAGAGGGCCAGTCTAGG	CCGTCATCTCTACCAGTTGCT
KDR	GCACAAAGTGACACGTTGAGAT	AGTGATCGGAAATGACACTGGA
LEFTY1	CTGAGCAAGGGCTCTCCAGTGG	AGCTGGCGATGACTGAACTGCTG
MEF2B	CACTGAGAGCGTCTACGTCCT	CTGCATGTCTGCCTTGTGC
MEF2C	ATGGATGAACGTAACAGACAGGT	CGGCTCGTTGTACTCCGTG
MESP1	TCGAAGTGGTTCCTTGGCAGAC	CCTCCTGCTTGCCTCAAAGTGTC
MESP2	CGCTGCGCCTGGCCATCCGCTACAT	GCCCCAAGGGGACCCCGCGAC
MYL2	TTGGGCGAGTGAACGTGAAAA	CCGAACGTAATCAGCCTTCAG
MYL7	CAGCATAGATGATCGAGAGGAGC	TAGGAGGTCCGAAGAGTTCAG
MYH3	CAACAATCCCTACGACTACGC	ACGTCAAAGGCACTATCGGTG
MYH6	CACTGATAACGCTTTTGATGTGC	TAGGCAGACTTGTCAGCCTCT
MYH7	TTTGAACAAGCCCAGATACAGG	CTCGGGGTCTGTCCCATTG
NKX2-5	CCCTGACCGATCCCACCTCAAC	GGCGGGCGACGGCGAGATAGC
NOS2	TCATCCGCTATGCTGGCTAC	CTCAGGGTCACGGCCATTG
PDGFRA	CCTGGTCTTAGGCTGTCTTCT	GCCAGCTCACTTCACTCTCC
PECAM1	GGAAAAGGCCCCAATACACTT	TAAAACGCGGTCCTGTTCCTC
SMAD2	CTCAAGCTCATCTAATCGTCCTG	ATGTCGTCCATCTTGCCATTC
SMAD3	ACGCAGACCTCGTCCTTCT	GAACGTCAACACCAAGTGCAT
SMARCD3	CCACCTCGTTGAGTGGCATC	AGCCTGTTGGTCTTCACATACT
SOX17	ACTCTGGCAGTCGCGGTAGTGGC	GCGGCGCAAGCAGGTGAAG
TAGLN2	CTGCTTGCAGGAACTGAGAGA	ACCCAGTGCCGAAAGGATG
TBX20	TCCAGATTCTCCTTTTACCG	TTCAGACTTCAGGTTGAGCA
TBX2	ACAGCTCACAGTGGATGGTAG	GCACTCGGGGTTTGTACTTATG
TBX3	TTAAAGTGAGATGTTCTGGGCTG	ACTATAATTCCCCTGCCACGTA
TBX5	CACTTCTCCGCTCACTTCACC	TGGCACGCCATGAGAGTAGA
TDGF1	AGATGGACGAGCAAATTCCTG	AAGATGGCCCGCTTCTCTTAC
TNNT1	TCTTCCGCTGCTCGAAATGTA	TGATCCCGCCAAAGATCCC
TNNT2	GAGTTGCAGGCGCTGATTG	TCTGGATGTAACCCCCAAAATG
WNT2	CCCCCGAGGTCAACTCTTC	CTAATGGCACGCATCACATCT
WNT11	CTGACCTCAAGACCCGATACC	CGAGTTCCGAGTCCTTCACAG
ZIC3	AGTAGCAGACGTGGTTGTTCT	CGTTATATGCGGCAGCCTATC

2.5 Western Blot and Co-Immunoprecipitation

RIPA buffer (50 mM Tris, pH8; 150 mM NaCl; 1% NP40; 0.5% Sodium Deoxycholate; 0.1% SDS) supplemented with Protease Inhibitor Cocktail Tablet (05892791001, Roche) were used to lyse the cells. Generally, 10^7 cells were lysed by 1 ml of RIPA buffer on ice for 30 minutes then sonicated for 20 seconds to shear DNA if sample was viscous. The protein concentrations were determined by BCA Protein Assay Kit (23225, Pierce). Samples were denatured by boiling for 10 minutes with NuPAGE LDS Sample Buffer (NP0007, Life Technologies). Novex NuPAGE Gel Electrophoresis System was used to conduct Western Blotting follow the manufacturer's instruction. Briefly, 10 to 50 μg of protein was loaded onto a 4-12% Bis-Tris Protein Gel and separated by electrophoresis. The protein was transferred to PVDF membranes for 90 minutes then blocked with 5% TBST milk for an hour. The PVDF membranes were incubated with the primary antibody at 4°C overnight gently rocking and secondary antibodies for an hour at room temperature the following day. Membranes were incubated with enhanced chemi-luminescence (ECL) (NEL10400, Perkin Elmer) and visualized by X-ray film exposure. The information of each antibody was listed in Table 3.

Co-Immunoprecipitation (Co-IP) was performed by the similar protocol of Western blot. The cells were lysed and Anti-HA Afftinity Gel (E6779, Sigma) or Anti-FLAG Affinity Gel (F2426, Sigma) were used for immunoprecipitation. Generally, 50 μl of affinity gel were used to precipitate 1000 μg of total cell lysate at 4°C overnight. Samples were eluted by 0.1M of acetic glycine solution (pH 2.8) from affinity gel and 1M of Tris

buffer (pH 8.0) were used to neutralize the elution reaction. The eluted sample and total cell lysates were subject to Western blot as previously discussed.

2.6 Flow Cytometry Analysis and Sorting

Cells carrying endogenous fluorescence such as eGFP, eYFP or Nkx2-5-tdTomato reporter were harvested and dissociated by trypsinization and resuspended in culture media. Cells were subject to fluorescence analysis or Fluorescence-Activated Cell Sorting (FACS). Sorted cells were collected in regular culture media.

Cells subjected to surface marker analysis were harvested and dissociated by trypsinization. Surface markers sensitive to trypsinization were harvested and dissociated by Accutase. Cells were resuspended at 1×10^5 cells/100 μl in Fixation Buffer (R&D) for 10 minutes at room temperature. Centrifuged cells and decanted the Fixation Buffer and wash with PBS 2 times. Resuspended cells in 100 μl of Fixation/Permeablization buffer and add 1-5 μg/ml of primary antibody (follow the recommended dilution from manual) and incubated on ice in the dark for 30 minutes. Washed the cells with PBS 2 times and add fluorescence conjugated secondary antibody or fluorescence conjugated streptavidin (if biotinylated primary antibody were used) following the recommended dilution and incubated 20 minutes on ice in the dark. Washed the cells with PBS 2 times and resuspended with 500 ml of PBS and subject to flow cytometry.

Table 3 Antibodies Used for Western Blot, Co-IP, and Flow Cytometry

Western Blot

Name	Manufactory	Catalog Number	Dilution
Mesp1	Abgent	AP16455c	1000
Mesp1	Sigma	SAB1302562	1000
MESP1	Abcam	ab77013	1000
MESP1	Abcam	ab171427	1000
HA	Covance	MMS-101P	2000
ACTC1	Abcam	ab28052	1000
TNNT2	Abcam	ab8295	500
MHC	U of Iowa	MF20	500
TNNI3	Cell Signaling	#4002	1000
GAPDH	Santa Cruz	sc-20357 HRP	5000
Rabbit IgG	Santa Cruz	sc-2301	5000
Mouse IgG	Santa Cruz	sc-2302	5000

Flow Cytometry

Name	Manufactory	Catalog Number	Dilution
CD31	BD Pharmingen	558094	100
CD31*	BD Pharmingen	560984	100
CD144	eBioscience	56-1449	100
CD309	R&D Systems	FAB357A	100
CD309*	BD Pharmingen	560872	100
CD34	BioLegend	343523	100
CD140a	BioLegend	323503	100
CD184	BioLegend	306504	100
CD117	BioLegend	313208	100

Immunofluorescence Staining

Name	Manufactory	Catalog Number	Dilution
GFP	Abcam	ab6658	800
Foxa2	DSHB	4C7	100
Gata1	Santa Cruz	sc-265	200
Nkx2-5	Abcam	ab35842	50
Runx1	Santa Cruz	sc-365644	200
Pea3	Abcam	ab70425	200

* Antibody used in Islas, Liu, Weng et, al. PNAS, 2012; **109**(32): 13016-21

2.7 Isolation of ChIP-Enriched DNA

This procedure is ideally started with 10-20 million cultured cells; however, we also applied the same general method for starting with 100k cells. Cells were first fixed using 1 % formaldehyde for 10 minutes and then quenched for 0.125 M glycine for 15 minutes. Nuclei were isolated by washing the cell pellets with 5mM PIPES, 85mM KCl, and 0.5% NP-40 buffer and then spinning for 3 minutes at 3000rpm. A 1% SDS buffer was used for both nuclear lysis and subsequent sonication to shear the chromatin. For sonication with the Bioruptor UCD-300 sonicator, we used a 50-100µl volume/sample in 1.5ml eppendorf tubes. Sonication was performed for 15 cycles at 1 minute/ cycle. We performed 3 sets of 5 consecutive cycles on the "high" setting, which corresponds to 5W. The target size of sheared DNA was 200 to 300 bp. After shearing to the ideal size, sheared chromatin was diluted in ChIP-IP buffer and pre-cleared for 1 hour at room temperature with Protein-G Dynabeads from Invitrogen. 100uL of pre-cleared sheared chromatin was used per immunoprecipitation with a specific antibody for Mesp1. Samples were probed overnight at 4°C gently rocking. The following day samples were then incubated with Dynabeads conjugated with Protein-G from Invitrogen for 1 hour at 4°C and 30 minutes at room temperature. Beads were then washed with Low and High salt washes and 2 final washes with Tris EDTA buffer before being eluted. Eluted samples were treated with RNAse and Proteinase K before being reverse cross-linked at 65°C overnight. DNA was purified using Qiagen Minelute columns following the recommended protocol. Resulting DNA was quantified using the picogreen assay from Promega and the Quantifluor fluorometer. Mesp1 ChIP-DNA was obtained from both

eYFP + and negative cells from 4 biological replicate samples along with 3 replicate samples from undifferentiated UH3 cells.

2.8 ChIP-Seq Libraries for Next Generation Sequencing

Resulting DNA fragments isolated from above were then used to generate Next Generation Sequencing (NGS) libraries for the Illumina HiSeq platform. We used the NEXTflex^{TM} ChIP-Seq kit from BIOO Scientific (5143-02) for generating paired end, barcoded libraries that could be multiplexed. NGS libraries were prepared following the manufacturer's recommended procedure up to the final PCR step. The final library amplification was done using KAPA biosciences HiFi polymerase. Final libraries were cleaned up using Ampure XP bead purification from Beckman Coulter and eluted in water. Eluted libraries were quantified using Promega's picogreen assay and Quantiflour fluorometer. Libraries were multiplexed and 30μL submitted for sequencing at a 1nM multiplexed concentration to MDAnderson's DNA core facility.

2.9 Sequencing Data Analysis

Sequencing raw data were aligned to the mouse reference genome (Ensembl 67) using bowtie with parameters: *-t -v 3 -a -m 1 --best --strata*. Duplicate and non-uniquely mapped reads were removed for downstream analysis. We used cisGenome to detect Mesp1 peaks. Similar to MACS (Model-based Analysis of ChIP-Seq), cisGenome applies a local Poisson background model and handles replicates. We used a local

Poisson P-value < 0.0001. We annotated the peaks into promoter (+/- 2kb within the TSS), integenic, intron, and exon.

In order to identify enriched motifs, FIMO (Find Individual Motif Occurrences) was used to scan for presence of known motifs in peaks with P-value < 0.001. A logistic regression model was used to model the motif occupancy in the peaks against a background set of sequences obtained by shuffling the peak sequences. The Z-values obtained from the regression were used to rank the motifs. High values of Z-values correspond to motifs with high rank.

2.10 RNA Isolation and NGS Library Preparation

RNA was isolated from undifferentiated and FACS sorted UH3 cells Days 5-8 post-differentiation. Cells were sorted as described above for EYFP+. Both the positive and negative populations were used for transcriptome analysis. RNA was extracted from 250k cells for each group using the RNAeasy kit from Qiagen following the manufacturer's recommended procedure. Following isolation the RNA was quantified using the Ribogreen quantification kit from Promega along with Promega's Quantifluor fluorometer.

A total of 100ng of the isolated RNA was used for generating the library for next generation sequencing using Illumina's HiSeq platform. Libraries were prepared using the ScriptSeq Complete Gold kit from Epicentre (SCL24EP) following the

manufacturers recommended procedure for low input starting material. Final libraries were quantified using the pico-green quantification kit from Promega along with Promega's Quantifluor fluorometer.

2.11 RNA-Seq Data Analysis

The raw reads from RNA-seq data were aligned to mouse genome (mm9) using bowtie and tophat2]. From the mapped reads, transcripts were assembled using cufflink. The abundances of transcripts are computed using cufflinks and expressed as FPKM (Fragments Per Kilobase of transcript per Million mapped reads). Assembled transcripts are obtained separately for 26 RNA-sequence samples (ES, Day 5, 6, 7 and 8 yfp+/yfp- samples). Using cuffmerge, the assembled transcripts from different samples are merged to obtain a unique list of assembled transcripts. To filter the annotated RNAs, annotated transcripts from UCSC genome browser (including the RefSeq genes) and Ensemble genome browser are obtained. Using cuffcompare, all annotated transcripts are removed from our unique list of assembled transcripts. After these steps, the remaining pool of transcripts is processed through a bioinformatics pipeline in order to remove transcripts, which are short or potentially protein coding.

2.12 Luciferase Assay

The Nkx2-5 promoter region was amplified from total mouse genomic DNA by PCR with specific primers to the region of interest. The promoter region was cloned into pGL3-Basic plasmid from Promega. HeLa or cos-1 cells were maintained by standard

protocol. The cell numbers were counted and found to be 1.8×10^5 cells/well and seeded into 24 well culture plate. Cells were transfected with 50 ng of pGL3-Nkx2-5 Luciferase reporter construct and 100 to 200 ng of activator plasmid. After 48 hours, the cells were harvested using 300 μl of CCLR (Cell Culture Lysis Reagent). 100 μl of substrate (E1500, Promega) was dispensed to 50 μl of lysate and luciferase activities were measured with a luminometer (BD Monolight 3010).

2.13 Immunofluorescence (IF) and Immunohistochemistry (IHC) Staining

Mouse embryos for indicated stage(s) were dissected and embedded in paraffin. Samples were sectioned into 5 μm sections using a microtome and subjected to immunofluorescence (IF) or immunohistochemistry (IHC) staining. Samples were rehydrated and sub-boiled for 10 minutes for antigens retrieval. The samples subjected to IF staining were incubated with primary antibodies at 4°C overnight then incubated with corresponding fluorescence conjugated, secondary antibodies at room temperature for an hour in the dark. The antibodies used and the dilution factors were listed in Table 3. The sections were mounted with mounting medium containing DAPI stain (H-1500, Vector Laboratory).

The samples subjected to IHC staining were incubated with biotinylated secondary antibodies for an hour at room temperature followed by horseradish peroxidase (HRP) conjugated streptavidin treatment for 15 minutes. DAB kit (SK4100, Vector Laboratory) was used to assess HRP activities. Sections were then dehydrated and then mounted onto

the glass slides. All images were taken using a Nikon Eclipse Ti Microscope and NIS Elements imaging software.

3. RESULTS*

3.1 Characterization of Mesp1 Expression Pattern

To study Mesp1 function, the expression pattern needs to be known. Previous studies demonstrated *Mesp1* is expressed only during E6.5-E8.5 in mouse embryos and only in the testis in adult mice. Here, the immunofluorescence staining was used to show the protein expression profile in early mouse embryo (Figure 1A-D). Mesp1 was expressed in the extra-embryonic and posterior regions of the mouse embryo at E6.5. At E7.5 and highly expressed in the posterior part, primitive streak and the mesodermal /endodermal cross region. At E8.5, the expression intensity decreased and was detected only in the cytoplasm of the epicardium. When the mouse embryo reached to E9.5, most tissues and regions were not expressing Mesp1. This result showed the Mesp1 expression pattern and location during early mouse embryo development.

The transient expression patterns were also observed in the mouse embryonic stem cell (ES cell or mESC) culture system. The mESC were differentiated by the hang drop method, an aggregating cell culture method to mimic embryo development. The qPCR data showed the expression profile from Day1 to Day10 (Figure 1E). *Mesp1* only

* Part of the reported data of this section is reprinted with permission from Islas, J.F., et al., *Transcription factors ETS2 and MESP1 transdifferentiate human dermal fibroblasts into cardiac progenitors*. Proc Natl Acad Sci U S A, 2012. **109**(32): p. 13016-21, Copyright [2012] by Proc Natl Acad Sci U S A.

expressed between Day3 to Day6 and peaks at Day4. Western blot also correlated with the qPCR data showing that the protein levels were about one day delay comparing to mRNA level (Figure 1F).

