Summer 8-15-2020

The role of CDX4 during patterning of definitive hemogenic mesoderm

John Philip Creamer
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Molecular Cell Biology

Dissertation Examination Committee:
Kyunghee Choi, Chair
Christopher Sturgeon, Co-Chair
Grant Challen
Kory Lavine
Samantha Morris

The Role of *CDX4* During Patterning of Definitive Hemogenic Mesoderm

by
J. Philip Creamer IV

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2020
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

Christopher Sturgeon was instrumental in assisting in this work as a mentor by providing direction, advice, performing experiments, and assisting with writing. Many past and present lab members assisted with this work, by giving advice, helping with/performing experiments, and feedback during presentations including Stephanie Luff, Carissa Dege, Kendra Sturgeon, Qihao Ren, Emily Lewis, Jolie Ho, and Tom Tenhula, From the laboratory of Todd Druley, Mark Valentine performed the analysis of bulk RNAseq found in the Chapter 2, while Samantha Morris assisted in processing the single cell transcriptomic samples found in Chapters 3 and 4 as well as in giving advice on analysis techniques. My committee members also provided guidance and mentorship that improved this work. The GTAC core facility provided next generation sequencing resources, the Pathology and Immunology FACS core with flow cytometry and sorting access, and the Center for Regenerative medicine provide computation cluster access.

This work was supported by a T32 training grant from the National Institutes of Health, National Heart, Lung, and Blood Institute (HL007088-41), an American Society of Hematology Scholar awarded to Christopher Sturgeon, and by startup funds given by Washington University.

.

J. Philip Creamer

Washington University in St. Louis

August 2020

Abstract of the Dissertation

The role of *CDX4* during patterning of definitive hemogenic mesoderm

by

J. Philip Creamer

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Cell Biology

Washington University in St. Louis, 2020

Kyunghee Choi, Chair

Christopher Sturgeon, Co-Chair

The current standard of treatment for a variety of hematopoietic malignancies and genetic disorders is allogenic bone marrow transplantation, where donor hematopoietic stem cells (HSCs) engraft within the host and give rise to all of them hematopoietic lineages necessary for homeostasis. In many cases, finding a compatible human leukocyte antigen (HLA) matching donor is not possible, due to the large amount of genetic variation at those loci, but with the advent of induced pluripotent stem cells (iPSCs), unlimited sources of patient matched cells can be derived. Hematopoietic differentiations of human pluripotent stem cells (hPSCs) have been shown to recapitulate the early development of the embryo, producing known progenitors such as the hemangioblast, but current efforts have been unable to produce an HSC without the use of transgene expression. This is partly due to the complex nature of human developmental hematopoiesis, which is known to contain multiple hematopoietic programs of varying potential. These programs or 'waves' can be generally fit into two categories, the first being

extraembryonic hematopoiesis that occurs within the yolk sac earliest in development and produces mainly transient, primitive progenitors that support the developing embryo. The second category is intraembryonic hematopoiesis, which occurs within the embryo proper, producing more mature progenitors, including the HSC, which arises from hemogenic endothelium (HE) in the dorsal aorta.

Our lab has developed an hPSCs differentiation model that can identify and specify WNT independent, extraembryonic, primitive hematopoietic progenitors, and WNT dependent, intraembryonic, definitive hematopoietic progenitors through stage specific modulation of WNT signaling during the mesodermal stage of differentiation. I have shown that mesodermal expression of *CDX4*, a caudal-like homeobox transcription factor, is an important regulator of the specification of definitive HE, by utilizing a doxycycline inducible *CDX4* hPSC line and a *CDX4*$^{y/-}$ KO hPSC line. In this work, I have demonstrated that *CDX4* acts to induce canonical gene targets, such as medial *HOXA* genes, in different subsets of hemogenic mesoderm, likely impacting the specification of definitive HE. Surprisingly, *TBX20*, a cardiomyocyte transcription factor, was found to be a negatively regulated *CDX4* target, suggesting that *CDX4* might also play a role in regulating cardiac specification. I performed cardiomyocyte differentiations and demonstrated that *CDX4*$^{y/-}$ KO lead to a significant expansion of cardiomyocytes over WT and that mesodermal *CDX4* expression abrogated this expansion. Additionally, single cell transcriptomics revealed that *CDX4*+ mesoderm also expresses *CD1D*, and when functionally characterized, definitive CD1D+ hemogenic mesoderm contained nearly all lymphoid, erythroid, and myeloid potential.

These results will impact the field of hematopoietic differentiation, having extensively characterized the role of *CDX4* in hemogenic mesoderm, it's gene targets, and correlative

markers. Further studies will be able to leverage *CDX4* and its downstream targets, such as *HOXA* genes, to improve hematopoietic differentiations, and perhaps improve cardiomyocyte differentiations by reducing the expression of *CDX4*.

Chapter 1: Rationale, introduction to developmental hematopoiesis and homeobox genes

1.1 Hematopoietic stem cells

Hematopoietic stem cells (HSCs) have been rigorously studied for decades for use in life-saving transplantation, as well as to understand hematopoiesis, and how they contribute to leukemic transformation in various hematopoietic malignancies. Multipotent progenitors were first identified in the mouse bone marrow as cells capable of producing clonogenic colonies of multiple hematopoietic cell types in the spleen of irradiated mice[1-4]. Further work lead to the characterization of these bone marrow progenitors via flow cytometry and isolation by flow cytometry assisted cell sorting (FACS), by removing all cells positive for lineage specific markers for B cells (B220), granulocytes (Gr-1), monocytes (Mac-1) and T cells (CD4 and CD8), and positive for Thy1 (CD90)[5], Sca1[6], and c-Kit[7]. This combination of markers is termed the Lineage-Sca1+c-Kit+ (LSK) fraction in mice, and a subset of these multipotent progenitors can be termed as HSCs, being capable of long-term self-renewal, as well as the production of multiple distinct lineages necessary for maintenance of circulating blood cells from only a single cell[8]. Later work in mice showed that an HSC containing population can be further purified from more lineage restricted progenitors through the use of CD150, CD48, CD229, and CD244 SLAM family markers[9,10]. While nascent human HSCs share the expression of markers like CD34, VE-cadherin, CD45, C-KIT, and THY-1[11], the SLAM marker system is specific to mouse HSCs[12].

The ability of these cells to reconstitute the circulating blood system of an adult has been exploited as a method of treatment for hematopoietic malignancies[13,14] as well as anemias[15], and blood related genetic disorders[16,17]. Though transplantation can offer treatment and even cures for many of these patients, these life-saving procedures are hampered by the necessity for high or complete matching of the human leukocyte antigen (HLA) loci[18]. This is due to the high level of variability found in humans at the HLA locus[19], making the most likely donor candidates siblings of the patient; around 30% of patients will have a sibling with the correct HLA genotype to be an effective donor[20], though autologous transplant is preferred in some cases for treatment of lymphomas[21]. Because of these unmet deficits in properly matched donors for treatments of hematological disorders, alternative approaches are needed to close the gap in available treatment options.

1.2 Human pluripotent stem cells (hPSCs)

Since the discovery of mouse pluripotent stem cells capable of unlimited expansion *in vitro,* differentiation into all three germs layers[22], as well the ability to produce a viable embryo[23], researchers have been continuously investigating ways to make clinically relevant cell types. With the discovery of human pluripotent stem cells (hPSCs) derived first from human embryos[24] and later through the genetic reprogramming of somatic cells to create induced pluripotent stem cells (iPSCs)[25], it is now possible to produce patient specific iPSC lines with the potential to differentiate into any replacement cell type needed. Deriving iPSC lines can now be done without potentially risky gene editing[26] and would be syngeneic to patients, preventing immune rejection in transplanted cell types. Additionally, hPSCs can be used to study early human development which is very difficult to study for ethic and technical reasons, as they have been shown to faithfully recapitulate many aspects of both mouse and human hematopoietic

development[27-29]. Despite the lack of a reliably produced, long-term engraftable progenitor, hematopoietic differentiation protocols have been steadily improved. Although there have been studies that have been able to produce an engraftable HSC *in vitro*[30,31], they have relied on the usage of transgene over-expression, which would not be suitable for patient treatments and likely not faithfully represent the true developmental trajectory. Knowing the key steps in hematopoietic differentiation during embryological development will shed light on the processes that are necessary to be able to produce HSCs through *in vitro* differentiation in a transgene free manner. The modeling of hematopoietic differentiation in a dish using hPSCs will also help inform our understanding of human development where *in vivo* study is not possible.

1.3 Hematopoietic Development

Efforts to recapitulate early human hematopoietic development *in vitro* have been complicated by the existence of several distinct programs of hematopoiesis that are spatiotemporally segregated and given rise to lineages of differing cellular potential. There are multiple, separate waves of hematopoiesis in humans during early embryogenesis[32-34], being either extraembryonic or intraembryonic in origin. Early anatomical studies of the blood islands in the yolk sacs of avian embryos noted the possibility of a common progenitor for blood and vasculature term the hemangioblast[35,36]. Direct evidence of hemangioblast was seen until much later through *in vitro* mouse pluripotent stem cell differentiations that yielded colonies of mixed blood and endothelium in semi-solid media, termed blast colony-forming cells (BL-CFCs)[28], suggesting the presence of a cell capable of both lineages. Later *in vivo* mouse embryo studies also demonstrated the presence of a BL-CFC capable progenitor in the primitive streak, expressing KDR/Flk1 and *brachyury/T* [37]. The hemangioblast is considered the direct progenitor of primitive hematopoiesis, as BL-CFC potential is found before the formation of primitive

hematopoietic progenitors both *in vitro* and *in vivo,* as well as lineage tracing studies demonstrating primitive hematopoietic cells were derived from KDR+ cells[38]. It was also found in mouse PSCs and embryos that *ER71/Etv2* acts downstream of KDR/Flk1, as the loss of *Etv2* lead to pan hematopoietic and endothelium defects, providing additional evidence of common regulation of both lineages. *Etv2* activation was also shown to be necessary in a brief window in mesoderm formation before or during the expression of KDR/Flk1 and loss of *Etv2* could be rescued by *Scl* expression[39]. *Scl* expression in KDR+ progenitors has been shown to mark both blood and endothelium potential, further strengthening *Scl* as a regulator of the hemangioblast[40]. While these studies suggest a mesodermal progenitor that is dually potent for hematopoiesis and vasculature, there is some evidence against a common origin of endothelium and blood *in vivo*. Lineages studies utilizing injections of mouse blastocysts with mixtures mouse PSCs with recombinase activated fluorescent genes revealed that individual blood islands were nearly always mixed in origin[41], suggested that the hemangioblast could be a mixed population. Additionally, there is evidence of the production of endothelium from the hemangioblast, that then gives rise to hematopoietic cells in the mouse yolk sac[42], further questioning the presence of a truly, dually potent hemangioblast.

In humans, primitive hematopoietic progenitors emerge as early as day 19 of conception in humans[43]and give rise to primitive erythrocytes mainly expressing epsilon globins[44], megakaryocytes[45], macrophages[46], and microglia[47]. Some of the cell types arising from this program, such as the primitive erythrocytes and megakaryocytes are short lived and are eventually replaced[48], though microglia and macrophages can persist throughout life. This directly contrasts with the intraembryonic, definitive hematopoietic program that gives rise to an HSC capable of long-term erythroid-myeloid-lymphoid-megakaryocytic engraftment, first in the

aorta gonad mesonephros (AGM) region at ~ 5 weeks of development[33,49,50] and later in other hematopoietic tissues. There are, however, several hematopoietic programs that have been identified that produce lineages not found in the primitive hematopoietic program but are independent of the HSC. One of these is an extraembryonic lineages that has been extensively studied in the mouse termed the erytho-myeloid progenitor (EMP)[51], which can give rise to more mature erythrocytes expressing adult globins, granulocytes, and macrophages[52]. Additionally, another progenitor termed the lymphomyeloid primed progenitor (LMPP) has also been found in the mouse yolk sac[53], giving rise to myeloid and lymphoid progenitors, which had been thought to be restricted to the HSC. The production of mature definitive erythrocytes with adult β-globin expression[54] and lymphoid progenitors which were previously thought to be HSC dependent, but yolk sac in origin, complicates these definitions. Whether to term these progenitors as 'definitive' or 'primitive' is a difficult question, as well as whether they are present in the human. Recent work has shown that the mouse EMP is also capable of producing natural killer cells, another canonically believed lymphoid cell type, that also had a correlate EMP-like population in hPSC differentiations[34]. There is also evidence for progenitors capable of adult β-globin expression in the early human fetal liver at 6 weeks, just as the first HSC's are emerging and presumably have not had enough time to fully differentiate[55]. This work doesn't directly prove there is a human EMP, though it provides supportive evidence for its existence. All of these different progenitor populations are summarized in **Figure 1.1**, adapted from Ditadi *et al* (2017)[32].

1.4 Hemogenic Endothelium

Hematopoiesis is largely understood to arise from hemogenic endothelium (HE), which undergoes a process known as the endothelial to hematopoietic transition (EHT)[56-59]. HE is

generally understood to express endothelial markers like CD31/34, but also *RUNX1*, a hematopoietic transcription factor[60], which primes the endothelium for the EHT[59]. HE gives rise to the nascent HSC's which bud from the dorsal floor of the aorta in clusters and enter circulation and eventually populate the fetal liver[61], where they mature and expand; these cells will then give rise to all circulating erythroid, myeloid, lymphoid, and megakaryocytic lineages necessary to sustain the developing fetus, and eventually colonize the bone marrow[62]. Hemogenic endothelium is known to not be exclusive to the dorsa aorta; though the hemangioblast also gives rise to hematopoietic progenitors and endothelium, there is disagreement in the literature on the origins of HE in relation to the hemangioblast, with the hemangioblast potentially giving rise to HE[42]. Since not all HE gives rise to the HSC, it cannot be used as a term exclusive to definitive hematopoiesis. Complicating matters further, there has also been reports of HE capable of producing blood progenitors of varying potential in the human liver and fetal bone marrow[63], as well HSCs from the mouse placenta[64].

What leads certain HE to give rise to an engrafting HSC, multipotent progenitors, or more lineage restricted progenitors is not well known, though genetic studies in *MYB*[52] and *RUNX1*[65] KO mice suggests that EMP/HSC-producing HE relies on these transcription factors, as only primitive hematopoiesis is completely unaffected in these models. Additionally, notch signaling is thought to be required for intraembryonic-HSC producing HE, whereas it appears to be dispensable for extraembryonic (EMP and primitive) lineages[66-68]. Notch signaling is important for the formation of endothelium in dorsal aorta where HE will ultimately arise[69], but notch suppression is actually required for the emergence of the HSC from HE[70]. Recent work using single cell transcriptomics of early human embryos computationally identified two different types of intraembryonic HE, one that appeared to be HSC-primed, and another earlier HE

population that lacked arterial markers[71]. Because HE that gives rise to the HSC is found in the dorsal aorta, it has been assumed that HE arises from arterial endothelium, though in the literature there is evidence both supporting an independent origin for HE[72], and it originating from arterial endothelium[73].

1.5 Hematopoietic differentiations from hPSCs

Current methods of differentiating hPSCs into hematopoietic progenitors are built on decades of work first pioneered in mouse embryonic stem cells. Early techniques relied on culture in serum containing media with free floating aggregates known as embryoid bodies, which induced the production of many mesodermal lineages including endothelium, blood, and cardiomyocytes[74]. Further work characterized the hematopoietic lineages found in these early mouse differentiation studies, demonstrating the development of primitive erythrocytes and myeloid lineages, that could be increased with treatment of hematopoietic cytokines such as EPO, IL-3, and M-CSF/GM-CSF[75]. These differentiations follow *in vivo* hematopoietic development, where primitive hematopoietic progenitors are produced first, followed by the later emergency of more mature definitive erythroid progenitors[76,77]. After the isolation of human pluripotent stem cells, serum based differentiation approaches were also able to were also of giving rise to the hemangioblast in BL-CFC assays[78], much like mouse PSC differentiations[28], indicative of the primitive program. Later advances in the human system allowed for the development of serum free media that utilized addition of cytokines including bFGF, BMP4, and VEGF[79-81]. Usage of serum free media allows for dissection of the precise signaling requirements needed in order to specify hematopoietic progenitors, as the numerous factors contained in serum made understanding these requirements difficult.

Although it was possible to generate lymphoid cells from mouse embryonic stem cell differentiations using serum containing media and hypoxic condtions[82], hPSCs require a coculture system with OP9 mouse stroma cells expressing notch ligands to specify T cells[83]. Despite later advances in cytokine based serum-free conditions that utilized TGFβ treatment to specify an endothelial populations capable of definitive erythrocytes and myeloid cells, the production of lymphoid lineages still requires the co-culture with OP9 stroma with notch ligands[29]. Nonetheless, just as in development, this work demonstrated that hPSC differentiations can produce both primitive/EMP erytho-myeloid potent progenitors and definitive erythro-myeloid-lymphoid progenitors if the correct signaling modulations are made in a stage specific manner.

While this serum free approach in hPSCs gives rise to primitive/EMP CD34+CD43+CD45+ hematopoietic progenitors and CD34+CD31+VECAD+CD43-CD45- definitive HE at certain timepoints[29], it was not possible to purely produce only one or the other. WNT signaling had been implicated in hematopoietic differentiations previously[84], but it was found that stage specific addition of either WNT antagonists/activin A or WNT agonists during mesoderm formation lead to the specification of a primitive/EMP hemogenic mesoderm or definitive mesoderm respectively[85]. The primitive hemogenic mesoderm was marked by KDR and glycophorin A (CD235a), giving rise to the hemangioblast and erythro-myeloid potential, while the definitive hemogenic mesoderm was KDR+ and CD235a-, giving rise to lymphoid, erythroid, and myeloid potential. This work allowed for precise control of the hematopoietic lineages produced in culture, differences in signaling requirements, and identification of distinct and easily segregable mesodermal origins for different waves of hematopoiesis. Interestingly, mouse embryonic stem cell differentiations also are influenced by agonizing WNT signaling, but this

instead promotes the primitive program and represents a key difference between the two systems[86].

Further advances in hPSC differentiation also identified a unique population that contained HE that was marked by CD34+CD43-CD184-CD73-, that on a single cell level was capable of lymphoid, erythroid, and myeloid potential and segregated from arterial and venous endothelium[87]. These culture conditions are summarized in **Figure 1.2** adapted from Sturgeon *et al.* (2014)[72]. With these improvements in technique, it is now possible to precisely interrogate the genetic regulation behind the specification of hemogenic mesoderm of varying potential.

