Epigenetic regulatory mechanisms that govern cardiovascular development

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Javier Eli Sierra-Pagan

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Daniel J. Garry, M.D., Ph.D.

June 2022

© Javier Eli Sierra-Pagan 2022
ALL RIGHTS RESERVED

ACKNOWLEDGEMENTS

First, I would like to thank mi advisor, Daniel J. Garry, for taking a chance and allowing me to work in his laboratory. He introduced me to a whole new area of research within cardiology, particularly genomics and bioinformatics. He never said no to any of my ideas, was incredibly supportive and I thank him for that.

Second, I would like to thank current and previous members of the Garry Lab for their unconditional support and help throughout the past four years. Particularly, I would like to thank Bhairab Singh, Satyabrata Das and Thijs A. Larson, you gave me so much and I will forever be thankful. Thank you for supporting me every day and teaching me so much.

To all the members of the Lillehei Heart Institute, University of Minnesota Genome Center and collaborators from the University of Minnesota in general (Tamas Alexy), thank for collaborating and helping out with experiments and reagents when needed.

I would also like to thank Jonathan A. Epstein and Arnaldo Diaz for believing in me when I was an undergraduate student. You both recruited me to the University of Pennsylvania to work in one of the best cardiovascular research laboratories in the nation. You both gave me the space and opportunity for me to grow as a scientist. I would not be an MD/PhD student if not for you two.

To my clinical mentor Kurt Prins, thank you for welcoming me into your clinic. I have learned so much about pulmonary hypertension and patient care from you and Sasha Prisco. I'm fortunate to have chosen to spend the last two years of my PhD in your clinic.

To my thesis committee (Daniel J. Garry, Bruce Blazar, Peter Bitterman, Scott McIvor and Gregory Vercellotti), thank you for pushing me to do the best science I could do and for teaching me so much. I couldn't have asked for a better committee.

To the MSTP program and the leadership, thank you. Particularly Yoji Shimizu, thank you for taking a chance with me six years ago. You have changed my life by allowing me to pursue my dreams of becoming a physician scientist.

To my MSTP cohort, thank you for all the great memories and for being amazing classmates. I had a lot of fun and can't wait to see what we all end up doing with our careers.

To my family and friends. You all have supported me throughout the last 29 years of my life. I couldn't have done this without you all.

Lastly, to my partner Sara and my dog Darwin. Thank you for bringing so much happiness and joy to my life on a daily basis. I'm the luckiest to have you both in my life.

DEDICATION

This thesis is dedicated to most influential group of people in my life, my family. You all have shaped me over the years and have been my guiding lights. I love you all dearly. This achievement is both yours and mine.

To my parents Javier Sierra-Fonseca and Elizabeth Pagan-Rivera. You two have and always will be role models and the source of my inspiration, perseverance, and confidence to achieve anything I put my mind to. You have given me the World, the opportunity to dream and I am forever thankful for that.

To my sister Ignerys Sierra-Pagan. You have always pushed me to be better, stronger and to work harder than anybody else. You have always been my best and worst critic. As my older sister and best friend, you have truly influenced and changed how I approach my daily life. I cannot imagine doing this without your support. You have given me the blessing and joy of being an uncle to two beautiful nephews Valentina and Sebastian. They are the best gift you have ever given me and I will forever be thankful for them.

To my partner Sara Bolivar Wagers. Thank you for accompanying me through this process. You have been a blessing and I consider myself the luckiest person to have you with me. You and Darwin are the coolest group to be around. I'm the happiest person with you both.

Most importantly, to my late grandfather William Pagan-Ocasio. Thank you for giving me a childhood full of joy and so many happy memories. You were my motivation for pursuing a career in cardiology and my passion for this area of research only grows stronger every day because of you. I love you and I miss you every day abuelo William.

ABSTRACT

Cardiovascular disease (CVD) remains the number one cause of death in the United States and the World. The clinical outcomes of patients with heart failure, a form of CVD, remain poor because current clinical therapies do not address a critical feature of heart failure which is the loss of functional cardiac muscle. To decrease the morbidity and mortality in these patients, several strategies are being developed to replace the loss of functional cardiac muscle with new one. Two attractive strategies for treating CVD involve converting cardiac fibroblasts (reprogramming) into functional muscle or vascular cells and promoting cell cycle re-entry of adult cardiomyocytes following cardiac injury to replace the dead muscle. While the adult mammalian heart has limited regenerative potential following injury, the embryonic and neonatal mammalian heart has a remarkable regenerative capacity. Therefore, our goal for these studies was to define new factors and mechanisms that could enhance repair and regeneration in the adult mammalian heart. To this end, in this thesis, we identified novel epigenetic regulatory mechanisms that govern cardiovascular development, particularly within the vascular and cardiac muscle lineages. Our first finding is that we identified that ETV2 functions as a pioneer transcription factor that relaxes closed chromatin and regulates endothelial development. We did this by comparing engineered embryonic stem cell (mESCs) differentiation and reprogramming models (MEFs) with multi-omics techniques, we demonstrated that ETV2 was able to bind nucleosomal DNA and recruit BRG1. The recruitment of BRG1 led to the remodeling of chromatin around endothelial genes and helped to maintain an open

configuration, resulting in increased H3K27ac deposition. Our second finding is that we discovered a signaling cascade where ETV2 regulates RHOJ expression during endothelial progenitor cell migration. We did this by combining computational genomics (RNAseq, ATACseq and ChIPseq) to discover that ETV2 regulates migratory pathways through the expression of RHOJ, particularly in developing endothelial progenitor cells (E7.75 and E8.5 mouse embryos and developing mESCs). Our third finding is that we identified FOXK1 as an essential transcriptional and epigenetic regulator of cardiovascular development. We used in mESCs that lacked FOXK1 expression and discovered that in its absence, cardiac muscle development is significantly affected, both at the transcriptional and chromatin level. Mechanistically, we also discovered that FOXK1 represses Wnt signaling, particularly *Wnt6*, to promote the development of cardiac progenitor cells. ETV2 has the capacity to reprogram fibroblasts to mature vascular cells and our findings identified new mechanisms we can explore to better reprogram cardiac fibroblasts. Additionally, FOXK1 is a known cell cycle regulator and together we these findings, becomes an attractive molecule that could be used to promote cell cycle re-entry of adult cardiomyocytes following injury. Altogether these studies provide exciting data for the field of cardiac regeneration but future studies will be needed in vivo to determine the potential benefit of these molecules following cardiac injury.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i-ii
DEDICATIONS	iii
ABSTRACT	iv-v
TABLE OF CONTENTS	vi
LIST OF FIGURES	vii-ix
LIST OF TABLES	x
CHAPTER 1: INTRODUCTION	1-8
CHAPTER 2: THE REGULATORY ROLE OF PIONEER FACTORS DURING CARDIOVASCULAR LINEAGE SPECIFICATION	9-20
CHAPTER 3: ETV2 FUNCTIONS AS A PIONEER FACTOR TO REGULATE AND REPROGRAM THE ENDOTHELIAL LINEAGE	21-95
CHAPTER 4: ETV2-*Rhoj* CASCADE REGULATES ENDOTHELIAL PROGENITOR CELL MIGRATION DURING EMBRYOGENESIS	96-160
CHAPTER 5: FOXK1 REGULATES WNT SIGNALING TO PROMOTE CARDIOGENESIS	161-197
CHAPTER 6: DISCUSSION & FUTURE DIRECTIONS	198-208
REFERENCES	209-224

LIST OF FIGURES

1. Figure 1-1	Schematic of cardiac development	8
2. Figure 2-1	Pioneer factors drive lineage specification	20
3. Figure 3-1	ETV2 promotes the endothelial program in both MEFs and EBs	58-59
4. Figure 3-2	ETV2 targets nucleosomes during reprogramming	60-61
5. Figure 3-3	ETV2 recruits BRG1 for chromatin remodeling	62-63
6. Figure 3-4	Phased nucleosomes are established surrounding ETV2 peaks	64-65
7. Figure 3-5	*Brg1* knockdown resulted in a significant decrease in cells expressing FLK1 during reprogramming	66
8. Figure 3-6	ETV2 requires BRG1 to activate downstream genes during reprogramming	67
9. Figure 3-7	*Brg1* conditional knockout in ES/EBs	68-69
10. Extended Data Fig. 3-1	Characterization of mouse embryonic fibroblasts that overexpress ETV2 (*iHA-Etv2* MEFs) following the addition of doxycycline to reprogram MEFs to endothelial cells	70-71
11. Extended Data Fig. 3-2	Expression of endothelial transcripts in the FLK1^{+} cell population at day 7 post-ETV2 induction during MEF reprogramming	72-73
12. Extended Data Fig. 3-3	Expression profile of immune response related genes and significance of pathways in reprogrammed MEFs are upregulated post induction of ETV2	74
13. Extended Data Fig. 3-4	ETV2 overexpression promotes endothelial lineage development in *iHA-Etv2* ES/EBs	75
14. Extended Data Fig. 3-5	Commonly up- and down-regulated genes in FLK1+ cell populations from ETV2 induced ES/EB differentiation and MEF reprogramming	76
15. Extended Data Fig. 3-6	Combined RNA-seq and ATAC-seq analysis during EB and MEF reprogramming	77-78
16. Extended Data Fig. 3-7	The ETV2 bound sites at day 1 post-Etv2 induction in MEFs target the nucleosomes and the analysis of ETV2 ChIP-seq peaks during EB and MEF reprogramming	79-80
17. Extended Data Fig. 3-8	*Brg1* knockdown in *iHA-Etv2* MEFs using shRNA lentiviral particles	81
18. Extended Data Fig. 3-9	*Dek*, *Znhit1* and *Cdh8* knockdown does not impact ETV2 mediated endothelial reprogramming	82
19. Extended Data Fig. 3-10	ETV2 requires BRG1 to activate downstream genes during reprogramming	83
20. Supplementary Figure 3-1	ETV2 requires BRG1 to activate downstream genes during reprogramming. GSEA plots comparing clusters 1, 2, 3 and 4 out of the seven clusters in Figure 1d demonstrates that the Inflammatory response is	84-85

	significantly upregulated for clusters 1 and 3 compared to clusters 2 and 4	
21. Supplementary Figure 3-2	Gene set enrichment analysis (GSEA) for HATs, HDACs, Inflammatory response and NF-kappaB signaling pathway during ES/EB differentiation and MEF reprogramming	86-87
22. Supplementary Figure 3-3	Gene set enrichment analysis (GSEA) plots obtained from previously published scRNA-seq data for MEF reprogramming shows significant upregulation of the Inflammatory response pathway during GMT induced cardiac reprogramming	88-89
23. Supplementary Figure 3-4	Gene ontology annotation identifies commonly up-regulated genes in both ES/EB differentiation and MEF reprogramming following ETV2 overexpression	90
24. Supplementary Figure 3-5	The transcription factors whose motif associated chromatin accessibility and expressions were consistently changed in both EB and MEF on Etv2 induction	91
25. Supplementary Figure 3-6	Comparison of the single cell expression profiles of undifferentiated MEFs and *Brg1* KD MEFs	92
26. Figure 4-1	RNAseq, ChIPseq and ATACseq analyses showed enrichment of cell migration program following the overexpression of ETV2 in the ESC/EB system	123-124
27. Figure 4-2	ETV2 promotes cell migration	125-126
28. Figure 4-3	Co-expression of *Rhoj* and *Etv2* during embryogenesis	127-128
29. Figure 4-4	ETV2 regulates the expression of *Rhoj*	129
30. Figure 4-5	ETV2 is an upstream regulator of *Rhoj*	130
31. Figure 4-6	ETV2-*Rhoj* network regulates migration	131-132
32. Supplementary Figure 4-I	Top 10 Gene Ontology (GO) biological processes	133
33. Supplementary Figure 4-II	ETV2 overexpression promotes migration by increasing the formation of sprouts	134
34. Supplementary Figure 4-III	Characterization of inducible over-expression of mouse embryonic fibroblasts (*iHA-Etv2* MEFs)	135
35. Supplementary Figure 4-IV	*Gata1* transcript is expressed in the blood lineage	136
36. Supplementary Figure 4-V	Expression analysis of *Rhoj* during ESC/EB differentiation	137

37. Supplementary Figure 4-VI	ETV2 binds to the upstream region of the *Rhoj* promoter	138
38. Supplementary Figure 4-VII	Screening for *Rhoj* shRNA clones using lentiviruses expressing shRNAs against *Rhoj*	139
39. Supplementary Figure 4-VIII	Validation of the ETV2-*Rhoj* network using a second shRNA (#3)	140
40. Figure 5-1	FOXK1 regulates mesodermal progenitor cell development	184
41. Figure 5-2	FOXK1 regulates cardiac developmental transcriptional networks	185
42. Figure 5-3	FOXK1 is an epigenetic regulator of cardiac development	186
43. Figure 5-4	FOXK1 regulates cardiogenesis in differentiating EBs	187
44. Figure 5-5	FOXK1 regulates Wnt signaling to promote cardiogenesis	188-189
45. Figure 5-6	FOXK1 is a transcriptional repressor of Wnt signaling	190
46. Figure 5-7	FOXK1 labels developing cardiac cells in the mouse embryo	191
47. Supplementary Figure 5-1	FOXK1 protein expression in ES/EBs	192
48. Supplementary Figure 5-2	Flow cytometry gating strategy	193
49. Supplementary Figure 5-3	RNAseq analysis of D3 and D5 EBs	194
50. Supplementary Figure 5-4	ATACseq analysis and RNAseq integration analysis of D3 and D5 EBs	195
51. Supplementary Figure 5-5	RNAseq analysis of D10 EBs	196
52. Supplementary Figure 5-6	*Wnt3a* expression in differentiating EBs	197

LIST OF TABLES

1. Supplementary table 3-1	Taqman qPCR probes	93
2. Supplementary table 3-2	Re-ChIP primers	94
3. Supplementary table 3-3	ChIP qPCR primers for ELK3	95
4. Supplementary table 4-1	Bulk-RNAseq analysis for the cell migratory genes (GO:0048870) arranged based on the relative expression at 6h and 12h time point following -Dox and +Dox treatment during ES/EB differentiation	141-150
5. Supplementary table 4-2	ETV2 ChIPseq peak analysis of migratory genes	151-159
6. Supplementary table 4-3	Taqman probe sets	160

CHAPTER 1: INTRODUCTION

Cardiovascular disease and regeneration

Cardiovascular disease (CVD) is the leading cause of death in the world and current therapies are limited due to the inability to promote remuscularization of the injured heart. In the absence of remuscularization, the injured heart forms a scar following a myocardial infarction (1) which can progress towards heart failure (2-5). These cardiovascular diseases are chronic, debilitating, lethal and warrant the development of novel therapies. Insufficient vasculature and impaired perfusion are critical factors affecting the morbidity and mortality observed in CVD (6, 7). The adult mammalian heart has a limited regenerative capacity and intense interest has been directed towards enhancing this process (2, 8-10). One approach is to promote vasculogenesis in order to promote heart regeneration (11, 12). Clinical trials using exogenous factors such as vascular endothelial growth factor (VEGF) to treat ischemia have shown conflicting results that have been attributed to the inability of such factors to drive vasculogenesis (12). Another approach is to promote cell cycle re-entry of adult cardiomyocytes following MI. Adult cardiomyocytes are arrested in the cell cycle (G_0) and finding ways to promote cell cycle re-entry can be beneficial for the treatment of CVD (2, 9). Unlike adult hearts, embryonic and neonatal hearts can fully regenerate following injury (2, 9, 13, 14). Therefore, understanding the mechanisms that drive this regenerative potential early on can lead to the development of new therapies for CVD by promoting vasculogenesis or adult cardiomyocyte proliferation. In this thesis, we identify novel epigenetic regulatory mechanisms that can impact the development of therapies for CVD.

Cardiovascular development

Genetic mutations can perturb the developing cardiovascular multipotent progenitors, which can result in congenital heart disease (CHD) (15-22). CHD is the most common genetic birth defect as it afflicts approximately 1% of live births and has considerable morbidity and mortality (22-24). Therefore, it is essential to decipher the regulatory pathways that govern the specification and differentiation of mesodermal progenitors and use this information to develop targeted therapies for congenital cardiovascular (CV) diseases. Cardiovascular development is a complex well-orchestrated process governed by transcriptional regulators and signaling pathways. Cardiovascular development includes the specification, proliferation, migration and differentiation of cardiac progenitors that become electrically coupled and ultimately form a functional syncytium (15, 19-22). The mesodermal progenitor cells coalesce to form the cardiac crescent and then migrate and fuse at the midline to form the linear heart tube that consists of an inner endocardial (noncontractile) cell layer and an outer layer of cardiomyocytes that are capable of contractility (Figure 1). The heart is composed of progenitors from the primary heart field which gives rise to the left and right atria and the left ventricle while the progenitors associated with the secondary heart field give rise to the right ventricle and outflow tract (as well as some contribution to the atria) (25). Understanding the molecular mechanisms that drive regulate the development of all the structures within this lineage will be essential for the generation of new therapies for CVD.

Pioneer factors and master regulators promote lineage development and cellular reprogramming

Essential transcription factors, better known as master regulators, regulate the fate and lineage commitment of progenitor cells during development. These transcription factors are commonly identified by loss and gain of function genetic studies, regulate the lineage commitment events, and can convert/reprogram cells (fibroblasts) to different lineages (26, 27). The first ever described example of a master regulator is the MYOD family of transcription factors, who are key regulators of the myogenic lineage (28). MYOD was the first examples of a transcription factor identified to reprogram cells (fibroblasts) in vitro. Another example of a master regulator is PDX1, an important regulator of pancreatic development (29). The list of master regulators that we have accumulated to date is extensive (26). In addition to MYOD (muscle) and PDX1 (pancreas), other master regulators such as ETV2 (blood & vasculature), MESP1 (cardiovascular lineage), among others have been identified (29-35). These lineage specific transcription factors or master regulators are crucial for the different combinations of reprogramming factor cocktails we use in the field of regenerative medicine in order to develop organ specific cellular therapies (36). However, among these master regulators, a small subset of transcription factors known as pioneer factors, initiate lineage specific regulatory events to open up compacted (heterochromatic) DNA and both developmental and reprogramming processes (36-38). Pioneer factors have the capacity to reprogram cell fate by binding transcriptionally silent

genes in a closed chromatin state. By relaxing condensed chromatin, these factors enable other transcription factors to access their binding motifs to collectively activate gene expression, resulting in reprogramming (39, 40). Importantly, pioneer factor binding occurs prior to lineage specification/commitment (41, 42). Some examples of pioneer factors include: OCT3/4, SOX2 and KLF4 for iPSC generation and FOXA for hepatocyte reprogramming (43-45). In this thesis, we identify ETV2 as a novel pioneer transcription factor for the endothelial lineage.

ETV2 is an essential regulator of hematoendothelial lineage development

Our lab and others have established that ETV2 is an essential transcription factor for the development of cardiac, endothelial and hematopoietic lineages (46-51). It is expressed transiently and very early during embryogenesis (50). Global loss of ETV2 leads to the absence of all blood and vasculature resulting in lethality of the E9.5 mouse embryo (46-48). Progenitor cells expressing ETV2 daughter endocardial/endothelial and hematopoietic lineages in the WT background during development (50). Furthermore, when crossed into the *Etv2* mutant background, the *Etv2*-EYFP progenitor cells daughtered cardiomyocytes in the absence of *Etv2* (50). *Etv2* has been shown to be responsive to BMP, WNT and NOTCH signaling pathways and synergizes with FOXC2 to regulate the endothelial program by directly targeting *Scl*, *Notch4*, *Cdh5*, *Tie2* and *Flk1* (48, 52). Furthermore, ETV2 overexpression promotes vasculogenesis by reprogramming both mouse and human fibroblasts (46-48, 53-55). This ability of ETV2 to reprogram fibroblasts to endothelial cells points to its master regulatory role and the need to decipher its

role. In this thesis, we present work characterizing the role of ETV2 as a pioneer transcription factor for the endothelial lineage.

Forkhead factors are important developmental regulators

Forkhead box (FOX) proteins are a superfamily of evolutionarily transcription factors that harbor a relatively conserved DNA binding domain (the *forkhead/winged helix* domain) (56). Members of this *forkhead/winged helix* transcription factor superfamily have been shown to have essential roles during embryogenesis in lineage fate decisions, cell cycle kinetics, aging, metabolism, stem cell regulation and epigenetics (pioneer factors) (41, 45, 56). Furthermore, many of these FOX factors have been shown to have roles in cancer proliferation and tumorigenesis (57). The Foxk family consists of two members, FOXK1 and FOXK2, which have a shared structure. Our laboratory discovered FOXK1 as a transcription factor whose expression was restricted to striated muscle (cardiac and skeletal muscle) during developing (58). FOXK1 is a known regulator of myogenic stem cells (satellite cells), regeneration, cancer and cell cycle kinetics (58-70). FOXK1 KO embryos are largely lethal at E9.5 and rarely produce growth restricted viable offspring (incomplete penetrance) that are nonviable following skeletal muscle injury (59, 64). The role of FOXK2 is less well examined, but studies support that FOXK1 and FOXK2 are oncoproteins for colorectal cancer, renal cell cancer, glioma, non-small cell cancer and others (71). Nevertheless, while FOXK1 is expressed in the heart during development, the role for FOXK1 in the developing heart is unknown and warrants investigation. In this thesis, we

present data demonstrating that FOXK1 is an essential transcriptional and epigenetic regulator of cardiovascular development.

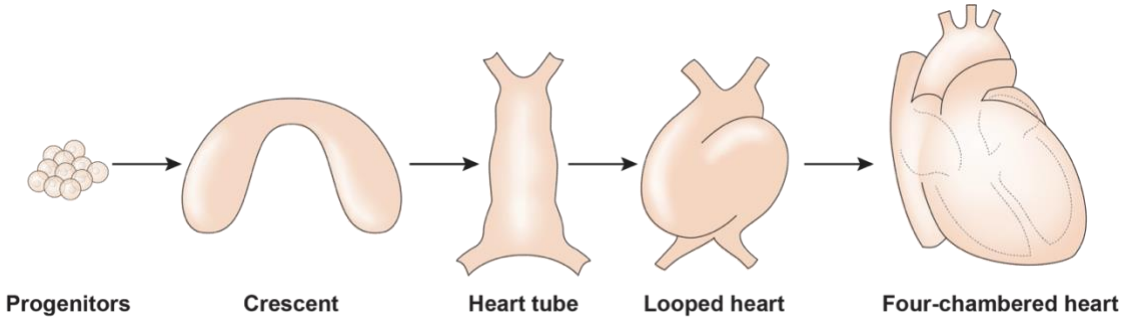


Figure 1. Schematic of cardiac development. Mesodermal progenitors coalesce to form the cardiac crescent (E7.5), the linear heart tube (E8.0), the looped heart (E9.0 to E9.5) and the four-chambered heart (E10.5).

CHAPTER 2: THE REGULATORY ROLE OF PIONEER FACTORS DURING CARDIOVASCULAR LINEAGE SPECIFICATION

Coordinated role of networks and lineage specification during cardiovascular development

Cardiovascular development is a complex and well-coordinated process that requires the specification, proliferation, migration and differentiation of progenitor cells that become coupled to form a functional syncytium within the heart (15, 19-22). Progenitor cells arising from the mesodermal layer form the early cardiac crescent and later fuse to form the linear heart tube. Two different progenitor cell populations known as the primary and secondary heart fields contribute to different structures within the mature heart, which only adds to the complexity of this process (25). Different transcription factors and signaling pathways have been described to be key regulators of cardiovascular development. A class of transcription factors known as master regulatory genes control the development of cellular lineages during differentiation and cellular reprogramming. However, within this group of master regulators, a small subset of transcription factors known as pioneer factors, have the unique capacity to bind and remodel silent and compacted regions of chromatin (nucleosomal DNA) to drive the expression of lineage specific genes that allow for development and reprogramming to occur. Because of their unique capacity to bind nucleosomal DNA and drive lineage development, pioneer factors are very important in the field of regenerative medicine and cancer biology. The focus of this review is to discuss what is known about pioneer factors within the cardiovascular lineage, particularly during development and reprogramming in an effort to generate new therapies for cardiovascular disease.

Role of master regulators during development

Essential transcription factors, better known as master regulators, regulate cell fate and lineage commitment development. These transcription factors are commonly identified by loss and gain of function genetic studies, regulate the start of lineage commitment events, and can convert/reprogram cells (fibroblasts) to different lineages (26, 27). The prototypic example of a master regulator is the MYOD family of transcription factors that are key regulators of the myogenic lineage (28). MYOD was one of the first examples of a transcription factor capable of reprogramming cells (fibroblasts). This was done in experiments where the insertion/overexpression of a cDNA construct containing *MyoD* into a fibroblast led to the formation of myoblast like cells (29-31, 55). Another example of a master regulator is PDX1, an important regulator of pancreas development (29). Global knockout of *Pdx1* in the mouse leads to the absence of the pancreas and just like MYOD, ectopic overexpression of PDX1 can convert cells to pancreatic acinar cells (72, 73). The list of master regulators that we have accumulated to date is extensive. In addition to MYOD (muscle) and PDX1 (pancreas), other master regulators such as ETV2 (blood & vasculature), MESP1 (cardiovascular lineage), OCT4/SOX2/NANOG (pluripotency), SCL/TAL1 (blood), HIF1 (hypoxia) have been identified (29-35). These lineage specific transcription factors or master regulators are crucial for the different combinations of reprogramming factors we use in the field of regenerative medicine in order to develop organ specific cellular therapies (36). However, among these master regulators, a small subset of transcription factors known as pioneer factors, initiate lineage specific regulatory

events to open up compacted (heterochromatic) DNA and both developmental and reprogramming processes (36-38). The focus of this review is to discuss pioneer factors and their role in the cardiovascular lineage, particularly focusing on the recently characterized endothelial specific pioneer transcription factor ETV2 (74).

Role of Pioneer factors during lineage specification

Pioneer factors are a specialized group of lineage-specific transcription factors that bind heterochromatic regions of DNA to promote chromatin relaxation and recruit non-pioneer transcription factors for lineage development or reprogramming to occur (Figure 1) (37, 38, 75). This is made possible because of their unique capacity to scan heterochromatin, recognize partial DNA motifs that are exposed on the surface on nucleosomes and bind to them. The complexes that are formed (chromatin remodeling and non-pioneer factors) following the binding of a pioneer factor are cell type specific and dictate the diverse mechanisms by which pioneer factors can regulate lineage specification and development. Pioneer factors were originally discovered while studying how liver specific regulatory complexes bound to heterochromatin early on during development (76, 77). In doing so, FOXA1 became the first pioneer factor described that regulates hepatic induction during early embryonic development (37, 78). A particular feature about FOXA1 is that its DNA binding domain shares a similar structure to that of linker histones, which enables this pioneer transcription factor to displace linker histones from nucleosomes to remodel chromatin and promote liver development (41, 45). The discovery of FOXA1 as a pioneer factor has led to the identification and characterization of other pioneer

factors. Perhaps, the most recognized examples of pioneer factors are Oct4, Sox2 and KLF4 (79), which promote the reprogramming of terminally differentiated fibroblasts to induced pluripotent stem cells (iPSCs) (43, 44). While c-Myc is also necessary for the reprogramming process of fibroblasts to iPSCs, unlike OSK that can bind partial DNA motifs in nucleosomal DNA in enhancers, c-Myc binds accessible regions in promoters and not nucleosomal DNA (37, 43, 44).

Role of chromatin modifying factors for the function of Pioneer factors

While pioneer factors are required for the initial binding to nucleosomal DNA, cooperation with other (non-pioneer factors) is needed in order to drive lineage development and reprogramming (1, 37, 75). Two important events need to take place following the binding of a pioneer factor 1) chromatin relaxation and 2) recruitment/interaction with other transcription factors. These two events enable the effects of pioneer factors and lineage development to occur by amplifying the signal and providing context dependent mechanisms in different regions of the genome (37, 75). Chromatin relaxation is a crucial step during lineage development where pioneer factors have been shown to promote remodeling by themselves (FOXA1) or with the assistance of the SWI/SNF complex (37, 75). The SWI/SNF complex of proteins is one of the most studied chromatin remodeling complexes. This complex is known to increase DNA accessibility to regulate the development or reprogramming of pluripotent, neuronal, cardiac and endothelial cells (74, 80). BRG1, the ATPase subunit of the SWI/SNF complex, is an important regulator of early embryonic development

as *Brg1* null embryos die pre-implantation (81). BRG1 is also an important regulator of cardiovascular development and disease as shown by previous studies using in vitro differentiation and in vivo mouse studies (82, 83). The role of BRG1 as an important chromatin remodeler is also highlighted by the fact that it interacts with at least four different pioneer factors (OCT4, GATA3, ISL1 and ETV2) to regulate chromatin remodeling, two of which are important regulators of cardiovascular development (74, 84-86).

Pioneer factors in the cardiovascular lineage

The cardiovascular lineage is composed of multiple cellular lineages, such as the muscle, vascular/endothelial and hematopoietic lineage (46, 48, 50, 87-89). While many master regulators have been described to have an important role in coordinated the development the cardiovascular lineage, few pioneer factors have been identified within this lineage (26, 75). In part, this is due to the complexity of the process that is developing all of the different cellular lineages and structures within the cardiovascular system (21, 90). These pioneer factors are ISL1, GATA4 and ETV2, and in this section we will discuss the data supporting their pioneer role and function in regulating the cardiovascular lineage.

ISL1 is an important regulator for the development of the second heart field (SHF), who was recently identified as a pioneer factor. *Isl1* KO mice lack the right ventricle, outflow tract and portions of the atria because of its role as an important regulator of SHF cardiac progenitor cells (CPCs) (91-94). In the recent work published by Gao et al., they mechanistically described that ISL1, like other

pioneer factors, regulates the development of SHF CPCs by binding nucleosomal DNA and relaxing chromatin by forming a complex with BRG1-BAF60C (84). They used both in vivo and in vitro assays to demonstrate that ISL1 binds nucleosomal DNA to regulate SHF development.

GATA4 is another important master regulator of cardiovascular development. Loss of *Gata4* has been shown to lead to cardiac defects related to looping, septation and chamber development of the heart (95-100). Additionally, GATA4 has the capacity (along with other master regulators) to reprogram fibroblasts to induced cardiomyocytes (iCMs) in vitro and in vivo (10, 101-104). While GATA4 is a key regulator of cardiovascular development, its pioneering factor function has only been described in the development of liver precursor cells and reprogramming of fibroblasts to hepatic like cells (78, 105). A recent study combined scRNAseq, ATACseq, ChIPseq and machine learning to better understand the molecular mechanisms governing iCM reprogramming using GATA4, MEF2C and TBX5 (GMT) and concluded that MEF2C and TBX5, but not GATA4 bind heterochromatin and promote chromatin remodeling during reprogramming (106). While these data do not support GATA4 is a pioneer factor for the cardiac muscle lineage, it does not completely rule it out as more studies need to be conducted, particularly during development to understand the heterochromatin binding and chromatin remodeling capabilities of GATA4 during cardiovascular development.

More recently, we identified ETV2 as a novel pioneer factor for the cardiovascular lineage that regulates and reprograms the endothelium (74).

ETV2 is an essential transcription factor for the development of endothelial and hematopoietic lineages (47, 51, 53, 107-113). ETV2 is expressed in developing progenitors that give rise to endocardial/endothelial and hematopoietic lineages, while repressing other lineages such as the cardiac lineage (47, 50, 113). Loss of *Etv2* is lethal by E8.5 in the developing mouse embryos due to lack of any vascular or blood development and congenital heart defects in aborted developing human fetuses have been reported to harbor *Etv2* mutations (46, 48, 114). Additionally, ETV2 overexpression reprograms fibroblasts to endothelial cells both in vitro and in vivo (55). Our recent findings characterized the molecular mechanism by which ETV2 regulates the endothelial lineage as a pioneer factor (74). ETV2 is capable of binding nucleosomal DNA and remodel chromatin independent of its cellular context, whether it is fibroblast reprogramming or mouse embryonic stem cell (mESC) differentiation into endothelial progenitor cells. We characterized this by combining scRNA, ATAC, NOMEseq, ChIPseq and in vitro nucleosomal binding assays to unequivocally demonstrate that ETV2 binds nucleosomal DNA during endothelial lineage reprogramming/development. Similar to ISL1, ETV2 recruits and directly interacts with BRG1. BRG1 is an essential co-factor for ETV2 to function as a pioneer factor as its knockdown and conditional knockout significantly affected the ability of ETV2 to remodel chromatin and drive endothelial lineage formation in both reprogrammed fibroblasts and differentiating mESCs, respectively. Additionally, we demonstrated that this interaction was important for enacting epigenetic changes during endothelial lineage development such as the deposition of

histone 3 lysine 27 acetylation (H3k27Ac) in regions surrounding ETV2-BRG1 bound sites. Lastly, we identified ELK3 as a co-factor that is recruited to ETV2-BRG1 bound sites following chromatin remodeling that may play an important role in endothelial cell development. Since the expression of ETV2 is transient during development, understanding how other downstream co-factors (i.e. ELK3) take over the developmental machinery following its downregulation will be important for the development of therapeutic strategies using ETV2 that aim to develop mature vasculature that can be used for ischemic diseases such as the transplantation of human vasculature (108, 115).

Unlike ISL1 and GATA4, the reported pioneer function of ETV2 in the cardiovascular lineage is independent of its cellular context, whether it is cellular differentiation or reprogramming, it functions similarly in both. Future work needs to focus on further characterizing the molecular mechanisms driving endothelial cell development/reprogramming by ETV2 to enhance therapeutic approaches to develop mature vasculature for ischemic diseases. While ETV2 is an essential regulator of hematopoietic development, we did not identify ETV2 as a pioneer factor for hematopoietic lineages and therefore hypothesize that other co-factors and pioneer factors might help in this developmental process. For example, EBF1, PU.1 and C/EBPα regulate hematopoietic development and act as pioneer factors for the B cells, DN3 t cells and macrophages (116-123). Whether these factors are regulated by ETV2 early on or they act independently of ETV2 remains to regulate the development of hematopoietic lineages remains to be elucidated. Identifying this pioneer role for ETV2 has big implications for the

development of regenerative therapies that aim to generate vasculature for ischemic diseases, particularly in the cardiovascular system (55, 108). Additionally, although not the focus of this review, developing therapies that target pioneer factors in cancer will be very important. With the known role ETV2 plays in cancer, understanding whether its ability to remodel silent/compacted chromatin as a pioneer factor plays an important role in angiogenesis, particularly if we can target it and inhibit it (124-126). Additionally, it would be interesting to determine whether or not BRG1 or another chromatin remodeler also forms a complex with ETV2 in this context.

Conclusion

More studies are coming out these days claiming they have characterized a novel pioneer transcription factor and we only expect this number to rise, just like it happened with master regulators ever since the term was initially coined (26). This is in part due to the advances in molecular biology that make it easier to characterize cells at the single cell level during embryogenesis, allow us to identify DNA binding sites for transcription factors (TF) and more importantly allow us to define the chromatin dynamics surrounding the DNA binding sites of such TFs. Particularly, with the development of the Assay for Transposase-Accessible Chromatin followed by sequencing (ATAC-seq), you can define the chromatin landscape of differentiating or reprogramming cells with very few cells (50,000 cells or less) and claim that a TF is a pioneer factor (127, 128). While ATAC-seq characterization of cell populations can be insightful, we caution the reader that a more in-depth characterization is needed when characterizing a

pioneer factor. To do this, three criteria need to be fulfilled 1) pioneer factors need to bind nucleosomal DNA in vivo (sequencing) and in vitro (nucleosomal binding assay), 2) pioneer factors need to promote chromatin remodeling around DNA binding sites by themselves or by interacting with chromatin remodelers and 3) pioneer factors need to enact global epigenetic changes and recruit other co-factors that further promote the development or reprogramming of a cellular lineage.

More work will be needed within the cardiovascular field to further identify pioneer factors that regulate the different cellular lineages and structures (i.e. primary versus secondary heart field) that are present within it. For example, while ETV2 sits at the top of the endothelial lineage developmental hierarchy, ISL1 and GATA4 are two of many regulators of the cardiac muscle lineage with very specific functions. We predict that multiple pioneer factors, and not only one like ETV2 for endothelium, will regulate cardiac muscle development and reprogramming. Another cellular lineage that we did not discuss here is smooth muscle, where no pioneer factor has yet to be identified. Pioneer factors can be powerful tools for the development of regenerative therapies whose goal is to generate mature and functional cell lineages. Understanding the molecular mechanisms that drive lineage development by these and other pioneer factors within the cardiovascular lineage will be instrumental because coupling these pioneer factors along with chromatin remodelers and downstream targets genes can amplify the molecular effect needed to better develop regenerative therapies for cardiovascular disease.

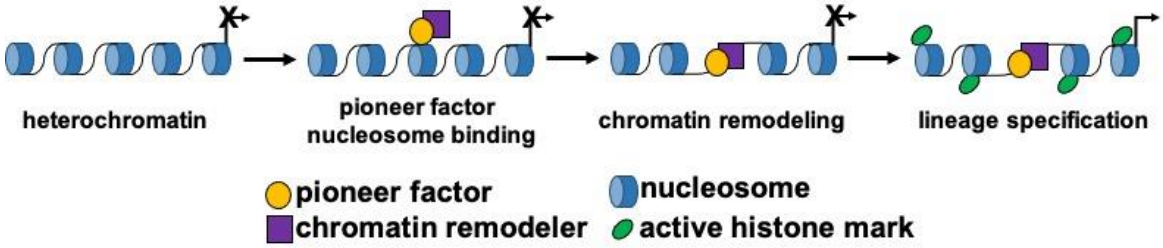


Figure 1. Pioneer factors drive lineage specification. Schematic model depicting the function of pioneer transcription factors during cellular lineage specification. Pioneer factors first bind to nucleosomal DNA and then remodel chromatin by themselves or by recruiting a chromatin remodeling factor. These steps lead to the activation of gene expression and changes to the epigenetic landscape surrounding the DNA binding sites of the pioneer factor.

CHAPTER 3: ETV2 FUNCTIONS AS A PIONEER FACTOR TO REGULATE AND REPROGRAM THE ENDOTHELIAL LINEAGE

Introduction

Ischemic heart and vascular diseases are common, deadly, and result in considerable morbidity and mortality (129). Progression of these diseases results in end stage heart failure, peripheral arterial diseases or limb amputation. Moreover, solid organ cancer progression is typically associated with angiogenesis and new treatment strategies have focused on the repression of vascular growth (130). Collectively, the management of these diseases requires a deeper understanding of the mechanisms that govern endothelial and vascular lineage determination (131-133).

During development, pioneer transcription factors have the essential and unique role of opening new regulatory chromatin landscapes on genomic DNA (38). Pioneer factors can recognize their target DNA sequences in compacted chromatin, recruit chromatin remodelers and trigger the relaxation of the adjacent chromatin landscape to provide accessibility to non-pioneer transcription factors (37, 44). Through the relaxation of the chromatin, pioneer factors enable other transcription factors to sequentially access their binding motifs and collectively activate gene expression resulting in the reprogramming of cell fate (39). A limited number of pioneer factors have been defined and all of these factors have been shown to have important roles in various biological processes. Examples of pioneer factors include the pluripotent factors OSK (OCT4, SOX2, and KLF4) for iPSC reprogramming, FOXA for hepatocyte reprogramming, PAX7 for pituitary development and ASCL1 for neurogenesis (1, 76, 134).

Studies from our laboratory and others have identified ETV2 as an essential transcription factor for the development of cardiac, endothelial and hematopoietic lineages (49, 50, 70, 79, 110-113, 135-137). Using the 3.9kb *Etv2-Cre* and *Rosa-EYFP* reporter alleles in engineered mouse models, *Etv2*-EYFP progenitors have been shown to give rise to endocardial/endothelial and hematopoietic lineages (138). Furthermore, when crossed into an *Etv2* mutant background, the *Etv2*-EYFP progenitor cells gave rise to cardiomyocytes in the absence of ETV2*,* showing that ETV2 elicits a cell fate choice. *Etv2* has been shown to be responsive to BMP, WNT and NOTCH signaling pathways and synergizes with FOXC2 to regulate the endothelial program by directly targeting *Lmo2, Cdh5*, *Tie2* and *Flk1* (48, 112, 139). Collectively, these and other studies support the role for ETV2 to function as a master regulator and also support the notion that ETV2 may function as a pioneer factor in endothelial development (55).

The mammalian SWI/SNF-related chromatin-remodeling complex consists of one of two different ATPases (brahma or BRM vs. brahma related gene-1 or SMARCA4/BRG1) along with Brg1 associated factor subunits (BAFs) (82). Using gene disruption strategies, BRM has been shown to be non-essential for hematoendothelial development (81, 140). In contrast, the global knockout of *Brg1* resulted in early lethality prior to embryo implantation (81). In addition, the conditional knockout of *Brg1* using the *Tie2-Cre* transgenic mouse line resulted in

midgestational lethality with perturbed hematoendothelial development (141). Previous studies have also established a mechanism whereby pioneer factors function in the recruitment of BRG1, which facilitates the chromatin relaxation to promote transcription factor binding and potentiates transcriptional activation of gene expression (85). Therefore, our goal was to define the role of ETV2 as a pioneer factor and its interaction with BRG1 during cellular reprogramming and differentiation.

Results

ETV2 reprograms fibroblasts into endothelial cells

We isolated E13.5 embryonic fibroblasts from the *iHA-Etv2* mouse that inducibly overexpresses ETV2 upon administration of doxycycline (Dox). These cells uniformly expressed fibroblast markers (THY1.2, CD44 and CD29) and lacked hematoendothelial (HE) marker expression by FACS in the absence of Dox (Extended Data Fig. 1a-b). Using western blot analysis, we further demonstrated that ETV2 was expressed within 3 hrs post-Dox treatment (Extended Data Fig. 1c). By 24 hrs, ETV2 overexpression resulted in more than a 50-fold increase in cells expressing FLK1/TIE2 by FACS (Extended Data Fig. 1d-g) and these endothelial cells were characterized by Ac-LDL uptake, endothelial tube assays and induction of downstream endothelial target gene expression (Figure 1g, Extended Data Fig. 1i-m).

To further investigate the molecular dynamics of ETV2 inducible reprogramming, we captured and sequenced 13,677 cells from 24 hrs, 48 hrs and 7 days post-induction of ETV2 in mouse embryonic fibroblasts (MEFs). We also performed single cell RNA-seq of 948 undifferentiated MEFs as well as 827 sorted FLK1^{+} cells from day 7 reprogrammed cells, and identified seven distinct cell clusters (Figure 1a, 1c-e). We noted that during the first 48 hrs of reprogramming, even though *Etv2* was significantly up-regulated as early as 24 hrs, endothelial markers such as *Flk1*, *Lmo2*, *Emcn*, *Cdh5* and *Sox18* were significantly activated only in a subpopulation of cells from day 1 and day 2 (clusters #1 and #3) but not in other cell clusters (clusters #2 and #4) (Figure 1d, Extended Data Fig. 2a-b). Pathway analysis associated the increased cell cycle activity and expression of chromatin organization related genes in cluster #1 with the initiation of ETV2 induced reprogramming, as observed in ASCL1 driven neuronal reprogramming (Extended Data Fig. 2c) (142). At day 7 of reprogramming, FACS analysis showed that 17% of the cells were FLK1^{+} cells (Extended Data Fig. 1d-e). The scRNA-seq showed that the FLK1^{+} cells at day 7 formed a unique cell population (cluster #7). Endothelial marker genes such as *Lmo2* and *Emcn* were robustly expressed in cluster #7 cells, while the fibroblast markers such as *Cd44* and *Fosl1* were down-regulated (Figure 1f). The pathway analysis confirmed that transcripts with functions related to vasculature development were more abundantly expressed in cluster #7 compared with the remaining clusters (Extended Data Fig. 1h). In summary, the scRNA-seq analysis supported the hypothesis that ETV2 overexpression in MEFs activated

the endothelial gene expression network and endothelial tube behavior as a cell biological response.

We identified two subpopulations of reprogrammed MEFs, where one cluster was responsive to ETV2 induction (cluster #1) while the other was not (cluster #2; Extended Data Fig. 2b). The GSEA analysis suggested that the inflammatory signaling pathway (GO:0006954) was significantly enriched in the up-regulated genes in cluster #1 MEFs (GSEA p-value=2.53e-05), which suggested that MEF subpopulations with an elevated inflammatory activity may reprogram better, supporting a role for immune response signaling in cell fate transitions (Extended Data Fig. 2d-g, Supplementary Figure 1) (143-145). Moreover, the immune response-related signaling pathway was significantly up-regulated at D1 post ETV2 induction, and mildly up-regulated in D7 Flk1^{+} cells (Extended Data Fig. 3) (144, 146, 147). We also observed elevated immune response signaling at 3 days of GMT induced cardiac reprogramming, and significantly decreased immune response signaling at later stages of OSKM induced iPSC reprogramming (day 16) and Ascl1 induced neural reprograming (day 22) (Supplementary Figure 2 and 3). These analyses suggest the need for future studies aimed at establishing a comprehensive understanding of the immune response pathways at different stages of cellular reprogramming directed towards different lineages.

The embryoid body formation from embryonic stem (ES) cells has been used extensively for studying the role of ETV2 in endothelial development (50). To define the molecular programs commonly or differentially responsive to ETV2 induction in EBs and MEFs, we induced ETV2 at day 2 of EB differentiation (prior to the onset of endogenous ETV2 expression) and performed bulk RNA-seq of sorted FLK1^{+} cells at 12 hrs post-induction (Figure 1b and Extended Data Fig. 4). Compared with the gene expression profiles in D2.5 EBs without induction, we identified 554 and 1,507 genes were up- and down-regulated in both EBs and MEFs, respectively (Extended Data Fig. 5a-b and Supplementary Figure 5). As expected, the commonly up-regulated genes were closely related to vasculature development (Extended Data Fig. 5c-f).

To examine the chromatin accessibility changes following ETV2 overexpression, we performed ATAC-seq at 24 hrs, 48 hrs and 7 days post-induction in MEFs and 12 hrs post-induction in EBs (Figure 1a-b). The transcription factor (TF) associated chromatin accessibility analysis (chromVAR) suggested that although significant batch effects existed between MEFs and EBs (PC1 in Figure 1h), FLK1^{+} cells from day 7 MEFs and day 2.5 EBs post-ETV2 induction shared a common global chromatin accessibility pattern (PC2 in Figure 1h). This included 113 and 246 TF genes whose chromatin accessibility was determined by ATAC-seq, which were commonly increased and decreased in MEFs and EBs, respectively (Extended Data Fig. 6a-f). By integrating the RNA-seq and ATAC-seq datasets from MEFs and EBs, we identified 13 TFs whose expression levels

and TF associated accessibility were consistently increased in FLK1$^+$ cells in both MEFs and EBs, and 18 TFs, which were decreased in FLK1$^+$ cells (Figure 1i and Extended Data Fig. 6g-h). Interestingly, mesodermal factors such as *Msx2*, *Eomes* and *Foxk2* were among the genes that were consistently down-regulated in both MEFs and EBs, suggesting the potential role of ETV2 as a suppressor of non-endothelial lineage development. We found that the TF motifs such as NRF1 and MYC, but not ETV2, were significantly enriched in the 5kb region surrounding the transcription start sites of the commonly down-regulated genes in both MEFs and EBs, suggesting that the down-regulation of these mesodermal factors were indirectly regulated by ETV2 (Extended Data Fig. 6g-i). Collectively, these results demonstrated that induction of ETV2 promoted the specification and cell differentiation towards the endothelial lineage and the suppression of non-endothelial lineages (Extended Data Fig. 5e-f, Supplementary Figure 5), in two distinct cellular environments (MEFs and EBs), supporting the notion that ETV2 functioned as a master regulator to drive endothelial lineage development.

ETV2 targets nucleosomes during reprogramming

To investigate how ETV2 binding reshapes the genomic accessibility landscape and drives endothelial lineage differentiation in different cellular environments, we performed ETV2 ChIP-seq at 24 hrs, 48 hrs and 7 days post-ETV2 induction in reprogrammed MEFs, and ETV2 ChIP-seq at 3 hrs and 12 hrs post-ETV2 induction in developing EBs at day 2. The initial ETV2 binding events were captured using ETV2 ChIP-seq at 24 hrs in MEFs and 3 hrs in EBs, resulting in

131,001 and 18,024 peaks, respectively (Figure 2a). The 11,751 common ETV2 peaks overlapped between MEFs and EBs and represented the majority (65.2%) of the ETV2 peaks in EBs, and were located predominantly in the promoter region. In contrast, the EB and MEF specific ETV2 peaks were more distributed at the distal intergenic regions (Figure 2b). One of the key features of pioneer factors is their capacity to target nucleosomes (44). To test whether ETV2 was able to target the nucleosome, we first examined the nucleosome profiles in undifferentiated MEFs using published MNase-seq datasets (148). We divided the ETV2 ChIP-seq peaks from MEFs into four quartiles based on the mean MNase-seq signals of the central 200-bp region, and we used the first and the fourth quartile to represent the nucleosome free region (NFR) and the nucleosome occupied region (148), respectively (Figure 2c and Extended Data Fig. 7a). The NOR quartile represented the ETV2 peaks with their summits located at the nucleosome centers, suggesting that ETV2, like other reported pioneer factors can target nucleosomes (44). Similarly, we divided the ETV2 ChIP-seq peaks at 3 and 12 hrs post-induction EBs into NFR (5,291 peaks) and NOR (8,843 peaks) groups according to the local V-plot and fragment size profiles of ATAC-seq day 2.5 EBs without ETV2 induction (Extended Data Fig. 7b-d). We found that similar to the MEF reprogramming, ETV2 also targeted nucleosome centers at day 2.5 EB differentiation (Figure 2d). Our results using MEFs and EBs suggested that ETV2 was intrinsically able to target nucleosomal sites independent of its cellular context.

We performed NOMe-seq to analyze the nucleosome profiles of control and D1 MEFs post ETV2 induction. We focused on ~44k ETV2 ChIP-seq peaks at D1 of reprogramming that also contained the canonical ETV2 motif at the center. We used the GCH sites (GCA/GCT/GCC) in the NOMe-seq to determine the chromatin accessibility and calculated ratio of non-methylated cytosines (C) in GC dinucleotides of every GCH site (149). We identified 5,320 ETV2 binding sites that were statistically determined as NORs in MEFs. Among them, 4,744 (89.1%) became nucleosome-free while 576 (10.9%) remained as NORs at D1 of reprogramming (Figure 2e). These results suggested that the majority of the ETV2 binding sites that were occupied by nucleosomes in MEFs became nucleosome free following ETV2 induction at D1, supporting our main conclusion that ETV2 as a pioneer factor targets nucleosomes and relaxes chromatin.

Sequence motif analysis identified a common GGAAAT motif that was significantly more enriched in NFRs compared with the NORs in both MEFs and EBs (Fisher's adjusted p-value=6.0E-05 and 6.8E-5) (Figure 2f and Extended Data Fig. 7e). This NFR motif has additional terminal "AT" nucleotides compared with the canonical ETV2 motif. These findings suggested that ETV2, similar to other pioneer factors, such as OCT4, SOX2 and KLF4, was able to target nucleosome-enriched sites using partial or degenerate motifs, and targeted their full canonical motif in nucleosome-depleted sites (44).

Moreover, the ETV2-targeted nucleosome regions were characterized by low BRG1 ChIP-seq signals at 24 hrs post-induction in MEFs and 3 hrs post-induction in EBs, indicating that the initial recognition of the nucleosome does not require BRG1 (Figure 2c-d). Interestingly, the ETV2-targeted nucleosomes had a distinct H3K27ac surrounding pattern in MEFs and EBs: ETV2 targeted H3K27ac-depleted nucleosomes in MEFs and H3K27ac-enriched ones in EBs, which suggested that ETV2 can target nucleosomes in different H3K27ac-contexts.

ETV2 is able to physically bind nucleosomes

Since nucleosome binding is a hallmark ability of pioneer factors, we assessed whether ETV2 was able to bind nucleosomes using *in vitro* binding assays. We identified a region upstream of *Lmo2* that was enriched for ETV2 ChIP-seq peaks and binding motifs as well as MNase-seq and ATAC-seq peaks, suggesting both ETV2 binding and nucleosome occupancy (Figure 2g). We generated a 159-bp PCR product of the endogenous NOR containing ETV2 binding motifs (Cy5-labeled-*Lmo2*-DNA*).* As expected, recombinant ETV2 protein could bind to Cy5-labeled *Lmo2*-DNA probes without any histones (Figure 2h). Upon assembling the Lmo2 sequence into nucleosomes (Figure 2i) by salt gradient dilution with purified recombinant mouse histones, we observed that ETV2 could bind the *Lmo2* nucleosomes comparable to the free DNA (Figure 2i). Altogether, these data demonstrated that ETV2 physically binds nucleosomes *in vitro*, further supporting our hypothesis that it functions as a pioneer factor.

ETV2 recruits BRG1 for chromatin remodeling

To mechanistically define the secondary consequences of how ETV2 targets the genome to drive endothelial lineage development during MEF reprogramming and ES/EB differentiation, we performed ETV2 ChIP-seq at 1 (D1), 2 (D2) and 7 (D7) days post-ETV2 induction in reprogrammed MEFs and ETV2 ChIP-seq at 3 hrs and 12 hrs post-induction in differentiated EBs. In total, we identified 154,468 and 19,651 non-overlapping ETV2 peaks in MEFs and EBs, respectively. Similar to the OSK-binding sites during fibroblast reprogramming to iPSCs (148), more than 80% of the unique ETV2 peaks were only present at early stages (D1 for MEFs and 3 hrs for EBs) but not sustained during later stages (D2, D7 for MEFs and 12 hrs for EBs; Extended Data Fig. 7f-g). We divided the ETV2 peaks into "early", "late" and "sustained" groups according to whether ETV2 peaks were present in the early stage, the late stage or throughout the entire MEF reprogramming or ES/EB differentiation period (Figure 3a-b). We found that for both EB and MEFs, more than 40% of the "sustained" ETV2 peaks and more than 55% of the "late" ETV2 peaks contained partial ETV2 binding sites within 50 bp of the peak summits, and that the binding sites were closer to the peak summit. In comparison to EBs, only 13% of the "early" ETV2 peaks had partial ETV2 motifs, suggesting the early ETV2 binding events in MEFs were dominated by pioneer factor scanning of chromatin shown to be important for this class of TFs to search for their genomic targets (37, 150-152). We hypothesized that the "late" and the "sustained" ETV2 peaks would be

important for ETV2 to activate downstream endothelial genes. We then identified BRG1 as one of the top ATP dependent chromatin remodeling enzymes to interact with ETV2 from mass spectrometry analysis of the protein complex pulled down using an ETV2 antibody (Figure 3c). Using co-immunoprecipitation in EBs and GST pulldown assays in vitro, we further validated that ETV2 physically and directly interacted with BRG1 (Figure 3d-f). BRG1 ChIP-seq results showed that in both EBs and MEFs, the "late" and "sustained" ETV2 peaks were coupled with increasing levels of BRG1 in the late stages of EB differentiation and MEF reprogramming, while the "sustained" ETV2 peaks were associated with detectable levels of BRG1 at the early stage (Figure 3g-h). Re-ChIP assays were performed using EBs which validated the co-occupancy of ETV2 and BRG1 to regulatory regions of known ETV2 downstream targets such as *Flt1*, *Lmo2* and *Tcf12* (Figure 3i). Note that during both MEF reprogramming and EB differentiation, the establishment of "late" and "sustained" ETV2 peaks were also coupled with increasing H3K27ac enrichment, characterized by a "dip" at the summit of ETV2 peaks at late stage. The establishment of H3K27ac modification surrounding the pioneer factor bound sites was also found in ASCL1 driven neural differentiation and OSKM driven MEF reprogramming (148, 153). We found there was significant overlap between the "late" ETV2 peaks in EBs and MEFs, as well as the nearby genes (Figure 3j). Moreover, the "late" ETV2 peaks were located near a higher proportion of endothelial genes, in contrast to the "early" ETV2 peaks (Extended Data Fig. 7h). These results suggested that in both MEFs and EBs, the "late" and "sustained" ETV2 bindings recruits BRG1,

promotes H3K27ac deposition and chromatin remodeling to activate downstream endothelial programs.

Phased nucleosomes are established surrounding ETV2 peaks

To better characterize the nucleosome arrays from the ATAC-seq data, we used a variational autoencoder (VAE) model to learn the latent representation of ATAC-seq V-plots. We applied the VAE to 18,214 ETV2 ChIP-seq peaks (with canonical ETV2 motifs that did not overlap with promoters) during MEF reprogramming and inferred the fragment size distribution of 100-bp region over the summit of the ChIP-seq peaks, and identified six clusters of V-plot according to the central fragment size distribution (Figure 4a). The six clusters included three types of V-plots where the central ETV2 sites were nucleosome free (C1, C3 and C4), and three types of V-plots where the central ETV2 sites were nucleosome occupied (C2, C5, and C6), represented by the aggregated V-plot and NucleoATAC (154) profiles from each cluster (Figure 4b-4e). Six clusters of V-plots also showed a distinct nucleosome array pattern surrounding the ETV2 summit. Next, we examined whether the "early", "late" and "sustained" ETV2 peaks have different associated V-plot patterns. While the early ETV2 peaks have a relatively similar distribution of V-plot clusters compared with the background peaks, the late and sustained ETV2 peaks have a significantly increased proportion of C1 (NFR1) and C2 (Nuc1) V-plot patterns during reprogramming (Figure 4f). These results demonstrated that ETV2 was able to alter the chromatin structure by opening and closing the chromatin. Specifically,

we found that a majority of the newly obtained C1 V-plot cluster with phased nucleosomes, came from nucleosome occupied C2 cluster, and C3, a V-plot cluster where the central ETV2 motif has less phased flanking nucleosomes (Figure 4g).

Like ASCL1 induced neural reprogramming, we also found nucleosomal arrays that were generated through positioning of nucleosomes surrounding the ETV2 binding motifs in the FLK1^{+} cells in both MEFs and EBs (Figure 4h-4i) (155). These results demonstrated that ETV2 was able to open the nucleosome occupied by chromatin and establish phased nucleosomes in the flanking region[49].

ETV2 requires BRG1 for endothelial cell reprogramming

Next, we knocked down *Brg1* in *iHA-Etv2* MEFs using shRNAs 48 hrs prior to the induction of ETV2, and continued ETV2-induction at 2-, 4- and 7-days post-induction (Extended Data Fig. 8a). We found that the BRG1 protein levels and the percent of FLK1^{+} cells were significantly reduced at 1, 2- and 7-days post-induction, compared with control ETV2-induced MEF reprogramming (Figure 5a-b and Extended Data Fig. 8b). Additionally, the expression levels of downstream endothelial genes were also significantly reduced at 7 days post-induction when *Brg1* was knocked down, suggesting that BRG1 was required for ETV2 activation of downstream endothelial genes (Extended Data Fig. 8c-l). We also examined *Chd8, Dek* and *Znhit* as other potential chromatin-remodeling

enzymes that were expressed in *iHA-Etv2* MEFs (Figure 3c and Extended Data Fig. 9a-c) and interacted with ETV2 from our mass spectrometry analysis, and did not find their knockdown of affected endothelial cell reprogramming (Extended Data Fig. 9d-h) unlike our studies with *Brg1* (Figure 5a-b and Extended Data Fig. 8).

scRNA-seq data showed that knocking down *Brg1* in MEFs prior to reprogramming significantly impacted cell proliferation (156). After correcting for these cell cycle effects, the *Brg1* KD MEFs and control MEFs clustered together (Supplementary Figure 6). The scRNA-seq and gene expression analysis (Figure 5c-f and Extended Data Fig. 8c-l) at D7 post-ETV2 induction in *Brg1* KD MEFs showed reduced expression levels of key endothelial genes such as *Flk1/Kdr*, while *Etv2* expression levels remained unchanged. Additionally, the FLK1^{+} cells in *Brg1* KD MEFs and control MEFs were separate populations following ETV2 induction (Figure 5c-f). These results suggested that BRG1 is critical for ETV2 to activate downstream endothelial genes during endothelial cell reprogramming.