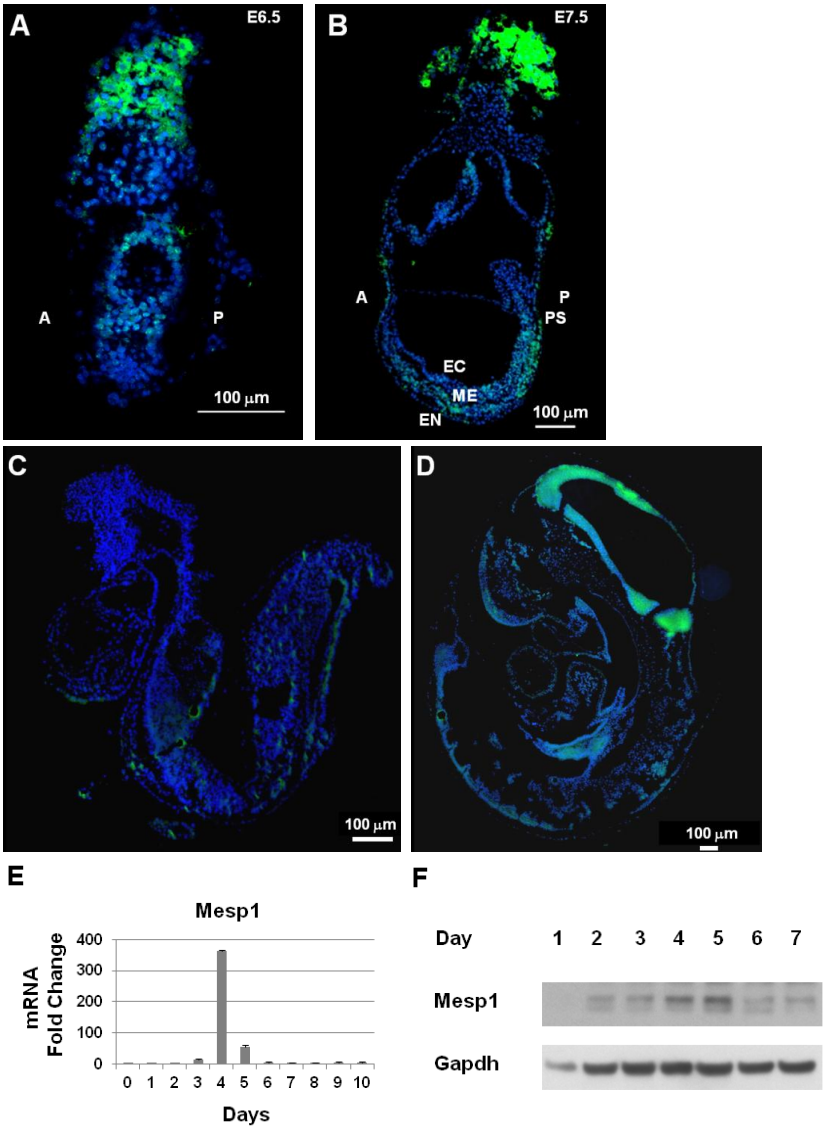


Figure 1 Expression Pattern of Mesp1 in Early Mouse Embryo and EB.
(A)-(D) Immunofluorescence staining of Mesp1 at E6.5-E9.5 of mouse embryo by anti-MESP1 antibody (ab77013, Abcam). (E) Fold Change of mRNA at different days of AB2.2 embryoid body. (F) Western blot of different days of embryoid body by anti-Mesp1 (AP16455c, Abgent) and anti-Gapdh antibodies.

3.2 Generation of Mesp1 Lineage Tracing ES Cell Lines

To further understand what lineage population Mesp1 is contributing, a lineage tracing system was designed. Mice carrying Mesp1$^{Cre/+}$ locus were crossed with Rosa26$^{EYFP/+}$ mice and the resulting ES cells were isolated from inner cell mass of mice embryo to obtain an immortalized ES cell line named UH3. To be brief, when cells express Mesp1, it also expresses homologous recombinase Cre. The Cre recombinase then recognizes the *flox* sequence and excises the stop codon between two flox sites to allow eYFP permanent expression. UH3 cells were grown and maintained under standard ES culture conditions and were subject to differentiation by the aforementioned hang drop method. Once Mesp1 was expressed during differentiation, the eYFP allowed us to visually trace the population by fluorescence techniques (Figure 2A). Here flow cytometry was used to monitor the Mesp1-eYFP population. Similar to previous studies, UH3 cells were differentiated and EBs harvested on different days were analyzed by flow cytometry to characterize trace the lineage via a time series. The eYFP fluorescence expression started at Day4 and reached to 3-5% of the total cell population after Day5 (Figure 2B) while the percentage ration seemed not change through culture. These results showed the UH3 cells marked the Mesp1 expressing population during differentiation and the eYFP allowed us to monitor and isolate the Mesp1 expressing population for further experiments.

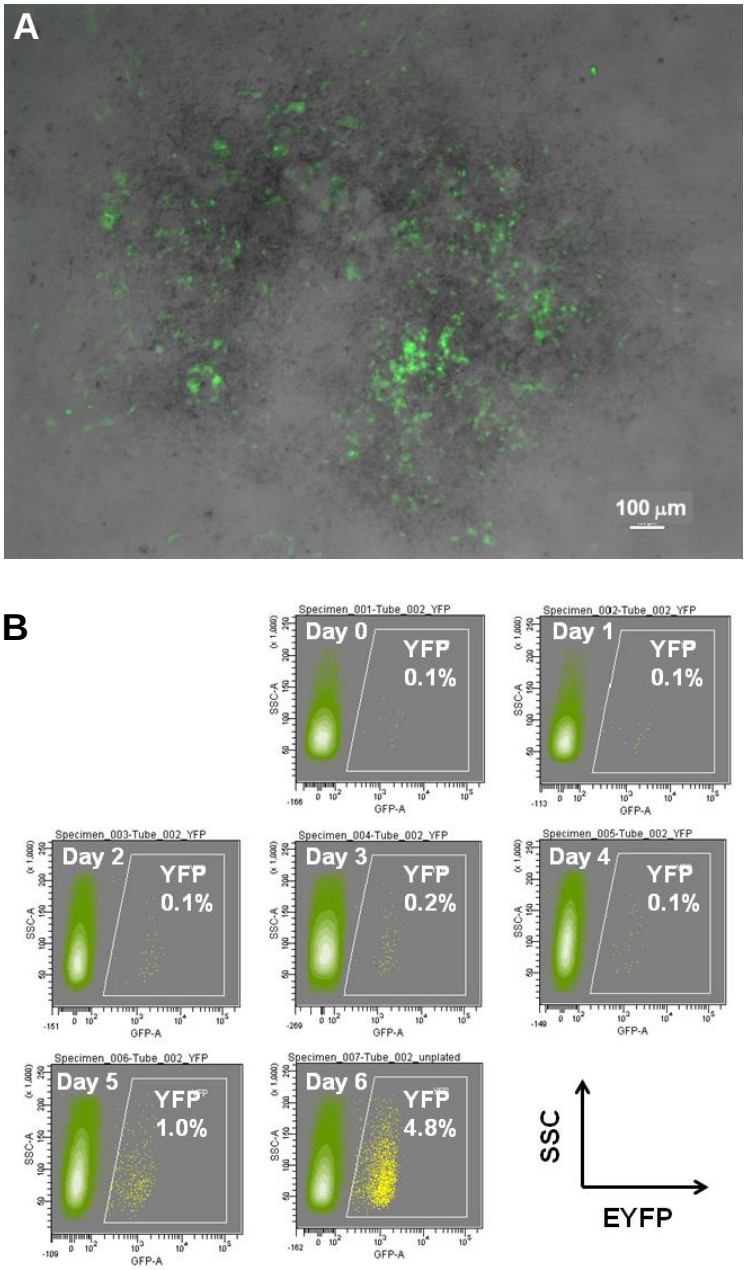


Figure 2 UH3 Cells Expressed eYFP After Differentiation.
(A) Fluorescence microscopy of embryoid body at Day7 of UH3 cells showing eYFP. Scale bar is 100 μm. (B) Flow cytometry analysis of different days of UH3 EB showing the percentages of eYFP population.

3.3 Identify Direct Targets of Mesp1 by ChIP-Sequencing

To study the direct downstream targets of Mesp1, Chromatin Immunoprecipitation of Mesp1 bound targets followed by next generation sequencing (ChIP-seq) was performed. One of the major challenges to study Mesp1 targets was the transient expression of Mesp1 and the percentage of Mesp1 expressing cells in the overall population was small. Mesp1 expressed between Days 4 -7 and gave rise to only 5% within the entire EB cell population. To solve this problem, UH3 cells were used to trace the Mesp1 expressing population by monitoring the eYFP fluorescence. UH3 cells were differentiated by hanging drop method and FACS (Fluorescience-Activated Cell Sorting) sorting was performed at Day5. The eYFP positive population was collected for the Mesp1 ChIP experiment (Figure 3).

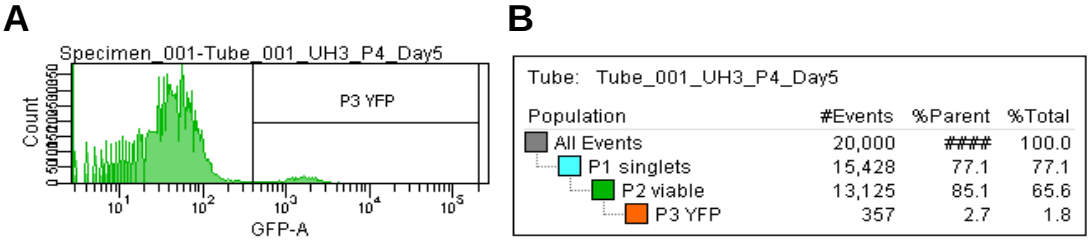


Figure 3 FACS Sorted eYFP Cells.
(A) UH3 cells were differentiated by hanging drop and FACS sorting performed at Day 5. (B) Around 1-3% of total population of positive cells was sorted out.

Another challenge of the ChIP technique in addition to the low cell number was the specificity of the antibody. A good antibody recognizes the target protein specifically

with minimal non-specific binding to eliminate false positive targets during the pull down step of ChIP experiment. To solve this problem, several commercial antibodies were tested by western blot and the optimal one (ab77013, Abcam) was chosen for the ChIP experiment (Figure 4).

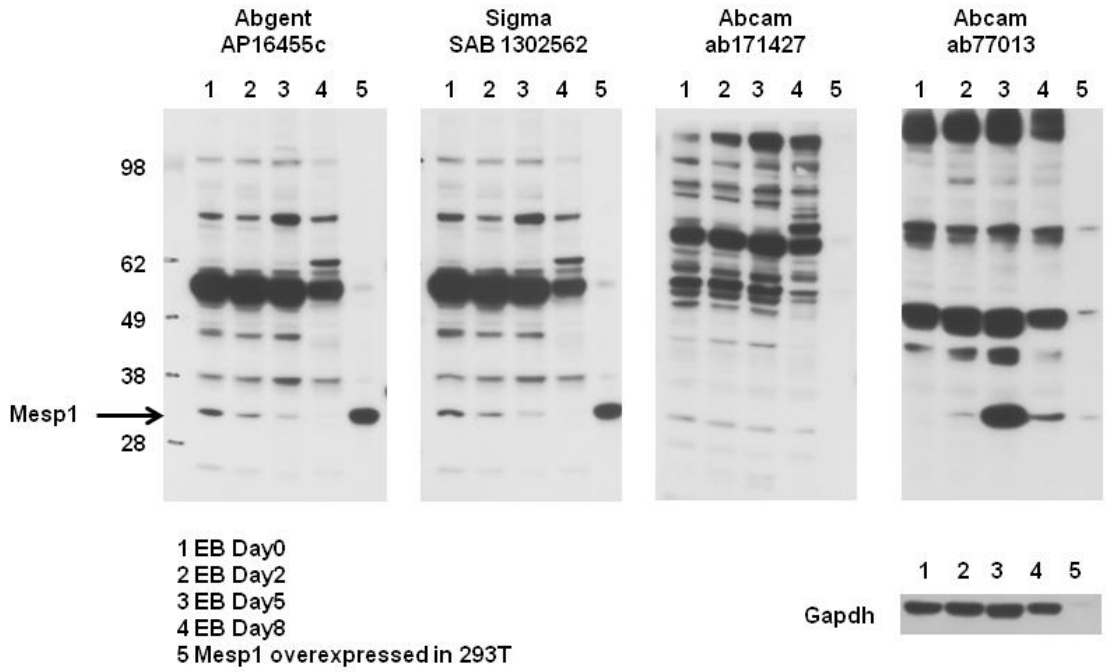


Figure 4 Commercial Mesp1 Antibody Comparison by Western Blot.
UH3 cells were differentiated by hanging drop method and collect at different days. Four Mesp1 commercial antibodies were used for western blot. To mark the accurate size of Mesp1 protein, 293T cells were transiently transfected with constitutive Mesp1 expressing plasmid for 48 hours and cell lysates were subjected to western blot analysis.

The eYFP cells were FACS sorted and lysed for chromatin immunoprecipitation by the optimized Mesp1 antibody. The chromatin DNA was sheared to the optimal size ranging between 200 and 400 fragment sizes and subject to sequencing library preparation and Next Generation Sequencing.

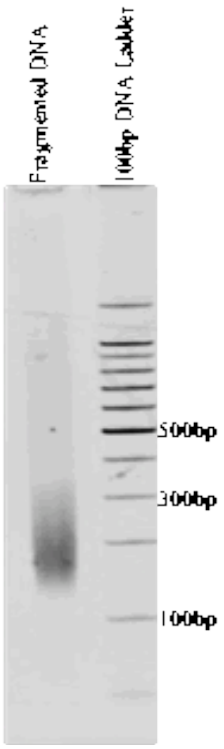


Figure 5 DNA Fragment Size Verification.
UH3 cells were cross-linked with 1% formalin and chromatin was sonicated for 15 cycles with the Bioruptor UCD300 on the High setting. 50µL of sheared input chromatin sample was taken from total input sheared chromatin and treated with RNase A and Proteinase K and reverse cross-linked to release DNA bound fragments from proteins. DNA was cleaned up with a Qiagen DNA Minelute column and separated on a 2% Agarose gel by electrophoresis for 45 minutes at a constant 110V.

After Next Generation Sequencing, data were manipulated by software cisGenome. First, Gene ontology analysis was performed (Table 4). Within the top picked pathways from the GO analysis, we confirmed that Mesp1 targets were enriched in regulating developmental pathways including heart, vasculature, blood vessel, and epithelial cell development. Within the target list, many cardiac genes, such as Mef2c, Isl1, Tbx20 and Acvr2a, were direct targets of Mesp1 (Table 5). This suggests Mesp1 regulates cardiac development through these transcription networks. Not only did we find cardiac related gene targets but several endoderm genes, such as Gata6, Sox17, Foxa2, Pitx2 and Sall4, were also found to be direct Mesp1 targets. This suggests Mesp1 may have a role not only in cardiac mesoderm development but also in definitive endoderm development. Moreover, Mesp1 is also involved in early germ layer differentiation pathways such as meso/mesoendoderm development. This finding suggests Mesp1 may specify bi-potent populations and this population derives into both mesoderm and endoderm populations.