1.6 Common mesodermal origins for HE and cardiac fates

In the gastrulating embryo, multiple lineages and tissues will be produced from the nascent mesoderm, including the heart, blood, somites, and endothelium. There are many types of mesoderm formed, including axial mesoderm, paraxial mesoderm, intermediate mesoderm, and lateral plate mesoderm[88]; lateral plate mesoderm gives rise several of these lineages, include cardiac cells, hematopoietic cells, and endothelium. Because of these common origins, these lineages share signaling requirements, such as WNT, Nodal, and BMP[89-91], for their specification and development. Cardiac and hematoendothelial progenitors also share expression of Mesp1, a transcription factor necessary for early mesoderm formation[92], as both the heart and HSCs were marked as Mesp1+ in lineage tracing studies[93]. It is not clear if there is a single cell type that is plastic that gives rise to both cardiac and blood/endothelium or if there is a population of mesoderm containing a mixture of cardiac and hematoendothelial progenitors. Expression of Brachury (*T*) and Flk1 (KDR) can segregate these lineages in *in vitro* PSC differentiations, suggesting independent origins[94], but transgenic activation of notch signaling was able to respecify hemangioblast containing populations to cardiomyocytes[95]. Additionally,

KO of *Scl*, a critical pan-hematopoietic transcription factor, lead to the ectopic formation of cardiomyocytes in the yolk sacs of mouse embryos from CD31+ endothelium[96]. *Scl* has already been identified as a regulator of the hemangioblast, the progenitor of endothelium and primitive hematopoiesis[40] as well as *Etv2*, which activates *Scl*[39]. These studies suggest that within the developing lateral plate mesoderm, there is a critical balance of a variety of factors that can swing the balance between hematoendothelial and cardiac lineages. From the literature available, it is unclear if there is a single cell type that gives rise to hematopoietic, endothelium, and cardiac lineages, or whether current known mesodermal markers simply represent a mixed population of already specified mesodermal subtypes.

1.7 Homeobox genes and hematopoiesis

Homeobox genes were first discovered in *Drosophila*, while trying to clone the genetic locus responsible for antennapedia mutants, where legs would grow from the head instead of antennae[97]. The *antennapedia* gene contains a helix-turn-helix protein structure, which can interact directly with DNA at a specific sequence, named the "homeobox", of which many homologous and paralogous genes have since been found in vertebrates and invertebrates[98,99]. These include the HOX genes, which in mice and humans are linearly organized into four clusters (A-D) with 39 genes[100] that play important roles in axial patterning, limb-specification, and early development. There are also the ParaHox genes, named for their similarity to HOX gene clusters and probability as a paralog when discovered as a similar cluster in lancets[101]. The ParaHox genes found in mice and humans are the GSX, PDX, and CDX gene families[102], which have been broadly categorized as influencing brain/head, mid sectional, and caudal development respectively[103]. Generally, CDX genes have been found to activate during gastrulation in the early vertebrate embryo and work to activate the expression of HOX genes in a concerted

pattern of expression anterior to posterior necessary for proper body plan formation[104,105]. In humans and mice, there are three CDX genes *CDX1, CDX2,* and *CDX4* which have been shown to have somewhat compensatory functions[106], extending as far complete complementation of *Cdx1* with an in-frame *Cdx2* sequence in mice, leading to a normal phenotype[107]. CDX and HOX genes have also been shown to be regulated by variety of different signaling pathways in a variety of different contexts, including fibroblast growth factor (FGF)[108], WNT[109], retinoic acid[110], and transforming growth factor beta (TGF-β)[111]. Among their various roles in development, HOX and CDX genes have been shown to be important for hematopoiesis, in particular HSC-dependent definitive hematopoiesis. In the zebrafish, *cdx4* mutants had early defects in all hematopoiesis that could be partially rescued with the expression of certain zebrafish *hoxa* and *hoxb* genes[112], suggesting that *cdx4* plays an important role in the specification of hematopoietic progenitors. Further in mouse embryos and embryonic stem cells seems to suggest that loss of *Cdx1/2/4* all resulted in defects in specification of primitive/definitive hematopoietic progenitors coupled with a suppression in HoxA/B gene expression[113,114], but with limited effects on HSCs[115]. HOXA genes have been specifically implicated in the regulation of the definitive hematopoietic program in both mouse and hPSC differentiation models. In particular, the medial HOXA genes, *HOXA5,7,9*, have been shown to be critical for human fetal liver HSC function, as well as hematopoietic progenitors differentiated from hPSCs[116]. In the mouse and murine embryonic stem cells, the formation of definitive HE has also been shown to be regulated by *HoxA3*, forming a regulatory relationship with *Runx1*, a critical hematopoietic transcription factor necessary for HSC formation[117]. This is mirrored in transcriptional profiling of FACS purified human AGM that displays a distinct *HOXA+* gene signature, which is comparatively found in definitive hematopoietic progenitors differentiated

from hPSCs[118]. Within transgene dependent hPSC hematopoietic differentiations, *HOX5/7/9* are also necessary for forming expanded lineage engraftable progenitors from less mature progenitors[30,119], suggesting that *HOXA* expression is likely required for the specification of an HSC from HE. Collected together, these results suggest that *CDX*/*HOXA* genes are critical for not only the HSC, but also for the proper patterning of HE that gives rise to the HSC *in vivo* and *in vitro*.

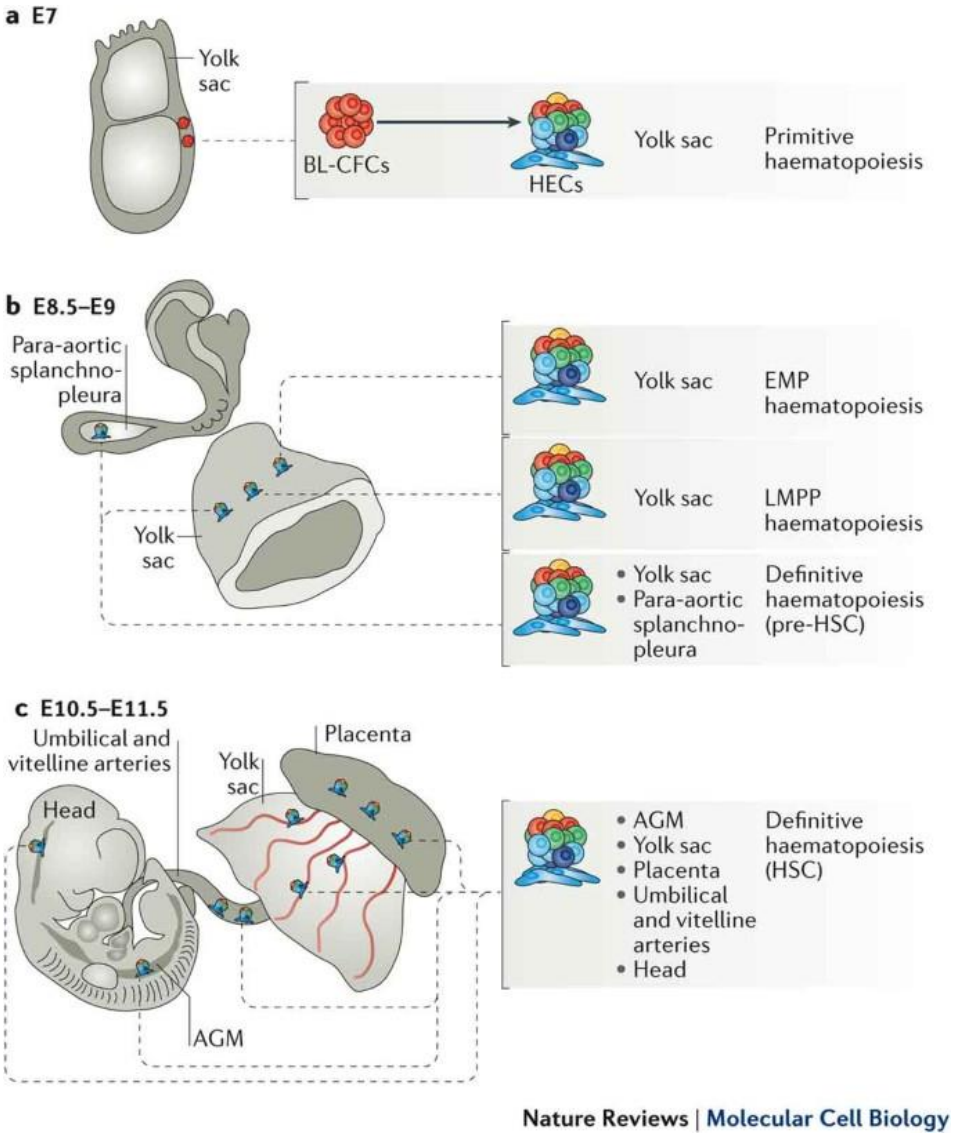


Figure 1.1: Identified hematopoietic programs in mouse development

Adapted from Ditadi *et al.* 2016[32], demonstrating the timing and development of hematopoietic progenitors of varying potential from different physical locations in the embryo.

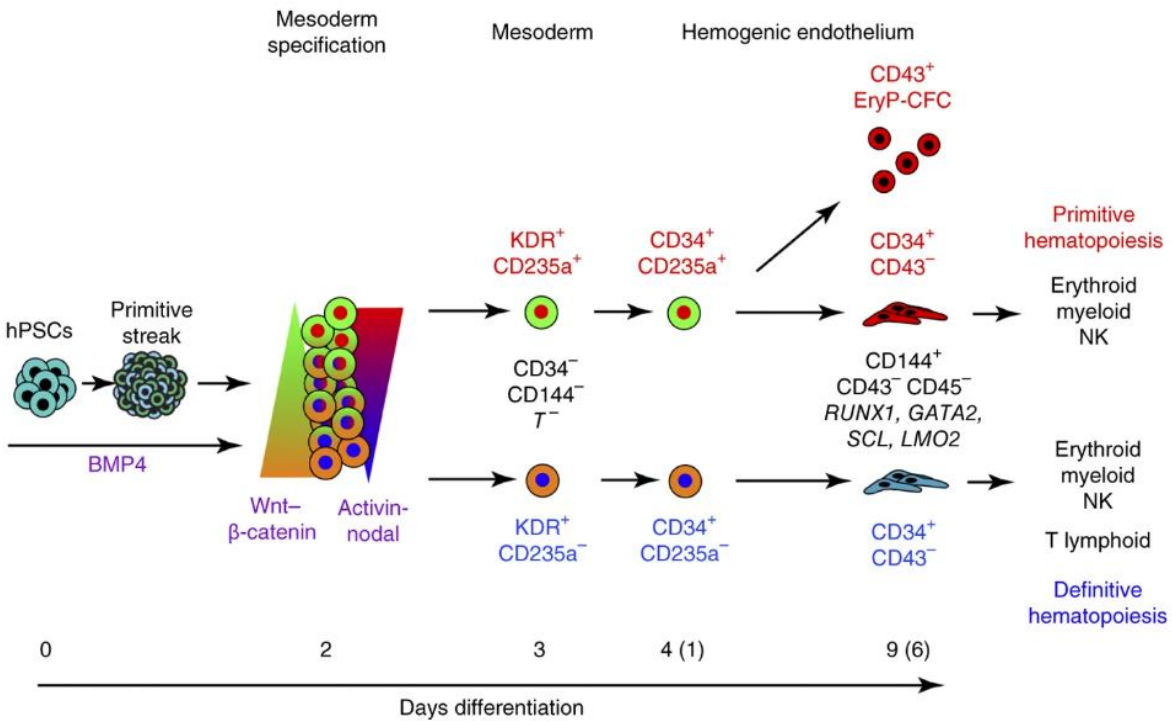


Figure 1.2: Schematic of hPSCs hematopoietic differentiation model system
Adapted from Stugeon *et al.* 2014[85], this details the differentiation used throughout this thesis to specify primitive/EMP hemogenic mesoderm and definitive hemogenic mesoderm and the lineages that arise from these mesodermal populations.

Chapter 2: Human definitive hematopoietic specification from pluripotent stem cells is regulated by mesodermal expression of CDX4

J. Philip Creamer, Carissa Dege, Qihao Ren, Jolie T. K. Ho, Mark C. Valentine, Todd E. Druley, Christopher M. Sturgeon

Blood (2017) 129 (22): 2988–2992

2.1 Acknowledgments

This article was published in the journal Blood in 2017 with the following contributions: J.P.C., Q.R., and C.D. designed and performed experiments, and analyzed the data; J.T.K.H. and M.C.V. analyzed expression data; T.E.D. supervised expression analyses; and C.M.S. designed and supervised the research project, performed experiments, analyzed the data, and wrote the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

This work was supported by T32 training grants from the National Institutes of Health, National Heart, Lung, and Blood Institute (HL007088-41) (J.P.C. and C.D.) and (HG000045) (M.C.V.); T.E.D. is supported by a grant from the Children's Discovery Institute of Washington University and St. Louis Children's Hospital; and C.M.S. is supported by an American Society of Hematology Scholar Award.

2.2 Abstract

The generation of hematopoietic stem cells from human pluripotent stem cells (hPSCs) is a major goal for regenerative medicine. Achieving this goal is complicated by our incomplete understanding of the mechanism regulating definitive hematopoietic specification. We used our stage-specific hPSC differentiation method to obtain and identify, via CD235a expression, mesoderm harboring exclusively primitive or definitive hematopoietic potential to understand the genetic regulation of definitive hematopoietic specification. Whole-transcriptome gene expression analyses on WNT-dependent KDR+CD235a− definitive hematopoietic mesoderm and WNT-independent KDR+CD235a+ primitive hematopoietic mesoderm revealed strong *CDX* gene expression within definitive hematopoietic mesoderm. Temporal expression analyses revealed that CDX4 was expressed exclusively within definitive hematopoietic KDR+CD235a− mesoderm in a WNT- and fibroblast growth factor-dependent manner. We found that exogenous *CDX4* expression exclusively during mesoderm specification resulted in a >90% repression in primitive hematopoietic potential, but conferred fivefold greater definitive hematopoietic potential, similar to that observed following WNT stimulation. In contrast, *CDX4* knockout hPSCs had intact primitive hematopoietic potential, but exhibited a fivefold decrease in multilineage definitive hematopoietic potential. Taken together, these findings indicate that *CDX4* is a critical transcription factor in the regulation of human definitive hematopoietic specification, and provides a mechanistic basis for WNT-mediated definitive hematopoietic specification from hPSCs.

2.3 Introduction

The generation of hematopoietic stem cells (HSCs) from human pluripotent stem cells (hPSCs) is a major goal for regenerative medicine. To reproducibly achieve this goal, we must first

understand human hematopoietic ontogeny. Embryonic hematopoiesis is classically defined by the spatiotemporal emergence of at least 2 distinct programs.[33] The first, primitive hematopoiesis, does not give rise to HSCs,[120] but instead transiently gives rise to a limited subset of lineages, including HBE-expressing erythroblasts or primitive erythroid colony-forming cells (EryP-CFCs), which can be reliably used as an indicator of the primitive hematopoietic program.[29,85] Shortly thereafter, the definitive program emerges and gives rise to all lineages found in the adult, including the HSC.[49,121] When differentiating hPSCs, the development of the definitive program can be distinguished from the primitive program by using in vitro assays for HBG-expressing erythroblasts and T lymphocytes.[29] However, the genetic regulation of human definitive hematopoietic specification, both in vivo and in vitro, remains unclear.

We recently developed a method to independently specify progenitors of the primitive or definitive hematopoietic programs from hPSCs via stage-specific manipulation of canonical WNT signaling.[85] This approach generates a WNT-dependent KDR+CD235a− mesodermal population that gives rise to CD34+ definitive hematopoietic progenitors, as well as a WNT-independent KDR+CD235a+ mesodermal population that gives rise to CD43+ primitive hematopoietic progenitors. Using this tractable system, we interrogated the transcriptional regulation of definitive hematopoietic specification at its earliest identifiable mesodermal progenitor, and identified CDX4 as a critical regulator of human definitive hematopoietic progenitor specification.

2.4 Methods

Culture and differentiation. The hPSC line H1 (WA01; WiCell) was maintained on irradiated mouse embryonicfibroblasts in hESC media as described previously[78]. hPSCs were differentiated

as described previously[72,85]. Briefly, hPSCs were MEF-depleted by culturing on Matrigel (BD Biosciences) in hESC media for 24 hr. Embryoid bodies were generated by treating hPSCs with trypsin-EDTA (0.05%) for 1 min. Cells were detached by scraping to form small aggregates (6–10 cells). Embryoid bodies were resuspended in SFD[122] supplemented with L-glutamine(2 mM), ascorbic acid (1 mM), monothioglycerol (MTG, 4x10-4 M; Sigma), holo-transferrin (150 µg/mL), BMP-4 (10 ng/mL), bFGF (5 ng/mL), Activin A (1 ng/mL), and either DMSO (vehicle control; 0.1%), CHIR99021 (3 µM), or IWP2 (3 µM), as indicated in **Figure 1A**. Following 72 hrs of differentiation, embryoid bodies were washed with IMDM and then placed in StemPro-34 media supplemented with L-glutamine (2 mM), ascorbic acid (1 mM), monothioglycerol (MTG, 4 × 10−4 M; Sigma-Aldrich), holo-transferrin (150 µg/mL), VEGF (15 ng/mL), IL-6 (10 ng/mL), IGF-1 (25 ng/mL), IL-11 (5 ng/mL), SCF (50 ng/mL), and EPO (2 U/mL final). Cultures were maintained in a 5% CO_2/5% O_2/90% N_2 environment. All recombinant factors are human and were purchased from R&D Systems (Minneapolis, MN) except EPO and IGF-1 (Peprotech). Analysis of hematopoietic colony potential via Methocult (StemCellTechnologies) was performed as described previously[29,87].

Genome engineering of hPSCs. Generation of the inducible hCDX4 hPSC line was performed similar to that previously described, using a 3xFLAG-hCDX4 cDNA[123]. *CDX4* knockout hPSCs, with a 77bp deletion and frameshift mutation in the first exon of *CDX4*, were generated using CRISPR/Cas9 technology. CRISPR guides (http://crispr.mit.edu/)[124] were inserted into the MLM3636 plasmid (Addgene 43860), and along with Cas9 (Addgene 43945) were nucleofected into H1 hPSCs (Lonza). Single colonies were expanded and screened by PCR.