ETV2 requires BRG1 to relax chromatin and recruit co-factors

To further examine the global impact of *Brg1* knockdown (KD) in the pioneer function of ETV2, we performed ATAC-seq before ETV2-induction (D0), 1 day (D1) and 7 days (D7) post-induction of ETV2 following *Brg1* KD in MEFs. We found that 81.3% of "sustained" ETV2 ChIP-seq peaks that were present at both D1 and D7 post-induction of ETV2 in control MEFs had significantly reduced

chromatin accessibility at D7 in *Brg1* KD MEFs (Extended Data Fig. 10a). Knockdown of *Brg1* affected sustained ETV2 binding events that targeted both open and closed chromatin in undifferentiated MEFs and reduced the sustained ETV2 binding at D7 in *Brg1* KD MEFs (Figure 6a and Extended Data Fig. 10b-c) supporting the notion that BRG1 is important for the stabilization of ETV2 binding in a comparable fashion to the BRG1 mediated stabilization of OCT4 binding as previously described (85). ChromVAR showed that reprogrammed *Brg1* KD MEFs at D7 had significantly less partial or canonical ETV2 motifs containing ATAC-seq reads, compared to control MEFs and FLK1^{+} sorted cells (Figure 6b). Moreover, the chromatin at ETV2 binding sites affected by the knockdown of *Brg1* were more likely located at the distal intergenic regions (Figure 6c). These results demonstrate that ETV2 acts as a pioneer factor in targeting closed chromatin, but requires BRG1 for subsequent steps of chromatin opening and gene activation. To further examine how knocking down *Brg1* affected the chromatin accessibility surrounding ETV2 binding sites during the first 24 hrs following ETV2-induction, we identified 1,204 ETV2 binding sites that were closed in control MEFs and became open at D1 post-ETV2 induction in control MEFs. The aggregated V-plot showed that these ETV2 binding sites were still closed and occupied by nucleosomes at D1 post-ETV2 induction in *Brg1* KD MEFs (Figure 6d). These findings further supported the notion that ETV2 requires BRG1 to promote chromatin decompaction and fulfill its pioneer function.

To further investigate the importance of BRG1 for ETV2 to function as a pioneer factor, we performed a conditional deletion (KO) of *Brg1* using *Brg1*$^{f/f}$*;ActinCreER* ESCs treated with 4-hydroxytamoxifen (4-OHT) during mesodermal culture conditions (Figure 7a-b) (157). Flow cytometry analysis at D4 of EB differentiation showed a significant decrease in HE lineage development in *Brg1* KO cells compared to control (Figure 7c). Furthermore, analysis of mature endothelial cell surface markers such as TIE2 demonstrated that the loss of *Brg1* significantly affected endothelial lineage development (Figure 7d). This latter finding was further reinforced with sprouting assays of *Brg1*$^{f/f}$*;Actin-CreER* differentiating EBs that showed a significant reduction of sprouts in cells lacking *Brg1 (*Figure 7e). We then performed ATAC-seq analysis using control and *Brg1* KO EBs at D4 and used chromVAR to identify TF motifs and examine whether chromatin accessibility was significantly different between the control and *Brg1* KO samples. Similar to the *Brg1* KD ATAC-seq in MEF reprogramming (Figure 5b), we found that the loss of *Brg1* significantly reduced the global ETV2-associated chromatin accessibility (Figure 7f). Furthermore, examination of 7,807 sustained ETV2 ChIP-seq peaks in EBs (ETV2 peaks that were present at 3h and 12h post induction) revealed significantly reduced chromatin accessibility (Figure 7g). These results demonstrated that BRG1 was critical for ETV2 to perform its pioneer function and relax the chromatin landscape during EB differentiation.

To understand whether BRG1 was important for ETV2 to relax chromatin and allow other factors to bind and drive endothelial lineage development, we examined the V-plots of ELK3, a downstream target of ETV2. ELK3 motif centric genomic regions showed different V-plots between control and *Brg1 KO* D4 EBs, demonstrating an increase of mono-nucleosome reads at the ELK3 binding sites in *Brg1 KO* D4 EBs, suggesting that the absence of BRG1 may disrupt nucleosomal relaxation at ELK3 binding sites (Figure 7h). To test this hypothesis, we performed ChIP qPCR analysis using ELK3 and found a diminished recruitment to ETV2-BRG1 co-occupied DNA binding sites in *Brg1 KO* EBs (Figure 7i). This highlighted the capacity of ETV2 as a pioneer factor to recruit co-factors during endothelial lineage specification in collaboration with BRG1. Overall, these results suggested that ETV2 recruits BRG1 to remodel closed chromatin, which is critical for ETV2 to recruit other co-factors to activate the downstream endothelial gene network and fulfill its role as a pioneer factor (Figure 7j).

Discussion

In the present study, we used two distinct biological systems: ES/EB differentiation and MEF reprogramming, to define the role of ETV2 as a pioneer factor that regulates endothelial fate and development. To our knowledge, no other study has used these diverse systems together to define pioneer factors and reprogramming capabilities. Even though these two model systems have very different global expression, chromatin accessibility and epigenetic profiles,

we found similar molecular programs and downstream genes that were regulated following ETV2 induction. We showed that, during endothelial cell reprogramming (MEFs) and differentiation (EBs), ETV2 targeted closed chromatin domains independent of BRG1, confirming the key characteristic of a pioneer factor. Then ETV2 recruited BRG1 and functioned together as a complex to relax closed chromatin and recruit other co-factors such as ELK3. Similar to OCT4 in OSKM induced reprogramming, the late binding events were coupled with increased BRG1 occupancy (148). Additionally, ATAC-seq analysis following the knockdown of *Brg1* in MEFs or the knockout of *Brg1* in EBs showed that the maintenance of the majority of the sustained open chromatin states elicited by ETV2 binding required BRG1. While our results demonstrated a critical role for BRG1 in the pioneer function of ETV2, they do not rule out the involvement of other chromatin modifying factors during endothelial lineage specification. Furthermore, since cellular reprogramming requires complex signaling transduction and collaborative integration, other pathways such as inflammatory signaling may also play an important role in endothelial cell reprogramming and lineage specification. We identified two clusters of MEFs with distinct immunogenic profiles that responded to ETV2 overexpression differently, highlighting the important role of the immune system in regulating transcriptional, epigenetic and phenotypic changes during cellular reprogramming. These mechanisms uncovering the role of ETV2 as a pioneer factor for the endothelial lineage further enhance our understanding of endothelial and vascular development, regeneration and angiogenesis. Our studies the definition of

pioneer factors and regulatory pathways that govern the specification and differentiation of endothelial progenitors will serve as a platform for designing pro- and anti-angiogenic strategies in the setting of cardiovascular disease and solid tumors, respectively (130).

Methods

All animal handling, experimental procedures and ethical guidelines were approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Minnesota.

Mouse embryonic stem cell lines. *iHA-Etv2* ESCs (113) and *Brg1*$^{f/f}$*;Actin-CreER* ESCs were cultured in media containing 15% fetal bovine serum (FBS), 2 mM Glutamax, 1X penicillin/streptomycin, 0.1 mM β-mercaptoethanol, and 1,000 U/mL LIF (Millipore), at 37°C in 5% CO2 together with irradiated mouse embryonic fibroblasts as the feeder layer. We differentiated our *iHA-Etv2* ESCs into embryoid bodies (EBs) using mesodermal conditions as previously described (113). Briefly, ESCs were dissociated into single cells using 0.25% trypsin and plated for 50 minutes to remove fibroblast cells from the feeder layer (de-MEF). Following de-MEF, ESCs were differentiated using the shaking method in media containing 15% FBS, 1X penicillin / streptomycin, 1X GlutaMAX, 50 µg/ml Fe-saturated transferrin, 450 mM monothioglycerol, 50 µg/ml ascorbic acid in IMDM. *iHA-Etv2* EBs were treated with doxycycline (0.5 µg/ml) at day 2 of differentiation to overexpress ETV2 and harvested following 3 and 12 hrs of doxycycline

induction of ETV2. To conditionally delete *Brg1*, *Brg1*$^{fl/fl}$;*Actin-CreER* ESCs (82, 157), shaking EB cultures were treated with 200 nM 4-hydroxytamoxifen (4-OHT) every 2 days from D0-D4.

***iHA-Etv2* MEF cell line.** *iHA-Etv2* MEFs were isolated from E13.5 embryos using a previously described methodology and their purity was assessed by flow cytometry using fibroblast markers (158). *iHA-Etv2* MEFs were cultured in media containing 10% fetal bovine serum (FBS), 2 mM glutamine, 1X penicillin / streptomycin and 1X non-essential amino acids at 37°C in 5% CO2. Doxycycline (1 μg/ml) was added to *iHA-Etv2* MEFs in order to reprogram fibroblasts to endothelial cells. Media containing doxycycline was changed every 48 hrs and reprogrammed cells were harvested 24 hrs, 48 hrs and 7 days following the addition of doxycycline for analysis.

Fluorescent-Activated Cell Sorting (FACS) analysis. Harvesting and staining of *iHA-Etv2* MEFs and EBs was performed as previously described (159) and analyzed or sorted using a FACSAria (BD) machine. The antibodies used for FACS include: Flk1-APC (1:200, Cat# 560070, Lot# 8298981), platelet-derived growth factor alpha receptor-a (Pdgfra)-PE (1:1000, Cat# 4315814, Lot# 2049418), Flk1-PE (1:300, Cat# 12582181, Lot# 1930444), Tie2-APC (1:400, Cat# 124010, Lot# B231548), CD31-FITC (1:1000, Cat# 11-0311-82, Lot# 4291915), CD41-FITC (1:1000, Cat# 11-0411-82, Lot# 4295908), CD45-FITC (1:1000, Cat# 553079, Lot# 7096563), CD144-AF647 (1:1000, Cat# 51-1441-80,

Lot# E028392), Tie2-PE (1:1000, Cat# 12-5987-81, Lot# 4339859), Thy1.2-APC (1:1000, Cat# 561974, Lot# 5208833), CD44-PE (1:1000, Cat# 553134, Lot# 6308790) and CD29-APC (1:1000, Cat# 17-0291-80, Lot# E07122-1631).. Propidium iodide (1:1000, Cat# 1423090, Lot# 1325708) was used to gate for live cells during FACS analysis.

RNA isolation and quantitative PCR (qPCR) analysis. Total RNA was isolated from iHA-Etv2 MEFs and EBs using the RNeasy kit (Qiagen, Cat# 74104) according to the manufacturer's protocol. Briefly, cells were lysed in RLT-lysis buffer, followed by a column-based purification process and on-column DNA digestion. Complementary DNA (cDNA) was synthesized using the SuperScript IV VILO kit (Thermo Fisher Scientific, Cat# 11756050) according to the manufacturer's protocol. Quantitative polymerase chain reaction (qPCR) was performed with ABI Taqman probe sets. Taqman probes used in this study include VIC-labeled GAPDH:4352339E, FAM-labeled Etv2: mm01176581_g1, Flk1: mm00440099_m1, PECAM (Cd31): mm0124616 7_m1, Tie2: mm01256892_ m1, Cdh5: mm00486938_m1, Lmo2: Mm00493153_m1, Sox18: Mm00656049_gH, Smarca4 (Brg1): Mm01151944_m1, Emcn: Mm00497495_m1, Mmp9: Mm00442991_m1, Hopx: Mm00558630_m1, Otor: Mm00498571_m1, Lax1: Mm00556050_m1, Chd8: Mm01316316_m1, Dek: Mm01351566_m1 and Znhit: Mm01201686_m1 (See Supplementary Table 1).

Western blot analysis. Western blots were performed as described previously (107). Briefly, cell lysates from were obtained control and reprogrammed *iHA-Etv2* MEFs following ETV2 induction (+Dox) at various time points. These were lysed in ice-cold lysis buffer for 30 minutes and centrifuged at 9,300 x g for 10 minutes at 4 °C. Equal amounts of protein were loaded on 10% SDS-polyacrylamide gels. PVDF (polyvinylidene fluoride) membranes were blocked with 5% milk protein and incubated with a rabbit-HA antibody (1:1000, Cat# C29F4, Lot# 1), rabbit-BRG1 antibody (1:1000, Cat# ab110641, Lot# GR3208604-18) and rabbit-GAPDH antibody (1:1000, Cat# D16H11, Lot# 7) overnight at 4 °C. The membrane was subsequently incubated with a goat-anti-rabbit (1:2000, Cat#SC-2004; Lot# G247) HRP (horseradish peroxidase)-conjugated secondary antibody and visualized using the Pico luminescence kit (Invitrogen) according to manufacturer's instructions. The protein bands were visualized and imaged using the Image Lab software.

ETV2 protein expression and purification. 6XHis tagged ETV2 recombinant protein was expressed and purified from BL21 Star (DE3) E. coli cells. Briefly, bacteria were grown in LB medium containing 50 μg/ml kanamycin at 37 degrees with shaking at 200 rpm. When the OD_{600} value reached 0.6-0.8, ETV2 expression was induced using IPTG (0.5 mM) for 4 hrs at 37 degrees. Bacterial cells were lysed using a sonicator (3 sec on, 6 sec off for a total of 3 min at 600 w) and the target protein was purified with 1ml Ni-NTAcolumn. Denaturing buffer (50 mM Tris-HCl, 8M Urea, pH 8.0) was used as the column equilibration buffer

and wash buffer. The target protein was eluted with a stepwise gradient of imidazole. Recombinant ETV2 was refolded by dialysis using PBS, 10% Glycerol, 0.5M L-Arginine, pH 7.4. The dialysis was performed using a 14 kDa cut-off dialysis membrane for 4 hours and changed with fresh buffer for an additional 16 hours. After dialysis, the sample was centrifuged at 15,700 x g for 30 min and filtered through a 0.22 µm filter.

Nucleosome reconstitution and DNA binding reactions. The 160 bp Lmo2 DNA fragment corresponds to the genomic location: mm10-chr2: 103,906,846 – 103,907,005

TGTCTATTTTCACCTCAGGCAAAGGGACAGGTTCCCTCAGGCCACTGGAGCCCACCTCAAGAGTCTCTTTGCCAGGGAAATTCTCCAATGAAGGTTTCTGGTGAGTGGTTTGCCTGGGCTCTGCTGTCAGGGGCCATGATGTGCCTGCTGTTTATGCAAC

The DNA sequence was created by PCR with end-labeled primers (Fwd Primer: TGTTCAATTCAAAATGCCTGTCT; Rev Primer: ACATCATGGCCCCTGACAG). The 160 bp fluorescent-tagged DNA fragments were PCR cleaned using the NEB Monarch PCR & DNA Cleanup Kit (Cat # T1030L). Recombinant human histones H3, H4, H2A, and H2B were individually purified from Rosetta (DE3) pLysS bacteria and assembled into octamers as previously described (160). Nucleosomes were assembled by mixing optimal molar ratios of histone octamers to Cy5-labeled DNA in assembly buffer (10mM HEPES pH 7.5, 1mM EDTA, 0.1 mg/ml BSA, 2-β mercaptoethanol, 2M NaCl). Then, the salt was

slowly removed by dialysis using dialysis buffers with the same composition as the assembly buffer but different salt concentrations. First, 2M NaCl Dialysis buffer was slowly diluted overnight with 0M NaCl buffer to a final concentration of 600mM NaCl. Then, the reaction was incubated on 10mM NaCl dialysis buffer for 2h twice. Nucleosomes were purified by ultracentrifugation (147,215 x g for 18 h at 4°C) in centrifugation buffer (50mM HEPES pH 7.5, 1mM EDTA, 0.03mg/mL 2-β mercaptoethanol) containing a glycerol gradient ranging from 10% to 40%. The fractions containing nucleosomes were then pooled, the glycerol was removed by dialysis, and the final sample was concentrated by filtration. Quantification was done by denaturing the nucleosomes in a 1% SDS solution for 10min at 65ºC; then, nucleosomal DNA was quantified using a QuBit 2.0. The EMSAs were performed by mixing purified ETV2 and 1nM of the reconstituted nucleosomes in a binding solution (10 mM Tris-HCl pH7.5, 1 mM MgCl2, 10 mM ZnCl2, 1 mM DTT, 50 mM KCl, 3 mg/ml BSA, 5% Glycerol) and incubating the binding reaction at room temperature for 30min. Then, samples were immediately loaded and ran into a 5% Native PAGE gel with 0.5X TBE running buffer at 100V. Cy5-labeled bands were observed using an Amersham Typhoon RGB Biomolecular Imager (161).

Single cell barcoding, library preparation, and sequencing. Single cell RNA sequencing (scRNA seq) was performed on *iHA-Etv2* MEF reprogrammed cells with and without *Brg1* shRNA knockdown using the previously described protocol (162). Briefly, single cells from reprogrammed D1, D2, D7 and D7-Flk1^{+} sorted

iHA-Etv2 MEFs were barcoded using the 10x Chromium Single Cell platform, and cDNA libraries were prepared according to the manufacturer's protocol (Single Cell 30v3, 10x Genomics, USA) at the University of Minnesota Genomics Center. Final libraries were analyzed on an Agilent Bioanalyzer High Sensitivity DNA chip for qualitative control purposes. cDNA libraries were then sequenced on a NovaSeq S4 Illumina platform (2x150 bp) aiming for 50,000 reads per cell.

Immunoprecipitation and mass spectrometry analysis. Isolation of the ETV2 complex was performed using the iHA-Etv2 mES cells. Doxycycline was added on day 3 of differentiation at a final concentration of 1µg/ml. Day4 EBs were disaggregated using trypsin and gentle shaking in a 37°C water bath followed by inactivation with FBS and gentle resuspension with a pipette. After centrifuge (5 min at 500g), cell pellets were washed with ice cold PBS and incubated 30 min in 10 Packed Cell Volumes (PCV) of ice-cold PBS supplemented with 5µg/ml Digitonin (Sigma-Aldrich). Nuclear extraction was performed as described earlier (163, 164). Nuclear extract was transferred into 3ml Slide-A-Lyzer® G2 Dialysis Cassette – cutoff 7000 MWCO (Thermoscientific) and dialyzed overnight at +4°C with gentle stirring using Dialysis buffer (10mM TRIS HCl pH 7.3, 100mM NaCl, 1.5mM $MgCl_2$, 0.1mM EDTA, 10% Glycerol). Nuclear extracts were recovered, centrifuged 30 min at +4°C 16000g to remove precipitated proteins and precleared 2h with 100µl of Protein G dynabeads. 3 mg of nuclear extract was then incubated with either 15µg of anti-ETV2 antibody (Abcam, ab181847) or control mouse IgG overnight at 4°C with gentle rotation followed by incubation

with 105µl of Protein G dynabeads for 2 hours at 4°C. The beads were then washed five times with PBS containing 0.1% Tween-20. Proteins bound to the dynabeads were eluted in LDS sample buffer (Invitrogen). Eluents from the immunoprecipitation were analyzed by Western blot using HA antibody. Pulled down protein were separated using a 4-15% gradient polyacrylamide gel (Biorad) and stained using the Pierce silver stain for Mass spectrometry (Thermo 24600). The lanes from the gels were cut out from the ETV2 and control IgG pulldown lanes individually and submitted for mass spectrometry analysis at the Taplin Mass Spectrometry facility (Harvard Medical School). Two biological replicates were submitted for the analysis.

shRNA knockdown of *Brg1, Dek, Chd8 and Znhit* in *iHA-Etv2* MEFs.

Lentiviral particles were obtained using unique 21-mer shRNA against *Brg1* (pLKO.1 -TRC shRNA library, Oligo ID TRCN0000071383, Cat # RMM3981-97059770), *Dek* (pLKO.1 -TRC shRNA library, Oligo ID TRCN0000086420, Cat # RMM3981-97074717), *Chd8* (pGIPZ -GIPZ shRNA library, Oligo ID V2LMM_101581, Cat # RMM4431-98762531) and *Znhit* (pGIPZ -GIPZ shRNA library, Oligo ID V3LMM_501875, Cat # RMM4431-99958032). We generated these lentiviral particles using standard protocols in HEK293 T cells and tested each one of them by infecting NIH 3T3 MEFs (165). Cultured *iHA-Etv2* MEFs were infected with lentiviruses using the Lentiblast Premium reagent (OZBiosciences) as per the manufacturer's instruction. After 72 hrs of infection,

doxycycline was added to cells in order to reprogram them. *iHA-Etv2* MEFs were harvested for analysis 24 hrs and 7 days following the addition of doxycycline.

Analysis of scRNA-seq. The sequencing reads were mapped to the mouse genome (mm10) using TopHat (v2.0.13) and the raw read counts were obtained by HTSeq (v0.6.0) with default parameters (166, 167). We removed the genes that were detected in less than 3 cells. We used scrublet (v0.1) to detect the doublet (expected_doublet_rate=0.1, min_counts=2, min_cells=3, min_gene_variability_pctl=85, n_prpin_comps=30), and removed the cells where the doublet scores were greater than 0.25 (168). The single cells were then scaled and normalized by Seruat 3 (169). We used Seurat's ScaleData to regress out the effects of number of RNA counts, the number of detected genes, the percent of mitochondria genes, the percent of ribosomal genes, and cell cycle scores. We used the top 2,000 high variable genes (HVG) detected by Seurat for the dimension reduction analysis performed by scVI (170). We used UMAP (Uniform Manifold Approximation and Projection) for visualizing the scRNA-seq (171). There were 3,539, 2,936 and 7,202 high quality single cells from 24 hours, 48 hours and 7 days post-induction of Etv2 in mouse embryonic fibroblasts (MEFs), as well as 948 undifferentiated MEFs as well as 827 sorted Flk1^{+} cells from day 7 reprogrammed cells remaining for the downstream analysis.

Endothelial sprouting, acetylated-LDL (Ac-LDL) uptake and tube formation assay. Sprout formation assays were performed using *Brg1$^{f/f}$;Actin Cre-ERT2* EBs and a method previously described. Briefly, *Brg1$^{f/f}$;Actin Cre-ERT2* ESCs were differentiated into EBs as previously described (113). Following 4-OHT treatment from D0-D4, 30 EBs were transferred to a 12-well plate coated with low growth factor matrigel and supplemented with 50 ng/mL of VEGF. Sprouts were imaged and quantified using an inverted brightfield microscope at day 3 of plating. Ac-LDL update and tube formation assays were performed as previously described (172). Briefly, following doxycycline treatment to overexpress ETV2, *iHA-Etv2* MEFs were incubated with 10 mg/ml of Alexa594 conjugated AcLDL (Invitrogen) for 4 hr at 37C and analyzed by either fluoresecent microscopy. Vascular tube formation assays were performed by transfering MEFs onto growth factor-reduced Matrigel (BD Biosciences) and cultured in endothelial medium with 40 ng/ml VEGF165 at 37C in 5% CO2.

Analysis of bulk RNA-seq. The sequencing reads were mapped to the mouse genome (mm10) using TopHat (v2.0.13) and the raw read counts were obtained by HTSeq (v0.6.0) with default parameters (166, 167). The read counts data were normalized by DESeq2, followed by differential expression analysis (173).

Generation of ATAC-seq libraries and sequencing. EBs were collected at different time points and disaggregated in 0.25% trypsin at 37°C for 3 min incubation with gentle agitation followed by inactivation with culture medium

containing 10% FBS. Cells were collected by centrifugation at 500 x g for 5 minutes, washed once with ice-cold PBS. *iHA-Etv2* MEFs were harvested from the culture dishes at different time points by treating the cells with 0.25% trypsin at 37°C for 4 min incubation followed by neutralizing and collecting the cells with culture medium containing 10% FBS. ATAC reaction was performed with 50,000 cells as previously described (174) using the Tn5 transposase (Illumina) and libraries were created at the University of Minnesota Genome Center. Libraries were then sequenced on a NextSeq Illumina platform (2x50 bp) aiming for 25 million reads per sample.

ChIP, Re-ChIP and generation of ChIP-seq libraries. EB or MEF cells were harvested as described above. ChIP assays for ETV2 and H3K27ac were performed from 2 x 10^6 MEFs or 1 x 10^7 EB cells, respectively. Briefly, cells were crosslinked with 1% formaldehyde for 10 min at room temperature and the reaction was quenched by glycine at a final concentration of 0.125 M. For the BRG1 studies, cells were first crosslinked in 2 mM disuccinimidyl glutarate (DSG; Life Technologies: Cat. #20593) for 30 min then in 1% formaldehyde for 10 min, quenched with glycine for 5 min (175). The remainder of the ChIP protocol was performed following the protocol described by Magli et al. We used antibodies against ETV2 (5ug ab per 25ug of chromatin, Cat# ab181847, Lot# GR229868-10), H3k27ac (2ug ab per 25ug of chromatin, Cat# ab4729, Lot# GR3231988-1) and BRG1 antibody (5ug ab per 25ug of chromatin, Cat# ab110641, Lot# GR3208604-18). and ELK3 (2.5ug ab per 25ug of chromatin, Cat# NBP2-16315,

Lot# 40492) and control rabbit IgG (Cat #P120-101, Lot# 12) for the respective pulldowns. For both of the pulldowns, protein A Dynabeads were added to the ChIP reactions and incubated for 30 minutes at room temperature. Magnetic beads were washed and chromatin was eluted. After crosslinking reversal, RNase A, and proteinase K treatment, ChIP DNA was extracted with the Min-Elute PCR purification kit (Qiagen). ChIP DNA was quantified with Quant-it PicoGreen dsDNA Assay Kit (Life Technologies). Sequencing libraries were prepared using equal amount of ChIP (176) ThruPLEX® DNA-Seq Kit (Takara R400675) and SMARTer® DNA Unique Dual Index Kit - 24U Set A (Takara R400665). Sequencing was performed at the University of Minnesota Genome Center with the The NextSeq 550 high-throughput benchtop sequencer. The Re-ChIP protocol was performed using the protocol described by Furlan-Magaril and Recillas-Targa (176). The first round of pulldown was performed using the antibody against BRG1 and the second pulldown used the ETV2 antibody. Rabbit IgG was used as a negative control. PCR was used to verify the binding of ETV2 and BRG1 to small regulatory regions of *Flt1*, *Lmo2* and *Tcf12* using the eluted DNA and primers listed in Supplementary Table 2 (176). ChIP-qPCR was performed using SYBR green mastermix to analyze the recruitment of ELK3 using decrosslinked chromatin from ELK3 antibody pulldown of D4 *control* and *Brg1* KO EBs. The primers listed in Supplementary Table 3 were used for this assay.

ATAC-seq, ChIP-seq and MNase-seq analysis. The sequencing reads where mapped to the mouse genome (mm10) using Bowtie2 (v2.2.4) (177). The ATAC-seq peaks and ChIP-seq peaks were called by MACS2 (v2.1.1) (178). The ATAC-seq lied within the blacklisted genomic regions for functional genomics analysis were excluded (179). ChromVAR (v1.10) were used for transcription factor-based chromatin accessibility analysis. 321 transcription factors compiled in the Homer database were used for the chromVAR analysis. The pathway analysis was performed using R package ChIPseeker (180).

Nucleosome Occupancy and Methylome Sequencing (NOMe-seq): NOMe-seq was performed in duplicate using 250,000 cells from control and day 1 ETV2 reprogrammed MEFs per sample following the protocol described in Lay et al**.** (181). The reagents used in the studies were GpC Methyltransferase (M.CviPI) (NEB # M0227), Klenow Fragment (3'→5' exo-) (NEB # M0210S), EZ DNA Methylation™ Kit (VWR # 77001-534), End-It™ DNA End-Repair Kit (VWR # 75927-936), Rapid DNA Ligation Kit (ThermoFisher # K1422), xGen™ Methyl UDI-UMI Adapters, 1-16 (116 10006644), xGen Library Amplification Primer Mix (116 1077677), KAPA HiFi HotStart Uracil+ Kit (Roche # 07959052001). 200-500 bp size gel purified libraries were sequenced at the University of Minnesota Genome Center with the The NOVASeq S4 flow cell with 150 bp paired end reads.

NOMe-seq analysis. The sequencing reads were mapped to the mouse genome mm10 using Biscuit (https://github.com/zhou-lab/biscuit) (v0.3.16). The GC methylation information was extracted using the pipeline for NOMe-seq provided by Biscuit. Only the DNA methylation information in GCH (H represents A, T and C nucleotides) sites was used to call NDR and nucleosome (149). The genome was sized to 200 bp sliding windows with 20bp steps, and the proportion of non-artificially methylated GpC dinucleotides of every GCH site in each window were averaged and p-values (test of equal or given proportions) for the enrichment of non-artificially methylated GpC dinucleotides were calculated compared with the background. We used the proportion test (test of equal or given proportions) to compare the average GpC^m ratio surrounding the 200 bp flanking region of each ETV2 site to the background, and statistically determined whether any GCH site was protected by a nucleosome. Only the significant windows with p-values passed the cutoffs (0.05) were considered as significantly nucleosome occupied region (148). Additional downstream NOMe-seq analysis to visualize the nucleosome accessibility was performed using the aaRon R package (v0.9.5) (182).

V-plot analysis from ATAC-seq. We used a variational autoencoder (VAE) model implemented in R package SeATAC (https://github.com/gongx030/seatac) to map the V-plots centering at each peak to low dimensional space. The VAE model learned the latent representation of ATAC-seq V-plots using a CNN (convolutional neural networks) based encoder and a devolution decoder. The

nucleosome and NFR regions were determined by separating the V-plots on the low dimensional space into two groups. The V-plots clustering was performed by the Louvain clustering algorithm of a kNN (k-nearest neighbor) graph of the latent representation of V-plots (k=200),

Motif analysis. We used Discriminative DNA Motif Discovery algorithm (DREME, MEME suite 5.3.1) for *de novo* motif discovery (183). We focused on motifs occurring at the central 200 bp of Etv2 peaks, using central motif enrichment analysis (CentriMo, MEME suite 5.3.1) (183, 184). We quantified the occurrences of the first hits that returned with the most statistical significance within the Etv2 sites using Find Individual Motif Occurrences (FIMO, MEME suite 5.3.1) (185). Motifs that showed most central enrichment were considered. Moreover, the newly discovered motifs were compared to the JASPAR and UniPROBE motif databases using the Motif comparison tool (MEME) (186).

Data analysis. No data were excluded from these studies and all attempts at replication for standard assays (that is, FACS, qPCR, methylcellulose colony-forming assays and immunohistochemistry) were successful. Data collection and analysis were not performed blind to the conditions of the experiments.

Data availability. The scRNA-seq, bulk RNA-seq, ATAC-seq, ChIP-seq data and NOMe-seq of Etv2 induced EB differentiation and MEF reprogramming are deposited at NCBI Gene Expression Omnibus (GEO) database with the

accession GSE185684 (GSE168521: ChIP-seq of Etv2 induced MEF reprogramming and EB differentiation; GSE168636: ATAC-seq of Etv2 induced MEF reprogramming and EB differentiation; GSE185682: Bulk RNA-seq of Etv2 induced EB differentiation; GSE185683: Single cell RNA-seq of Etv2 induced MEFs reprogramming). All data will be available upon request. All unique materials used in these studies are readily available from the authors or from commercial sources (see Supplementary Tables 1-3). The MEF MNase-seq is from GSE90893. The MEF histone modification ChIP-seq of H3K9me3, H3K27me3, H3K9ac, H3K4me2, H3K4me1 and HDAC1 is from GSE90893. The H3K27ac ChIP-seq in Brg1 KD EB is from GSE71509. The scRNA-seq of iPSC reprogramming is from GSE100344. The scRNA-seq of cardiac fibroblast reprogramming is from GSE98567. The scRNA-seq of neural reprogramming is from GSE67310. The mass spectrometry data is available from GitHub webpage (https://github.com/gongx030/Etv2_pioneer).

Code availability. Codes pertaining to important analyses in this study are available from GitHub webpage (https://github.com/gongx030/Etv2_pioneer).

Footnotes

This work has already been accepted for publication in *Nature Cell Biology*. Reprinted from *Nature Cell Biology*. Wuming Gong, Satyabrata Das, Javier E. Sierra-Pagan Erik Skie, E, Nikita Dsouza, Thijs A. Larson, Mary G. Garry, Edgar Luzete-Monteiro, Kenneth S. Zaret, Daniel J. Garry. "ETV2 functions as a

pioneer factor to regulate and reprogram the endothelial lineage." Copyright 2022 with permission from *Nature Cell Biology*.

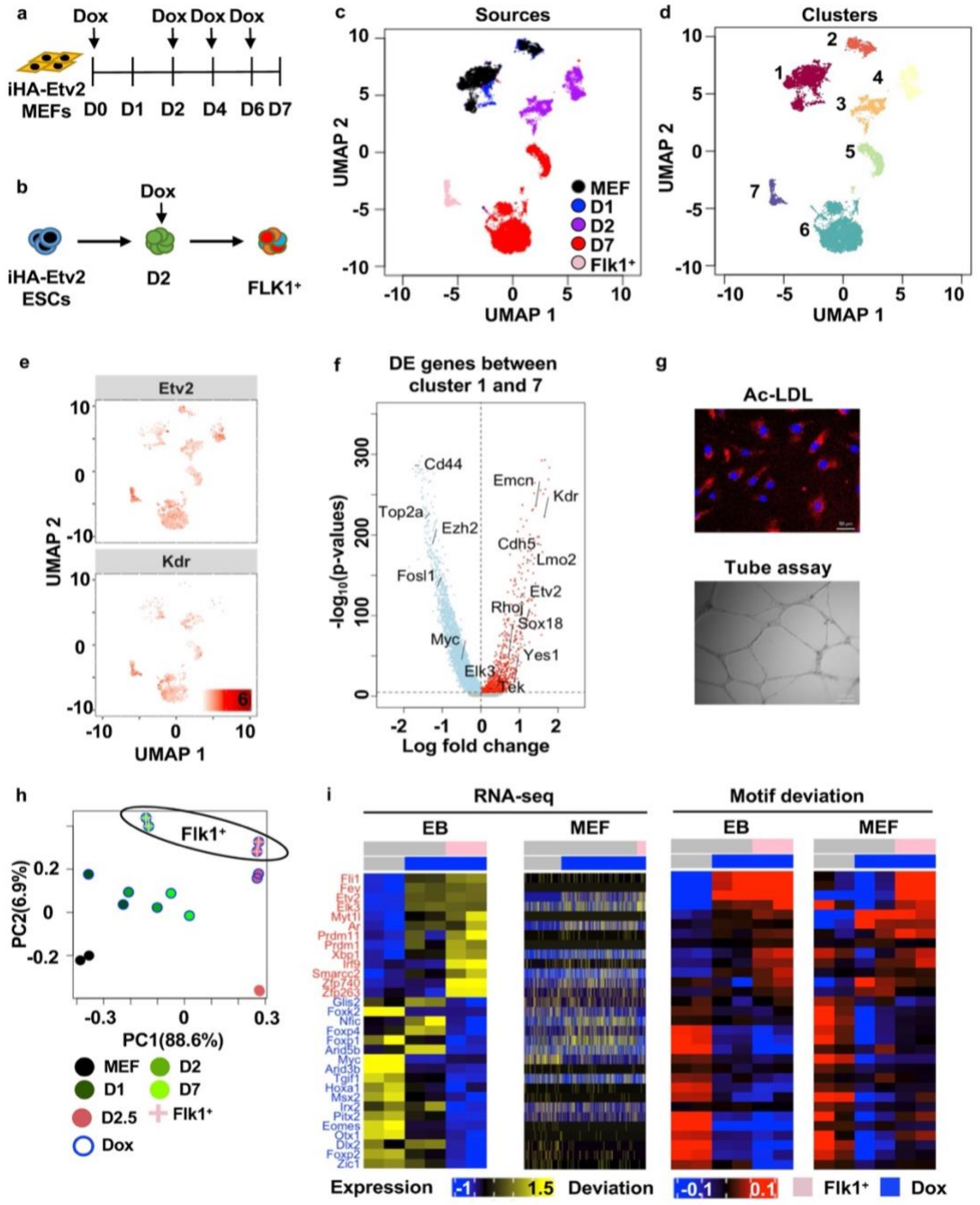

a
Dox
Dox Dox Dox
iHA-Etv2
MEFs
D0 D1 D2 D4 D6 D7
b
Dox
iHA-Etv2
ESCs
D2
FLK1+
c
Sources
UMAP 2
UMAP 1
MEF
D1
D2
D7
Flk1+
d
Clusters
1
2
3
4
5
6
7
e
Etv2
Kdr
UMAP 2
UMAP 1
f
DE genes between
cluster 1 and 7
Cd44
Top2a
Ezh2
Fosl1
Myc
Elk3
Emcn
Kdr
Cdh5
Lmo2
Etv2
Rhoj
Sox18
Yes1
Tek
-log10(p-values)
Log fold change
g
Ac-LDL
Tube assay
h
Flk1+
PC2(6.9%)
PC1(88.6%)
MEF
D1
D2.5
D2
D7
Flk1+
Dox
i
RNA-seq
Motif deviation
EB
MEF
EB
MEF
Fli1
Fev
Etv2
Elk3
Myt1l
Ar
Prdm11
Prdm1
Xbp1
Irf9
Smarcc2
Zfp740
Zfp263
Glis2
Foxk2
Nfic
Foxp4
Foxp1
Arid5b
Myc
Arid3b
Tgif1
Hoxa1
Msx2
Irx2
Pitx2
Eomes
Otx1
Dlx2
Foxp2
Zic1
Expression
-1
1.5
Deviation
-0.1
0.1
Flk1+
Dox

Figure 1. ETV2 promotes the endothelial program in both MEFs and EBs. **(a)** Schematic of the reprogramming strategy in *iHA-Etv2* MEFs by overexpression of ETV2 with doxycycline (Dox). **(b)** Schematic of the differentiation of embryoid body (134) and induction of ETV2 at day 2.5 of differentiation. **(c-d)** The UMAP plot shows the scRNA-seq of 948 undifferentiated MEFs, 3,539 reprogrammed cells at 24 hrs, 2,936 cells at 48 hrs and 7,202 cells at 7 days and 827 FLK1^{+}/KDR cells at 7 days post-induction of ETV2 in MEFs (one biological replicate per condition). The dimension reduction analysis by scVI, followed by uniform manifold approximation and projection (UMAP) and k-means clustering identified seven distinct cell clusters. The color represents **(c)** the cell sources and **(d)** cell clusters. **(e)** The expression profiles of ETV2 and FLK1/KDR. **(f)** The volcano plot of genes differentially expressed between cluster 1 and cluster 7. The *p*-values were determined by two-sided Wilcoxon rank sum test of the normalized read counts. **(g)** The Ac-LDL uptake experiment (upper panel) and the endothelial tube assays (lower panel) performed with the FLK1+ sorted cells7 days following ETV2 induction demonstrates an endothelial phenotype (Data shown here represent 3 independent experiments). **(h)** The PCA of the variations of transcription factor (TF) associated chromatin accessibility of the ATAC-seq of MEF reprogramming (MEFs, 24 hrs, 48 hrs and 7 days post-induction) and EB differentiation (2.5 days and 3 hrs post induction). The TF deviations were inferred by chromVAR (Data shown here represent two biological replicates) **(i)** The 31 TF expression levels and motif associated chromatin accessibility consistently showed directional change in both EBs and MEFs (13 up-regulated TFs and 18 down-regulated TFs). Data shown here represent the average of two biological replicates.

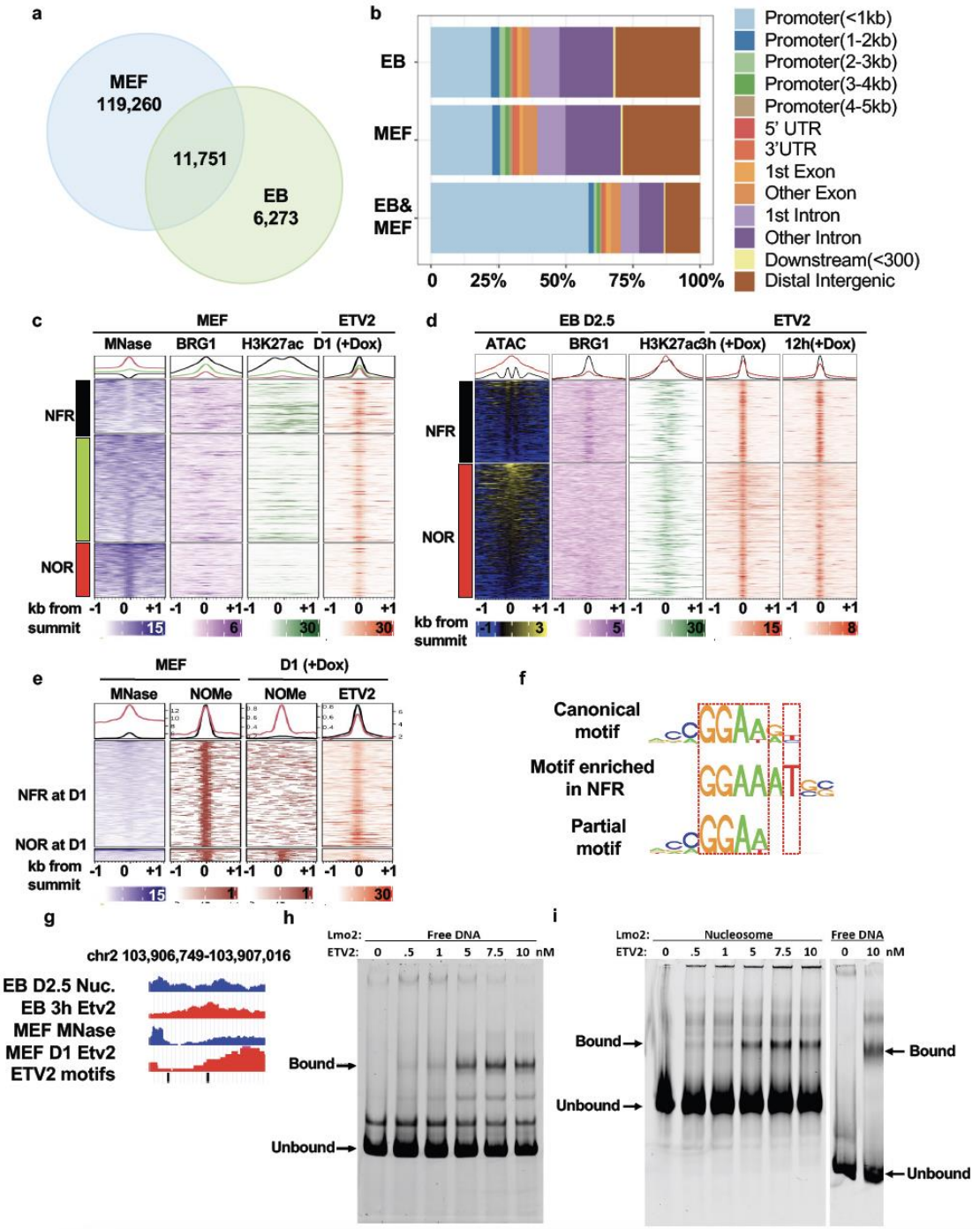

a
MEF
119,260
11,751
EB
6,273
b
EB
MEF
EB&
MEF
0
25%
50%
75%
100%
Promoter(<1kb)
Promoter(1-2kb)
Promoter(2-3kb)
Promoter(3-4kb)
Promoter(4-5kb)
5' UTR
3'UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream(<300)
Distal Intergenic
c
MEF
ETV2
MNase
BRG1
H3K27ac
D1 (+Dox)
NFR
NOR
kb from summit
d
EB D2.5
ETV2
ATAC
BRG1
H3K27ac
3h (+Dox)
12h(+Dox)
e
MEF
D1 (+Dox)
MNase
NOMe
NOMe
ETV2
NFR at D1
NOR at D1
f
Canonical motif
Motif enriched in NFR
Partial motif
g
chr2 103,906,749-103,907,016
EB D2.5 Nuc.
EB 3h Etv2
MEF MNase
MEF D1 Etv2
ETV2 motifs
h
Lmo2:
Free DNA
ETV2:
0 .5 1 5 7.5 10 nM
Bound
Unbound
i
Lmo2:
Nucleosome
Free DNA
ETV2:
0 .5 1 5 7.5 10
0 10 nM
Bound
Unbound

Figure 2. ETV2 targets nucleosomes during reprogramming. (a) The Venn diagram shows 131,001 and 18,024 Etv2 ChIP-seq peaks at 24 hrs post-induction during MEF reprogramming and 3 hrs post-induction in day 2.5 EBs, respectively. There were 11,751 common ETV2 peaks overlapped between MEFs and EBs. **(b)** The genomic distribution of EB specific, MEF specific and common ETV2 peaks. **(c)** The heatmap shows the read density of MNase-seq, BRG1 and H3K27ac ChIP-seq in MEFs, surrounding 131,001 ETV2 ChIP-seq peaks at 24 hrs post-induction during MEF reprogramming. The ETV2 peaks were divided into four quartiles based on the mean MNase-seq signals of the central 200-bp region. The first (lowest mean signal) and the fourth quartile (highest mean signal) were used to represent the nucleosome free region (NFR) and nucleosome. **(d)** The heatmap shows the ratio of NFR / nucleosome read density, read density of BRG1 and H3K27ac ChIP-seq in EBs (day 2.5), surrounding 18,024 ETV2 ChIP-seq at 3 hrs post-induction. The ETV2 peaks were divided into NFR and nucleosome groups according to the local V-plot and fragment size profiles of ATAC-seq day 2.5 EBs without ETV2 induction. **(e)** The heatmap generated using NOMe-seq shows among 5,320 ETV2 binding sites that were nucleosome occupied at undifferentiated MEF, 4,744 became significantly nucleosome-free while 576 stayed NOR at D1 of reprogramming. The NOMe-seq is represented as the proportion of non-artificially methylated GpC dinucleotides. **(f)** Sequence motif analysis by DREME and CentriMo identified a common GGAAAT motif that were significantly more enriched in NFR regions compared with the nucleosomes in both MEFs and EBs (Fisher's adjusted p-value=6.0E-05 and 6.8E-5). **(g)** A region upstream of *Lmo2* that was enriched for nucleosomes in both cell types, as measured by MNase-seq and ATAC-seq was selected to perform in vitro nucleosomal binding assays (Chr2: 103,906,749-103,907,016). This locus was also bound by ETV2 as indicated by Etv2 ChIP-seq in both EBs and MEFs. **(h)** Recombinant ETV2 protein bound to Cy5-labeled *Lmo2*-DNA probes without any histones (free DNA). **(i)** EMSA shows *Lmo2* free DNA, reconstituted *in vitro* assembled *Lmo2* nucleosomes and ETV2 binding to nucleosomes. All lanes are from the same gel. Data shown in 2c,2d and 2e represent the average of two biological replicates.

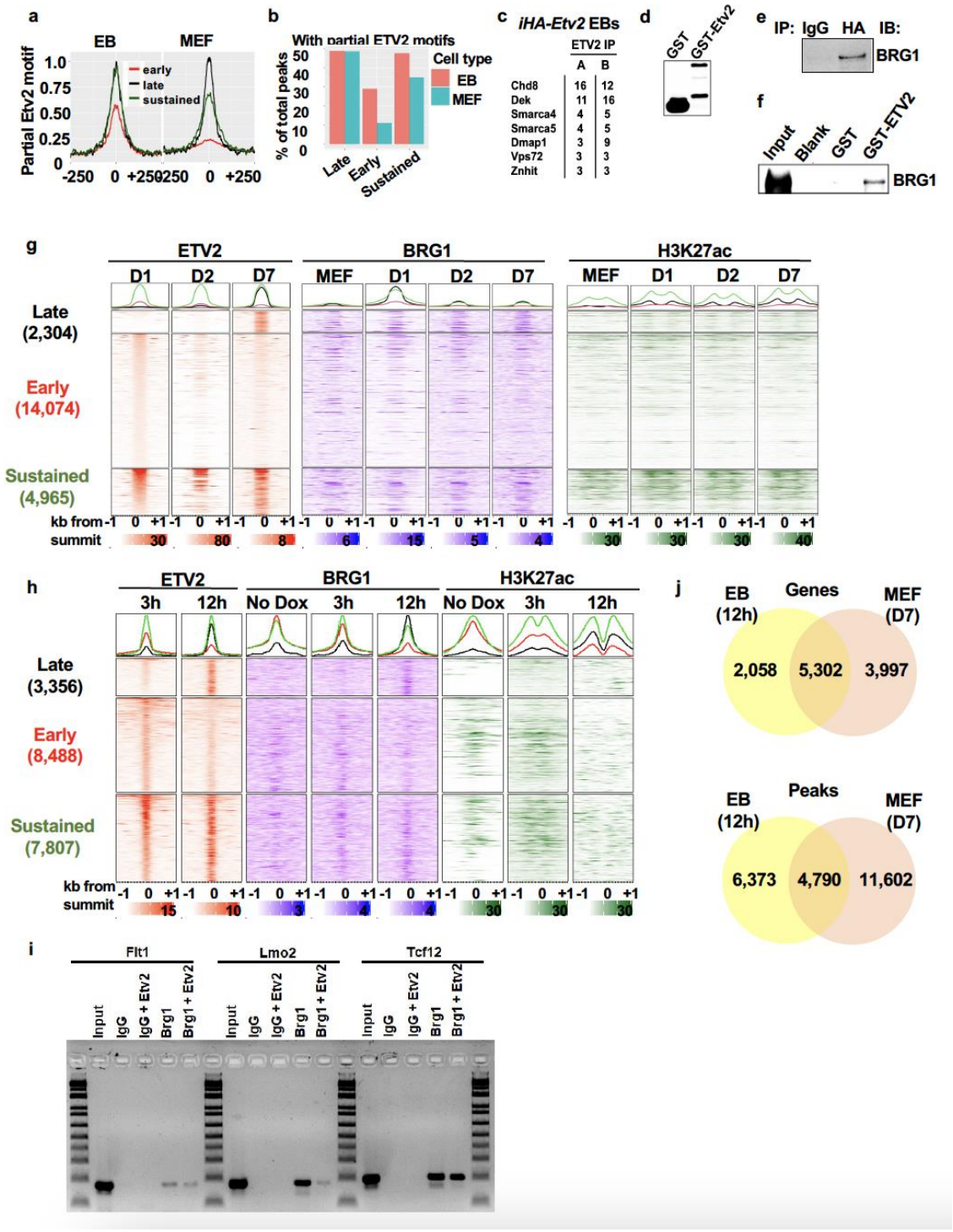

a
EB
MEF
Partial Etv2 motif
early
late
sustained
b
With partial ETV2 motifs
% of total peaks
Cell type
EB
MEF
Late
Early
Sustained
c
iHA-Etv2 EBs
ETV2 IP
A B
Chd8 16 12
Dek 11 16
Smarca4 4 5
Smarca5 4 5
Dmap1 3 9
Vps72 3 3
Znhit 3 3
d
GST
GST-Etv2
e
IP: IgG HA IB:
BRG1
f
Input
Blank
GST
GST-ETV2
BRG1
g
ETV2
BRG1
H3K27ac
D1 D2 D7
MEF D1 D2 D7
MEF D1 D2 D7
Late (2,304)
Early (14,074)
Sustained (4,965)
kb from summit
h
ETV2
BRG1
H3K27ac
3h 12h
No Dox 3h 12h
No Dox 3h 12h
Late (3,356)
Early (8,488)
Sustained (7,807)
kb from summit
j
EB (12h)
Genes
MEF (D7)
2,058 5,302 3,997
EB (12h)
Peaks
MEF (D7)
6,373 4,790 11,602
i
Flt1
Lmo2
Tcf12
Input
IgG
IgG + Etv2
Brg1
Brg1 + Etv2

Figure 3. ETV2 recruits BRG1 for chromatin remodeling. (a) The average partial Etv2 motif scores in 500 bp regions surrounding the summit of the "early', "late" and "sustained" Etv2 peaks in EB differentiation and MEF reprogramming. **(b)** The percent of the "early', "late" and "sustained" Etv2 peaks in EBs and MEFs include partial Etv2 motifs in 50bp regions surrounding the summits. **(c)** List of chromatin remodelers enriched in the ETV2 pulldown samples over control IgG pulldown samples detected by mass spectrometry. **(d)** Empty GST and GST-ETV2 proteins purified from bacterial cultures and probed by a GST-antibody. **(e)** Co-immunoprecipitation of BRG1 using a HA antibody to pull down HA-Etv2 from Dox-induced D4 EBs. **(f)** The GST-pulldown assay was performed with in vitro translated BRG1 protein. The pulled down materials were analyzed by western blotting with the BRG1 antibody [representative blots (d-f) from 3 independent experiments with similar results]. **(g)** The heatmap shows the fold enrichment of ETV2, BRG1 and H3K27ac ChIP-seq in undifferentiated MEFs, at 24 hrs, 48 hrs and 7 days post-induction of ETV2. Each site is centered at the 21,343 non-overlapping ETV2 ChIP-seq summits (with ETV2 motif) during MEF reprogramming. **(h)** The heatmap shows the fold enrichment of ETV2, BRG1 and H3K27ac ChIP-seq, in uninduced day 2.5 EBs and at 3 hrs and 12 hrs post-induction of ETV2 in day 2 EBs. Each site is centered at the 19,651 non-overlapping ETV2 ChIP-seq summits during EB differentiation. **(i)** Re-ChIP by PCR shows the co-binding of ETV2 and BRG1 in regulatory regions of *Flt1*, *Lmo2* and *Tcf12,* respectively. The IgG and IgG + Etv2 pulldowns were used as negative controls for the ChIP and the input chromatin was used a positive control for the PCR (representative blot from 2 biological replicates and 2 independent experiments). **(j)** The Venn diagram shows the overlap between late ETV2 peaks and nearby genes in MEFs and EBs. Data shown in 3a, 3b, 3g and 3h represent the average of two biological replicates.

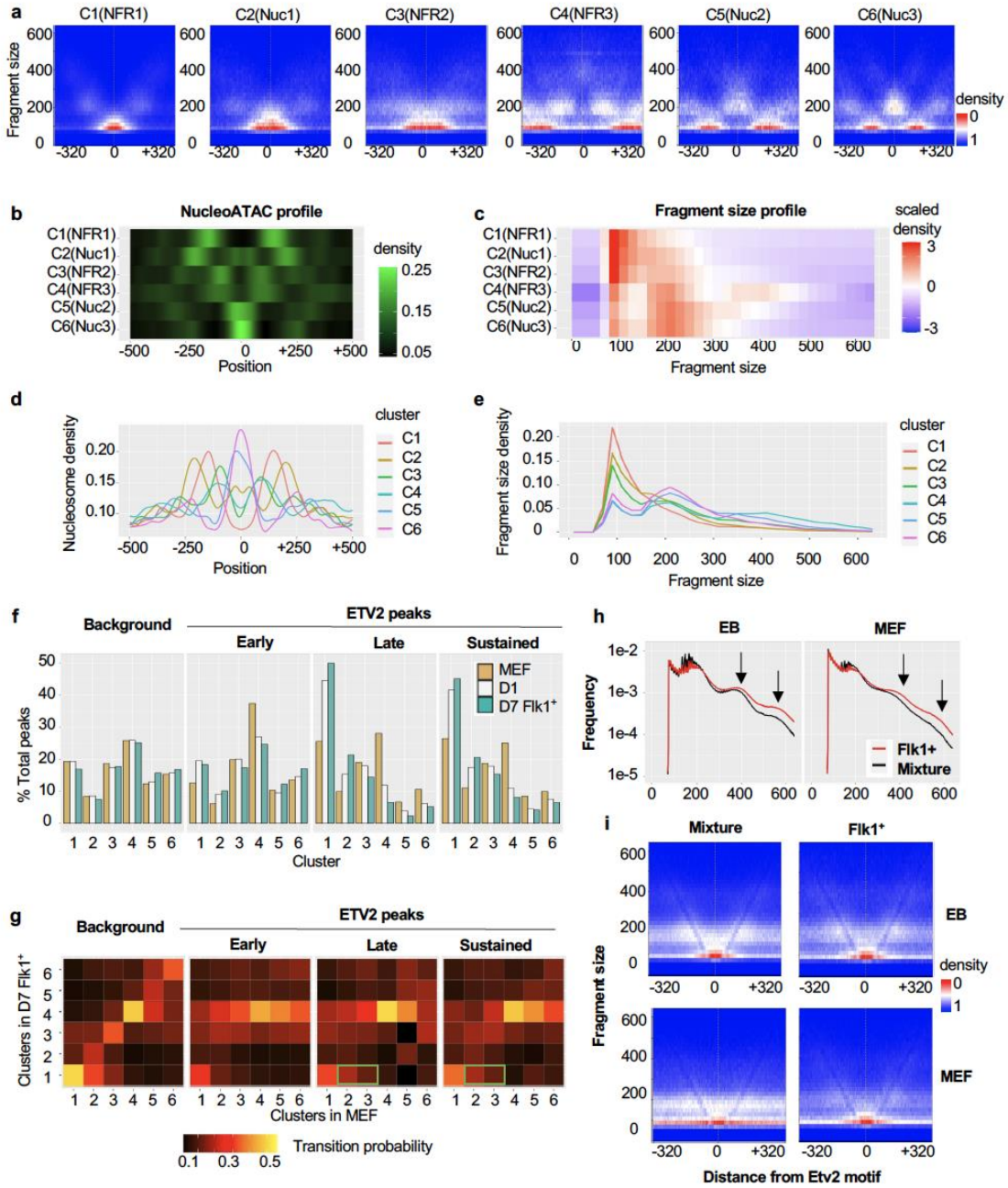

a
C1(NFR1)
C2(Nuc1)
C3(NFR2)
C4(NFR3)
C5(Nuc2)
C6(Nuc3)
Fragment size
density
b
NucleoATAC profile
Position
c
Fragment size profile
scaled density
Fragment size
d
Nucleosome density
cluster
e
Fragment size density
f
Background
ETV2 peaks
Early
Late
Sustained
MEF
D1
D7 Flk1+
% Total peaks
Cluster
g
Clusters in D7 Flk1+
Clusters in MEF
Transition probability
h
EB
MEF
Frequency
Flk1+
Mixture
i
Mixture
Flk1+
Fragment size
Distance from Etv2 motif

Figure 4. Phased nucleosomes are established surrounding ETV2 peaks. **(a)** We applied the VAE to 18,214 Etv2 ChIP-seq peaks during MEF reprogramming and identified six clusters of V-plot according to the central fragment size distribution. The six clusters included three types of V-plots where the central Etv2 sites were nucleosome free (C1, C3 and C4), and three types of V-plots where the central Etv2 sites were nucleosome occupied (C2, C5, and C6), represented by the **(b and d)** aggregated V-plot and **(c and e)** NucleoATAC **(f)** The bar plot shows the proportion of each V-plot clusters in early, late and sustained Etv2 peaks, as well as in the background peaks. **(g)** The cluster labels of early, late and sustained ETV2 peak that were changed from MEF to D7 Flk1$^+$ cells. **(h)** The fragment size distribution of ATAC-seq of FLK1$^+$ cells vs. the mixture population at 12 hours post-ETV2 induction during EB reprogramming and day 7 post-ETV2 induction during MEF reprogramming. In both conditions, the mono-nucleosomes and the di-nucleosomes were significantly increased in the FLK1$^+$ cell populations. **(i)** The aggregated V-plot whose centers are the ETV2 bound sites at FLK1$^+$ cell populations at 12 hours post-ETV2 induction in EB, and at 7 days post-ETV2 induction. Data shown here are based on the average of two biological replicates.

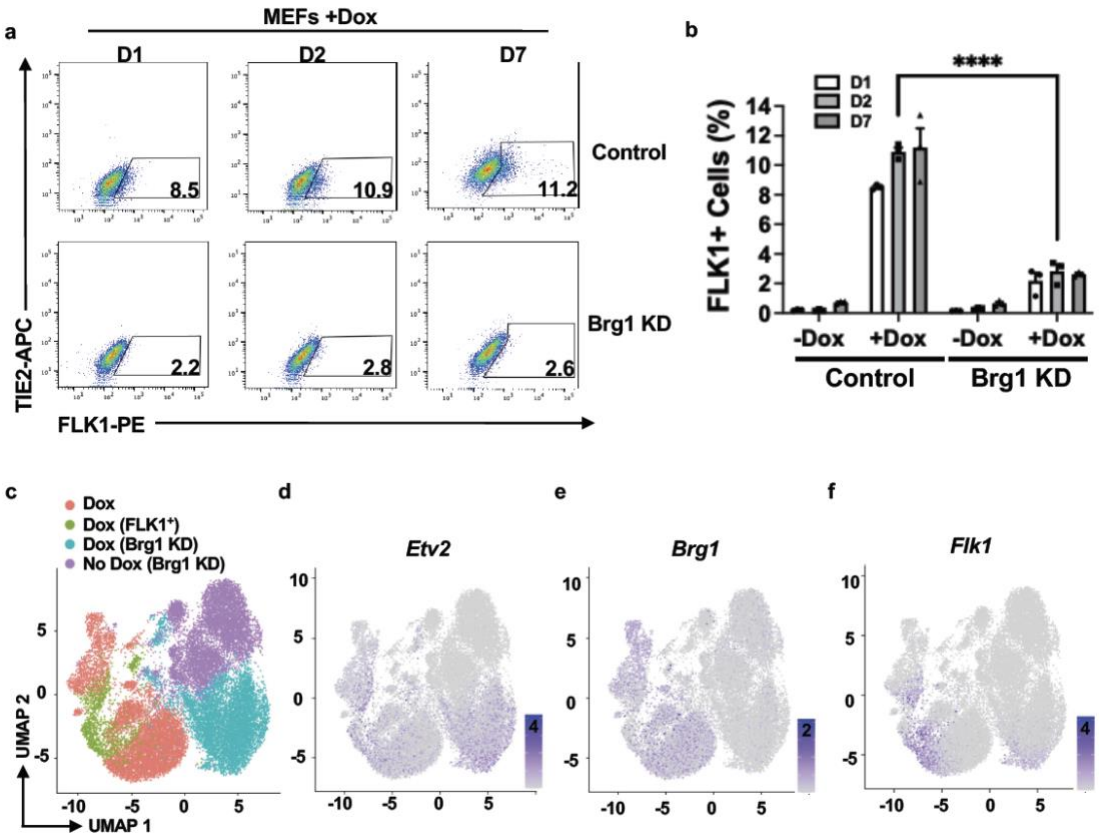


Figure 5. *Brg1* knockdown resulted in a significant decrease in cells expressing FLK1 during reprogramming. (50) *Brg1* knockdown resulted in a significant decrease (p=1.8e-14) in cells expressing FLK1 on D1, D2 and D7 of reprogramming as measured by flow cytometry (n=3 biological replicates; 2way ANOVA with multiple comparison ****P=1.0x10-15). Data are presented as mean±SEM. **(c-f)** The UMAP plot shows the scRNA-seq of 8,838 cells from day 7 post-induction in control MEFs, 1,502 FLK1^{+} cells from day 7 post-ETV2 induction in MEFs, 8,248 cells from day 7 post-ETV2 induction in *Brg1* KD MEFs, and 8,034 cells at day 7 in *Brg1* KD MEFs (without induction) (one biological replicate per condition). The dimension reduction analysis by scVI, followed by uniform manifold approximation and projection (UMAP). The color represents **(c)** the cell sources and **(d-f)** the expression levels.