3.4 Analyze Mesp1 Binding Motifs

It is known that transcription factors recognize genomic DNA through conserved DNA binding motifs. Mesp1 is a basic helix-loop-helix (bHLH) transcription factor that recognizes DNA through the E-Box (CANNTG). Mesp1 ChIP-seq results were further analyzed through JASPAR database by comparing the existing motif from different transcription factors. Over a hundred existing motifs including Myc, Myod1, Hand1 and Runx1 were identified from JASPAR database (Table 6). The motif analysis showed the classical E-box (CANNTG) was identified in the list. It supports Mesp1 regulates the

Table 4 Gene Ontology Analysis of Mesp1 Target Genes from ChIP-Sequencing

	Term ID	Term	P Value	FDR
BP	GO:0007507	heart development	1.30E-07	2.47E-04
BP	GO:0001944	vasculature development	2.78E-07	5.27E-04
BP	GO:0051094	positive regulation of developmental process	3.28E-07	6.23E-04
BP	GO:0045941	positive regulation of transcription	6.64E-07	1.26E-03
BP	GO:0001568	blood vessel development	7.82E-07	1.48E-03
BP	GO:0045597	positive regulation of cell differentiation	5.38E-06	1.02E-02
BP	GO:0045165	cell fate commitment	7.17E-06	1.36E-02
BP	GO:0048568	embryonic organ development	8.38E-06	1.59E-02
BP	GO:0060429	epithelium development	6.69E-05	1.27E-01
BP	GO:0045449	regulation of transcription	4.97E-04	9.40E-01
BP	GO:0030324	lung development	2.19E-03	4.07E+00
BP	GO:0016055	Wnt receptor signaling pathway	4.93E-03	8.97E+00
BP	GO:0048762	mesenchymal cell differentiation	8.04E-03	1.42E+01
BP	GO:0048738	cardiac muscle tissue development	1.98E-02	3.16E+01
BP	GO:0003007	heart morphogenesis	2.92E-02	4.30E+01
BP	GO:0055010	ventricular cardiac muscle morphogenesis	3.12E-02	4.53E+01
BP	GO:0007512	adult heart development	3.51E-02	4.92E+01
BP	GO:0033077	T cell differentiation in the thymus	3.83E-02	5.24E+01
BP	GO:0002064	epithelial cell development	4.21E-02	5.59E+01
BP	GO:0048864	stem cell development	4.87E-02	6.13E+01
BP	GO:0001755	neural crest cell migration	4.87E-02	6.13E+01
BP	GO:0048863	stem cell differentiation	5.98E-02	6.90E+01
BP	GO:0060425	lung morphogenesis	7.64E-02	7.79E+01
BP	GO:0051145	smooth muscle cell differentiation	7.73E-02	7.83E+01
BP	GO:0045823	positive regulation of heart contraction	7.73E-02	7.83E+01
BP	GO:0007498	mesoderm development	9.43E-02	8.47E+01
CC	GO:0031012	extracellular matrix	1.28E-06	1.93E-03
CC	GO:0034704	calcium channel complex	7.39E-03	1.06E+01
CC	GO:0005667	transcription factor complex	4.94E-02	5.34E+01
MF	GO:0015267	channel activity	6.58E-10	1.12E-06
MF	GO:0022836	gated channel activity	1.86E-09	3.17E-06
MF	GO:0005509	calcium ion binding	1.65E-08	2.81E-05
MF	GO:0005249	voltage-gated potassium channel activity	1.86E-06	3.16E-03
MF	GO:0005244	voltage-gated ion channel activity	4.50E-06	7.66E-03
MF	GO:0022832	voltage-gated channel activity	4.50E-06	7.66E-03
MF	GO:0003700	transcription factor activity	7.98E-05	1.36E-01
MF	GO:0005262	calcium channel activity	7.07E-04	1.20E+00
MF	GO:0003712	transcription cofactor activity	5.27E-02	6.02E+01

BP: Biological Process; CC: Cellular Component; MF: Molecular Function

Table 5 Target Genes in Mesp1 ChIP-Seq Results

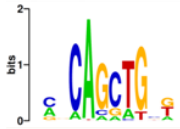
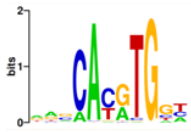
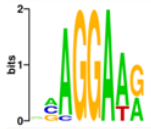
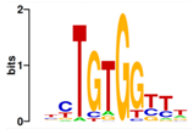
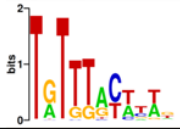
Acvr2a	Acvr2b	Bmp2	Bmp4	Cd34
Cdh11	Celf1	Cited2	Cxcr4	Dkk1
Dvl2	Ehf	Elf2	Elf4	Elk3
Erg	Ets1	Ets2	Etv1	Etv4
Etv5	Etv6	Ezh2	Fgf10	Foxa2
Fst	Gadd45a	Gadd45b	Gadd45g	Gata6
Gja1	Gli2	Gli3	Hand1	Hand2
Hes1	Hey1	Hey2	Hif1a	Id2
Isl1	Mef2c	Nkx6-1	Nppb	Pax3
Pdgfa	Pitx2	Runx1	Runx2	Sall4
Shh	Smad2	Smad3	Smad4	Smarcd3
Sox17	Tbx20	Tgfb2	Tgfbr1	Tgfbr2
Tnnt2	Wnt3a	Zic3		

downstream targets through binding to the conserved regions of these genes. The conserved binding regions also implied the potential factors within the same transcription complexes.

3.5 Validation of Mesp1 Targets by Overexpressing Mesp1

The ChIP-seq analysis revealed direct targets of Mesp1. The occupancy of certain genomic regions by transcription factors usually results in up or down regulation of transcription activities of downstream genes. It may also recruit the transcription complex to regulate the epigenetic pattern of these genes to change the transcription activities. Therefore it is necessary to measure the gene activity of downstream gene targets to understand the molecular mechanism behind how Mesp1 regulates downstream networks.

Table 6 DNA Binding Motif from Mesp1 ChIP-seq Results

MOTIF ID	Motif name	Motif	Number of Peaks with motif	Number of random peaks with motifs	P-Value	Z-values
MA0048.1	NHLH1		3162	2089	8.84E-238	3.29E+01
MA0059.1	MYC::MAX		2091	2051	4.27E-168	2.76E+01
MA0475.1	FLI1		3979	2858	5.42E-250	3.38E+01
MA0002.2	RUNX1		4706	2883	7.00E-280	3.57E+01
MA0047.2	Foxa2		4005	3265	1.78E-204	3.05E+01

To answer this question, an inducible Mesp1 construct was generated and introduced into mES cells. AB2.2 ES cells were transduced a Doxycycline-inducible V5-Mesp1-eGFP construct by lentiviral transduction. ES cells were cultured under standard procedure and Doxycycline was used (1 μg/ml) to induce Mesp1 expression. Flow cytometry were used to monitor the eGFP expression after induction. Western blot were also perform to monitor the Messp1 and V5 protein expression (Figure 6).

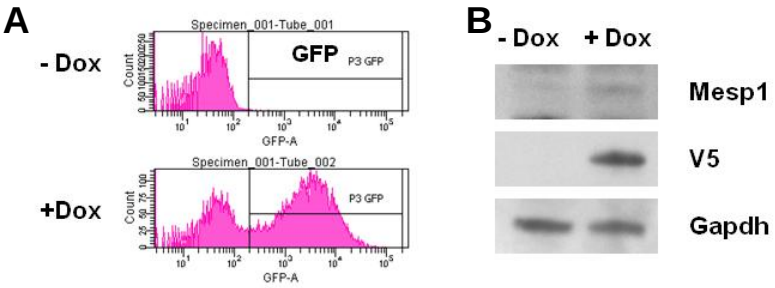


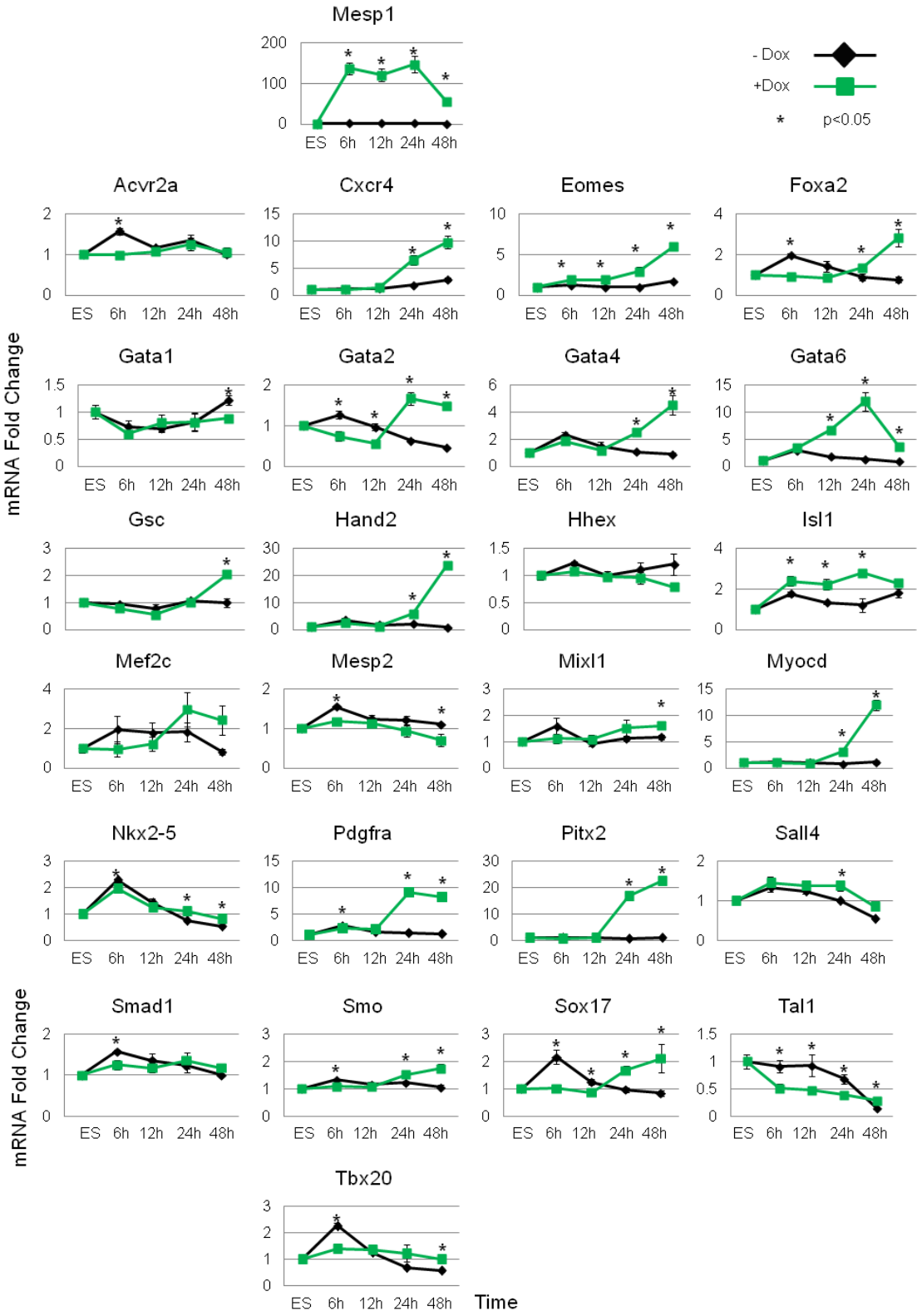
Figure 6 Overexpressing Mesp1 by Doxycycline Induction in ES Cells.
(A) Flow Cytometry showed ES cells expressed eGFP after Doxycycline induction. (B) Western blot showed Mesp1 and V5 were expressed after Doxycycline induction.

After validating Mesp1 overexpression, ES cells were induced by Doxycycline and total RNA was isolated at indicated time points to analyze the gene expression. Quantitative real-time PCR showed the Mesoendoderm marker genes Gata6, Pitx2, Sox17, Foxa2, and Gsc were highly up regulated after 48 hours (Figure 7). Cardiac transcription factors such as Gata4, Hand2, and Myocd were activated after 24 hours of induction. Cardiac surface markers Cxcr4 and Pdgfra were also induced after 24 hours. These results confirmed that Mesp1 up-regulated the transcription activities of downstream target genes through genomic gene transcription activation.

3.6 Gene Expression Characterization of Mesp1 Derived Cells by RNA-Seq

Another way to confirm the ChIP-seq results was by monitoring the gene expression profiles in differentiating EB at the time that Mesp1 is expressed. UH3 mouse ES cells were used as a model to address this question. UH3 ES cells were differentiated by thehang drop method and the eYFP positive cells were subject to FACS sorting from Day5 to Day8. RNA was isolated and next generation sequencing was performed. This

Figure 7 Genes Regulated After Mesp1 Overexpression.
Mesp1 were induced by Doxycycline (1 μg/ml) treatment and sample were collected at indicated times. Quantitative realtime PCR were performed. Expression intensity was normalized by internal control Gapdh. Data represent relative expression comparing to undifferentiated ES cells as mean ± standard error of mean (SEM) from three biological triplicates.


Mesp1
- Dox
+Dox
* p<0.05
Acvr2a
Cxcr4
Eomes
Foxa2
Gata1
Gata2
Gata4
Gata6
Gsc
Hand2
Hhex
Isl1
Mef2c
Mesp2
Mixl1
Myocd
Nkx2-5
Pdgfra
Pitx2
Sall4
Smad1
Smo
Sox17
Tal1
Tbx20
mRNA Fold Change
mRNA Fold Change
ES 6h 12h 24h 48h
Time

allowed us to trace Mesp1 population and monitor genes regulated by Mesp1. The RNA-seq results showed genes relative to germ layer specification such as Mesp1, Eomes, T, and Mixl1 were down regulated through differential days (Figure 8A). This implied that these genes would be negatively regulated once Mesp1 specified the progenitors. On the other hand, the cardiac genes such as Nkx2-5, Mef2c, Tbx5, and Gata4 were up-regulated (Figure 8B).

Interestingly, genes that were considered as endoderm markers, such as Sox17 and Foxa2, were not down regulated right after Mesp1 expression (Figure 8C). These results suggest that Mesp1 is not directly down regulated in endoderm development; instead, it might play roles involved in positively regulating other germ layer development.

3.7 Mesp1 Contributes to Multiple Lineages

The roles of Mesp1 during developmental processes were identified. However, did all Mesp1 committed cells differentiate to the same lineage? To answer this question, UH3 cells were used as a lineage tracing model. Once UH3 cells expressed Mesp1, it also expressed eYFP permanently and eYFP protein can be recognized by anti-GFP antibody by using immunofluorescence staining. UH3 cells were differentiated by the hang drop method and EBs from samples were “dropped” or plated on Day5 and maintained through Day8. The samples were incubated with anti-GFP antibody and other antibodies that recognize different tissue specific markers. The results showed the markers from different germ layers were co-localized with GFP positive cells. CD31 is a surface

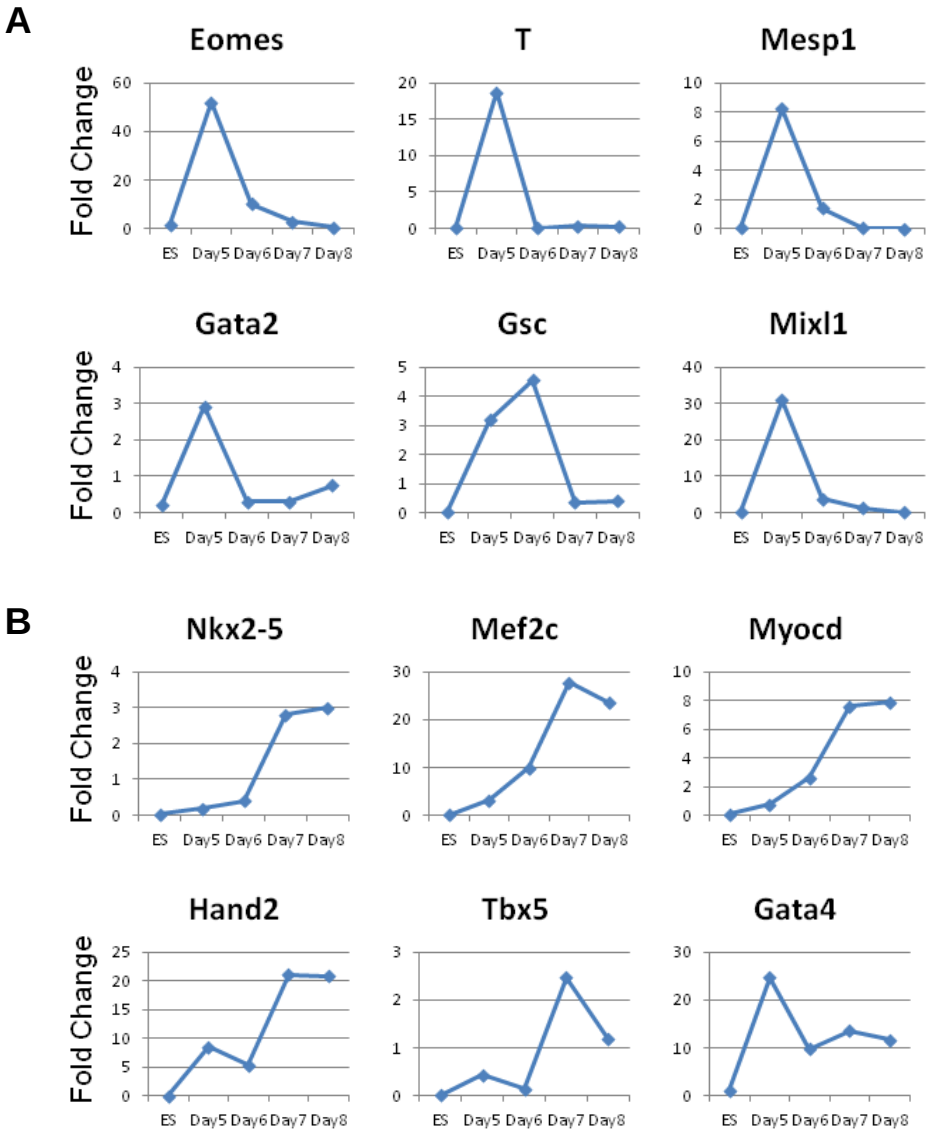


Figure 8 Gene Expressions Profiling by RNA-Seq of Mesp1-eYFP Cells at Different Stages. (A) Germ layer specific genes were down regulated in later stage of EB. (B) Cardiac genes were up regulated in later stage of EB

C

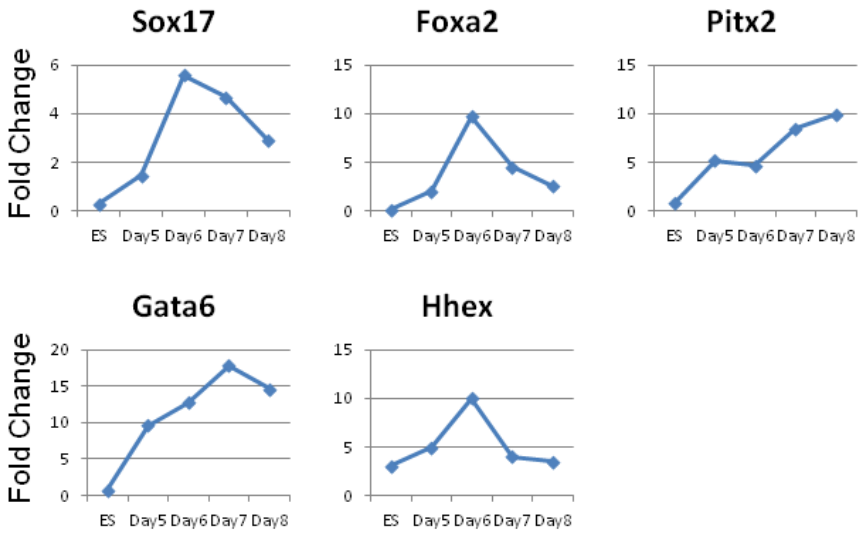


Figure 8 (Continued).
(C) Endoderm markers at different stage of EB showed these marker were not direct down regulated at Day5 and Day6.

marker of endothelial cells and a majority of hematopoietic stem cells. Endoderm marker Foxa2 was also co-localized with GFP (Figure 9). This data suggested that Mesp1 committed cells contributed to not only mesoderm but also endoderm tissue. Similar results can be confirmed *in vivo* (Figure 10). The immunofluorescence staining of sections in early mouse embryo showed GFP cells were present in the cardiac crescent. These results suggested that Mesp1 daughter cells contribute to mesoderm tissues such as heart, hematopoietic cells, endothelial cells and endoderm tissue(s) such as pancreas.