Endothelial-to-hematopoietic transition assay. CD34+CD43- hemogenic endothelium was isolated by FACS and allowed to undergo the endothelial-to-hematopoietic transition as

described previously[72,87]. Briefly, cells (CD34+CD43- or CD34+CD43-CD73-CD184- cells) were aggregated overnight at a density of 2x105 cells/mL in StemPro-34 media supplemented with L-glutamine (2 mM), ascorbic acid (1 mM), monothioglycerol (MTG, 4 ×10−4 M; Sigma-Aldrich), holo-transferrin (150 µg/mL), TPO (30 ng/mL), IL- 3 (30 ng/mL), SCF (100 ng/mL), IL-6 (10 ng/mL), IL-11 (5 ng/mL), IGF-1 (25 ng/mL), EPO (2 U/mL), VEGF (5 ng/mL), bFGF (5 ng/mL), BMP4 (10 ng/mL), FLT3L (10 ng/mL), and SHH (20 ng/mL). Aggregates were spotted onto Matrigel-coated plasticware and were cultured for additional 9 days. Cultures were maintained in a 5% CO_2/5% O_2/90% N_2 environment. Hematoendothelial cultures were harvested by trypsinization and assessed for hematopoietic potential by Methocult.

OP9-DL4 co-culture assay for T-lymphoid potential. OP9 cells expressing Delta-like 4 (OP9-DL4) were generated and described previously[125,126]. Isolated CD34+ CD43- cells were added to individual wells of a 6-well plate containing OP9-DL4 cells, and cultured for 21-28 days as described previously[29]. Briefly, cells were cultured in a-MEM supplemented with 20% FBS, SCF (30ng/mL, first 5 days only), FLT3L (5 ng/mL) and IL-7 (5 ng/mL). Every four days co-cultures were transferred onto fresh OP9-DL4 cells by vigorous pipetting and passaging through a 40 µm cell strainer. Limiting dilution assays of CD34+CD43- cells was performed on either control (WT) or *CDX4*$^{-/Y}$ (KO) hPSCs. 160,000 (n = 2), 100,000 (n = 2), 30,000 (n = 4), 10,000 (n = 4), 3,000 (n = 4), 1,000 (n = 4), 300 (n = 4) or 100 (n = 4) cells were plated for T-lymphoid assay (biological n = 3). Cultures were assayed following 28 days of co-culture, as above, for the presence of a CD45+CD4+CD8+ population. Progenitor frequency was calculated by Extreme Limiting Dilution Analysis (http://bioinf.wehi.edu.au/software/elda/)[127].

Gene expression analyses. Total RNA was prepared for whole-transcriptome sequencing using the Clontech SMARTer kit, and was sequenced with an Illumina HiSeq 2500 with 1x50 single

reads. Reads were uploaded to the Sequence Read Archive and are available as part of BioProject PRJNA35244, accession number SRP093125. RNA-seq data was analyzed using the kallisto/sleuth software suite (https://github.com/pachterlab/sleuth)[128]. qRT-PCR was performed as previously described[85]. Briefly, total RNA was isolated with the RNAqueous RNA Isolation Kit (Ambion), followed immediately by transcription into cDNA using random hexamers and Oligo (dT) with Superscript III Reverse Transcriptase (Invitrogen). Real-time quantitative PCR was performed on a StepOnePlus thermocycle (Applied Biosystems), using Power Green SYBR mix (Invitrogen). Primer sequences are available upon request. Gene expression was evaluated as DeltaCt relative to control (*ACTB*).

Globin analyses. Following 10 days (EryP-CFC) or 14 days (BFU-E) of erythroblast maturation in Methocult (StemCellTechnologies), individual colonies were picked my mouth pipetting with a capillary tube. 10 colonies were pooled per biological replicate and assessed for HBG and HBE expression by qRT-PCR, as previously described[85].

Flow Cytometry and Cell Sorting. The antibodies used are all as previously described[29,85,87]. KDR (clone 89106), CD4 (clone RPA-T4), CD8 (clone RPA-T8), CD34-APC (clone 8G12), CD34- PE-CY7 (clone 4H11), CD43 (clone 1G10), CD45 (clone 2D1), CD56 (clone B159), CD73 (clone AD2), CD184 (clone 12G5) and CD235a (clone HIR-2). All antibodies were purchased from BD Biosciences (San Diego, CA) except CD34-PECY7 purchased from eBioscience, CD184 purchased from Biolegend and KDR purchased from R&D systems. Cells were sorted with a FACSAria™II (BD) cell sorter and analyzed on a LSRii (BD) cytometer.

2.3 Results and discussion

2.3.1 Mesodermal *CDX4* expression is specific to definitive hematopoietic specification

Given that our hPSC differentiation system gives rise to populations of mesoderm harboring exclusively primitive or exclusively definitive hematopoietic progenitors (**Figure 2.1A**)[85] we asked whether these populations could identify which transcription factor(s) regulate definitive hematopoietic specification within early mesoderm. We isolated by FACS KDR+CD235a− and KDR+CD235a+ mesoderm, generated by CHIR99021 or IWP2 treatment, respectively (**Figure 2.1B)**[85] and performed whole-transcriptome expression analyses. Differential gene expression analysis[128] revealed significant enrichment of the CDX and HOX genes within definitive hematopoietic mesoderm (**Figure 2.1C**, **Figure 2.3**). *CDX1*, *CDX2*, and *CDX4* were all highly expressed in definitive, but not primitive, hematopoietic mesoderm, and have been previously identified as being expressed during hPSC-derived definitive hematopoietic specification.[118]

Interestingly, qRT-PCR analyses of each gene over the first 6 days of differentiation revealed that *CDX1* and *CDX2* are expressed within 24 hours of differentiation, whereas *CDX4* was instead upregulated twofold at the time of CHIR99021 treatment (**Figure 2.1D**). CDX expression immediately decreased following CHIR99021 removal. This suggested that *CDX1* or *CDX2* may not specifically regulate definitive hematopoietic progenitor specification, but instead regulate mesendoderm specification, and thus may affect the emergence of both programs, whereas *CDX4* expression correlates with definitive hematopoietic progenitor specification by WNT signaling.

In addition to WNT signaling, differentiation cultures at this stage employ the use of BMP4 and basic fibroblast growth factor (FGF)[85,87] both of which contribute to the expression of *CDX*

genes in early mesoderm.[84,108] Although inhibition of BMP4 signaling with recombinant NOGGIN caused a complete block in mesoderm formation (not shown), the inhibition of basic FGF signaling with PD173074 at the same time as CHIR99021 treatment led to the emergence of a CD34+CD43− population that lacked a distinct CD73−CD184− HE (**Figure 2.1E**). [87,129] Interestingly, FGF receptor inhibition (FGFRi) had no effect on the expression of *CDX1* or *CDX2*, but did repress *CDX4* expression (Figure 1F). Collectively, these results suggested that *CDX4* is a relevant transcriptional target during human definitive hematopoietic specification.

2.3.2 *CDX4* is required for efficient human definitive hematopoietic specification

In both mouse and zebrafish models, cdx4 deficiency has been implicated in the regulation of embryonic hematopoiesis.[84,112-114,130,131]Although *Cdx4*$^{-/-}$ mice showed no significant definitive hematopoietic defects,[115] zebrafish exhibit a moderate decrease in definitive hematopoietic potential that was significantly enhanced when combined with *cdx1* knockdown[112,130,131] suggesting species-specific usage of different *Cdx* genes during definitive hematopoietic specification. As we observed definitive hematopoietic mesoderm-specific expression of *CDX4* (**Figure 2.1C**), we reasoned that the role of *CDX4* in human embryonic hematopoiesis may be elucidated using our hPSC model system, by monitoring the ontogeny of primitive and definitive hematopoietic progenitors from early mesoderm.[85]

We first generated an inducible expression system using the AAVS1 locus[123] to allow for *CDX4* expression at any stage of differentiation. Exogenous *CDX4* expression from days 2 to 3 of differentiation (**Figure 2.4A**) repressed the specification of the primitive hematopoietic progenitors, as we observed a dramatic decrease in the emergence of CD43+ primitive hematopoietic progenitors[132] (**Figure 2.2A**), and a 10-fold decrease in detectable EryP-CFCs

(Figure 2B). Because this resembled CHIR99021 treatment of differentiation cultures,[85] which causes *CDX4* expression (**Figure 2.1C-D**), we asked whether induction of *CDX4* expression during the same window of time would yield a similar effect as CHIR99021 treatment, resulting in an enrichment of definitive hematopoietic specification.[85] Therefore, differentiation cultures did not have WNT signaling manipulated, so as to allow for the specification of both programs (**Figure 2.1A**). Under these conditions, the definitive hematopoietic progenitors were still specified with doxycycline treatment, because functional CD34+CD43−CD73−CD184− HE[87,129] was still specified with similar efficiency to control conditions (**Figure 2.2C-D**, **Figure 2.4B**). However, *CDX4* overexpression caused a functional increase within this population, as indicated by a fivefold increase in definitive erythroid progenitors, similar to that observed following CHIR99021 treatment (**Figure 2.2E**). Collectively, these observations indicate that mesodermal *CDX4* expression recapitulates WNT activation during mesoderm specification[85] resulting in definitive hematopoietic specification and primitive hematopoietic repression.

We next used a *CDX4* knockout ($CDX4^{-/Y}$) hPSC line, which completely lacked WT *CDX4* expression, whereas *CDX1* and *CDX2* were still expressed (**Figure 2.2F**, **Figure 2.4D**). Under all differentiation conditions (**Figure 2.1A**), $CDX4^{-/Y}$ cultures gave rise to approximately threefold more CD43+ cells (**Figure 2.2G**). $CDX4^{-/Y}$ hPSCs gave rise to threefold more EryP-CFCs in the absence of WNT signaling (**Figure 2.2H**), likely due to the complete absence of *CDX4* expression in comparison with control IWP2-treated cultures (**Figure 2.2F**). CHIR99021 treatment still repressed primitive hematopoiesis, suggesting other β-catenin transcriptional targets also repress primitive hematopoietic specification. In contrast, when CHIR99021-derived definitive CD34+CD43− cells were assessed for hematopoietic potential, the $CDX4^{-/Y}$ cells exhibited an approximately sevenfold reduction in HBG-expressing BFU-E potential, and a

fourfold decrease in myeloid potential (**Figure 2.2I, Figure 2.4C**). Further, limiting dilution analyses of T-lymphoid potential revealed a fivefold decrease in lymphoid potential in *CDX4*$^{-/Y}$ CD34+ cells (**Figure 2.2J**), indicating a broad decrease in multilineage definitive hematopoietic potential.

To determine the cause of this, we asked whether *CDX4*$^{-/Y}$ hPSCs have either defective HE specification or an impaired endothelial-to-hematopoietic transition. *CDX4*$^{-/Y}$ CD34+CD43− cells exhibited an average fivefold reduction in the specification of CD73−CD184− HE (**Figure 2.2K-L),** indicating they have impaired definitive hematopoietic specification. However, this HE population remained functional, giving rise to definitive BFU-E with similar efficiency to control hPSCs (**Figure 2.2M**). Because both *CDX1* and *CDX2* were still expressed under these conditions (**Figure 2.2F**), it is possible that either may be redundantly[106] contributing to definitive hematopoietic specification in the absence of *CDX4*, similar to that observed in zebrafish.[115] Collectively, these results establish that *CDX4* is a key regulator of specifying human definitive hematopoietic progenitors within mesoderm, by regulating the specification of HE, and thus provides a mechanistic basis for WNT-mediated definitive hematopoietic progenitor specification.[85]

CDX/HOX expression has been suggested to be essential for definitive hematopoietic specification from hPSCs.[118,133] Together, our data demonstrate the importance of *CDX4* expression for human definitive, but not primitive, hematopoietic progenitor specification within mesoderm, prior to HE emergence. With this insight, it will be possible to interrogate the intrinsic and extrinsic regulators of human definitive hematopoietic specification, so as to ultimately increase hPSC-derived definitive hematopoiesis for regenerative medicine applications.

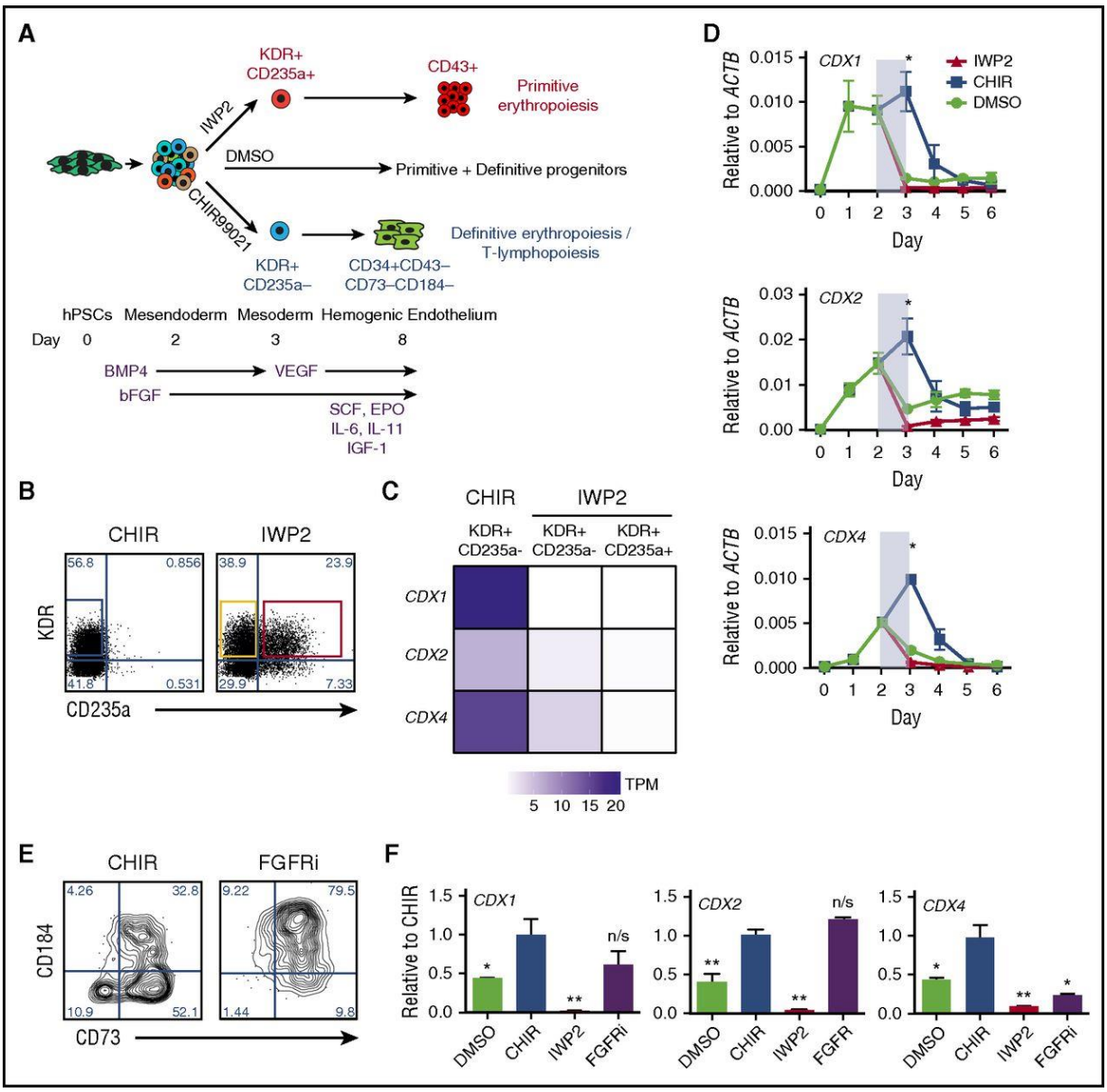


Figure 2.1: *CDX4* is expressed at the onset of definitive hematopoietic progenitor specification within mesoderm.

(**A**) Differentiation schematic and hematopoietic progenitor identification. hPSCs are differentiated using a serum-free, stroma-free approach, with stage-specific application of WNT signal manipulation. Inhibition of WNT signaling within mesendoderm with 3 μM IWP2 leads to the generation of KDR+CD235a+ mesodermal population, which gives rise to CD43+ primitive hematopoietic progenitors, whereas WNT activation with 3 μM CHIR99021 generates a

KDR+CD235a− mesodermal population that gives rise to CD34+CD43−CD73−CD184− HE. No manipulation of WNT signaling leads to a heterogeneous population of primitive and definitive hematopoietic progenitors. (**B**) Representative cell-sorting strategy employed for RNA-seq analyses. Mesoderm harboring definitive (blue) or primitive (red) progenitors were isolated by FACS. (**C**) Heatmap of *CDX* gene expression within different mesodermal populations, as determined by RNA-seq. n = 4. (**D**) qRT-PCR analyses of *CDX1* (top), *CDX2* (middle), and *CDX4* (bottom) expression during the first 6 days of differentiation as in panel A. Period of WNT manipulation is shaded in light blue. n ≥ 3 mean ± standard error of the mean (SEM). Student t test compared with DMSO control: *P < .05. (**E**) Representative flow cytometric analysis of CD73 and CD184 expression, gated on CD34+CD43− cells following either CHIR99021 (CHIR) treatment or CHIR + 1 μM PD173074 (FGFRi) treatment as in panel A. (**F**) qRT-PCR analyses of *CDX1* (left), *CDX2* (middle), and *CDX4* (right) expression on day 3 of differentiation, following treatment with either vehicle (DMSO), CHIR99021 (CHIR), IWP2, or PD173074 (FGFRi) as in panel A. Normalized to CHIR treatment. n = 3 mean ± SEM. Student t test compared with CHIR treatment: *P < .05; **P < .01. BMP4, bone morphogenetic protein 4; DMSO, dimethylsulfoxide; EPO, erythropoietin; IGF-1, insulin-like growth factor 1; IL-6, interleukin-6; qRT-PCR, quantitative reverse transcription polymerase chain reaction; RNA-seq, RNA sequencing; SCF, stem cell factor; TPM, transcripts per million; VEGF, vascular endothelial growth factor.