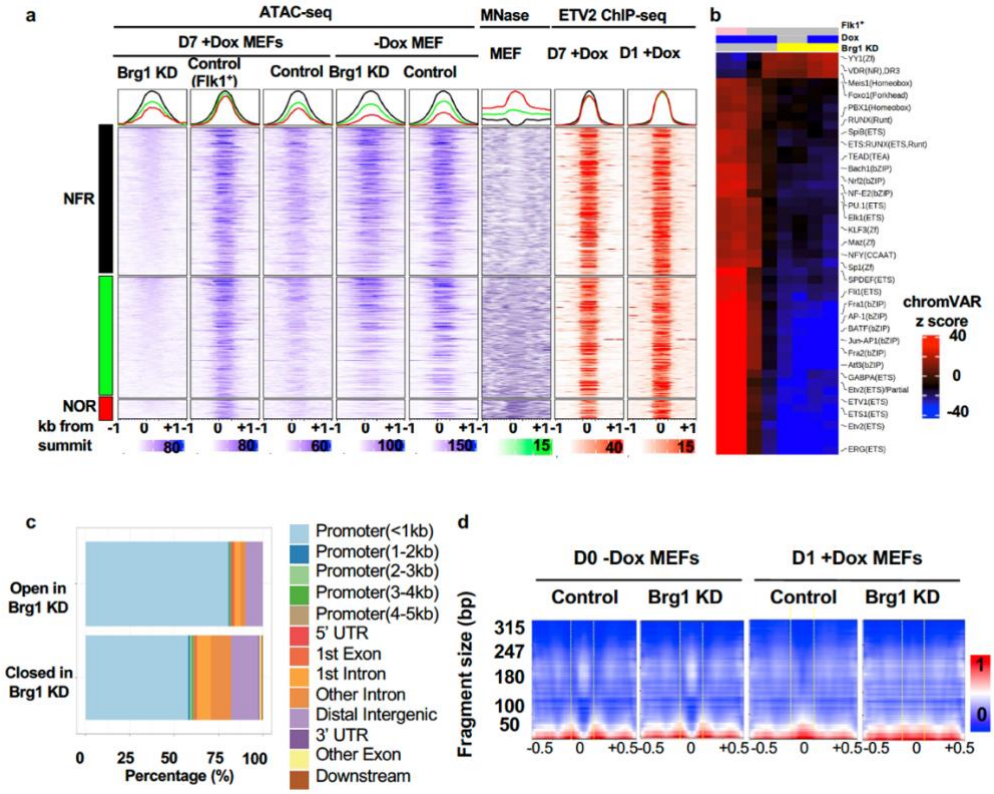


Figure 6. ETV2 requires BRG1 to activate downstream genes during reprogramming. (a) The heatmap shows the piled up ATAC-seq signal surrounding the summit of 12,170 sustained ETV2 ChIP-seq peaks that were present at day 1 and day 7 post-induction of ETV2 in control MEFs (the sustained Etv2 peaks). The ATAC-seq data include undifferentiated MEFs, day 7 post-ETV2 induction, and FLK1$^+$ cells from day 7 post-ETV2 induction in MEFs. We also include the ATAC-seq data from undifferentiated *Brg1* KD (knockdown) in MEFs and day 7 post-ETV2 induction with *Brg1* KD in MEFs. The sustained ETV2 peaks were divided into three groups: NFR (187), uncertain (green) and NOR (132), according to the chromatin accessibility in undifferentiated MEF. **(b)** The heatmap shows the transcription factors where motif associated chromatin accessibility were significantly changed at day 7 post-ETV2 induction in MEFs (unsorted MEFs or FLK1$^+$ cells), or the *Brg1* KD MEFs (adjusted *p*-value of two-sided χ^2 test in chromVAR <1e-200). The colors red and blue indicate the enrichment and the deficiency of the transcription factor associated ATAC-seq reads in associated conditions, respectively. **(c)** The genomic distribution of open ETV2 peaks and closed ETV2 peaks at 7 days post-ETV2 induction in *Brg1* KD MEFs. These ETV2 peaks were present in both day 1 and day 7 post-ETV2 induction in MEFs. **(d)** The aggregated V-plot of 1,204 ETV2 binding sites that were closed in MEFs and became open at day 1 post-ETV2 induction in MEFs. Top left: control MEFs; top right: *Brg1* KD MEFs; bottom left: 24 hrs post-ETV2 induction in MEFs; bottom right: 24 hrs post-ETV2 induction with *Brg1* KD MEFs. Data shown here represent the average of two biological replicates.

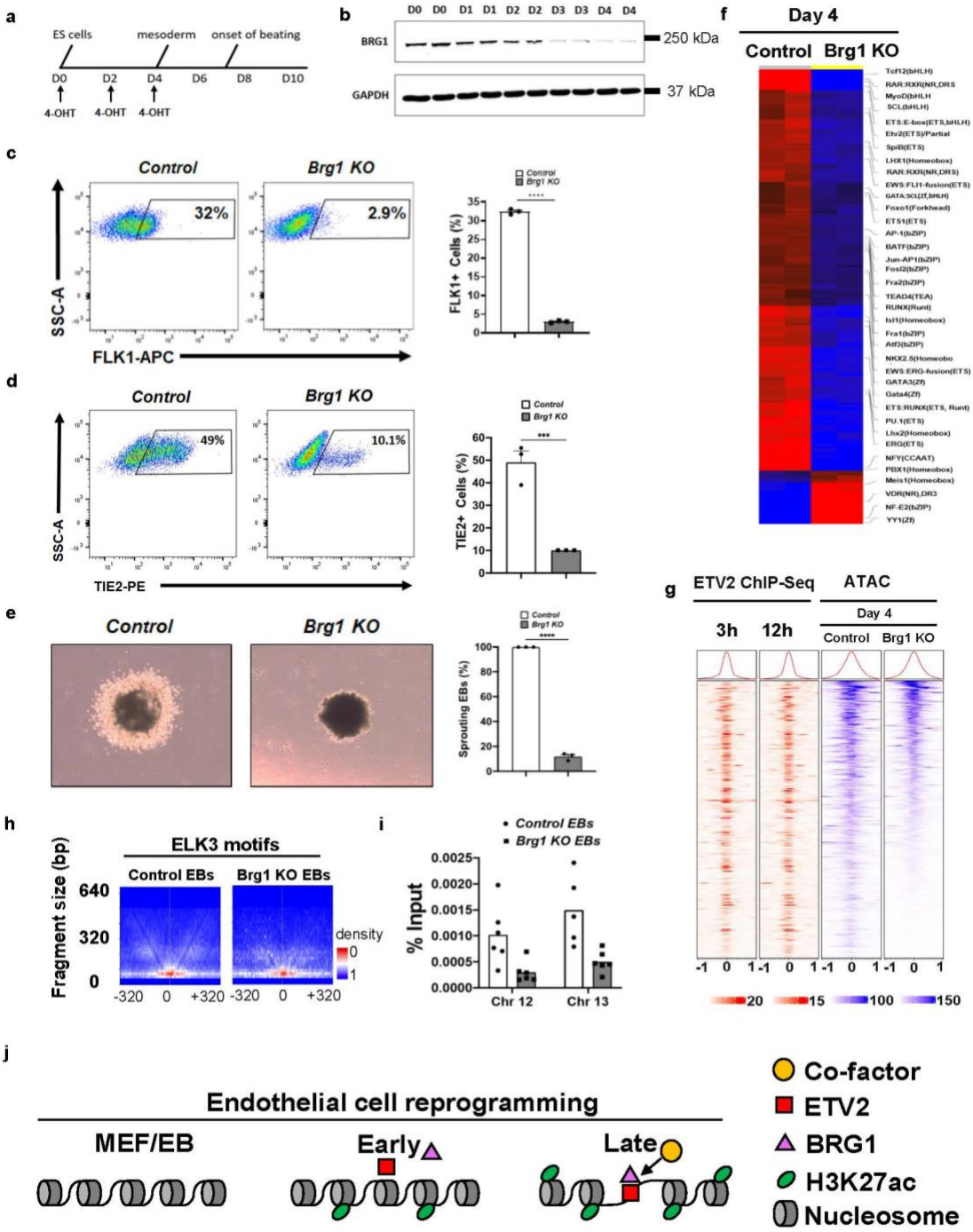
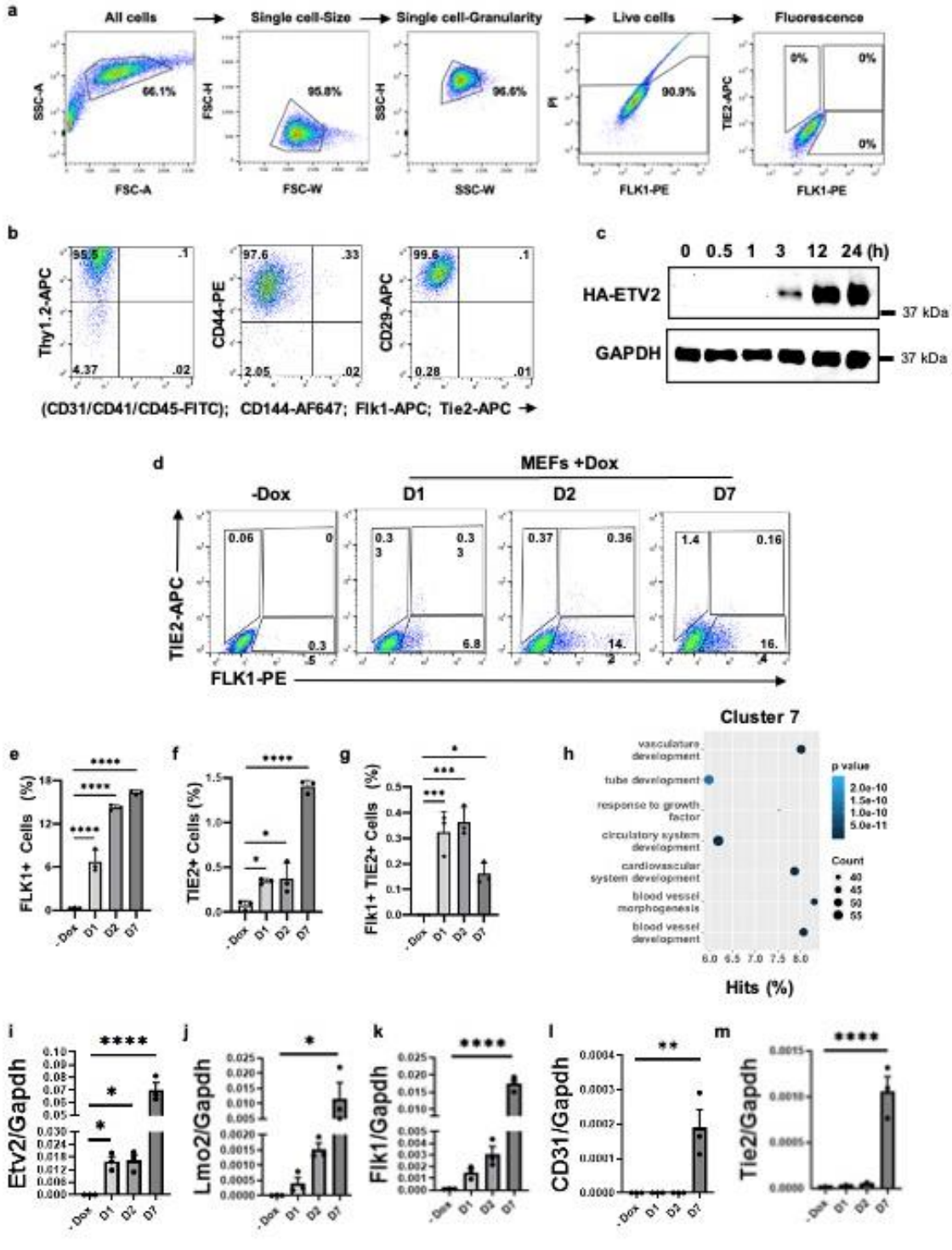
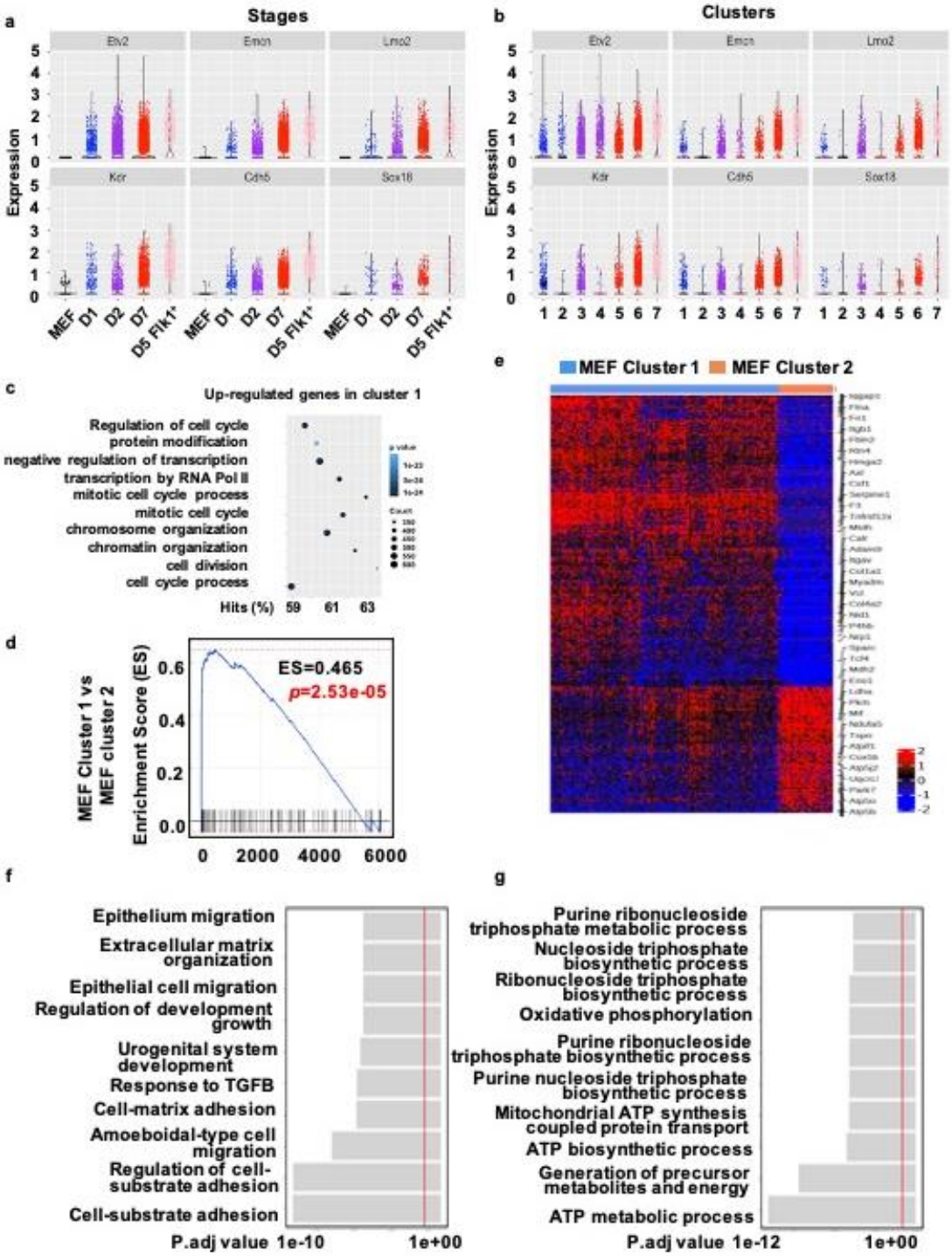

a
ES cells
mesoderm
onset of beating
D0
D2
D4
D6
D8
D10
4-OHT
4-OHT
4-OHT
b
D0 D0 D1 D1 D2 D2 D3 D3 D4 D4
BRG1
250 kDa
GAPDH
37 kDa
f
Day 4
Control
Brg1 KO
c
Control
Brg1 KO
32%
2.9%
SSC-A
FLK1-APC
FLK1+ Cells (%)
d
Control
Brg1 KO
49%
10.1%
SSC-A
TIE2-PE
TIE2+ Cells (%)
e
Control
Brg1 KO
Sprouting EBs (%)
g
ETV2 ChIP-Seq
ATAC
3h
12h
Day 4
Control
Brg1 KO
h
ELK3 motifs
Control EBs
Brg1 KO EBs
Fragment size (bp)
640
320
0
density
-320
0
+320
i
Control EBs
Brg1 KO EBs
% Input
Chr 12
Chr 13
j
Endothelial cell reprogramming
MEF/EB
Early
Late
Co-factor
ETV2
BRG1
H3K27ac
Nucleosome

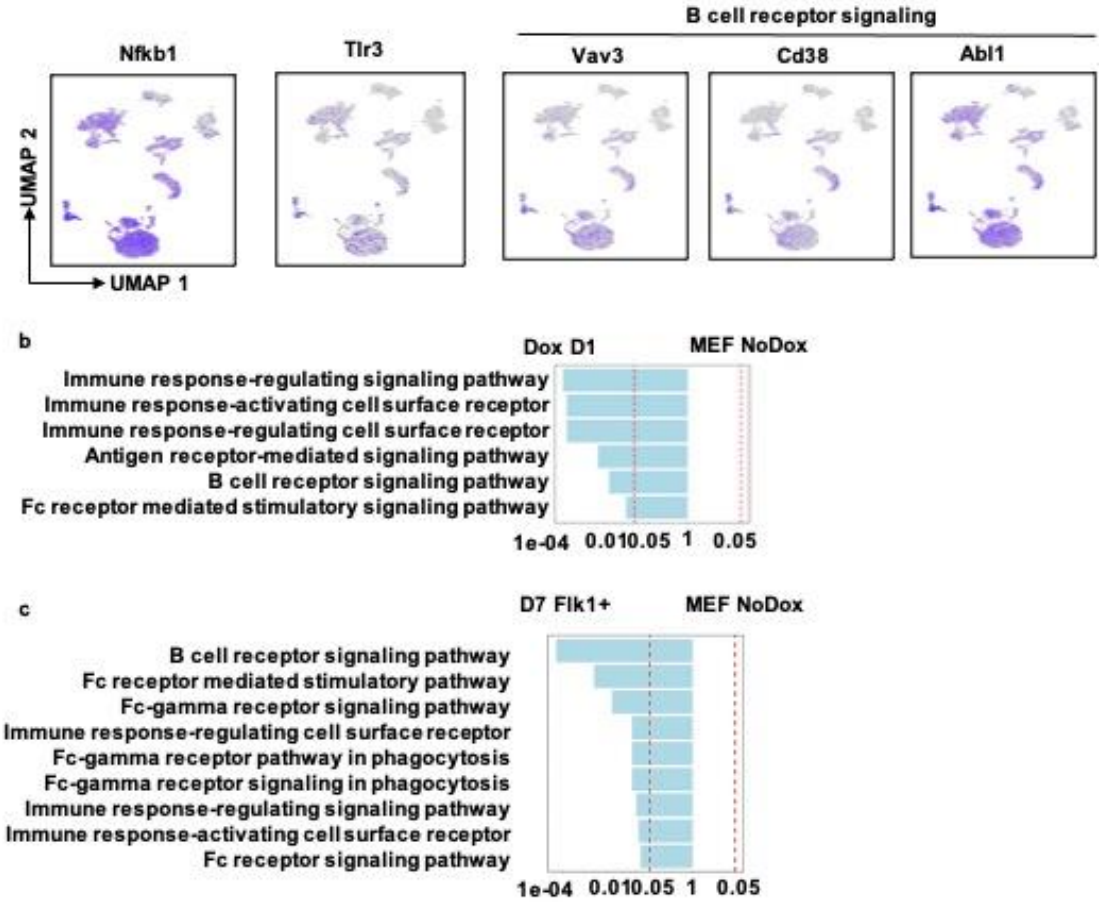
Figure 7. *Brg1* conditional knockout in ES/EBs. (a) Schematic of Brg1 conditional knockout using 4-OHT in Brg1$^{f/f;}$Actin-CreER shaking EBs undergoing mesodermal differentiation. **(b)** Western blot analysis of BRG1 expression in *Brg1$^{f/f}$;Actin-CreER* shaking EBs exposed to 200nM 4-OHT from D0-D4 of differentiation [representative blots (b) from 3 independent experiments with similar results]. **(c)** Flow cytometry analysis of D4 EBs comparing FLK1 expression between Control and *Brg1* KO cells. Conditional knockout of Brg1 in EBs significantly affects the generation FLK1+ progenitor cells (n=3 biological replicates; one-tailed unpaired t test ****P<1.0x10-4). **(d)** Flow cytometry analysis of D6 EBs comparing TIE2 expression between Control and Brg1 KO cells. Conditional knockout of Brg1 in EBs significantly affects the generation mature endothelial cells (n=3 biological replicates; one-tailed unpaired t test ***P=*8*.0x10-4). **(e)** Analysis of sprouts formed from D4 EBs and analyzed 72 hrs later showed a significant decreased in sprout formation in cells lacking Brg1 (n=3 biological replicates; one-tailed unpaired t test ****P<1.0x10-4). Data are presented as mean±SEM. **(f)** The heatmap shows chromatin accessibility for transcription factors for control vs. *Brg1* knockout during ES/EB differentiation. The heatmap indicates an increase (132) or decrease (blue) in accessibility for transcription factors with a significant change. **(g)** The heatmap shows the ChIP-seq of sustained ETV2 binding sites in EB differentiation, and ATAC-seq of Control and *Brg1* KO EBs at day 4. **(h)** ATAC-seq V-plots of the genomic regions (640 bp) that are centered at ELK3 motifs. **(i)** qPCR analysis using ELK3 ChIP demonstrates that in the absence of BRG1, ELK3 recruitment to ETV2-BRG1 co-occupied DNA binding sites is significantly reduced (n=2 biological replicates; 6 technical replicates) at two independent locations, chr12:12697258-12697757 and chr13:75076251-75076750 of the mouse genome. Data are presented as mean±SEM. **(j)** The model of ETV2 induced reprogramming. ETV2 targets the nucleosome during the early stages of reprogramming. ETV2 recruits BRG1 to maintain and stabilize the ETV2 binding near the endothelial genes, coupled with the increase of local H3K27ac levels. Data shown in 7f, 7g and 7h represent the average of two biological replicates.


a
All cells
Single cell-Size
Single cell-Granularity
Live cells
Fluorescence
SSC-A
FSC-A
66.1%
FSC-H
FSC-W
95.8%
SSC-H
SSC-W
96.6%
PI
FLK1-PE
90.9%
TIE2-APC
FLK1-PE
0%
0%
0%
b
Thy1.2-APC
CD44-PE
CD29-APC
(CD31/CD41/CD45-FITC); CD144-AF647; Flk1-APC; Tie2-APC
c
0 0.5 1 3 12 24 (h)
HA-ETV2
GAPDH
37 kDa
37 kDa
d
MEFs +Dox
-Dox
D1
D2
D7
TIE2-APC
FLK1-PE
e
FLK1+ Cells (%)
f
TIE2+ Cells (%)
g
Flk1+ TIE2+ Cells (%)
h
Cluster 7
vasculature development
tube development
response to growth factor
circulatory system development
cardiovascular system development
blood vessel morphogenesis
blood vessel development
p value
Count
Hits (%)
i
Etv2/Gapdh
j
Lmo2/Gapdh
k
Flk1/Gapdh
l
CD31/Gapdh
m
Tie2/Gapdh

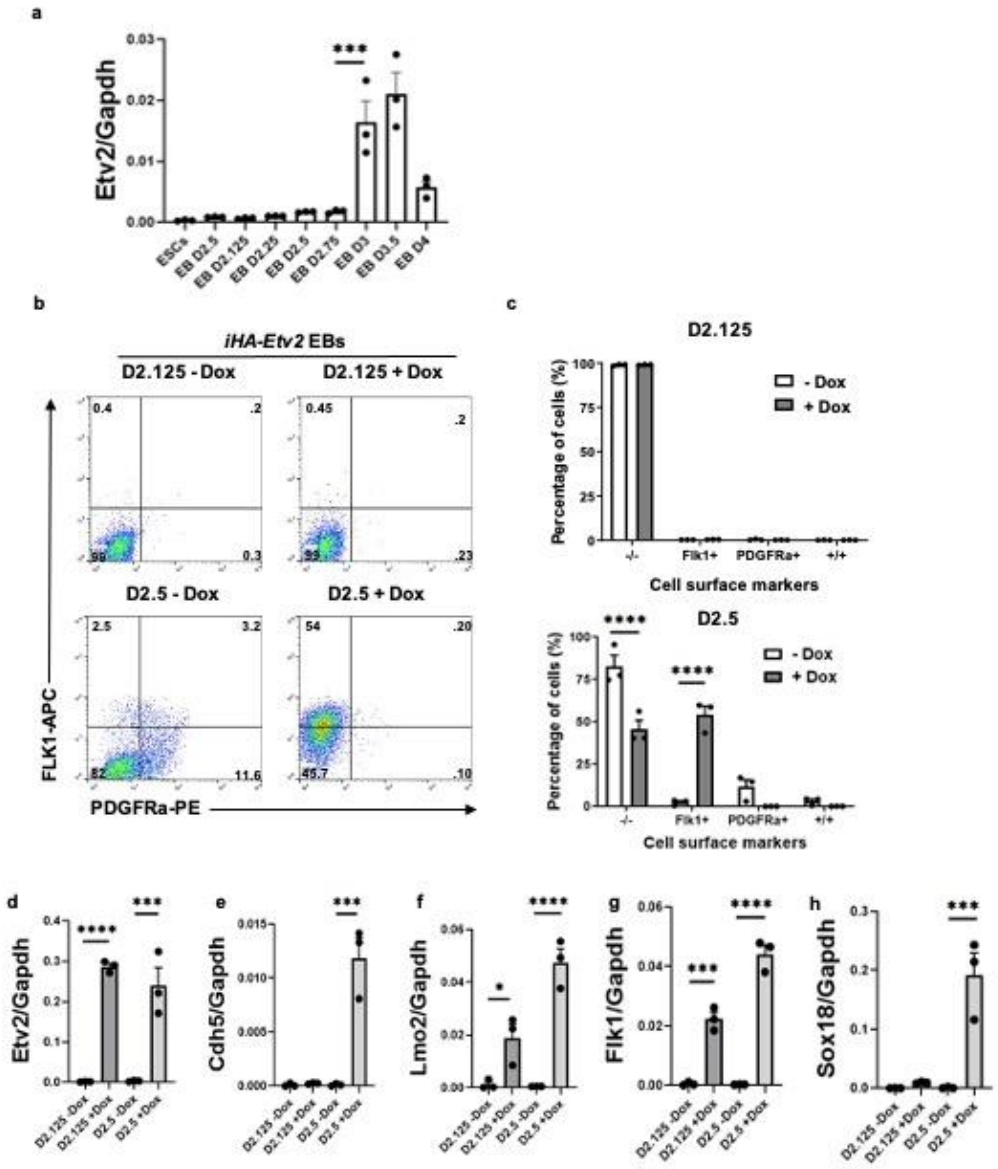
Extended Data Fig. 1. Characterization of mouse embryonic fibroblasts that overexpress ETV2 (*iHA-Etv2* MEFs) following the addition of doxycycline to reprogram MEFs to endothelial cells. (a) The gating strategy for iHA-Etv2 MEF FACS characterization is outlined in the five profiles. **(b)** We isolated embryonic fibroblasts from this mouse line and demonstrated that this cell population uniformly expressed fibroblast markers (Thy1.2, CD44 and CD29). **(c)** The western blot analysis showed that ETV2 was robustly expressed within 3 hrs post-Dox treatment [representative blots (c) from 3 independent experiments with similar results]. **(d-g)** ETV2 overexpression resulted in an increase in cells expressing FLK1/TIE2, as measured by FACS (n=3 biological replicates; *P<0.05). **(h)** The biological processes that are significantly associated with the up-regulated genes in cluster 7 ($FLK1^+$ cells at day 7 of reprogramming) compared with cluster 1 (undifferentiated MEFs). **(I-m)** qPCR experiments showed the increased expression levels of endothelial genes **(i)** *Etv2* (*P=0.0423, *P=0.0377, ****P<1x10-4), **(j)** *Lmo2* (*P=0.0496), **(k)** *Flk1* (****P<1x10-4), **(l)** *CD31* (**P=0.0044) and **(m)** *Tie2* (****P<1x10-4) at 1 day, 2 days, and 7 days post induction of ETV2, as well as the no Dox control (n=3 biological replicates; one-way ANOVA with multiple comparison). Data are presented as mean±SEM.


a
Stages
Etv2
Emcn
Lmo2
Kdr
Cdh5
Sox18
Expression
MEF
D1
D2
D7
D5 Flk1+
b
Clusters
c
Up-regulated genes in cluster 1
Regulation of cell cycle
protein modification
negative regulation of transcription
transcription by RNA Pol II
mitotic cell cycle process
mitotic cell cycle
chromosome organization
chromatin organization
cell division
cell cycle process
Hits (%)
d
MEF Cluster 1 vs MEF cluster 2
Enrichment Score (ES)
ES=0.465
p=2.53e-05
e
MEF Cluster 1
MEF Cluster 2
f
Epithelium migration
Extracellular matrix organization
Epithelial cell migration
Regulation of development growth
Urogenital system development
Response to TGFB
Cell-matrix adhesion
Amoeboidal-type cell migration
Regulation of cell-substrate adhesion
Cell-substrate adhesion
P.adj value 1e-10
1e+00
g
Purine ribonucleoside triphosphate metabolic process
Nucleoside triphosphate biosynthetic process
Ribonucleoside triphosphate biosynthetic process
Oxidative phosphorylation
Purine ribonucleoside triphosphate biosynthetic process
Purine nucleoside triphosphate biosynthetic process
Mitochondrial ATP synthesis coupled protein transport
ATP biosynthetic process
Generation of precursor metabolites and energy
ATP metabolic process
P.adj value 1e-12
1e+00

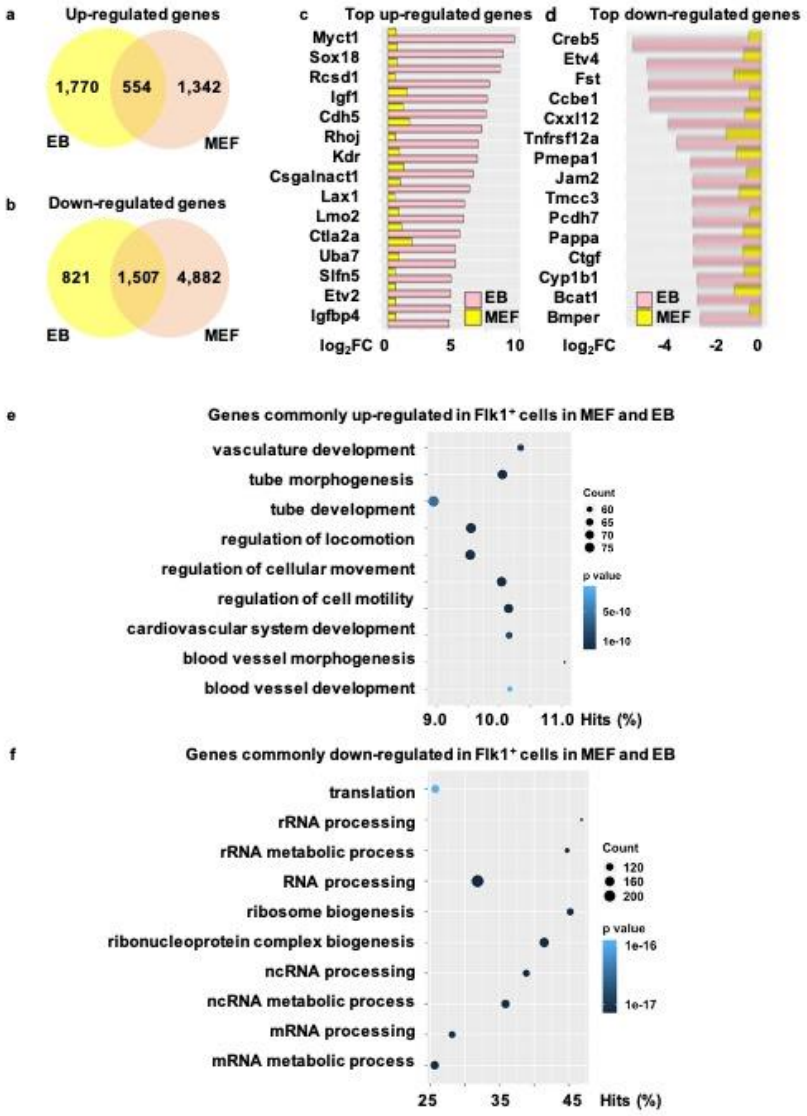
Extended Data Fig. 2. Expression of endothelial transcripts in the FLK1$^+$ cell population at day 7 post-ETV2 induction during MEF reprogramming. **(a)** The violin plots show the scaled expression levels of endothelial markers such as Etv2, Emcn, Lmo2, Flk1/Kdr, Cdh5 and Sox18 in MEFs, day 1, day 2, day 7 post-ETV2 induction, as well as the FLK1$^+$ cells from day 7. The y-axis indicates the gene expression levels scaled and normalized by Seurat. **(b)** The violin plots show the scaled expression levels of endothelial markers in seven cell clusters. The y-axis indicates the gene expression levels scaled and normalized by Seurat. The one-sided enrichment test was used to evaluate the significance of pathway enrichment. The p-value adjustment was performed by B-H Procedure. **(c)** The biological processes that are significantly associated with the up-regulated in genes in cluster 1 (undifferentiated MEFs) compared with the rest of the cell populations. There are 3,562, 948, 2,936, 7,202 and 827 single cells from undifferentiated MEFs, MEFs with day 1, day 2, day 7 post-ETV2 induction, and FLK1$^+$ cell population of day 7 post-ETV2 induction, respectively. **(d)** GSEA plot indicates significant upregulation of the inflammatory response in MEFs (cluster 1). The y-axis representing the enrichment score (ES) for each gene and x-axis indicates the gene rank in the ordered list. The default GSEA permutation test was used to evaluate the significance of gene set enrichment. No *p*-value adjustment was performed. **(e)** Heatmap representing the gene expression levels scaled by Seurat for upregulated (132) and downregulated (blue) genes in cluster 1 and cluster 2. **(f-g)** The bar plots show top 10 significant pathways for cluster 1 and cluster 2 for MEFs. The y-axis represents the p adjusted values obtained from over-representation test in Gene Ontology enrichment analysis. The one-sided enrichment test was used to evaluate the significance of Gene Ontology enrichment. The p-value adjustment was performed by B-H Procedure.



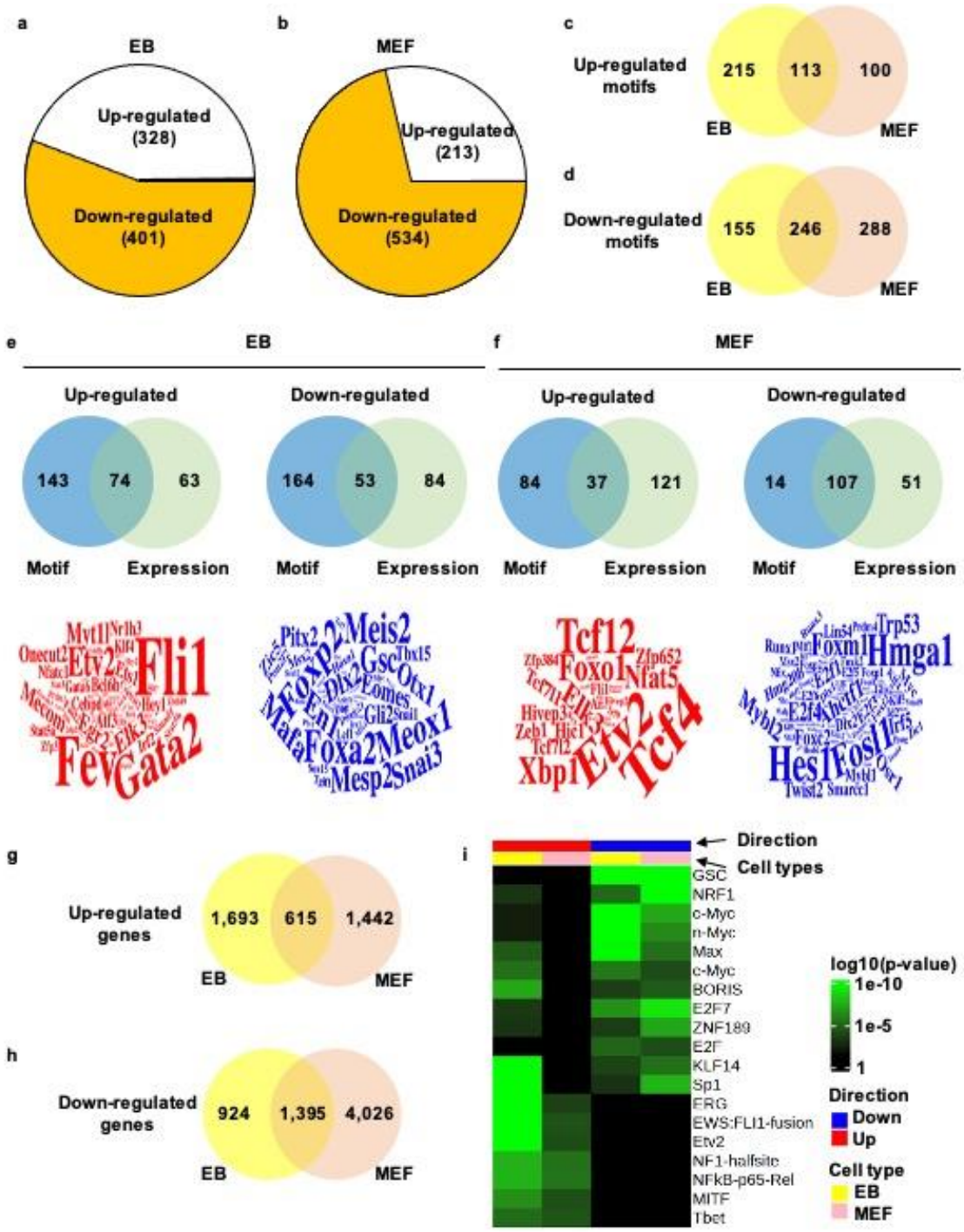
Extended Data Fig. 3. Expression profile of immune response related genes and significance of pathways in reprogrammed MEFs are upregulated post induction of ETV2. (a) The UMAP shows expression profiles of Tlr3, Nfkb1 and Vav3, Cd38 and Abl1 (members of B cell receptor signaling pathway) in undifferentiated MEFs and post Etv2 induction day 1, day2, day 7 and Flk1+ cells at day 7. **(b-c)** The bar plot shows immune response related pathways significantly upregulated in **(b)** Flk1+ cells from day 7 and **(c)** day 1 post Etv2 induction compared to MEFs. The y-axis indicates the p-value showing significance for each pathway obtained from Gene Ontology enrichment. The one-sided enrichment test was used to evaluate the significance of Gene Ontology enrichment. The p-value adjustment was performed by B-H Procedure.



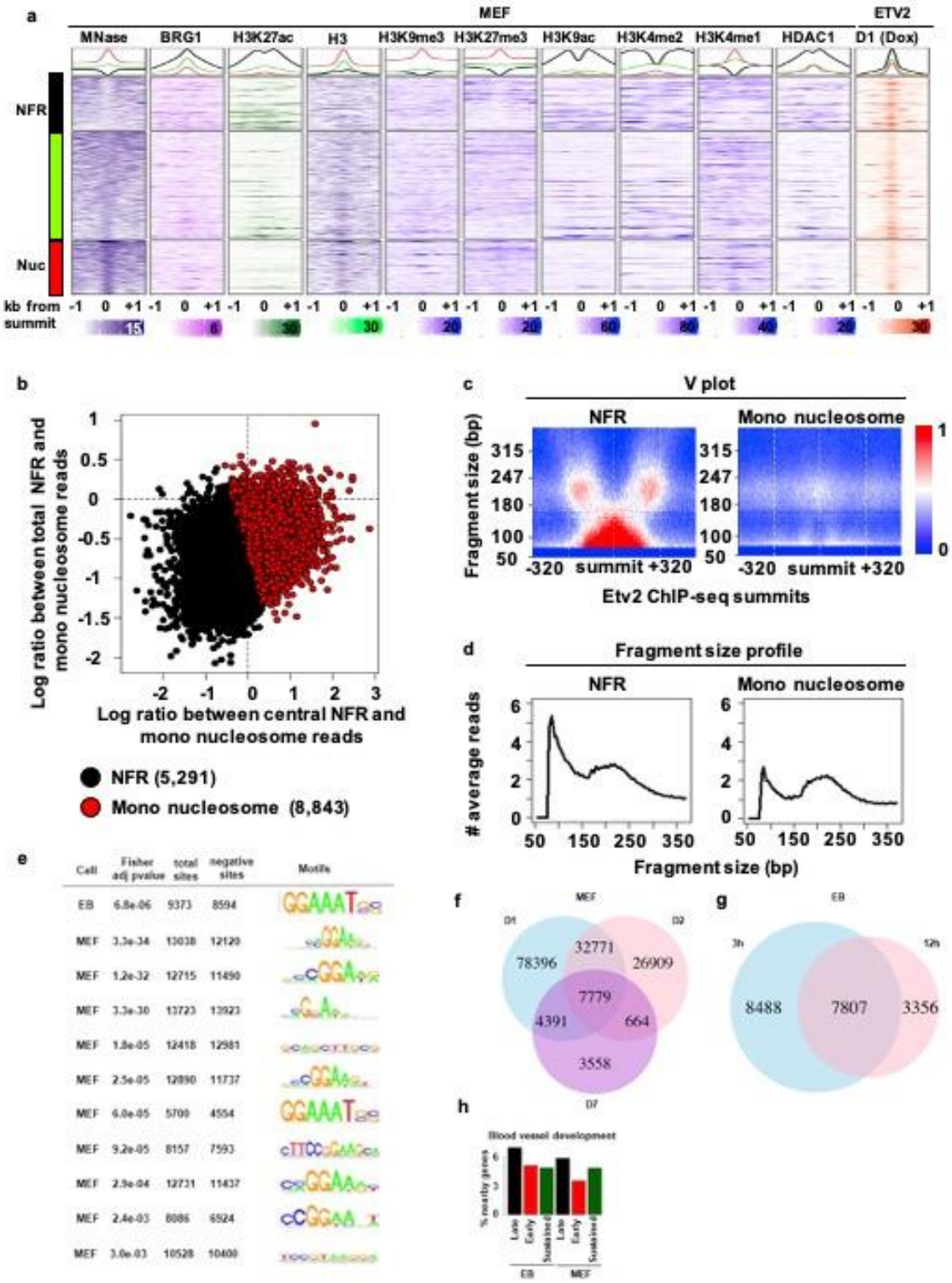
Extended Data Fig. 4. ETV2 overexpression promotes endothelial lineage development in *iHA-Etv2* ES/EBs. (a) Etv2 endogenous expression during differentiation of ES/EBs (n=3 biological replicates; one-way ANOVA with multiple comparison ***P=2x10-4). **(b-c)** FACS analysis of *iHA-Etv2* EBs after 3h (D2.125) or 12h (D2.5) of the overexpression of ETV2 (+Dox). Note a significant induction of the endothelial lineage (FLK1^{+} cells) following 12h of ETV2 overexpression (n=3 biological replicates; 2way ANOVA with multiple comparison ****P<1x10-4). **(d-h)** qPCR experiments showed increased expression levels of endothelial genes **(d)** *Etv2* (****P<1x10-4, ***P=0.0004)*,* **(e)** *Cdh5* (***P=0.0001), **(f)** *Lmo2* (*P=0.041, ****P<1x10-4), **(g)** *Flk1* (***P=0.0002, ****P<1x10-4) and **(h)** *Sox18* (***P=0.0005) following the induction of ETV2 (+Dox). (n=3 biological replicates; one-way ANOVA with multiple comparison). Data are presented as mean±SEM



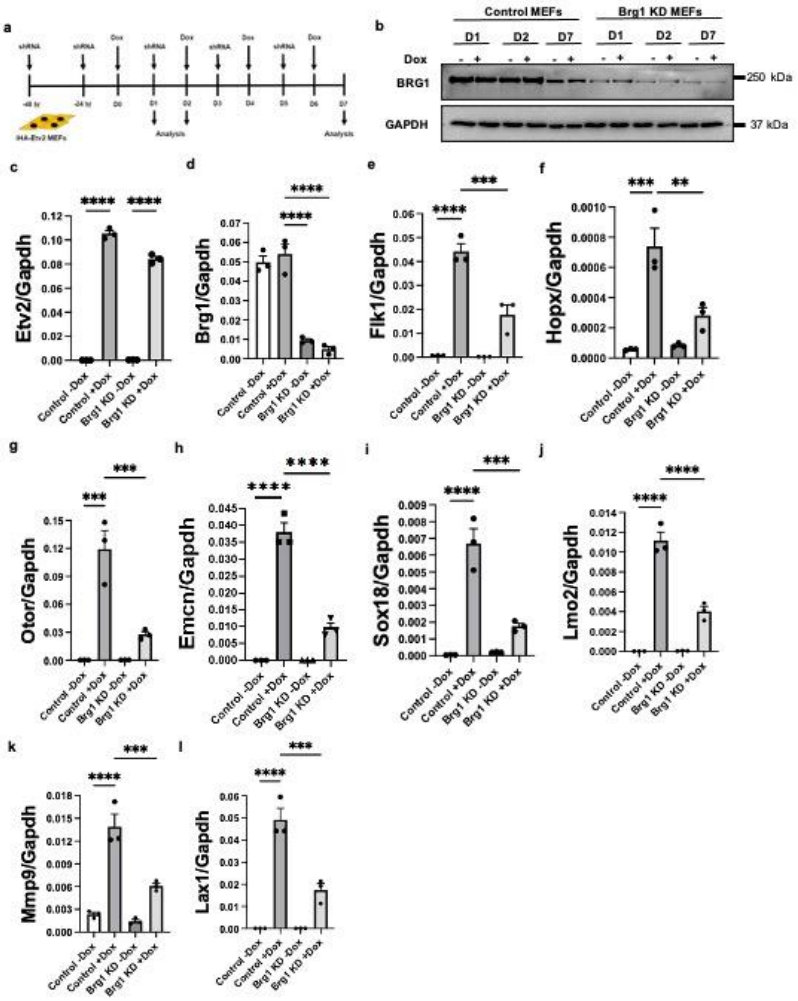
Extended Data Fig. 5. Commonly up- and down-regulated genes in FLK1+ cell populations from ETV2 induced ES/EB differentiation and MEF reprogramming. (50) The Venn diagrams show the overlap of commonly up- and down-regulated genes during EB differentiation and MEF reprogramming**.** **(c-d)** Top commonly up- and down-regulated genes during EB differentiation and MEF reprogramming. The y-axis shows the log2 fold change of gene expression between FLK1^{+} cell populations and the baseline conditions (D2.5 EB in ES/EB differentiation, and undifferentiated MEFs during MEF reprogramming). **(e-f)** The pathways that are significantly associated with commonly up- and down-regulated genes during ES/EB differentiation and MEF reprogramming are highlighted. The one-sided enrichment test was used to evaluate the significance of Gene Ontology enrichment. The p-value adjustment was performed by B-H Procedure.


a
EB
Up-regulated (328)
Down-regulated (401)
b
MEF
Up-regulated (213)
Down-regulated (534)
c
Up-regulated motifs
215
113
100
EB
MEF
d
Down-regulated motifs
155
246
288
EB
MEF
e
EB
Up-regulated
143
74
63
Motif
Expression
Down-regulated
164
53
84
Motif
Expression
f
MEF
Up-regulated
84
37
121
Motif
Expression
Down-regulated
14
107
51
Motif
Expression
g
Up-regulated genes
1,693
615
1,442
EB
MEF
h
Down-regulated genes
924
1,395
4,026
EB
MEF
i
Direction
Cell types
GSC
NRF1
c-Myc
n-Myc
Max
c-Myc
BORIS
E2F7
ZNF189
E2F
KLF14
Sp1
ERG
EWS:FLI1-fusion
Etv2
NF1-halfsite
NFkB-p65-Rel
MITF
Tbet
log10(p-value)
1e-10
1e-5
1
Direction
Down
Up
Cell type
EB
MEF

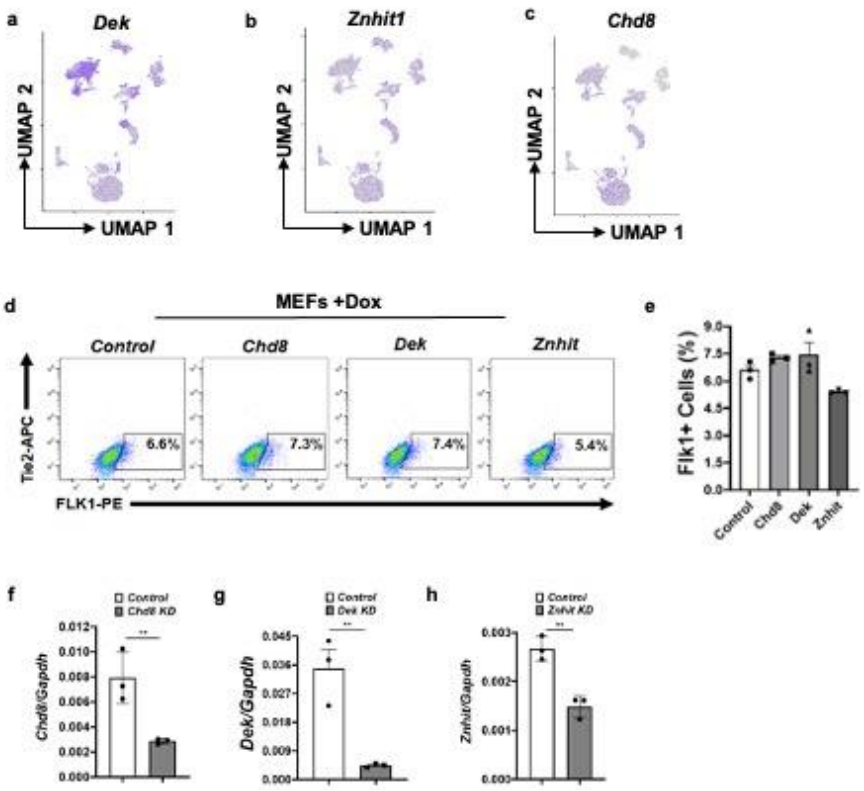
Extended Data Fig. 6. Combined RNA-seq and ATAC-seq analysis during EB and MEF reprogramming. (a) The number of transcription factors whose motifs associated chromatin accessibility were significantly increased or decreased in the FLK1$^+$ cell populations at 12 hours post-Etv2 induction compared with D2.5 EBs. **(b)** The number of transcription factors whose motifs associated chromatin accessibility were significantly increased or decreased in the FLK1$^+$ cell population at day 7 post-ETV2 induction compared with undifferentiated MEFs. **(c-d)** The number of transcription factors whose motif associated chromatin accessibility that were commonly increased or decreased during EB and MEF reprogramming. **(e)** The transcription factors whose RNA-seq expression levels and motifs associated chromatin accessibility that were both up-regulated or down-regulated during EB reprogramming (FLK1$^+$ cell from EBs at 12 hours post-ETV2 induction vs. day 2.5 EB). **(f)** The transcription factors whose RNA-seq expression levels and motifs associated chromatin accessibility that were both up-regulated or down-regulated during MEF reprogramming (FLK1$^+$ cell population at day 7 post-ETV2 induction vs. undifferentiated MEFs). **(g-h)** The commonly up- and down-regulated genes between EBs and MEFs. **(i)** The transcription factor motifs that are significantly enriched in 5k region surrounding the transcription start sites of the commonly up- and down-regulated genes in EBs and MEFs. The TF enrichment was evaluated by the two-sided χ^2 test in chromVAR. Data shown in 6i represent the average of two biological replicates. The p-value adjustment was performed by B-H Procedure.


a
MEF
ETV2
MNase
BRG1
H3K27ac
H3
H3K9me3
H3K27me3
H3K9ac
H3K4me2
H3K4me1
HDAC1
D1 (Dox)
NFR
Nuc
kb from summit
b
Log ratio between total NFR and mono nucleosome reads
Log ratio between central NFR and mono nucleosome reads
NFR (5,291)
Mono nucleosome (8,843)
c
V plot
NFR
Mono nucleosome
Fragment size (bp)
summit
Etv2 ChIP-seq summits
d
Fragment size profile
NFR
Mono nucleosome
average reads
Fragment size (bp)
e
Cell
Fisher adj pvalue
total sites
negative sites
Motifs
f
MEF
D1
D2
D7
78396
32771
26909
7779
4391
664
3558
g
EB
3h
12h
8488
7807
3356
h
Blood vessel development
% nearby genes
Late
Early
Sustained
EB
MEF

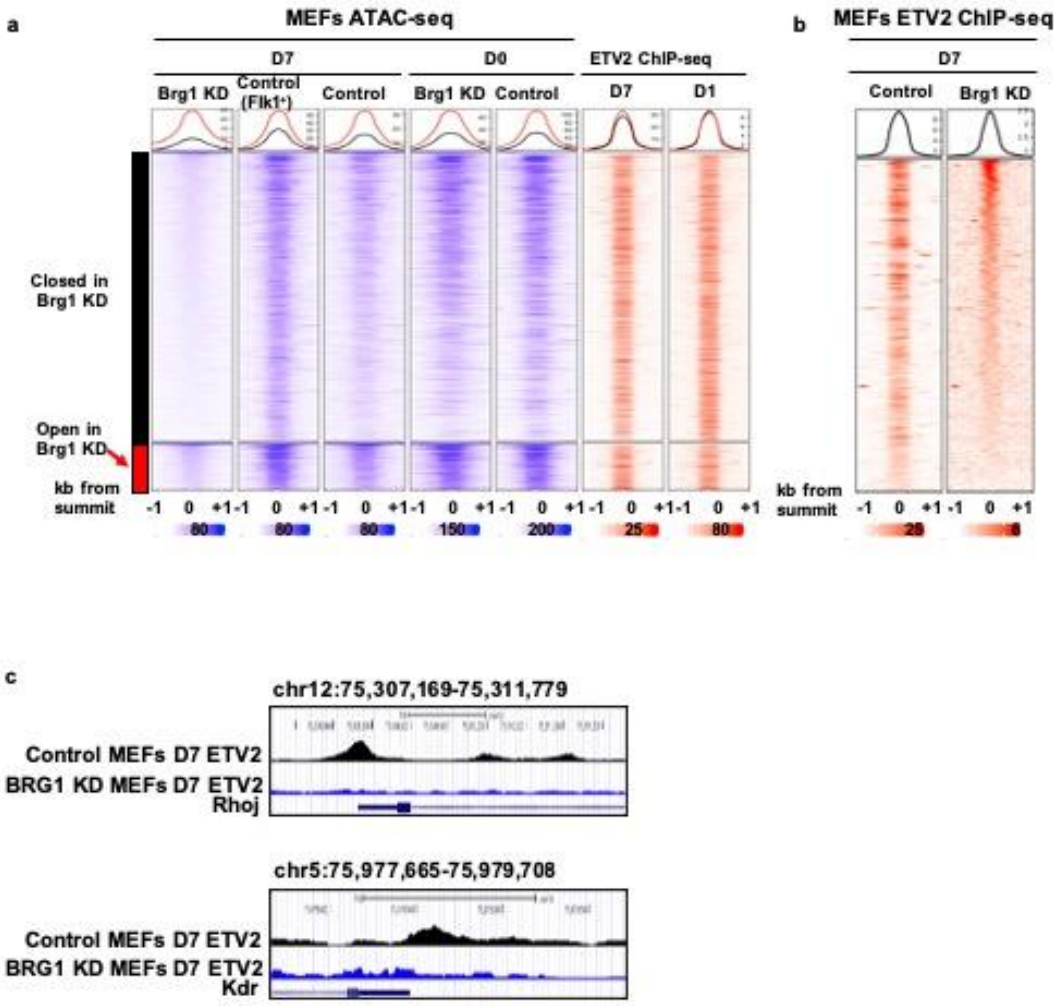
Extended Data Fig. 7. The ETV2 bound sites at day 1 post-Etv2 induction in MEFs target the nucleosomes and the analysis of ETV2 ChIP-seq peaks during EB and MEF reprogramming. (a) The MNase-seq, BRG1, H3K27ac, H3, H3K9me3, H3K27me3, H3K9ac, H3K4me3, H4K7me1 and Hdac1 ChIP-seq signals surrounding the ETV2 bound sites at day1 post-ETV2 induction during MEF reprogramming. The ETV2 binding sites were split into nucleosome and nucleosome free region (NFR) according to the MNase-seq signals in undifferentiated MEFs. Data shown here represent the average of two biological replicates. **(b)** The latent representation of ATAC-seq V-plots (-320bp to + 320bp) where the centers are nucleosome free or occupied by mono nucleosome. **(c)** The aggregated ETV2 bound sites centric V-plot whose centers were occupied by mono nucleosomes or nucleosome free. **(d)** The fragment size profile of ETV2 bound sites centric region (-320bp to +320bp) where the centers are nucleosome free or occupied by mono nucleosomes. **(e)** Motif analysis of ETV2 bound sites during EB reprogramming (3 hours post-ETV2 induction) and MEF reprogramming (24 hours post-ETV2 induction). The table shows the significantly enriched motifs in ETV2 bound sites during EB and MEF reprogramming. (**f)** The overlap of ETV2 bound sites at day 1, day 2 and day 7 post-ETV2 induction during MEF reprogramming. **(g)** The overlap of ETV2 bound sites at 3 hours and 12 hours post-ETV2 induction during EB reprogramming. **(h)** The bar plot shows the percent of genes located near the late, early and sustained ETV2 bound sites related to blood vessel development.