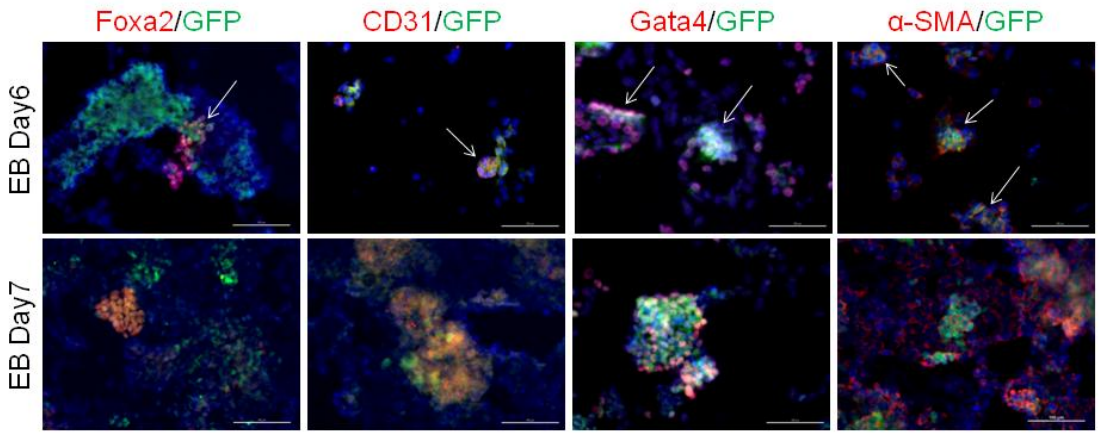


Figure 9 Immunofluorescence Staining of Embryoid Body by Different Lineage Markers. UH3 cells were differentiated and samples were fixed at Day6 and Day7. The scale bar is 100 μm.

3.8 Identify the Association Partners of Mesp1

Transcription factors regulate their downstream genes by binding to the regulatory regions such as promoter and/or enhancer regions on the genomic DNA. Theycan also form transcription complexes with histone modifiers or other transcription factors. The regulation of different downstream genes can be achieved by interacting with different interaction partners. Mesp1 is a bHLH transcription factor that may form a homodimer with itself or a heterodimer with other transcription factors. The biological functions of Mesp1 have been well characterized while the molecular details such as structural domain(s) and interaction partners are not well understood. Studying the interaction partners of Mesp1 will give us more information as to how Mesp1 regulates downstream pathways.

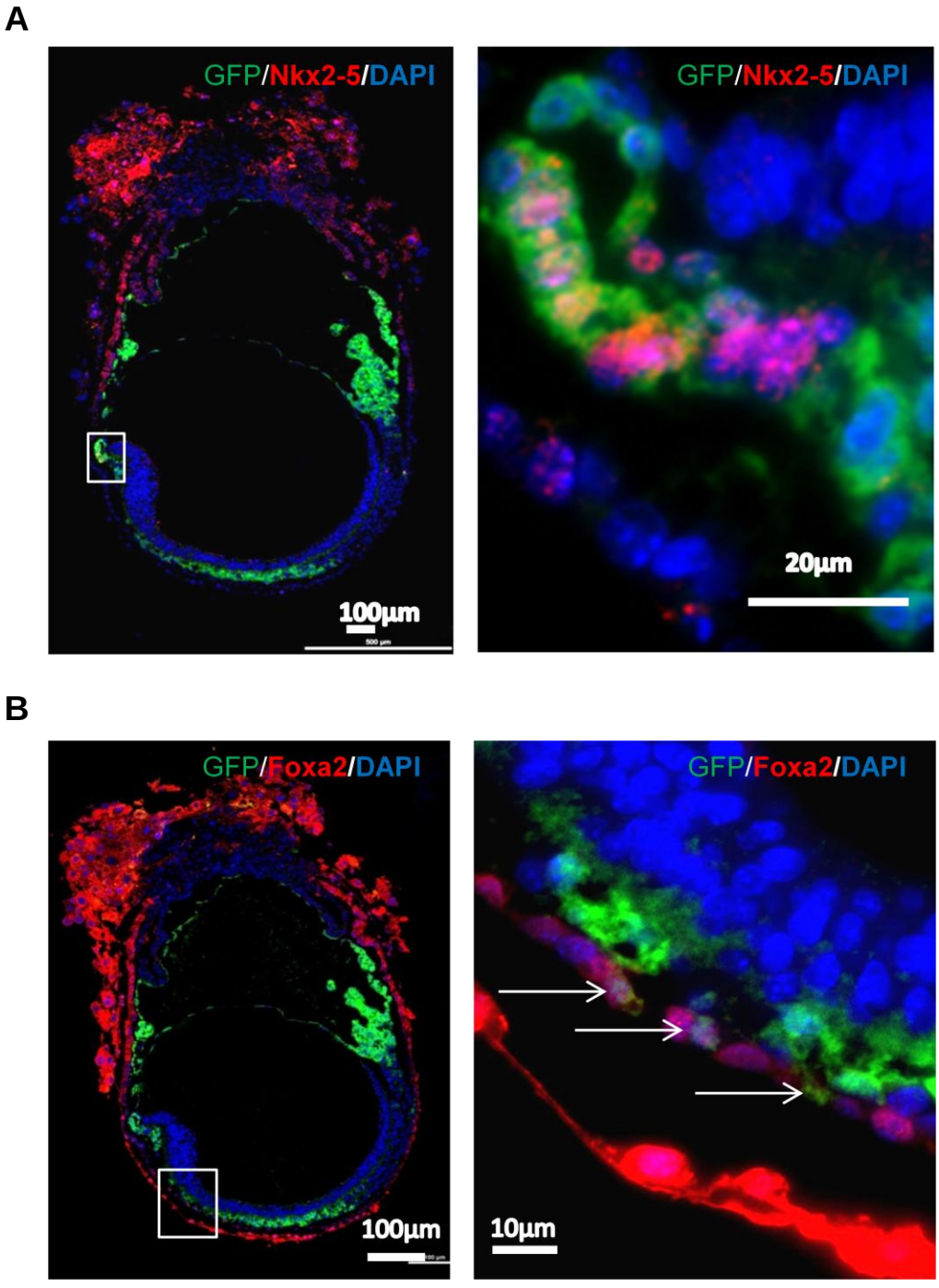
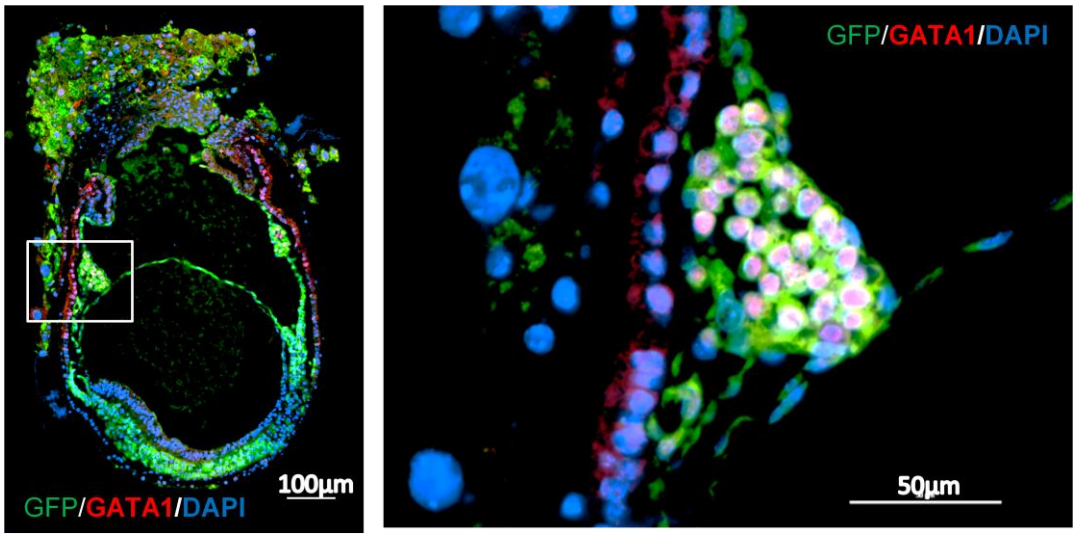


Figure 10 Immunofluorescence Staining of Different Lineage Markers in Mouse Embryo. (A) The co-staining of GFP and Nkx2-5 at E7.5 in sagittal orientation (B) The co-stianing of GFP and Foxa2 at E7.5 in sagittal orientation (Luo and Weng, manuscript in preparation).

C



D

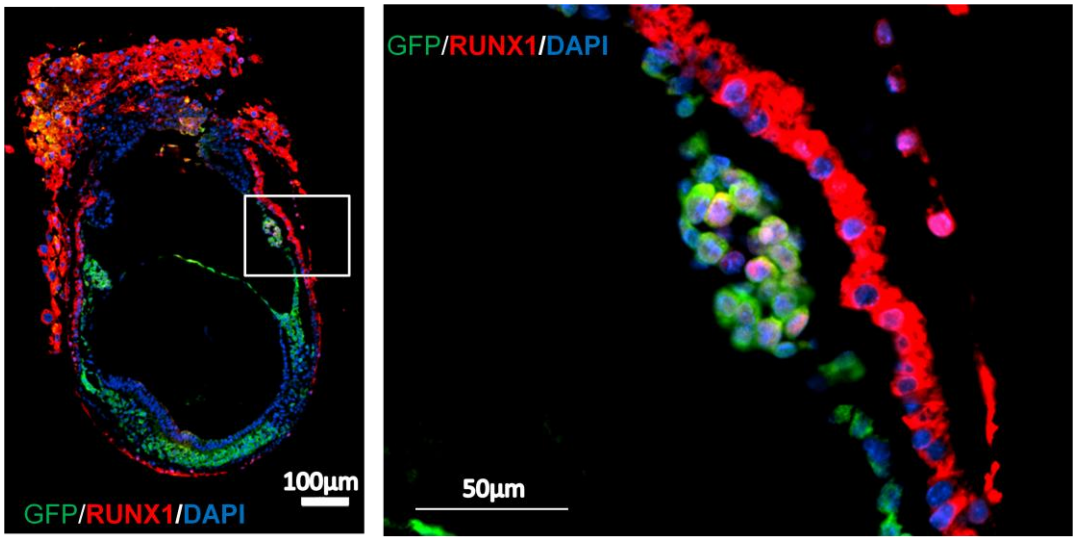


Figure 10 (continued).
(C) The co-staining of GFP and Nkx2-5 at E7.5 in sagittal orientation (B) The co-stianing of GFP and Foxa2 at E7.5 in sagittal orientation (Luo and Weng, manuscript in preparation).

In previous motif related studies several binding motifs of well known transcription factors such as bHLH transcription factors, Fox transcription factors, Gata transcription factors and ETS transcription factors have been revealed. This implies that Mesp1 may form transcription complexes with these transcription factors and the specific interactions may impact the regulation of downstream genes. Within these motif lists, Runx1 and ETS transcription factors are important markers for hematopoietic differentiation; Nkx2-5 is the most critical marker for cardiac mesoderm specification; Fox families are essential for endothelial cells differentiation.

The protein-protein interactions can be confirmed by co-immunoprecipitation technique (Co-IP). Here, Mesp1 was overexpressed with other potential interaction partners. The results also showed Mesp1 interacting with several ETS factors including ETV1, ETV4, ETV7, ELF4, ELF5, and SPDEF (Figure 11). Although ETS transcription factors are not from the bHLH family, they were able to physically interact with Mesp1. This suggests Mesp1 may regulate some downstream target genes through interacting with ETS factors in different genomic locations or with respect to certain developmental time points. The detailed mechanism(s) needs further studies to support this theory. Most of the ETS factors play roles in hematopoietic differentiation such as the proliferation and maturation of B cells, T cells, NK cells, and lymphocytes. It implies the possible roles of Mesp1 during hematopoietic development. We also found most bHLH transcription factors that play roles in second heart field differentiation such as Hand1 and Hand2 physical interacted with Mesp1. This provides supporting evidence that Mesp1 may

regulate cardiac differentiation through interacting with Hand2 and guide the maturation of cardiomyocytes.

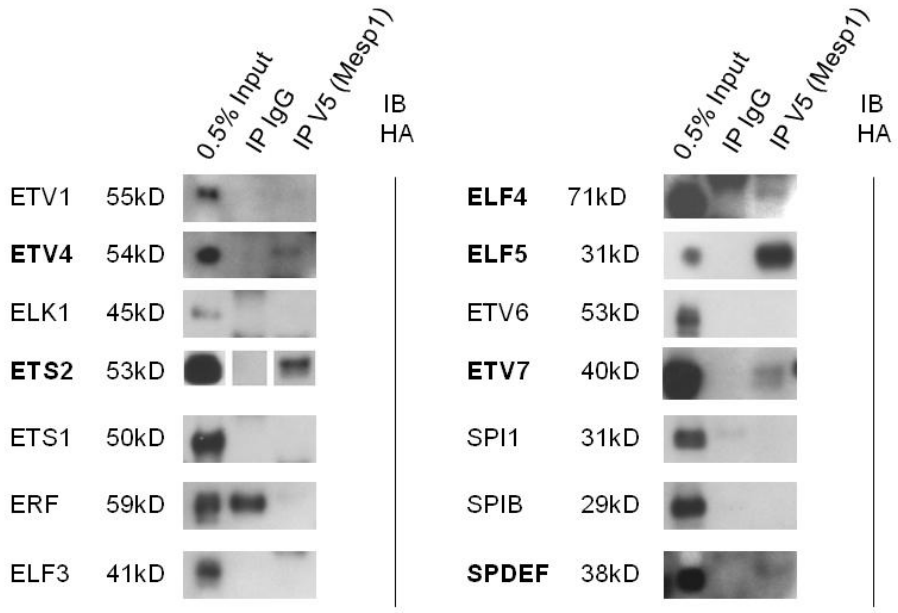


Figure 11 Physical Interactions of Mesp1 and ETS Transcription Factors.
Co-immunoprecipitation of Mesp1 and ETS transcription factors showed ETV4, ETS2, ELF4, ELF5, ETV7 and SPDEF physical interact with Mesp1. Both Mesp1 and ETS factors were cloned into constitutive expression vector pcDNA3. ETS factors were designed to carry HA tag and Mesp1 were V5-tagged.Transfections of both Mesp1 and ETS factors in cos-1 cells by FuGene HD and sample were collected 48 hours after transfection. Cells were lysed by RIPA buffer and V5-conjugated agarose beads were used to precipitate interaction complex. Anti-mouse IgG was used as control during precipitation process Western blot was performed after co-immunoprecipitation and anti-HA antibody was used.

3.9 Characterization of the Gene Expression Profiles of Different ETS Factors

The potential roles of ETS factors in cardiac development were identified previously. However the expressions of ETS factors were found to be ubiquitous. To narrow down

the factors that play roles specifically during cardiac development, screening through different stages of developmental processes were performed. Similar to Mesp1 expression patterns in EB, qRT-PCR on different days of EB differentiation were performed to understand the expression profiles of ETS factors (Figure 12). The results showed the factors ETV4, ETV5, and SPIB were enriched in undifferentiated ES cells or in the early stage of EB differentiation while other factors such as ETS1, ETS2 and FLI1 were up regulated in EBs during later stages. Still, others such as ERG and SPDEF were found to express only in a short time frame.