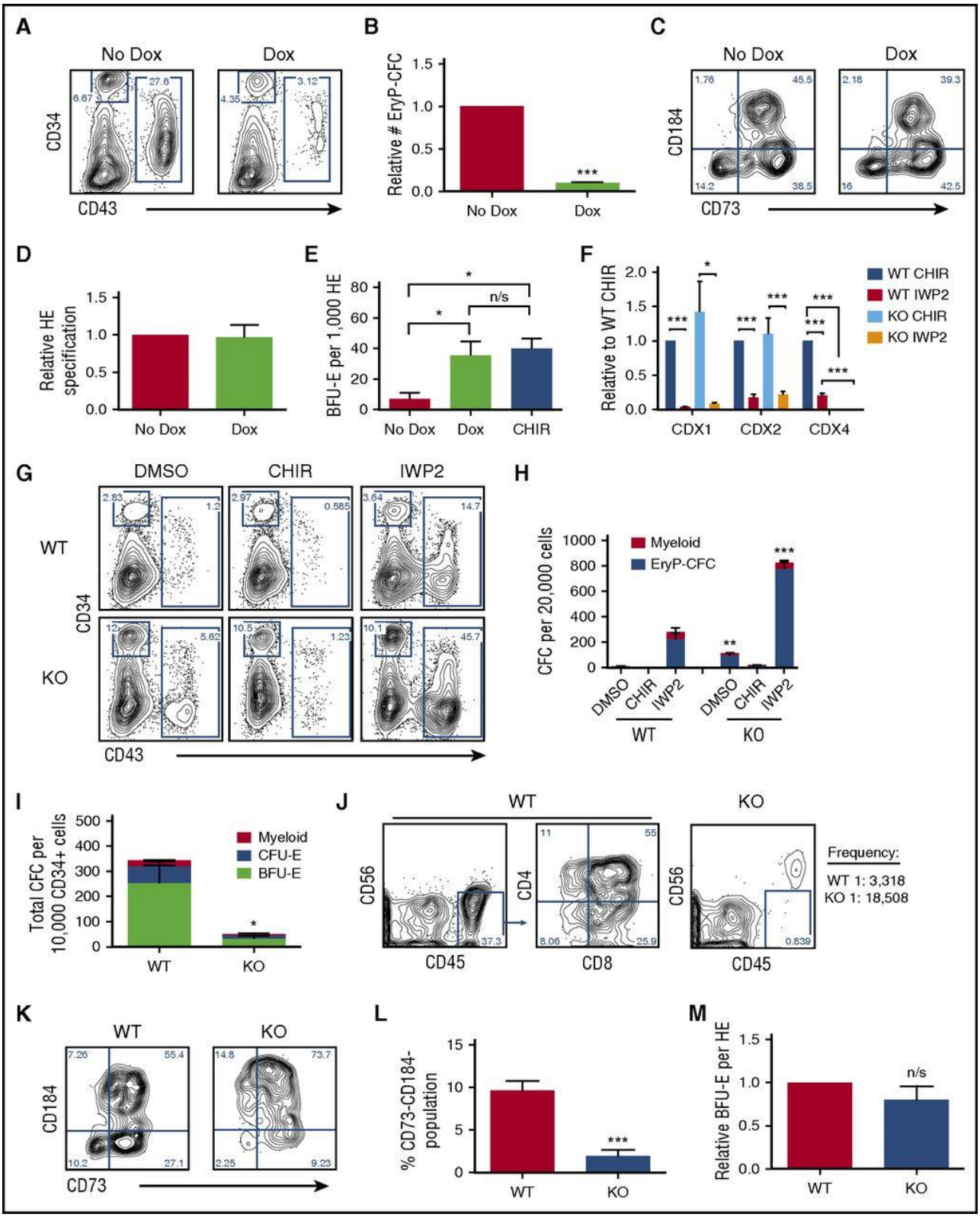


Figure 2.2: Mesodermal CDX4 expression is critical for efficient definitive hemogenic endothelial specification.

(**A-E**) Analyses of primitive and definitive hematopoietic potential following doxycycline-inducible *CDX4* expression via the AAVS1 locus. (**A**) Representative flow cytometric analysis of CD34 and CD43 expression on day 9 of differentiation, following IWP2 and doxycycline treatment from days 2 to 3 of differentiation. (**B**) Normalized EryP-CFC potential at day 9 of differentiation as in panel A. n = 3 mean ± SEM. Student t test: ***P < .001. (**C**) Representative flow cytometric analysis of CD73 and CD184 expression within CD34+CD43− cells obtained on day 8 of differentiation, following DMSO treatment, with or without doxycycline, from days 2 to 3 as in Figure 1A. (**D**) Quantification of CD73−CD184− HE as in panel C. n = 4. Student t test: P > .05. (**E**) Quantification of definitive erythroid burst forming unit (BFU-E) potential from 1000 CD34+CD43−CD73−CD184− hemogenic endothelial progenitors, following 9 days of hemato-endothelial culture to promote the endothelial-to-hematopoietic transition. n = 3 mean ± SEM. Student t test compared with DMSO: *P < .05. (**F-M**) Analyses of primitive and definitive hematopoietic potential from *CDX4*$^{-/Y}$ hPSCs. (F) qRT-PCR analysis of *CDX1*, *CDX2*, and *CDX4* expression on day 3 of differentiation within wild-type (WT) and *CDX4*$^{-/Y}$ (knockout [KO]) hPSCs, following CHIR99021 or IWP2 treatment as in Figure 1A. n = 4 mean ± SEM. Student t test: *P < .05; ***P < .001. (**G**) Representative CD34 and CD43 flow cytometric analyses on day 9 of differentiation following WNT manipulation from days 2 to 3 as in Figure 1A. (H) Primitive hematopoietic potential within day 9 differentiation cultures of WT and *CDX4*$^{-/Y}$ (KO) hPSCs, following WNT manipulation as in Figure 1A. n ≥ 4 mean ± SEM. Student t test compared across WT and KO, per condition: **P < .01; ***P < .001. (**I**) Quantification of definitive erythro-myeloid colony-forming potential from CHIR99021-derived CD34+ progenitors, following 9 days of hemato-endothelial culture to promote the endothelial-

to-hematopoietic transition. Cultures were treated with CHIR99021 from days 2 to 3 as in Figure 1A. Normalized to 10 000 CD34+CD43− day 8 input cells. n ≥5 mean ± SEM. Student t test: *P < .05. (**J**) Representative flow cytometric analysis of T-lymphoid potential of WT and *CDX4*$^{-/Y}$ (KO) hPSCs, under limiting dilution conditions. Shown is T-lymphoid analyses following 28 days OP9−DL4 coculture under T-lymphopoiesis promoting conditions. Input population shown is 10 000 CD34+ progenitors. Limiting dilution analyses indicate WT CD34+ cells possess lymphoid progenitors at a 1:3318 frequency, and KO cells at 1:18 508 frequency. n = 3. P = .00224. (**K**) Representative flow cytometric analysis of CD73 and CD184 expression within CD34+CD43− cells obtained on day 8 of differentiation, following CHIR99021 treatment from days 2 to 3 as in Figure 1A. (**L**) Quantification of CD73−CD184− HE as in (**K**). n > 6. Student t test: ***P < .001. (**M**) Normalized definitive BFU-E potential from CD34+CD43−CD73−CD184− hemogenic endothelial progenitors, following 9 days of hemato-endothelial culture to promote the endothelial-to-hematopoietic transition. n = 5 mean ± SEM. Student t test: P > .05. CFU-E, erythroid colony-forming unit; DOX, doxycycline; n/s, not significant.

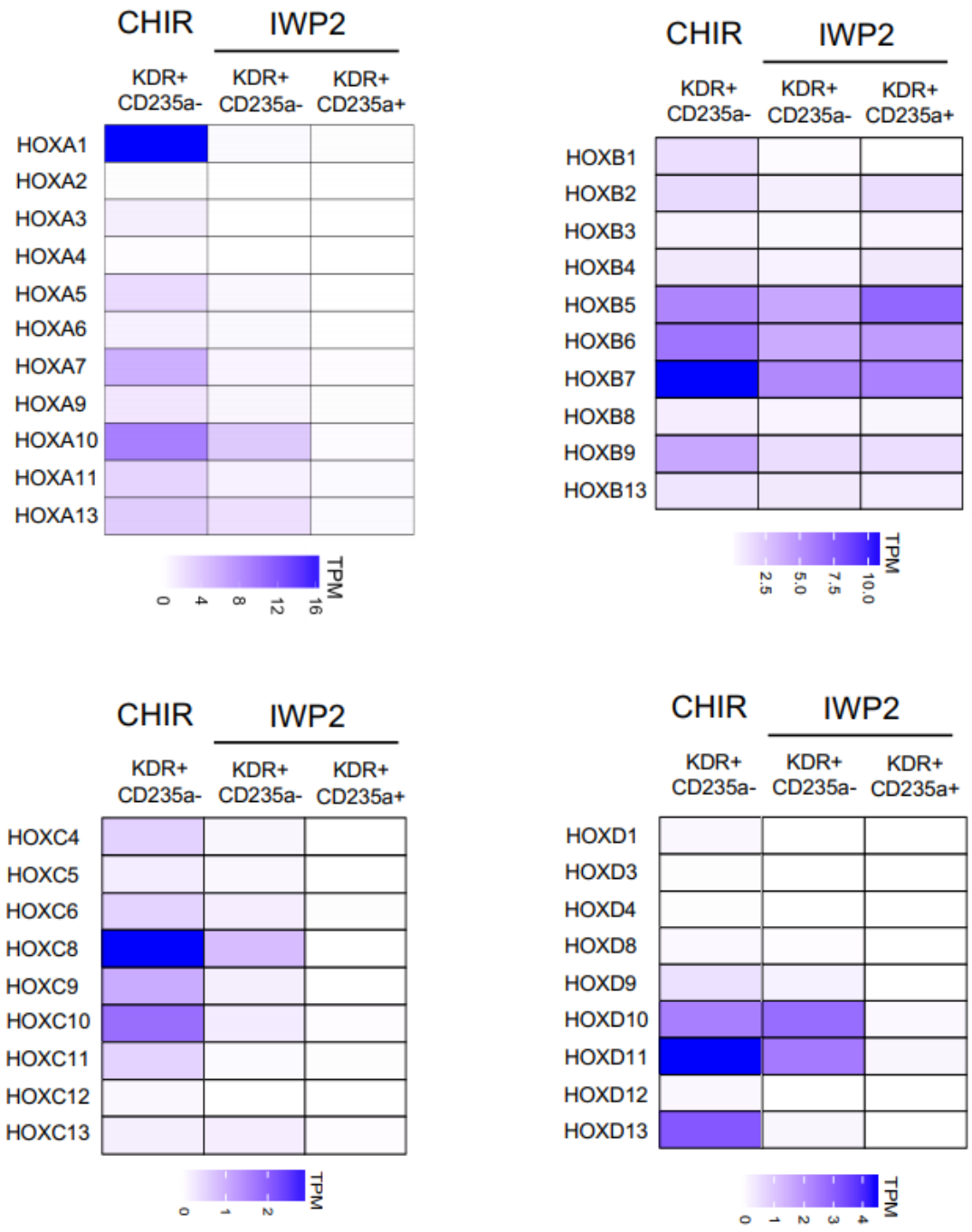


Figure 2.3: *HOX* gene expression differences in mesoderm

Heatmap of HOX gene expression within different mesodermal populations, as determined by RNA-seq. Populations obtained as in Figure 1B. TPM: Transcripts Per Million. n = 4.

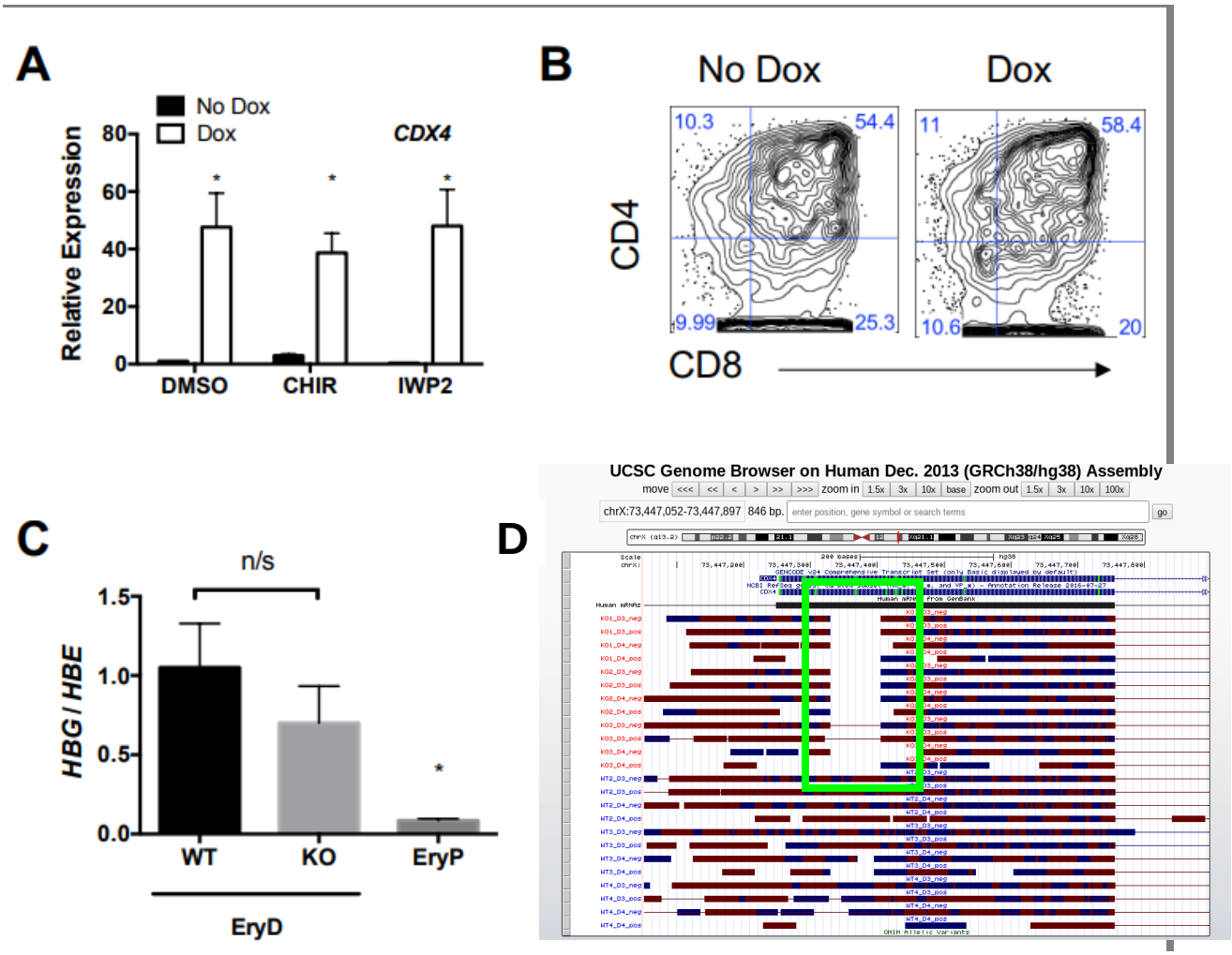


Figure 2.4: Characterization of *iCDX4* and *CDX4*$^{y/-}$ hPSC lines

(**A**) qRT-PCR analysis of *CDX4* expression following 2 µg/mL doxycycline treatment from days 2-3 of differentiation, as in Figure 1A. n = 3 mean ± SEM. Students t-test compared between no doxycycline and doxycycline treated cultures * $p < 0.05$. (**B**) T-lymphoid potential of the CD34+CD43- populations derived from day 8 of differentiation cultures, following treatment or not with doxycycline from days 2-3 of differentiation. Representative CD4 and CD8 flow cytometric analysis following 21+ days OP9-DL4 co-culture under T-lymphopoiesis promoting conditions. Gated on CD45+CD56- population. n = 2. (**C**) Ratio of HBG/HBE expression, as determined by qRT-PCR, within erythroid colonies derived as in (Figure 2H; "EryP") and

(Figure 2I; “EryD”). n = 4 mean ± SEM. Students t-test * $p < 0.05$. (**D**) Genome browser snapshot of the *CDX4* locus, showing next generation sequencing of *CDX4*$^{y/-}$ hPSCs (in red) compared to WT (in blue), demonstrating the 77 bp deletion as a result of CRISPR Cas9 cleavage.

Chapter 3: Targets of *CDX4* and their role in promoting definitive hematopoietic specification and repression of cardiogenic fate

3.1 Abstract

After having demonstrated that *CDX4* is a critical regulator of definitive hematopoiesis, an interesting question remained: what gene expression changes are resulting from activation of WNT/FGF dependent *CDX4* expression, in the context of definitive hemogenic mesoderm? *CDX4* in human and animal models is generally known act upstream of *HOX* genes, in particular *HOXA* genes[112,113,118,131] in hematopoietic development. I hypothesized that *CDX4* is binding directly to target genes responsible for the specification of hemogenic endothelium (HE), such as in the *HOXA* cluster, which has shown to be critical for HE[117] and HSC function[116,118]. In order to find these transcriptional targets of *CDX4*, I utilized a *CDX4*$^{y/-}$, *iCDX4* human pluripotent stem cell (hPSC) line, which overexpresses 3xFLAG-tagged *CDX4* upon doxycycline (DOX induction, to perform paired RNAseq and chromatin immunoprecipitation sequencing (ChIPseq) on 3xFLAG-CDX4 bound genomic DNA. The sequencing revealed areas of CDX4 binding during definitive hematopoietic specification in mesoderm, but also allowed me to explore CDX4 binding in the context of primitive hemogenic mesoderm, as I have shown that *CDX4* expression represses the primitive program. By using FACS, I isolated each of these populations: KDR+CD235a-CD184- definitive, retinoic acid independent hemogenic mesoderm, KDR+CD235a-CD184+ retinoic acid dependent hemogenic mesoderm, and KDR+CD235a+ primitive hemogenic mesoderm with or without CDX4 expression by manipulating DOX during

mesoderm formation. As expected, there was broad binding across the *HOXA* cluster in both definitive hemogenic mesoderm's (CD184-/+), but interestingly, there was population specific *HOXA5/7/9* gene expression as a result of *CDX4* induction. *CYP26A1* were also identified as being bound upregulated by *CDX4*, which suggests influence on retinoic acid signaling respective. *CDX4* was also found to be activating *CDX2*, possibly in a feedback loop within KDR+CD235a+ primitive hemogenic mesoderm. Intriguingly, *TBX20*, an important cardiac transcription factor, was found to be significantly downregulated and CDX4-bound in all three populations as a result of *CDX4* expression, suggesting that *CDX4* expression results in suppression of cardiac lineages. Other cardiomyocyte related genes, such as *MESP1* and *NKX2-5*, were also suppressed but not *CDX4* bound in KDR+CD235a+ mesoderm, which has been identified as a the progenitor of ventricular cardiomyocytes in previous work[134]. Further studies utilizing WT and *CDX4*$^{y/-}$, *iCDX4* lines differentiated under cardiac promoting conditions revealed that loss of *CDX4* lead to a large increase in cardiomyocyte formation and rescue of *CDX4* expression during mesoderm formation reversed this increase. These data revealed that *CDX4* is regulating *HOXA* genes as expected within hemogenic mesoderm, but is also suppressing the specification of cardiac progenitors, through the downregulation of cardiomyocyte transcription factors like *TBX20.*

3.2 Methods

Culture and differentiation. Hematopoietic differentiations and functional assays were performed according to the methods in **Chapter 2**, with the addition of SB431542 (6 μM) alongside CHIR99021[118]**.**

For cardiac promoting conditions, alterations to the base protocol were adapted from Lee *et al.*[134] Briefly, after aggregation of dissociated hPSCs, cells were suspended in SFD media the same as

hematopoietic conditions, except containing BMP-4 (5 ng/mL) and Activin A (10 ng/mL) at indicated concentrations and without addition of CHIR99021 or IWP2. Following 72 hours of differentiation, cells were washed and resuspending in StemPro-34 media as hematopoietic with differentiations, except with only VEGF (10ng/mL) and IWP2 (3 μM) for 48 hours. Cells were then washed, and media exchanged every 3 days until day 12, when cells were transitioned to 5% CO_2 normoxia environment. Media was continued to be exchanged every 3 days, until analysis was performed at day 20.