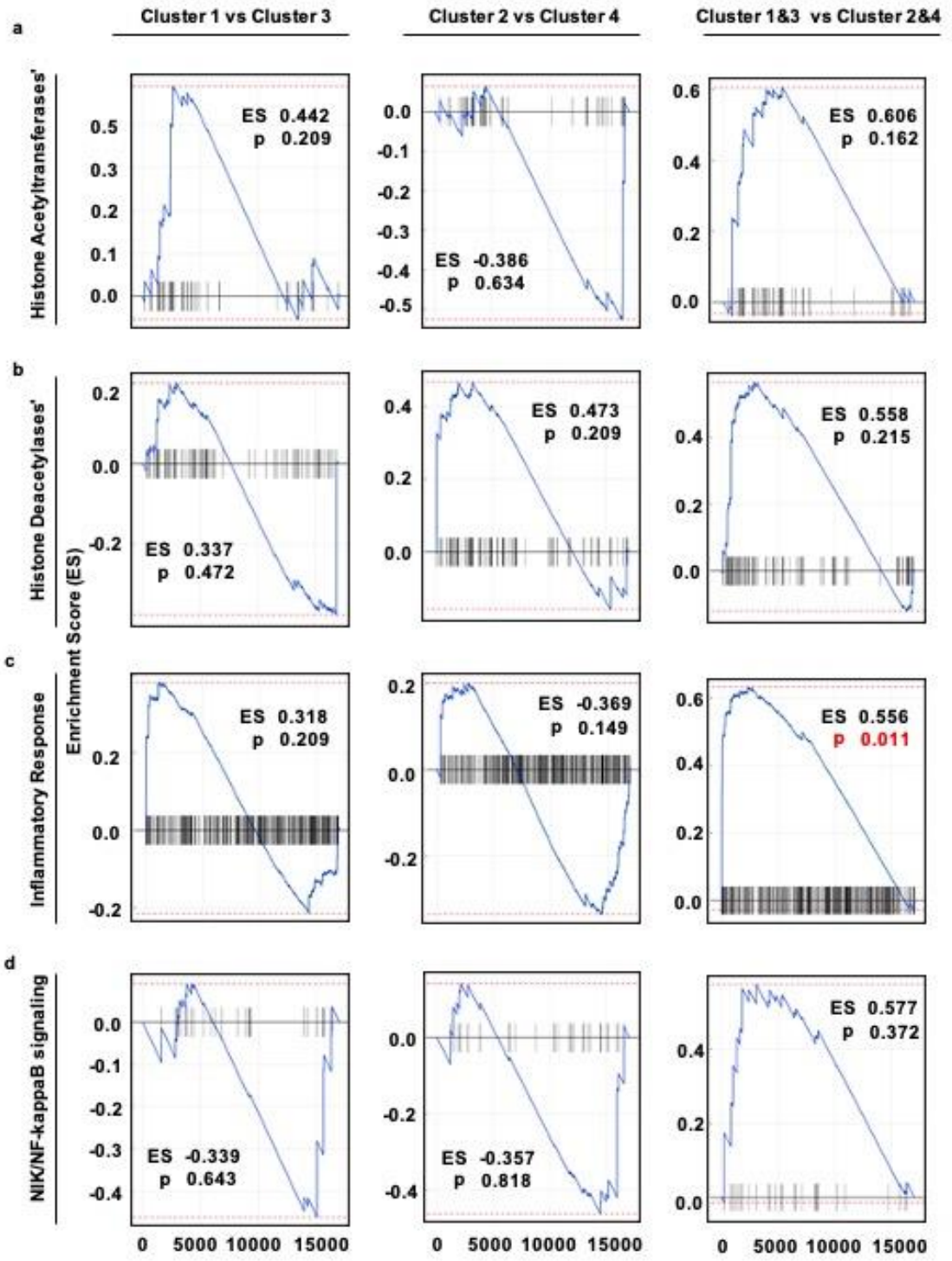
Extended Data Fig. 8. *Brg1* knockdown in *iHA-Etv2* MEFs using shRNA lentiviral particles. (a) Schematic diagram of shRNA lentiviral knockdown of *Brg1* in *iHA-Etv2* MEFs. Briefly, MEFs were exposed to shRNA particles 72 hrs before reprogramming was started and cells were collected for analysis and sequencing at various time points throughout the reprogramming process (D1, D2 and D7). shRNA particles were added throughout the reprogramming process to ensure BRG1 expression was not increased. **(b)** Western blot analysis of BRG1 expression in *iHA-Etv2* MEFs exposed to normal reprogramming media versus media with shRNA particles against Brg1 [representative blots (b) from 3 independent experiments with similar results]. **(c-l)** Compared to control, shRNA knockdown of **(d)** *Brg1* (****P<1x10-4) in the context of **(c)** *Etv2* overexpression (****P<1x10-4) leads to a significant decrease in the expression of **(e)** *Flk1* (****P<1x10-4, ***P=0.0005), **(f)** *Hopx* (***P=0.0004, **P=0.0052), **(g)** *Otor* (***P=0.0001, ***P=0.0009), **(h)** *Emcn* (****P<1x10-4), **(i)** *Sox18* (****P<1x10-4, ***P=0.0003), **(j)** *Lmo2* (****P<1x10-4), **(k)** *Mmp9* (****P<1x10-4, ***P=0.001) and **(l)** *Lax1* (****P<1x10-4, ***P=0.0003) transcripts, which are upregulated at D7 following overexpression of ETV2 in *iHA-Etv2* MEFs (n=3 biological replicates; one-way ANOVA with multiple comparison). Data are presented as mean±SEM.



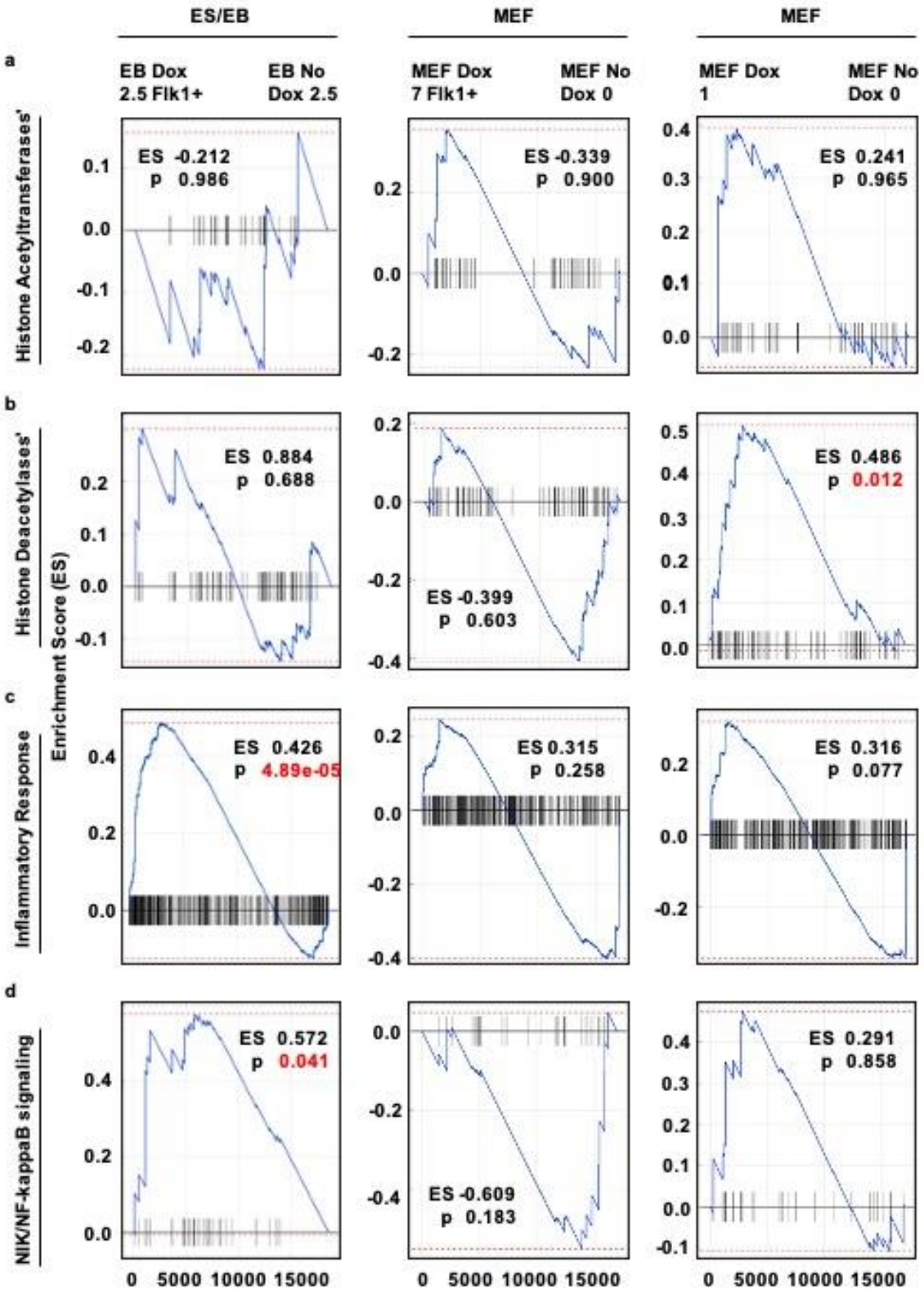
Extended Data Fig. 9. *Dek*, *Znhit1* and *Cdh8* knockdown does not impact ETV2 mediated endothelial reprogramming. (a-c) The expression profiles of **(a)** *Dek*, **(b)** *Znhit1* and **(c)** *Cdh8* expression during MEF reprogramming. **(d-e)** *iHA-Etv2* MEFs were exposed to *Chd8, Dek* and *Znhit* shRNA particles 72 hrs before ETV2 mediated reprogramming was started and cells were collected for analysis 24 hrs (D1) following ETV2 overexpression. Flow cytometry analysis shows that knockdown of *Chd8, Dek* and *Znhit* does not affect MEF reprogramming mediated by ETV2 (n=3 biological replicates). **(f-h)** qPCR analysis shows efficient knockdown of *Chd8* (**P=0.0068)*, Dek* (**P=0.0037) and *Znhit* (**P=0.0017) in *iHA-Etv2* MEFs (n=3 biological replicates; one-tailed unpaired t test ****P<1.0x10-4). Data are presented as mean±SEM.



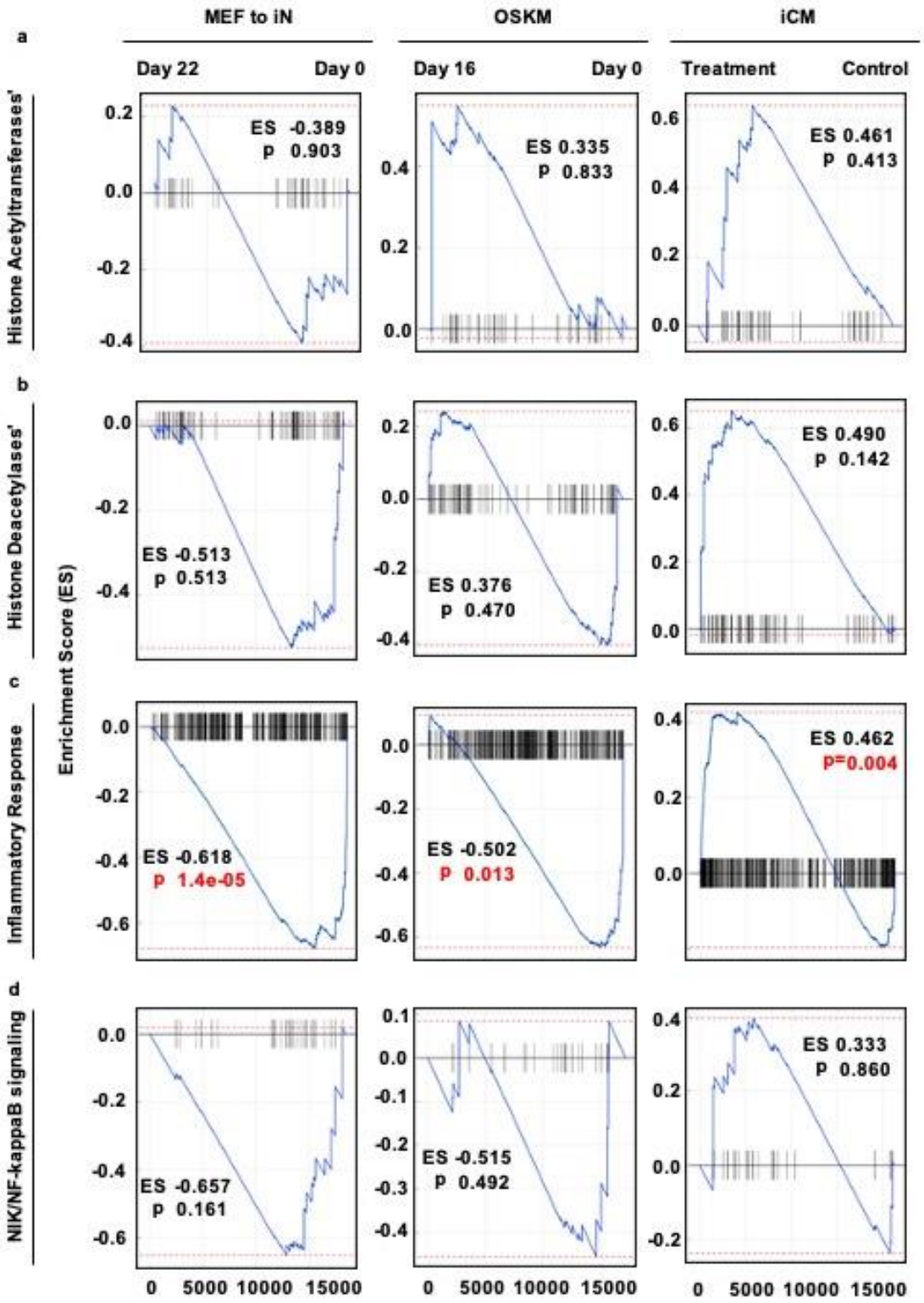
Extended Data Fig. 10. ETV2 requires BRG1 to activate downstream genes during reprogramming. (a) The heatmap shows the ATAC-seq signal surrounding the summit of 12,170 sustained ETV2 ChIP-seq peaks that were present at day 1 and day 7 post-induction of ETV2 in control MEFs (the sustained Etv2 peaks). The ATAC-seq data include undifferentiated MEFs, day 7 post-ETV2 induction, and FLK1$^+$ cells from day 7 post-ETV2 induction in MEFs. We also include the ATAC-seq data from undifferentiated *Brg1* KD (knockdown) in MEFs and day 7 post-ETV2 induction with *Brg1* KD in MEFs. The sustained ETV2 peaks were divided into two groups: open (132) or closed (187) at day 7 post- induction of *Brg1* KD in MEFs. **(b)** Heatmap shows ETV2 ChIP-seq signal surrounding 4,965 sustained ETV2 ChIP-seq peaks present in the D7 post ETV2 induction in WT MEFs and Brg1 KD MEFs. **(c)** UCSC genome browser tracks show the ETV2 ChIP-seq signal surrounding Etv2 ChIP-seq peaks at the promoter region of two endothelial genes Rhoj and Kdr. Data shown in 10a and 10b represent the average of two biological replicates.


a
Cluster 1 vs Cluster 3
Cluster 2 vs Cluster 4
Cluster 1&3 vs Cluster 2&4
Histone Acetyltransferases'
ES 0.442
p 0.209
ES -0.386
p 0.634
ES 0.606
p 0.162
b
Histone Deacetylases'
ES 0.337
p 0.472
ES 0.473
p 0.209
ES 0.558
p 0.215
c
Inflammatory Response
ES 0.318
p 0.209
ES -0.369
p 0.149
ES 0.556
p 0.011
d
NIK/NF-kappaB signaling
ES -0.339
p 0.643
ES -0.357
p 0.818
ES 0.577
p 0.372
Enrichment Score (ES)
0 5000 10000 15000

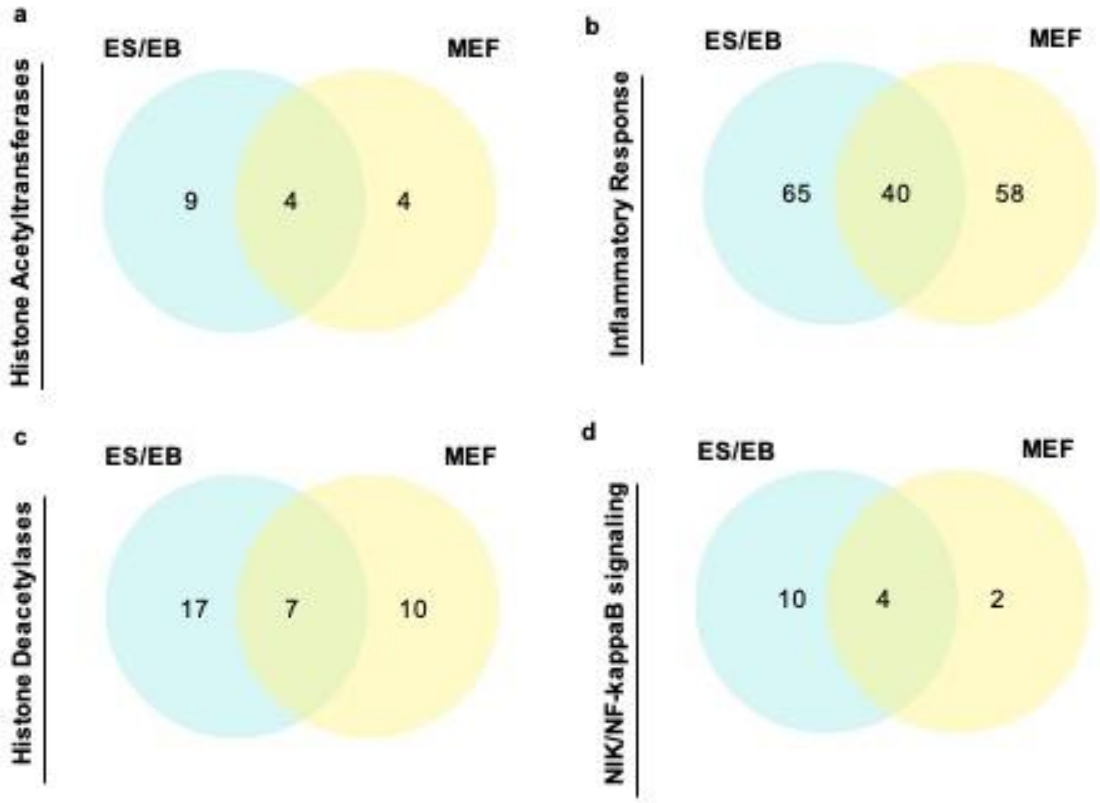
Supplementary Figure 1. ETV2 requires BRG1 to activate downstream genes during reprogramming. GSEA plots comparing clusters 1, 2, 3 and 4 out of the seven clusters in Figure 1d demonstrates that the Inflammatory response is significantly upregulated for clusters 1 and 3 compared to clusters 2 and 4. (a-d) The plots show enrichment scores for each pathway, with the y-axis representing the enrichment score (ES) for each gene and x-axis indicates the gene rank in the ordered list. The highlighted p value in red represents the degree of significance ($p < 0.05$). The default GSEA permutation test was used to evaluate the significance of gene set enrichment. No *p*-value adjustment was performed.


ES/EB
MEF
MEF
a
EB Dox 2.5 Flk1+
EB No Dox 2.5
MEF Dox 7 Flk1+
MEF No Dox 0
MEF Dox 1
MEF No Dox 0
Histone Acetyltransferases'
ES -0.212
p 0.986
ES -0.339
p 0.900
ES 0.241
p 0.965
b
Histone Deacetylases'
ES 0.884
p 0.688
ES -0.399
p 0.603
ES 0.486
p 0.012
c
Inflammatory Response
Enrichment Score (ES)
ES 0.426
p 4.89e-05
ES 0.315
p 0.258
ES 0.316
p 0.077
d
NIK/NF-kappaB signaling
ES 0.572
p 0.041
ES -0.609
p 0.183
ES 0.291
p 0.858
0 5000 10000 15000
0 5000 10000 15000
0 5000 10000 15000

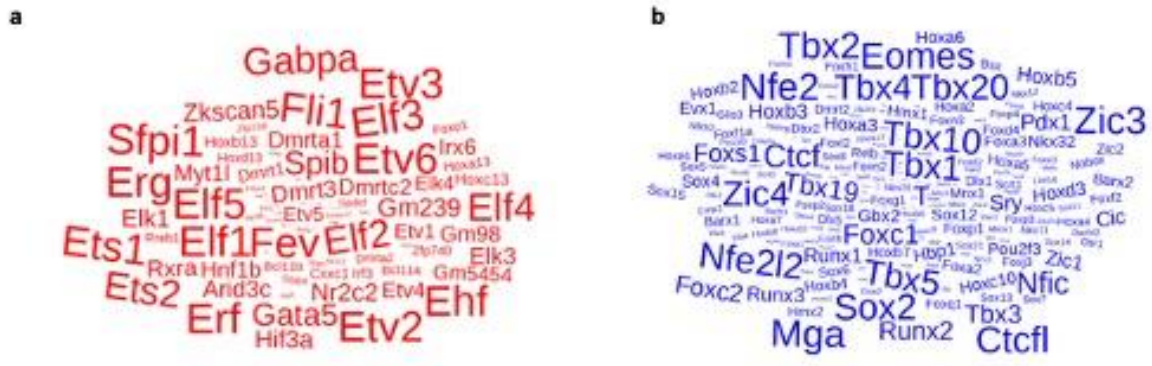
Supplementary Figure 2. Gene set enrichment analysis (GSEA) for HATs, HDACs, Inflammatory response and NF-kappaB signaling pathway during ES/EB differentiation and MEF reprogramming. (a-d) GSEA plots showing the enrichment score profile where the positive enrichment score represents enrichment in Flk1+ cells from day 2.5 of Etv2 induced EBs compared to controls (uninduced EBs) as well as enrichment in day 7 and day 1 post Etv2 induction compared to MEFs. The y-axis indicates the enrichment score (ES) and x-axis indicates the gene rank in the ordered list. The highlighted p value in red represents the degree of significance ($p < 0.05$). **(a)** HATs are not significant in both ES/EB and MEFs. **(b)** HDACs are significantly upregulated in MEFs day 1 post Etv2 induction. **(c)** Inflammatory and **(d)** NIK-NF-kappaB signaling pathways are significantly upregulated in day 2.5 Flk1+ cells in Etv2 induced EB compared to day 2.5 uninduced EB during ES/EB differentiation. The default GSEA permutation test was used to evaluate the significance of gene set enrichment. No *p*-value adjustment was performed.


a
MEF to iN
OSKM
iCM
Day 22
Day 0
Day 16
Day 0
Treatment
Control
Histone Acetyltransferases'
ES -0.389
p 0.903
ES 0.335
p 0.833
ES 0.461
p 0.413
b
Histone Deacetylases'
ES -0.513
p 0.513
ES 0.376
p 0.470
ES 0.490
p 0.142
c
Inflammatory Response
ES -0.618
p 1.4e-05
ES -0.502
p 0.013
ES 0.462
p=0.004
d
NIK/NF-kappaB signaling
ES -0.657
p 0.161
ES -0.515
p 0.492
ES 0.333
p 0.860
Enrichment Score (ES)
0
5000
10000
15000

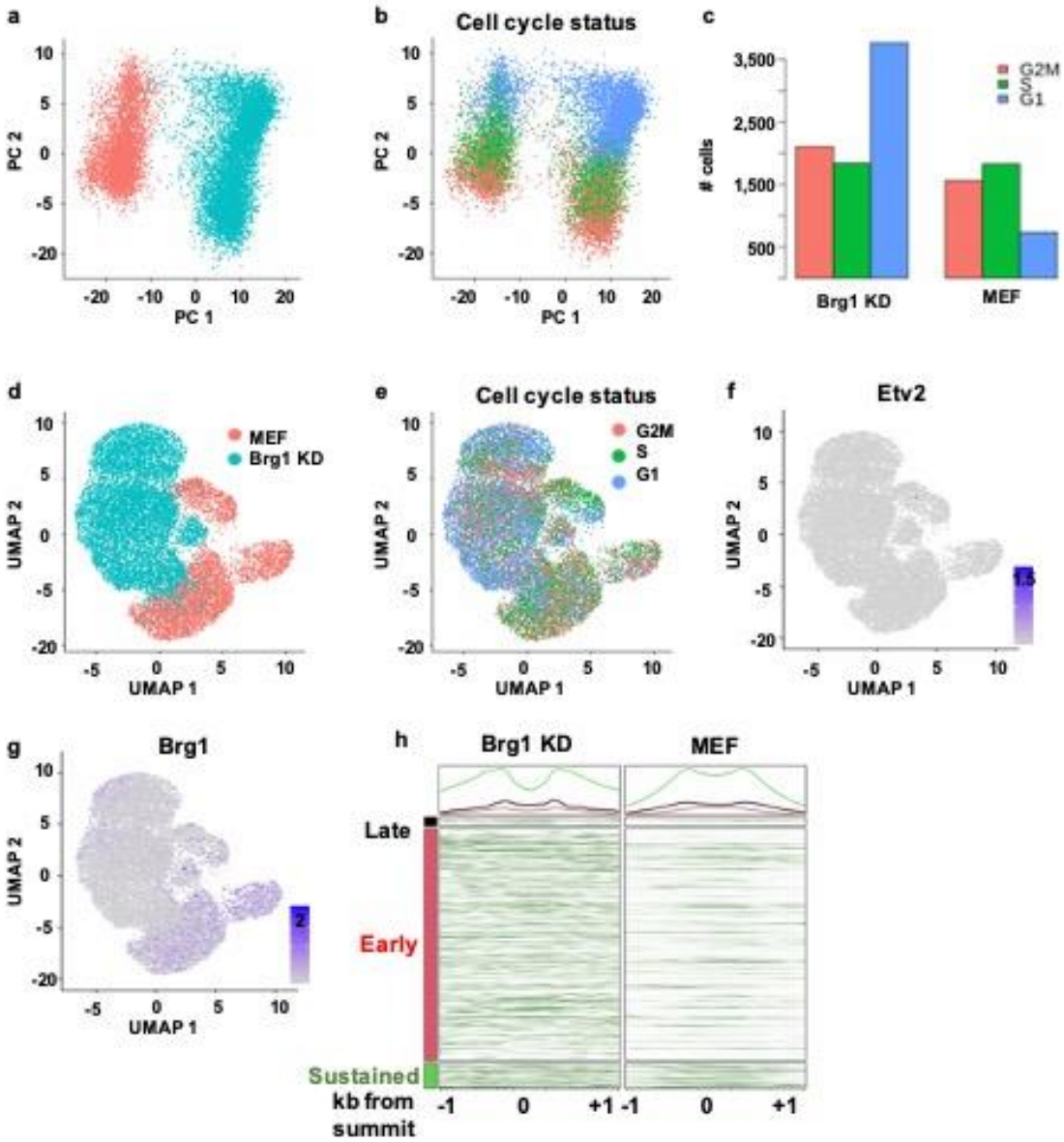
Supplementary Figure 3. Gene set enrichment analysis (GSEA) plots obtained from previously published scRNA-seq data for MEF reprogramming shows significant upregulation of the Inflammatory response pathway during GMT induced cardiac reprogramming. (a-d) GSEA plots for HATs, HDACs, Inflammatory response and NIK-NF-kappaB signaling pathways. **(c)** Inflammatory response is significantly downregulated in day 16 OSKM driven reprogramming and day 22 Ascl1 induced neuronal reprogramming compared to day 0 undifferentiated MEFs. Significant upregulation of the Inflammatory response is observed in the GMT induced cardiac reprogramming compared to the control. The default GSEA permutation test was used to evaluate the significance of gene set enrichment. No *p*-value adjustment was performed.



Supplementary Figure 4. Gene ontology annotation identifies commonly up-regulated genes in both ES/EB differentiation and MEF reprogramming following ETV2 overexpression. (a-d) Venn diagram representing the overlap of common genes upregulated in ES/EB and MEF reprogramming post Etv2 induction, also overlapping with the genes from gene ontology terms for **(a)** histone acetyltransferase **(b)** histone deacetylase **(c)** inflammatory response related genes and **(d)** NF-kappaB signaling pathways.



Supplementary Figure 5. The transcription factors whose motif associated chromatin accessibility and expressions were consistently changed in both EB and MEF on Etv2 induction. **(a)** The transcription factors whose motif associated chromatin accessibility and expressions were up-regulated in both EB and MEF on Etv2 induction. **(b)** The transcription factors whose motif associated chromatin accessibility and expressions were down-regulated in both EB and MEF on Etv2 induction.



Supplementary Figure 6. Comparison of the single cell expression profiles of undifferentiated MEFs and *Brg1* KD MEFs. (a-c) The PCA analysis showed that the undifferentiated MEFs and *Brg1* KD MEFs have significantly different cell cycle programs, where the cells in the G1 phase are increased with *Brg1* KD in MEFs. **(d-e)** The correction of cell cycle effects when combining the single cell RNA-seq data of undifferentiated MEFs and *Brg1* KD MEFs. **(f-g)** The expression levels of *Etv2* and *Brg1* in cell-cycle effected adjusted single cell RNA-seq data from undifferentiated MEFs and *Brg1* KD MEFs**. (h)** The H3K27ac ChIP-seq signals of undifferentiated MEFs and *Brg1* KD MEFs surrounding the early, late and sustained ETV2 bound sites in ETV2 induced MEF reprogramming.

Supplementary Table 1 – Taqman qPCR probes

No	Gene	Cat#	Concentration	Species	Fluorophore
1	Mmp9	Mm00442991_m1	20X	mouse	FAM
2	Etv2	Mm01176581_g1	20X	mouse	FAM
3	Gapdh	4352339E	20X	mouse	VIC
4	Lmo2	Mm00493153_m1	20X	mouse	FAM
5	Flk1 (Kdr)	mm00440099_m1	20X	mouse	FAM
6	Cd31	Mm01242576_m1	20X	mouse	FAM
7	Tie2	mm01256892_ m1	20X	mouse	FAM
8	Cdh5	mm00486938_m1	20X	mouse	FAM
9	Sox18	Mm00656049_gH	20X	mouse	FAM
10	Brg1 (Smarca4)	Mm01151944_m1	20X	mouse	FAM
11	Hopx	Mm00558630_m1	20X	mouse	FAM
12	Otor	Mm00498571_m1	20X	mouse	FAM
13	Emcn	Mm00497495_m1	20X	mouse	FAM
14	Lax1	Mm00556050_m1	20X	mouse	FAM
15	Chd8	Mm01316316_m1	20X	mouse	FAM
16	Dek	Mm01351566_m1	20X	mouse	FAM
17	Znhit	Mm01201686_m1	20X	mouse	FAM

Supplementary Table 2 – Re-ChIP primers

No	Gene	Forward Primer	Reverse Primer
1	Lmo2	TACAGGAGAAGGAGGGCTGA	CAATGTCAGGCAGCAGGTTG
2	Flt1	GCCTCCATTCCACCACTTGA	GCCTCCCTCAAGACTGTTCC
3	Tcf12	ACAGTGTCTCAGAGGCTCCC	CCCCAGTTAGCCCTGTTTCC

Supplementary Table 3 – ChIP qPCR primers for ELK3

No	Locus	Forward Primer	Reverse Primer
1	Chr12	CCCAGCCTCACTATGTCCAG	GGATTTGCTCTCGTGGTCTT
2	Chr13	ATAAGCAAGCTGGGACAGATG	TGACAAAACTTCTTTTCCCTCA

CHAPTER 4: ETV2-*Rhoj* CASCADE REGULATES ENDOTHELIAL PROGENITOR CELL MIGRATION DURING EMBRYOGENESIS

Introduction

ETV2, an Ets-related transcription factor, is expressed transiently in primitive angioblasts and regulates hematoendothelial (HE) lineage specification during embryogenesis (113). Genetic deletion of *Etv2* results in embryonic lethality by E9.5, due to the complete absence of HE lineages (46, 48). Studies have demonstrated that ETV2 regulates multiple cellular processes such as proliferation, differentiation and commitment of HE progenitor cells (46, 51, 109, 113). These functional roles are achieved through the interactions of ETV2 with several factors including GATA2, FOXC2 and others during HE development (52, 137). Although these studies have provided important insights regarding the role of ETV2 in endothelial precursors and their regulation, the mechanisms whereby ETV2 regulates these progenitors are unclear.

Endothelial precursors or angioblasts from the lateral plate mesoderm converge in the midline of developing embryos to form the primary vascular plexus (188). Convergence of these progenitors is dependent on their migratory phenotype, which is regulated by multiple signaling pathways and transcription factors (188). The Rho GTPase family and related members are involved in endothelial migration and guidance (189). Rho GTPases are activated by binding to GTP in exchange for GDP, a mechanism mediated by guanine nucleotide exchange factors (189). The activity of Rho GTPases is regulated in a spatio-temporal fashion during vasculogenesis under normal and pathogical states (190, 191).

RHOJ, a member of the Rho GTPase subfamily, is highly expressed in endothelial cells and is required for focal adhesion numbers and actomyosin contractility (192). Several studies have demonstrated that RHOJ interacts with the GIT-PIX complex to regulate focal adhesion disassembly in endothelial cells (189). Furthermore, RHOJ has been shown to regulate the activity of CDC42 and RAC1 during lamellipodia formation (193). The knockout of *Rhoj* results in delayed radial growth and defective vascular structures (189). Similarly, the conditional knockout of *Rhoj* in the endothelial lineage results in reduced growth and abnormal vascular development at E10.5 (194). These studies support an important role for RHOJ in endothelial lineages but the mechanisms regulating RHOJ expression in the endothelial lineage is unclear.

In the present study, we define a new role for ETV2 as a regulator of cell migration. We demonstrate that ETV2 upregulates cellular migratory networks by binding to and enhancing chromatin accessibility of genes that govern cell migration. Furthermore, we show that ETV2 acts as an upstream regulator of *Rhoj* and regulates its expression in endothelial progenitors. Collectively, our genomics, biochemical, molecular and rescue experiments define an ETV2-*Rhoj* cascade that functions to regulate cell migration.

Material and Methods

Embryo harvesting and microscopy. All animal studies were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Minnesota. All methods were performed in accordance with the relevant guidelines and regulations. Time-mated pregnant (*Etv2*-promoter driving EYFP transgenic lines (50)) mice were used for embryo harvest, imaging and FACS-sorting experiments at E8.5. For imaging experiments, these embryos were fixed for 1 hour at 4°C in 4% paraformaldehyde (94), washed twice in PBS and imaged on a Zeiss Axio Imager inverted microscope and processed using Adobe Photoshop CS6 software. For $EYFP^-$ and EYFP+ cell-sorting experiments, the E8.5 embryos were dissociated using 0.25% trypsin for 2-3 min at 37ºC and the cells were resuspended in 2%FBS/PBS buffer. Live cells were gated using propidium iodide (PI) staining.(109)

***iHA-Etv2* embryonic stem (ES) cell/embryoid body (134) and mouse embryonic fibroblast (MEF) cultures.** Wildtype ESCs and doxycycline inducible *iHA-Etv2* ESCs (113) were cultured in media containing 15% fetal bovine serum (FBS), 2 mM Glutamax, 1X penicillin/streptomycin, 0.1 mM β-mercaptoethanol, and 1,000 U/mL LIF (Millipore), at 37°C in 5% CO2 together with irradiated embryonic fibroblasts as the feeder layer. For ESC/EB differentiation, ESCs were

differentiated into embryoid bodies (EBs) using mesodermal differentiation media as previously described (113). Briefly, ESCs were dissociated into a single cell suspension using 0.25% trypsin, plated for 30-40 minutes to remove fibroblast cells (de-MEF) and differentiated using the shaking method in differentiation media containing 15% FBS, 1X penicillin/streptomycin, 1X GlutaMAX, 50 µg/ml Fe-saturated transferrin, 450 mM monothioglycerol, 50 µg/ml ascorbic acid in IMDM. The embryoid bodies were treated with doxycycline (0.5 µg/ml) between day 2 and day 4 of differentiation, as specified for each experiment and harvested.(113) *iHA-Etv2* MEFs were isolated from E13.5 embryos (Supplementary Figure II) using a previously described method(158) and cultured in media containing 10% fetal bovine serum (FBS), 2 mM glutamine, 1X penicillin/streptomycin and 1X non-essential amino acids at 37°C in 5% CO2. Doxycycline (1 ug/ml) was added to *iHA-Etv2* MEFs for migration assays and transcriptional analysis.

Western blot analysis. Western blot analysis was performed as described previously (195). Briefly, cell lysates from *iHA-Etv2* ESCs, EBs and MEFs were obtained following –Dox and +Dox treatment at various time points and were lysed in ice-cold lysis buffer for 30 minutes and centrifuged at 10,000rpm for 10 min at 4°C. Equal amounts of protein were loaded on 10% SDS-polyacrylamide gels. The PVDF membrane was blocked with 5% (w/v) milk protein and incubated with a rabbit-HA-antibody [Cell Signaling (C29F4; 1:2000)], mouse-RhoJ-antibody [Santa Cruz (sc-81936; 1:500)] and Gapdh antibody [Cell

Signaling (D16H11; 1:2000)] for an overnight period at 4ºC. The membrane was subsequently incubated with anti-rabbit [Santa Cruz Biotechnology (SC-2004; 1:2,000)] or anti-mouse [Santa Cruz Biotechnology (SC-2005; 1:2,000)] HRP-conjugated secondary antibody and visualized using the Pico luminescence kit (Invitrogen, USA) according to the manufacturer's instructions. The protein bands were visualized and imaged using Image Lab software.

RNA isolation and qPCR analysis. Total RNA was isolated from MEFs, FACS-sorted cells (~25,000 cells) or cells from EBs using the RNeasy kit (Qiagen) according to the manufacturer's protocol. Briefly, cells were lysed in RLT-lysis buffer, followed by a column-based purification process and on-column DNA digestion to remove any traces of DNA. cDNA was synthesized using the SuperScript IV VILO kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Quantitative PCR (qPCR) was performed with ABI Taqman probe sets. The list of taqman probes used is provided in Supplementary Table III.

Electrophoretic mobility shift assay (EMSA). The protocol used for EMSA was described previously (109). Briefly, HA-tagged ETV2 was synthesized using pcDNA3.1-Etv2-HA vectors using the TNT Quick Coupled Transcription/Translation System (Promega, Madison, WI) according to the manufacturer's protocol. DNA oligonucleotides corresponding to wild-type *Rhoj* promoter sequences (Site#1, Site#2, and Site#3) or mutant sequences with an

AGG>TTC mutation in the putative Etv2 binding sites were synthesized (Integrated DNA Technologies, Coralville, IA). Top-strand, wildtype oligonucleotides were synthesized with and without the IRDye® 700 fluorophore (Integrated DNA Technologies, Coralville, IA). The following oligonucleotides sequences were used for the EMSA experiment. *Rhoj* promoter region for Site#1 WT Strand (196) labeled: IRD700-GGGATAAAGCAGGAAGTTTGACAGC; Site#1 WT Strand (196) unlabeled: GGGATAAAGCAGGAAGTTTGACAGC; Site#1 WT Strand (bottom) unlabeled: GCTGTCAAACTTCCTGCTTTATCCC; Site#1 Mutant Strand (196) unlabeled: GGGATAAAGCttcAAGTTTGACAGC; Site#1 Mutant Strand (bottom) unlabeled: GCTGTCAAACTTgaaGCTTTATCCC; Site#2 WT Strand (196) labeled: IRD700-GGAATGCAGCAGGAAACCCACGATT; Site#2 WT Strand (196) unlabeled: GGAATGCAGCAGGAAACCCACGATT; Site#2 WT Strand (bottom) unlabeled: AATCGTGGGTTTCCTGCTGCATTCC; Site#2 Mutant Strand (196) unlabeled: GGAATGCAGCttcAAACCCACGATT; Site#2 Mutant Strand (bottom) unlabeled: AATCGTGGGTTTgaaGCTGCATTCC; Site#3 WT Strand (196) labeled: IRD700-AACCCACGATTTCCTGACACTCGGC; Site#3 WT Strand (196) unlabeled: AACCCACGATTTCCTGACACTCGGC; Site#3 WT Strand (bottom) unlabeled: GCCGAGTGTCAGGAAATCGTGGGTT; Site#3 Mutant Strand (196) unlabeled: AACCCACGATTTgaaGACACTCGGC; Site#3 Mutant Strand (bottom) unlabeled: GCCGAGTGTCttcAAATCGTGGGTT.

The complimentary WT or mutant oligos were annealed to generate labeled probe and unlabeled competitor DNA. *In vitro* synthesized HA-ETV2 (1µL) was pre-bound with 250 ng of poly dI-dC (Sigma) in binding buffer (50 mM Tris pH 7.6, 80 mM NaCl, 8% glycerol) at room temperature for 10 minutes. Pre-binding reactions included 5 nmol of unlabeled competitor oligo as appropriate. For supershift assays, pre-binding of ETV2 was performed in the presence of active or heat-inactivated anti-human ETV2 antibody (ER71 (N-15), catalog #sc-164278; Santa Cruz Biotechnology, Inc., Dallas, TX). IRDye® 700-labelled probe (100 fmol) was then added to the pre-binding reaction and then incubated at room temperature for 15 minutes. DNA-Protein complexes were resolved on a 6% non-denaturing polyacrylamide gel in 0.5x TBE (40 mM Tris pH 8.3, 45 mM boric acid, and 1 mM EDTA) at room temperature. Fluorescence was detected using an Odyssey CLx imager (LI-COR Biosciences, Lincoln, NE).

Bioinformatics analyses. The method used for the bioinformatics analyses was previously described (109). Briefly, peak analysis using MACS2 version 2.2.6 was performed on our *iHA-Etv2* ES/EB Flk1 positive cell sorted ATACseq data following 3h Dox treatment with a q-value cutoff of 0.05, normalized to the no Dox Flk1 positive condition. We re-analyzed our previously published bulk RNAseq data (137) of the *iHA-Etv2* ES/EB system following 6h and 12h of Dox treatment. Differential gene analysis was performed using the R package, DESeq2 7 (v1.18.1), to obtain normalized counts, fold change, and p-values. Cell migratory genes were considered significant if the p-value was less than 0.001

and absolute fold change was greater than 2 Normalized expression was log-transformed and scaled to generate heatmaps. We used the goseq R package for Gene Ontology analysis (197). Significance values were determined using the Fisher Exact test. The size of the dot indicates the number of the genes annotated with corresponding GO terms, and the color of the dot indicates the *p*-value of the Fisher's exact test of the pathway analysis.

Luciferase assays. The transcriptional assays were undertaken using our previously published protocols (51). The promoter-reporter constructs (*Rhoj-Luc*) were generated with luciferase (*Luc*) under the control of either a 0.5 kb fragment of the *Rhoj* promoter harboring three evolutionarily conserved Etv2 binding motifs. The mutant promoter-reporter constructs were generated using the same method but substituted mutant oligos for ETV2 binding sites. The *Rhoj* promoter region was amplified using PCR and subcloned into the pGL3 vector to generate *pGL3-Rhoj-Luc*. HEK/293T cells were grown in Dulbecco's modified Eagle's complete medium supplemented with 10% FBS and 1X penicillin/streptomycin (ThermoFisher Scientific). HEK/293T cells were trypsinized using 0.25% trypsin and 1×10^5 cells were plated in each well of a 12-well plate and co-transfected with wildtype (WT) or mutant (140) *pGL3-Rhoj-Luc* and increasing amounts of ETV2 expression plasmid using Lipofectamine 3000 (Life Technologies) as per manufacturer's protocol. Cells were transfected with 10 ng of pRL-CMV (Promega) expressing Renilla luciferase as an internal control. Cells were

harvested 36 hours after transfection and luciferase activity quantified using the Dual Luciferase Stop-Glo System (Promega).

Chromatin Immunoprecipitation (ChIP). *iHA-Etv2* MEFs were used for ChIP using the protocol as previously described (109). Briefly, MEFs were dissociated into single cells using 0.25% trypsin, fixed with 1% formaldehyde at room temperature for 10 min, and quenched in 0.125 M glycine. Following the cross-linking step, cell pellets (1-2 × 10^7) were resuspended in 5 mL lysis buffer 1 (250 mM Hepes pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton X-100) for 10 minutes with rotation, followed by centrifugation and resuspension in 5 mL lysis buffer 2 (10 mM Tris [pH 8], 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) for 10 minutes with rotation. Samples were centrifuged and resuspended in 1.5 mL lysis buffer 3 (0.5% N-Laurylsarcosine, 0.1% Na-Deoxycholate, 0.5 mM EGTA, 1 mM EDTA, 100 mM NaCl, 10 mM Tris-HCl [pH 8], followed by sonication to achieve 200-500 basepair (bp) DNA fragments using an ultrasonicator. The sonicated lysate was centrifuged for 10 minutes at 16,000 g at 4 C and 150 μl 10% Triton X-100 was added to the soluble lysate. The chromatin lysate was precleared with protein G dynabeads and incubated with 10 μg anti-HA antibody (Sigma 12CA5) overnight at 4°C with rotation. Subsequently, 15μl protein G-dynabeads were added and incubated for 15 minutes at room temperature. The beads were then washed five times with cold RIPA wash buffer (1% Triton X-100, 150 mM NaCl, 0.1% SDS, 1 mM EDTA, 50 mM Tris-HCl [pH 8], 0.5% Na-Deoxycholate), and then TE buffer. All buffers were supplemented

with a protease inhibitor cocktail (Sigma P8340). Precipitated chromatin complexes were eluted in 50 µl elution buffer (50 mM Tris [pH 8], 10mM EDTA, 1% SDS) at 65° C for 10 minutes, and decrosslinked overnight at 65°C. Samples were then treated with RNase A for 2 h followed by proteinase K treatment for an additional 2h period. DNA was purified with the PCR purification kit (Qiagen) and qPCR was performed using specific primers. The primers we used for our study were: *Rhoj* Promoter Fwd: 5'- GTTCCCCAGAAGTCCAAACA-3'; *Rhoj* Promoter Rev: 5'- CTTGCCGAGTGTCAGGAAAT -3'; Intergenic control Fwd: 5'-TGGGCATATCCCTGGAGCTT-3'; Intergenic control Rev: 5'-GGCCATCCCACAGTCACAAC-3'; *Gapdh* promoter Fwd: 5'-CATGGCCTTCCGTGTTCCTA-3'; *Gapdh* promoter Rev: 5-CTGGTCCTCAGTGTAGCCCAA-3'.

ESC migration and scratch assay. For the ESC migration assay, *iHA-Etv2* ESCs were plated on 6-well plate containing sterile glass coverslips in the center of the well. Following 100% confluence, the coverslips were removed to establish an empty zone. The wells were washed twice with PBS followed by culturing the cells in the absence (-Dox) and presence (+Dox) of doxycycline for a 24h period in the ES culture media containing 15% fetal bovine serum (FBS), 2 mM Glutamax, 1X penicillin/streptomycin, 0.1 mM β-mercaptoethanol, and 1,000 U/mL LIF (Millipore), at 37°C in 5% CO2. The migrating ESCs were imaged every 12h at the same region of the well. For the migration (scratch) assay, passage (P) P7-P8 *iHA-Etv2* MEF cells were grown to 100% confluence in 35mm

tissue culture dish and the scratch was performed using P200 pipette tips. Each well was washed with PBS twice and the cells were culture in the absence (-Dox) and presence (+Dox) of doxycycline for a 12h period. Each well was imaged every 6h period at the same region of the well.

Sprouting assay. Sprout formation assays were performed using *iHA-Etv2* ES/EBs and a method previously described (110). Briefly, *iHA-Etv2* ESCs were differentiated into EBs as previously described (113), and doxycycline (Dox) was added from day 2 to day 3. Following Dox treatment, 30-50 day 3 EBs were transferred to a 12-well plate coated with Low Growth Factor Matrigel and supplemented with 50ng/ml of VEGF in serum free media. Sprouts were imaged and quantified using an inverted brightfield microscope at day 3 and day 6 of plating. No Dox and no VEGF were used as controls in these experiments.

Statistical Analysis. All experiments were repeated at least three times and values presented are mean ± standard error of the mean (35). Statistical significance was determined using the Student's *t*-test when comparing 2 groups and one-way ANOVA with multiple comparisons when comparing more than 2 groups. The Tukey test was used for multiple comparisons of more than 2 groups in conjunction with one-way ANOVA (post-hoc analysis). A p-value $\leq$ 0.05 was considered a significant change and was highlighted in each panel by an asterisk. For the bioinformatics analyses, significance was determined by using

the Fisher Exact Test. Normality and variance were not tested to determine whether the applied parametric tests were appropriate for these analyses.

Results

ETV2 regulates cell migratory networks

ETV2 is an essential transcription factor that is expressed transiently in the earliest endothelial progenitors (53). These progenitors migrate to and from the primitive vascular plexus and the mechanisms that govern these migratory networks are incompletely defined. To further evaluate the role of ETV2 in angioblasts, we investigated and mined our previously published RNAseq datasets (137), obtained from differentiating embryonic stem cells (ESCs) and embryoid bodies (EBs) following doxycycline-induced (+Dox) overexpression of ETV2 (*iHA-Etv2*-ESCs) for 6h or 12h periods on day 3 (D3) of differentiation. Multiple transcripts of cell migration genes including: Mmp9, Mmrn2, Egfl7, RhoV and Pik3cd were significantly up-regulated in the +Dox conditions compared to -Dox conditions (Figure 1A and Supplementary Table I). To investigate whether ETV2 transcriptionally activates these genes, we analyzed publicly available ETV2 ChIPseq datasets during ESC/EB differentiation (49). We examined the promoter regions of genes (TSS +/- 5kb) for ETV2 ChIPseq peaks and found ETV2 binding sites in the promoters of 5,195 genes. We found that cell migration genes [Gene Ontology (GO)-classification for cell migration (GO: 0048870) genes] have higher ETV2 ChIPseq peaks near their TSS region than non-migration genes (335 out of 1,543 genes, Fisher' exact test $p < 0.001$) (Figure

1B). We also evaluated the association between all other GO Biological Process (BP) terms and genes that were located near the ETV2 ChIP-seq peaks or ATAC-seq peaks (D3 EB). Among 4,473 GO BP terms that have at least 20 annotated genes, cell motility (GO:0048870) was the most significantly associated term associated with the Etv2 ChIP-seq peaks by Fisher's exact test (Supplementary Figure IA). In comparison, cell motility (GO:0048870) ranked as the 134th most significantly associated term with the genes located near the ATAC-seq peaks (Fisher's exact test p-value=3.9E-64). The GO terms that were most significantly associated with the ATAC-seq peaks include: cellular response to stress, cell cycle, regulation of expression, etc. (Supplementary Figure IB), suggesting these biological processes also played important roles in reshaping the chromatin accessibility landscape during EB differentiation induced by Etv2 induction.

To further examine whether the binding of ETV2 to migratory gene promoters affects chromatin accessibility, we analyzed the previously published ATACseq experiments from D3 *iHA-Etv2* ESC/EBs with (+Dox) or without Dox (-Dox) treatment for 3h (109). Similar to the ChIPseq analysis, the ATACseq analysis showed significantly higher presence of ATACseq peaks near the TSS of migratory genes, compared to background genes (non-migratory genes) (Fisher's exact test $p < 0.001$) (Figure 1C). Moreover, the combined analysis of RNAseq, ChIPseq and ATACseq data showed that the genes up-regulated following ETV2 overexpression were significantly associated with cell mobility,

cell migration, RhoA signaling and Rho GTPase related function and consistently have more ETV2 ChIPseq peaks and ATACseq peaks near the TSS region (Figure 1D and 1E). These results indicated that ETV2 could promote migratory processes during differentiation by transcriptionally activating and relaxing the chromatin near the TSS of endothelial genes. Based on these results, we hypothesized that ETV2 regulated cell migration by the modulation of migratory networks during differentiation.

ETV2 promotes cell migration in both embryonic stem cells (ESCs) and mouse embryonic fibroblasts (MEFs)

Having established the positive correlation of ETV2 induction with expression of cell migration genes, we analyzed mouse embryos expressing EYFP under the control of the 3.9kb *Etv2* promoter fragment (*Etv2-EYFP*) at E8.5. We have previously shown that these $EYFP^+$ cells in transgenic embryos mark the earliest HE lineages (angioblasts) in the embryo (50). Wholemount analysis of *Etv2-EYFP* embryos showed loss of migration of EYFP+ cells from the peripheral regions to the dorsal aortae in *Etv2* knockout embryos (Figure 2A; n = 3 replicates). These results supported the notion that ETV2 plays a critical role in angioblast migration *in vivo*. Next, we directly tested whether the overexpression of ETV2 could induce cell migration in an *in vitro* setting. It has been previously reported that the expression of ETV2 is absent in ESCs and is transient in differentiating EBs, with the highest expression between D3 and D4 of EB differentiation (50, 51, 198). To examine the role of ETV2 in cell migration,

we plated *iHA-Etv2* ESCs in a monolayer and performed scratch assays in the absence (-Dox) or presence (+Dox) of doxycycline. Western blot analysis using cell lysates from -Dox and +Dox *iHA-Etv2-ESCs* revealed robust increase in the levels of the exogenous ETV2 following Dox-mediated induction of Etv2 at the 24 h time-point (Figure 2B). We found minimal migration of ESCs in the absence of doxycycline (Figure 2C-D; n = 4 replicates). Imaging and quantification of the distance migrated revealed that the overexpression of ETV2 resulted in increased migration of ESCs compared to control (Figure 2D; n = 4 replicates; *$p < 0.05$). Next, to validate the migratory role of ETV2 in endothelial progenitors, we utilized the *iHA-Etv2* ES/EB system and induced ETV2 between D2-D3 (24h period) and performed sprout formation assays using a Matrigel-sandwich assay supplemented with VEGF (50ng/ml) in serum free conditions. Analysis of sprout formation at day 3 and day 6 of plating showed ~10% of EBs with sprout formation in the -Dox condition. Notably, we observed a robust increase in the percent of EBs with sprout formation in the +Dox condition relative to -Dox condition, supporting its role in migration of endothelial progenitors (Supplementary Figure II). To further confirm the role of ETV2 in cell migration, we utilized a heterologous system of isolated mouse embryonic fibroblasts (MEFs) obtained from E13.5 embryos by crossing *Rosa26-rtTA* with *TRE-iHA-Etv2* genetic mouse lines. These isolated cells (*iHA-Etv2* MEFs) from *Rosa26-rtTA;TRE-iHA-Etv2* embryos robustly expressed ETV2 following the addition of doxycycline (Supplementary Figure III). Induction of ETV2 protein in these isolated MEFs was further confirmed by western blot using an HA antibody

(Figure 2E). Using these cells, we performed scratch assays and monitored their migration in the absence (-Dox) or presence (+Dox) of doxycycline for 12h. Our analysis revealed that the migratory properties of *iHA-Etv2* MEFs were enhanced following the overexpression of ETV2 (Figure 2F-G; n = 5 replicates; **$p < 0.01$). To further confirm these migratory results, we undertook qPCR analysis using RNA isolated from migrating *iHA-Etv2* MEFs in the presence or absence of doxycycline. Similar to the bulk RNAseq dataset analysis using ESC/EB database (Figure 1), our qPCR analysis revealed that the levels of multiple migratory gene transcripts, including *Mmp9* and *PlexinD1* were increased in +Dox MEFs compared to -Dox controls (Figure 2H, I, n = 3 replicates; **$p < 0.01$). Overall, these results indicated that overexpression of ETV2 promoted cell migration in several distinct model systems.

***Rhoj* is co-expressed with *Etv2* in endothelial progenitors**

To further define direct downstream targets of ETV2, we analyzed the bulk RNAseq datasets from *iHA-Etv2* ES/EBs (D3) in the presence (+Dox) (6 hr or 12 hr time periods) or absence (-Dox) of doxycycline. Initially, we clustered our published bulk RNAseq datasets(137) based on the following criteria: i) up- or down-regulated by >1.5 fold difference between -Dox and +Dox conditions; ii) significant differential expression ($p < 0.05$); iii) location of ETV2 binding motifs in the up/downstream regions of their transcriptional start site; and iv) annotation using the GO-classification for cell migration. Based on these criteria, we identified *Rhoj* as one of the top-ranked and most enriched candidates involved

in cell migration (Figure 1A and Supplementary Table I). To decipher the correlation between *Etv2* and *Rhoj* expression, we analyzed the single cell RNAseq (scRNAseq) datasets obtained from the *Etv2-EYFP* progenitors at three embryonic stages, E7.25, E7.75 and E8.25 (50). We used uniform manifold approximation and projection (UMAP) to visualize the single cells and found three distinct cell clusters: progenitor cells that are mostly from E7.25, endothelial lineages and hematopoietic lineages that are mostly from later stages E7.75 and E8.25 (Figure 3A-D, Supplementary Figure IV). Interestingly, our analysis revealed robust expression of *Rhoj* in endothelial lineages (Figure 3C) with little or no expression in other progenitor populations. Next, we undertook qPCR analysis using RNA isolated from FACS-sorted EYFP^{-} and EYFP^{+} cells from *Etv2*-EYFP E8.5 transgenic mouse embryos. Our analysis revealed robust expression of *Rhoj* in EYFP^{+} relative to EYFP^{-} cells (Figure 3E; n = 3 replicates; **$p < 0.01$). We then utilized the *3.9kb Etv2* promoter driving *zsGreen-DR (Etv2-zsGreen1-DR)* ESC/EB system (198) and sorted zsGreen^{-} and zsGreen^{+} cells at D3 and D4 of differentiation and performed qPCR for *Rhoj*. The zsGreen^{+} cell lineages at D3 and D4 represent the early angioblast populations during ESC/EB differentiation. Our results indicated that *Rhoj* transcripts were robustly expressed in the zsGreen^{+} cell populations compared to zsGreen^{-} cells at D3 and D4 of differentiation (Figure 3F; n = 3 replicates; *$p < 0.05$). Based on these results, we hypothesized that *Rhoj* was restricted to the HE progenitors during development and differentiation. To confirm this hypothesis, we differentiated wildtype ESCs and undertook qPCR analysis using RNA isolated from FACS-

sorted Flk1^{+}/Pdgfra^{-} (lateral plate mesoderm), Flk1^{+}/Pdgfra^{+} (cardiac mesoderm), Flk1^{-}/Pdgfra^{+} (paraxial mesoderm) and Flk1^{-}/Pdgfra^{-} (other lineages) cell populations (Figure 3G). Our analysis showed a relatively robust and restricted enrichment of *Rhoj* in the Flk1^{+}/Pdgfra^{-} population as compared to other lineages (Figure 3H; n = 3 replicates; **$p < 0.01$). In the Flk1^{+}/Pdgfra^{-} sorted cells, we found *Etv2* to be highly expressed (Figure 3I; n = 3 replicates; *$p < 0.05$). qPCR analysis for other Ets family members revealed no change in *Ets1*, but a significant enrichment of both *Fli1* and *Erg*, which are known downstream targets of ETV2 in Flk1^{+}/Pdgfra^{-} sorted cells (Figure 3J-L; n = 3 replicates; *$p < 0.05$). Altogether, these data support the notion that *Etv2* and *Rhoj* are coexpressed in HE lineages.

ETV2 is an upstream regulator of *Rhoj* gene expression

Based on the coexpression analysis of *Etv2* and *Rhoj*, we examined whether the expression of *Rhoj* was regulated by ETV2 during embryogenesis. Initially, we analyzed the ETV2 ChIPseq dataset (49) and found a strong peak in the promoter region of the *Rhoj* gene. We then utilized ATACseq datasets (109) from differentiating *iHA-Etv2* ESCs/EBs and found an ATACseq peak and ChIPseq peak in the +Dox condition compared to -Dox controls (Figure 4A). Moreover, the expression of *Rhoj* transcript paralleled the expression of *Etv2* during ESC/EB differentiation (198) (Supplementary Figure V). These bioinformatics analyses revealed that ETV2 could potentially bind to the promoter region of the *Rhoj* gene to modulate its expression. To explore this possibility, we

utilized the *iHA-Etv2* ESC/EB system and performed qPCR analysis of D4 EBs in the presence (+Dox) and absence (-Dox) of doxycycline. Our data revealed that Dox-mediated overexpression of ETV2 resulted in a robust (~10-fold) increase in the level of *Rhoj* transcripts (Figure 4B; n = 3 replicates; **$p < 0.01$). Next, we performed qPCR analysis using *Etv2* null ESC/EBs and found significantly reduced levels of *Rhoj* transcripts in *Etv2* null D4 EBs compared to wildtype EBs (Figure 4C; n = 3 replicates; **$p < 0.01$). To test these findings *in vivo*, we performed qPCR analysis for *Rhoj* transcripts using RNA isolated from E8.5 *Etv2* wildtype and null mouse embryos or yolk sacs (YS) (Figure 4D; n = 3 replicates; *$p < 0.05$). We observed robust expression of *Rhoj* in wildtype embryos but significantly lower expression in *Etv2* null embryos (Figure 4D; n = 3 replicates; *$p<0.05$). To further examine the ability of ETV2 to regulate *Rhoj* gene expression, we undertook qPCR analysis using the *iHA-Etv2* MEFs, in the presence (+Dox) or absence (-Dox) of doxycycline. Our data revealed that Dox-mediated overexpression of ETV2 resulted in increased expression of *Rhoj* as compared to -Dox condition (Figure 4E; n = 3 replicates; *$p < 0.05$).