The qRT-PCR data showed the expression profiles of ETS factors in EB in a temporal manner. The mesoderm formation occurs in very early embryonic stages. Therefore characterizing ETS factors in early developmental stages becomes important. Using ETS antibodies unique to individual ETS family members, we were able to perform immunofluorescence staining. At E6.5, ETV4 was present in the blood islands, mesoderm and endoderm (Figure 13). At E7.5, the ETV4 expression was present in the posterior region and migrates through the mesoderm layer toward the anterior region. This result suggested that although ETV4 expression intensity decreases through development, it was still detectable in early embryonic stages and may play roles in early mesodermal development.

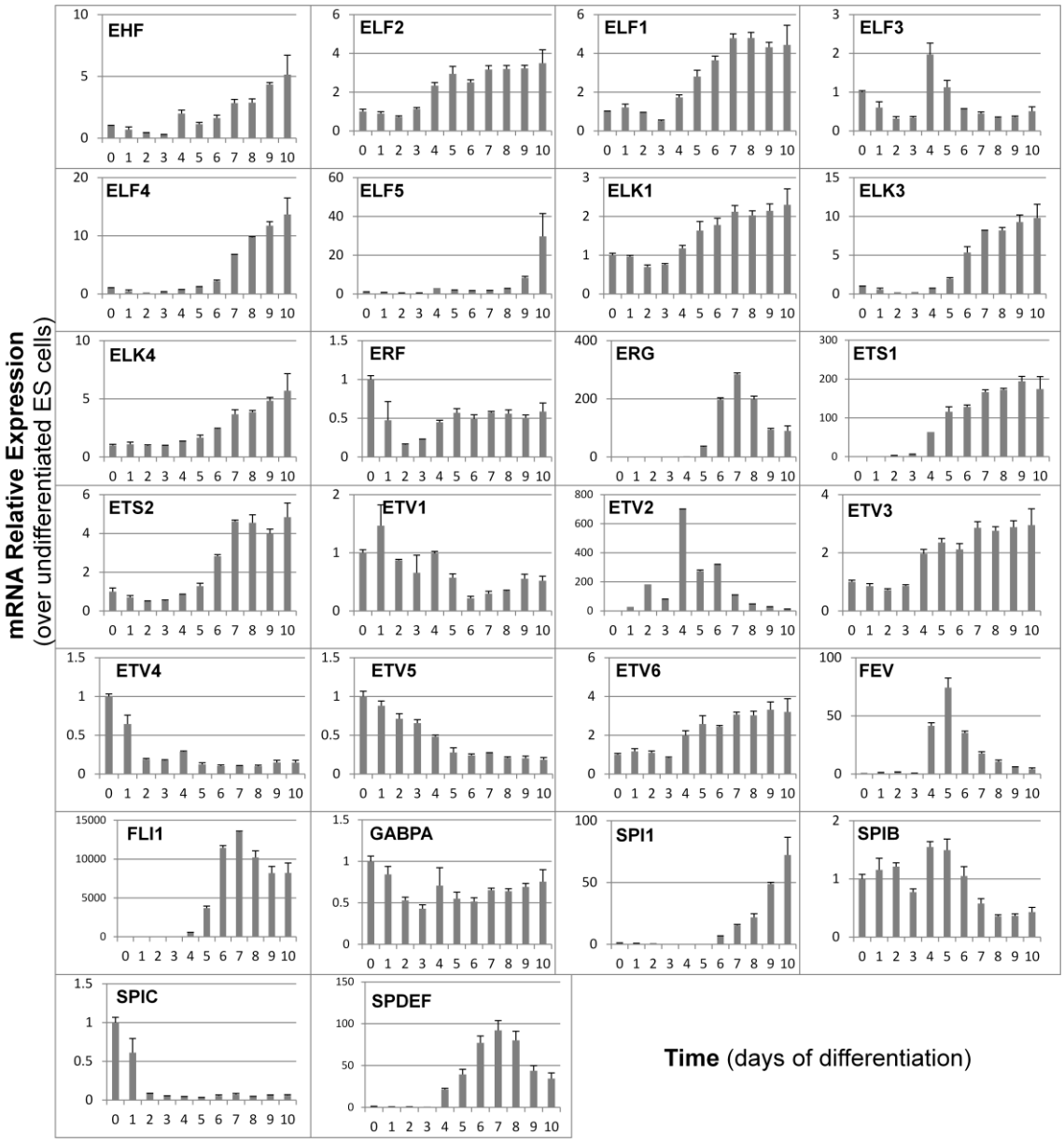


Figure 12 Gene Expression Profiles of ETS Transcription Factors In Embryoid Bodies.
Mouse ES cells were differentiated by the hang drop method and samples were collect at indicated times. Quantitative Real-time PCR was performed to measure mRNA expressed normalized to the internal control, Gapdh, and compared to undifferentiated ES cells. Data were presented as mean ± standard error of mean from three biological replicates.

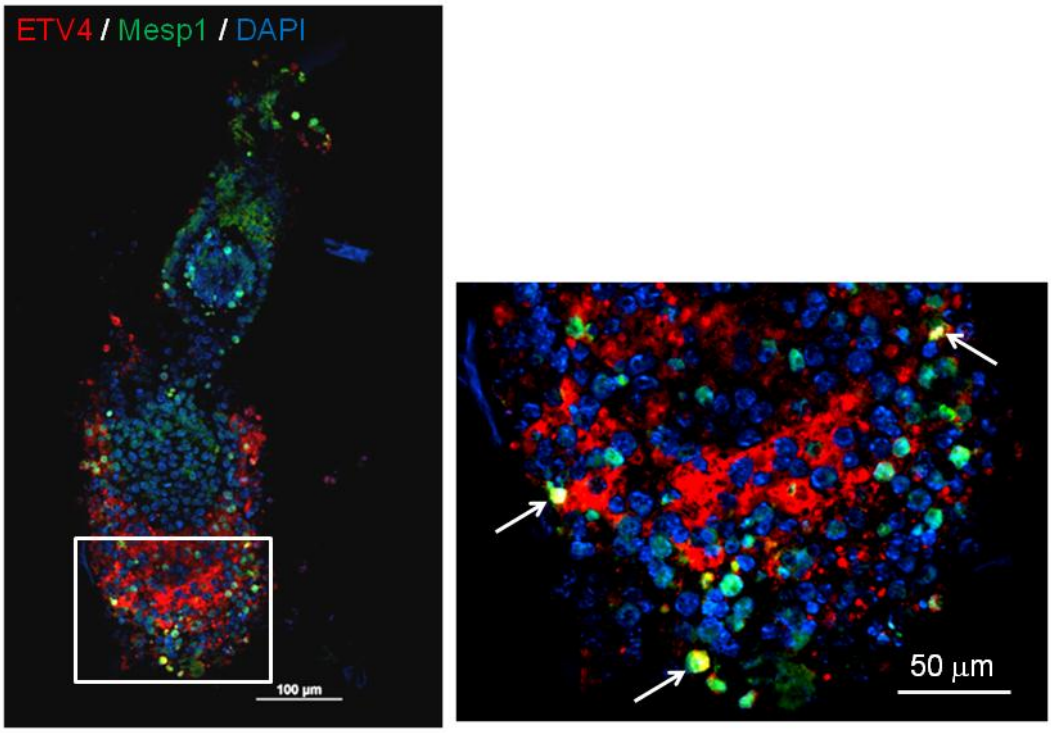


Figure 13 Immunofluorescence Staining of ETV4 in Early Mouse Embryo.
Mouse Embryo At E6.5 was sectioned in sagittal orientation and stained ETV4 (red) and Mesp1 (green).

3.10 Combinatorial Regulation of ETS Transcription Factors and Mesp1

The ETS factors' expression profiles helped us to understand which ETS family members may play roles in regulating the development of cardiac progenitors along with Mesp1. Next, molecular and functional studies were needed. Luciferase assays were carried out to answer these questions. Our previous study showed Nkx2-5 playing a critical role in cardiac crescent specification. Moreover, the lacZ staining showed the enhancer region, which is -3000 base pairs upstream of the transcription start site, is the most critical region during heart development. It is highly acetylated during heart formation which means Nkx2-5 is controlled through this enhancer region. This Nkx2-5

enhancer region was cloned into a luciferase reporter construct to study what is regulating its transcriptional activity. The results showed both ETS2 and Mesp1 increased the transcription activity of Nkx2-5 (Figure 14). The rest of the ETS factors showed only marginal activation via this enhancer region. Interestingly when combining Mesp1 expression with other ETS factors, some of them showed additive effects such as ETV4, ETV7 and SPDEF. It implied ETS transcription factors might regulate cardiac gene Nkx2-5 synergistically along with Mesp1. This result helped us to focus on those factors that were able to activate Nkx2-5 through the specified enhancer region, which may further activate the cardiogenesis pathway.

3.11 Gene Regulation by Overexpressing ETS Factors

To better study the biological function(s) of ETS factors, gain-of-function studies were conducted. In this study, Normal Human Dermal Fibroblast (NHDF) cells were chosen as our model system, and it was also chosen for performing reprogramming experiments. First of all, we cloned the ETS factors into a lenti-viral plasmid system and overexpressed these genes by lentiviral transduction. The cells were cultured for seven days to allow expression of ETS factors and responding genes. The RNA was isolated from these cells and cDNA microarrays were performed. The microarrays showed groups with differential gene expression regulated by ETS factors (Figure 15). The genes were clustered by their definition of germ layer differentiation. From the clustered heap map, ETV4 and ETS2 shared the most similarity in mesoderm and endoderm development.

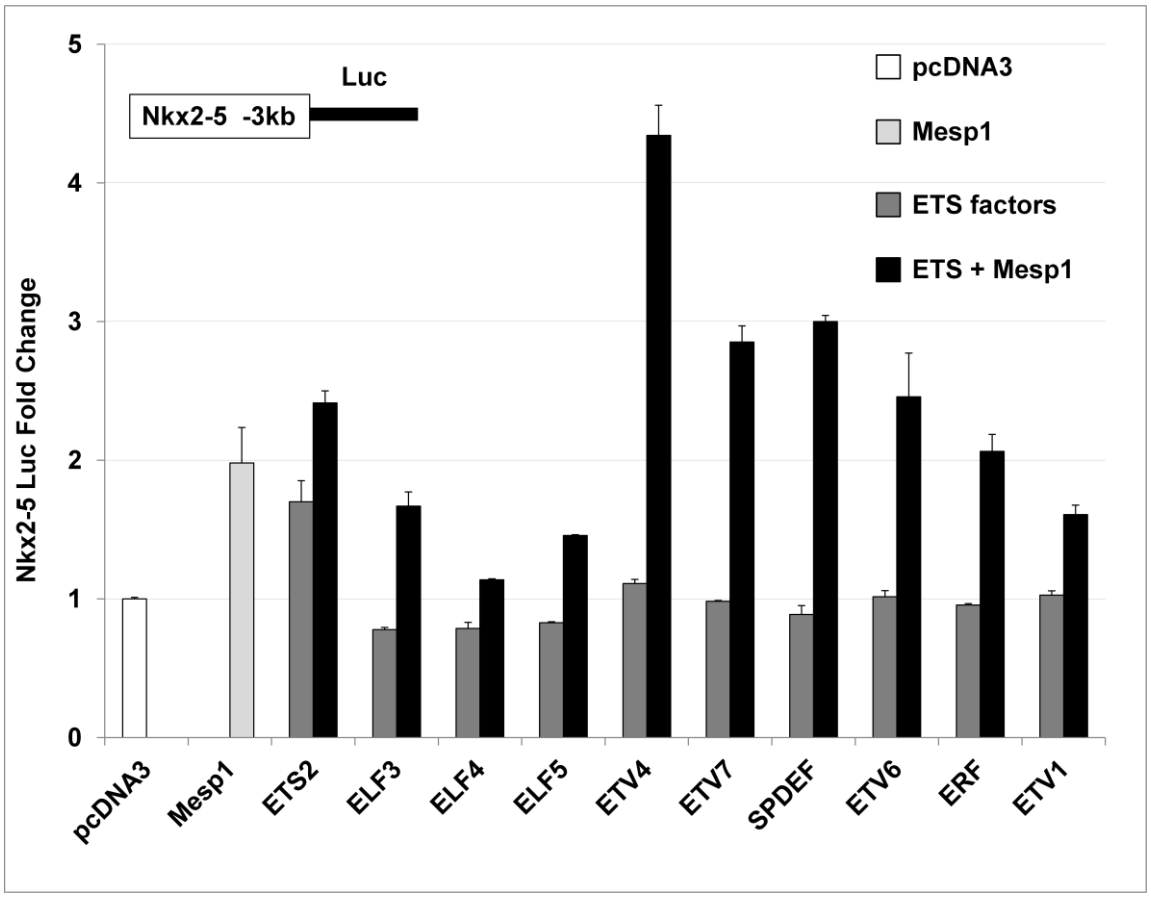


Figure 14 ETS Factors and Mesp1 Activate Nkx2-5 Luciferase Activity.
Equal amount of both Mesp1 and ETS factors were transfected into HeLa cells by FuGENE HD. Samples were collect and analyzed 48 hours after transfection. Data were presented as mean ± standard error of mean from three biological replicates.

A

B

ETS Factors	Differential Genes
ETS2	3705
SPIB	2135
SPDEF	1377
ERF	440
ELF5	1630
ETV4	4668
PU.1	2646
ELK1	389
ETV5	918
GABPA	578

C

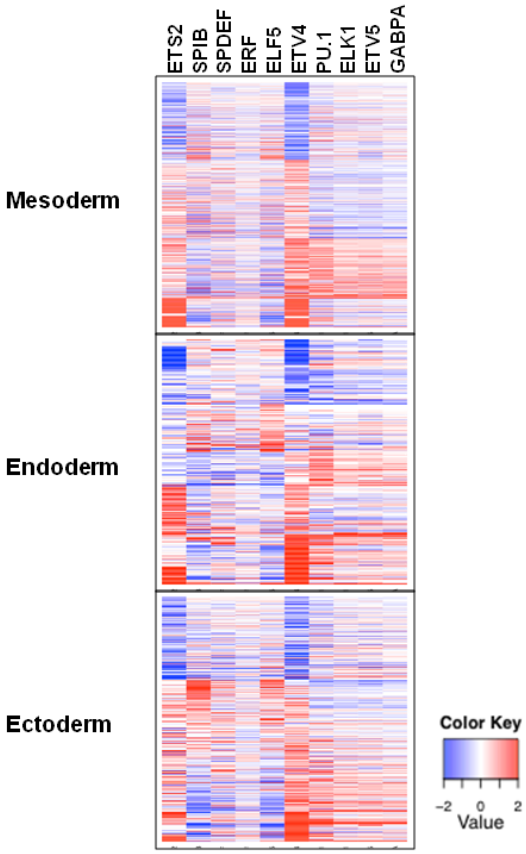


Figure 15 Genes Regulation After Overexpressing ETS Factors.
(A) Protocol used to overexpress ETS factors in NHDF. (B) Differential genes regulated by different ETS factors. (C) Cluster heat map showed the genes regulated by ETS factors.

3.12 Characterization of Cell Surface Marker Expression after ETS Factor Overexpression

Another way to characterize the cell population was by the cell surface markers. Cell surface markers were Cluster of Differentiation (CD) expressed on cytoplasm membranes that are either receptors for corresponding signal cascade or adhesion molecule to determine cell adhesion or migration. Previous results showed ETS factors activated essential genes for cardiogenesis. Here the surface markers CD31 (PECAM),

CD34, CD117 (c-kit), CD140a (PDGFRa), CD144 (VE-Cadherin), CD184 (CXCR4) and CD309 (Flk-1, KDR) were chosen to analyze the reprogramming effects (Figure 16).

The results showed Mesp1 activates surface markers that include CD34, CD140a, CD184 and CD309. ETV4 was shown to activate CD31, CD34, CD144, CD184 and CD309. When we combined ETV4 with Mesp1, we found ETV4 and Mesp1 additively activated CD31 from 5% to 7% a 2% increase in this surface marker within the total cell population. ETV4 and Mesp1 activated CD184 3.4% and 6.9% respectively but increased to 7.5% when combined. This result suggests the ETS factors and Mesp1 were capable of activating critical cardiac markers CD31 and CD184. It also supports the concept of the combinatorial regulation by ETS transcription factors and Mesp1.

3.13 Reprogramming of NHDF by Introducing Mesp1 and ETS Factors

Our studies revealed the expression pattern, physical interactions, and biological functions of Mesp1 and ETS factors. The next step was introducing Mesp1 and the ETS factors into somatic cells to see whether they beneficially regulate the genes necessary for cardiogenesis, most importantly, to see if differentiated cells can be converted into another cell fate. In order to reprogram the somatic cells, the commercially available Normal Human Dermal Fibroblast from Lonza was chosen. This eliminated the variation between different samples during the isolation process and avoided cell contamination issues.