Genome engineering of hPSCs. Generation of the *CDX4*$^{y/-}$, *iCDX4* recue line was generated using the same strategy as the *iCDX4* found in **Chapter 2**, but into the already established *CDX4*$^{y/-}$ KO line also detailed in **Chapter 2**.

scRNA-seq analyses. Cells from each day 3 differentiation culture condition were methanol-fixed as previously described[135]. Libraries were prepared following the manufacturer's instruction using the 10X Genomics Chromium Single Cell 3′ Library and Gel Bead Kit v2 (PN-120237), Chromium Single Cell 3′ Chip kit v2 (PN-120236), and Chromium i7 Multiplex Kit (PN-120262). 17,000 cells were loaded per lane of the chip, capturing >6000 cells per transcriptome. cDNA libraries were sequenced on an Illumina HiSeq 3000. Sequencing reads were processed using the Cell Ranger software pipeline (version 2.1.0). Using Seurat[136] (version 3.0.2) implemented in R (version 3.5.1), the dataset was filtered by removing genes expressed in fewer than 3 cells, and retaining cells with unique gene counts between 200 and 6000. The remaining UMI counts were log normalized and mitochondrial UMI counts were regressed out. Principal component analysis was used to generate t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and project (UMAP) plots. Monocle[137] (version 2.10.1) was used for pseudotime analysis. First size factors and dispersions were estimated, and

then genes were filtered with expression 10 cells. Doublets were removed by filtering out cells with 24813 total RNA. Cell clustering and trajectory construction were performed using an unsupervised approach.

Western Blotting. Cells were harvested and then lysed in RIPA-LS buffer containing protease inhibitors with sonication on ice. After lysis, protein was quantified using BCA quantification with a BSA standard curve and ~20 μg of protein was boiled in loading buffer and loaded per sample for PAGE. The protein was then transferred via semi-dry method onto PVDF membrane. After blocking with BSA, the membrane was staining overnight at 4C using Flag M2 and αTubulin antibody. After washing with BSA, the membranes were then probe with LI-COR IRDye secondary antibodies at 4C for 1 hour, followed by additional BSA washes. Membranes were imagined on a LI-COR Odyssey.

Paired ChIP/RNAseq experiments. After relevant cell types were isolated $0.75x10^6$ cells per sample were processed according to the beginning steps of Chipmentation[138]. Briefly, after snap freezeing at -80C, cells were washed and then formaldehyde fixed at 1% while rocking for 10 minutes at room temperature and quenched by adding glycine to 0.125 M concentration for an additional 5 minutes at room temp. After centrifugation, cells were resuspended in sonication buffer (0.25 % SDS) with protease inhibitors and subjected to Covaris microtube sonication, producing 200-700 bp fragments as assessed by Tapestation analysis. Cells were then diluted with 1:1.5 with equilibration buffer (10mM Tris, 233 mM NaCl, 1.66 % TritonX-100, 0.166 % DOX, 1 mM EDTA, with protease inhibitors). Cells were then IP'd rotating overnight at 4C in RIPA-LS + inhibitor with anti-FLAG antibody and the next day combined with Protein G Dynabeads for 4 hours. After using magnetic separation to remove the supernatant, the beads were then successively washed in twice in RIPA-LS, twice in RIPA-HS, twice in RIPA-LiCL,

and then once in 10 mM Tris pH8. To elute the chromatin, beads were resuspended in ChIP elution buffer with proteinase K and incubated at 55C for 1 hour and 65C for 6-10 hours. The supernatant was then transferred to a DNA lo-bind tube, followed by a SPRI bead DNA cleanup and sequencing performed by core facility. Total RNA was prepared for whole-transcriptome sequencing using the Clontech SMARTer kit using between 0.75-2.5x10^5 cells as input. Both ChIP and RNAseq samples were sequenced on an Ilumina Novaseq S4 XP with 2x150 reads. Reads from both sample sets was aligned to hg38 reference using the STAR aligner[139], for RNAseq reads were counted using Subread[140] and ChIP peaks were called using MACS2[141].

Gene Expression analyses. RT-qPCR was performed as described in **Chapter 2**.

Flow Cytometry and Cell Sorting. Flow cytometry was performed for hematopoietic differentiations as previously described, using methods from **Chapter 2**. Intracellular staining for CTNT was performed as previously described[134]; briefly, aggregates on day 20 were dissociated using Collagenase II (0.5mg/ml) overnight, followed by 8 minute Trypsin (0.25%) incubation at 37C, halted with 5% FCS in PBS. Cells were then fixed for 15 minutes at 4C with 4% PFA in PBS and washed, followed by permeabilization with 90% methanol for 20 minutes at 4C. Cells were then washed, stained with primary antibody at 4C overnight and then washed and stained with secondary antibody. PDGFRα antibody (clone αR1) was obtained from BD PharMingen and cTNT (clone 13-11) was obtained from ThermoFisher.

3.3 Discovery of a retinoic-acid dependent hematopoietic program in hPSCs

In order to better understand the exact cell types that exist within the definitive hemogenic mesoderm produced within our culture system, our lab performed single cell RNAseq of bulk

cultures under CHIR99021 and SB431542 (a TGFβ inhibitor, shown to improve *CDX/HOXA* expression[118]) treatment at day 3 of differentiation. After processing through the 10X Genomics Chromium system and Illumina sequencing, approximately 6,000 unique cells were obtained after filtering out cells with low read count. To identify these cell types and understand the relationships between then, the monocle2 software[137] was utilized to generate a pseudotime trajectory in an unbiased manner (**Figure 3.1A**, adapted from Luff *et al.* under review[142]). This trajectory orders cells based on similarities in gene expression into a low dimensional space to allow for inference on possible cells state and progenitor/progeny relationships. Each of the 'branches' determined by the algorithm were assessed for their cells state by looking for common germ layer markers, leading to the identification of pluripotent cells (*SOX2/NANOG*), ectoderm (*TFAP2A, DLX5, KRT7*), mesendoderm (*SOX2/T*) endoderm (*FOXA2, SOX17*), and mesoderm (*KDR*). The cells types farthest along in psuedotime are ectoderm, endoderm, and mesoderm, following well understood conventions on gastrulation and the development of the different germ layers[143].

One of the more interesting observations from this pseudotime analysis is that there are two distinct mesoderm branches formed that diverge as pseudotime progresses. Branchpoint analysis of these two mesodermal populations revealed differential expression of several genes as the cells progressed from pluripotency to either mesoderm populations (**Figure 3.1B**, Luff *et al.*[142]). *CDX4* and *CYP26A1* expression appears to be maintained in expression in one branch, where *CXCR4* (CD184) and *ALDH1A2* are expressed in the other opposing branch. *CYP26A1* and *ALDH1A2* are important regulators of retinoic acid signaling. Retinoic acid signaling has been known to play critical roles in embryonic development, in particular in the context of neurogenesis, eye development, and limb-bud fomation[144-146], but also has been implicated has

having a crucial role in the formation of the HSC[147]. Retinoic acid signaling depends on ingestion of Vitamin A (retinol) which is converted first to retinal after transport into the cytosol by various retinol dehydrogenase (RDH) and then converted to all trans retinoic acid (ATRA) by alcohol dehydrogenase enzymes (ALDH); ATRA is then bound by RA receptors (RARs) and retinoid X receptors (RXRs), which translocate to the nucleus and affect gene transcription[146]. ATRA is then degraded via a cytochrome P450 (CYP26) enzymes, leading to a halt in RAR/RXR based gene transcription.

The presence of a high *ALDH1A2*, low *CYP26A1* population, that could be segregated by *CXCR4* (CD184) lead the hypothesis that retinoic acid signaling might be playing a role in this particular mesodermal population. Initial flow cytometry experiments demonstrated that there was KDR+CD235a-CD184+ mesoderm at day 3 within our CSB treated differentiations, confirming these bioinformatic observations (**Figure 3.1C**, Luff *et al.*[142]). FACS isolation of either KDR+CD184-/+ populations, followed by reaggregation and continued culture for another five days under previous established conditions[72] lead to the production of CD34+ endothelium from both CD184- (P1) and CD184+ (P2) cells. To test for hematopoietic potential, these CD34+ cells were FACS isolated and either seeded onto OP9-DL4 stroma to test for lymphoid/T-cell potential[29] or into an EHT culture[87] followed by plating into methylcellulose to test for erythroid/myeloid potential. No T-cells were found when CD184+ derived endothelium was tested for lymphoid potential, while robust CD4/8+ T-cells were derived from CD184- mesoderm. Similarly, only small numbers of mainly myeloid colonies were derived from CD184+ mesoderm, whereas there were large numbers of erythroid and myeloid colonies formed from CD184- mesoderm (**Figure 3.1D**, Luff *et al.*[142]). This observation correlated with previous observations of the dependence of definitive HE on mesodermal expression of *CDX4*, which is

more highly expressed within CD184- mesoderm. However, with the knowledge that CD184 expression correlated with *ALDH1A2* expression, it was hypothesized that CD184+ mesoderm might require the addition of retinol (ROH) for specification of hematopoietic progenitors. Addition of ROH to the culture after FACS isolation (P2') again lead to the formation of CD34+ endothelium from CD184+ mesoderm, but was able to unlock T-cell/lymphoid potential from this endothelium. Similarly, ROH treated CD184+ mesoderm was able to be produce CD34+ endothelium that give rise to robust erythroid and myeloid colonies after the EHT. Addition of ROH to CD184- hemogenic mesoderm did not significantly affect the erythroid, myeloid, or lymphoid potential, suggesting that this population is largely independent of retinoic acid signaling, correlating with its low levels of *ALDH1A2* and higher level of *CYP26A1* expression.

The discovery of this novel ROH dependent hemogenic mesoderm raised interesting questions in terms of the genetic regulation behind the emergence of these populations. *CDX4* expression is heavily downregulated in CD184+ mesoderm, whereas its expression is maintained in CD184- mesoderm. Does *CDX4* expression promote a ROH independent hemogenic mesoderm or is it necessary for the emergency of a ROH dependent hemogenic mesoderm? Are there more critical gene expression differences that segregate these two populations and is it possible that *CDX4* could regulating gene expression in different ways in these two populations? In order to answer these questions and better understand the precise genetic targets of *CDX4* in these different hemogenic progenitors, genetic tools will be needed.

3.4 Establishment of a *CDX4*$^{-/y}$, *iCDX4* rescue line

After demonstrating the critical role that *CDX4* plays in the specification of definitive hemogenic mesoderm and its suppression of the primitive program (**Chapter 2**)[148], the stage specific role of *CDX4* was not completely addressed by a *CDX4*$^{-/y}$ KO line or the doxycycline (DOX) inducible

iCDX4 line. The possibility of *CDX4* expression impacting hematopoiesis at later stages of hematopoietic differentiation could not be entirely ruled out. To address whether mesodermal expression only of *CDX4* would lead to the specification of hemogenic endothelium, a *CDX4*$^{-/y}$, *iCDX4* line was generated, combining the *AAVS1*-based[123] DOX-inducible *CDX4* with the endogenous *CDX4*$^{y/-}$ KO line. This genetic system allows for precise control of *CDX4* expression, as there is no endogenous expression and the DOX inducible *AAVS1-CDX4* allows for tunable *CDX4* expression. The exogenous *CDX4* produced is triple tagged with a FLAG motif at the 5' end, allowing for pulldown of CDX4 protein using anti-FLAG antibodies. This would allow for chromatin immunoprecipitation (ChIP) of CDX4-bound chromatin, to find where in the genome CDX4 is binding in different mesodermal populations. Commercial antibodies against human CDX4 showed poor binding affinity and could only be detected through overexpression of CDX4 (data not shown), necessitating the use a FLAG-tagged construct. Before ChIP could be performed, it was necessary to shown that 3xFLAG-CDX4 protein can be efficiently expressed, detected, and immunoprecipitated.

To establish that 3xFLAG-CDX4 could be efficiently expressed, protein was isolated from *CDX4*$^{-/y}$, *iCDX4* hPSC's treated with or without 1 μg/ml of DOX and a western blot performed (**Figure 3.2A**). The detection of a large anti-FLAG band at ~42 kD was close to the expected size of CDX4 protein based on size estimates from commercially available antibodies and ExPASy[149]. Having established that 3xFLAG-CDX4 can be expressed and detected at the protein level, confirmation of effective immunoprecipitation (IP) was needed. Bulk day 3 differentiated *CDX4*$^{-/y}$, *iCDX4* hPSC's were harvested for protein, which was then immunoprecipitated used anti-FLAG protein with magnetic beads. The flow-through from the IP was kept and run alongside the original lysates, confirming the initial presence of 3xFLAG-CDX4, as well as

pulldown of CDX4 (**Figure 3.2B**). Additionally, sonication conditions that yielded 100-500 bp fragments useful for next generation sequencing and were compatible with the buffers and cell types used was determined (**Figure 3.2C**). Finally, functional confirmation of the *CDX4*$^{-/y}$, *iCDX4* hPSC was obtained by induction of *CDX4* during mesoderm formation and then functional assessment of erythroid and myeloid potential, as was done previously with the *iCDX4* only line[148]. Hematopoietic methylcellulose colony forming assays (**Figure 3.2D**) revealed that mesoderm induction of *CDX4* was able to increase the functional output of hemogenic endothelium, over non-induced, endogenous *CDX4*$^{-/y}$ KO.

3.5 Finding *CDX4* targets through ChIP/RNAseq

Once the technical conditions for ChIP were found and *CDX4*$^{-/y}$, *iCDX4* hPSC's were functionally confirmed, combined ChIP and RNAseq could be performed on the relevant mesodermal populations. Since a unique, retinol responsive KDR+CD235a-CD184+ hemogenic mesoderm was discovered through observations from our lab (Luff *et al*, in review[142]), analysis of this population and the potential role that *CDX4* could play in its formation was of great interest. FACS isolation of both CD184- ROH independent and CD184+ ROH dependent definitive hemogenic mesoderm would allow for assessment of the role *CDX4* in both of these populations simultaneously. Additionally, understanding how enforced mesodermal expression of *CDX4* results in repression of the primitive program could yield insight into other key mesodermal regulators of definitive/primitive hematopoiesis. Answering these questions will require FACS isolation of three different populations: KDR+CD184- ROH independent, KDR+CD184+ ROH dependent, and KDR+CD235a+ primitive hemogenic mesoderms (**Figure 3.3A**). Paired RNAseq of each of these populations in biological triplicate would allow for correlation of CDX4-bound chromatin with changes in gene expression, to rule out the

possibility of CDX4 binding, but not inducing transcription. A control population without DOX/*CDX4* induction would serve as the control population for ChIP and RNAseq, creating technically paired samples, resulting in a total of six different populations. Because of the technical limitations of sorting large amounts of these mesodermal populations that are approximately ~30-40% of the total numbers of cells within the culture, a ChIP protocol that was designed to use a low input was (in part) adapted for this experimental design[138]. A target population of $0.75x10^6$ cells was sorted for each ChIP replicate and between $0.75\text{-}2.5x10^5$ sorted for each RNAseq replicated.

After the ChIP was performed and the RNA extracted, the samples were submitted for a target of ~30 million reads on an Illumina Novaseq with 150bp paired end reads. The ChIP samples were aligned to the hg38 reference using the STAR aligner[139], and triplicate samples were combined to increase the read counts at any given locus and aid in peak calling. To find peaks where CDX4 protein was bound, the program MACS2 was used to call peaks in CD184-, CD184+, and CD235a+ sets in comparisons to background reads provided by the no DOX control[141]. The RNAseq reads were aligned to hg38 reference using STAR[139] and reads counted and assigned to genes through Subread[140]. Global analysis of regions of significant CDX4-bound regions was visualized using deepTools[150], demonstrating that the majority of CDX4 binding in all three populations was just upstream of the transcriptional start site (TSS), as expected for a transcription factor (**Figure 3.3B**). Interestingly, the KDR+CD235a+ ChIP samples had more dispersed binding just downstream of the TSS when compared to either of the definitive hemogenic mesoderm (CD184-/+). This could possibly be due to the exogenous expression of *CDX4* that is not normally found within this population, perhaps causing spurious binding beyond the TSS.

3.6 *HOXA* and retinoic acid processing genes are *CDX4* targets

To find more precise targets of *CDX4*, significant peaks called by the MACS2 software were fed into GREAT analysis[151], correlating peaks with the closest nearby genes. Within a shorter given range (5 kilobases upstream, 1 KB down), peaks could be associated with more than one gene's TSS, otherwise longer-range interactions were captured by extending the range up to 1000 KB until the TSS of a gene was reached. Approximately 2105 genes were associated with a CDX4 peak in the KDR+CD184- samples, 1435 genes for the KDR+CD184+ samples, and 8757 genes for the KDR+CD235a+ samples. To narrow down these large numbers of gene-associated peaks and determine which are potentially influencing transcription, changes in RNAseq for each sample set were cross referenced (RPKM > 1, fold change > 1.5, p value < 0.05) with these peaks. After this analysis, the CD184- samples had 15 *CDX4* gene putative targets (up and downregulated), the CD184+ samples had 71 targets, and the CD235a+ samples had 892 putative targets (**Table 1**, top 10).