These results supported the hypothesis that *Rhoj* is a downstream effector of ETV2 important in the regulation of endothelial progenitor cell migration. To monitor whether ETV2 could regulate the expression of *Rhoj*, we analyzed the upstream region of the *Rhoj* gene and identified evolutionary conserved ETV2 binding motifs among various species (Figure 5A). These bioinformatics analyses suggested that ETV2 could potentially bind to the *Rhoj* promoter. Therefore, we

undertook transcriptional assays using the 0.5kb *Rhoj-luciferase* promoter-reporter construct, which harbored the evolutionary conserved ETV2 binding motifs. Co-transfection of the *Rhoj-promoter*-reporter plasmid with an *Etv2* expression plasmid resulted in a robust increase in luciferase activity (~90-fold) relative to control conditions (Figure 5B; n = 3 replicates; **p < 0.01). Mutation of all three ETV2 binding motifs resulted in complete attenuation of the activation of the Rhoj-*promoter-reporter* construct by ETV2 (Figure 5B; n = 3 replicates; **p < 0.01). Next, we performed electrophoretic mobility gel shift assays (EMSAs) using IRdye-labelled double stranded DNA oligonucleotides (oligos) that harbored the conserved ETV2 binding motifs in the *Rhoj* promoter region. We found that incubation of *in vitro* synthesized ETV2 protein with an IRdye-labelled *Rhoj* promoter oligo (probe) led to the formation of a protein-DNA complex (arrowhead), [Figure 5C (site #1), D (site #3) and Supplementary Figure VI], which was blocked by the addition of an unlabeled oligo (competitor) but not by a mutant competitor, indicating that the binding of ETV2 to these oligos was sequence specific. Furthermore, addition of an ETV2 specific antibody to this protein-DNA complex supershifted (asterisk) the complex formation, but this supershift was not observed when using a denatured (heat-inactivated; h.i.) anti-ETV2 antibody, indicating specific binding of ETV2 to the promoter region of *Rhoj* [Figure 5C (site #1), D (site #3) and Supplementary Figure VI]. To examine the in vivo binding potential of ETV2 to the *Rhoj* promoter, we performed chromatin immunoprecipitation (ChIP) using Dox-induced cell lysates from *iHA-Etv2* MEFs. Our ChIP-qPCR analysis revealed ~7-fold enrichment of ETV2 in the *Rhoj*

promoter region relative to the *Gapdh* promoter as the control promoter (Figure 5E). The binding of ETV2 to the *Rhoj* promoter region was highly specific as ChIP-qPCR for an intergenic region did not show any enrichment (Figure 5E; n = 3 replicates; **p < 0.01). Overall, these results indicated that ETV2 could bind the *Rhoj* promoter and regulate its gene expression during embryogenesis.

ETV2 rescues *Rhoj* knockdown mediated migratory defects

Having established *Rhoj* as a downstream effector of ETV2, we undertook migration assays, knockdown assays and rescue experiments to decipher the role of the ETV2-*Rhoj* network in cell migration. Initially, we generated three different shRNA lentiviral clones for *Rhoj* and tested their ability to knockdown *Rhoj* expression (Supplementary Figure VII). Infection of MEFs using GFP-reporter lentiviral particles showed >85% infectivity at the 24h time-period. qPCR analysis showed that *Rhoj* lentiviral clones #1 and #3 were most effective at knocking down *Rhoj* (Supplementary Figure VII). Our western blot analysis using RHOJ specific antibodies showed a reduction in the protein levels of RHOJ using both lentiviral *Rhoj* shRNA constructs #1 and #3 (Supplementary Figure VII). To undertake rescue experiments, we infected *iHA-Etv2* MEFs using *Rhoj* shRNA clone #1 for 24h, and then performed migration assays (i.e. scratch assay) in the absence (-Dox) or presence (+Dox) of doxycycline for 6h to overexpress ETV2. Since we found a significant difference in the migratory properties of *iHA-Etv2* MEFs following ETV2 induction (+Dox) as early as 6h (Figure 2C), we decided to undertake knockdown and rescue experiments at the earliest time period (i.e. 6h)

of Dox treatment to decipher whether ETV2 could rescue migration defects due to *Rhoj* knockdown. Induction of ETV2 showed increased migration of MEFs compared to controls (-Dox) (Figure 6A, 6B; compare panel (2) and (6) in 6A; n = 3 replicates; **p < 0.01). Moreover, the knockdown of *Rhoj* resulted in reduced cell migration (Figure 6A, 6B; compare panel (2) and (4) in 6A; n = 3 replicates; **p < 0.01). Notably, we found that this migratory defect due to *Rhoj* knockdown was partially rescued by overexpression of ETV2 in *iHA-Etv2* MEFs (Figure 6A, 6B; compare panel (4) and (8) in 6A; n = 3 replicates; **p < 0.01). To verify whether the rescue of perturbed migration was due to an ETV2-*Rhoj* network, we undertook qPCR experiments. We found higher expression of *Rhoj* following ETV2 overexpression and intermediate expression following co-expression of ETV2 and *Rhoj* shRNA compared to control conditions (Figure 6C; n = 3 replicates; **p<0.01). Similarly, rescue experiments using another shRNA construct (#3) resulted in a similar migratory phenotype and essentially phenocopied the results from the shRNA construct #1 (Supplementary Figure VIII). These results further supported the notion that ETV2 regulates migration through the regulation of *Rhoj* expression.

Discussion

The process of cell migration is critical for embryonic development, wound repair, immune response, tumor formation and metastasis (189, 199). A number of transcription factors and signaling pathways have been implicated in these processes, however, the mechanisms that govern endothelial progenitor cell

migration are still unclear (79, 200-204). We and others have previously shown the essential requirement of *Etv2* in hematoendothelial development during embryogenesis but its functional role(s) are incompletely defined (46, 48, 50, 113). In the present study, we have used computational genomics, cellular and molecular biological techniques to make several fundamental discoveries to decipher the role of ETV2 as a regulator of cell migration.

First, we identified the role of ETV2 in the regulation of cell migratory networks. We and others have shown that ETV2 is expressed in the earliest hematoendothelial progenitors (46, 48, 50, 113). These endothelial progenitors migrate to and from the primary vascular plexus in a coordinated fashion (205). Importantly, in the absence of *Etv2*, progenitors (identified using the *Etv2-EYFP* transgenic reporter mouse model) were shown to be present but mislocalized supporting the notion that ETV2 has a role in cell migration. We also observed that the overexpression of ETV2 led to increased cell migration in both EBs and MEFs. While these results provide strong evidence for a functional role for ETV2 in the regulation of cell migration, we also recognize that other factors may also contribute and regulate cell migration of endothelial progenitors during embryogenesis and will be the focus of future studies.

Our next discovery defined a mechanism whereby ETV2 regulated migration. We identified ETV2 as an upstream regulator of the *Rhoj* gene in the endothelial lineage during migration. In the present study, we showed that the

overexpression of ETV2 resulted in ~90-fold increase in *Rhoj-promoter*-reporter luciferase activity. Furthermore, our qPCR analysis using FACS-sorted cells revealed that *Rhoj* was robustly expressed in endothelial progenitors (angioblasts) and not in the non-endothelial lineages. Based on these results, we proposed that ETV2 was an important regulator of *Rhoj* in endothelial progenitors. Analysis of other Ets-family members revealed enrichment of *Fli1* and *Erg* but not *Ets1* in the lateral plate mesoderm (FLK1+ cells) population, indicating a possible role for these factors in endothelial cell migration. *Fli1 and Erg are* downstream targets of ETV2 and contribute to hemato-endothelial lineage development (206, 207). Due to the transient expression of ETV2 during embryogenesis (50), it is possible that ETV2 initiates the migration process in the angioblast populations and subsequently, other Ets-factors including *Fli1* and *Erg* maintain cellular migration in the mature endothelial population at later stages of developments as well as postnatally. Future studies will focus on understanding the role of *Fli1* and *Erg* in endothelial progenitor cell migration. Additional emphasis should be given to identification of novel pathways in migration as well as the promotion of the *Etv2-Rhoj* network in the context of ischemia or the repression of the *Etv2-Rhoj* network in the context of tumor vascular biology and angiogenesis, as both of these factors have been shown to play a role in cancer (192, 208-210).

Our recent findings defined a role for ETV2 in the regulation of *Yes1* gene expression and the Hippo signaling pathway during embryogenesis (109). In a context-dependent fashion, we demonstrated that ETV2 also regulated cell

proliferation. Previous studies have shown that cell migration and proliferation are highly coordinated and coregulated processes observed in wound healing, cancer, tissue regeneration or hypoxic environments (211-214). Furthermore, effectors such as FGF2, Shh, mir-221, AMPK, YAP and others have been shown to coregulate cell migration and cell proliferation in a context dependent fashion (196, 215-218). Therefore, our studies support the hypothesis that ETV2 plays a critical role in the coregulation of hematoendothelial progenitor cell proliferation and migration mediated via YES1 and RHOJ, respectively. Future studies will be needed to determine whether the Hippo signaling pathway modulates RHOJ expression and cell migration and to define any feedback mechanisms whereby the Hippo signaling pathway and RHOJ impact ETV2 expression.

In summary, we defined a novel role for ETV2 in the regulation of cell migration and showed that ETV2 is a direct upstream activator of *Rhoj* gene expression and together this cascade regulates HE progenitor cell migration. Future studies that target this cascade may serve as a platform for therapies that could benefit from increased (i.e. ischemic diseases) or decreased (i.e. tumorigenesis) HE progenitor cell migration.

Footnotes

This work has already been published. Reprinted from *Atherosclerosis Thrombosis & Vascular Biology*. December 2020, Volume 40, Issue 12. Singh, Bhairab N., Javier E. Sierra-Pagan, Wuming Gong, Satyabrata Das, Joshua WM Theisen, Erik Skie, Mary G. Garry, and Daniel J. Garry. "ETV2 (Ets Variant

Transcription Factor 2)-Rhoj cascade regulates endothelial progenitor cell migration during embryogenesis." Copyright 2020 with permission from *Atherosclerosis Thrombosis & Vascular Biology*.

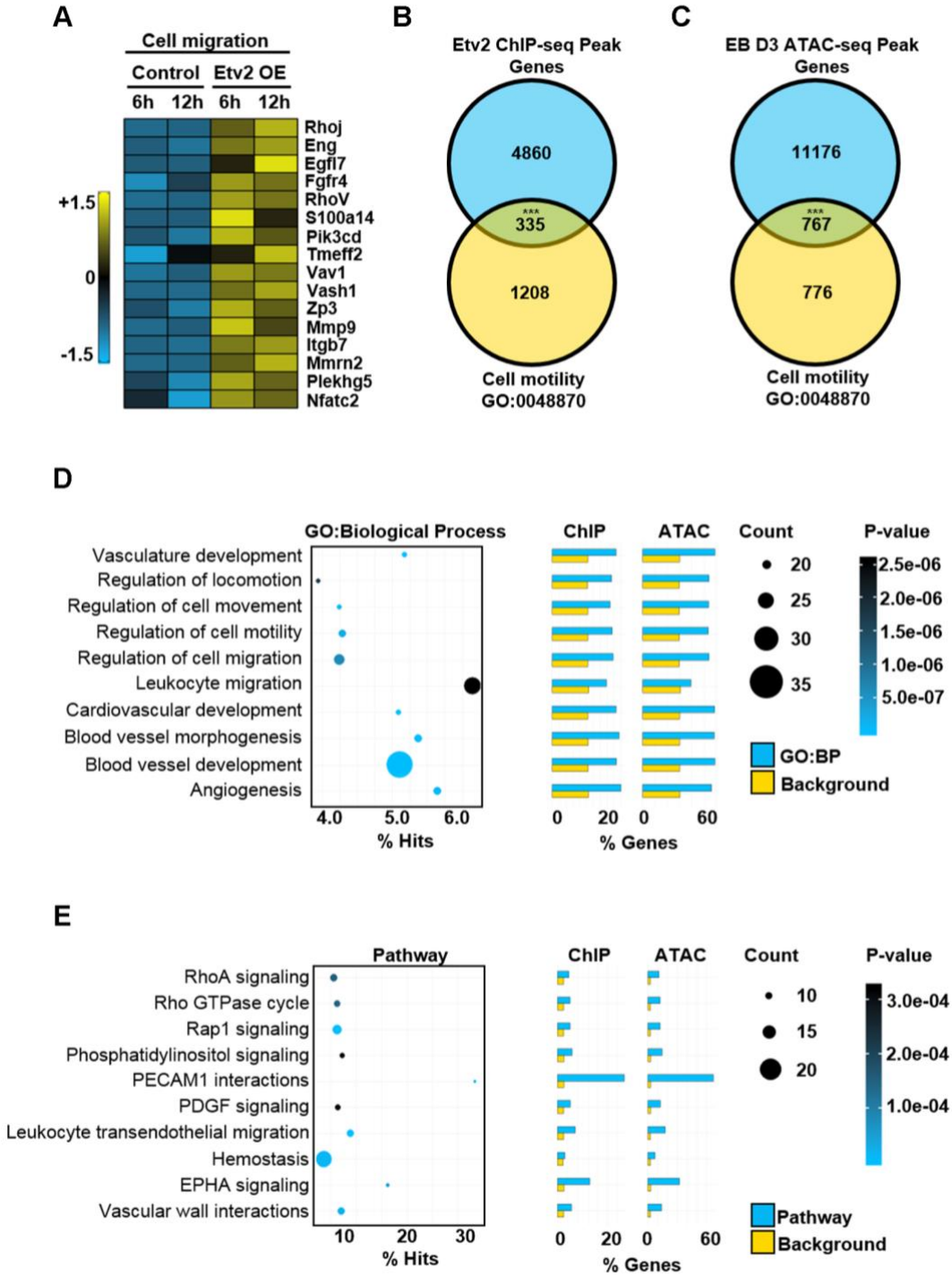

A
Cell migration
Control
Etv2 OE
6h
12h
6h
12h
Rhoj
Eng
Egfl7
Fgfr4
RhoV
S100a14
Pik3cd
Tmeff2
Vav1
Vash1
Zp3
Mmp9
Itgb7
Mmrn2
Plekhg5
Nfatc2
+1.5
0
-1.5
B
Etv2 ChIP-seq Peak Genes
4860
***
335
1208
Cell motility
GO:0048870
C
EB D3 ATAC-seq Peak Genes
11176
***
767
776
Cell motility
GO:0048870
D
GO:Biological Process
Vasculature development
Regulation of locomotion
Regulation of cell movement
Regulation of cell motility
Regulation of cell migration
Leukocyte migration
Cardiovascular development
Blood vessel morphogenesis
Blood vessel development
Angiogenesis
4.0
5.0
6.0
% Hits
ChIP
ATAC
0
20
0
60
% Genes
Count
20
25
30
35
P-value
2.5e-06
2.0e-06
1.5e-06
1.0e-06
5.0e-07
GO:BP
Background
E
Pathway
RhoA signaling
Rho GTPase cycle
Rap1 signaling
Phosphatidylinositol signaling
PECAM1 interactions
PDGF signaling
Leukocyte transendothelial migration
Hemostasis
EPHA signaling
Vascular wall interactions
10
20
30
% Hits
ChIP
ATAC
0
20
0
60
% Genes
Count
10
15
20
P-value
3.0e-04
2.0e-04
1.0e-04
Pathway
Background

Figure 1. RNAseq, ChIPseq and ATACseq analyses showed enrichment of cell migration program following the overexpression of ETV2 in the ESC/EB system. (A) Heatmap analysis of bulk RNAseq datasets using *iHA-Etv2* ES/EBs showing increased expression of cell migration genes (GO:0048870) in Dox-induced EBs relative to uninduced EBs. Note the increased expression of cell migration genes following the overexpression of HA-ETV2 at both 6h and 12h time points. (B-C) Venn diagram of the overlap between genes associated with (B) ETV2 ChIPseq peaks and (C) EB D3 ATACseq peaks and genes annotated to cell migration (GO:0048870). Note that there was a significantly higher percentage of ChIPseq and ATACseq peaks near the transcriptional start sites (TSS) of cell migratory genes compared to background genes in the Dox-treated samples (Fisher's exact test $p < 0.001$). (D-E) The biological process and pathways that are significantly enriched in the up-regulated genes following ETV2 overexpression in the ESC/EB system. The genes associated with these biological processes or pathways have consistently higher proportion of ETV2 ChIPseq peaks and ATACseq peaks near their TSS. The size of the dot indicates the number of the genes annotated with corresponding GO terms, and the color of the dot indicates the *p*-value of the Fisher's exact test of the pathway analysis. The y-axis (% Hits) indicates the percent of up-regulated genes that have the annotation of corresponding GO terms.

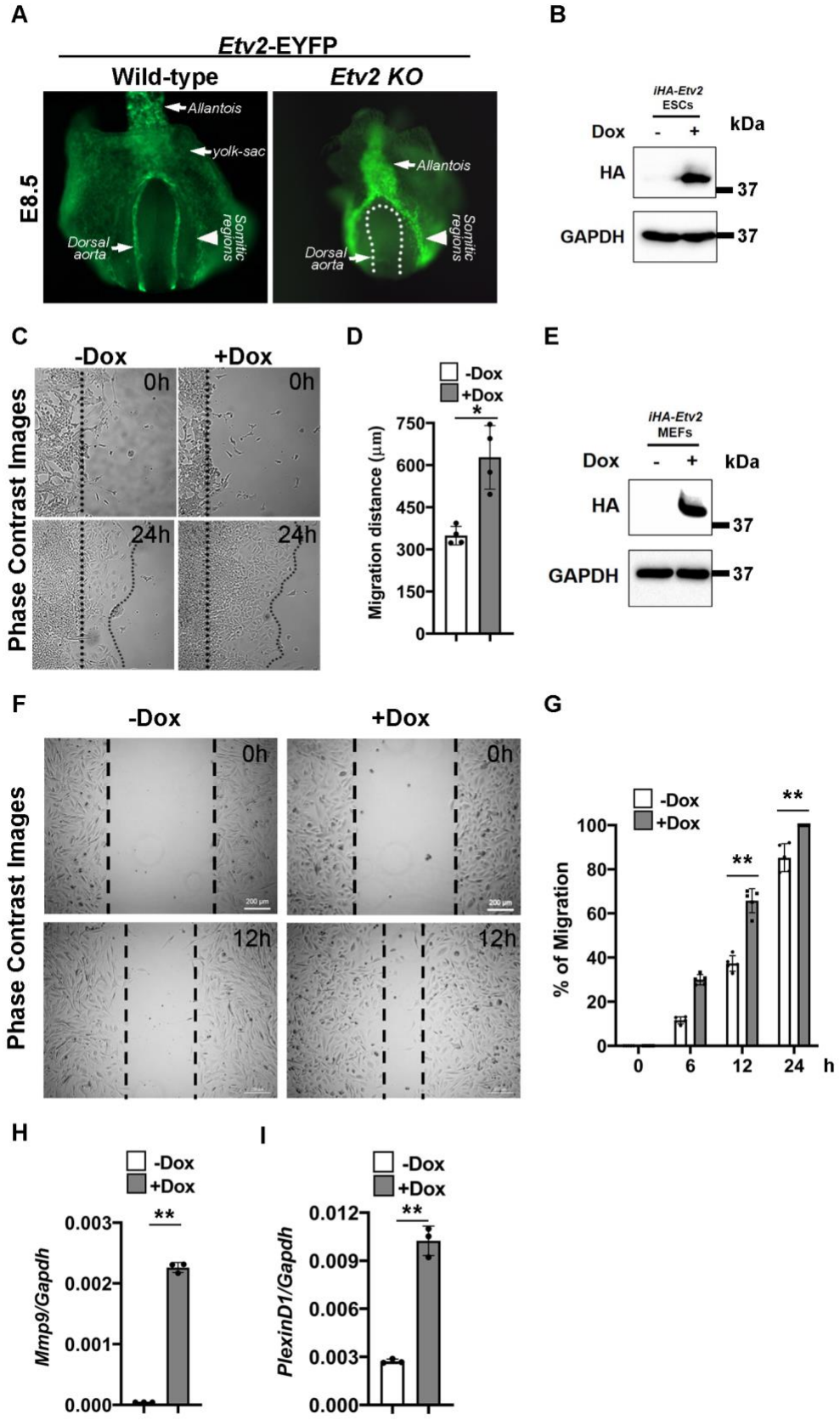

A
Etv2-EYFP
Wild-type
Etv2 KO
E8.5
Allantois
yolk-sac
Dorsal aorta
Somitic regions
B
iHA-Etv2 ESCs
Dox
kDa
HA
GAPDH
37
C
-Dox
+Dox
Phase Contrast Images
0h
24h
D
Migration distance (µm)
750
600
450
300
150
0
E
iHA-Etv2 MEFs
F
12h
G
% of Migration
100
80
60
40
20
0
6
24
h
H
Mmp9/Gapdh
0.003
0.002
0.001
0.000
I
PlexinD1/Gapdh
0.012
0.009
0.006

Figure 2. ETV2 promotes cell migration. (A) Representative wholemount fluorescence microscopy image of the transgenic *Etv2-EYFP* mouse embryos(50) at E8.5. The endothelial progenitors line the lumina of vascular structures including the dorsal aorta, yolk-sac vessels and allantois region. The dotted line indicates the expected position of the dorsal aorta, which was absent in the *Etv2* KO embryo(50) (n = 3 replicates). (B) Western blot analysis of exogenous ETV2 using an anti-HA antibody with cell lysate from -Dox and +Dox condition from *iHA-Etv2* ESCs. (C) Migration assays with *iHA-Etv2* ESCs in the absence (-Dox) and presence (+Dox) of doxycycline for a 24h period (n = 4 replicates). (D) Quantification of ESC migration. Migration was significantly enhanced in the +Dox condition. Data are presented as mean ± SEM (n = 4 replicates; *$p < 0.05$). Significance was determined by Student's *t*-test with two-tailed distribution and equal variance. (E) Western blot analysis of exogenous ETV2 using an anti-HA antibody with cell lysate from -Dox and +Dox treated *iHA-Etv2* MEFs. (F) Phase contrast microscopic images of *iHA-Etv2* MEFs in the absence (-Dox) and presence (+Dox) conditions for a 12h period following initiation of the scratch assay. (G) Quantification of *iHA-Etv2* MEFs migration in the -Dox and +Dox conditions. Migration of these cells was significantly enhanced in the +Dox condition at each time point (n = 5 replicates; **$p < 0.01$). Significance was determined by one-way ANOVA with multiple comparison. (H, I) qPCR analysis of cell migratory gene expression, including Mmp9 and PlexinD1 using RNA isolated from *iHA-Etv2* MEFs in the absence (-Dox) or presence (+Dox) of doxycycline for 24h. Note increased expression of these genes in the Dox-treated samples compared to -Dox (n = 3 replicates; **$p < 0.01$). Significance was determined by Student's *t*-test with two-tailed distribution and equal variance. Data are presented as mean ± SEM.

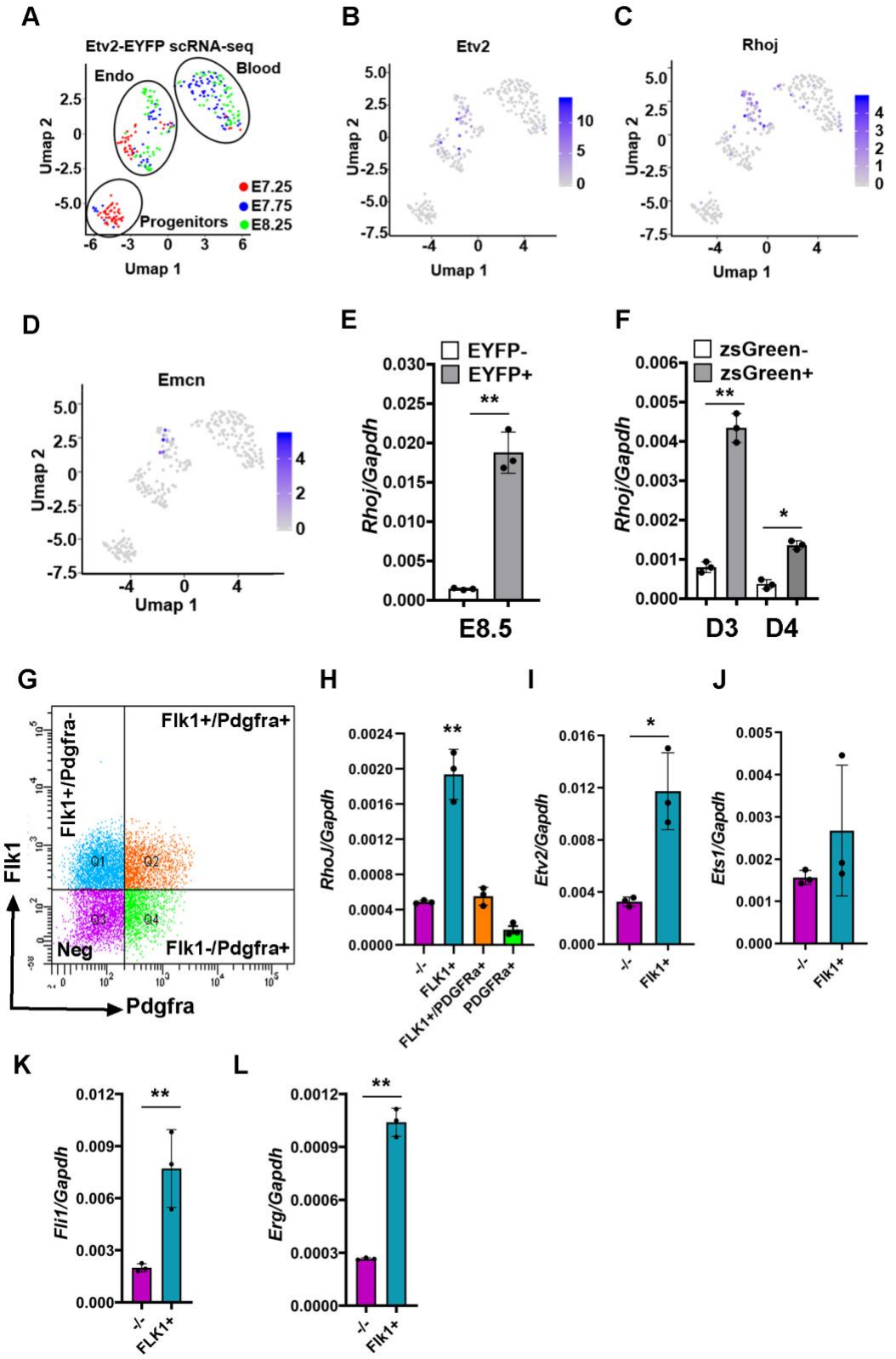

A
Etv2-EYFP scRNA-seq
Endo
Blood
Progenitors
E7.25
E7.75
E8.25
Umap 1
Umap 2
B
Etv2
C
Rhoj
D
Emcn
E
EYFP-
EYFP+
Rhoj/Gapdh
E8.5
F
zsGreen-
zsGreen+
D3
D4
G
Flk1+/Pdgfra+
Flk1+/Pdgfra-
Flk1-/Pdgfra+
Neg
Flk1
Pdgfra
H
RhoJ/Gapdh
-/-
FLK1+
FLK1+/PDGFRa+
PDGFRa+
I
Etv2/Gapdh
J
Ets1/Gapdh
K
Fli1/Gapdh
L
Erg/Gapdh

Figure 3. Co-expression of *Rhoj* and *Etv2* during embryogenesis. (A-D) Uniform manifold approximation and projection (UMAP) visualization of single cell RNAseq data from *Etv2-EYFP* embryos at E7.25, E7.75, and E8.25. (A) The single cells were divided into three distinct groups: progenitors, endothelial lineage and hematopoietic lineage. (C) *Rhoj* has a similar expression pattern to (B) endogenous *Etv2* and (D) endothelial marker *Emcn*. (E) qPCR analysis for *Rhoj* transcripts from EYFP^{-} and EYFP^{+} cells sorted from *Etv2*-EYFP embryos at E8.5. Note the robust enrichment of *Rhoj* expression in the EYFP^{+} cells relative to the EYFP^{-} cells (n = 3 replicates; **$p < 0.01$). Significance was determined by Student's *t*-test with two-tailed distribution and equal variance. (F) qPCR analysis for *Rhoj* transcripts from zsGreen^{-} and zsGreen^{+} sorted cells using the *Etv2*- zsGreen1-DR ES/EB system at D3 and D4 of differentiation. Note a significant enrichment of *Rhoj* in zsGreen^{+} cells relative to zsGreen^{-} cells (n = 3 replicates; *$p < 0.05$). Significance was determined by one-way ANOVA with multiple comparison. (G) FACS profile of mesodermal derivatives using a wildtype ESC/EB system. (H) qPCR analysis for *Rhoj* transcripts from Flk1^{-}/Pdgfra^{-} (other lineages), Flk1^{+}/Pdgfra^{-} (lateral plate mesoderm), Flk1^{+}/Pdgfra^{+} (cardiac mesoderm) and Flk1^{-}/Pdgfra^{+} (paraxial mesoderm) sorted cells using the wildtype ES/EB system. Note a significant enrichment of *Rhoj* in the Flk1^{+}/Pdgfra^{-} (lateral plate mesoderm) cells relative to the other lineages (n = 3 replicates; **$p < 0.01$). Significance was determined by one-way ANOVA with multiple comparison. (I, J) qPCR analysis for *Etv2* and *Ets1* transcripts from negative and Flk1^{+}/Pdgfra^{-} (hemato-endothelial mesoderm) sorted cells using the wildtype ES/EB system. Note the enrichment of *Etv2* but not *Ets1* transcripts in the Flk1^{+}/Pdgfra^{-} (lateral plate mesoderm) cells relative to the negative populations (n = 3 replicates; *$p < 0.05$). Significance was determined by Student's *t*-test with two-tailed distribution and equal variance. (K, L) qPCR analysis for *Fli1* and *Erg* transcripts from negative and Flk1^{+}/Pdgfra^{-} (hemato-endothelial mesoderm) sorted cells using the wildtype ES/EB system. Note a significant enrichment of both *Fli1* and *Erg* transcripts in the Flk1^{+}/Pdgfra^{-} (lateral plate mesoderm) cells relative to the negative populations (n = 3 replicates; *$p < 0.05$). Significance was determined by Student's *t*-test with two-tailed distribution and equal variance. Data are presented as mean $\pm$ SEM.

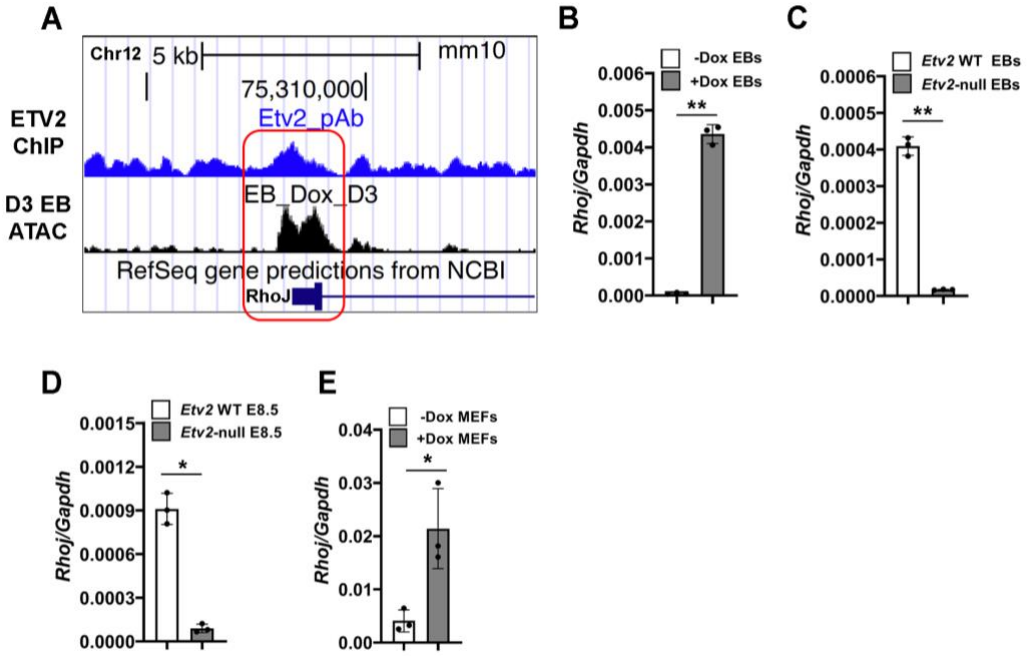


Figure 4. ETV2 regulates the expression of *Rhoj*. (A) The transcriptional start site region of *Rhoj* contains an ETV2 ChIPseq peak and an ATACseq peak (+Dox vs. -Dox). (B) qPCR analysis for *Rhoj* transcripts using the uninduced (-Dox) and induced (+Dox) D4 EBs from *iHA-Etv2* ESCs. Note a robust enrichment of *Rhoj* in the +Dox EBs as compared to -Dox (n = 3 replicates; **p < 0.01). (C) qPCR analysis for *Rhoj* transcripts using the wildtype and *Etv2 null* D4 EBs. The levels of *Rhoj* transcripts were significantly reduced in the *Etv2 null* EBs (n = 3 replicates; **p < 0.01). (D) qPCR analysis for *Rhoj* transcripts using the E8.5 *Etv2* wildtype and null embryos. Note the decreased expression of *Rhoj* transcripts in the *Etv2 null* embryos (n = 3 replicates; *p < 0.05). (E) qPCR analysis for *Rhoj* transcripts using RNA from uninduced (-Dox) and induced (+Dox) *iHA-Etv2* MEFs. Note a robust enrichment of *Rhoj* in the +Dox MEFs as compared to -Dox (n = 3 replicates; *p < 0.05). Significance was determined by Student's *t*-test with two-tailed distribution and equal variance. Data are presented as mean ± SEM.

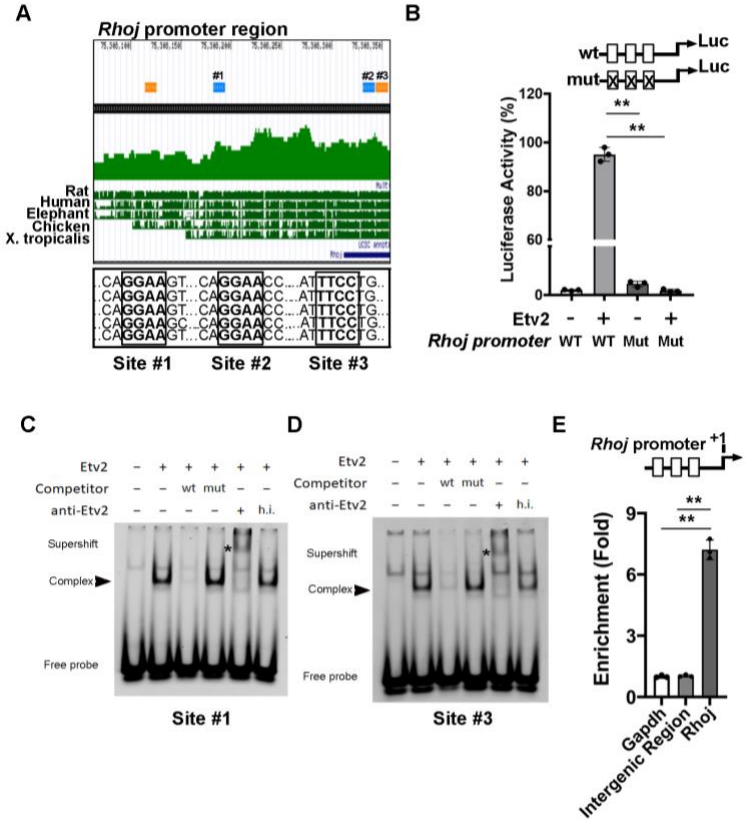


Figure 5. ETV2 is an upstream regulator of *Rhoj*. (A) Evolutionary conservation of the upstream promoter fragment of the *Rhoj* gene locus. Note the high conservation (green peaks) of the ETV2 binding motif across various species. (B) Luciferase reporter constructs using the *Rhoj* promoter (0.5 kb) harboring wildtype (wt; open box) or mutant (mut; crossed box) ETV2 binding motifs. ETV2 enhanced the transcriptional activity of the promoter-reporter construct, which was abrogated by mutating all three ETV2 binding sites (n = 3 replicates; **p < 0.01). (C, D) EMSA showing ETV2 bound to the Ets binding site in the *Rhoj* promoter region. IRdye-labeled probes containing the putative binding sites were incubated with *in vitro* synthesized HA-ETV2 protein to form a specific complex with the oligo (lane 2; arrowhead), which is competed with wildtype unlabeled oligos (lane 3) but not with mutant oligos (lane 4). Addition of the HA-antibody supershifted the complex but not with heat-inactivated (h.i.) antibody (asterisk), indicating specificity of the complex. Panels C & D represent site #1 and #3, respectively. (E) Top: Schematic of the upstream region of the *Rhoj* promoter showing the ETV2 binding sites (open boxes). Bottom: ChIP analysis of Dox-inducible *iHA-Etv2* MEFs using an HA antibody (n = 3 replicates; **p < 0.01). ChIP assay for the *Gapdh* promoter was used as a control. ChIP assay using an intergenic region was performed to validate the specificity. Significance was determined by Student's *t*-test with two-tailed distribution and equal variance (panel B) and one-way ANOVA with multiple comparison (panel E). Data are presented as mean ± SEM.

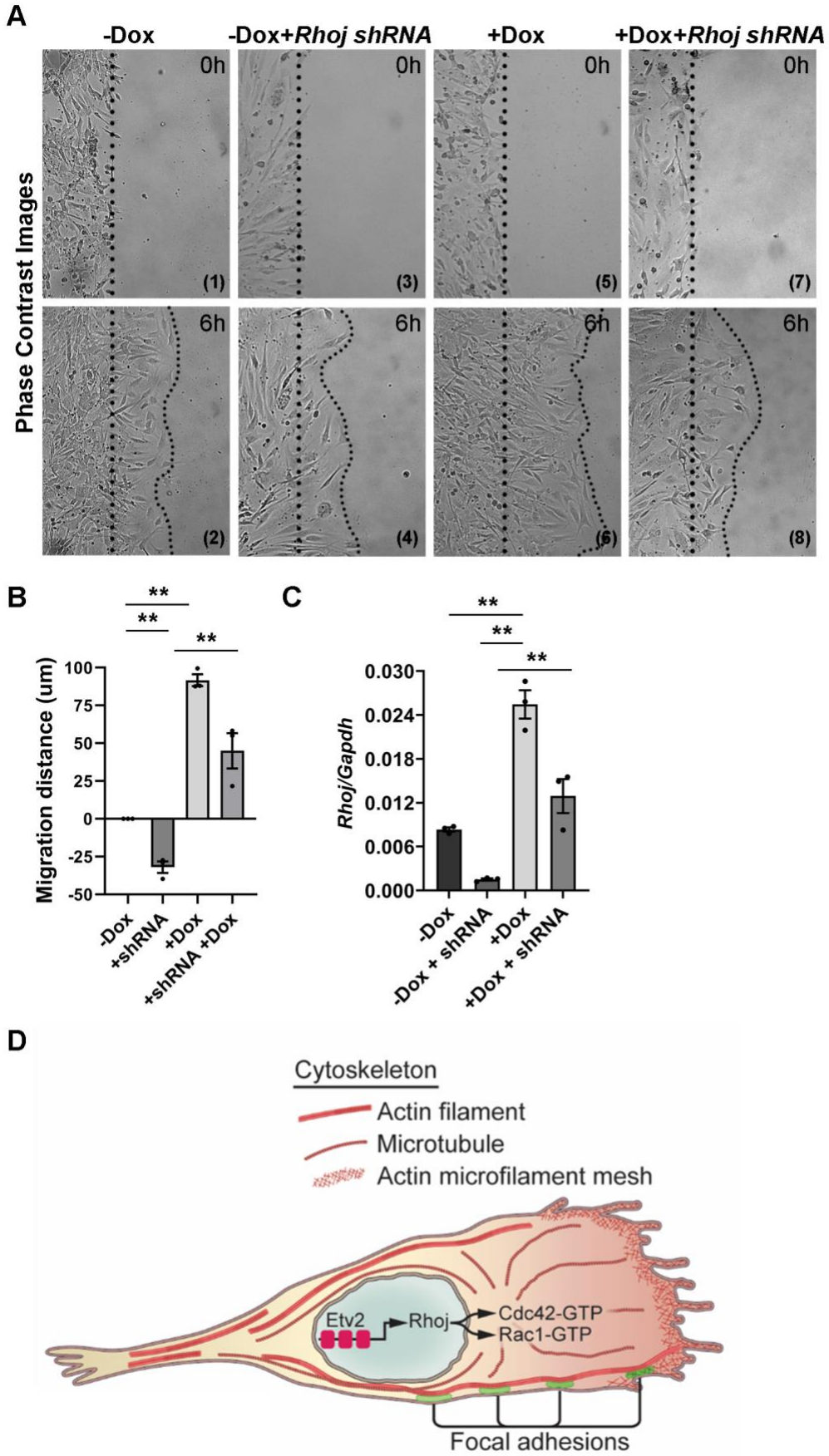
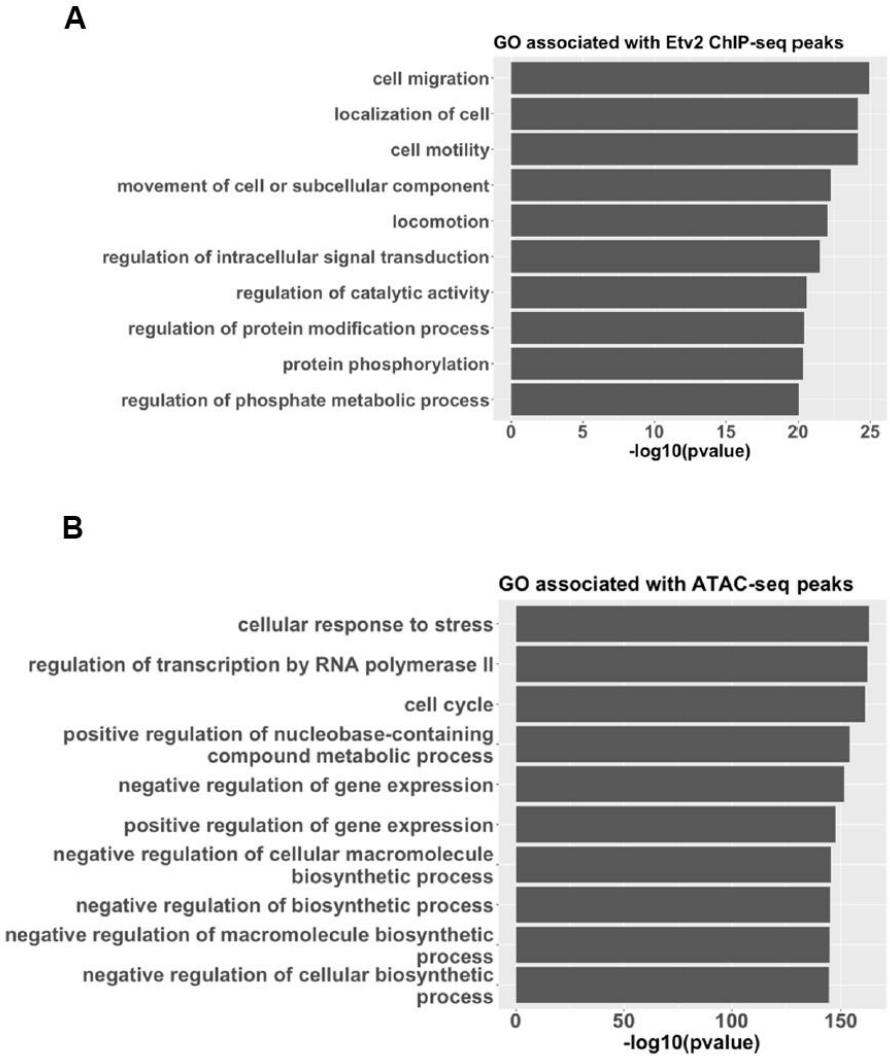
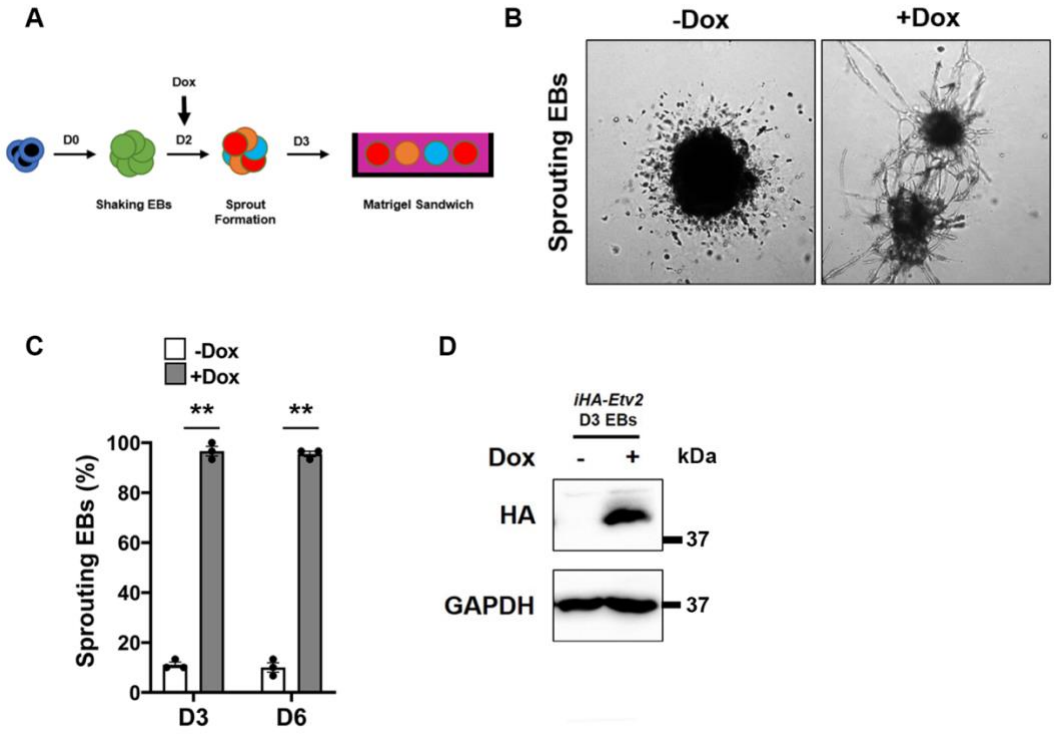

A
-Dox
-Dox+Rhoj shRNA
+Dox
+Dox+Rhoj shRNA
Phase Contrast Images
0h
6h
(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
B
Migration distance (um)
100
75
50
25
0
-25
-50
**
-Dox
+shRNA
+Dox
+shRNA +Dox
C
Rhoj/Gapdh
0.030
0.024
0.018
0.012
0.006
0.000
-Dox
-Dox + shRNA
+Dox
+Dox + shRNA
D
Cytoskeleton
Actin filament
Microtubule
Actin microfilament mesh
Etv2
Rhoj
Cdc42-GTP
Rac1-GTP
Focal adhesions

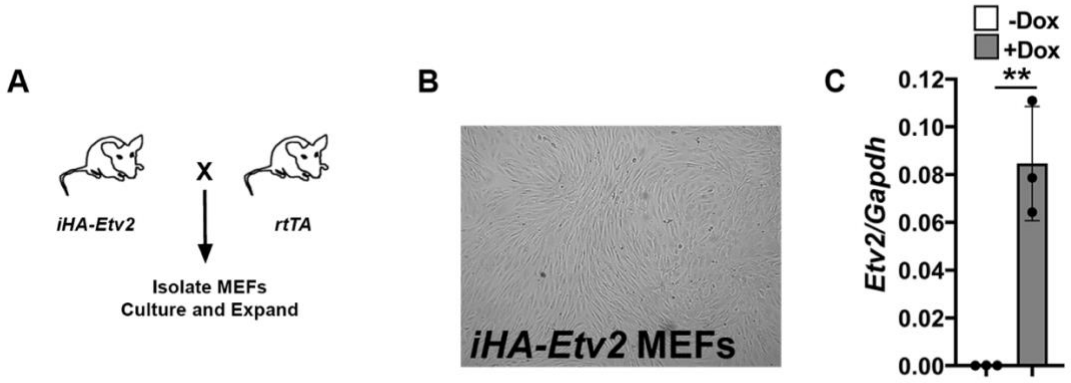
Figure 6. ETV2-*Rhoj* network regulates migration. (A) Migration (scratch) assay using the *iHA-Etv2* MEFs in the absence [-Dox (1, 2)] and presence [+Dox (5, 6)] alone or following infection with lentiviral vectors expressing *shRNA* #1 for *Rhoj* [-Dox+shRNA #1 (3, 4)] and [+Dox+shRNA #1 (7, 8)] (n = 3 replicates). (B) Quantification of migration in *iHA-Etv2* MEFs in the conditions described in panel A. Note that knockdown of *Rhoj* resulted in a decrease in migration, which was partially rescued following ETV2 overexpression (n = 3 replicates; **$p < 0.01$). (C) qPCR analysis of *Rhoj* transcripts demonstrating that the induction of ETV2 resulted in increased expression of *Rhoj* (n = 3 replicates; **$p < 0.01$). (D) Schematic showing the ETV2-RHOJ network in the regulation of cell migration. Based on our data, we propose that ETV2 transactivates *Rhoj* gene expression in endothelial progenitors that promotes other downstream effectors to induce HE progenitor cell migration. Significance was determined by one-way ANOVA with multiple comparison. Data are presented as mean $\pm$ SEM.



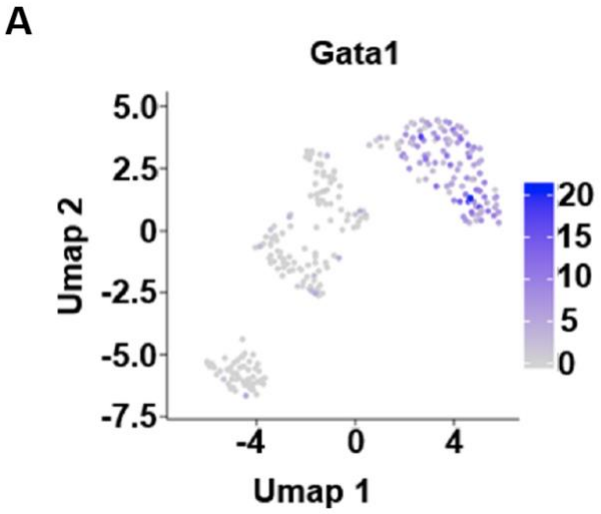
Supplementary Figure I. Top 10 Gene Ontology (GO) biological processes that are significantly associated with the genes located within -500 bp to +500 bp of (A) the Etv2 ChIP-seq peaks and (B) the ATAC-seq peaks of ES/EB differentiation. The *y*-axis indicates the -log10 of Fisher's exact test *p*-values.



Supplementary Figure II. ETV2 overexpression promotes migration by increasing the formation of sprouts. (A) Schematic diagram showing how *iHA-Etv2* ESCs were differentiated into embryoid bodies, followed by the addition of doxycycline and the formation of sprouts in a Matrigel sandwich. (B) Brightfield images of EBs 72 hrs after being placed in the Matrigel sandwich to promote the formation of sprouts. Note the formation of sprouts between EBs in the +Dox compared to no sprouts in the -Dox condition. (C) Quantification of the number of sprouts observed in each condition 3 and 6 days after being placed in the Matrigel sandwich (n = 3 replicates; **$p < 0.01$). (D) Western blot analysis of ETV2 in *iHA-Etv2* EBs following Dox treatment for 24 hrs. Significance was determined by Student's *t*-test with two-tailed distribution and equal variance. Data are presented as mean + SEM

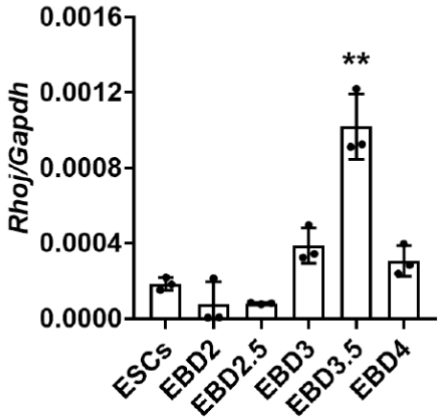


Supplementary Figure III. Characterization of inducible over-expression of mouse embryonic fibroblasts (*iHA-Etv2* MEFs). (A) Schematic showing the cross between mice carrying the *TRE-HA-Etv2* cassette with mice carrying the *Rosa26-rtTA* cassette. (B) Brightfield image of *iHA-Etv2* MEFs. (C) qPCR analysis of *Etv2* transcripts using RNA isolated from MEFs following treatment with –Dox and +Dox for a 24h period (n = 3 replicates; *$p < 0.05$). Significance was determined by Student's *t*-test with two-tailed distribution and equal variance. Data are presented as mean ± SEM.

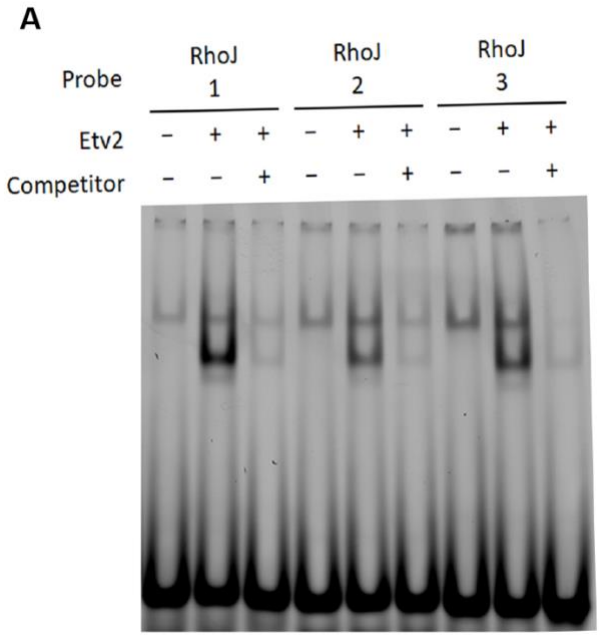


Supplementary Figure IV. *Gata1* transcript is expressed in the blood lineage. (A) Uniform manifold approximation and projection (UMAP) visualization of single cell RNAseq data from *Etv2-EYFP* embryos at E7.25, E7.75, and E8.25. *Gata1* is significantly enriched in the blood lineage with little to no expression in the other populations, which is distinct from the expression profile of both *Rhoj* and *Emcn*.

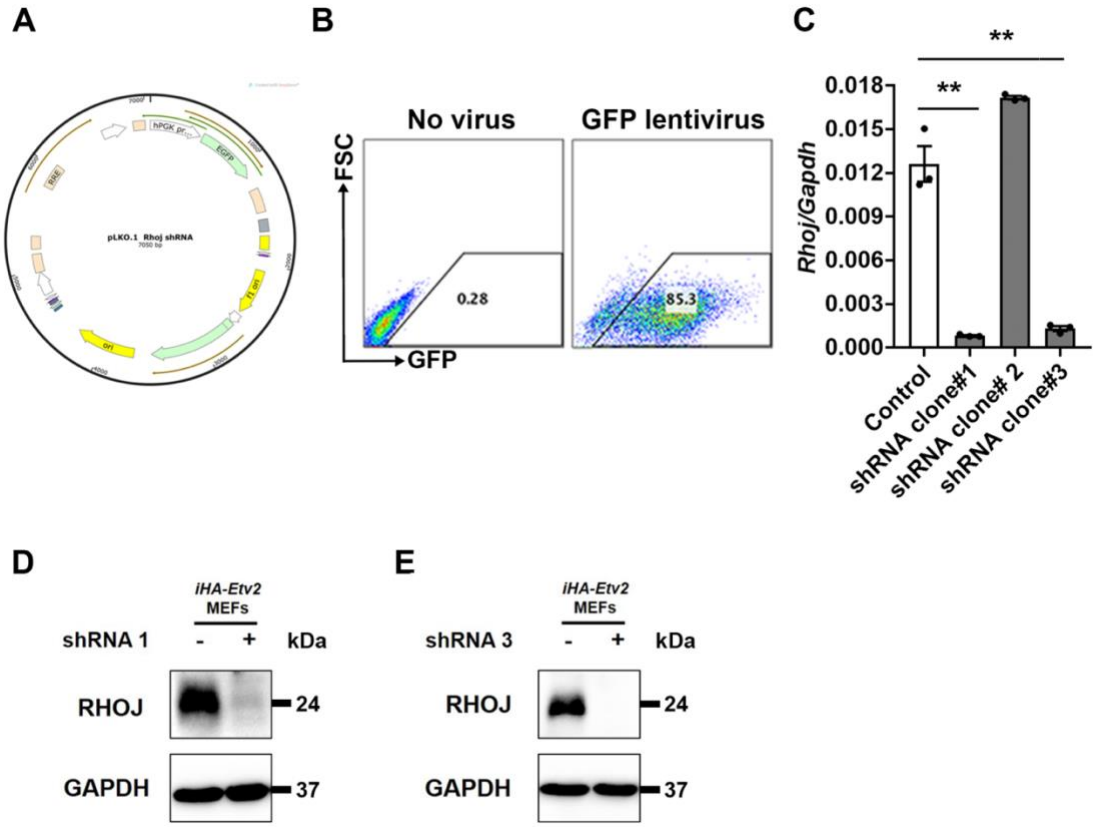
A



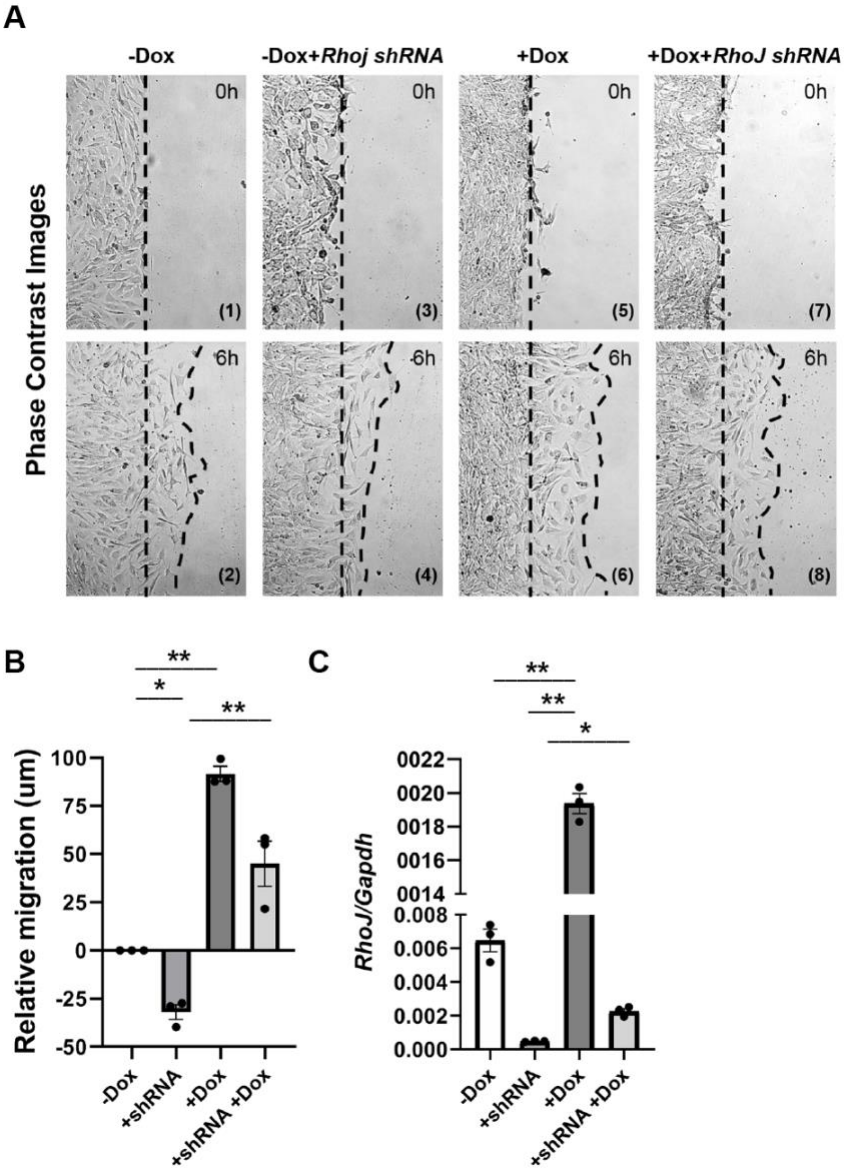
Supplementary Figure V. Expression analysis of *Rhoj* during ESC/EB differentiation. (A) qPCR analysis of *Rhoj* at various stages of ESC/EB differentiation (n = 3 replicates; **p < 0.01). Significance was determined by ONE-WAY ANOVA with multiple comparison. Data are presented as mean ± SEM.



Supplementary Figure VI. ETV2 binds to the upstream region of the *Rhoj* promoter. (A) Gel-shift assay showing ETV2 bound to the three Ets binding sites in the *Rhoj* promoter region. IRdye-labeled probes containing the putative binding sites were incubated with *in vitro* synthesized HA-ETV2 protein to form a specific complex with the oligo (lane 2; lane 5; and lane 8), which is competed with wildtype unlabeled oligos (lane 3; lane 6; and lane 9).