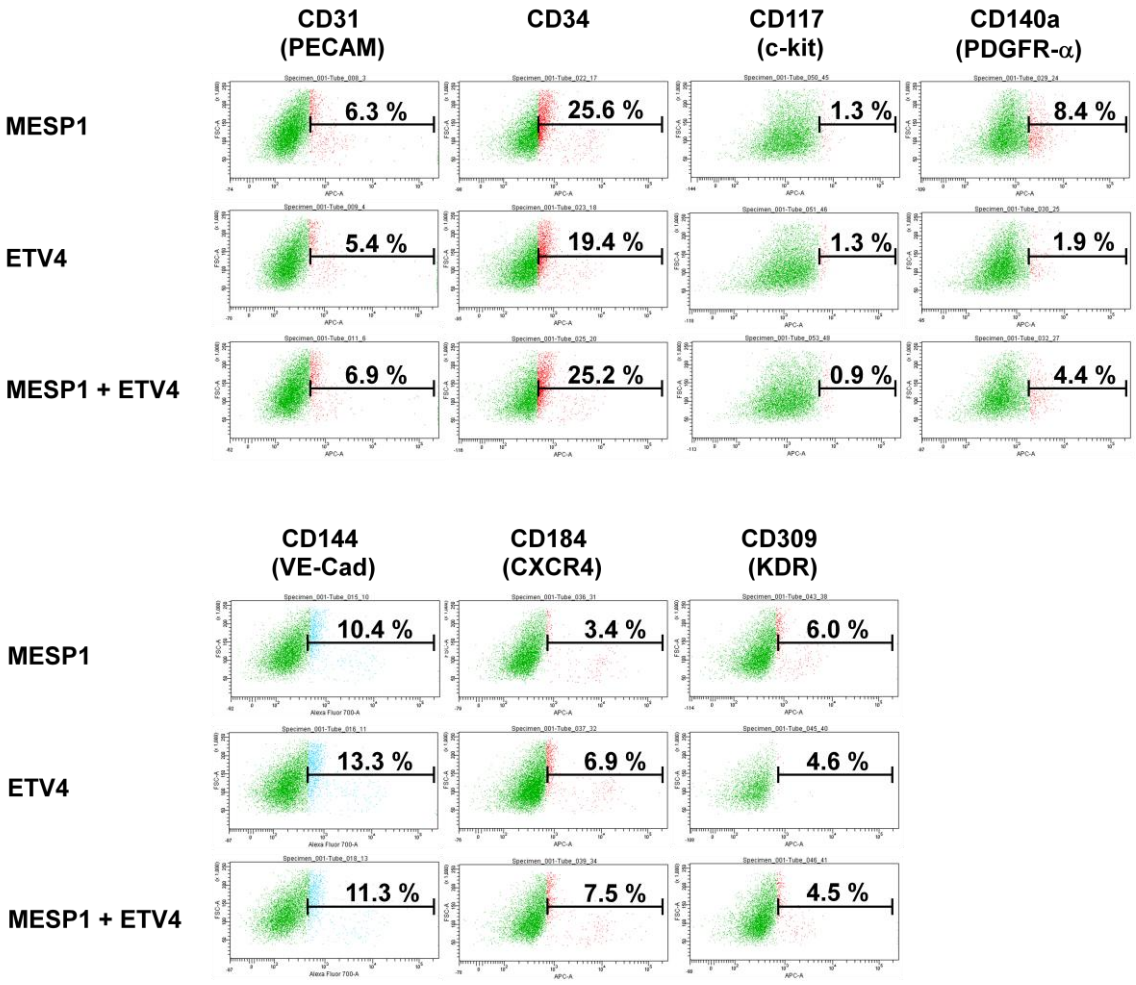


Figure 16 Flow Cytometry Analysis Of Surface Marker in Reprogrammed NHDF.
Specified transcription factors were transduced into NHDF by lenti-viral infection. Samples were harvested at Day 7 post transduction and stained with indicated surface markers for FACS sorting.

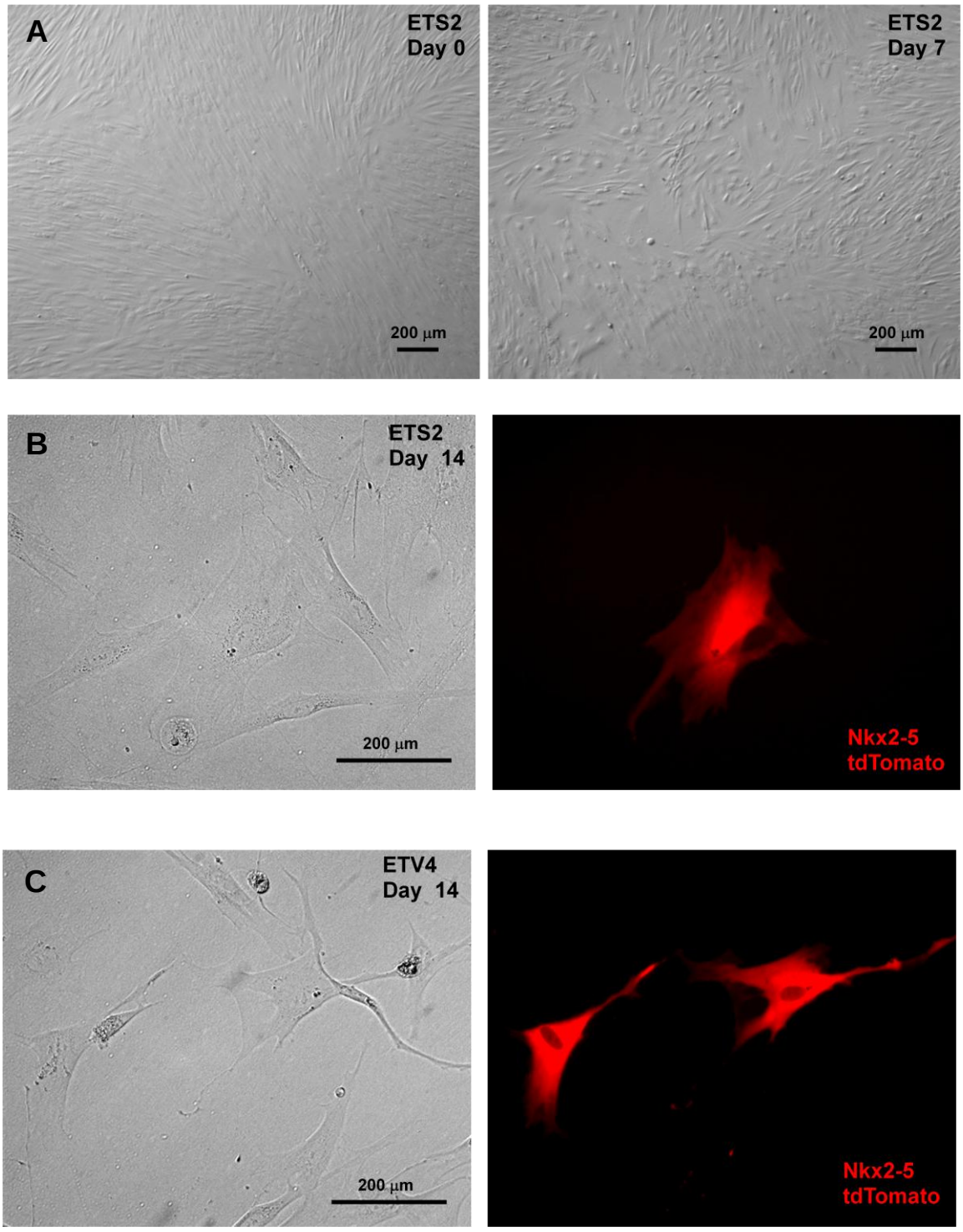
To ensure proper gene delivery, Mesp1 and ETS factors were cloned into constitutive or Doxycycline-inducible lenti-viral plasmid constructs. Once the cells were transduced,

the cells can be isolated to ensure protein expression. This enabled the cells to be cultured for the long-term process required for cell conversion.

To trace the conversion efficiency, an Nkx2-5 tdTomato reporter was used to monitor the reprogramming progress. Our previous research suggested the expression of Nkx2-5 during heart formation in early mouse embryo was regulated by the acetylation of the enhancer region between -3000 and -2500 base pairs upstream of the transcription start site. This region was used to develop a fluorescence reporter to monitor Nkx2-5 expression lifetime and therefore an internal marker that allowed us the physically follow the conversion progression.

The NHDF cells were maintained in normal growth conditions. The NHDF cells were infected with the lenti-viral particles carrying Mesp1 and ETS factors (ETV4 and ETS2 in this part) once cell confluency reached 80%. After infection, the media was changed to cardiac media alpha-MEM and cells were maintained until ready for analysis. One week post-infection, the morphologies of the NHDF cells changed from long hair like fibroblasts to epithelial shaped cells or rounded cells (Figure 17). The cells also expressed the tdTomato fluorescence reporter indicating the cells had activated and were expressing Nkx2-5. However the cells did not form mature and beating cardiomyocytes. This result suggested that using lenti-viral particles to convert NHDF cells was possible, but the protocol needed further refinement and optimization in order to achieve beating cardiomyocytes.

Figure 17 Reprogramming of NDHF by Mesp1 and ETS Factors.
(A) Contrast images showed NHDF in Day0 and Day7. (B) Contrast and fluorescence images showed NHDF transduced with ETS2 in Day 14. (C) Contrast and fluorescence images showed NHDF transduced with ETV4 and Mesp1 in Day14


A
ETS2
Day 0
200 μm
ETS2
Day 7
200 μm
B
ETS2
Day 14
200 μm
Nkx2-5
tdTomato
C
ETV4
Day 14
200 μm
Nkx2-5
tdTomato

3.14 Reprogramming NHDF Into Cardiac Progenitors by ETS2 and Mesp1

The reprogramming of adult cells toward other cell lineages requires not only endogenous gene expression but also environment cues such as cytokine treatments and culture conditions. After screening all ETS factors, ETS2 was chosen to reprogram the NHDF. Similar to previous part, the NHDF was transduced with lentiviral particles containing Mesp1 and ETS2. The results showed ETS2 converted the cells into rounded cells within one week (Figure 18). The cell surface markers also expressed CD31 and CD309. Quantitative real-time PCR also showed the cardiac genes NKX2-5, MEF2C, FLK1, PECAM and TDGF were highly expressed in the reprogrammed cells. These results showed ETS2 and Mesp1 successfully reprogrammed NHDF cells into another cell type specifically into cardiomyocytes.

3.15 Addition of Activin A and BMP2 Increases the Reprogramming Efficiency

The conversion of NHDF cells into replicating stages was achieved by transducing Mesp1 and ETS2. However, the cells seemed not to be able to form fully mature cardiomyocytes. It was reported the maturation of cardiomyocytes required the inhibition of the Wnt signaling pathway such as adding Activin A and BMP2. The continued expression of differentiation genes, ETS2 and Mesp1, was also not beneficial to the maturation of cells. Therefore a Doxycycline inducible system was designed to control the activation of these genes to only the early stages. Similar to the previous protocol, the lenti-viruses carrying ETS2 and Mesp1 were infected into NHDF cells but under an inducible promoter. These cells were treated with Doxycycline for four days.

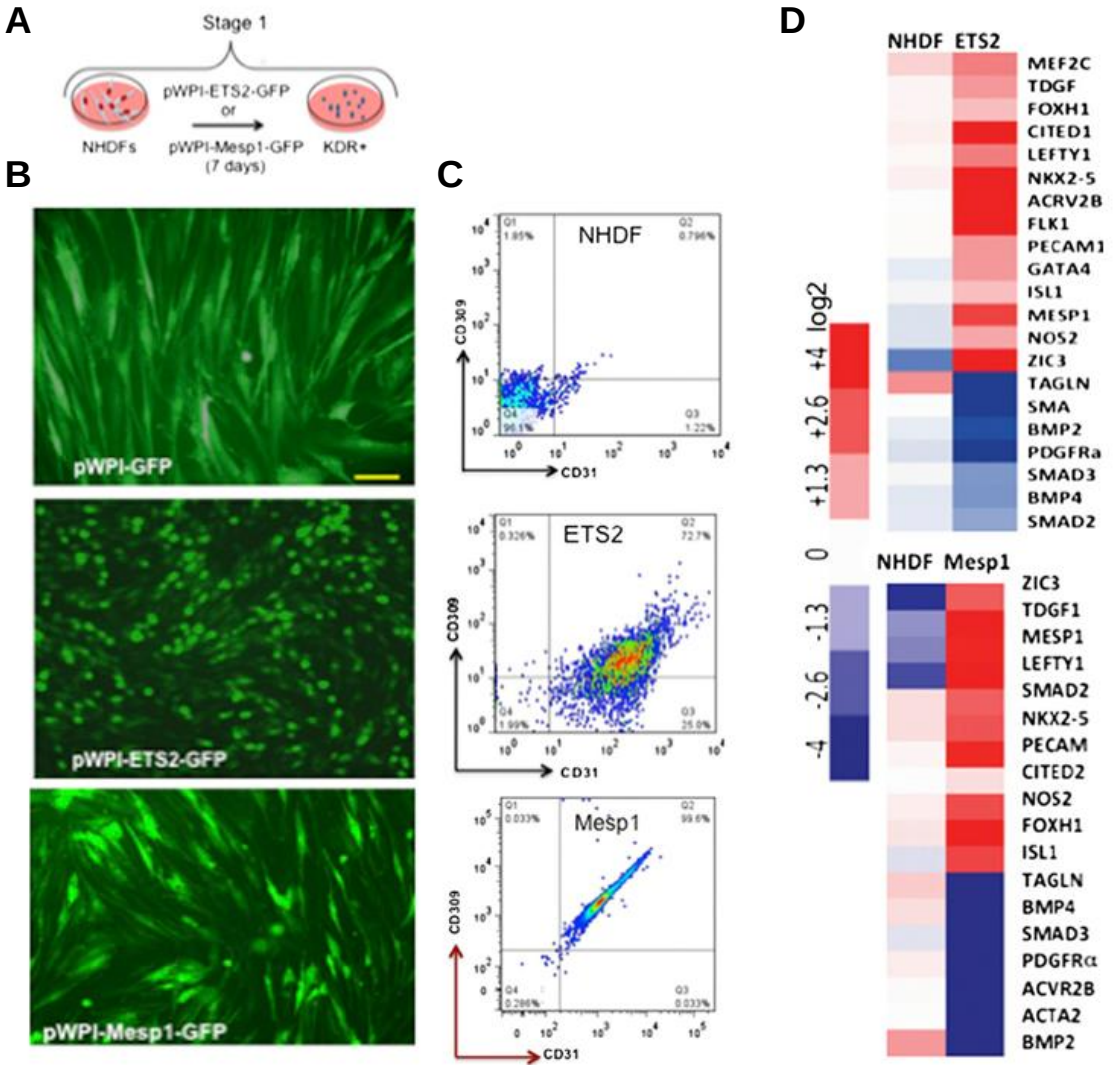


Figure 18 ETS2 and Mesp1 Induce NHDF into KDR Positive Cells.
(A) Protocol used to reprogram NHDF. (B) Morphology change of NHDF after reprogramming. (C) Flow cytometry analysis showed reprogrammed cells expressed CD31 and CD309. (D) Heat map of qPCR results showed gene expression profile after reprogramming. Data modified from Islas, Liu, Weng, et al. PNAS, 2012; **109**(32): 13016-21.

Following this, Activin A and BMP2 were then added for another 2 days. Cells were collected at Day 8 and subject to multiple analyses. The results showed two days of

Activin A and BMP2 treatment reached the optimal tdTomato fluorescence expression of Nkx2-5 reporter (Figure 19). Quantitative real-time PCR also showed the cardiac genes NKX2-5, GATA4, HAND2, and TNNT2 were the most highly expressed in combination of ETS2, Mesp1, Activin A, and BMP2.