Because of the established relationship of *CDX* and *HOXA* genes in hematopoietic context[105,112,116,117,130,133], the *HOXA* cluster was examined first. Large numbers of significant peaks were found across the majority of the locus, in all three sample sets (**Figure 3.4A**), particular near *HOXA4*, *5*, *7*, and *9*. Interesting, *HOXA* gene expression changes as a result of *CDX4* induction was differential depending what type of hemogenic mesoderm it was expressed in (**Figure 3.4B**). Despite strong peaks near *HOXA4*, there was poor expression in all sample sets and did not change as a result of *CDX4* induction (data not shown), while *HOXA7* was induced in the CD184- mesoderm and *HOXA5/7/9* were induced in the CD184+ mesoderm. The limited change in expression in the CD235a+ population was expected as this population has little to no

HOXA expression under basal conditions and the lack of WNT signaling might result in a lack of cofactors necessary to effect gene transcription.

With the understanding that KDR+CD184-/+ hemogenic mesoderms have established differences in expression of retinol processing enzymes (**Figure 3.1**, Luff *et al.*[142]), genes that were important for ROH signaling were assessed as *CDX4* targets. *CYP26A1*, the gene responsible for breaking down retinoic acid, was found to have a significant peak in the CD184+ mesoderm, but not in the other populations (**Figure 3.4C**). This correlated with a significant increase of *CYP26A1* to comparable levels found in the CD184- retinoic acid independent hemogenic mesoderm, which were not significantly impacted by *CDX4* induction (**Figure 3.4D**). Although the levels of *CYP26A1* expression are still much lower in comparison to CD235a+ primitive hemogenic mesoderm, this suggests that *CDX4* might be influencing the responsiveness of the *CD184+* mesoderm to retinoic acid signaling.

3.7 *CDX4* is regulating *CDX2* and *TBX20*

Investigating the other genes found to be *CDX4* targets within the CD184- mesoderm, where the highest levels of *CDX4* expression are found endogenously, revealed that *CDX2* was a target of *CDX4*. (**Figure 3.5A**), and RNAseq revealed a modest, but significant increase in *CDX2* expression as a result of *CDX4* induction (**Figure 3.5B**). In the mouse embryo, exogenous expression of *Cdx2* was found to bind to the *Cdx4* promoter and induce *Cdx4* expression through WNT mediated feedback[152]. Interestingly, significantly upregulation of *CDX2* was not found in either CD184-/+ mesoderm, possibly because of already high levels of expression due to WNT activation, which is known to induce *CDX* expression[153]. This suggests there might be a *CDX2*-*CDX4* feedback loop, though *CDX2* acting as a transcription factor for *CDX4* has not been established in the context of human hemogenic mesoderm.

Looking at the *CDX4* targets identified in the CD184- mesoderm, it was surprising to see 2 out of 15 genes downregulated as a result of *CDX4* induction, as well as having nearby *CDX4* binding peaks. *CDX* genes are canonically transcription factors that lead to gene induction, though there might be some evidence of *CDX* genes acting as repressor in concert with epigenetic regulators such *BRG1*-SWI-SNF[154], which has been known to occasionally act in a repressor role[155]. One of the two genes identified as *CDX4* repressed was *TBX20*, an important cardiac transcription factor necessary for proper heart structure[156,157]. A significant peak for both CD184- and CD235a+ mesoderm was found downstream of *TBX20* in the intron of another gene (**Figure 3.5C**), interestingly all three mesodermal populations also displayed a decrease in *TBX20* expression as a result of *CDX4* induction (**Figure 3.5D**). Although the same peak was not considered significant for the CD184+ mesoderm, this could be due to noise or variation in samples that lead to the MACS software as not calling it significant. This possible suppression of a cardiac transcription factor in a CD235a+ mesoderm is intriguing, as recent work has shown that ventricular cardiomyocytes arise from a CD235a+ mesoderm in an hPSC differentiation model system[134]. This is supported by studies in mouse embryos/embryonic stem cells[158] as well as the zebrafish[159] that show that *Cdx1/2/4* gene expression negatively regulates cardiogenesis, resulting in disruptions in heart structure and decreased cardiac related gene expression. Since *CDX2* was also upregulated through enforced *CDX4* and both genes are implicated in downregulating various cardiomyocyte transcription factors, examination of the CD235a+ RNAseq alone revealed downregulation (but not CDX4 binding) for *NKX2-5*, *MESP2*, and interestingly *GPYA* (CD235a). *NKX2*-5 is a critical cardiac transcription factor[160] and *MESP1/2* are known to mark mesoderm capable of cardiac specification[92]. Since these genes were not

bound by CDX4, the decrease due to enforced *CDX4* expression (**Figure 3.5E**) is due to indirect effects from *CDX4*, possibly as a result of increased *CDX2* expression.

3.8 *CDX4* expression leads to a decrease in cardiac specification

Having discovered evidence that *CDX4* expression in certain mesodermal populations leads to repression of several cardiomyocyte transcription factors, it became obvious to test whether mesodermal *CDX4* expression could lead to a functional repression in the formation of cardiac progenitors. In order to test this hypothesis, both WT and *CDX4*$^{y/-}$, *iCDX4* hPSC's were differentiated under established conditions for the specification of PDGFRα+CD235a+ ventricular cardiac mesoderm[134]. Mesoderm was specified from days 0-3 under very similar conditions as hemogenic mesoderm, but with lower levels of BMP4 (5 ng/ml) and higher levels activin A (ACTA, 10 ng/ml) and from days 3-5, treatment of the WNT inhibitor IWP2. On day 5, PDGFRα and CD235a expression was assessed via flow cytometry demonstrating that WT and *CDX4*$^{y/-}$, *iCDX4* (with or without DOX from days 2-3) hPSCs were all capable of producing PDGFRα+CD235a+ cardiac mesoderm at ~30% efficiency (**Figure 3.6A**). Seeing no phenotypic differences in the mesoderm produced, qPCR analysis of bulk day 5 culture revealed that expression of *TBX5*, an important cardiomyocyte transcription factor[161], was significantly reduced as a result of *CDX4* expression during days 2-3 (**Figure 3.6B**). Interestingly, there was no difference in *TBX20* or *NKX2-5* expression at day 5, although DOX induced *CDX4* expression had already dropped to nominal levels, perhaps allowing for recovery of expression in these genes. Despite no functional differences in CD235a based on flow cytometry, there was a decrease in *GYPA* (CD235a) expression as a result of *CDX4* induction. This could be due to a

long half-life of CD235a protein, which in some cell types could be up to 48hrs[162], and might be reduced later in the differentiation.

Further maturation of the cardiac differentiations to days 8-12 produced spontaneously beating embryoid bodies (EBs), indicating that cardiomyocytes were present. WT differentiations had sparse numbers of beating EBs, while the *CDX4*$^{y/-}$, *iCDX4* hPSC's had a significantly higher amount, that was reduced when DOX was applied from days 2-3. While informative, this observation was not easily quantifiable, and the presence of cardiomyocytes can be better assessed through intra-cellular flow cytometry of cardiac specific troponin T (cTNT). On day 20, differentiations were harvested, fixed/permeabilized, and stained for CD90 (to exclude endothelium) and cTNT. WT hPSCs had a low level of cardiomyocytes present, at ~5% CD90-cTNT+ cells at day 20 (**Figure 3.7A**), in comparison to ~30% CD90-cTNT+ cells in *CDX4*$^{y/-}$, *iCDX4* differentiations. This expansion in cTNT+ cardiomyocytes was reduced to WT levels after induction of *CDX4* through application of DOX only from day's 2-3 (**Figure 3.7B**), demonstrating that mesodermal *CDX4* expression can suppresses the formation of cardiomyocytes, and its loss leads to their expansion. qPCR of day 20 differentiations also demonstrated a reduction in *cTNT* transcriptionally and expression of *MYL2*, a ventricular specific cardiomyocyte gene[163] (**Figure 3.7C**). Collectively, these results demonstrate that mesodermal *CDX4* leads to a repression in ventricular cardiomyocyte specification, likely directly through the repression of *TBX20* and indirectly through *CDX2*.

3.9 Discussion

The data generated from these ChIP/RNAseq experiments yielded several insights into how *CDX4* regulates the development of multiple mesoderm-derived lineages. As expected based on the literature surrounding *CDX-HOXA* genes in the context of hematopoiesis[112,117-119,130,133],

CDX4 is indeed leading to induction of several *HOXA* genes, in particular medial *HOXA* genes (5/7/9). The medial *HOXA* genes have been demonstrated to be induced in both CD184- retinoic acid independent hemogenic mesoderm (*HOXA7*) and within CD184+ retinoic acid dependent mesoderm (*HOXA5/7/9*). This differential response to induction of *CDX4* suggests that other cell intrinsic factors could be influencing which *HOXA* genes are being induced when *CDX4* binding within the locus. The *HOXA* locus has complicated gene regulatory mechanisms that influence which genes are induced, generally in a colinear manner starting with *HOXA1-HOXA13* in an anterior to posterior pattern within in the embryo[105]. This is thought, in part to be due to CTCF-mediated looping that results in PRC-dependent gene silencing and further impacted by retinoic acid signaling[164]. Since neither of these three mesodermal progenitor populations was exposed to ROH or retinoic acid before being harvested for sequencing, it is unlikely at this particular stage of development that ROH signaling is influencing the population specific *HOXA* gene expression.

Based on the basal level of *HOXA5* and *HOXA9* expression in the KDR+CD184- uninduced *CDX4*$^{y/-}$, *iCDX4* mesoderm, it appears that *CDX4* induction in the KDR+CD184+ mesoderm resulted in expression levels that match KDR+CD184- mesoderm. *CYP26A1* expression was also induced to a similar level as the KDR+CD184- population, when *CDX4* was enforced in the KDR+CD184+ mesoderm. These data suggest that when *CDX4* is exogenously expressed within KDR+CD184+ retinoic acid dependent mesoderm, it takes on a *HOXA*/*CYP26A1* gene expression profile similar to CD184- retinoic acid independent mesoderm. Although extensive functional characterization would be needed, this suggests CD184+ mesoderm could be converted into a ROH-independent hemogenic mesoderm, where the exogenous expression of *CDX4* causes an override of the ROH-dependent program. Genetic-epistasis studies would be

needed to fully confirm the population-dependent nature of induction of *HOXA5/7/9* by *CDX4*, which could be addressed through complementation of the *CDX4*$^{y/-}$ KO with inducible *HOXA5/7/9* constructs.

Perhaps the most unexpected gene target of *CDX4* was *TBX20*, an important cardiac transcription factor[156,157], which was consistently downregulated by *CDX4* expression in all three mesodermal populations. The most striking *CDX4* binding peak however, was the CD235a+ mesoderm, which having been shown to be the progenitors for ventricular cardiomyocytes[134], leading me to hypothesize that *CDX4* could be repressing the cardiomyocyte program. Having demonstrated that mesodermal *CDX4* expression leads to the suppression of cardiac specification, it is still unclear if this suppression is due to a lack of contribution from KDR+CD235a+ mesoderm. In cardiac differentiations, day 5 is when CD235a+ cells are assessed, after WNT inhibition from days 3 to 5. It is unclear whether CD235a+ mesoderm maintains CD235a expression and gives rise to the CD235a+ cells found on day 5, or whether this is a different population that emerges later. FACS isolation of KDR+CD235a+ mesoderm on day 3, followed by continued culture could address the question of where ventricular cardiomyocyte progenitors are coming from on day 3.

Another unaddressed question in these experiments was the role *CDX4* plays in the specification of atrial cardiomyocytes, which have been shown to be dependent on ROH-signaling for their formation[134]. Since CD184+ mesoderm has been shown by our lab to be dependent on ROH for the specification of hematopoietic progenitors, perhaps this population also contains the progenitors for atrial cardiomyocytes[134]. Again, FACS isolation of CD184+ mesoderm under cardiogenic conditions on day 3, followed by treatment of ROH could perhaps answer this; the role of *CDX4* in ROH-dependent hematopoiesis has also not been established.

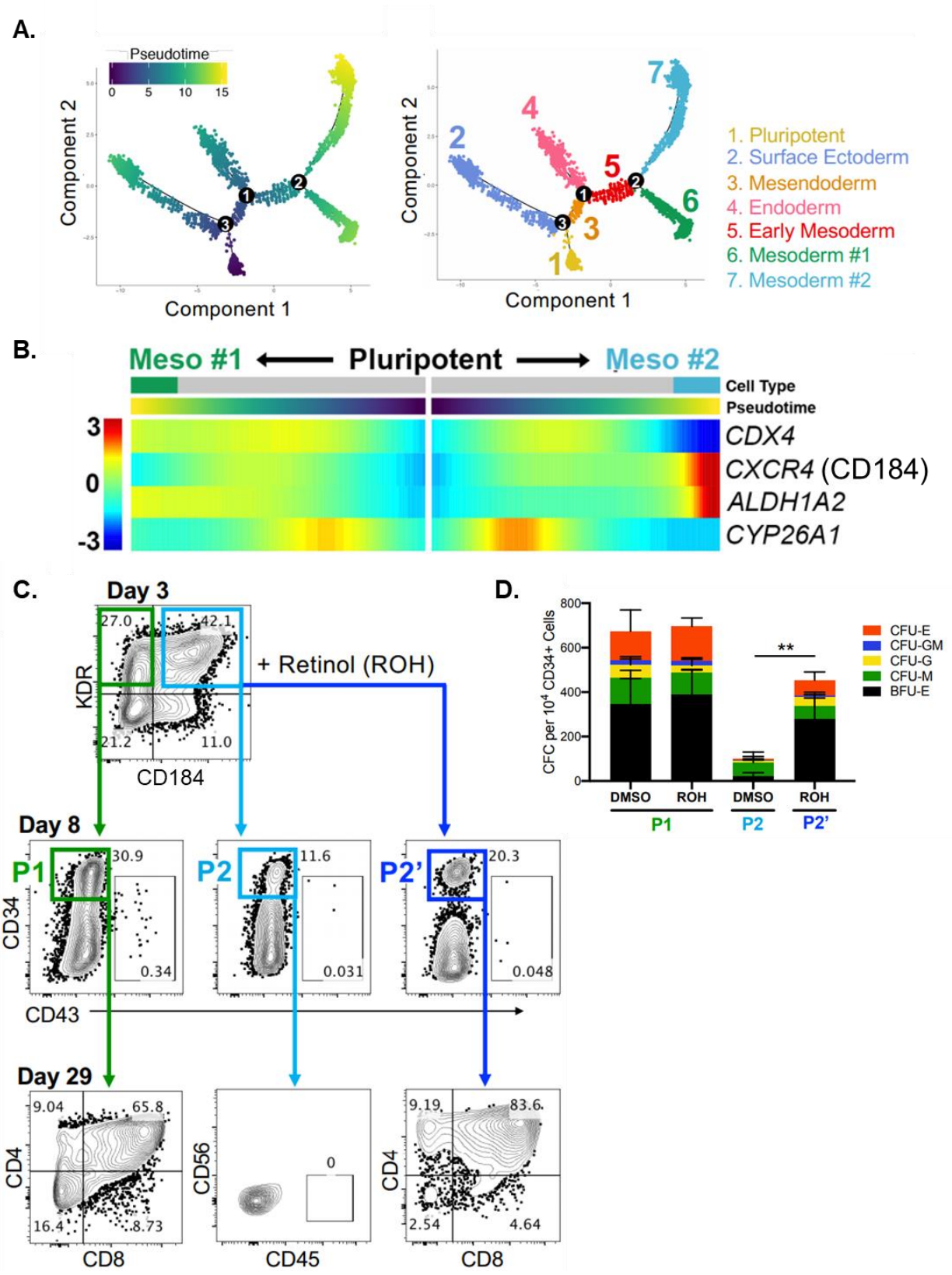


Figure 3.1: CD184 expression marks a retinol responsive, *CDX4*^{lo} mesodermal population

(from Luff *et al*., in review[142]) **A.** scRNAseq was performed on bulk culture treated with CHIR99021 and SB431542 (TGFβ inhibitor) at day 3, after the specification of mesoderm to better assess the culture heterogeneity and to find hemogenic mesoderm. Briefly, the cells were dissociated via trypsin and fixed in methanol, before being processed via 10X Genomics Chromimum 3' kit, capturing >6000 cells. The libraries were sequencing via an Illumina HiSeq 3000 and processed via the Cell Ranger pipeline. The R package Monocle2 was used to generate an unsupervised pseudotime plot with branchpoints. **B.** A heatmap of expression in cells across the peusdotime plot, ranging from population 1 (pluripotency) to either mesoderm populations identified after the branchpoint, showing an inverse correlation between *CDX4* and *CXCR4* (CD184) and *ALDH1A2*. **C.** Flow cytometry of day 3 mesoderm after CHIR99021/SB431542 treatment, showing two different mesodermal populations, CD184- (P1) and CD184+ (P2). FACS isolation of these two populations and continued culture shows both give rise to CD34+ endothelium populations, which there then assessed for lymphoid potential by further coculture with OP9-DL4 for 29 days and harvested for flow cytometry. **D.** All three populations were FACS isolated at day 8 (CD34+) and placed into an EHT assay for an additional 9 days and then plated into hematopoietic methylcellulose media. The numbers of colonies were then assessed after 12-14 days and the numbers were counted of burst forming units erythroid (BFU-E) and colony forming units (CFU) of erythroid (E), granulocyte (G), myeloid (M), and mixed granulocyte/myeloid (GM). n = 3, ** $p < 0.01$ via students t test of BFU-E. All Error bars SEM.