Supplementary Figure VII. Screening for *Rhoj* shRNA clones using lentiviruses expressing shRNAs against *Rhoj*. (A) Map of the lentiviral vector for the different shRNA clones of *Rhoj*. (B) FACS profile showing the percentage of infected fibroblasts as assayed by GFP+ cell percentage. (C) qPCR analysis of *Rhoj* following knockdown of *Rhoj* using three different shRNA clones. Note clones #1 and #3 were able to robustly reduce the levels of Rhoj transcripts (n = 3 replicates; **$p < 0.01$). (D, E) Western blot analysis of RHOJ in *iHA-Etv2* MEFs following treatment with shRNA #1 and #3, respectively. Significance was determined by one-way ANOVA with multiple comparison. Data are presented as mean ± SEM.



Supplementary Figure VIII. Validation of the ETV2-*Rhoj* network using a second shRNA (#3). (A) Migration (scratch) assay using the *iHA-Etv2* MEFs in the absence [-Dox (1, 2)] and presence [+Dox (5, 6)] alone or following infection with lentiviral vectors expressing *shRNA #3* for *Rhoj* [-Dox+shRNA #3 (3, 4)] and [+Dox+shRNA #3 (7, 8)] (n = 3 replicates). (B) Quantification of migration in *iHA-Etv2* MEFs in the conditions described in panel A. Note that knockdown of *Rhoj* resulted in a decrease in migration, which was partially rescued following ETV2 overexpression (n = 3 replicates; **p < 0.01). (C) qPCR analysis of *Rhoj* transcripts demonstrate that the induction of ETV2 resulted in increased expression of *Rhoj* (n = 3 replicates; **p < 0.05). Significance was determined by one-way ANOVA with multiple comparison. Data are presented as mean ± SEM.

Supplementary Table I. Bulk-RNAseq analysis for the cell migratory genes (GO:0048870) arranged based on the relative expression at 6h and 12h time point following -Dox and +Dox treatment during ES/EB differentiation.(137)

	6hEtv2_minus_HA_Dox	12hEtv2_minus_HA_Dox	6hEtv2_Dox	12hEtv2_Dox	pvalue
T	-1.76939897	-1.224116806	-0.359301215	-0.133598003	1.01E-85
FLT4	-1.016959657	-0.746895364	1.28534735	1.345196105	3.49E-59
SEMA6A	0.293596476	0.660969573	-1.071597016	-0.801258055	2.46E-55
HDAC7	-1.743030143	-1.373436808	0.772342204	1.108684354	5.69E-34
ENG	-1.137196226	-1.099932563	1.16297549	1.420105566	1.21E-27
MSX2	-1.87037454	-1.685412991	-0.274706267	-0.015007123	8.98E-26
PECAM1	1.281401521	0.939113002	0.924235201	0.838942308	9.76E-26
DOCK5	1.330740982	1.349613496	0.987072861	0.691291146	6.39E-25
EPHA2	1.522784757	1.514726304	-0.488302512	-0.807946392	1.15E-24
WASF2	-0.086321523	-0.501117085	1.927853132	1.137911414	7.40E-24
NRP2	-1.872964551	-1.493860883	0.512261028	0.869276719	8.27E-21
CRB2	-1.626665363	-1.341661066	-1.005623243	-0.229412321	2.66E-19
EGR1	-0.048489038	0.04657574	-0.952699128	-1.192644287	7.03E-18
PIK3R1	-1.3647752	-1.455273619	1.214226189	1.217212084	2.92E-16
AMOTL1	-1.360547286	-1.118008233	1.209442039	1.328293187	4.03E-16
MYO1C	-0.118442815	-0.297434161	1.04561572	1.42679185	5.17E-16
BMPER	-1.388782871	-1.983815648	-0.207382368	-0.370446417	5.48E-16
PDGFRB	-1.508992556	-1.384230389	-0.625550376	-0.834621918	7.06E-15
LRP1	-1.51635963	-1.000347345	-0.369238613	-0.504182773	2.68E-14
GPSM3	0.35871509	-0.294895399	1.229452219	1.342885255	1.24E-13
RHOJ	-1.936980155	-1.249224665	0.888946057	1.18729874	1.78E-13
VASH1	-0.907074134	-0.570069475	1.354321408	1.739876594	2.72E-13
FGF8	-1.573893714	-0.879248886	-0.822323303	-0.920250069	6.03E-13
MMP9	-1.479844687	-1.182289938	1.567022766	0.746719334	1.12E-12
AKAP12	1.479906705	1.561951653	0.544089576	0.1322819	1.45E-12
SLIT1	-1.492148884	-1.77059411	-0.198237876	-0.456822446	2.97E-12
NAV1	-1.398157561	-0.874017216	1.618459959	1.137525772	8.39E-12
LAMA3	1.166707028	-0.226916059	1.424639384	0.369533422	1.80E-11
BAX	1.11961546	0.98460321	0.702676368	0.935423539	3.89E-11
RBPJ	1.463800334	0.766640388	0.958176191	0.376753661	8.18E-11
COL18A1	1.550321828	1.373546358	0.478809969	0.247786163	8.48E-11
AMOTL2	-1.79248671	-0.683356193	-0.923256174	-0.699596267	4.93E-10
EFNB1	-0.558914225	-0.80418775	-1.63498714	-0.770976588	6.36E-10
LDB2	0.847923851	0.841259836	-1.47419343	-0.639890806	2.38E-09
MAGI2	0.191609178	0.598156914	-1.662108665	-1.72413128	4.87E-09
KIT	-2.125570629	-1.199291026	-0.161862584	0.004505814	5.84E-09

APCDD1	-1.567406714	-1.287417861	-1.020308136	-0.209319925	8.08E-09
SIRPA	-1.779617588	-0.321106529	-1.529334202	-0.220067221	8.84E-09
SCARB1	-1.566868275	-0.785826502	0.575913958	1.392139467	1.56E-08
GLIPR2	-1.246968833	-1.135336158	1.125304672	0.967763427	2.29E-08
WT1	1.776470822	1.45646615	0.586942304	-0.127007507	3.59E-08
SP1	-0.391248634	-0.444536041	2.347117312	0.810146535	4.38E-08
DDX58	-0.321594456	-0.952787337	2.085944962	0.640308851	7.25E-08
TNS3	1.837568512	1.218474592	0.92813266	-0.221899013	9.12E-08
HAS2	-1.567865819	-1.296602944	-1.139101887	-0.007311662	1.12E-07
PTP4A3	0.422855194	0.119670996	1.069404218	1.208755471	1.36E-07
EGFLAM	-2.211928057	-1.0052088	0.040816799	-0.045085081	4.18E-07
CD34	-0.106886822	-0.101521112	0.664664796	2.240722021	4.55E-07
SMAD3	-1.257833096	0.451238356	-1.490643834	-0.759055681	8.17E-07
TRP53INP1	0.593274002	1.239699014	1.511469107	0.909484409	8.78E-07
S1PR1	-1.528267614	-1.896731362	0.875416363	1.177943821	1.01E-06
PLXND1	-1.572121499	-1.563226665	0.810706008	1.435248593	1.15E-06
CNTN2	0.011430266	0.378267634	-1.224692829	-1.568797871	1.54E-06
INSM1	1.024164184	0.933372118	0.15180168	1.430908281	1.78E-06
PDE4B	0.918903036	0.740159779	1.717513806	0.266738675	2.33E-06
SRF	0.238900482	0.158274577	-1.566944893	-0.630294031	3.22E-06
APBB1	-0.756951222	-0.553263222	-1.653136264	-1.088551773	3.50E-06
LMO4	0.550910898	0.355658165	-1.563461922	-1.974252473	4.29E-06
CER1	-2.107703761	-1.247766682	-0.183471049	-0.322625754	5.01E-06
KALRN	-1.809222267	-1.702690385	0.502338532	0.684138003	6.46E-06
MCAM	1.476660072	1.443622755	0.89524486	0.364516308	7.00E-06
FSCN1	-2.011742772	-1.3535112	-0.420339994	0.284688712	8.08E-06
SH3KBP1	-1.361442593	-0.583744599	0.024512528	1.570374235	8.12E-06
RRAS	-1.253518898	-0.229752155	-1.272003212	-0.809596727	1.27E-05
NFATC2	0.892717848	0.166341577	1.327078368	0.893072898	1.72E-05
TEK	-1.141756687	-1.010516316	0.654805551	1.765666507	2.36E-05
SERPINF1	0.478342484	0.847689601	1.385785883	1.030266955	2.49E-05
RAPGEF3	-0.416565456	-0.237004967	1.738923144	1.402426594	2.76E-05
VCL	-0.393748155	-1.515169554	-0.1803301	-1.582184174	3.85E-05
PLA2G7	-0.246165384	-1.338311116	-0.245997719	-1.934949846	4.69E-05
RHBDF1	-0.333420138	-0.239992639	-0.085250508	1.211292148	4.97E-05
ARAP3	-1.497832434	-1.32801477	0.597567065	1.646960014	5.62E-05
JAK2	-1.330729335	-0.922480686	0.343327347	1.670942783	6.51E-05
VAV3	-1.338593972	-1.572831614	0.400433109	1.53855313	7.54E-05
SRGAP1	-1.103243067	-1.687135209	0.544142219	1.758353759	0.000111
PPARD	-0.782941619	-0.638303165	1.148048253	1.116500328	0.000119
ITGA3	1.255061746	0.995864191	1.251197802	0.314733887	0.000126

IQCG	1.524766569	1.15311259	0.836474053	0.918731985	0.000129
EMILIN1	-1.029651888	-1.674469483	0.406542222	1.942636125	0.00014
CSPG4	0.009050482	-0.965678691	-1.355809908	-0.74057858	0.000146
LYST	0.377390374	1.123570369	0.738800156	0.661517209	0.000151
IGF1R	-0.808104826	0.79773441	-0.675738383	-0.078789764	0.000169
PLVAP	0.293620149	0.258664445	0.257212854	2.243844888	0.000207
PLXNA4	-1.571085759	-1.95437975	0.027982117	-0.239654852	0.000276
BCAR1	1.872890084	1.731042631	0.058179751	0.068700603	0.000294
OSGIN1	-0.29881364	-0.340586834	1.205727104	0.592863379	0.000365
PDGFA	-1.715484788	-1.41527287	-0.738449358	-0.190777047	0.000378
MAP2K2	-1.765554363	-0.893872173	-0.691145653	-1.100107963	0.000389
ITGB7	1.215366476	0.492026066	0.647593354	0.851217458	0.000421
GCNT1	-1.800962444	-1.33217307	0.797799949	1.195951297	0.000451
ELMO1	0.812558758	0.628578396	1.17809684	1.013566513	0.000469
HSPG2	-0.645200399	0.18239527	0.248144735	-1.287438251	0.000497
FGR	-0.62182146	-0.131640586	1.394822573	1.145245563	0.000529
ICAM1	1.808848422	0.886653237	1.020710321	-0.301267473	0.000546
GPNMB	-0.58785631	-1.198584315	-1.714061532	-0.451027822	0.00059
EGFL7	-1.10538854	-0.641594248	0.472421462	1.884900679	0.000591
PRKCD	-1.145957338	-0.918688889	-1.309897011	-0.618653044	0.000607
PTPRJ	0.493799929	0.030233779	0.331709069	-1.684824252	0.000634
ITGB1	1.282617347	1.3152281	-1.273078157	-0.685290712	0.000744
SPNS2	-1.710859991	-1.147731876	0.709793729	1.742016549	0.000754
SFRP2	1.57446325	1.760758119	0.291385054	0.023841409	0.000771
TRP53	0.480189368	0.463568298	1.62051143	0.521372626	0.000789
CTTN	0.968690728	0.30446086	-1.846622573	-1.453783824	0.000808
F2R	-1.28710189	-1.317262619	0.293079341	1.43304399	0.001176
RIPK3	-0.804375805	-1.133573988	0.583242485	0.715508028	0.001303
LRP6	-1.131182673	-0.051441442	-0.290360342	0.318099515	0.001787
GATA2	-1.195209326	-1.438572673	-0.36841759	0.296087517	0.002045
VAV1	-0.879237783	-1.342104378	1.260546276	1.033149666	0.002487
LRP5	-0.254516857	0.160962071	0.953594026	1.218827613	0.002743
TGFB1	-1.849243731	-1.480608339	0.016234653	0.321480958	0.003591
PMP22	-2.205493661	-1.317872606	-0.044004384	-0.060671435	0.003636

SRCIN1	0.922984898	1.591142461	0.584086443	0.570837052	0.003638
ARC	1.802850138	1.412806687	0.45135065	-0.224110205	0.004019
ARF6	0.309984935	-0.125669355	0.428831392	2.215122703	0.004091
BCL11B	0.46773255	0.615639851	-1.951147082	-0.166636042	0.004117
NOTCH1	-0.953306363	-0.928172938	0.812919512	1.188576013	0.00425
SPATA13	-0.17172874	-0.533514307	1.73355989	0.688777732	0.004332
PTPRM	-1.929596791	-1.512589166	-0.150698301	0.150764555	0.004431
EPS8	0.65276909	1.046415549	0.583804717	1.690528997	0.00461
APPL2	-0.829283652	-0.810893713	0.563235062	1.067013003	0.004991
ACTR3	0.133607457	0.000556189	0.850580125	1.420081316	0.005105
MAP2K5	-0.214421167	0.048928642	-1.420407687	-0.221331361	0.006882
SWAP70	1.147263128	1.809364648	-0.276917078	0.761278937	0.006883
ABI2	0.79058045	1.143151089	0.173552832	-1.452393948	0.00775
PHACTR4	-1.18333858	-0.831930432	-0.303207175	-0.553664145	0.008573
CDC42BPB	-1.172184568	-1.475786556	0.433337062	1.083390573	0.008593
PTPRF	1.640548663	1.720333408	-0.227347785	-0.076446052	0.009237
WNT11	-2.203432167	-1.231595346	-0.302144167	0.149800213	0.009327
GSK3A	0.796456357	0.804991173	-1.449250907	-1.750142884	0.010298
CD40	-2.211593936	-1.137718065	0.036552192	0.901176555	0.01055
PSEN1	-1.693297253	-1.395534275	-0.479165046	0.05753871	0.011046
ST14	1.109728322	1.8836235	-0.199236442	0.171779286	0.011858
EPHA1	-1.529005102	-0.935686341	0.342699266	-0.546744881	0.011945
PLAA	0.086989949	0.129431645	1.394780017	1.339603931	0.012408
RND2	0.148548026	-0.104803795	0.059081812	0.512127392	0.012907
CDH5	-1.544188832	-1.504840129	0.588846071	1.521961341	0.013263
LGALS9	1.826507763	1.543551516	0.167700218	-0.429115138	0.014111
AU040320	-1.093269371	-0.298979673	-1.652694241	0.429866742	0.014324
EPHB4	-1.221065712	-0.684803134	-0.261920069	1.172008949	0.015846
EMX2	-1.002929989	-1.806214384	-1.254295433	0.031165312	0.017163
DAB2IP	-1.521730437	-1.323388532	1.114858869	1.637546635	0.018739
SPRY2	1.842671955	1.86568802	-0.313169156	-0.673469958	0.019502
ANXA6	-1.244790848	-0.839306807	0.212684326	1.299479698	0.020032
KANK2	1.076075002	1.135435459	0.272063822	0.039213213	0.020452
P2RY2	-0.742711273	-0.782205324	0.80749464	1.539288433	0.021119

ITGA2B	-2.01620778	-1.659719631	0.53543182	0.538785224	0.021595
STRAP	-0.498924919	-1.197672619	-0.022886554	-0.312530374	0.022031
TRIM32	-1.77928415	-1.432483408	-0.647763494	0.234217901	0.022605
PLXNC1	-1.884752107	-1.593966577	0.500233205	0.936236235	0.022902
TIE1	-1.524437072	-1.932420595	0.454569347	1.324181156	0.02548
SPEF1	0.519635702	-0.288709544	-1.970671137	-0.42102248	0.029753
NOD2	-1.110544181	-2.306576459	-0.106681913	0.186554921	0.032262
ARPC2	-1.414062783	-1.256825298	0.539828776	1.535539726	0.035972
GFRA3	-1.796247593	-1.796247593	-0.269975596	0.638948164	0.037982
CD300A	0.495622281	-0.019358118	1.141770846	0.426177169	0.039039
ARHGAP18	-1.439527824	-0.404706884	-0.363560623	-0.074414956	0.039816
RAC1	-0.182341574	-0.422100451	-0.922951671	-0.279969511	0.040775
MAP3K1	-2.066769783	-1.279270036	-0.131099461	-0.064936818	0.041616
NCKAP1	-0.394140922	-0.623364685	-0.391495573	1.89369439	0.042107
ZSWIM6	1.411748689	0.227307393	1.601679584	0.204127598	0.042484
PRKD2	-1.79255545	-1.210271829	1.046301371	0.475671938	0.043676
SEMA3F	-1.604444633	-1.831871008	0.428996045	0.534081071	0.046556
FBLN1	-2.112791006	-1.422780255	0.027483553	0.473453933	0.052648
SEMA6B	-1.681380619	-1.78186298	-0.096983766	-0.252486684	0.054067
CCR4	0.371491162	-0.168840754	0.953016192	1.182742308	0.058203
MYC	0.52323323	-0.930959939	-0.385499949	-1.252199145	0.060176
WNK1	-1.084483928	-0.729676295	0.466399049	-0.250065659	0.06282
MAP4K4	-1.944252742	-1.329499539	0.015265391	0.897732804	0.064206
FUBP1	1.32218532	0.919377683	-1.433230412	0.191420051	0.065161
RAP2A	-1.815836023	-1.863886274	0.222646417	0.235130614	0.06568
CDH13	0.549677776	0.778827913	0.008314134	-2.711125495	0.066393
TRPV4	-1.219085917	-1.241466697	0.78442503	1.125025435	0.071089
PAK2	1.606501913	1.277139716	0.254015157	0.508741035	0.074283
TRIB1	-1.154452502	-0.391098257	0.522690052	0.634836383	0.076456
SERPINE1	-1.49868225	-0.97917721	1.729637308	1.094319683	0.07754
KDR	-1.911502299	-1.783868238	0.541407936	0.623778934	0.081588
ZC3H12A	-0.740497447	-0.403227281	1.436470823	-0.03847924	0.087164
SRC	-1.694174907	-1.248692073	-0.879898273	-0.049085501	0.088548
CACNA1E	-0.574266472	-2.303114935	0.658341712	1.064427493	0.0905
CD47	-0.359050635	-1.13985934	0.270525027	1.422323151	0.096941

PAWR	0.982856612	0.495768092	-1.422476817	-0.438827026	0.099207
ANLN	0.985306158	-0.330730648	-0.75340811	0.70497293	0.104922
GNAI2	-1.849297404	-1.042968422	-0.360861221	0.555579063	0.110756
CDK5R1	0.041194345	-0.642575607	1.017284424	1.185497533	0.114018
PRKCE	-1.552650041	-1.220612832	-1.27622228	0.388803918	0.117617
AVL9	0.634206225	1.690923522	0.340791104	0.628575649	0.119836
GAS6	-1.570927021	-1.721360923	-0.174136084	0.819814879	0.123623
MBOAT7	0.187758245	-0.104051184	-0.603548383	-1.798602618	0.126151
SCHIP1	1.94887225	1.354134916	0.352873384	-0.568138227	0.129236
LIMCH1	-1.361512312	-1.968273513	0.142163197	1.04513041	0.133637
PPP3CA	-1.899197911	-1.544298515	0.110890966	0.287344491	0.141006
MADCAM1	0.524648756	1.213914277	-0.372293555	-1.204824255	0.149999
IER2	1.599644995	1.539419696	0.32415857	0.195992006	0.150931
DDIT4	-1.516344191	-0.68787485	-0.132467565	-1.008126899	0.154097
EGFR	0.32559466	-0.368188274	-1.615518326	-1.188962604	0.156177
ZEB2	-1.834427451	-1.482057983	-0.577720662	0.579467308	0.157193
NDEL1	-0.544189958	0.54658392	-0.20758588	0.659384396	0.157363
CLIC4	-1.462021399	-0.714335107	0.053141353	0.038934444	0.157938
GBF1	-1.689051572	-1.227120609	0.593263778	-0.615674857	0.158576
BMPR2	-1.102884543	-1.091296516	0.720312545	0.76721543	0.160364
TGFBR2	-1.316087812	-1.810063869	0.419778669	0.928272438	0.165582
ITGA2	0.354593926	-1.746871364	0.681007186	1.182257461	0.166117
DAPK3	-1.553077263	0.664121435	-0.977535997	-0.874619414	0.16733
CORO1B	-0.516857742	0.595075111	-0.893727537	1.615624823	0.168754
RERE	-1.004318091	0.330155435	0.53473465	-0.097257673	0.169631
SPECC1L	-0.912261852	-0.659169791	-0.141257422	2.013213973	0.174926
MTOR	-0.658855276	-1.921231803	0.44534715	-0.698815029	0.176277
MGAT3	1.33107426	0.946175522	-2.011248807	-0.80652178	0.177857
PFN1	-0.627962392	0.272201207	0.647997681	0.983035258	0.178187
IL24	1.120354518	1.062816726	-0.755957747	-0.755957747	0.188482
TWIST2	-0.215957162	-0.907363925	-0.624126907	-1.773151814	0.194831
SYNE2	-0.37326566	0.985511765	-0.482606941	0.565828668	0.199845
GTPBP4	1.213454144	0.087182634	1.076525256	0.481618939	0.203281
ADCY3	-1.155714368	-1.628760104	0.658640104	1.115202281	0.206724

ITGB2	-2.32186932	-1.148946513	0.774422655	0.557014415	0.208802
APBB2	1.381916077	1.082700613	0.330782469	-1.291675529	0.210693
CAP1	-0.53891182	-0.524958469	-1.214103202	1.605614905	0.216166
IL16	-0.327046877	-2.730272878	0.626235212	0.603084756	0.219651
MAP2K3	1.792457088	-0.975686936	0.121166679	0.729681651	0.222803
CDH1	1.773300867	1.714042361	0.130731737	-0.392790433	0.225938
PRPF40A	0.606131602	-0.002031536	1.53992009	0.248754522	0.243359
ATP5A1	1.058195523	0.682250958	0.080669007	0.082231413	0.251364
PAFAH1B1	0.593039035	1.947376335	-1.103350966	-0.172885377	0.256638
ARL13B	-0.906635975	-0.126867208	0.189361835	1.027344407	0.259073
CTSH	-2.175053539	-1.137008206	0.369979448	1.028247093	0.272422
NR4A1	-0.12615441	0.333425211	-1.139084206	-0.16262973	0.282257
WWC1	1.175638145	1.678115002	-0.608176568	-0.539769684	0.282307
PKN3	-0.103891173	-0.486103192	-0.689828494	2.121339013	0.283067
PRR5	-1.859773719	-1.355424405	0.100623784	0.084028144	0.284424
STAT5A	-1.779266323	-0.70826814	-0.602814415	1.018298379	0.287262
PIN1	0.659051473	0.73679589	-0.65376816	-0.999658461	0.301951
NOX4	-0.090831787	-0.06832877	0.857479264	1.344289237	0.304
ELMO2	-1.582554041	-1.388020927	-0.568397626	1.328563593	0.307785
MYO10	1.732385854	1.065367961	0.370678767	0.005954823	0.308402
ITGB4	1.849978428	1.278075181	-0.259299273	0.12403421	0.309196
CADM4	0.942734	1.337186571	-0.629821781	-0.710011172	0.309591
DUSP22	-0.022071308	-0.420906549	-1.504270136	-0.534877297	0.322733
BBS4	0.97908319	0.875275732	0.094100343	-1.02875878	0.324149
CARD10	-2.063612724	-1.202245095	-0.29059684	0.069093188	0.330112
GAB2	-1.975715147	-1.094132938	0.003458282	1.555084973	0.330641
ARPC5	-0.498517679	-0.343776117	0.162886799	0.723071743	0.333458
SDCCAG8	-0.521226625	0.850594702	-1.027619996	-0.886299805	0.337337
PTPRO	0.182515572	-2.536469569	0.65093387	0.809215979	0.342552
IQCF1	1.290566487	-0.740185138	-0.740185138	-0.740185138	0.345445
DCC	-0.395785747	-0.250769569	0.763306186	0.435590742	0.35001
FAT2	-0.568173466	-0.236653308	1.258989015	-0.690649794	0.356032
CORO6	1.596745644	1.177871636	0.033245092	-0.656809127	0.361747
SKAP1	-1.548420471	-1.121142458	-0.06753126	0.710083973	0.363867

PTP4A1	0.324833756	-0.838349012	0.12055886	0.769220445	0.36498
ARF4	1.57202991	-0.215529393	-0.199676063	-0.88863492	0.384394
NTN5	-0.50022144	-0.124589551	0.831583857	1.574442353	0.387919
PDCD10	0.54717773	1.122376005	-0.250430602	1.334800094	0.399104
APOA1	-0.957033904	-2.488157015	0.468873909	0.755216207	0.401433
CD200	1.346418859	0.011773435	-0.508954306	0.804863934	0.42284
TRPM2	0.745843733	0.151018653	-2.478372419	0.695695051	0.42486
CX3CL1	-1.736998609	-2.006288254	0.425327292	0.316375425	0.425434
DEPDC1B	-1.589827691	-1.928648399	0.397550292	0.846074558	0.430568
ATP7A	0.770914303	-0.41671685	-0.482847134	1.60917567	0.468381
ACTB	-0.838049773	-0.703934804	0.067289036	-1.574125407	0.470734
PTPRG	-0.170096757	0.051760402	0.50788727	-0.245306323	0.475069
RHOD	-2.415014922	-0.232365336	-0.264891418	0.604443742	0.475138
SCRIB	0.393919756	0.598567301	-0.398598671	0.53854752	0.477222
RAB13	-1.815442799	-0.52951473	-0.814084579	0.372189209	0.481649
RPS19	-0.302658485	2.112571958	-0.334321378	-0.219503159	0.496207
JUN	-1.379943556	-0.30829282	-1.010750522	1.543345359	0.500316
TBX1	-0.378530284	-0.877641389	0.105746738	1.817499422	0.502619
RFFL	-0.560461723	-1.00867569	0.054566309	0.280723729	0.508602
P2RX4	-2.080614227	-1.035199939	-0.00218153	1.014315723	0.511812
CCL8	-0.455874348	-0.455874348	-0.455874348	-0.455874348	0.515088
PLAT	-1.043097987	-1.619294934	-0.308435592	0.495494112	0.525482
ITGA9	-1.418637542	-1.777199815	-0.270250944	0.867348495	0.534447
SLC9A3R1	-0.601613768	-0.009332293	0.599679392	-0.315569614	0.53629
IL12A	-0.316227766	-0.316227766	-0.316227766	-0.316227766	0.537422
ITGB6	-0.316227766	-0.316227766	-0.316227766	-0.316227766	0.537422
GJA1	0.481567874	1.517129126	0.001934202	-1.224121708	0.546994
ACTA2	1.144925352	1.474666223	-0.226171366	-1.398697462	0.55434
PARD3	-0.524213672	-0.922694636	-0.962284288	0.746875094	0.558037
PTPN23	-0.663803574	1.063917809	0.161710414	-0.412978187	0.558883
ARHGEF2	-0.954322958	-0.529500101	-0.055405576	-0.141825011	0.56745
CSF1	-1.349570076	-0.92682497	-1.391490734	0.698785377	0.569037
EPHB3	1.779300215	0.94707621	0.23085844	-0.902105338	0.577183
SPARC	-2.068517567	-1.570915997	0.192147137	0.646140685	0.585706
PIK3CA	-1.290063058	-0.963617771	0.796112704	0.57031247	0.604178

BMP2	-1.110570443	-2.456132814	0.354285201	0.519748314	0.605919
ANGPT2	0.433257572	-1.160981341	0.986648646	0.800280416	0.616187
DNAIC1	1.831668763	1.088098009	-0.195913439	0.700347326	0.623415
SSTR4	-1.090469512	-1.090469512	0.074313222	1.397078645	0.644283
LRRC6	-1.726532492	0.26766171	0.986776724	-1.726532492	0.657699
PAX3	0.58944413	-1.665189376	0.445595581	0.785525169	0.661381
LDHC	-2.538316816	-0.353076343	0.106059576	0.27514699	0.66487
CXCL14	-0.449592968	1.153150253	-0.486615413	0.130082561	0.694511
CORO1C	1.408109834	0.846496102	0.758208037	-1.037103567	0.695611
IFT46	-1.490774453	-1.845721003	0.301687871	-0.224197007	0.705717
HYAL2	-1.809994007	-1.662086734	0.193468481	0.721714928	0.715616
TNN	1.412202895	-0.892048135	0.860541079	-0.892048135	0.743174
CXCR5	-1.734262004	0.454640313	0.839809812	-1.734262004	0.766917
TMIGD1	-0.468635077	-0.468635077	-0.468635077	2.202600202	0.770167
DLC1	-0.68245634	-1.81182239	1.209743131	0.023941797	0.777676
UTS2	-0.316227766	-0.316227766	-0.316227766	-0.316227766	0.779579
TBXA2R	-1.774618978	-1.599933421	-0.241554057	0.198225816	0.783197
P2RY1	-1.630546993	-0.688022731	-0.537560931	1.413843366	0.799791
MEF2C	-1.673370819	-1.875151453	-0.151864493	1.176105679	0.802858
CCR7	-2.471092359	-0.022463661	-0.251417675	0.183870242	0.803992
RDX	0.298214271	-0.483327282	-0.793585042	0.722651104	0.807327
SOCS7	-0.909175356	0.230773147	-1.131291133	-0.734018986	0.816222
DPCD	-1.754353357	-1.397675285	-0.482975033	0.842112218	0.823394
ADAM17	0.412956872	-1.411900766	-0.556183881	1.260821937	0.823718
P2RY12	-0.463169347	-0.463169347	2.310370651	-0.463169347	0.833621
PLXNB2	-1.583856481	-1.252828562	0.997040108	0.265381815	0.848608
ASPM	-0.133943897	-0.029808653	0.126860755	0.5326141	0.868813
CCR1	-0.905323228	-0.905323228	0.506764197	-0.905323228	0.889899
FLT1	-1.908786384	-1.562133868	-0.13475813	0.673602214	0.901006
HRH1	0.489411016	-1.080963459	-0.299059758	1.600692648	0.905867
ASAP3	0.997996647	1.687706357	-0.219164439	0.045183972	0.922285
RNF20	0.19174501	0.703138113	-0.854026565	1.350316679	0.933783
CREB3	1.624273235	-1.078274969	-1.1297397	0.287774632	0.959918
TMEM18	-1.695109713	-1.350560724	0.34569884	-0.655825336	0.968534

DCDC2A	-1.370748016	0.188822467	-0.232505412	1.689776347	0.973502
EDNRB	0.927784633	1.378220834	0.159975332	-0.478375016	0.97952
GPR18	2.846049894	-0.316227766	-0.316227766	-0.316227766	1
BMP10	-0.316227766	-0.316227766	-0.316227766	-0.316227766	1
PF4	-0.316227766	-0.316227766	-0.316227766	-0.316227766	1

Supplementary Table II. ETV2 ChIPseq peak analysis of migratory genes.(49)

	seqnames	start	end	SYMBOL	distancetoFeature
10	chr11	6542177	6542325	Itga2b	4058
43	chr11	20179101	20179185	Pecam1	15634
79	chr11	32486915	32486985	Cd300a	31545
82	chr11	32642365	32642539	Slc9a3r1	-359
94	chr11	35024502	35024709	Itgb4	-96722
150	chr11	50240269	50240582	Adcy3	-1653
165	chr11	51822847	51823130	Adam17	34806
175	chr11	53335955	53336103	Tmem18	9722
201	chr11	57158892	57158974	Arf6	147505
214	chr11	59080786	59080925	Rhoj	58384
218	chr11	59458919	59459067	Syne2	-7733
244	chr11	62661050	62661203	Psen1	12386
247	chr1	59765389	59765597	Arpc2	1989
256	chr11	63122034	63122182	Vash1	-6948
290	chr11	68901535	68901772	Bcl11b	-48
297	chr11	69245807	69245991	Cdc42bpb	5208
320	chr11	70306245	70306472	Gtpbp4	16383
326	chr11	70605132	70605255	Lyst	8079
330	chr11	70653505	70653887	Elmo1	1139
345	chr11	71995937	71996278	Dcdc2a	-26818
354	chr11	72535971	72536040	Dusp22	14535
380	chr1	62665039	62665311	Pax3	4200
395	chr11	76893580	76893846	Msx2	-8579
407	chr11	78245658	78245876	Cxcl14	-88
439	chr11	82183328	82183548	Mef2c	68143
455	chr11	86757450	86757650	F2r	115
484	chr11	93310099	93310168	Pik3r1	103225
496	chr11	95104373	95104547	Zswim6	-15746
499	chr11	95688548	95688755	Depdc1b	10595
505	chr11	96166854	96166927	Map3k1	-10697
509	chr11	96562327	96562491	Itga2	97734
516	chr11	96777674	96777743	Ptprg	-115
528	chr11	97825339	97825487	Vcl	-15441
541	chr11	98963612	98963760	Arf4	17168
548	chr11	99126935	99127177	Prkcd	28142
597	chr11	102877478	102877547	Ripk3	3507
605	chr11	103286133	103286214	Spata13	-18388
632	chr11	106310328	106310397	Dock5	4434

661	chr11	109439986	109440167	Rgcc	-1841
676	chr11	114244887	114245035	Ednrb	-6385
682	chr11	114679280	114679557	Spry2	3849
690	chr11	115219426	115219635	Rap2a	-31567
696	chr11	115475528	115475729	Gpr18	-139
710	chr11	116218977	116219054	Egflam	2974
719	chr11	116759191	116759262	Myo10	7024
724	chr1	73923241	73923552	Twist2	-58775
759	chr11	120281323	120281570	Has2	6836
772	chr11	121658443	121658591	Trib1	14710
776	chr12	3235284	3235539	Myc	-220
781	chr12	3239372	3239454	Lrrc6	3868
786	chr12	4234218	4234384	Ago2	76
789	chr12	4391336	4391493	Ptp4a3	-3146
793	chr12	8208072	8208215	Arc	-35
803	chr12	12593436	12593742	Scrib	76481
823	chr12	21119623	21119793	Card10	76664
833	chr12	25240301	25240598	Mgat3	-42294
856	chr12	31917330	31917547	Prr5	33205
860	chr12	32142782	32142859	Fbln1	-19080
875	chr12	36521881	36522174	Plxnb2	-17476
889	chr12	54102563	54102811	Rapgef3	4322
893	chr12	54187743	54187961	Hdac7	16117
911	chr12	65857356	65857504	Bin2	-111020
918	chr12	69296597	69296745	Nr4a1	-84
921	chr12	69406968	69407037	Itgb7	-34500
925	chr12	69761787	69761980	Sp1	2157
970	chr12	76216771	76217051	Tbx1	29975
980	chr12	77762196	77762267	Ephb3	52613
1016	chr12	83073759	83073832	Pak2	25115
1021	chr12	84285224	84285495	Iqcg	-8
1026	chr12	85094044	85094115	Kalrn	7670
1050	chr12	89953320	89953691	Cd200	-114779
1054	chr12	91728543	91728612	Cd47	-42430
1067	chr12	99971740	99971809	Arl13b	7241
1114	chr1	87755792	87756079	Actr3	-78
1117	chr12	110155321	110155492	T	2756
1174	chr13	3768661	3769055	Ppard	-1939
1206	chr13	17879088	17879302	Gpsm3	-15767
1241	chr13	25050473	25050593	Pla2g7	-5531
1255	chr13	30045237	30045309	Srf	-13440

1285	chr13	35968894	35968969	Sema6b	-3886
1296	chr13	38527837	38527969	Vav1	434
1300	chr13	40892094	40892394	Fer	-17905
1302	chr1	90599320	90599559	Il24	-3661
1307	chr13	41703661	41703988	Ptprm	54218
1342	chr13	49276811	49277064	Prkce	-28105
1372	chr13	53920116	53920310	Lama3	-4368
1402	chr1	93624561	93624721	Cntn2	-10632
1414	chr13	60670643	60670891	Gfra3	28853
1415	chr13	60738028	60738171	Egr1	-33647
1431	chr13	64432461	64432530	Arap3	-59
1445	chr13	70907650	70907813	Sema6a	24702
1471	chr13	83572074	83572347	Pdgfrb	68040
1484	chr13	89496626	89496859	Apcdd1	-43170
1500	chr13	95161643	95161717	Dcc	-46469
1519	chr13	97984149	97984297	Atp5a1	-107487
1534	chr13	99414487	99414576	Lrp5	-1611
1539	chr13	99814219	99814303	Coro1b	86872
1541	chr13	99851102	99851274	Rhod	49989
1580	chr1	105544680	105544863	Nav1	19898
1595	chr13	108980350	108980430	Gcnt1	309747
1612	chr13	110903374	110903554	Jak2	-125
1621	chr13	112084512	112084745	Acta2	-53604
1640	chr14	7991194	7991342	Slit1	3379
1649	chr14	10289149	10289384	Dpcd	90799
1650	chr14	10453587	10453675	Fgf8	-25087
1653	chr14	11868706	11869066	Gbf1	1112
1685	chr14	21032222	21032483	Emx2	20286
1713	chr1	119125006	119125120	Aspm	-32580
1761	chr14	34918732	34918856	Notch1	24123
1762	chr14	35192971	35193042	Egfl7	26467
1787	chr14	46679058	46679376	Pkn3	-5283
1808	chr14	52305006	52305223	Eng	122
1826	chr14	55634159	55634289	Dab2ip	1798
1832	chr14	56393771	56394016	Crb2	-2209
1843	chr14	58812739	58812897	Zeb2	67489
1857	chr1	125559389	125559490	Arpc5	-1206
1858	chr14	61804627	61804775	Prpf40a	-116749
1864	chr14	62526783	62526890	Itgb6	5760
1890	chr1	23518329	23518449	Ptp4a1	38790
1924	chr1	127230482	127230841	Cacna1e	-11609

1930	chr14	70153660	70153967	Nckap1	151
1949	chr14	73123819	73124007	Ptprj	782
1996	chr14	102810141	102810299	Wt1	4042
2055	chr14	124277607	124277764	Sirpa	-121084
2070	chr15	7010425	7010617	Spef1	14388
2085	chr15	10869931	10870025	Bmp2	-47506
2090	chr1	131249083	131249197	Tnn	-3825
2110	chr15	28105995	28106287	Insm1	12457
2116	chr15	30610682	30610848	Sstr4	9168
2160	chr15	43838335	43838483	Src	29874
2197	chr15	59567944	59568113	Mmp9	-50646
2199	chr15	59638151	59638404	Cd40	-10199
2203	chr15	59685342	59685411	Elmo2	-13229
2230	chr15	68072492	68072679	Nfatc2	-129688
2277	chr15	76617295	76617532	Pex2	89
2329	chr15	81803134	81803229	Pik3ca	-521
2374	chr15	86214346	86214496	P2ry12	-113
2376	chr15	86813873	86814095	P2ry1	76047
2384	chr15	88875439	88875508	Schip1	13253
2386	chr15	89170211	89170430	Il12a	-7303
2394	chr15	89544207	89544297	Pdcd10	-20530
2410	chr15	96533822	96533985	Sfrp2	35920
2429	chr15	98050604	98050680	Arhgef2	-19148
2445	chr15	99333051	99333228	Rab13	16484
2452	chr15	100039125	100039209	S100a8	461
2499	chr16	4288409	4288500	Csf1	32271
2504	chr16	4648259	4648340	Vav3	8270
2505	chr16	4656577	4656729	S1pr1	16588
2523	chr16	9162727	9162875	Ppp3ca	3957
2524	chr1	136694949	136695227	Sdccag8	-1262
2533	chr16	11008126	11008552	Lmo4	13740
2547	chr16	14391624	14391704	Fubp1	8817
2565	chr16	18718200	18718363	Trp53inp1	-35984
2606	chr16	26737167	26737315	Ddx58	18335
2614	chr16	29911394	29911568	Dnaic1	35360
2618	chr16	30308833	30308915	Creb3	1946
2619	chr16	30458327	30458413	Glipr2	5739
2641	chr16	31200154	31200334	Rnf20	1091
2684	chr16	37731192	37731316	Trim32	-33852
2701	chr16	42456716	42456915	Cer1	-116065
2737	chr16	57582123	57582298	Plaa	24741

2739	chr16	57915594	57915717	Tek	12661
2740	chr16	59498750	59498829	Jun	26975
2773	chr16	89830278	89830468	Pde4b	-196138
2832	chr17	5090085	5090154	Ptprf	22025
2834	chr17	5283044	5283213	Tie1	-105719
2837	chr17	5425333	5425564	Cfap57	3869
2844	chr17	6173102	6173249	Cap1	25683
2863	chr17	10395930	10396081	Zc3h12a	-60286
2870	chr17	10740243	10740364	AU040320	-67591
2906	chr17	24162663	24162739	Phactr4	1103
2912	chr17	25064514	25064707	Fgr	19077
2915	chr17	25184447	25184544	Wasf2	-107
2936	chr17	26547856	26548043	Clic4	5096
2947	chr17	27416633	27416719	Asap3	18486
2951	chr17	27820780	27820869	Hspg2	-132
2972	chr17	29315759	29316024	Epha2	-1921
2994	chr17	31217900	31218022	Mtor	10027
3003	chr17	32771003	32771072	Rere	-16805
3007	chr17	33754694	33754938	Uts2	-5298
3048	chr17	35916285	35916798	Magi2	104
3076	chr17	44160622	44160784	Emilin1	25854
3115	chr17	50331817	50332114	Ldb2	-7435
3122	chr17	51258349	51258595	Rbpj	-21287
3134	chr17	56005572	56005751	Apbb2	12
3137	chr17	56133797	56133907	Limch1	6546
3148	chr17	56442727	56442797	Kit	-23818
3150	chr17	56613325	56613535	Kdr	-91
3167	chr17	66344218	66344302	Pf4	-24706
3168	chr1	160646159	160646230	Cd34	586
3217	chr17	78903263	78903411	Coro1c	3729
3228	chr17	83113771	83114015	Trpv4	6970
3246	chr1	163443923	163444238	Akap12	-40254
3263	chr17	88551085	88551274	P2rx4	18058
3284	chr18	5424462	5424704	Scarb1	-53980
3315	chr18	11817216	11817441	Serpine1	-22004
3317	chr18	12109845	12110003	Ephb4	11692
3326	chr18	12956193	12956301	Pdgfa	341
3328	chr18	13495967	13496049	Gper1	105433
3344	chr18	17680715	17680881	Actb	-4596
3345	chr18	20609981	20610129	Fscn1	51907
3351	chr18	20996827	20996989	Rac1	-4514

3359	chr18	23815583	23815803	Flt1	1788
3423	chr18	35730396	35730605	Plxna4	-8040
3448	chr18	37997578	37997733	Epha1	1391
3454	chr18	38285060	38285315	Gpnmb	523
3478	chr18	46280772	46280841	Avl9	221
3519	chr18	61060803	61060974	Bmp10	15653
3525	chr18	61550443	61550621	Gata2	-4831
3560	chr18	67299822	67299894	Hrh1	10636
3565	chr18	68131325	68131537	Plxnd1	-12951
3578	chr18	71647094	71647242	Wnk1	703975
3621	chr18	80628799	80628960	Lrp6	-22988
3624	chr1	172829106	172829306	Arhgap18	14532
3630	chr18	82244834	82244950	Ptpro	13676
3631	chr18	82479929	82480030	Eps8	4783
3632	chr18	82529304	82529399	Strap	-5283
3654	chr19	3847465	3847534	Mboat7	-4308
3679	chr19	5661543	5661813	Prkd2	2164
3695	chr19	6558066	6558263	Cadm4	-5714
3696	chr19	6666913	6666993	Rps19	7635
3703	chr19	7400714	7400932	Gsk3a	-1637
3706	chr19	8702682	8702769	Tgfb1	-20519
3769	chr19	28695268	28695348	Rras	-15191
3776	chr19	29641170	29641318	Bax	7245
3781	chr19	30371980	30372193	Ntn5	107100
3788	chr19	33083063	33083138	Ldhc	12991
3807	chr19	37493966	37494374	Igf1r	-26913
3838	chr19	42676991	42677193	Il16	22911
3841	chr19	43896403	43896483	Cemip	-24070
3843	chr19	44281144	44281233	Nox4	8873
3853	chr19	45767953	45768114	Gab2	15884
3861	chr19	46621924	46622120	Wnt11	22840
3875	chr19	47951181	47951337	P2ry2	13506
3883	chr19	53263968	53264227	Apbb1	1437
3889	chr19	55011459	55011614	Swap70	-74735
3946	chr1	180861143	180861336	Gja1	-6685
3997	chr2	20693659	20693872	Cttn	19575
4011	chr2	25171742	25171866	Gas6	-9016
4021	chr2	26475095	26475358	Angpt2	-4976
4031	chr2	26578633	26578761	Plat	-1381
4046	chr1	182670598	182670848	Ddit4	-9341
4055	chr2	29365281	29365350	Dlc1	18462

4097	chr2	32715011	32715389	Plvap	-1935
4136	chr2	37780267	37780424	Ier2	4018
4153	chr2	48397229	48397438	Nod2	9192
4167	chr2	53098356	53098446	Cx3cl1	93346
4174	chr2	60810716	60810911	Cfap20	-88073
4184	chr2	62441025	62441282	Cdh5	117
4208	chr2	71055700	71055854	Cdh1	-117
4237	chr2	73834815	73834891	Bcar1	-6984
4257	chr1	187447770	187447933	Specc1l	-107
4263	chr2	80581652	80581721	Cdh13	-272
4265	chr2	80771512	80771615	Osgin1	63923
4315	chr2	92618097	92618301	Pard3	18390
4320	chr2	92680934	92681083	Itgb1	-61296
4328	chr2	93214253	93214401	Amotl1	26705
4334	chr1	37344670	37344848	Map4k4	-10498
4335	chr1	189731802	189731883	Col18a1	3534
4340	chr2	93791774	93791940	Pin1	-2796
4342	chr2	93849559	93849805	Icam1	384
4351	chr2	94583456	94583527	Kank2	7965
4354	chr2	102037135	102037387	Anln	-2494
4358	chr2	103463556	103463653	Bmper	21604
4362	chr2	103900506	103900627	St14	-11518
4379	chr1	190032802	190032951	Itgb2	-110439
4390	chr1	190419609	190419757	Trpm2	15964
4395	chr2	118594327	118594403	Mcam	248
4398	chr2	119161444	119161723	Cxcr5	-4410
4402	chr2	120128592	120128938	Ift46	26014
4407	chr2	120970666	120970771	Apoa1	49
4416	chr2	122118268	122118456	Plet1	118
4419	chr2	124705095	124705264	Rdx	-3560
4423	chr1	190542031	190542193	Madcam1	-52295
4430	chr2	125724610	125724722	Cspg4	-462
4451	chr2	129595947	129596016	Bbs4	3112
4458	chr1	191075314	191075475	Elane	3366
4459	chr2	130519316	130519490	Map2k5	-4679
4462	chr2	130597791	130597996	Smad3	-15237
4519	chr2	143533932	143534071	Ctsh	-12224
4550	chr2	151947694	151947846	Amotl2	2947
4566	chr2	153432685	153432794	Iqcf1	18836
4570	chr2	153495944	153496017	Hyal2	3154
4572	chr2	154198263	154198343	Gnai2	7445

4573	chr2	154517034	154517104	Sema3f	-30051
4586	chr2	156312495	156312617	Ptpn23	252
4594	chr2	157329581	157329746	Ccr4	17925
4601	chr2	158141957	158142388	Tgfbr2	4479
4607	chr2	158539844	158539913	Itga9	27048
4631	chr2	162885473	162885542	Ccr1	14050
4635	chr1	194927937	194928137	Map2k2	-10882
4646	chr1	194949472	194949567	Dapk3	10653
4650	chr2	164942781	164943135	Efnb1	2001
4653	chr2	165056043	165056237	Atp7a	416
4660	chr2	165499160	165499248	Sh3kbp1	-4627
4680	chr10	3718302	3718603	Tbxa2r	-22062
4746	chr10	4107613	4108044	Appl2	3300
4990	chr10	8501079	8501295	Cfap54	17746
5057	chr10	11149365	11149540	Plxnc1	-62
5212	chr10	14705388	14705612	Pawr	-203
5468	chr10	20541321	20541392	Srgap1	-21553
5534	chr10	21632821	21632894	Lrp1	39676
5913	chr10	34937180	34937249	Tns3	38924
5979	chr10	37837793	37837941	Egfr	-93887
6390	chr10	52273109	52273436	Rhbdf1	39490
6568	chr10	59221656	59221964	Wwc1	-297
6813	chr10	62665315	62665573	Flt4	-4488
7212	chr10	75651969	75652067	Anxa6	-7961
7246	chr10	76629052	76629128	Fat2	-5422
7257	chr10	77137071	77137140	Sparc	-29
7468	chr10	79946700	79946771	Map2k3	-13452
7656	chr10	80856357	80856535	Pmp22	1082
7891	chr10	83805555	83805723	Ndel1	6010
8035	chr10	85473471	85473684	Trp53	86657
8168	chr10	89611062	89611282	Pfn1	10191
8312	chr10	93294896	93295087	Spns2	16239
8401	chr10	93419751	93419904	Pafah1b1	-13356
8468	chr10	95414777	95414846	Serpinf1	2403
8512	chr10	95969847	95969916	Myo1c	29197
8601	chr10	99228982	99229145	Tmigd1	-76
8634	chr10	99461762	99461912	Coro6	-1532
8778	chr1	53851382	53851605	Bmpr2	1958
8801	chr10	107548831	107548938	Lgals9	-54094
8867	chr10	111100087	111100330	Cdk5r1	2784
8935	chr10	111670686	111670777	Ccl8	-20874

8979	chr10	111972557	111972736	Rffl	-107
9107	chr1	53993937	53994207	Abi2	-106996
9130	chr10	120508538	120508607	Itga3	-32069
9141	chr10	121461716	121461800	Skap1	978
9149	chr10	122008833	122008965	Socs7	38482
9152	chr10	122526101	122526172	Srcin1	20397
9166	chr10	128097725	128097965	Ccr7	6148
9178	chr11	3573183	3573547	Stat5a	-32571
9184	chr1	55193425	55193611	Nrp2	-32372
9190	chr11	5496351	5496423	Rnd2	-905

Supplementary Table III. Taqman probe sets.

No	Symbol	Cat#	Lot#	Concentration	Species
1	Mmp9	Mm00442991_m1	P120224-002 A04	20X	mouse
2	Plxnd1	Mm01184367_m1	P190814-008 F02	20X	mouse
3	RhoJ	Mm00502666_m1	1725932	20X	mouse
4	Etv2	Mm01176581_g1	P181129-023 C09	20X	mouse
5	Ets1	Mm01175819_m1	1534642	20X	mouse
6	Gapdh	4352339E	1708051	20X	mouse
7	Fli1	Mm00484410_m1	1169024	20X	mouse
8	Erg	Mm01214244_m1	1690648	20X	mouse

CHAPTER 5: FOXK1 REGULATES WNT SIGNALING TO PROMOTE CARDIOGENESIS

Introduction

Congenital heart disease is the most common genetic birth defect affecting approximately 1% of live births and having considerable morbidity and mortality.(22, 23) Therefore, it is essential to decipher the regulatory pathways that govern the specification and differentiation of mesodermal progenitors and to use this information to develop targeted therapies for congenital cardiovascular diseases. The cardiovascular system consists of multiple cell lineages including: the hematopoietic, vascular and muscle lineages.(87-89) The genesis of each of these lineages from a common germ layer requires precise and coordinated regulatory signals to take place during embryonic development.(87) The transcriptional regulators and signaling pathways that govern the development of these lineages are incompletely defined and warrant further investigation.

The Wnt signaling pathway has been shown to play essential roles in the development of the brain, limb, blood, endothelium and heart.(219) More importantly, the Wnt/β-catenin signaling pathway has been shown to have positive and negative modulatory effects on the specification and the differentiation of mesodermal lineages.(220-222) This biphasic role for Wnt/β-catenin signaling has been demonstrated in zebrafish, mouse embryos and differentiating mouse embryonic stem cells. In this fashion, Wnt signaling has been shown to be procardiogenic in early precardiac mesoderm and inhibitory to cardiogenesis during the later stages of cardiac differentiation.(223) Moreover, β-catenin has been reported to be an upstream activator of *Isl1* gene expression in the heart.(224) While the Wnt/β-catenin signaling pathway has a binary role

during cardiogenesis, the overexpression of the Wnt signaling cascade is associated with an expansion of the hematopoietic and endothelial lineages.(222) Therefore, the activity and functional role of the Wnt signaling pathway is context dependent and, in part, modulated through protein-protein interactions. For example, BMPs, Tgfβ and others have been shown to interact with Wnt signaling and have a combinatorial role during development and regeneration.(220)

Forkhead box (FOX) proteins are a family of evolutionarily conserved transcription factors that share a DNA binding domain known as the *forkhead/winged helix* domain.(56) Members of this family have been shown to have essential roles during embryogenesis in lineage fate decisions, cell cycle kinetics, aging, metabolism, stem cell regulation and chromatin remodeling (pioneer factors).(41, 45, 56) Furthermore, many of these FOX factors have been shown to have important roles in cancer proliferation and tumorigenesis.(57) The Foxk family consists of two members, FOXK1 and FOXK2, which have a shared structure.(225) FOXK1 was discovered to be a transcription factor restricted to the striated muscle (cardiac and skeletal muscle) group during development in somites and the heart.(58) Previous work has characterized the role of FOXK1 as an important regulator of myogenic stem cells (satellite cells) proliferation following injury.(59, 63, 64, 67-69, 226) While the role of FOXK1 has been extensively studied in skeletal muscle, the role for FOXK1 in cardiac muscle and cardiogenesis is unknown and warrants further investigation.

In the present study, we characterized the role of FOXK1 during mesodermal development. Using *Foxk1* KO ES/EBs and *Foxk1*-EYFP transgenic mice, we identify FOXK1 as an important transcriptional and epigenetic regulator of cardiogenesis. Mechanistically, our findings demonstrate that FOXK1 promotes cardiogenesis by repressing the Wnt/β-catenin signaling pathway. This study identifies a novel role for FOXK1 in the regulation and specification of mesodermal lineages during development and the interaction and regulation of distinct signaling pathways during cardiogenesis. These results identify FOXK1 as an essential transcriptional and epigenetic regulator of cardiovascular development.

Results

FOXK1 regulates mesodermal progenitor cell development. While previous studies have demonstrated that FOXK1 expression during development is restricted to striated muscle (skeletal and cardiac)(58),the role for FOXK1 in the development of the cardiac lineage has yet to be characterized. We hypothesized that FOXK1 was an important regulator for mesodermal and cardiac progenitor cells. To test this hypothesis, we engineered *Foxk1* KO mouse embryonic stem cells (ESCs) to examine the role of FOXK1 during mesoderm and the cardiac lineage development in vitro.

Lineage-tracing studies using the embryonic stem cell/embryoid body (ES/EB) system support the existence of cardiovascular multipotent progenitors that can give rise to endothelial, myocardial, smooth muscle and hematopoietic lineages.(92, 135, 227-229) Using a previously published mesodermal and

cardiac differentiation protocol for shaking ES/EBs(230), we first differentiated control and *Foxk1* KO mESCs to examine the expression of FOXK1 (Figure 1A and 1B and Figure S1). Both qPCR and western blot assays demonstrated that FOXK1 expression peaks at approximately day 5 (D5) of differentiation (Figure 1B and Figure S1), a time period where mesoderm is formed in these EBs and at a time when progenitors start differentiating towards the cardiac, hematoendothelial and skeletal muscle lineage.(82, 231) To determine whether FOXK1 plays an important role in the development of mesoderm, we analyzed the lineage commitment of control and *Foxk1* KO EBs at Day 3 (D3) by staining for FLK1 and PDGFRα using flow cytometry (Figure 1C and Figure S2). We observed that in the absence of FOXK1, skeletal myogenic (FLK1-PDGFRα+) and cardiac (FLK1+PDGFRα+), but not hematoendothelial (FLK1+PDGFRα-) progenitors were absent on D3 compared to the control EBs (Figure 1C). To further characterize this phenotype, we examined later time periods of differentiation (Figure 1D and 1E). Similar to D3, flow cytometry analysis of D5 and D7 EBs demonstrated a significant reduction in cardiac and skeletal muscle progenitors in the *Foxk1* KO EBs compared to their respective controls (Figure 1D and 1E). However, unlike D3, D5 and D7 *Foxk1* KO EBs demonstrated a significant increase in blood and vascular progenitor cells (FLK1+PDGFRα-), suggesting a cell fate change in the cardiac/skeletal muscle progenitor cells that can no longer form due to the absence of FOXK1. Based on these results, we hypothesized that FOXK1 was an important mesodermal regulator of cardiac and skeletal muscle progenitor cell development.

FOXK1 regulates cardiac developmental transcriptional networks. Having established a role for FOXK1 in mesodermal progenitor cell development, we isolated RNA and performed bulk RNA sequencing (RNAseq) from D3 and D5 differentiating EBs to define the transcriptional networks of these cells before and after mesoderm formation in both control and *Foxk1* KO EBs. We identified 644 and 951 upregulated genes in the control D3 and D5 groups compared to the *Foxk1* KO groups, respectively. GO pathway analysis of D3 differentiating EBs identified mesodermal and cardiovascular developmental networks in the top 5 signaling pathways enriched in the control but not the *Foxk1* KO EBs (Figure 2A and Figure S3A). Among the top upregulated genes in the control EBs we identified important regulators of early cardiovascular development such as Mesp1, Mesp2, Isl1 and Hand2 (Figure 2B and Figure S3B). Furthermore, GO pathway analysis of D5 EBs further highlighted cardiac developmental networks in the top 5 enriched pathways in the control EBs compared to the *Foxk1* KO EBs, with the top genes enriched in the control group also highlighting important regulators of cardiovascular development such as Hand1, Hand2 and Tbx5 (Figure 2C and 2D and Figure S3C and S3D). When we examined the *Foxk1* KO group, we identified 703 and 1,164 upregulated genes in the D3 and D5 *Foxk1* KO groups, respectively (Figure S3). GO pathway analysis of the genes upregulated in the KO group demonstrated that vasculature development was significantly enriched in both the D3 and D5 *Foxk1* KO EBs (Figure S3E and S3F). These results further supported the notion that in the absence of FOXK1,

cardiac and myogenic progenitor cells were redirected to a vascular fate as we observed a significant increase FLK1+ cells in the D5 and D7 *Foxk1* KO group (Figure 1D and 1E). While previous studies have characterized the role of FOXK1 as a regulator of myogenic stem cell development,(67, 70) these RNAseq analysis identified a novel function for FOXK1 as an essential regulator of cardiovascular development.

FOXK1 modulates dynamic chromatin accessibility of cardiac genes. Forkhead transcription factors have been shown to play important epigenetic regulatory roles during development, reprogramming (pioneer factors) and tumorigenesis.(41, 45) Additionally, previous work has demonstrated that FOXK1 can interact with histone deacetylase 3 (HDAC3) to regulate skeletal muscle regeneration and antiviral immune responses.(62, 232) Therefore, we hypothesized that FOXK1 regulated cardiovascular development by modulating chromatin accessibility in differentiating EBs. Using the assay for transposase-accesible chromatin using sequencing (ATACseq),(174) we profiled any chromatin changes that occurred during mesodermal and cardiac differentiation in the absence of FOXK1 during EB differentiation (D3 and D5 EBs). We extracted FOXK1 motif binding positions using motifmatchr and the ATAC-seq peaks, and based on the nucleosome signal obtained from NucleoATAC within the center 200 bp region, we divided the ATAC-seq analysis into 4 regions (all containing FOXK1 motifs) - nucleosome free regions (NFR) in both control and *Foxk1* KO, (2) NFR in control and nucleosome occupied regions (148) in *Foxk1*

KO, (169) NOR in control and NFR in *Foxk1* KO, and NFR in both (Figure 3A and 3B). More than 43% of the chromatin surrounding regions containing a FOXK1 motif were significantly affected (open or closed) by the absence of FOXK1 at D3 (2,693/6,196) and D5 (3,004/6,196) (Figure 3A and 3B). Our analysis demonstrated reduced chromatin accessibility in 1,527 regions at D3 and 2,062 regions at D5 in the *Foxk1* KO group compared to control (Figure 3A and 3B). Additionally, the absence of FOXK1 led to an increase in chromatin accessibility in 1,166 regions at D3 and 942 regions at D5 in the Foxk1 KO group compared to control (Figure 3A and 3B). These results suggested that FOXK1 has an important role as an epigenetic regulator in promoting both chromatin relaxation and compaction during differentiation, just like other forkhead transcription factors. However, its role as a promoter of chromatin relaxation is more significant based on the number of regions affected (Figure 3A and 3B).

Integration of the RNAseq and ATACseq datasets from D3 and D5 EBs allowed for the identification of transcription factors whose expression and chromatin accessibility was significantly affected (Figure 3C through 3F). Analysis of this integration further identified important cardiac regulatory transcription factors whose expression and chromatin accessibility was significantly reduced due to the absence of FOXK1 (Figure 3C through 3F and Figure S4). These results further supported our hypothesis that FOXK1 has an essential role during cardiovascular development and identified a novel role for FOXK1 as an epigenetic regulator of cardiovascular development by regulating chromatin dynamics.