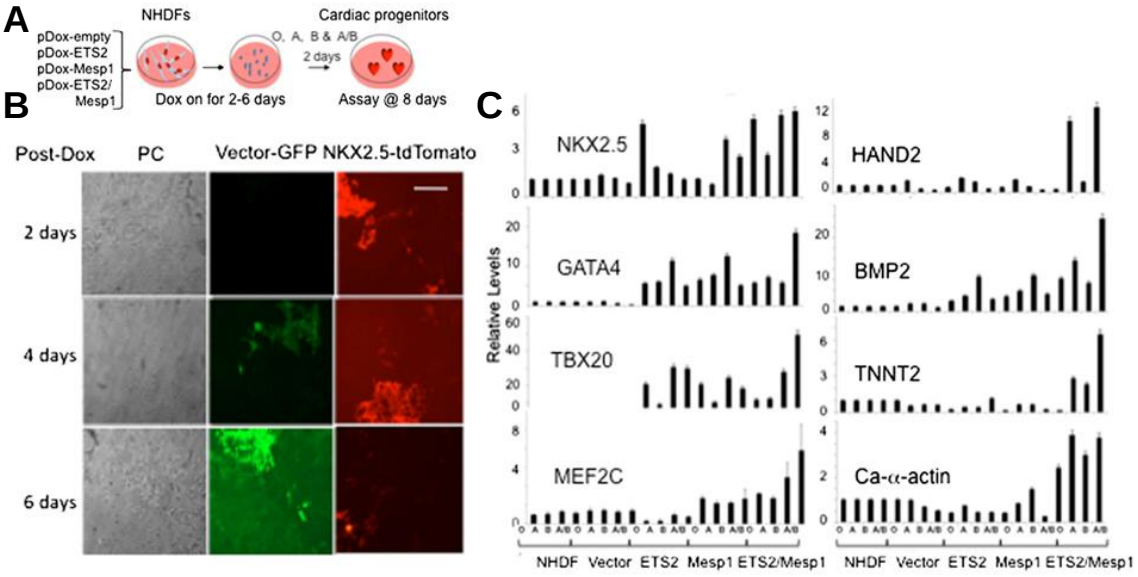


Figure 19 Activin A and BMP4 Treatment Induced NHDF into Cardiac Progenitors.
(A) Schematic showing the culture conditions during the reprogramming process. (B) Fluorescence of reporters indicative of the reprogramming progress. (C) The qPCR of core cardiac genes expression in combination of ETS2, MESP1, Activin A and BMP2. Data modified from Islas, Liu, Weng, et al. PNAS, 2012; **109**(32): 13016-21.

3.16 TAT-Protein Treatment Reprogrammed NHDF Cells into Cardiac Progenitors

The lentiviral treatment was suitable for basic research purposes because of their high transduction efficiency. However, the leakiness of gene expression may result in

unwanted and long term expression of proliferating genes and was shown to not benefit to differentiation process. It may also result in oncogenic like consequences if the genes were constitutively expressed. To solve this problem, we designed a TAT-protein delivery system. Both ETS2 and Mesp1 were cloned into TAT-protein plasmids, and the proteins were purified from bacteria. The TAT-proteins were delivered to NHDF sequentially and similar cell converting was observed. The results showed the surface markers CD31 and CD309 were expressed (Figure 20). The qRT-PCR results showed essential cardiac genes were activated. Western blot analysis also confirmed the protein expression of cardiac structural proteins such as MHC and TNNI3. Long term culture experiments showed the cells formed mature sarcomeric structures. These results showed that transient treatments of ETS2 and Mesp1 along with the addition of Activin A and BMP2 were able to convert the NHDF cells and could increase the maturation process. This novel treatment and long term culturing methods the NHDF cells eventually were able to form mature sarcomeric cardiomyocytes.

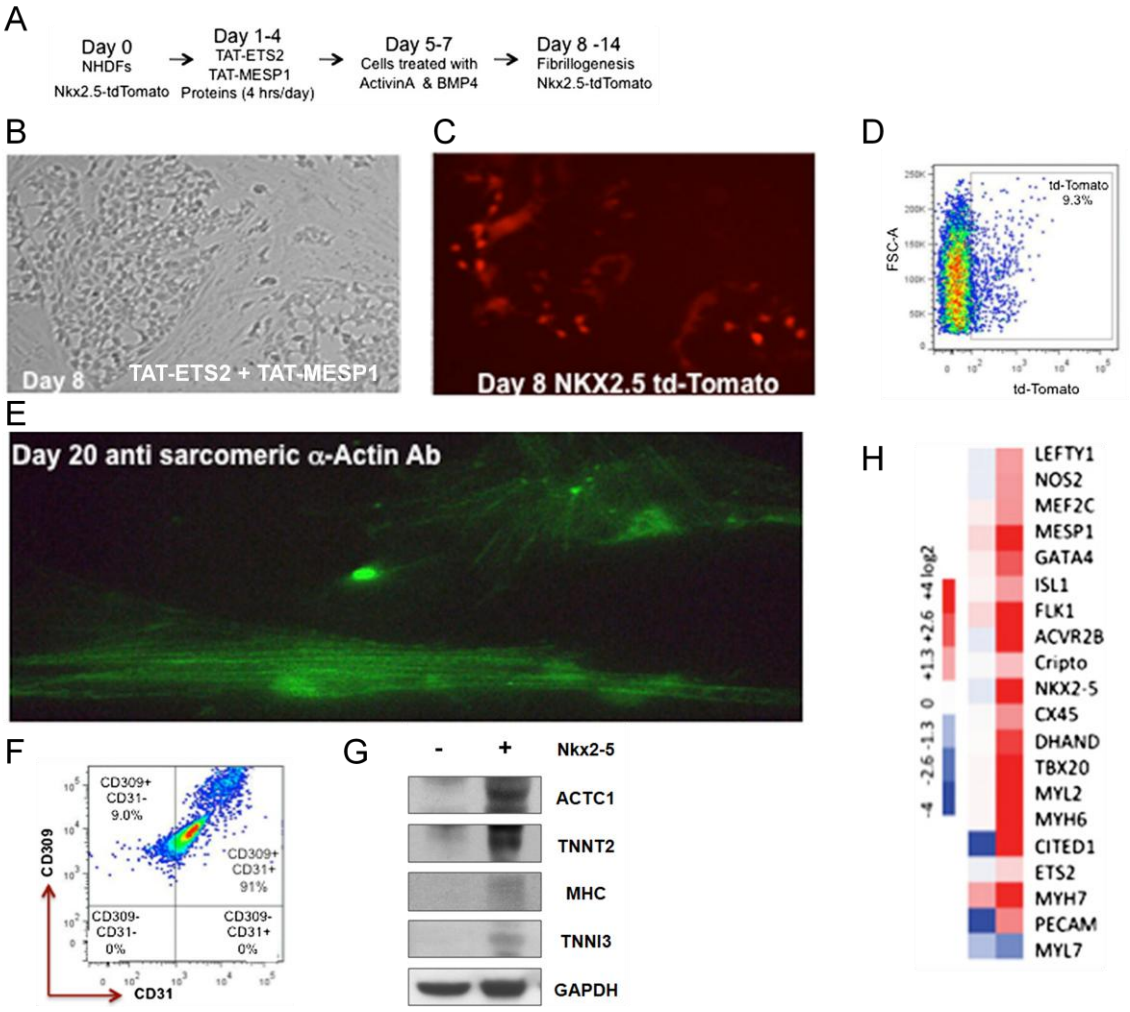


Figure 20 TAT-Protein Induced NHDF into Cardiac Progenitors.
(A) Schematic protocol for reprogramming by TAT-protein (B) Phase contrast image showed the converting of NHDF. (C). The converted cells expressed NKX2-5 tdTomato fluorescence. (D) Flow cytometry analysis showed the NKX2-5 tdTomato expression. (E) The immunofluorescence staining of sarcomeric a-actin (F) The flow cytometry analysis showed the cells expressed CD31 and CD309 (G) Western blot showed the core cardiac genes expression in NKX2-5 tdTomato enriched population. (H) Heat map of qPCR results showed the core cardiac gene expression in converted cells. Data modified from Islas, Liu, Weng, et al. PNAS, 2012; **109**(32): 13016-21.

4 DISCUSSION AND CONCLUSION

4.1 Discussion

4.1.1 The Expression Pattern of Mesp1 Can Be Further Studied

Previous studies demonstrated E8.5 was the last stage of mouse embryo development that the mRNA of *Mesp1* can be detected by *in situ* hybridization. Our immunofluorescence staining showed Mesp1 protein was elevated at E6.5 and E7.5 and barely detectable at E8.5 and E9.5. Interestingly, the staining at E7.5 showed the localization of Mesp1 in some cells was located in cytoplasm, not in the nucleus. Mesp1 is a transcription factor and is considered to be functional in nucleus. Whether Mesp1 localized in the cytoplasm has any functional implications is interesting and should be characterized in future studies.

The Mesp1 lineage tracing system allowed us to answer the question of what are the Mesp1 derived tissues. It has been considered that Mesp1 leads to driving regions of the heart. However, our immunofluorescence staining showed tissues derived from endoderm such as pancreas also stained positive for Mesp1 in addition to those resulting from mesoderm. This finding suggests the roles of Mesp1 during development are not just involved in mesoderm development but also endoderm specification. Mesp1 may specify a bi-potent lineage and this lineage may contribute to both heart formation as well as endoderm tissues.

Another possibility is Mesp1 is reactivated in a later stage of embryonic development. Pancreas development occurs after E10.5 and Mesp1 could be expressed in those stages. Our results showed that Mesp1 is expressed between E6.5-E8.5 but further staining of later embryonic stages will be needed to answer this question.

4.1.2 The Mesp1 Binding Site Validation along with Further Investigation of Non-Coding Gene Targets

The Mesp1 ChIP-seq results were carried out using an endogenous Mesp1 specific antibody on Mesp1 eYFP population from UH3 cells. One of the challenges was Mesp-eYFP are only three to five percent of the total population. The low cell count made it difficult to isolate enough genomic DNA for generating a sequencing library. Therefore, the enriched signals from the sequence data were harder to determine compared to a more robust sample like a histone marker. This increased the chance of false positive targets. To solve this issue, regular ChIP-PCR should be performed to validate the targets we identified. We also used different antibodies specific to Mesp1 from different companies targeting unique epitopes of Mesp1 and compared the results. Those targets appearing as significant between both data sets were scored with a higher confidence since they overlapped.

One of the advantages of the next generation sequencing is that the results include all enriched binding regions including those associated with non-coding genes. It has been the general opinion that transcription factors regulate downstream targets through

binding on the promoter or enhancer region and recruiting the transcription complex to alter or control the transcription activity. The next generation sequencing allows us to identify the regulation region in unbiased way. Our sequencing results suggested most of the binding regions of Mesp1 localized in non-coding area. This suggests Mesp1 could regulate non-coding genes such as miRNAs or lncRNAs (long non-coding RNA). One of the examples in our previous study showed SRF drove miR-210 expression that regulated cardiac development through the Shh signaling pathway. It would be interesting to identify the regulatory miRNA families and potential lncRNAs involved in cardiac differentiation that are also downstream targets of Mesp1.

4.1.3 Characterization of Mesp1 Binding Partners

Our studies showed Mesp1 is necessary for to not only mesoderm but also endoderm lineage specification. How Mesp1 specifies multiple lineages becomes one of the most interesting topics. During embryonic development, Mesp1 daughter cells may receive different developmental signals such as Wnt or BMP signal gradients with respect to their physical location and polarity within the developing embryo. The other possibility as to how Mesp1 contributes to different lineages is by interacting with different binding partners that lead to different or altered regulation of certain downstream genes. Therefore, to identify the interaction partners of Mesp1 in developmental stages could help solve this mystery. One way to achieve this is to pull down the interaction complexes at different stages of embryoid bodies post-differentiation using a Mesp1 specific antibody and performing Mass Spectrometry analysis to identify the co-

immunoprecipitated proteins that were bound to Mesp1. This information will further reveal the molecular mechanism(s) as to how Mesp1 guides committed cells toward maturation stages. It also helps us to understand whether there are combinatorial regulation elements.

4.1.4 Study the Molecular Mechanism of Regulation of Mesp1

One of most interesting characteristics of Mesp1 is that it is expressed in a very narrow time frame during developmental stage or it is transiently expressed at a specific time and is not expressed outside of the specific time frame. How to control its expression and how to down-regulate the transcription and degrade Mesp1 protein is another interesting topic. The repression of gene transcription can be achieved at the genomic and at post-transcriptional levels by genomic repression of the enhancer and/or promoter region, histone methylation, and/or miRNA inhibition. Identifying potential regulating factors such as downstream transcription factors or histone modifiers can help us understand how Mesp1 expression is so tightly regulated endogenously. Second, searching potential miRNA binding sites on Mesp1's 3'untranslated region (UTR) will provide a starting list of miRNAs that are regulating Mesp1 post-transcriptionally by blocking translation.

Another way to regulate signalling pathways is by post-translational regulation. One example is the Notch signal pathway. This signaling cascade needs to be switched quickly during somitogenesis. It was shown that Mesp2 can repress Notch signalling

pathway by forming a heterodimer with Maml1 and causeing subsequent Maml1 degradation through proteasomal degradation. Protein degradation is more efficient compared to RNA repression in that it elicits a much faster cellular response. Is it possible that Mesp1 can be degraded by proteasomal degradation through interacting with certain protein binding partners? What interaction partners can cause Mesp1 destabilization and what and where is the degradation recognition domain on Mesp1? This molecular mechanism can help us to explain the transient expression of Mesp1 and how the developmental timing in regulating Mesp1 can be so specific. It also helps us to extend the half-life of Mesp1 if Mesp1 protein is needed for clinical purposes.

4.1.5 Multiple Roles of ETS Factors

There are twenty-seven ETS transcription factors in the ETS family. They share the same conserved DNA binding site while regulating unique gene sets. The Nkx2-5 luciferase assay provided a fast screening method to understand their potential roles in cardiac development. Microarrays in NHDF also provided an overview of changes in global gene expression patterns implicating different regulatory pathways in adult fibroblasts. However the mechanism involved in how individual factors play different roles with respect to time and location need to be achieved by further studies. Gain-of-function and loss-of-function studies are the best ways to answer the signal pathways they are involved with. The deletion of a single ETS factor in a mouse model results in different phenotypes. For example, loss of Spi1, Spib, and Fli1 leads to defective B cells. Loss of Elf4, Elk4, and Ets1 causes T or NK cells to be defective. Loss of Etv6, Etv2,

and Erg leads to faulty hematopoietic stem cell development. How these ETS factors coordinate the transcription networks with Mesp1 to guide mesoderm development such as heart formation and hematopoietic stem cell maturation would be an interesting field to study.

4.1.6 Combination of ETS Factors or Other Cardiac Genes to Reprogram Cells into Cardiac Progenitors

Our data showed the effect of individual ETS factors on NHDF cells such as gene expression and cell surface marker profiling. It seemed ETS2 and ETV4 were able to reprogram these cells into the cardiac lineage the best among the entire set of ETS factors that we screened. However, whether a single gene is sufficient or a combination of more than two factors may work better is still unexplored. There was a report that showed the deletion of ETV1, ETV4 and ETV5 leads to a complete absence of heart tissue in Zebra fish. The deletion of both Mesp1 and Mesp2 also showed a loss of cardiac formation while a single deletion of Mesp1 only led to cardia bifida. The function of Mesp1 can be compensated by Mesp2 in early cardiac mesoderm specification. The function of a transcription factor can be compensated by another, factor from the same family in some cases. Therefore, it is worth trying various combinations of ETS transcription factors that are in the same sub-family to see which members are acceptable replacements for others.

4.1.7 Reprogramming Human Adult Cells into Cardiac Progenitors

More and more studies provide evidence that shows switching of terminally differentiated cells into another fate is practical. Introducing transcription factors such as Gata4, Mef2c, Tbx5, and Hand2 converted fibroblasts into cardiomyocytes *in vitro* and *in vivo* in mouse models. By adding miR-1 and miR-133 to the previous recipe successfully converted neonatal foreskin into cardiac cells. A recent report showed that stressful environments such as a low pH bath can also convert cells into stem-like cells even more efficiently.

The concept of reprogramming adult cells is more and more accepted in the scientific community. However, the reprogramming processes usually take at least a week to a month to reach mature and beating cardiomyocytes. It becomes a big challenge to reprogram sufficient cells in a short time for clinical usage. A second concern is that the reprogrammed cells from patients are assumed to maintain the same major histocompatibility complex (MHC). However, there is a report showing the MHC will alter after the reprogramming process. It is critical to evaluate the MHC in reprogrammed cells before launching any clinical studies. A third concern is to control proliferating cells from forming teratomas. The pluripotent cells transplanted into animals may derive to all three germ layers and become a teratoma, which is not beneficial to repair and creates more problems than it solves. Although we showed the ETS and Mesp1 derived cells may contribute to mesoderm and endoderm, no evidence

showing ectoderm programs was observed. This needs to be further confirmed by *in vivo* studies to prevent any oncogenic issues that may arise.

4.2 Conclusion

Mesp1 has been considered to play the top hierarchal role during cardiogenesis. Our studies provide a detailed mechanism from multiple molecular approaches. The Mesp1 expression pattern in early mouse embryo was first revealed. The protein expression starts from the posterior region of E6.5. At E7.5, Mesp1 migrates toward the anterior embryonic region from the primitive streak. It migrates and contributes to the mesoderm layer and our newly discovered endoderm layer. By E8.5 and afterward, Mesp1 is not significantly detectable. Similar results were demonstrated in both mRNA and protein levels in embryoid bodies *in vitro*.

A Mesp1 lineage tracing ES cell line was also established and characterized. By using this cell line, we were able to purify the Mesp1 expressing population by flow cytometry and able to perform next generation sequencing with Mesp1 specific, endogenous antibody. The sequencing results confirmed the developmental roles during cardiogenesis. Mesp1 target genes involved in cardiac development included *Gata4*, *Myocd*, *Hand2,* and *Cxcr4* and these were confirmed with a Mesp1 over-expressed system. Moreover, the results suggested Mesp1 also regulates genes in endoderm development such as *Foxa2*, *Gata6,* and *Pitx2*. The lineage tracing ES cell line confirmed the endoderm marker Foxa2 was present in Mesp1 daughter cells. We showed

Mesp1 expresses in early mesoderm and endoderm and provided the information of the binding region of downstream targets.