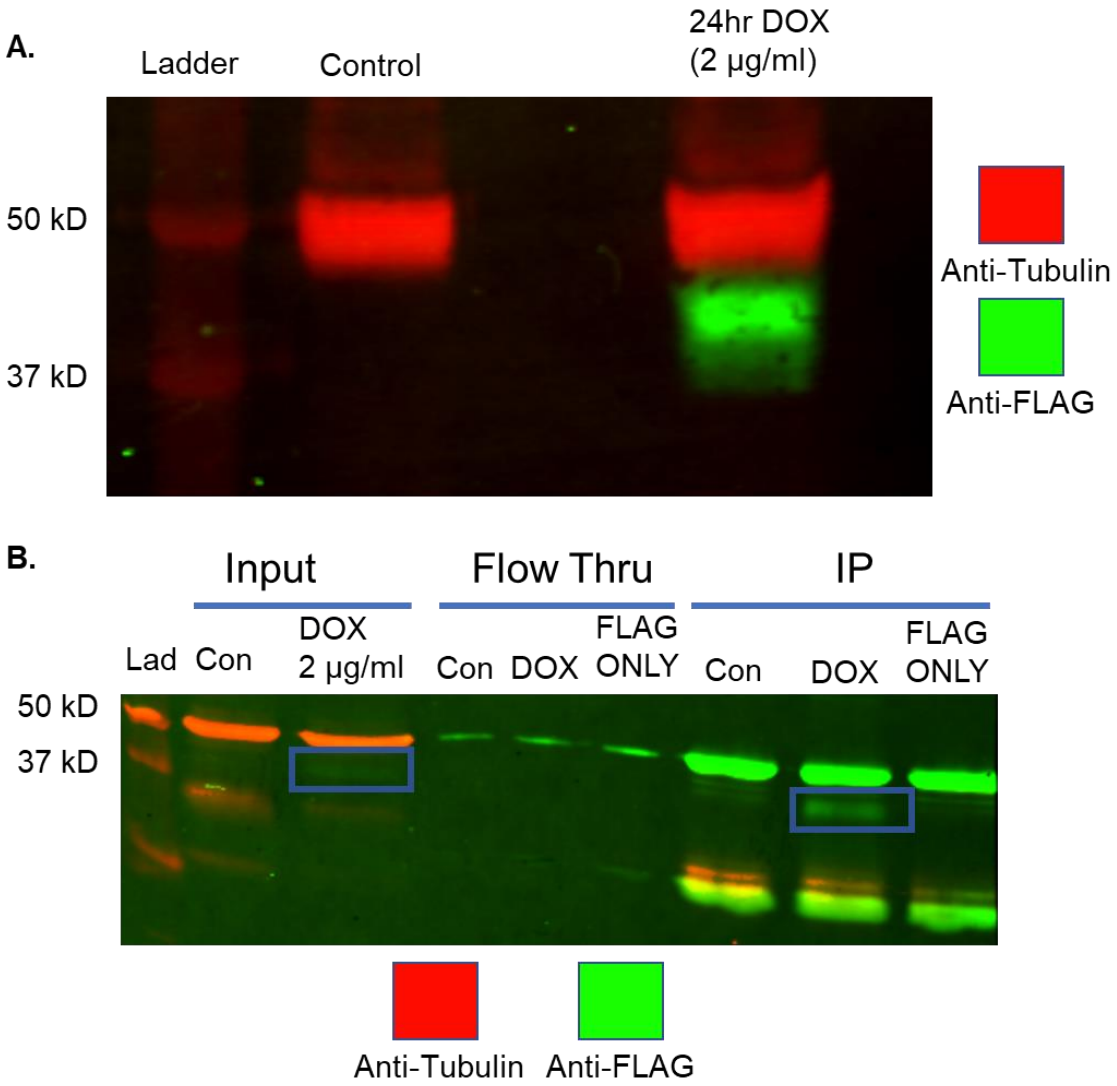
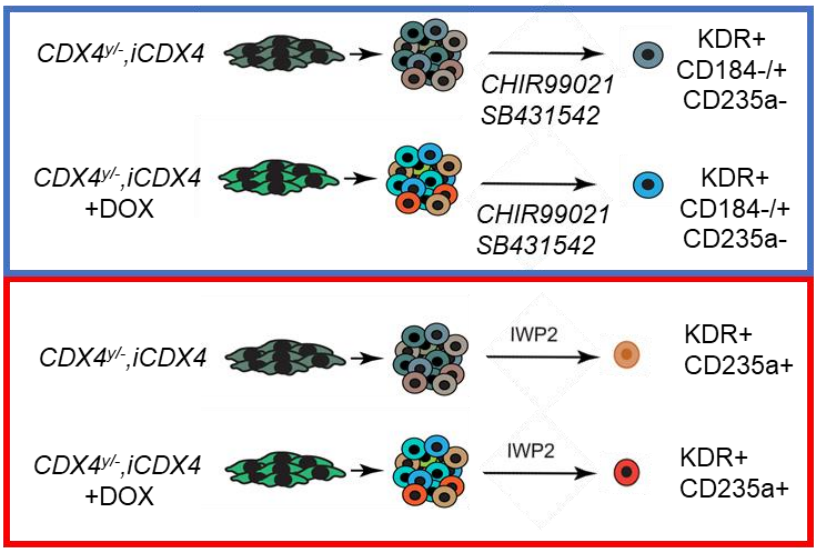


Figure 3.2: Doxycycline inducible, FLAG-CDX4 can be immunoprecipitated effectively in day 3 mesoderm

A. Protein isolated from *CDX4*$^{y/-}$, *iCDX4* hPSCs was isolated and quantified, with or without 24 hours of 2 μg/ml doxycycline (DOX) and 20 μg of protein was loaded to each well from the corresponding sample for SDS-PAGE. After western blot transfer, the membrane was blocked and probed with anti-Tubulin and ant-FLAG primary antibodies, IRDye secondary antibodies, and imaged on a LICOR Odyssey. **B.** Western blot of protein (20 μg) isolated from day 3 bulk differentiation from *CDX4*$^{y/-}$, *iCDX4* hPSCs for immunoprecipation (IP), which was preformed using anti-FLAG antibody and Protein G bound magnetic beads. DOX treatment was performed for 30 hours, from Day 2 to Day 3. Input denotes protein samples pre-IP, flow thru denotes the supernatant removed after the IP, and IP denotes the protein removed from the beads. CDX4 bands highlighted in blue. **C.** TapeStation size analysis of genomic DNA fragments of day 3 differentiated *CDX4*$^{y/-}$, *iCDX4* hPSCs after lysis and formaldehyde crosslinking. Far left lane ladder, middle control with no DOX added, and 2 μg/ml of DOX from days 2-3. **D.** *CDX4*$^{y/-}$, *iCDX4* hPSCs were differentiated as described before with CHIR99021 (CHIR), with or without the addition of 2 μg/ml DOX from days 2 to 3 and continued until day 8 of culture. The cells were harvested and CD34+CD43-CD184-CD73- HE was sorted as previously described and plated and reaggregated for an additional 9 days in EHT culture. Cells were then harvested and plated in hematopoietic methylcellulose media for 10-12 days and the numbers were counted of burst forming units erythroid (BFU-E) and colony forming units (CFU) of erythroid (E) and myeloid (M) colonies. * indicates $p < 0.05$ via student's T test between BFU-E, n =5, error bars SEM.

A.



B.

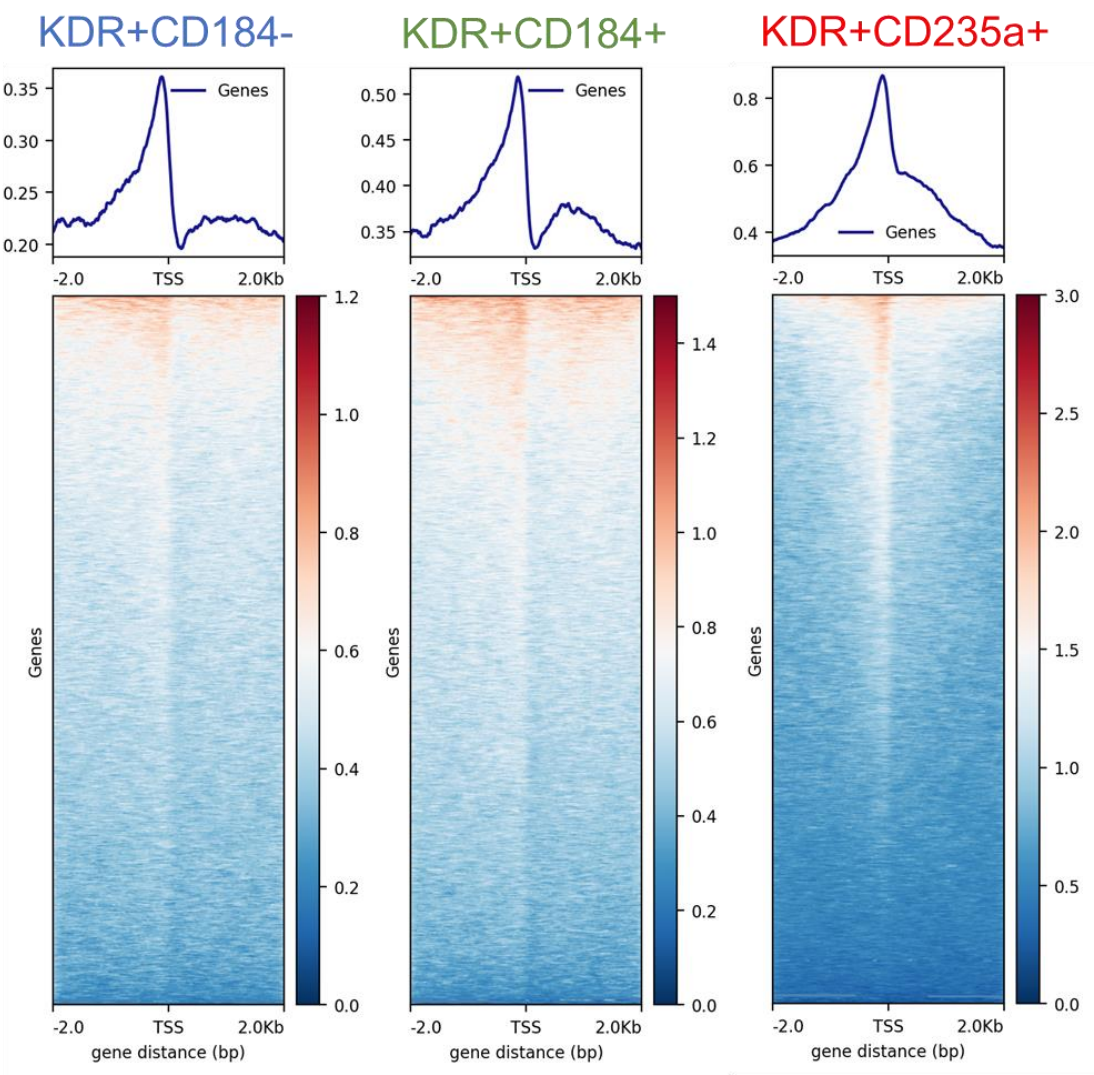


Figure 3.3: Finding CDX4 gene targets through ChIP/RNAseq

A. Schematic of the different populations isolated for paired ChIP/RNAseq; the *CDX4*$^{y/-}$, *iCDX4* hPSC line was cultured under four different conditions: CHIR99021/SB431542 treatment to induce definitive hemogenic mesoderm, with or without 1 μg/ml doxycycline (DOX) from days 2-3 to induce *iCDX4*, or IWP2 treatment to induce primitive hemogenic mesoderm, with or without DOX. KDR+CD235a-CD184- and KDR+CD235a-CD184+ cells were then FACS isolated from the definitive cultures and KDR+CD235a+ isolated from the primitive cultures. These six populations were isolated in biological triplicate and subjected to paired RNAseq/ChIPseq via anti-FLAG pulldown of CDX4 protein. **B.** After alignment of the ChIPseq reads, the software package deepTools was used to generate heatmaps of pulled down reads across all known genes in the genomes by taking each of the three populations respective populations and normalizing to the paired biological control (no DOX), where red indicates greater relative enrichment and blue low enrichment. The histogram plot at top was generated by taking the relative enrichment at positions from -2 to +2 kilobases from the transcriptional start across all genes, scale arbitrary.

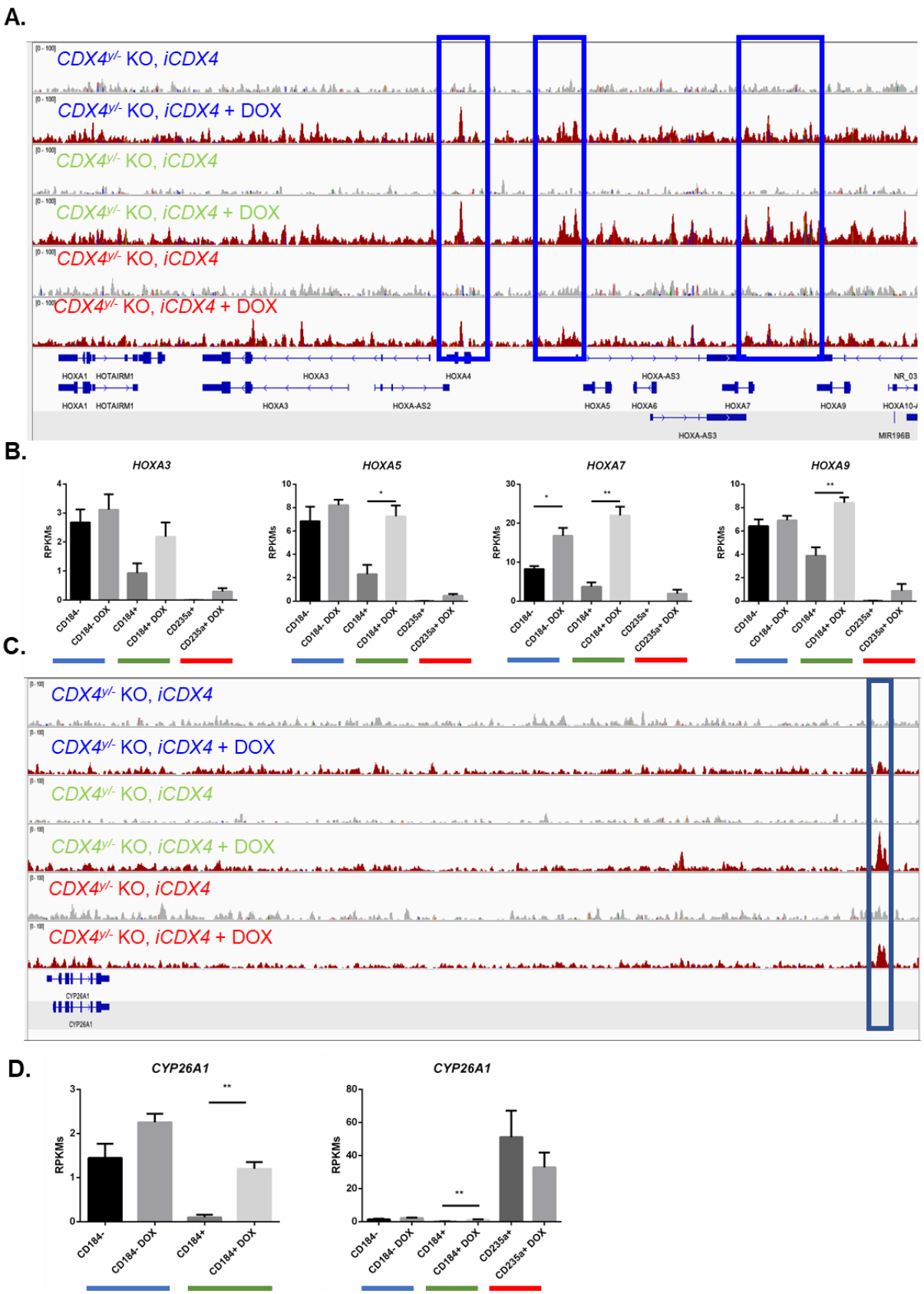


Figure 3.4: *CDX4* is binding at *HOXA* and *CYP26A1* loci, but inducing expression differentially across populations

A. Genome browser view of the *HOXA* cluster found on chromosome 7, displaying the *CDX4* ChIP reads for KDR+CD184- (Blue), KDR+CD184+ (Green), and KDR+CD235a+(Red) samples, with the CDX4 DOX induced samples read pileup in maroon, compared to controls in grey. Highlighted are several regions denoting highly significant peaks called by the MACS2 software in the samples when compared to the no DOX controls. **B.** Gene expression of *HOXA* genes in RPKMs from the RNAseq paired with the ChIP samples; KDR+CD184- highlighted in blue, KDR+CD184+ highlighted in green, and KDR+CD235a+ highlighted in red. n = 3, error bars SEM, * is $p < 0.05$, ** $p < 0.01$ via students T test, error bars SEM. **C.** Genome browser view of the *CYP26A1* locus on chromosome 10, displaying read pileups from ChIPseq performed on KDR+CD184 (Blue), KDR+CD184+ (Green), and KDR+CD235a+(Red) samples. Sample tracks with DOX induced FLAG-CDX4 highlighted in maroon, compared to controls in grey. MACS2 significant peaks highlighted in blue. **D.** Gene expression of *CYP26A1* in RPKMs from the RNAseq paired with the ChIP samples; KDR+CD184- highlighted in blue, KDR+CD184+ highlighted in green, and KDR+CD235a+ highlighted in red. Left plot shown without KDR+CD235a+ samples and right plot with. n = 3, error bars SEM, ** is $p < 0.01$ via students t test, error bars SEM.