FOXK1 regulates cardiogenesis in differentiating EBs. After establishing that FOXK1 regulated cardiac transcriptional and epigenetic networks in mesodermal and cardiac progenitors, we hypothesized that the absence of FOXK1 would affect the differentiation and maturation of these early progenitors (Figure 4A). Therefore, we characterized D10 control and *Foxk1* KO EBs for the expression of cardiac Troponin T (cTnT) and their beating potential. *Foxk1* KO EBs demonstrated a significant reduction in the number of cTnT+ and beating EBs compared to control EBs (Figure 4B through 4D). Bulk RNAseq of *Foxk1* KO EBs demonstrated a failure to activate gene expression, which was important for cardiac muscle development compared to control EBs based on GO pathway analysis (Figure 4E). Among the genes affected in the *Foxk1* KO group, important cardiac transcription factors (Nkx2-5, Tbx5) and mature structural/functional cardiac genes (Myh6, Myh7, Ttn, Ryr2) were significantly downregulated compared to control EBs (Figure 4F), further demonstrating the important role FOXK1 plays during the differentiation and maturation of cardiac progenitor cells. Similar to D3 and D5 EBs, ATACseq analysis of D10 EBs also demonstrated a significant reduction in the chromatin accessibility of transcription factors important for the regulation of later stages of cardiac development such as Mef2c, Klf2, Klf3 and Klf13 (Figure 4G). GO pathway analysis of the genes upregulated in the D10 *Foxk1* KO group demonstrated that, similar to D3 and D5 EBs, vasculature development was a significantly enriched pathway, suggesting that the progenitor cells in the KO group were redirected to the vascular lineage and formed mature vascular cells (Figure S5). These results establish that

FOXK1 regulates not only the early development of cardiac progenitor cells, but also their differentiation and maturation at later stages.

FOXK1 regulates Wnt signaling in cardiac progenitor cells. Wnt is an essential signaling pathway for cardiogenesis(223, 224) and previous studies have demonstrated that *forkhead/winged helix* factors can modulate Wnt signaling.(233, 234) For example, FOXG1 has been shown to repress Wnt5a during brain development while FOXK1 has been reported to modulate Wnt signaling by translocating disheveled proteins to the nucleus.(233, 234) Therefore, we hypothesized that FOXK1 modulated Wnt signaling in cardiac progenitor cells. To test this hypothesis, we first evaluated the temporal expression of Wnt signaling in our EB differentiation system. qPCR analysis of *Wnt3a* demonstrated that our EB differentiation system recapitulated the biphasic role of Wnt signaling observed during cardiogenesis where it is expressed early (D3-D4) and repressed later (D5-D6) during development to allow for cardiogenesis to proceed (Figure S5). Note that the peak expression of *Wnt3a* (D4) precedes that of *Foxk1* (D5, Figure 1B). These results supported the hypothesis that FOXK1 modulated Wnt signaling at D5 (peak expression of Wnt signaling). To examine this hypothesis, we queried our D5 bulk RNA sequencing dataset and determined that in the absence of FOXK1, Wnt signaling pathways were enriched in the *Foxk1* KO over the control group (Figure 5A), suggesting that FOXK1 repressed Wnt signaling at distinct stages of cardiogenesis.

If FOXK1 acts as a repressor of Wnt signaling during cardiogenesis, we hypothesized that inhibition of the Wnt signaling pathway in the *Foxk1* KO EBs would rescue the perturbation of cardiogenesis (Figure 4B through 4D). To this end, we employed a defined Wnt signaling inhibitor (IWR1), added it to our differentiation conditions at D4 and characterized the EBs at D10 (Figure 5B). The addition of IWR1 to our differentiation conditions rescued both the cTnT+ percentage and beating potential of the *Foxk1* KO EBs as no significant differences were observed between control and *Foxk1* KO groups (Figure 5C and 5D). RNAseq analysis of D10 EBs demonstrated a significant overlap in the principal component analysis (PC2) between control (Ct) and *Foxk1* KO + IWR1 (KO+IWR1) but not the Foxk1 KO (KO) alone (no IWR1) group (Figure 5E). We further explored this overlap by identifying the commonly upregulated genes between control and *Foxk1* KO + IWR1 groups that were not upregulated in the *Foxk1* KO group to identify important regulators of cardiogenesis that were responsible for this rescue (Figure 5F). We identified 320 commonly upregulated genes unique to these two groups (Figure 5F). Muscle tissue development as well as important regulators of cardiovascular development such as *Nkx2-5*, *Shh* and *Ttn* were enriched between these two groups (Figure 5G and 5H).

FOXK1 is a transcriptional repressor of Wnt signaling. To further characterize the mechanism whereby FOXK1 represses Wnt signaling, we evaluated the role of FOXK1 as a transcriptional regulator of Wnt signaling. To this end, we first identified target genes that were dysregulated at D5 in the absence of FOXK1. Analysis of our D5 bulk RNAseq dataset identified Wnt

related genes that were dysregulated during mesodermal and cardiac differentiation in the absence of FOXK1 (Figure 6A). *Wnt6* was the top hit of the dysregulated genes, which we validated using qPCR to demonstrate that it was overexpressed not only at D5 but also at later stages of differentiation compared to control EBs (Figure 6A and 6B). WNT6 is an important regulator of cardiovascular development and its overexpression results in a hypomorphic heart in Xenopus, suggesting that downregulation of *Wnt6* is essential during cardiogenesis.(235) To determine whether FOXK1 is a transcriptional regulator of *Wnt6*, we examined the upstream region of the *Wnt6* gene to identify any potential FOXK1 DNA binding motifs. FOXK1 DNA binding motifs were identified and when combined with the ATACseq datasets, we confirmed that the chromatin surrounding this DNA motif was significantly more relaxed in the absence of FOXK1 (Figure 6C). Additionally, this region had a β-catenin DNA binding site(236), suggesting that in the absence of FOXK1, the chromatin upstream of the *Wnt6* gene cannot be closed to other factors such as β-catenin that promote *Wnt6* expression. To validate the capacity of FOXK1 to bind to the upstream region of the *Wnt6* gene, we performed chromatin immunoprecipitation (ChIP) with a FOXK1 antibody followed by qPCR of the selected region in D5 EBs. ChIP qPCR analysis using FOXK1 demonstrated a 3.6-fold enrichment for FOXK1 in the *Wnt6* binding site over the Gapdh site (Figure 6D). Because of the identified β-catenin DNA binding site upstream of *Wnt6* we also hypothesized that both FOXK1 and β-catenin could compete for binding to repress or activate Wnt signaling pathways, respectively. To this end, we used the TOP/FOP-flash

reporter system to evaluate the role of FOXK1 as a transcriptional repressor of Wnt signaling in the presence of β-catenin (Figure 6E). Analysis of this luciferase reporter demonstrated that FOXK1 significantly repressed β-catenin activity in a dose dependent fashion (Figure 6E), further confirming the role of FOXK1 as a transcriptional regulator of Wnt signaling. Collectively, these data demonstrate that mechanistically, FOXK1 acts as a transcriptional and epigenetic repressor of Wnt signaling to regulate cardiovascular development.

FOXK1 labels developing cardiac cells in the mouse embryo. To further validate our in vitro results with FOXK1, we used the previously characterized upstream 4.6kb *Foxk1* promoter(68) and engineered a 4.6kb *Foxk1*-EYFP transgenic mouse model to examine the expression of FOXK1 during development. Analysis of E9.5 mouse embryos demonstrated that the EYFP reporter recapitulated endogenous FOXK1 expression(58) with robust expression in the heart (Figure 7A). Immunohistochemical analysis of 9.5 to E11.5 mouse developing hearts demonstrated continuous labeling of developing cardiac cells (Nkx2.5+ cells) with the EYFP reporter (Figure 7B through 7D).

Discussion

The heart is the first organ to form and function during development and its proper formation is crucial for mammals to develop without any congenital anomalies or CHD(21, 88, 237). A number of transcription factors and signaling cascades have been identified and have an important role during cardiovascular development; however, the mechanisms that govern this process are still unclear. *Forkhead/winged helix* transcription factors have been shown to regulate key cellular processes such as cell proliferation, differentiation, chromatin remodeling, metabolism and others(237). We and others have shown that FOXK1 expression is restricted to striated muscle (skeletal and cardiac muscles) during development and that it is essential during skeletal muscle development and regeneration. In the present study, we have used gene disruption strategies, computational genomics, transgenic mice, cellular and molecular techniques to make several fundamental discoveries to decipher an important role for FOXK1 as a regulator of cardiovascular development.

First, we identified an important role for FOXK1 in the regulation of mesodermal progenitor cell, cardiac progenitor and skeletal muscle progenitor cell development. Transcriptional analysis of these developing progenitors further demonstrated a significant decrease in the expression of cardiovascular developmental pathways and key regulatory transcription factors such as Isl1, Hand1, Hand2 and Tbx5 in the absence of FOXK1. These data identified FOXK1 as a key regulator of cardiovascular development, particularly of first and second

heart field structures by regulating the expression of these transcription factors.(21) While deletion of other *forkhead* factors can lead to cardiac developmental defects, they appear to be due to secondary effects as their expression is not restricted to developing striated muscle, unlike FOXK1.(237)

Interestingly, in the absence of FOXK1, progenitor cells that could not form cardiac and skeletal muscle progenitors, instead formed vascular progenitor cells using flow cytometry. This observation was further validated with the RNAseq analysis that showed an enrichment of vascular developmental pathways in the cells lacking FOXK1. Similar fate changes in the absence of a transcription factor during mesodermal development has been observed as in the case of the master regulator ETV2, where *Etv2* KO mesodermal progenitors preferentially generated cardiac instead of vascular and blood progenitor cells.(50) These data suggested that one of the mechanisms by which FOXK1 promoted cardiovascular development was by repressing other lineages such as the vascular lineage.

Our next discovery identified FOXK1 as an epigenetic regulator of cardiovascular development. More than 40% of the chromatin surrounding regions with a FOXK1 DNA binding motif demonstrated a significant impact in the chromatin dynamics in the absence of FOXK1. More importantly, the absence of FOXK1 led to a decrease in the chromatin accessibility of the same key regulatory cardiovascular development genes identified in our RNAseq analysis (Isl1, Hand1/2 and Tbx5), along with other important regulators (Gata4, Tbx20).

Forkhead factors are known epigenetic regulators that can function by interacting with a chromatin remodeler or through their unique protein structure (resembling that of linker histones) that allows them to interact with heterochromatin and relax the chromatin landscape by displacing linker histones.(41, 45, 232) FOXK1 has been shown to play a role as an epigenetic regulator in skeletal muscle through interactions with histone deacetylases, however, the exact mechanism whereby FOXK1 regulates the chromatin landscape near cardiac genes remains to be elucidated. The SWI/SNF complex has been recently shown to play an important role in cardiac development by assisting in the chromatin remodeling process of different important cardiac regulators.(82, 231, 238) Whether this complex assists FOXK1 or whether FOXK1 remodels chromatin by itself like other *forkhead/winged helix* factors such as FOXA1(41) remains to be elucidated.

Another major finding from this work is that mechanistically, we have identified that FOXK1 regulates cardiovascular development by transcriptionally and epigenetically repressing Wnt signaling in cardiac progenitor cells and we identified *Wnt6* as a major downstream target of FOXK1. Absence of FOXK1 during mesodermal and cardiac differentiation led to a significant increase in the expression of *Wnt6* that persisted until later stages of differentiation compared to the control group. WNT6 is a known regulator of cardiovascular development whose expression is tightly regulated as continuous overexpression of *Wnt6* has been shown to be detrimental for cardiac development.(235) While these current studies identified FOXK1 as a repressor of Wnt signaling, others have shown that FOXK1 positively regulates Wnt signaling (in a tumorigenic setting), suggesting

that the function of FOXK1 can be context dependent.(234) Therefore, it will be important to identify upstream regulators of FOXK1 during cardiovascular development that might regulate this context dependent expression and function.

Our in vitro studies were further supported by the *Foxk1*-EYFP transgenic mouse model that labeled differentiating cardiac progenitor cells during embryogenesis. The bioinformatics data further reinforced the observation that EYFP and Nkx2.5 were coexpressed in the developing heart, suggesting that FOXK1 regulates the differentiation and maturation of both first and second heart field structures.

In summary, we have identified a novel role for FOXK1 in the regulation of cardiovascular development and showed that FOXK1 is a direct transcriptional and epigenetic repressor of Wnt signaling, particularly *Wnt6*. Future studies will be needed to identify upstream regulators of FOXK1 early during mesodermal specification.

Methods

Embryo Harvesting and Microscopy. All animal studies were approved by the Institutional Animal Care and Use Committee at the University of Minnesota, and all the methods were performed in accordance with relevant guidelines and regulations. Time-mated pregnant mice from the 4.6 kb *Foxk1*-promoter driving EYFP transgenic lines were used for embryo harvest, imaging, and immunohistochemistry experiments at E9.5. For imaging experiments, these embryos were fixed for 1 hour at 4 °C in 4% paraformaldehyde, washed twice in PBS, and imaged on a Zeiss Axio Imager inverted microscope.

Mouse embryonic stem cell lines. *iHA-Foxk1* ESCs (control) harboring a tetracycline responsive element upstream of HA tagged FOXK1 was generated as previously described.(239) *Foxk1* KO ESCs were isolated using standard techniques by harvesting blastocysts from Foxk1 heterozygous female mice bred with Foxk1 heterozygous males.(64, 240) *iHA-Foxk1* and *Foxk1* KO ESCs were cultured in media containing 15% fetal bovine serum (FBS), 2 mM Glutamax, 1X penicillin/streptomycin, 0.1 mM β-mercaptoethanol, and 1,000 U/mL LIF, at 37°C in 5% CO2 together with irradiated mouse embryonic fibroblasts as a feeder layer. We differentiated *control* and *Foxk1* KO ESCs into embryoid bodies (EBs) using mesodermal differentiation conditions as previously described.(113) Briefly, ESCs were separated into single cells using 0.25% trypsin and plated for 50 minutes to remove fibroblast cells from the feeder layer (de-MEF step). Following de-MEF, ESCs were differentiated in mesodermal media using the

shaking method in media containing 15% FBS (Foundation GeminiBio), 1X penicillin / streptomycin, 1X GlutaMAX, 50 µg/ml Fe-saturated transferrin, 450 mM monothioglycerol, 50 µg/ml ascorbic acid in IMDM media (Thermo). *iHA-Foxk1* EBs were treated with doxycycline (0.5 µg/ml) in the differentiation media to overexpress FOXK1. D10 EBs harvested for immunohistochemistry were fixed for 10 mins at 4 °C in 4% paraformaldehyde, washed twice in PBS and embedded in Optimal Cutting Temperature (O.C.T) solution before cyrosectioning.

Flow cytometry analysis. Harvesting and staining of *control* and *Foxk1* KO EBs was performed as previously described(159) and analyzed using a FACSAria (BD) machine. The antibodies used for FACS include: Flk1-APC (1:200, Cat# 560070, Lot# 8298981), platelet-derived growth factor alpha receptor-a (Pdgfra)-PE (1:1000, Cat# 4315814, Lot# 2049418), cTnT (1:100, Cat# MS-295-P1, Lot# 295P16048) and anti-mouse AlexaFluor 488 (1:400, Cat# 715-485-151, Lot# 94650). Propidium iodide (1:1000, Cat# 1423090, Lot# 1325708) were used to gate for live cells during flow cytometry analysis.

RNA isolation and quantitative PCR (qPCR) analysis. Total RNA was isolated from control and *Foxk1* KO EBs using the RNeasy kit (Qiagen, Cat# 74104) following the manufacturer's protocol. Briefly, EBs were lysed in RLT-lysis buffer, followed by a column-based purification process and on-column DNA digestion. Complementary DNA (cDNA) was synthesized using the SuperScript IV VILO kit (Thermo Fisher Scientific, Cat# 11756050) following the manufacturer's protocol.

Quantitative polymerase chain reaction (qPCR) was performed using ABI Taqman probe sets. Taqman probes used in this study include VIC-labeled GAPDH:4352339E, FAM-labeled Foxk1:Mm01195488_m1, Wnt3a:Mm0437337_m1 and Wnt6:Mm00437353_m1.

Western blot. Western blots were performed as previously described.(230) Briefly, embryoid body lysates from were obtained control and *Foxk1* KO EBs at various time points (Day 3, 5 and 7). These were lysed using ice-cold lysis buffer for 30 minutes and centrifuged at 9,300 x g for 10 minutes at 4 °C. Equal amounts of protein were loaded on 10% SDS-polyacrylamide gels. PVDF (polyvinylidene fluoride) membranes were blocked with 5% milk protein and incubated with a rabbit-FOXK1 antibody (1:1000, Cat# sc134550, Lot# D1911) and rabbit-GAPDH antibody (1:1000, Cat# D16H11, Lot# 7) overnight at 4 °C. The membrane was subsequently incubated with a goat-anti-rabbit (1:2000, Cat#SC-2004; Lot# G247) HRP (horseradish peroxidase)-conjugated secondary antibody and visualized using the Pico luminescence kit (Invitrogen) following the manufacturer's instructions. The protein bands were visualized and imaged using the Image Lab software.

Analysis of bulk RNA-seq. The sequencing reads of the bulk RNAseq data were mapped to the mouse genome (mm10) using Kallisto (0.46.2)(241) with default parameters. The read counts data were normalized by DESeq2, followed by differential expression analysis.(173)

Generation of ATAC-seq libraries, sequencing and analysis. Control and *Foxk1* KO EBs were collected at different time periods during differentiation and disaggregated in 0.25% trypsin at 37°C for 3 min incubation with gentle agitation followed by inactivation with culture medium containing 10% FBS. Cells were collected by centrifugation at 500g for 5 minutes at 4 degrees celcius and washed once with ice-cold PBS. ATAC reaction was performed with 50,000 cells as previously described(174) using the Tn5 transposase (Illumina) and libraries were created at the University of Minnesota Genome Center. Libraries were then sequenced on a NextSeq Illumina platform (2x50 bp) aiming for 25 million reads per sample. The sequencing reads where mapped to the mouse genome (mm10) using Bowtie2 (v2.2.4)(177) and ATAC-seq peaks were called using MACS2 (v2.1.1).(178) The ATAC-seq lied within the blacklisted genomic regions for functional genomics analysis were excluded.(179) ChromVAR (v1.10) was used for transcription factor-based chromatin accessibility analysis. We used NucleoATAC (v0.3.4)(154) to estimate the nucleosome signals at the Foxk1 motif binding positions.

Cardiobeating assays. ES/EBs were differentiatiated as described above using mesodermal conditions. At D8, EBs were transferred to a 96-well gelatin coated flat bottom plate containing EB media at a concentration of 1 EB/well. Individual beating potential (presence or absence of beating) was assessed and quantified at D10.

Transcriptional assays. Transcriptional assays were performed as previously described.(70) Briefly, HEK 293 T cells were cultured in 35 mm dishes containing DMEM supplemented with 10% fetal bovine serum and penicillin/streptomycin. 100-200k cells were seeded and transfected using the lipofectamine reagent and assayed for activity. Reporter assay was performed using the Promega Luciferase Assay System following the manufacturer's instructions. Luciferase activity was analyzed with dual-luciferase reporter assay system (Promega), and normalized with the Renilla luciferase. All transfection experiments were performed in triplicate and replicated three times.

Immunohistochemistry. Cryosectioning and immunohistochemical analyses were performed as previously reported(51) using the following antibodies: cTnT (1:100, Cat# MS-295-P1, Lot# 295P16048), GFP (1:500, Cat# ab13970, Lot# GR236651-19), anti-chicken AlexaFluor 488 (1:400, Cat# 703545155, Lot# 136424), anti-mouse AlexaFluor 594 (1:400, Cat# , Lot#) and anti-mouse AlexaFluor 488 (1:400, Cat# 715-485-151, Lot# 94650).

Data analysis. No data were excluded from these studies and all attempts at replication for standard assays (flow cytometry, qPCR, cardiogenic beating assays and immunohistochemistry) were successful. Data collection and analysis were not performed blind to the conditions of the experiments.

Data availability. The bulk RNA-seq and ATAC-seq datasets of differentiating EBs were deposited at NCBI Gene Expression Omnibus (GEO) database with the accession code ( ). All data will be available upon request. All unique materials used in these studies are readily available from the authors or from commercial sources.

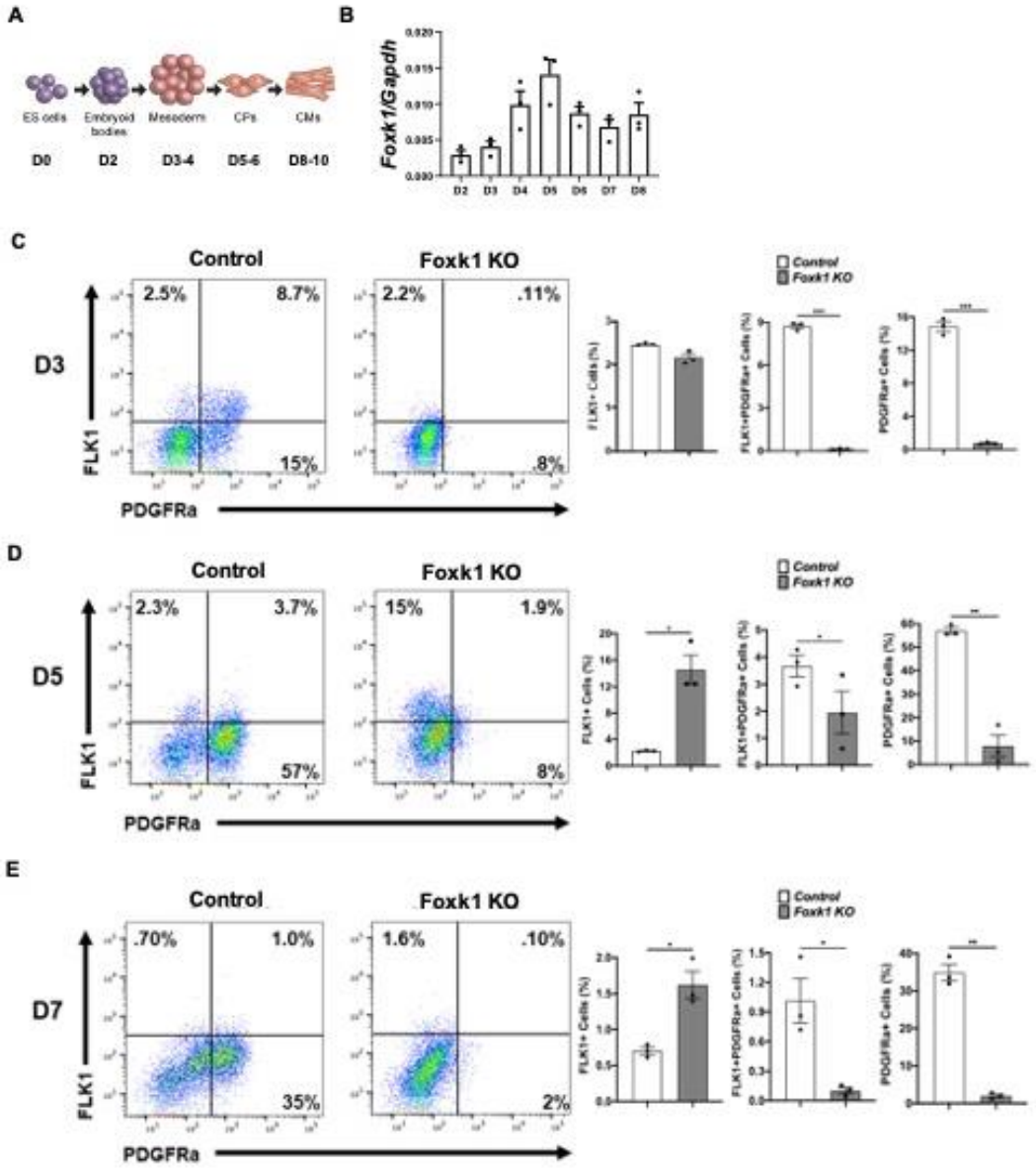


Figure 1. FOXK1 regulates mesodermal progenitor cell development. A, Schematic of embryoid body in vitro differentiation and cardiac milestones. **B**, *Foxk1* transcript expression during EB differentiation from day 2 (D2) to day 8 (D8). Note that the expression of *Foxk1* peaks at day 5 (D5). **C-E**, Representative flow cytometry profile of control and *Foxk1* KO EBs at D3, D5 and D7 of mesodermal differentiation protocol with quantitation of the results. Note that throughout differentiation, there is a significant defect in the ability of *Foxk1* null EBs to form mesodermal progenitors, particularly cardiac and skeletal myogenic (n = 3, *p < 0.05). Statistical test: Student's t-test. Data are presented as mean ± SEM.

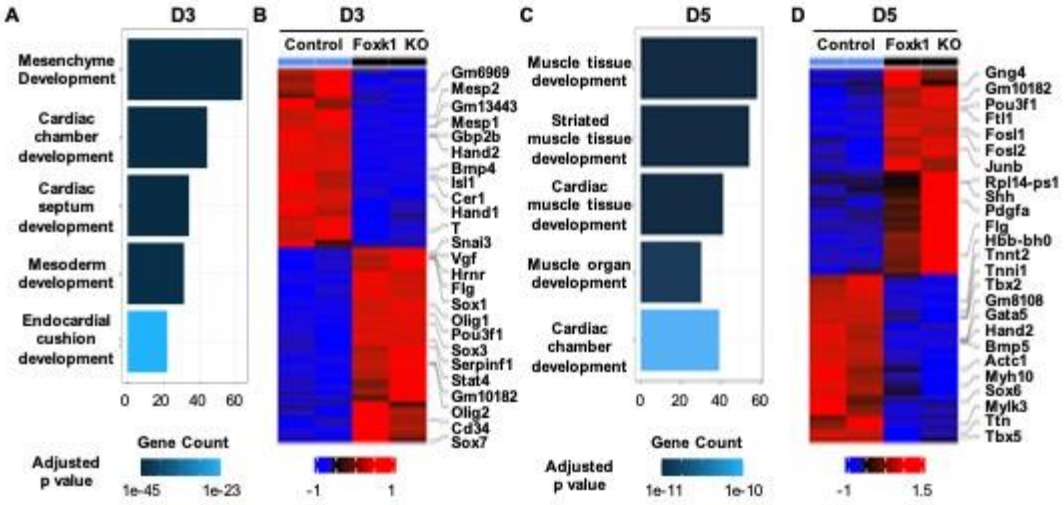


Figure 2. FOXK1 regulates cardiac developmental transcriptional networks. **A**,**C** GO pathway analysis highlights pathways and development related terms in D3 and D5 control EBs, the x-axis represents the counts of genes in each GO term. The color scale shows the increased significance of biological processes using the over-representation test with an adjusted p value < 0.05. **B,D** The heatmap represents upregulated and downregulated in the control EBs vs. the *Foxk1* KO EBs at D3 and D5, respectively. The heatmap color scheme key is provided, with red representing upregulated and blue representing downregulated genes.

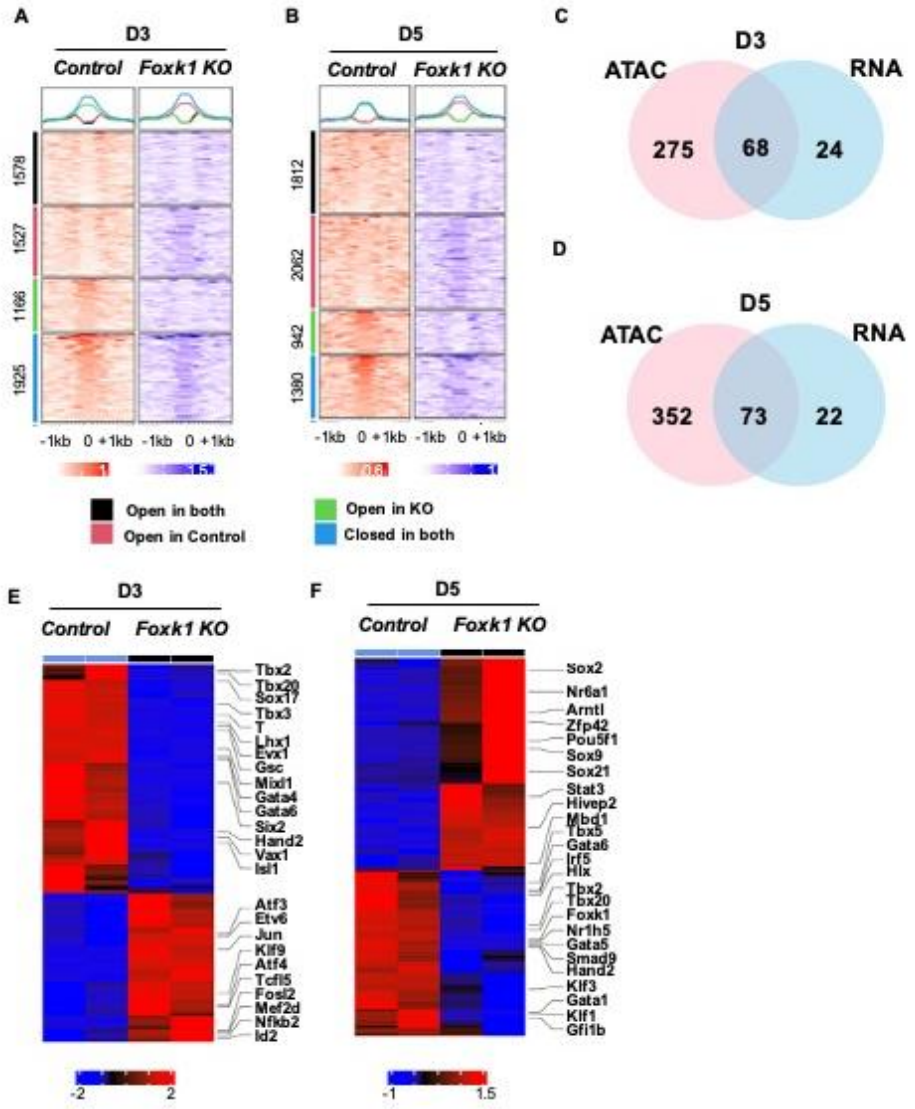


Figure 3. FOXK1 is an epigenetic regulator of cardiac development. A,B Enriched heatmap of the nucleoATAC data showing more nucleosome free regions (NFR) in control samples as compared to *Foxk1* KO samples at D3 and D5 at the Foxk1 binding sites. We divided the FOXK1 binding sites into 4 regions, showing NFR in both, NFR in control and nucleosome occupied region (148) in *Foxk1* KO, NOR in control and NFR in *Foxk1* KO, and NFR in both. **C,D**, Venn Diagram shows the overlap of increased accessibility and upregulated genes between ATACseq and RNAseq, respectively, in the control group at D3 and D5 over the *Foxk1* KO group. **E,F** The heatmap shows commonly expressed transcripts in both the ATACseq and RNAseq datasets upregulated and downregulated in the control EBs over the *Foxk1* KO EBs at D3 and D5. The heatmap color scheme key is provided, with red representing upregulated and blue representing downregulated genes.

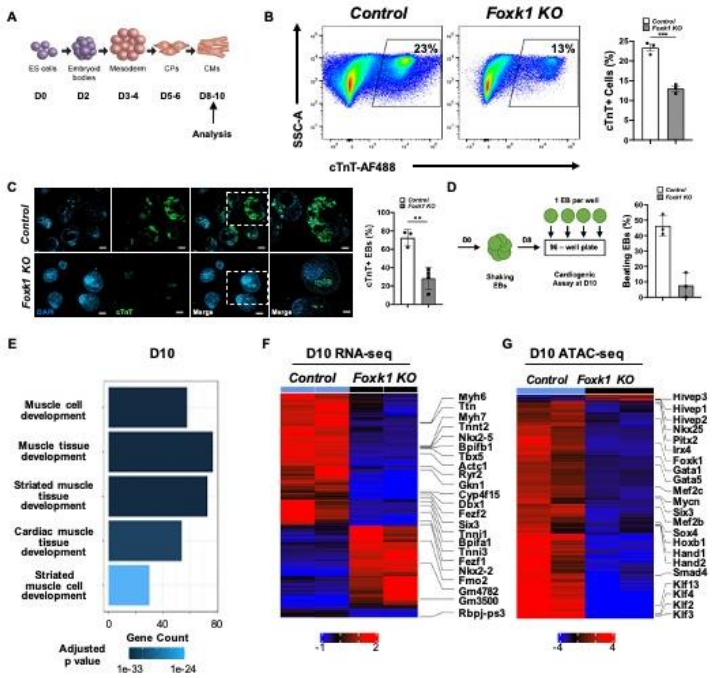


Figure 4. FOXK1 regulates cardiogenesis in differentiating EBs. A, Schematic of embryoid bodies (242) during in vitro differentiation with notation of cardiac milestones. **B**, Representative flow cytometry profile of control and *Foxk1* KO D10 EBs with quantification of the results. Note the significant decrease in cTnT+ cells in the *Foxk1* KO group compared to the control (n = 3, *p < 0.05). **C**, Immunohistochemical analysis of D10 EBs demonstrate that *Foxk1* KO EBs have perturbed cardiogenesis compared to the control group as assayed by cTnT staining. Quantification of the immunohistochemical results demonstrate normal cardiac differentiation in Day 10 EBs in control group that is significantly reduced in the absence of FOXK1 (n = 3, *p < 0.05). **D**, Schematic of control and *Foxk1* KO EBs, cardiogenic beating assay and quantification of results. Note the significant decrease in the number of beating EBs at D10 of differentiation in the *Foxk1* KO EB group compared to control (n = 3, *p < 0.05). Statistical test: Student's t-test. Data are presented as mean ± SEM. **E**, GO pathway analysis highlights pathways and development related terms in D10 control EBs, the x-axis represents the counts of genes in each GO term. The color scale shows the increased significance of the biological processes using the over-representation test with an adjusted p value < 0.05. **F**, The heatmap shows significantly (adjusted p-value < 0.05) differentially expressed transcripts upregulated and downregulated with a log2FoldChange between control EBs vs. *Foxk1* null EBs at D10. Red represents upregulation of transcripts and blue represents downregulation of transcripts. **G**, The heatmap represents a significant (adjusted p-value < 1e-04) change in chromatin accessibility at D10 for control and *Foxk1* KO. Red represents an increase in accessibility, and blue represents reduced accessibility for the transcription factor.

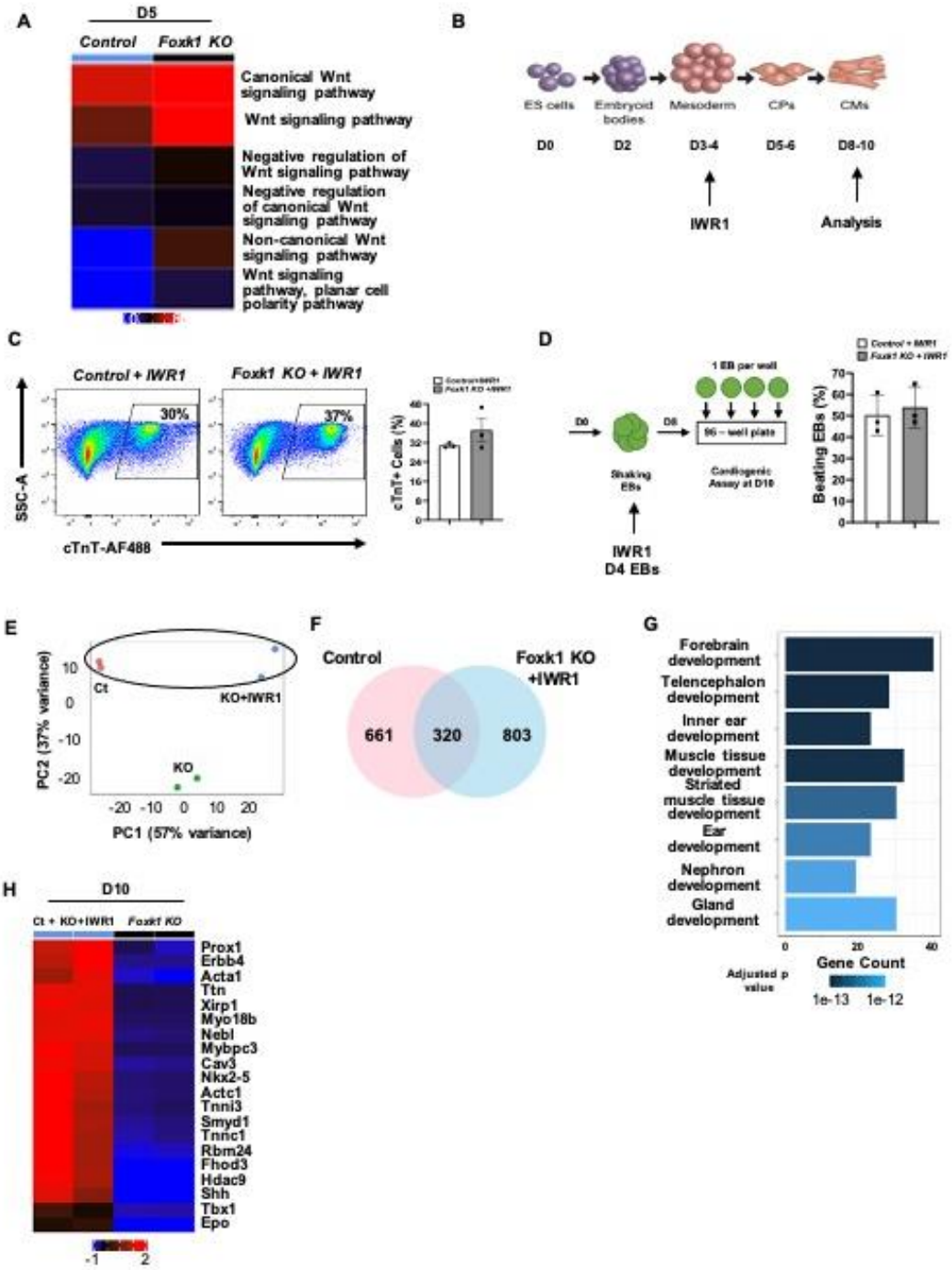

A
D5
Control
Foxk1 KO
Canonical Wnt signaling pathway
Wnt signaling pathway
Negative regulation of Wnt signaling pathway
Negative regulation of canonical Wnt signaling pathway
Non-canonical Wnt signaling pathway
Wnt signaling pathway, planar cell polarity pathway
B
ES cells
Embryoid bodies
Mesoderm
CPs
CMs
D0
D2
D3-4
D5-6
D8-10
IWR1
Analysis
C
Control + IWR1
Foxk1 KO + IWR1
30%
37%
SSC-A
cTnT-AF488
cTnT+ Cells (%)
D
1 EB per well
96 – well plate
Cardiogenic Assay at D10
Shaking EBs
IWR1
D4 EBs
Beating EBs (%)
E
PC2 (37% variance)
PC1 (57% variance)
Ct
KO+IWR1
KO
F
Control
Foxk1 KO +IWR1
661
320
803
G
Forebrain development
Telencephalon development
Inner ear development
Muscle tissue development
Striated muscle tissue development
Ear development
Nephron development
Gland development
Gene Count
Adjusted p value
1e-13
1e-12
H
D10
Ct + KO+IWR1
Foxk1 KO
Prox1
Erbb4
Acta1
Ttn
Xirp1
Myo18b
Nebl
Mybpc3
Cav3
Nkx2-5
Actc1
Tnni3
Smyd1
Tnnc1
Rbm24
Fhod3
Hdac9
Shh
Tbx1
Epo
-1
2

Figure 5. FOXK1 regulates Wnt signaling to promote cardiogenesis. A, RNAseq pathway analysis of Wnt signaling in D5 EBs comparing control and *Foxk1* KO groups. Wnt signaling is significantly upregulated in the *Foxk1* KO group over the control group and expression persists (in the absence of Foxk1) during differentiation. **B**, Schematic of EB in vitro differentiation with the Wnt signaling inhibitor (IWR1) and cardiac milestones noted. **C**, Representative flow cytometry profile of control + IWR1 and *Foxk1* KO + IWR1 D10 EBs with quantification of the results. Note that no significant differences were observed between the two groups (n = 3). **D**, Schematic of the beating assay of control + IWR1 and *Foxk1* KO + IWR1 D10 EBs with quantification of the results. Note that no significant differences were observed between the two groups (n = 3). Statistical test: Student's t-test. Data presented as mean ± SEM. **E**, The PCA of RNAseq of EB differentiation at D10 shows samples with similar gene expression cluster together. The second PCA (PC2) shows similarities between control and *Foxk1* KO+IWR1 (Ct: control; KO: *Foxk1* KO; KO+IWR1: *Foxk1* KO + IWR1). **F**, Venn diagram shows 320 genes up-regulated in both Ct and KO + IWR1 samples when compared to *Foxk1* KO samples. **G**, Top 10 GO pathways and development terms significantly enriched using the genes commonly upregulated in control and *Foxk1* KO + IWR1 conditions. The x-axis represents the counts of genes in each GO term. The color scale shows the increased significance of the biological processes using the over-representation test with an adjusted p value < 0.05. **H**, The heatmap shows top 20 genes from two GO ontology terms (muscle tissue development and striated muscle tissue development) commonly upregulated in control and *Foxk1* KO + IWR1 samples. Red represents upregulation of genes and blue represents downregulation of genes.

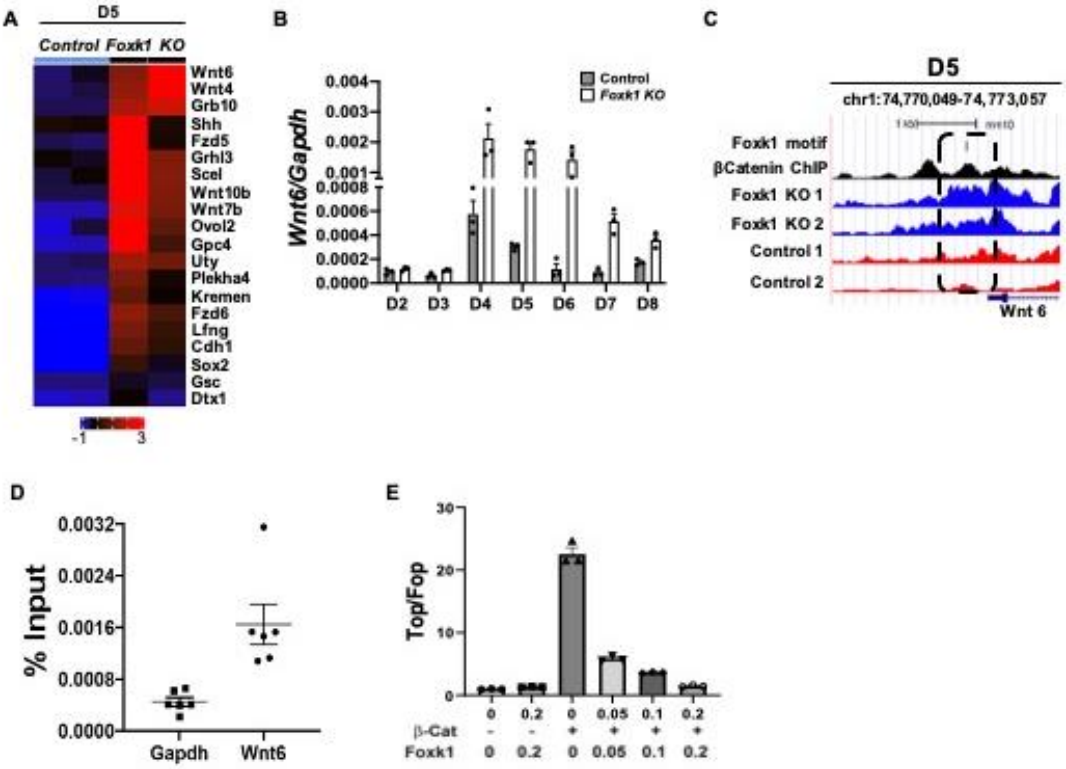


Figure 6. FOXK1 is a transcriptional repressor of Wnt signaling. A, Heatmap highlighting the top 20 transcription factors associated with the Wnt signaling pathway enriched in the *Foxk1* null D5 EB group from the RNA-seq dataset. **B**, Wnt6 transcript expression during EB differentiation from D2 to D8 in the presence and absence of FOXK1. Note that the expression of Wnt6 remains high at later stages of differentiation compared to the Control group where it is downregulated. **C**, A region upstream of *Wnt6* (chr1:74,770,049-74,773,057) that contains a FOXK1 DNA motif, β-catenin binding site and decreased chromatin accessibility in the absence of *Foxk1* compared to control EBs at D5. **D**, qPCR analysis using FOXK1 ChIP demonstrates a significant enrichment at the *Wnt6* upstream region compared to the Gapdh control region. **E**, β-Catenin transactivates the Top-flash reporter, which is inhibited by FOXK1. Statistical test: Student's t-test. Data presented as mean ± SEM.

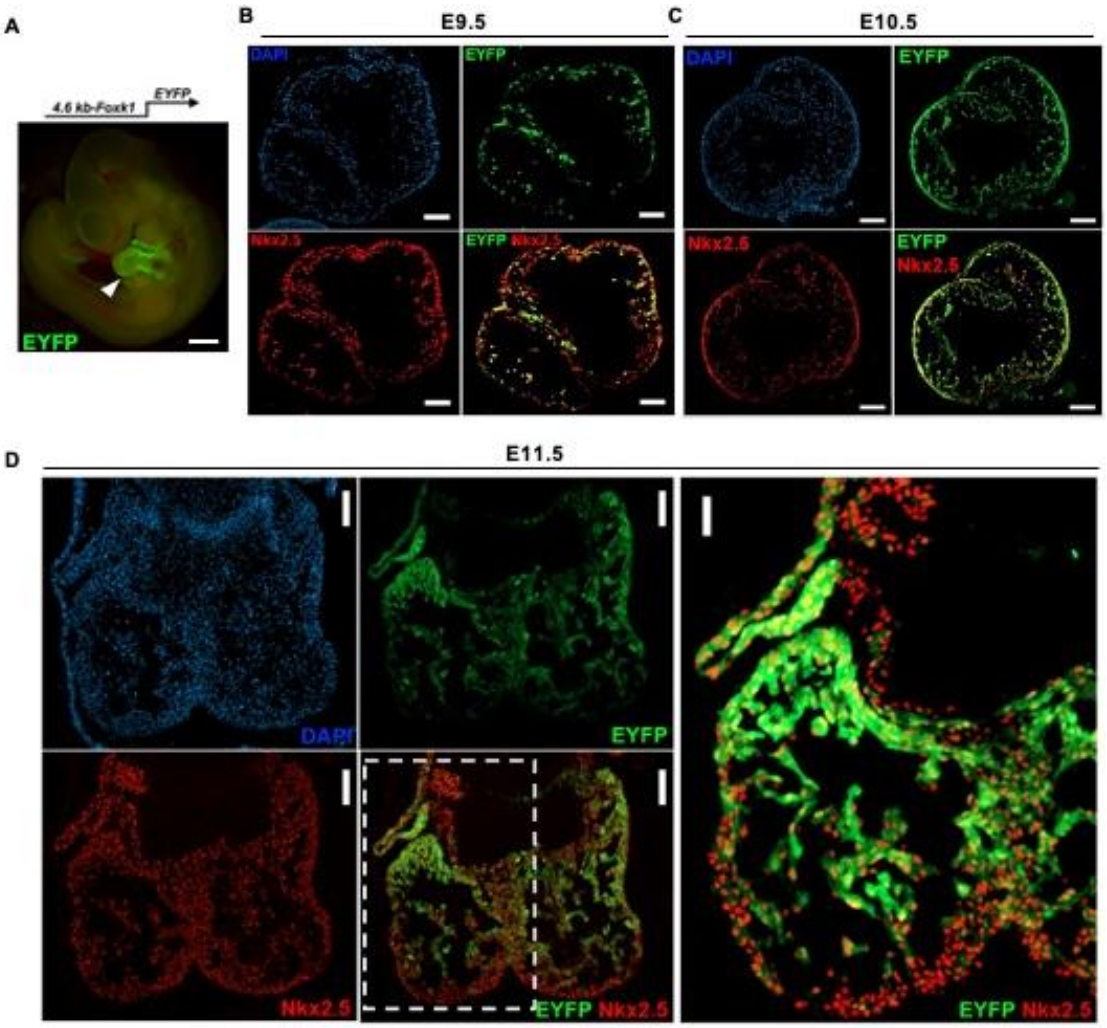
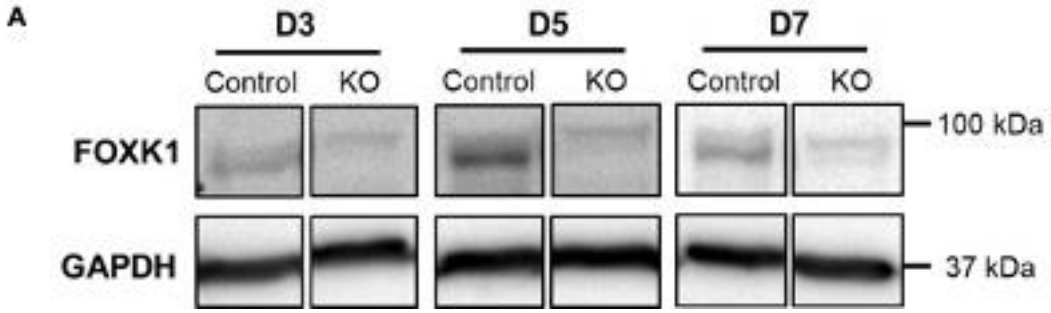
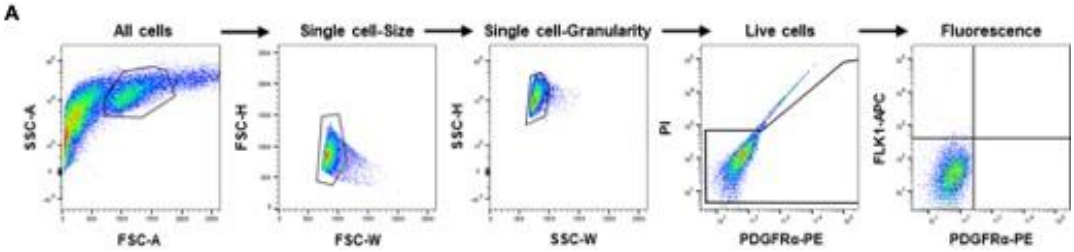


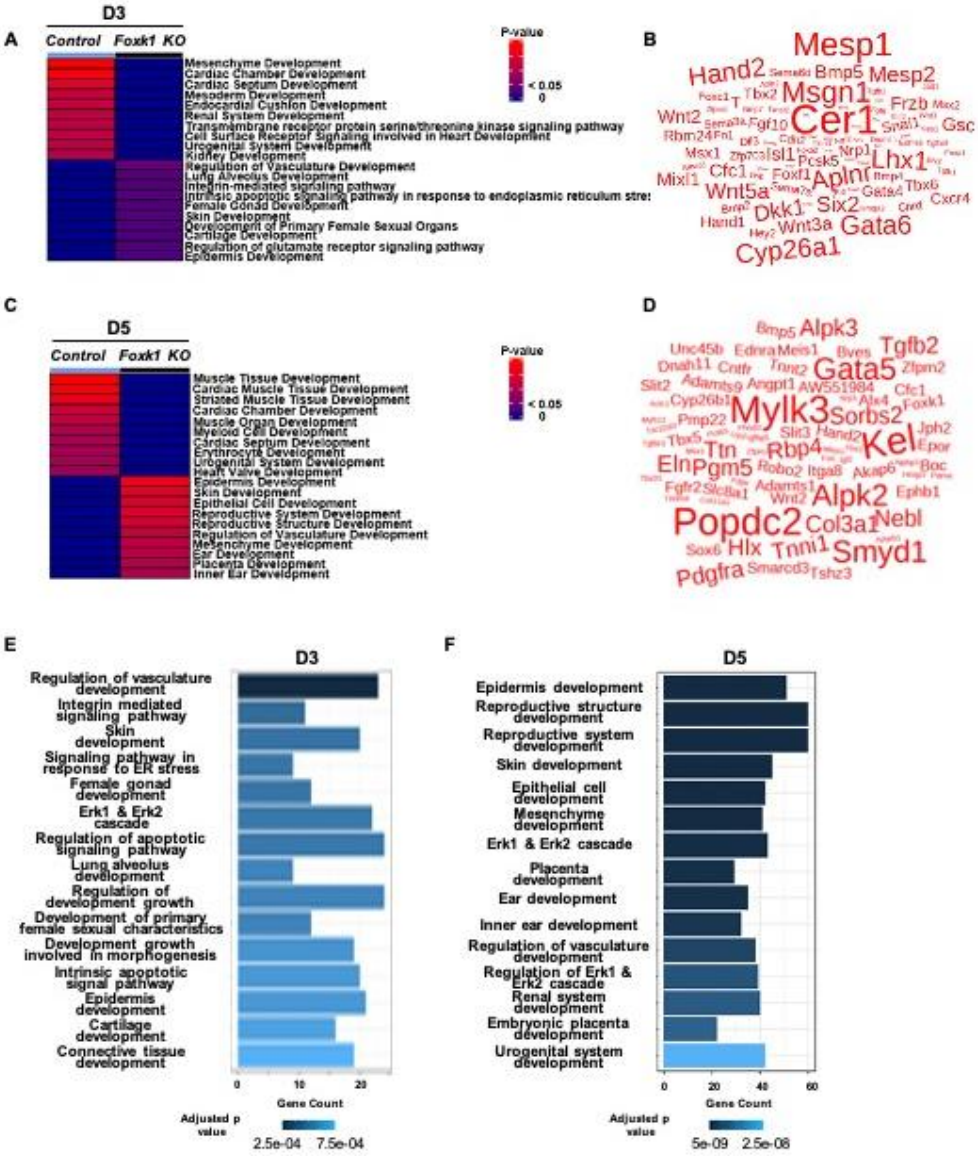
Figure 7. FOXK1 labels developing cardiac cells in the mouse embryo. A-D, Using the 4.6kb *Foxk1*-EYFP promoter-reporter transgenic mouse model, we demonstrated that this evolutionary conserved upstream fragment harbors modules that direct reporter expression to the developing heart in of E9.5, E10.5 and E11.5 mouse embryos. Immunohistochemical sections demonstrate that EYFP+ cells co-express Nkx2.5, particularly in the more differentiated trabeculae of the developing ventricles.



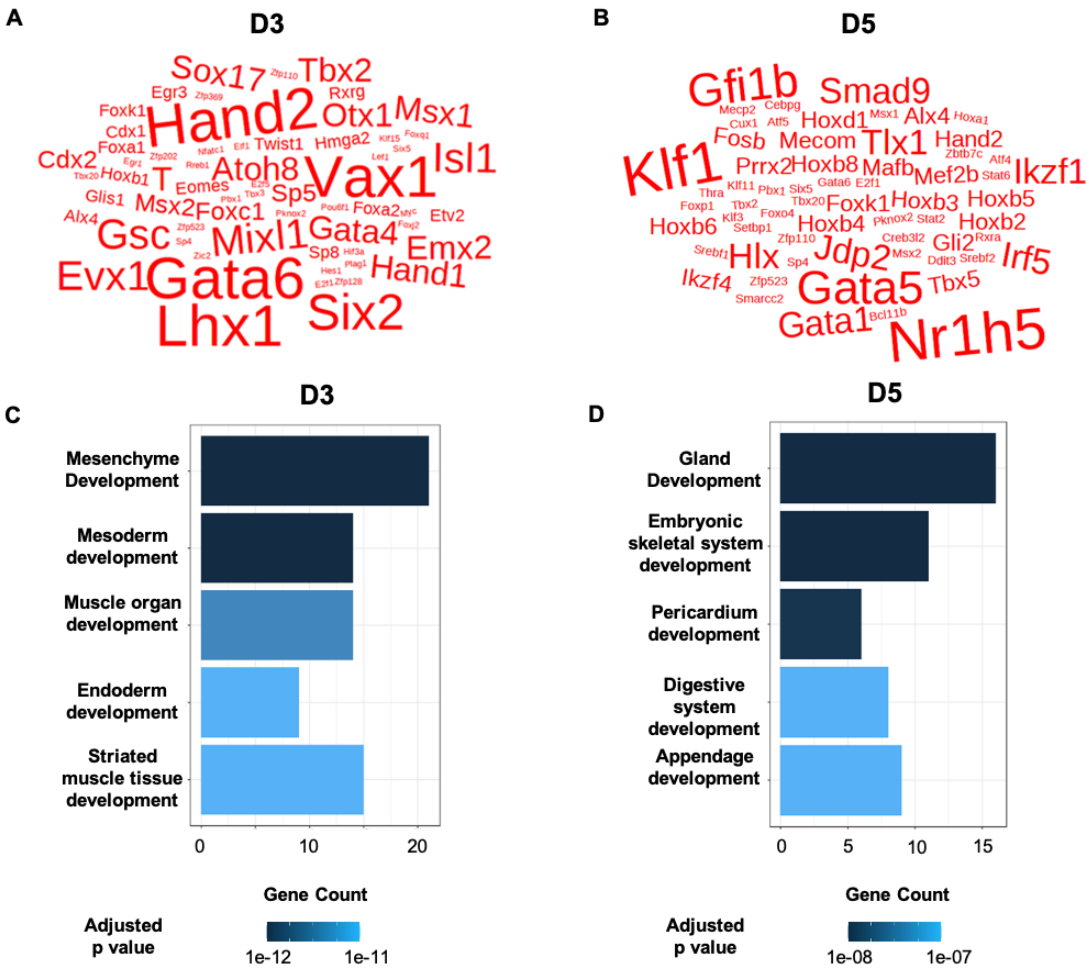
Supplementary Figure 1. FOXK1 protein expression in ES/EBs. A, FOXK1 protein expression in D3, D5 and D7 control and KO (*Foxk1* KO) EBs.



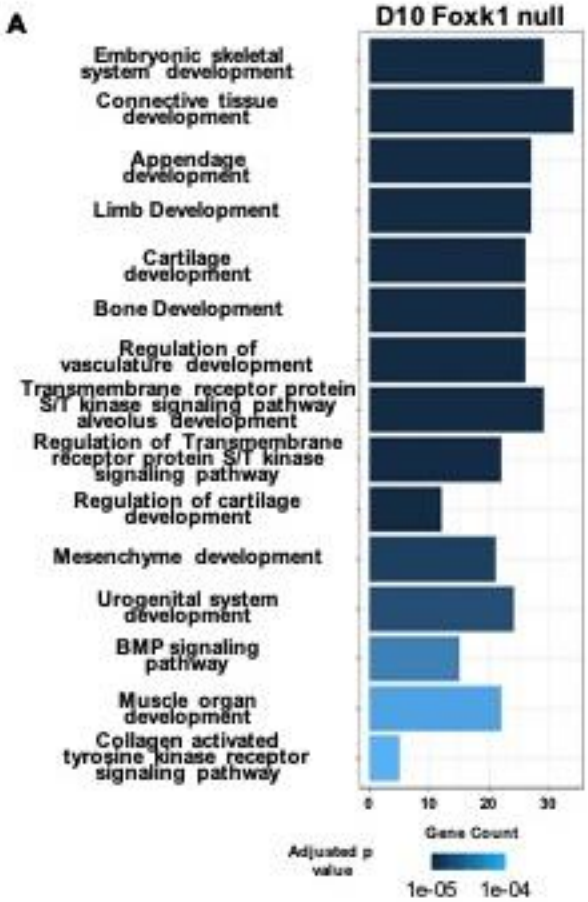
Supplementary Figure 2. Flow cytometry gating strategy. A, Flow cytometry gating strategy for differentiating embryoid bodies stained with antibodies.



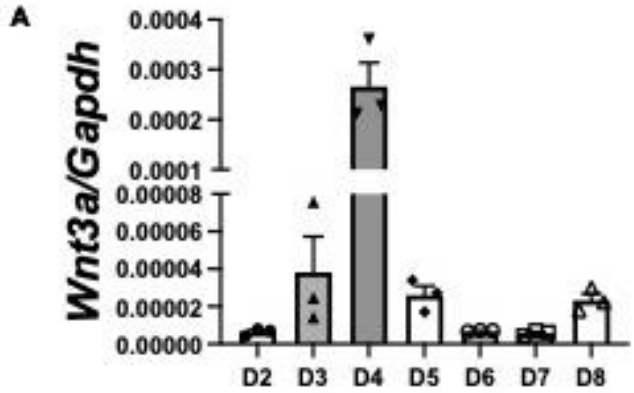
Supplementary Figure 3. RNAseq analysis of D3 and D5 EBs. A,**B**, The heatmap shows pathways significantly enriched in either control or *Foxk1* KO showing cardiac related pathways enriched in control D3 and D5. The pathways are significant with an adjusted p value < 0.05. Red color represents significance in the condition and blue represents pathways that are not significant. **C**,**D**, The word cloud shows genes up-regulated in control at D3 and D5 with the size directly representing a log2FC. **E**,**F**, The barplot shows GO pathway analysis highlighting pathways and development terms enriched in The *Foxk1* KO D3 and D5 respectively. The x-axis represents the counts of genes in each GO term. The color scale shows the increased significance of the biological processes using the over-representation test with an adjusted p value < 0.05.