The roles of the ETS family of transcription factors were focused on oncogenic processes of tumor formation and the development of hematopoietic cells. We showed the expression pattern in cardiac development via embryoid body development. Some ETS factors such as ETS2, ETV4, and SPDEF could physically interact with Mesp1 and up-regulated Nkx2-5 transcription activity. By transducing ETV4 or ETS2 with Mesp1 in human fibroblast cells led to up-regulation of essential genes for cardiogenesis such as *GATA4*, *NKX2-5*, *MEF2C* and *TNNT2*. ETV4 and ETS2 also up-regulated the cell surface marker CD184 (CXCR4). The screening of ETS factors provided their roles during cardiogenesis and this information could benefit to the reprogramming process of adult cells to be utilized in the future.

The reprogramming of adult cells toward other cell types requires genes activation and culture conditions. In our studies, introducing ETS2 and MESP1 by lentiviral particle infection or TAT-protein delivery and combining Activin A and BMP2 treatment was able to convert NHDF cells into cardiac progenitors. The reprogrammed cells were evaluated through the characterization of gene expression, surface marker presence, along with structural proteins and overall cell morphology.

To sum up, our studies provided detailed information of the molecular regulation of Mesp1 and ETS transcription factors during early cardiac development. Through understanding the transcriptional machinery, we further manipulated adult cells and evaluated the reprogrammed cells by multiple approaches. This work helps to solve how cardiac progenitors are regulated and provided a method for reprogramming adult cells for future clinical applications that could be used not only in cardiac related applications but also in other non-cardiac tissues such as the pancreas for treating disease such as diabetes.

REFERENCES

1. Kelly, B.B., J. Narula, and V. Fuster, *Recognizing global burden of cardiovascular disease and related chronic diseases.* Mt Sinai J Med, 2012. **79**(6): p. 632-40.

2. Go, A.S., et al., *Heart disease and stroke statistics--2014 update: a report from the american heart association.* Circulation, 2014. **129**(3): p. e28-e292.

3. Meade, T.W., et al., *Haemostatic function and cardiovascular death: early results of a prospective study.* Lancet, 1980. **1**(8177): p. 1050-4.

4. Ebrahim, S., et al., *Carotid plaque, intima media thickness, cardiovascular risk factors, and prevalent cardiovascular disease in men and women: the British Regional Heart Study.* Stroke, 1999. **30**(4): p. 841-50.

5. Kleiger, R.E., et al., *Decreased heart rate variability and its association with increased mortality after acute myocardial infarction.* Am J Cardiol, 1987. **59**(4): p. 256-62.

6. Taylor, A.J., J.E. Udelson, and V. Fuster, *Training cardiovascular fellows in cardiovascular magnetic resonance and vascular imaging; Current status following the core cardiovascular training symposium (COCATS-2) guidelines.* J Am Coll Cardiol, 2004. **43**(11): p. 2108-12.

7. Beyersdorf, F., *The use of controlled reperfusion strategies in cardiac surgery to minimize ischaemia/reperfusion damage.* Cardiovasc Res, 2009. **83**(2): p. 262-8.

8. Asante-Korang, A., et al., *Experience of FK506 immune suppression in pediatric heart transplantation: a study of long-term adverse effects.* J Heart Lung Transplant, 1996. **15**(4): p. 415-22.

9. Chinen, J. and W.T. Shearer, *Advances in basic and clinical immunology in 2009.* J Allergy Clin Immunol, 2010. **125**(3): p. 563-8.

10. Lloyd-Jones, D.M., *Cardiovascular risk prediction: basic concepts, current status, and future directions.* Circulation, 2010. **121**(15): p. 1768-77.

11. Poss, K.D., L.G. Wilson, and M.T. Keating, *Heart regeneration in zebrafish.* Science, 2002. **298**(5601): p. 2188-90.

12. Oberpriller, J.O. and J.C. Oberpriller, *Response of the adult newt ventricle to injury.* J Exp Zool, 1974. **187**(2): p. 249-53.

13. Mignone, J.L., et al., *Cardiogenesis from human embryonic stem cells.* Circ J, 2010. **74**(12): p. 2517-26.

14. Pijnappels, D.A., S. Gregoire, and S.M. Wu, *The integrative aspects of cardiac physiology and their implications for cell-based therapy.* Ann N Y Acad Sci, 2010. **1188**: p. 7-14.

15. Porrello, E.R., et al., *Transient regenerative potential of the neonatal mouse heart.* Science, 2011. **331**(6020): p. 1078-80.

16. Bergmann, O., et al., *Evidence for cardiomyocyte renewal in humans.* Science, 2009. **324**(5923): p. 98-102.

17. Senyo, S.E., et al., *Mammalian heart renewal by pre-existing cardiomyocytes.* Nature, 2013. **493**(7432): p. 433-6.

18. Rossant, J. and P.P. Tam, *Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse.* Development, 2009. **136**(5): p. 701-13.

19. Smith, J.C., et al., *Inducing factors and the control of mesodermal pattern in Xenopus laevis.* Development, 1989. **107 Suppl**: p. 149-59.

20. Zorn, A.M. and J.M. Wells, *Vertebrate endoderm development and organ formation.* Annu Rev Cell Dev Biol, 2009. **25**: p. 221-51.

21. Arnold, S.J. and E.J. Robertson, *Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo.* Nat Rev Mol Cell Biol, 2009. **10**(2): p. 91-103.

22. David, R., et al., *MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling.* Nat Cell Biol, 2008. **10**(3): p. 338-45.

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

24. Dodou, E., et al., *Mef2c is a direct transcriptional target of ISL1 and GATA factors in the anterior heart field during mouse embryonic development.* Development, 2004. **131**(16): p. 3931-42.

25. Srivastava, D., *Making or breaking the heart: from lineage determination to morphogenesis.* Cell, 2006. **126**(6): p. 1037-48.

26. Cai, C.L., et al., *Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart.* Dev Cell, 2003. **5**(6): p. 877-89.

27. Garry, D.J. and E.N. Olson, *A common progenitor at the heart of development.* Cell, 2006. **127**(6): p. 1101-4.

28. Kelly, R.G., N.A. Brown, and M.E. Buckingham, *The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm.* Dev Cell, 2001. **1**(3): p. 435-40.

29. Schultheiss, T.M., J.B. Burch, and A.B. Lassar, *A role for bone morphogenetic proteins in the induction of cardiac myogenesis.* Genes Dev, 1997. **11**(4): p. 451-62.

30. Ladd, A.N., T.A. Yatskievych, and P.B. Antin, *Regulation of avian cardiac myogenesis by activin/TGFbeta and bone morphogenetic proteins.* Dev Biol, 1998. **204**(2): p. 407-19.

31. Barron, M., M. Gao, and J. Lough, *Requirement for BMP and FGF signaling during cardiogenic induction in non-precardiac mesoderm is specific, transient, and cooperative.* Dev Dyn, 2000. **218**(2): p. 383-93.

32. Zhang, H. and A. Bradley, *Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development.* Development, 1996. **122**(10): p. 2977-86.

33. Winnier, G., et al., *Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse.* Genes Dev, 1995. **9**(17): p. 2105-16.

34. Matzuk, M.M., T.R. Kumar, and A. Bradley, *Different phenotypes for mice deficient in either activins or activin receptor type II.* Nature, 1995. **374**(6520): p. 356-60.

35. Kosaki, R., et al., *Left-right axis malformations associated with mutations in ACVR2B, the gene for human activin receptor type IIB.* Am J Med Genet, 1999. **82**(1): p. 70-6.

36. Eswarakumar, V.P., I. Lax, and J. Schlessinger, *Cellular signaling by fibroblast growth factor receptors.* Cytokine Growth Factor Rev, 2005. **16**(2): p. 139-49.

37. Dailey, L., et al., *Mechanisms underlying differential responses to FGF signaling.* Cytokine Growth Factor Rev, 2005. **16**(2): p. 233-47.

38. Davidson, B., et al., *FGF signaling delineates the cardiac progenitor field in the simple chordate, Ciona intestinalis.* Genes Dev, 2006. **20**(19): p. 2728-38.

39. Amaya, E., T.J. Musci, and M.W. Kirschner, *Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in Xenopus embryos.* Cell, 1991. **66**(2): p. 257-70.

40. Ornitz, D.M. and N. Itoh, *Fibroblast growth factors.* Genome Biol, 2001. **2**(3): p. REVIEWS3005.

41. Chien, K.R., I.J. Domian, and K.K. Parker, *Cardiogenesis and the complex biology of regenerative cardiovascular medicine.* Science, 2008. **322**(5907): p. 1494-7.

42. Powers, C.J., S.W. McLeskey, and A. Wellstein, *Fibroblast growth factors, their receptors and signaling.* Endocr Relat Cancer, 2000. **7**(3): p. 165-97.

43. Ornitz, D.M., *Regulation of chondrocyte growth and differentiation by fibroblast growth factor receptor 3.* Novartis Found Symp, 2001. **232**: p. 63-76; discussion 76-80, 272-82.

44. Slack, J.M., et al., *Mesoderm induction in early Xenopus embryos by heparin-binding growth factors.* Nature, 1987. **326**(6109): p. 197-200.

45. Sugi, Y., et al., *Fibroblast growth factor (FGF)-4 can induce proliferation of cardiac cushion mesenchymal cells during early valve leaflet formation.* Dev Biol, 2003. **258**(2): p. 252-63.

46. Yamaguchi, T.P., et al., *fgfr-1 is required for embryonic growth and mesodermal patterning during mouse gastrulation.* Genes Dev, 1994. **8**(24): p. 3032-44.

47. Duarte, A., et al., *Dosage-sensitive requirement for mouse Dll4 in artery development.* Genes Dev, 2004. **18**(20): p. 2474-8.

48. Loomes, K.M., et al., *The expression of Jagged1 in the developing mammalian heart correlates with cardiovascular disease in Alagille syndrome.* Hum Mol Genet, 1999. **8**(13): p. 2443-9.

49. Timmerman, L.A., et al., *Notch promotes epithelial-mesenchymal transition during cardiac development and oncogenic transformation.* Genes Dev, 2004. **18**(1): p. 99-115.

50. Uyttendaele, H., et al., *Notch4/int-3, a mammary proto-oncogene, is an endothelial cell-specific mammalian Notch gene.* Development, 1996. **122**(7): p. 2251-9.

51. Kokubo, H., et al., *Hesr1 and Hesr2 regulate atrioventricular boundary formation in the developing heart through the repression of Tbx2.* Development, 2007. **134**(4): p. 747-55.

52. Nusse, R., *The Wnt gene family in tumorigenesis and in normal development.* J Steroid Biochem Mol Biol, 1992. **43**(1-3): p. 9-12.

53. Logan, C.Y. and R. Nusse, *The Wnt signaling pathway in development and disease.* Annu Rev Cell Dev Biol, 2004. **20**: p. 781-810.

54. Karim, R., et al., *The significance of the Wnt pathway in the pathology of human cancers.* Pathology, 2004. **36**(2): p. 120-8.

55. Ueno, S., et al., *Biphasic role for Wnt/beta-catenin signaling in cardiac specification in zebrafish and embryonic stem cells.* Proc Natl Acad Sci U S A, 2007. **104**(23): p. 9685-90.

56. Kohn, A.D. and R.T. Moon, *Wnt and calcium signaling: beta-catenin-independent pathways.* Cell Calcium, 2005. **38**(3-4): p. 439-46.

57. Maye, P., et al., *Multiple mechanisms for Wnt11-mediated repression of the canonical Wnt signaling pathway.* J Biol Chem, 2004. **279**(23): p. 24659-65.

58. Marvin, M.J., et al., *Inhibition of Wnt activity induces heart formation from posterior mesoderm.* Genes Dev, 2001. **15**(3): p. 316-27.

59. Nakamura, T., et al., *A Wnt- and beta -catenin-dependent pathway for mammalian cardiac myogenesis.* Proc Natl Acad Sci U S A, 2003. **100**(10): p. 5834-9.

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

61. Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. **282**(5391): p. 1145-7.

62. Takeuchi, J.K. and B.G. Bruneau, *Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors.* Nature, 2009. **459**(7247): p. 708-11.

63. Kattman, S.J., et al., *Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines.* Cell Stem Cell, 2011. **8**(2): p. 228-40.

64. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.

65. Song, K., et al., *Heart repair by reprogramming non-myocytes with cardiac transcription factors.* Nature, 2012. **485**(7400): p. 599-604.

66. Qian, L., et al., *In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes.* Nature, 2012. **485**(7400): p. 593-8.

67. Saga, Y., et al., *MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation.* Development, 1996. **122**(9): p. 2769-78.

68. Saga, Y., et al., *MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube.* Development, 1999. **126**(15): p. 3437-47.

69. Saga, Y., S. Kitajima, and S. Miyagawa-Tomita, *Mesp1 expression is the earliest sign of cardiovascular development.* Trends Cardiovasc Med, 2000. **10**(8): p. 345-52.

70. Kitajima, S., et al., *MesP1 and MesP2 are essential for the development of cardiac mesoderm.* Development, 2000. **127**(15): p. 3215-26.

71. Lindsley, R.C., et al., *Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs.* Cell Stem Cell, 2008. **3**(1): p. 55-68.

72. Bondue, A., et al., *Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification.* Cell Stem Cell, 2008. **3**(1): p. 69-84.

73. McLaughlin, F., et al., *Combined genomic and antisense analysis reveals that the transcription factor Erg is implicated in endothelial cell differentiation.* Blood, 2001. **98**(12): p. 3332-9.

74. Spyropoulos, D.D., et al., *Hemorrhage, impaired hematopoiesis, and lethality in mouse embryos carrying a targeted disruption of the Fli1 transcription factor.* Mol Cell Biol, 2000. **20**(15): p. 5643-52.

APPENDIX

Working Model of Regulation of Mesp1 and ETS Transcription Factors

The schematic shows the working model of how Mesp1 coordinating with ETS transcription factors to regulate cardiac progenitors.

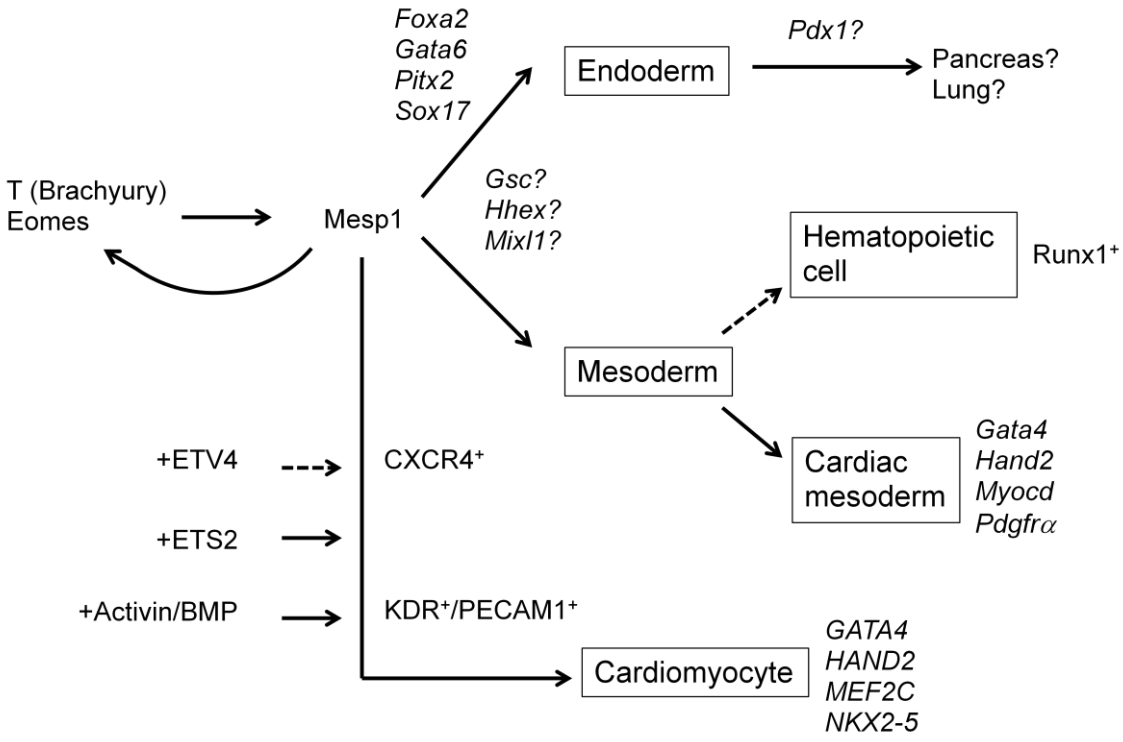


Figure A1 The Working Model of Regulation of Mesp1 and ETS Transcription Factors