Supplementary Figure 4. ATACseq analysis and RNAseq integration analysis of D3 and D5 EBs. A,**B**, The word cloud are commonly upregulated transcription factors between D3 and D5, respectively. The font size of the genes are proportional to the log2FoldChange values of the upregulated genes. C,D, Top 5 significantly enriched GO terms using genes that overlap between ATACseq and RNAseq. The x-axis represents the counts of genes in each GO term. The color scale shows the increased significance of the biological processes using the over-representation test with an adjusted p value < 0.05.



Supplementary Figure 5. RNAseq analysis of D10 EBs. A, The top 15 enriched GO pathways, highlighting pathways and development related terms in the *Foxk1* null at D10. The x-axis represents the counts of genes in each GO term. The color scale shows the increased significance of the biological processes using the over-representation test with an adjusted p value < 0.05.



Supplementary Figure 6. *Wnt3a* expression in differentiating EBs. A, Expression profile of *Wnt3a* throughout EB differentiation confirms a biphasic role of Wnt signaling during cardiogenesis in vitro in our differentiation protocol. Note that Wnt signaling is downregulated after D4 of differentiation. Data are presented as mean ± SEM.

CHAPTER 6: DISCUSSION & FUTURE DIRECTIONS

Summary

Cardiovascular disease (CVD) is the number one cause of death in the U.S and worldwide. The only curative disease for end stage cardiovascular disease is heart transplantation. Due of the shortage of organs available for transplantation, the development of curative therapies for CVD is desperately needed (2, 4, 243). In the field of regenerative medicine, we are interested understanding why adult mammalian heart has a limited regenerative capacity compared to other organs and early stages of development (2, 8-10). Particularly, we are interested in understanding why embryonic and neonatal hearts, unlike adult hearts, have the capacity to fully regenerate following cardiac injury (2, 9, 13, 14). Therefore, understanding the molecular signaling pathways governing this regenerative potential and designing efficacious, safe and potent delivery systems to express these molecular pathways can lead to the development of curative therapies for CVD. In this section of the thesis I will go over our discoveries, their significance and what I believe will be important moving forward in the field of regenerative medicine if we wish to translate these and other findings to the bedside.

In this thesis, we have identified novel epigenetic regulatory mechanisms that regulate cardiovascular development that we believe can have a big impact in the development of strategies to regenerate the adult heart following injury. Epigenetic regulation in cells have significant impacts in changing the identity of a cell. Global epigenetic changes are observed in every cell in numerous

processes such as movement, cell division, pathology, response to external stimuli, reprogramming, etc (244). Our end goal is to significantly impact cells in the heart following injury at the molecular level to either convert/reprogram to endothelial cells or push adult cardiomyocytes to re-enter the cell cycle and start proliferating (2). Exerting change at this scale in cells in the heart will require molecular changes at a global scale and I believe that targeting epigenetic regulatory mechanisms gives us a good chance (36).

Our first major finding is that we have identified ETV2 as a pioneer transcription factor for the endothelial lineage. Pioneer factors function to bind nucleosomal DNA and relax the chromatin landscape upstream of lineage specific genes. They can also repress lineages as well by maintaining a closed chromatin configuration or promoting the compaction of chromatin of previously closed regions (37). These factors reside at the very top of the hierarchical molecular cascade and can be powerful tools for us to develop reprogramming strategies to convert cardiac fibroblasts to functional vasculature following myocardial infarction. Pioneer factors also play an important role in cancer. They can open regions of chromatin that should remain silent as is the case of proto-oncogenes, and in this thesis, we have identified a new cascade involving ETV2 and RHOJ that regulates endothelial progenitor cell migration and angiogenesis. Whether this cascade has an impact in vivo remains to be seen, but as of now it is an attractive target for tumor angiogenesis. Finally, our last discovery involves the characterization of the transcription factor FOXK1 as a novel transcriptional

and epigenetic regulator for cardiovascular development. All these discoveries hold great potential for the field of cardiovascular regeneration and in this section, we discuss some of the implications of our studies, limitations and future directions.

ETV2 is a pioneer transcription factor for the endothelial lineage

We used two distinct biological systems: mouse embryonic stem cell/embryoid body (ES/EB) differentiation and fibroblast (MEF) reprogramming, to define the role of ETV2 as a pioneer factor that regulates endothelial fate and development. No other study in the field has used these diverse systems together to define pioneer factors and reprogramming capabilities. Even though these two model systems have very different global expression, chromatin accessibility and epigenetic profiles, we found similar molecular programs and downstream genes that were regulated following ETV2 induction. During endothelial cell reprogramming and differentiation, ETV2 targeted closed chromatin domains (in vitro and in vivo), confirming the key characteristic of a pioneer factor. Then ETV2 recruited BRG1, a chromatin remodeling enzyme and functioned together as a complex to relax closed chromatin and recruit other co-factors (i.e. ELK3). Loss of BRG1 significantly impacted the ability of ETV2 to reprogram and specify the endothelial lineage. This study was rigorous and used molecular biology techniques (ES/EB and MEFs), combined with computational genomics (ATACseq, bulk RNAseq, scRNAseq, ChIPseq and NOMEseq) and in vitro nucleosomal binding assays to unequivocally determine that ETV2 is a

pioneer factor. These results emphasize the importance of using ETV2 for the generation of mature vasculature in regenerative therapies and other areas of research. For example, recent work from our laboratory studying human:pig chimeras demonstrates that ETV2 can promote the formation of human vasculature in pig hosts that lack hematoendothelial lineages for future use in transplantation (108). While chimerism is still a key limitation in this study, the results are exciting because they bring us closer to the generation of mature tissues that can be transplantable. Some of the limitations observed with regards to chimerism can be solved in part by looking more deeply into all of the molecular and sequencing experiments we have performed demonstrating the efficient generation of mature vasculature cells in vitro.

Vasculogenesis and angiogenesis are known to be important for cardiac regeneration to occur in the neonatal mouse and therefore can be an important therapeutic avenue to explore for treating CVD (11). In terms of the applicability of using ETV2 to treat CVD, Lee et al., recently explored the ability of ETV2 to reprogram cells in the adult mouse heart following injury by overexpressing ETV2 using both lentiviruses and adeno-associated virus 9 (AAV9) (245). They demonstrated that overexpression of ETV2 in the injured heart promotes vascular regeneration and enhances cardiac repair (245). Although effective, lentiviruses and AAV9 target both non-dividing (cardiomyocytes) and dividing cells. Therefore, this study is flawed and while it showed encouraging data, we need to better assess the role of ETV2 and its ability to reprogram cardiac

fibroblasts in vivo before making any conclusions. To this end, more specific, potent, efficacious and safer (no risk of integration into the genome) strategies need to be considered when designing a therapy using ETV2 primarily due to its known role in cancer. Modified RNAs (modRNAs) that target specific cells in the heart have been recently developed and are expected to play a big role in the development of future therapies for CVD (246) and I envision that using an ETV2 modRNA will be impactful. modRNAs are an emerging technology for gene delivery that allows high gene expression in a variety of organs, including the heart of mice and pigs (247-249). While modRNA can be directly injected into muscle and it is able to be taken by cells and expressed, Rurik et al., recently demonstrated the capacity of reprogramming immune cells in vivo with modRNA packaged in a lipid nanoparticle to treat cardiac fibrosis (250). These results add an extra layer of complexity to the delivery mechanisms we can use to treat CVD by combining the technology with lipid nanoparticles, but nonetheless is very exciting for the field. The modifications made to modRNAs such as the replacement of uridine with psudouridine and cytidine with 5-methylcytidine, provides resistance to RNAses and evasion of the innate immune system, specifically toll-like receptors (247-249). More importantly, the delivery of a gene by modRNAs in vivo has been shown to lead to stable expression up to a one-week period when delivered into the heart (248). The efficacy and safety of modRNAs has been shown by their monumental impact on the development of extremely effective vaccines against SARS-CoV2 (251). Others have used modRNA to overexpress ETV2 in vitro and demonstrated its efficacy in producing

mature hematoendothelial lineages (133, 252). One important limitation of modRNAs is that they are very expensive to generate and not necessarily available to every group out there. Nonetheless, we propose that using an ETV2 modRNA that specifically targets fibroblasts will be impactful for the regeneration of the heart.

ETV2-RHOJ cascase regulates migration of endothelial cells

We also defined a novel role for ETV2 in the regulation of cell migration and demonstrated that ETV2 is a direct upstream activator of RHOJ to modulate its expression and together this cascade regulates endothelial progenitor cell migration. The impact of these findings, along with the characterized pioneer factor role of ETV2, is that these mechanisms and cascades are unique to the endothelial lineage and not blood. There is a need to develop better therapeutic strategies against angiogenesis in tumors to prevent growth and metastasis. The role of ETV2 and RHOJ, separately, in cancer has been defined already, however, putting this cascade together provides better therapeutic targets in different cancers, particularly in the context of angiogenesis (192, 208-210) (125, 209). For example, ETV2 has been implicated in glioblastoma and is thought to reprogram cells within tumors to form vasculature and provide oxygen and nutrients to the tumor (125). Due to the transient expression of ETV2 during embryogenesis, we hypothesize that understanding how downstream targets of ETV2 may play a role in tumorigenesis and become a target for therapies will be essential since we believe not all tumors will express ETV2 at the time of

analysis. All of the studies pertaining ETV2 and cancer have been done with patient samples, microRNAs and bulk RNAseq, therefore better analysis at the single cell level remains to be done in order to identify potential subsets or groups of cells that might play a bigger role in tumorigenesis and angiogenesis.

Our recent findings pertaining the role for ETV2 as a regulator of *Yes1* gene expression and the Hippo signaling pathway during embryogenesis (109) provides more relevance to targeting ETV2 in cancer. These studies identify ETV2 also as a regulator of cell proliferation. Previous studies have shown that cell migration and proliferation are highly coordinated and coregulated processes observed in cancer (211-214). Therefore, our studies support the hypothesis that ETV2 plays a critical role not only in development, but also in cancer by coregulating proliferation and migration via YES1 and RHOJ, respectively. Future studies will be needed to determine whether the Hippo signaling pathway modulates RHOJ expression and cell migration and to define any feedback mechanisms whereby the Hippo signaling pathway and RHOJ impact ETV2 expression.

FOXK1 is a novel regulator of cardiovascular development

The third major finding of this thesis is that we have identified a novel role for FOXK1 in the regulation of cardiovascular development. We have shown that FOXK1 acts as a transcriptional and epigenetic regulator of cardiovascular development by acting as a direct transcriptional and epigenetic repressor of the

Wnt/β signaling cascade, particularly the gene *Wnt6*. We used computational genomics to understand the impact that the absence of FOXK1 had in the development of mesodermal lineages, particularly the cardiac lineage. While these results are encouraging, further validation is needed using in vivo models and conditional knockout strategies since not all developmental pathways necessarily play a role in regeneration. The FOXK1 knockout mice are partially lethal by E9.5 and rarely produce offspring that can die following injury (64). Characterization of these mouse embryos is warranted to determine what is the cause of the lethality, which we hypothesize to be a cardiac phenotype. Additionally, developing a *Foxk1* floxed mouse will be important for dissecting the exact mechanism it plays in the heart when deleted there using a cardiac specific cre recombinase. While our in vivo studies using the 4.6 kb promoter of FOXK1 driving EYFP are encouraging, the above studies will be essential for us to continue to dissect these mechanisms in vivo.

Due to the known role of FOXK1 in regulating cell cycle kinetics, our findings in differentiating mESCs give us hope that FOXK1 might play a role during cardiac regeneration. Future studies need to focus on understanding the role of FOXK1 in the neonatal heart, particularly following injury. We already know that it plays a role in skeletal muscle injury, so we hypothesize that it might play a similar role in the neonatal mouse heart that possesses tremendous regenerative potential (64). One of the main limitations in regenerating the heart following injury is the inability to promote cell cycle re-entry of adult cardiomyocytes (2, 9). Therefore, overexpressing FOXK1 in the heart might be

beneficial just like ETV2, particularly if we can overexpress specifically in cardiomyocytes to overcome these limitations. Just like ETV2, utilizing a cardiomyocyte specific modRNA will be the correct way to go for these approaches (246, 253).

Concluding remarks

Taken together, the studies presented in this thesis provide novel pathways and signaling cascades that can have immediate impact in the development of therapies for cardiovascular diseases and oncological diseases. One of the main limitations in the field is the efficacy and safety of the delivery methods for overexpressing or repressing signaling pathways and cascades. With the advent of modRNAs, we believe that more therapies can be generated in order to enhance repair following injury and replace dead muscle with functional one. Additionally, any therapeutic strategies such as the ones proposed here with modRNAs will need to be conducted large animal model such as the pig or monkey. The reason for it is that while advantageous in many facets of scientific research, mice do not always accurately model human disease and the shortcomings of preclinical studies using mice are well recognized (254, 255). A lot of the studies cited in this thesis have been performed with cells or mice. However, large animal models such as pigs are key for cardiac translational research due to the many similarities they share with humans, both anatomically and physiologically (255, 256). Recent studies using porcine models to study cardiac regeneration have shown tremendous promise.

However, one big issue is that these have been primarily done by well-established faculty (more than 20-30 years in the field) and laboratories with significant funding from the NIH, DoD or the private sector (257-259). Young investigators cannot afford to perform these very expensive and high-risk experiments. Therefore, another important goal for the next 20-30 years in the field will be to enable everybody in the field to do these kinds of experiments. One solution that comes to mind is to find ways to increase the funding mechanisms provided by different agencies to allow for large animal models, modRNA generation and sequencing experiments that are instrumental for conducting the kind of research that is needed in the field of cardiac regeneration.

REFERENCES

1. Mayran A, Sochodolsky K, Khetchoumian K, Harris J, Gauthier Y, Bemmo A, et al. Pioneer and nonpioneer factor cooperation drives lineage specific chromatin opening. Nat Commun. 2019;10(1):3807.
2. Hashimoto H, Olson EN, Bassel-Duby R. Therapeutic approaches for cardiac regeneration and repair. Nat Rev Cardiol. 2018;15(10):585-600.
3. Malakar AK, Choudhury D, Halder B, Paul P, Uddin A, Chakraborty S. A review on coronary artery disease, its risk factors, and therapeutics. Journal of cellular physiology. 2019.
4. Nabel EG, Braunwald E. A tale of coronary artery disease and myocardial infarction. N Engl J Med. 2012;366(1):54-63.
5. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. Journal of the American College of Cardiology. 2017;70(1):1-25.
6. Ahuja P, Sdek P, MacLellan WR. Cardiac myocyte cell cycle control in development, disease, and regeneration. Physiological reviews. 2007;87(2):521-44.
7. Laflamme MA, Murry CE. Heart regeneration. Nature. 2011;473(7347):326-35.
8. Bergmann O, Bhardwaj RD, Bernard S, Zdunek S, Barnabe-Heider F, Walsh S, et al. Evidence for cardiomyocyte renewal in humans. Science (New York, NY). 2009;324(5923):98-102.
9. Porrello ER, Mahmoud AI, Simpson E, Hill JA, Richardson JA, Olson EN, et al. Transient regenerative potential of the neonatal mouse heart. Science. 2011;331(6020):1078-80.
10. Song K, Nam YJ, Luo X, Qi X, Tan W, Huang GN, et al. Heart repair by reprogramming non-myocytes with cardiac transcription factors. Nature. 2012;485(7400):599-604.
11. Aurora AB, Porrello ER, Tan W, Mahmoud AI, Hill JA, Bassel-Duby R, et al. Macrophages are required for neonatal heart regeneration. J Clin Invest. 2014;124(3):1382-92.
12. Gupta R, Tongers J, Losordo DW. Human studies of angiogenic gene therapy. Circulation research. 2009;105(8):724-36.
13. Riching AS, Song K. Cardiac Regeneration: New Insights Into the Frontier of Ischemic Heart Failure Therapy. Front Bioeng Biotechnol. 2020;8:637538.
14. Sadek H, Olson EN. Toward the Goal of Human Heart Regeneration. Cell Stem Cell. 2020;26(1):7-16.
15. Fishman MC, Olson EN. Parsing the heart: genetic modules for organ assembly. Cell. 1997;91(2):153-6.
16. Garry DJ, Martin CM. Cardiac regeneration: self-service at the pump. Circulation research. 2004;95(9):852-4.
17. McFadden DG, Olson EN. Heart development: learning from mistakes. Current opinion in genetics & development. 2002;12(3):328-35.
18. Olson EN, Schneider MD. Sizing up the heart: development redux in disease. Genes & development. 2003;17(16):1937-56.
19. Olson EN, Srivastava D. Molecular pathways controlling heart development. Science (New York, NY). 1996;272(5262):671-6.

20. Parmacek MS, Epstein JA. Pursuing Cardiac Progenitors: Regeneration Redux. Cell. 2005;120(3):295-8.
21. Srivastava D. Making or breaking the heart: from lineage determination to morphogenesis. Cell. 2006;126(6):1037-48.
22. Bruneau BG. The developmental genetics of congenital heart disease. Nature. 2008;451(7181):943-8.
23. Hoffman JI. Incidence of congenital heart disease: I. Postnatal incidence. Pediatric cardiology. 1995;16(3):103-13.
24. Rasmussen TL, Raveendran G, Zhang J, Garry DJ. Getting to the heart of myocardial stem cells and cell therapy. Circulation. 2011;123(16):1771-9.
25. Evans SM, Yelon D, Conlon FL, Kirby ML. Myocardial lineage development. Circulation research. 2010;107(12):1428-44.
26. Chan SS, Kyba M. What is a Master Regulator? J Stem Cell Res Ther. 2013;3.
27. Davis TL, Rebay I. Master regulators in development: Views from the Drosophila retinal determination and mammalian pluripotency gene networks. Dev Biol. 2017;421(2):93-107.
28. Tapscott SJ. The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. Development. 2005;132(12):2685-95.
29. Vinogradova TV, Sverdlov ED. PDX1: A Unique Pancreatic Master Regulator Constantly Changes Its Functions during Embryonic Development and Progression of Pancreatic Cancer. Biochemistry (Mosc). 2017;82(8):887-93.
30. Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell. 1987;51(6):987-1000.
31. Tapscott SJ, Davis RL, Thayer MJ, Cheng PF, Weintraub H, Lassar AB. MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. Science. 1988;242(4877):405-11.
32. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005;122(6):947-56.
33. Porcher C, Swat W, Rockwell K, Fujiwara Y, Alt FW, Orkin SH. The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. Cell. 1996;86(1):47-57.
34. Semenza GL. Hypoxia-inducible factor 1: master regulator of O2 homeostasis. Curr Opin Genet Dev. 1998;8(5):588-94.
35. Bondue A, Lapouge G, Paulissen C, Semeraro C, Iacovino M, Kyba M, et al. Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. Cell Stem Cell. 2008;3(1):69-84.
36. Wang H, Yang Y, Liu J, Qian L. Direct cell reprogramming: approaches, mechanisms and progress. Nat Rev Mol Cell Biol. 2021;22(6):410-24.
37. Zaret KS. Pioneer Transcription Factors Initiating Gene Network Changes. Annu Rev Genet. 2020;54:367-85.
38. Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. Genes Dev. 2011;25(21):2227-41.
39. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. Genes Dev. 2014;28(24):2679-92.

40. Zaret KS, Lerner J, Iwafuchi-Doi M. Chromatin Scanning by Dynamic Binding of Pioneer Factors. Mol Cell. 2016;62(5):665-7.
41. Iwafuchi-Doi M, Zaret KS. Cell fate control by pioneer transcription factors. Development. 2016;143(11):1833-7.
42. Zaret KS, Mango SE. Pioneer transcription factors, chromatin dynamics, and cell fate control. Current opinion in genetics & development. 2016;37:76-81.
43. Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell. 2012;151(5):994-1004.
44. Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell. 2015;161(3):555-68.
45. Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, et al. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. Mol Cell. 2016;62(1):79-91.
46. Ferdous A, Caprioli A, Iacovino M, Martin CM, Morris J, Richardson JA, et al. Nkx2-5 transactivates the Ets-related protein 71 gene and specifies an endothelial/endocardial fate in the developing embryo. Proc Natl Acad Sci U S A. 2009;106(3):814-9.
47. Koyano-Nakagawa N, Garry DJ. Etv2 as an essential regulator of mesodermal lineage development. Cardiovascular research. 2017;113(11):1294-306.
48. Lee D, Park C, Lee H, Lugus JJ, Kim SH, Arentson E, et al. ER71 acts downstream of BMP, Notch, and Wnt signaling in blood and vessel progenitor specification. Cell Stem Cell. 2008;2(5):497-507.
49. Liu F, Li D, Yu YY, Kang I, Cha MJ, Kim JY, et al. Induction of hematopoietic and endothelial cell program orchestrated by ETS transcription factor ER71/ETV2. EMBO Rep. 2015;16(5):654-69.
50. Rasmussen TL, Kweon J, Diekmann MA, Belema-Bedada F, Song Q, Bowlin K, et al. ER71 directs mesodermal fate decisions during embryogenesis. Development. 2011;138(21):4801-12.
51. Singh BN, Kawakami Y, Akiyama R, Rasmussen TL, Garry MG, Gong W, et al. The Etv2-miR-130a Network Regulates Mesodermal Specification. Cell Rep. 2015;13(5):915-23.
52. De Val S, Chi NC, Meadows SM, Minovitsky S, Anderson JP, Harris IS, et al. Combinatorial regulation of endothelial gene expression by ets and forkhead transcription factors. Cell. 2008;135(6):1053-64.
53. Garry DJ. Etv2 IS A MASTER REGULATOR OF HEMATOENDOTHELIAL LINEAGES. Transactions of the American Clinical and Climatological Association. 2016;127:212-23.
54. Han JK, Chang SH, Cho HJ, Choi SB, Ahn HS, Lee J, et al. Direct conversion of adult skin fibroblasts to endothelial cells by defined factors. Circulation. 2014;130(14):1168-78.
55. Lee S, Park C, Han JW, Kim JY, Cho K, Kim EJ, et al. Direct Reprogramming of Human Dermal Fibroblasts Into Endothelial Cells Using ER71/ETV2. Circ Res. 2017;120(5):848-61.

56. Golson ML, Kaestner KH. Fox transcription factors: from development to disease. Development (Cambridge, England). 2016;143(24):4558-70.
57. Liu Y, Ding W, Ge H, Ponnusamy M, Wang Q, Hao X, et al. FOXK transcription factors: Regulation and critical role in cancer. Cancer Lett. 2019;458:1-12.
58. Garry DJ, Yang Q, Bassel-Duby R, Williams RS. Persistent expression of MNF identifies myogenic stem cells in postnatal muscles. Developmental biology. 1997;188(2):280-94.
59. Hawke TJ, Jiang N, Garry DJ. Absence of p21CIP rescues myogenic progenitor cell proliferative and regenerative capacity in Foxk1 null mice. The Journal of biological chemistry. 2003;278(6):4015-20.
60. Shi X, Bowlin KM, Garry DJ. Fhl2 interacts with Foxk1 and corepresses Foxo4 activity in myogenic progenitors. Stem cells (Dayton, Ohio). 2010;28(3):462-9.
61. Alexander MS, Shi X, Voelker KA, Grange RW, Garcia JA, Hammer RE, et al. Foxj3 transcriptionally activates Mef2c and regulates adult skeletal muscle fiber type identity. Developmental biology. 2010;337(2):396-404.
62. Shi X, Seldin DC, Garry DJ. Foxk1 recruits the Sds3 complex and represses gene expression in myogenic progenitors. The Biochemical journal. 2012;446(3):349-57.
63. Hawke TJ, Garry DJ. Myogenic satellite cells: physiology to molecular biology. Journal of applied physiology (Bethesda, Md : 1985). 2001;91(2):534-51.
64. Garry DJ, Meeson A, Elterman J, Zhao Y, Yang P, Bassel-Duby R, et al. Myogenic stem cell function is impaired in mice lacking the forkhead/winged helix protein MNF. Proc Natl Acad Sci U S A. 2000;97(10):5416-21.
65. Hawke TJ, Meeson AP, Jiang N, Graham S, Hutcheson K, DiMaio JM, et al. p21 is essential for normal myogenic progenitor cell function in regenerating skeletal muscle. American journal of physiology Cell physiology. 2003;285(5):C1019-27.
66. Gallardo TD, Hammer RE, Garry DJ. RNA amplification and transcriptional profiling for analysis of stem cell populations. Genesis (New York, NY : 2000). 2003;37(2):57-63.
67. Shi X, Garry DJ. Sin3 interacts with Foxk1 and regulates myogenic progenitors. Molecular and cellular biochemistry. 2012;366(1-2):251-8.
68. Meeson AP, Shi X, Alexander MS, Williams RS, Allen RE, Jiang N, et al. Sox15 and Fhl3 transcriptionally coactivate Foxk1 and regulate myogenic progenitor cells. The EMBO journal. 2007;26(7):1902-12.
69. Yang Q, Kong Y, Rothermel B, Garry DJ, Bassel-Duby R, Williams RS. The winged-helix/forkhead protein myocyte nuclear factor beta (MNF-beta) forms a co-repressor complex with mammalian sin3B. The Biochemical journal. 2000;345 Pt 2:335-43.
70. Shi X, Wallis AM, Gerard RD, Voelker KA, Grange RW, DePinho RA, et al. Foxk1 promotes cell proliferation and represses myogenic differentiation by regulating Foxo4 and Mef2. J Cell Sci. 2012;125(Pt 22):5329-37.
71. Garry DJ, Maeng, G. and Garry, , 8(17), 1041. . Foxk1 regulates cancer progression. Annals of Translational Medicine. 2020:1041.
72. Ferber S, Halkin A, Cohen H, Ber I, Einav Y, Goldberg I, et al. Pancreatic and duodenal homeobox gene 1 induces expression of insulin genes in liver and ameliorates streptozotocin-induced hyperglycemia. Nat Med. 2000;6(5):568-72.

73. Heller RS, Stoffers DA, Bock T, Svenstrup K, Jensen J, Horn T, et al. Improved glucose tolerance and acinar dysmorphogenesis by targeted expression of transcription factor PDX-1 to the exocrine pancreas. Diabetes. 2001;50(7):1553-61.
74. Wuming Gong SD, Javier E. Sierra-Pagan Erik Skie, E, Nikita Dsouza, Thijs A. Larson, Mary G. Garry, Edgar Luzete-Monteiro, Kenneth S. Zaret, Daniel J. Garry. ETV2 functions as a pioneer factor to regulate and reprogram the endothelial lineage. 2022.
75. Balsalobre A, Drouin J. Pioneer factors as master regulators of the epigenome and cell fate. Nat Rev Mol Cell Biol. 2022.
76. Gualdi R, Bossard P, Zheng M, Hamada Y, Coleman JR, Zaret KS. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. Genes Dev. 1996;10(13):1670-82.
77. McPherson CE, Shim EY, Friedman DS, Zaret KS. An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. Cell. 1993;75(2):387-98.
78. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell. 2002;9(2):279-89.
79. Singh BN, Koyano-Nakagawa N, Gong W, Moskowitz IP, Weaver CV, Braunlin E, et al. A conserved HH-Gli1-Mycn network regulates heart regeneration from newt to human. Nat Commun. 2018;9(1):4237.
80. Cenik BK, Shilatifard A. COMPASS and SWI/SNF complexes in development and disease. Nat Rev Genet. 2021;22(1):38-58.
81. Bultman S, Gebuhr T, Yee D, La Mantia C, Nicholson J, Gilliam A, et al. A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. Mol Cell. 2000;6(6):1287-95.
82. Alexander JM, Hota SK, He D, Thomas S, Ho L, Pennacchio LA, et al. Brg1 modulates enhancer activation in mesoderm lineage commitment. Development. 2015;142(8):1418-30.
83. Hang CT, Yang J, Han P, Cheng HL, Shang C, Ashley E, et al. Chromatin regulation by Brg1 underlies heart muscle development and disease. Nature. 2010;466(7302):62-7.
84. Gao R, Liang X, Cheedipudi S, Cordero J, Jiang X, Zhang Q, et al. Pioneering function of Isl1 in the epigenetic control of cardiomyocyte cell fate. Cell Res. 2019;29(6):486-501.
85. King HW, Klose RJ. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. Elife. 2017;6.
86. Takaku M, Grimm SA, Shimbo T, Perera L, Menafra R, Stunnenberg HG, et al. GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. Genome Biol. 2016;17:36.
87. Ferretti E, Hadjantonakis AK. Mesoderm specification and diversification: from single cells to emergent tissues. Curr Opin Cell Biol. 2019;61:110-6.
88. Chang CP, Bruneau BG. Epigenetics and cardiovascular development. Annu Rev Physiol. 2012;74:41-68.

89. Tam PP, Behringer RR. Mouse gastrulation: the formation of a mammalian body plan. Mech Dev. 1997;68(1-2):3-25.
90. Srivastava D, Olson EN. A genetic blueprint for cardiac development. Nature. 2000;407(6801):221-6.
91. Kwon C, Qian L, Cheng P, Nigam V, Arnold J, Srivastava D. A regulatory pathway involving Notch1/beta-catenin/Isl1 determines cardiac progenitor cell fate. Nat Cell Biol. 2009;11(8):951-7.
92. Moretti A, Caron L, Nakano A, Lam JT, Bernshausen A, Chen Y, et al. Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. Cell. 2006;127(6):1151-65.
93. Laugwitz KL, Moretti A, Lam J, Gruber P, Chen Y, Woodard S, et al. Postnatal isl1+ cardioblasts enter fully differentiated cardiomyocyte lineages. Nature. 2005;433(7026):647-53.
94. Cai CL, Liang X, Shi Y, Chu PH, Pfaff SL, Chen J, 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):877-89.
95. Oka T, Maillet M, Watt AJ, Schwartz RJ, Aronow BJ, Duncan SA, et al. Cardiac-specific deletion of Gata4 reveals its requirement for hypertrophy, compensation, and myocyte viability. Circ Res. 2006;98(6):837-45.
96. Bisping E, Ikeda S, Kong SW, Tarnavski O, Bodyak N, McMullen JR, et al. Gata4 is required for maintenance of postnatal cardiac function and protection from pressure overload-induced heart failure. Proc Natl Acad Sci U S A. 2006;103(39):14471-6.
97. Zeisberg EM, Ma Q, Juraszek AL, Moses K, Schwartz RJ, Izumo S, et al. Morphogenesis of the right ventricle requires myocardial expression of Gata4. J Clin Invest. 2005;115(6):1522-31.
98. Watt AJ, Battle MA, Li J, Duncan SA. GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. Proc Natl Acad Sci U S A. 2004;101(34):12573-8.
99. Molkentin JD, Lin Q, Duncan SA, Olson EN. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. Genes Dev. 1997;11(8):1061-72.
100. Kuo CT, Morrisey EE, Anandappa R, Sigrist K, Lu MM, Parmacek MS, et al. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. Genes Dev. 1997;11(8):1048-60.
101. Fu JD, Srivastava D. Direct reprogramming of fibroblasts into cardiomyocytes for cardiac regenerative medicine. Circ J. 2015;79(2):245-54.
102. Nam YJ, Song K, Olson EN. Heart repair by cardiac reprogramming. Nat Med. 2013;19(4):413-5.
103. Fu JD, Stone NR, Liu L, Spencer CI, Qian L, Hayashi Y, et al. Direct reprogramming of human fibroblasts toward a cardiomyocyte-like state. Stem Cell Reports. 2013;1(3):235-47.
104. Ieda M, Fu JD, Delgado-Olguin P, Vedantham V, Hayashi Y, Bruneau BG, et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell. 2010;142(3):375-86.

105. Huang P, He Z, Ji S, Sun H, Xiang D, Liu C, et al. Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. Nature. 2011;475(7356):386-9.
106. Stone NR, Gifford CA, Thomas R, Pratt KJB, Samse-Knapp K, Mohamed TMA, et al. Context-Specific Transcription Factor Functions Regulate Epigenomic and Transcriptional Dynamics during Cardiac Reprogramming. Cell Stem Cell. 2019;25(1):87-102 e9.
107. Singh BN, Sierra-Pagan JE, Gong W, Das S, Theisen JWM, Skie E, et al. ETV2 (Ets Variant Transcription Factor 2)-Rhoj Cascade Regulates Endothelial Progenitor Cell Migration During Embryogenesis. Arterioscler Thromb Vasc Biol. 2020;40(12):2875-90.
108. Das S, Koyano-Nakagawa N, Gafni O, Maeng G, Singh BN, Rasmussen T, et al. Generation of human endothelium in pig embryos deficient in ETV2. Nat Biotechnol. 2020;38(3):297-302.
109. Singh BN, Gong W, Das S, Theisen JWM, Sierra-Pagan JE, Yannopoulos D, et al. Etv2 transcriptionally regulates Yes1 and promotes cell proliferation during embryogenesis. Sci Rep. 2019;9(1):9736.
110. Singh BN, Tahara N, Kawakami Y, Das S, Koyano-Nakagawa N, Gong W, et al. Etv2-miR-130a-Jarid2 cascade regulates vascular patterning during embryogenesis. PLoS One. 2017;12(12):e0189010.
111. Koyano-Nakagawa N, Shi X, Rasmussen TL, Das S, Walter CA, Garry DJ. Feedback Mechanisms Regulate Ets Variant 2 (Etv2) Gene Expression and Hematoendothelial Lineages. J Biol Chem. 2015;290(47):28107-19.
112. Rasmussen TL, Martin CM, Walter CA, Shi X, Perlingeiro R, Koyano-Nakagawa N, et al. Etv2 rescues Flk1 mutant embryoid bodies. Genesis. 2013;51(7):471-80.
113. Koyano-Nakagawa N, Kweon J, Iacovino M, Shi X, Rasmussen TL, Borges L, et al. Etv2 is expressed in the yolk sac hematopoietic and endothelial progenitors and regulates Lmo2 gene expression. Stem Cells. 2012;30(8):1611-23.
114. Basel-Salmon L, Ruhrman-Shahar N, Barel O, Hagari O, Marek-Yagel D, Azulai N, et al. Biallelic variants in ETV2 in a family with congenital heart defects, vertebral abnormalities and preaxial polydactyly. Eur J Med Genet. 2021;64(2):104124.
115. Garry DJ, Sierra-Pagan JE. Mechanisms that Govern Endothelial Lineage Development and Vasculogenesis. Advanced Technologies in Cardiovascular Bioengineering: Springer; 2022. p. 31-48.
116. Minderjahn J, Schmidt A, Fuchs A, Schill R, Raithel J, Babina M, et al. Mechanisms governing the pioneering and redistribution capabilities of the non-classical pioneer PU.1. Nat Commun. 2020;11(1):402.
117. Ungerback J, Hosokawa H, Wang X, Strid T, Williams BA, Sigvardsson M, et al. Pioneering, chromatin remodeling, and epigenetic constraint in early T-cell gene regulation by SPI1 (PU.1). Genome Res. 2018;28(10):1508-19.
118. Li R, Cauchy P, Ramamoorthy S, Boller S, Chavez L, Grosschedl R. Dynamic EBF1 occupancy directs sequential epigenetic and transcriptional events in B-cell programming. Genes Dev. 2018;32(2):96-111.
119. Boller S, Ramamoorthy S, Akbas D, Nechanitzky R, Burger L, Murr R, et al. Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. Immunity. 2016;44(3):527-41.

120. van Oevelen C, Collombet S, Vicent G, Hoogenkamp M, Lepoivre C, Badeaux A, et al. C/EBPalpha Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. Stem Cell Reports. 2015;5(2):232-47.
121. Barozzi I, Simonatto M, Bonifacio S, Yang L, Rohs R, Ghisletti S, et al. Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. Mol Cell. 2014;54(5):844-57.
122. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576-89.
123. Feng R, Desbordes SC, Xie H, Tillo ES, Pixley F, Stanley ER, et al. PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. Proc Natl Acad Sci U S A. 2008;105(16):6057-62.
124. Lee TJ, Kang HK, Berry JC, Joo HG, Park C, Miller MJ, et al. ER71/ETV2 Promotes Hair Regeneration from Chemotherapeutic Drug-Induced Hair Loss by Enhancing Angiogenesis. Biomol Ther (Seoul). 2021;29(5):545-50.
125. Zhao C, Gomez GA, Zhao Y, Yang Y, Cao D, Lu J, et al. ETV2 mediates endothelial transdifferentiation of glioblastoma. Signal transduction and targeted therapy. 2018;3(1):1-11.
126. Kabir AU, Lee T-J, Pan H, Berry JC, Krchma K, Wu J, et al. Requisite endothelial reactivation and effective siRNA nanoparticle targeting of Etv2/Er71 in tumor angiogenesis. JCI insight. 2018;3(8).
127. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature. 2015;523(7561):486-90.
128. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015;109:21 9 1- 9 9.
129. Ye L, Zimmermann WH, Garry DJ, Zhang J. Patching the heart: cardiac repair from within and outside. Circ Res. 2013;113(7):922-32.
130. Lugano R, Ramachandran M, Dimberg A. Tumor angiogenesis: causes, consequences, challenges and opportunities. Cell Mol Life Sci. 2020;77(9):1745-70.
131. Ginsberg M, James D, Ding BS, Nolan D, Geng F, Butler JM, et al. Efficient direct reprogramming of mature amniotic cells into endothelial cells by ETS factors and TGFbeta suppression. Cell. 2012;151(3):559-75.
132. Palikuqi B, Nguyen DT, Li G, Schreiner R, Pellegata AF, Liu Y, et al. Adaptable haemodynamic endothelial cells for organogenesis and tumorigenesis. Nature. 2020;585(7825):426-32.
133. Wang K, Lin RZ, Hong X, Ng AH, Lee CN, Neumeyer J, et al. Robust differentiation of human pluripotent stem cells into endothelial cells via temporal modulation of ETV2 with modified mRNA. Sci Adv. 2020;6(30):eaba7606.
134. Wapinski OL, Vierbuchen T, Qu K, Lee QY, Chanda S, Fuentes DR, et al. Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell. 2013;155(3):621-35.
135. Garry DJ, Olson EN. A common progenitor at the heart of development. Cell. 2006;127(6):1101-4.

136. Gong W, Rasmussen TL, Singh BN, Koyano-Nakagawa N, Pan W, Garry DJ. Dpath software reveals hierarchical haemato-endothelial lineages of Etv2 progenitors based on single-cell transcriptome analysis. Nat Commun. 2017;8:14362.
137. Shi X, Richard J, Zirbes KM, Gong W, Lin G, Kyba M, et al. Cooperative interaction of Etv2 and Gata2 regulates the development of endothelial and hematopoietic lineages. Dev Biol. 2014;389(2):208-18.
138. Rasmussen TL, Shi X, Wallis A, Kweon J, Zirbes KM, Koyano-Nakagawa N, et al. VEGF/Flk1 signaling cascade transactivates Etv2 gene expression. PLoS One. 2012;7(11):e50103.
139. Sumanas S, Choi K. ETS Transcription Factor ETV2/ER71/Etsrp in Hematopoietic and Vascular Development. Curr Top Dev Biol. 2016;118:77-111.
140. Griffin CT, Curtis CD, Davis RB, Muthukumar V, Magnuson T. The chromatin-remodeling enzyme BRG1 modulates vascular Wnt signaling at two levels. Proc Natl Acad Sci U S A. 2011;108(6):2282-7.
141. Griffin CT, Brennan J, Magnuson T. The chromatin-remodeling enzyme BRG1 plays an essential role in primitive erythropoiesis and vascular development. Development. 2008;135(3):493-500.
142. Treutlein B, Lee QY, Camp JG, Mall M, Koh W, Shariati SA, et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature. 2016;534(7607):391-5.
143. Chanda PK, Meng S, Lee J, Leung HE, Chen K, Cooke JP. Nuclear S-Nitrosylation Defines an Optimal Zone for Inducing Pluripotency. Circulation. 2019;140(13):1081-99.
144. Lee J, Sayed N, Hunter A, Au KF, Wong WH, Mocarski ES, et al. Activation of innate immunity is required for efficient nuclear reprogramming. Cell. 2012;151(3):547-58.
145. Sayed N, Wong WT, Ospino F, Meng S, Lee J, Jha A, et al. Transdifferentiation of human fibroblasts to endothelial cells: role of innate immunity. Circulation. 2015;131(3):300-9.
146. Zhou G, Meng S, Li Y, Ghebre YT, Cooke JP. Optimal ROS Signaling Is Critical for Nuclear Reprogramming. Cell Rep. 2016;15(5):919-25.
147. Zhou Y, Wang L, Liu Z, Alimohamadi S, Yin C, Liu J, et al. Comparative Gene Expression Analyses Reveal Distinct Molecular Signatures between Differentially Reprogrammed Cardiomyocytes. Cell Rep. 2017;20(13):3014-24.
148. Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, et al. Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell. 2017;168(3):442-59 e20.
149. Guo H, Hu B, Yan L, Yong J, Wu Y, Gao Y, et al. DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. Cell Res. 2017;27(2):165-83.
150. Caravaca JM, Donahue G, Becker JS, He X, Vinson C, Zaret KS. Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. Genes Dev. 2013;27(3):251-60.
151. Chen J, Zhang Z, Li L, Chen BC, Revyakin A, Hajj B, et al. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell. 2014;156(6):1274-85.

152. Sekiya T, Muthurajan UM, Luger K, Tulin AV, Zaret KS. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes Dev. 2009;23(7):804-9.
153. Aydin B, Kakumanu A, Rossillo M, Moreno-Estelles M, Garipler G, Ringstad N, et al. Proneural factors Ascl1 and Neurog2 contribute to neuronal subtype identities by establishing distinct chromatin landscapes. Nat Neurosci. 2019;22(6):897-908.
154. Schep AN, Buenrostro JD, Denny SK, Schwartz K, Sherlock G, Greenleaf WJ. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res. 2015;25(11):1757-70.
155. Wapinski OL, Lee QY, Chen AC, Li R, Corces MR, Ang CE, et al. Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons. Cell Rep. 2017;20(13):3236-47.
156. Marquez-Vilendrer SB, Thompson K, Lu L, Reisman D. Mechanism of BRG1 silencing in primary cancers. Oncotarget. 2016;7(35):56153-69.
157. Ho L, Jothi R, Ronan JL, Cui K, Zhao K, Crabtree GR. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc Natl Acad Sci U S A. 2009;106(13):5187-91.
158. Durkin ME, Qian X, Popescu NC, Lowy DR. Isolation of Mouse Embryo Fibroblasts. Bio Protoc. 2013;3(18).
159. Caprioli A, Koyano-Nakagawa N, Iacovino M, Shi X, Ferdous A, Harvey RP, et al. Nkx2-5 represses Gata1 gene expression and modulates the cellular fate of cardiac progenitors during embryogenesis. Circulation. 2011;123(15):1633-41.
160. Luger K, Rechsteiner TJ, Richmond TJ. Preparation of nucleosome core particle from recombinant histones. Methods Enzymol. 1999;304:3-19.
161. Tanaka Y, Tawaramoto-Sasanuma M, Kawaguchi S, Ohta T, Yoda K, Kurumizaka H, et al. Expression and purification of recombinant human histones. Methods. 2004;33(1):3-11.
162. Leavitt T, Hu MS, Borrelli MR, Januszyk M, Garcia JT, Ransom RC, et al. Prrx1 Fibroblasts Represent a Pro-fibrotic Lineage in the Mouse Ventral Dermis. Cell Rep. 2020;33(6):108356.
163. Das S, Lin D, Jena S, Shi A, Battina S, Hua DH, et al. Protection of retinal cells from ischemia by a novel gap junction inhibitor. Biochem Biophys Res Commun. 2008;373(4):504-8.
164. Magli A, Baik J, Mills LJ, Kwak IY, Dillon BS, Mondragon Gonzalez R, et al. Time-dependent Pax3-mediated chromatin remodeling and cooperation with Six4 and Tead2 specify the skeletal myogenic lineage in developing mesoderm. PLoS Biol. 2019;17(2):e3000153.
165. Tiscornia G, Singer O, Verma IM. Production and purification of lentiviral vectors. Nat Protoc. 2006;1(1):241-5.
166. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-9.
167. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105-11.
168. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. 2019;8(4):281-91 e9.

169. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019;177(7):1888-902 e21.
170. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods. 2018;15(12):1053-8.
171. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018.
172. Choi KD, Vodyanik MA, Togarrati PP, Suknuntha K, Kumar A, Samarjeet F, et al. Identification of the hemogenic endothelial progenitor and its direct precursor in human pluripotent stem cell differentiation cultures. Cell Rep. 2012;2(3):553-67.
173. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
174. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10(12):1213-8.
175. Magli A, Baik J, Pota P, Cordero CO, Kwak IY, Garry DJ, et al. Pax3 cooperates with Ldb1 to direct local chromosome architecture during myogenic lineage specification. Nat Commun. 2019;10(1):2316.
176. Furlan-Magaril M, Recillas-Targa F. Individual and Sequential Chromatin Immunoprecipitation Protocols. Methods Mol Biol. 2015;1334:205-18.
177. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.
178. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.
179. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep. 2019;9(1):9354.
180. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015;31(14):2382-3.
181. Lay FD, Kelly TK, Jones PA. Nucleosome Occupancy and Methylome Sequencing (NOMe-seq). Methods Mol Biol. 2018;1708:267-84.
182. Statham AL, Taberlay PC, Kelly TK, Jones PA, Clark SJ. Genome-wide nucleosome occupancy and DNA methylation profiling of four human cell lines. Genom Data. 2015;3:94-6.
183. Bailey TL, Machanick P. Inferring direct DNA binding from ChIP-seq. Nucleic Acids Res. 2012;40(17):e128.
184. Lesluyes T, Johnson J, Machanick P, Bailey TL. Differential motif enrichment analysis of paired ChIP-seq experiments. BMC Genomics. 2014;15:752.
185. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011;27(7):1017-8.
186. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(Web Server issue):W202-8.
187. De Val S, Black BL. Transcriptional control of endothelial cell development. Dev Cell. 2009;16(2):180-95.
188. Coffin JD, Poole TJ. Endothelial cell origin and migration in embryonic heart and cranial blood vessel development. Anat Rec. 1991;231(3):383-95.

189. Wilson E, Leszczynska K, Poulter NS, Edelmann F, Salisbury VA, Noy PJ, et al. RhoJ interacts with the GIT-PIX complex and regulates focal adhesion disassembly. J Cell Sci. 2014;127(Pt 14):3039-51.
190. Duman JG, Mulherkar S, Tu YK, J XC, Tolias KF. Mechanisms for spatiotemporal regulation of Rho-GTPase signaling at synapses. Neurosci Lett. 2015;601:4-10.
191. Wojnacki J, Quassollo G, Marzolo MP, Caceres A. Rho GTPases at the crossroad of signaling networks in mammals: impact of Rho-GTPases on microtubule organization and dynamics. Small GTPases. 2014;5:e28430.
192. Shi TT, Li G, Xiao HT. The Role of RhoJ in Endothelial Cell Biology and Tumor Pathology. Biomed Res Int. 2016;2016:6386412.
193. Ridley AJ. Rho GTPase signalling in cell migration. Curr Opin Cell Biol. 2015;36:103-12.
194. Fukushima Y, Okada M, Kataoka H, Hirashima M, Yoshida Y, Mann F, et al. Sema3E-PlexinD1 signaling selectively suppresses disoriented angiogenesis in ischemic retinopathy in mice. J Clin Invest. 2011;121(5):1974-85.
195. Singh BN, Rao KS, Rao Ch M. Ubiquitin-proteasome-mediated degradation and synthesis of MyoD is modulated by alphaB-crystallin, a small heat shock protein, during muscle differentiation. Biochim Biophys Acta. 2010;1803(2):288-99.
196. Kottakis F, Polytarchou C, Foltopoulou P, Sanidas I, Kampranis SC, Tsichlis PN. FGF-2 regulates cell proliferation, migration, and angiogenesis through an NDY1/KDM2B-miR-101-EZH2 pathway. Mol Cell. 2011;43(2):285-98.
197. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2):R14.
198. Koyano-Nakagawa N, Shi X, Rasmussen TL, Das S, Walter CA, Garry DJ. Feedback Mechanisms Regulate Ets Variant 2 (Etv2) Gene Expression and Hematoendothelial Lineages. J Biol Chem. 2015;290(47):28107-19.
199. Kim DH, Cho S, Wirtz D. Tight coupling between nucleus and cell migration through the perinuclear actin cap. J Cell Sci. 2014;127(Pt 11):2528-41.
200. Verheyden JM, Sun X. An Fgf/Gremlin inhibitory feedback loop triggers termination of limb bud outgrowth. Nature. 2008;454(7204):638-41.
201. Singh BN, Doyle MJ, Weaver CV, Koyano-Nakagawa N, Garry DJ. Hedgehog and Wnt coordinate signaling in myogenic progenitors and regulate limb regeneration. Dev Biol. 2012;371(1):23-34.
202. Zhao L, Borikova AL, Ben-Yair R, Guner-Ataman B, MacRae CA, Lee RT, et al. Notch signaling regulates cardiomyocyte proliferation during zebrafish heart regeneration. Proc Natl Acad Sci U S A. 2014;111(4):1403-8.
203. Smith A, Avaron F, Guay D, Padhi BK, Akimenko MA. Inhibition of BMP signaling during zebrafish fin regeneration disrupts fin growth and scleroblasts differentiation and function. Dev Biol. 2006;299(2):438-54.
204. Aguirre A, Montserrat N, Zacchigna S, Nivet E, Hishida T, Krause MN, et al. In vivo activation of a conserved microRNA program induces mammalian heart regeneration. Cell Stem Cell. 2014;15(5):589-604.
205. Schmidt A, Brixius K, Bloch W. Endothelial precursor cell migration during vasculogenesis. Circ Res. 2007;101(2):125-36.

206. Abedin MJ, Nguyen A, Jiang N, Perry CE, Shelton JM, Watson DK, et al. Fli1 acts downstream of Etv2 to govern cell survival and vascular homeostasis via positive autoregulation. Circ Res. 2014;114(11):1690-9.
207. Kanki Y, Nakaki R, Shimamura T, Matsunaga T, Yamamizu K, Katayama S, et al. Dynamically and epigenetically coordinated GATA/ETS/SOX transcription factor expression is indispensable for endothelial cell differentiation. Nucleic Acids Res. 2017;45(8):4344-58.
208. Kabir AU, Lee TJ, Pan H, Berry JC, Krchma K, Wu J, et al. Requisite endothelial reactivation and effective siRNA nanoparticle targeting of Etv2/Er71 in tumor angiogenesis. JCI Insight. 2018;3(8).
209. Kim C, Yang H, Fukushima Y, Saw PE, Lee J, Park JS, et al. Vascular RhoJ is an effective and selective target for tumor angiogenesis and vascular disruption. Cancer Cell. 2014;25(1):102-17.
210. Zhao C, Gomez GA, Zhao Y, Yang Y, Cao D, Lu J, et al. ETV2 mediates endothelial transdifferentiation of glioblastoma. Signal Transduct Target Ther. 2018;3:4.
211. Franz CM, Jones GE, Ridley AJ. Cell migration in development and disease. Dev Cell. 2002;2(2):153-8.
212. Hipfner DR, Cohen SM. Connecting proliferation and apoptosis in development and disease. Nat Rev Mol Cell Biol. 2004;5(10):805-15.
213. Shaw TJ, Martin P. Wound repair: a showcase for cell plasticity and migration. Curr Opin Cell Biol. 2016;42:29-37.
214. Rankin EB, Giaccia AJ. The role of hypoxia-inducible factors in tumorigenesis. Cell Death Differ. 2008;15(4):678-85.
215. Chen J, Zhou Q, Feng J, Zheng W, Du J, Meng X, et al. Activation of AMPK promotes thyroid cancer cell migration through its interaction with PKM2 and beta-catenin. Life Sci. 2019;239:116877.
216. Fu D, Lv X, Hua G, He C, Dong J, Lele SM, et al. YAP regulates cell proliferation, migration, and steroidogenesis in adult granulosa cell tumors. Endocr Relat Cancer. 2014;21(2):297-310.
217. Fu M, Lui VC, Sham MH, Pachnis V, Tam PK. Sonic hedgehog regulates the proliferation, differentiation, and migration of enteric neural crest cells in gut. J Cell Biol. 2004;166(5):673-84.
218. Nicoli S, Knyphausen CP, Zhu LJ, Lakshmanan A, Lawson ND. miR-221 is required for endothelial tip cell behaviors during vascular development. Dev Cell. 2012;22(2):418-29.
219. Li F, Chong ZZ, Maiese K. Winding through the WNT pathway during cellular development and demise. Histol Histopathol. 2006;21(1):103-24.
220. Clevers H. Wnt/beta-catenin signaling in development and disease. Cell. 2006;127(3):469-80.
221. Davis LA, Zur Nieden NI. Mesodermal fate decisions of a stem cell: the Wnt switch. Cellular and molecular life sciences : CMLS. 2008;65(17):2658-74.
222. Murry CE, Keller G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell. 2008;132(4):661-80.
223. Ueno S, Weidinger G, Osugi T, Kohn AD, Golob JL, Pabon L, et al. Biphasic role for Wnt/beta-catenin signaling in cardiac specification in zebrafish and embryonic stem

cells. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(23):9685-90.
224. Lin L, Cui L, Zhou W, Dufort D, Zhang X, Cai CL, et al. Beta-catenin directly regulates Islet1 expression in cardiovascular progenitors and is required for multiple aspects of cardiogenesis. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(22):9313-8.
225. Ji Z, Donaldson IJ, Liu J, Hayes A, Zeef LA, Sharrocks AD. The forkhead transcription factor FOXK2 promotes AP-1-mediated transcriptional regulation. Mol Cell Biol. 2012;32(2):385-98.
226. Shi X, Garry DJ. Myogenic regulatory factors transactivate the Tceal7 gene and modulate muscle differentiation. The Biochemical journal. 2010;428(2):213-21.
227. Kattman SJ, Huber TL, Keller GM. Multipotent flk-1+ cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages. Developmental cell. 2006;11(5):723-32.
228. Masino AM, Gallardo TD, Wilcox CA, Olson EN, Williams RS, Garry DJ. Transcriptional regulation of cardiac progenitor cell populations. Circ Res. 2004;95(4):389-97.
229. Wu SM, Fujiwara Y, Cibulsky SM, Clapham DE, Lien CL, Schultheiss TM, et al. Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart. Cell. 2006;127(6):1137-50.
230. Singh BN, Sierra-Pagan JE, Gong W, Das S, Theisen JWM, Skie E, et al. ETV2 (Ets Variant Transcription Factor 2)-Rhoj Cascade Regulates Endothelial Progenitor Cell Migration During Embryogenesis. Arteriosclerosis, Thrombosis, and Vascular Biology. 2020;40(12):2875-90.
231. Hota SK, Rao KS, Blair AP, Khalilimeybodi A, Hu KM, Thomas R, et al. Brahma safeguards canalization of cardiac mesoderm differentiation. Nature. 2022;602(7895):129-34.
232. Yang L, Chen S, Zhao Q, Pan C, Peng L, Han Y, et al. Histone deacetylase 3 contributes to the antiviral innate immunity of macrophages by interacting with FOXK1 to regulate STAT1/2 transcription. Cell Rep. 2022;38(4):110302.
233. Ni Y, Liu B, Wu X, Liu J, Ba R, Zhao C. FOXG1 Directly Suppresses Wnt5a During the Development of the Hippocampus. Neurosci Bull. 2021;37(3):298-310.
234. Wang W, Li X, Lee M, Jun S, Aziz KE, Feng L, et al. FOXKs promote Wnt/beta-catenin signaling by translocating DVL into the nucleus. Developmental cell. 2015;32(6):707-18.
235. Lavery DL, Martin J, Turnbull YD, Hoppler S. Wnt6 signaling regulates heart muscle development during organogenesis. Developmental biology. 2008;323(2):177-88.
236. Tao F, Soffers J, Hu D, Chen S, Gao X, Zhang Y, et al. beta-Catenin and Associated Proteins Regulate Lineage Differentiation in Ground State Mouse Embryonic Stem Cells. Stem Cell Reports. 2020;15(3):662-76.
237. Zhu H. Forkhead box transcription factors in embryonic heart development and congenital heart disease. Life Sci. 2016;144:194-201.
238. Hota SK, Bruneau BG. ATP-dependent chromatin remodeling during mammalian development. Development (Cambridge, England). 2016;143(16):2882-97.

239. Iacovino M, Bosnakovski D, Fey H, Rux D, Bajwa G, Mahen E, et al. Inducible cassette exchange: a rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. Stem cells (Dayton, Ohio). 2011;29(10):1580-8.
240. Bryja V, Bonilla S, Arenas E. Derivation of mouse embryonic stem cells. Nat Protoc. 2006;1(4):2082-7.
241. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34(5):525-7.
242. Sukonina V, Ma H, Zhang W, Bartesaghi S, Subhash S, Heglind M, et al. FOXK1 and FOXK2 regulate aerobic glycolysis. Nature. 2019;566(7743):279-83.
243. Malakar AK, Choudhury D, Halder B, Paul P, Uddin A, Chakraborty S. A review on coronary artery disease, its risk factors, and therapeutics. J Cell Physiol. 2019;234(10):16812-23.
244. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. Nat Rev Genet. 2016;17(8):487-500.
245. Lee S, Lee DH, Park BW, Kim R, Hoang AD, Woo SK, et al. In vivo transduction of ETV2 improves cardiac function and induces vascular regeneration following myocardial infarction. Experimental & molecular medicine. 2019;51(2):13.
246. Magadum A, Kurian AA, Chepurko E, Sassi Y, Hajjar RJ, Zangi L. Specific Modified mRNA Translation System. Circulation. 2020;142(25):2485-8.
247. Hadas Y, Katz MG, Bridges CR, Zangi L. Modified mRNA as a therapeutic tool to induce cardiac regeneration in ischemic heart disease. Wiley interdisciplinary reviews Systems biology and medicine. 2017;9(1).
248. Sultana N, Magadum A, Hadas Y, Kondrat J, Singh N, Youssef E, et al. Optimizing Cardiac Delivery of Modified mRNA. Molecular therapy : the journal of the American Society of Gene Therapy. 2017;25(6):1306-15.
249. Zangi L, Lui KO, von Gise A, Ma Q, Ebina W, Ptaszek LM, et al. Modified mRNA directs the fate of heart progenitor cells and induces vascular regeneration after myocardial infarction. Nature biotechnology. 2013;31(10):898-907.
250. Rurik JG, Tombacz I, Yadegari A, Mendez Fernandez PO, Shewale SV, Li L, et al. CAR T cells produced in vivo to treat cardiac injury. Science. 2022;375(6576):91-6.
251. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. N Engl J Med. 2021;384(5):403-16.
252. Brok-Volchanskaya VS, Bennin DA, Suknuntha K, Klemm LC, Huttenlocher A, Slukvin I. Effective and Rapid Generation of Functional Neutrophils from Induced Pluripotent Stem Cells Using ETV2-Modified mRNA. Stem Cell Reports. 2019;13(6):1099-110.
253. Magadum A, Singh N, Kurian AA, Munir I, Mehmood T, Brown K, et al. Pkm2 Regulates Cardiomyocyte Cell Cycle and Promotes Cardiac Regeneration. Circulation. 2020;141(15):1249-65.
254. Mak IW, Evaniew N, Ghert M. Lost in translation: animal models and clinical trials in cancer treatment. American journal of translational research. 2014;6(2):114-8.
255. Perleberg C, Kind A, Schnieke A. Genetically engineered pigs as models for human disease. Disease models & mechanisms. 2018;11(1).
256. Swindle MM, Makin A, Herron AJ, Clubb FJ, Jr., Frazier KS. Swine as models in biomedical research and toxicology testing. Veterinary pathology. 2012;49(2):344-56.

257. Liu S, Li K, Wagner Florencio L, Tang L, Heallen TR, Leach JP, et al. Gene therapy knockdown of Hippo signaling induces cardiomyocyte renewal in pigs after myocardial infarction. Sci Transl Med. 2021;13(600).
258. Zhao M, Zhang E, Wei Y, Zhou Y, Walcott GP, Zhang J. Apical Resection Prolongs the Cell Cycle Activity and Promotes Myocardial Regeneration After Left Ventricular Injury in Neonatal Pig. Circulation. 2020;142(9):913-6.
259. Zhu W, Zhang E, Zhao M, Chong Z, Fan C, Tang Y, et al. Regenerative Potential of Neonatal Porcine Hearts. Circulation. 2018;138(24):2809-16.