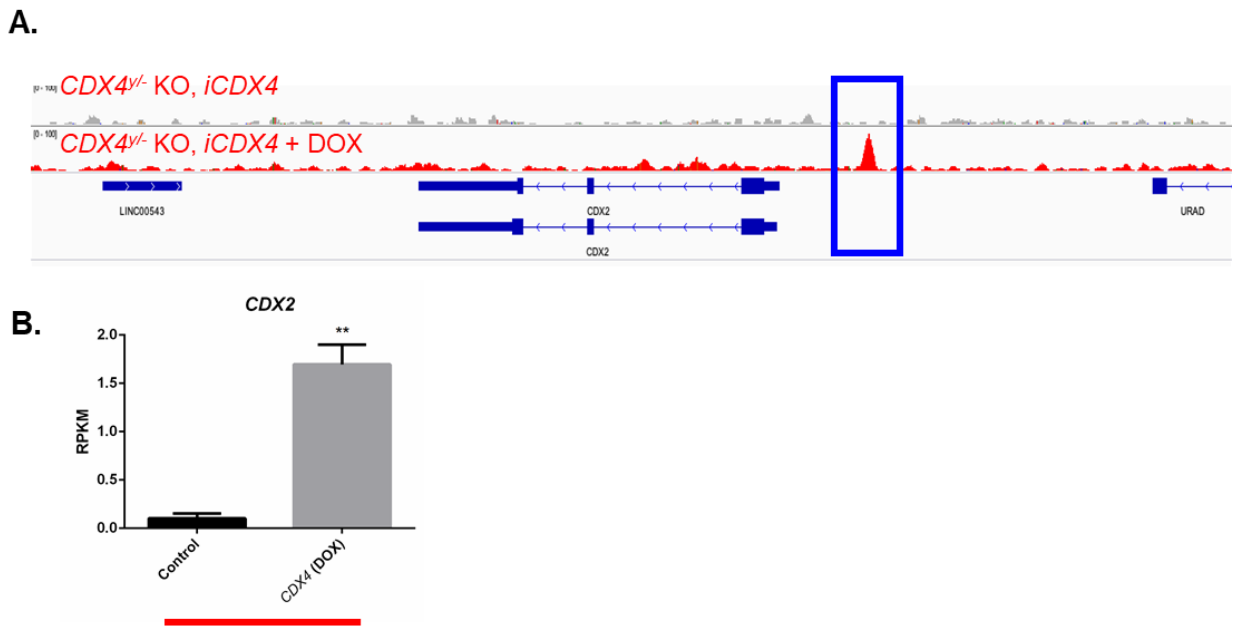


Figure 3.5: *CDX4* is regulating *CDX2* and *TBX20*, a cardiac transcription factor

Genome browser view of the *CDX2* locus on chromosome 13, displaying read pileups from ChIPseq performed on KDR+CD235a (Red). Sample tracks with DOX induced FLAG-CDX4 highlighted in maroon, compared to controls in grey. MACS2 significant peak highlighted in blue. **B.** Gene expression of *CDX2* in RPKMs from the RNAseq paired with the ChIP samples KDR+CD325a+ highlighted in red. n = 3, error bars SEM, * is $p < 0.05$ via students t test, error bars SEM. **C.** Genome browser view of the *TBX20* locus on chromosome 7, displaying read pileups from ChIPseq performed on KDR+CD184- (Blue) and KDR+CD235a+ (Red). Sample tracks with DOX induced FLAG-CDX4 highlighted in maroon, compared to controls in grey. MACS2 significant peaks highlighted in blue. **D.** Gene expression of *TBX20* in RPKMs from the RNAseq paired with the ChIP samples KDR+CD184- highlighted in blue, KDR+CD184+ in green, and KDR+CD235a+ in red. n = 3, error bars SEM, * is $p < 0.05$ via students t test, error bars SEM. **E.** Gene expression of various cardiac related transcription factors in RPKMs from the RNAseq paired with KDR+CD235a+ ChIP samples. n = 3, error bars SEM, * is $p < 0.05$, * $p < 0.01$ via student t test, error bars SEM.

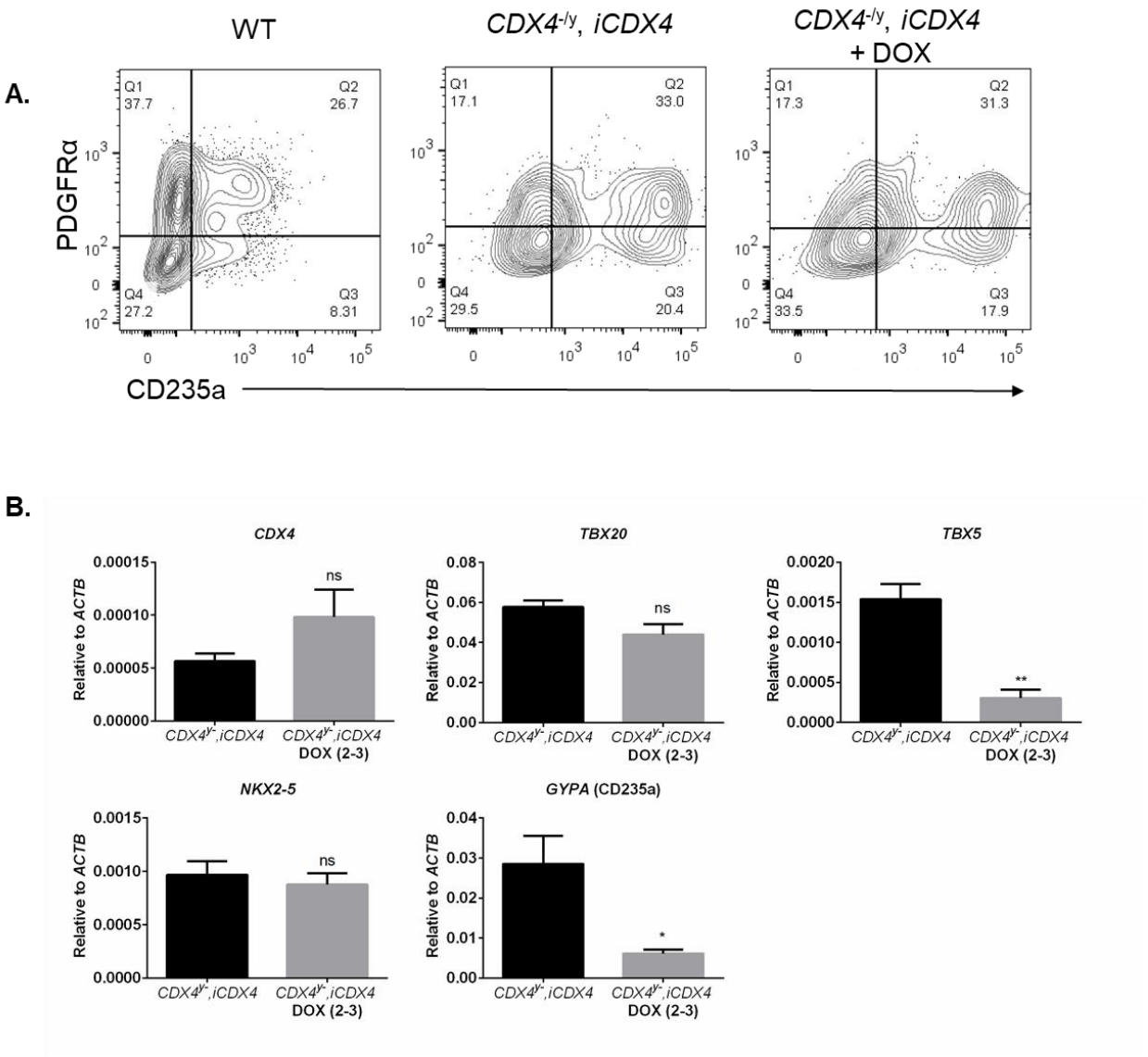


Figure 3.6: CDX4 expression during mesoderm formation leads to repression of cardiac factors under cardiogenic conditions

A. WT and *CDX4*$^{y/-}$,*iCDX4* hPSCs (w/ or w/out 1 μg/ml DOX from days 2-3) were differentiated under cardiogenic conditions under hypoxia similar to previous hematopoietic conditions, with the same media base except alterations to cytokine concentration. **B.** Bulk culture of *CDX4*$^{y/-}$,*iCDX4* hPSCs (w/ or w/out 1 μg/ml DOX from days 2-3) was harvest at day 5 for RNA and

cDNA generated for RT-PCR. All genes were normalized relative to *ACTB* expression, n = 3, * is $p < 0.05$, ** is $p < 0.01$, ns is not significant, via students t test, error bars SEM.

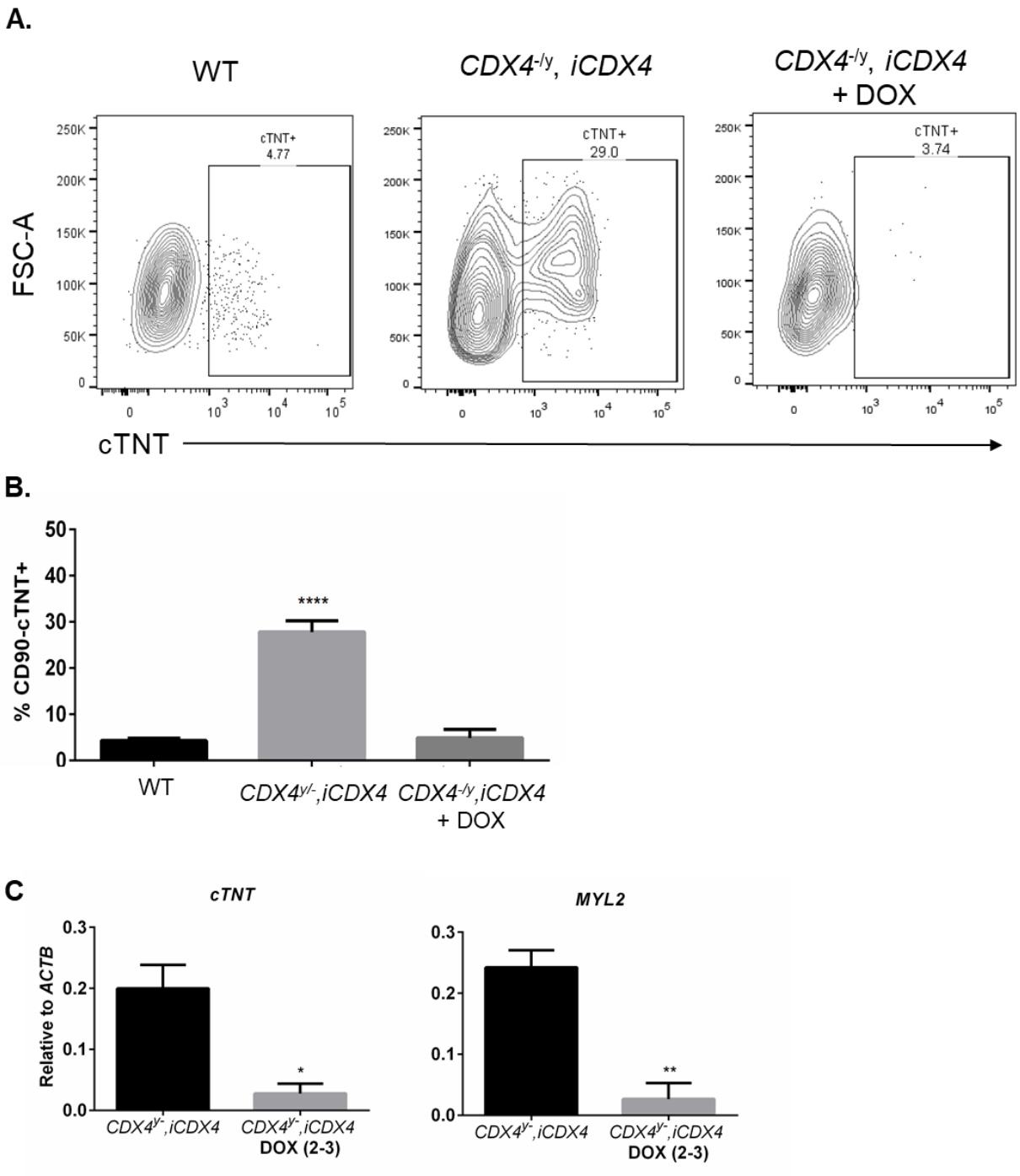


Figure 3.7: Mesodermal *CDX4* expression leads to suppression of cardiomyocyte formation

A. WT and *CDX4*$^{y/-}$,*iCDX4* hPSCs (w/ or w/out 1 μg/ml DOX from days 2-3) were differentiated under cardiogenic conditions under hypoxia similar to previous hematopoietic conditions, with the same media base except alterations to cytokine concentration. At day 20, cells were disassociated, fixed/permeabilized and then stained for CD90 and cTNT for flow cytometry, shown are all cells negative for CD90. **B.** Quantification of flow cytometry on day 20 differentiated WT and *CDX4*$^{y/-}$,*iCDX4* hPSCs (w/ or w/out 1 μg/ml DOX from days 2-3) for CD90-cTNT+ cells. WT n =2, *CDX4*$^{y/-}$,*iCDX4* n = 5. **** indicates $p < 0.0001$ between control and DOX treated *CDX4*$^{y/-}$,*iCDX4* hPSCs via students t test, error bars SEM. **C.** T20 cardiac differentiations from *CDX4*$^{y/-}$,*iCDX4* hPSCs (w/ or w/out 1 μg/ml DOX from days 2-3) were harvested for qPCR. Expression normalized to *ACTB*, n = 3, * $p < 0.05$, ** $p < 0.01$, error bars SEM.

Table 3.1: List of *CDX4* target genes in different hemogenic mesoderm populations

For each population, the genes that were associated (GREAT analysis[151]) with significant CDX4 peaks (MACS2[141]) and had a 1.5 fold change in RPKMs (>1, $p < 0.05$) when compared to their respective control population are listed below, segregated into genes that are upregulated with *CDX4* induction and those that were downregulated.

KDR+CD184-		KDR+CD184+		KDR+CD235a+	
Upregulated with *CDX4*	Downregulated with *CDX4*	Upregulated with *CDX4*	Downregulated with *CDX4*	Upregulated with *CDX4*	Downregulated with *CDX4*
GNRH2	ALPK2	ABCG2	ALDH1A2	43891	43901
GREB1	NFE2	ANGPT1	ALPK2	43893	ABTB2
HOXA7	SLN	ARHGAP10	HAS2	AAMDC	ACKR3
LMO2	TBX20	ARID5B	LRRN1	ABCC5	ADA
NAV1		C3orf52	PITX2	ABCG2	ADGRD1
PC		CEP97	PODXL	ABHD12	ALG1L
PDE4D		CYP26A1	RSPO2	ABHD5	ALPK2
PLXNA2		DHRS3	SBSPON	ADD2	AMOTL1
TGFBI		DUSP5	TBX20	ADGRE5	ANGPT2
TMPRSS11E		ELL2		ADGRL4	ANKH
TRPA1		EPB41L4A		ADIPOR1	ANKRD1
		EPHA4		AGPS	ANXA8L1
		ERC2		AGTPBP1	APCDD1L
		ETV1		AHCYL2	APLNR
		FAM188A		ALCAM	AQP3
		FST		ALDH1A1	ARHGAP32
		FZD6		ALDH3A2	ARHGAP40
		GAP43		ALDH4A1	ARL4D
		GOLIM4		ALDH6A1	ASH2L
		GREB1		ALDH9A1	B3GALT1
		GUCY1A3		ALS2	BAG3
		HAND2		AMIGO2	BCAR3
		HOXA5		ANKRD6	BIN1
		HOXA7		AP2B1	BMP2
		HOXA9		AP4E1	BMP4
		HOXB4		APLP1	BMPER
		HOXB7		APOL2	BRINP1
		HOXC4		APOOL	C14orf132
		HTR2B		ARHGAP10	C1GALT1

Chapter 4: CD1d marks *CDX4*+ mesoderm

4.1 Abstract

Having shown that *CDX4* is a critical regulator of definitive hemogenic mesoderm[148], a critical question remained: does *CDX4*+ mesoderm ultimately give rise to definitive hematopoietic mesoderm, or does it's expression act in supportive manner in other cell types that do not directly give rise to hematopoiesis? Creating a *CDX4* reporter that would label all *CDX4*+ cells with a fluorescent protein would allow for the purification of *CDX4* expressing mesoderm through flow cytometry and to ask which lineages they ultimately contribute to. Although the generation of a triple mCherry-*CDX4* reporter was ultimately unsuccessful and resulted in KO of *CDX4*, single cell transcriptomics allowed for the characterization of *CDX4* expressing mesoderm and the discovery of a surface marker, *CD1D*, that correlated strongly with *CDX4* and *HOXA* genes. Isolation of KDR+CD1d-/+ mesoderm under definitive hematopoietic conditions (CHIR99021+SB431542) confirmed that KDR+CD1d+ mesoderm was greatly enriched for expression of *CDX4* (~20 fold), as well as *HOXA7*, *HOXA9*, and *CDX1*. Although some mildly CD1d+ cells were found within KDR+CD235a+ primitive hemogenic mesoderm, functional characterization did not reveal any significant differences in potential, and qPCR of *CD1D* demonstrated that it was not appreciably expressed in this population. Both CD1d- and CD1d+ definitive hemogenic mesoderm gave rise to CD34+ endothelium, but only CD1d+ progenitors were able to give rise to CD4/CD8+ T-cells and were significantly enriched for erythroid and myeloid colony forming progenitors. These data demonstrate that definitive hematopoietic

potential is almost entirely restricted to CD1d/*CDX4*+ mesoderm and that CD1d is positive marker for definitive hemogenic mesoderm.

4.2 Methods

Culture and differentiation. Hematopoietic differentiations and functional assays were performed according to the methods in **Chapter 2**, with the addition of SB431542 (6 μM) alongside CHIR9902171.

Genome engineering of hPSCs. To generate the 3xmCherry-*CDX4* reporter line, the same *CDX4* CRISPR guides were used as before in **Chapter 2**, but also with plasmid based HDR template with homology arms containing the in-frame 3xmCherry construct.

scRNA-seq analyses. The scRNAseq datasets were analyzed as described in **Chapter 2**, but pseduotime analysis was omitted.

Gene Expression analyses. qRT-PCR was performed as described in **Chapter 2**.

Flow Cytometry and Cell Sorting. Flow cytometry was performed for hematopoietic differentiations as previously described, using methods from **Chapter 2**. CD1d antibody (Clone CD1d42) was obtained from BD Pharmigen.

4.3 Development of a *CDX4* fluorescent reporter

In order to address lineage questions of *CDX4* expressing mesoderm, a way to isolate and purify *CDX4*+ cells was needed. Due to technical problems associated with using traditional CRE-LOX recombination systems used in many lineage tracing studies in hPSCs[165,166], an in-frame fluorescent *CDX4* reporter design was selected. A design with triple repeat mCherry

(3xmCherry) and E2A linkers was used[167], to aid in robust fluorescence with lower levels of expression. The construct was designed to be inserted at the 5' end of *CDX4* via homologous recombination after CRISPR Cas9 mediated double strand breaks (**Figure 4.1A**). After PCR genotyping and sanger sequencing confirming correct insertion of the construct, the 3xmCherry hPSC line was differentiated under basal, high WNT (CHIR99021), and low WNT (IWP2) conditions and the presence of mCherry fluorescence assessed via microscopy and flow cytometry (**Figure 4.1B**). Treatment with CHIR should have induced *CDX4*-dependent mCherry expression but unexpectedly, there was no mCherry signal under any culture conditions. Continued culture with the 3xmCherry line to day 8 revealed there was a distinct lack of CD34+CD43-CD73-CD184 hemogenic endothelium (HE) under all conditions. Taken together, these observations suggest there was a disruption in transcription for endogenous *CDX4* as a result of insertion of the 3xmCherry construct, as the *CDX4*$^{y/-}$ KO hPSCs also displayed a similarly large reduction in HE. This was likely due to insertion at the 5' end of *CDX4*, possibly disrupting a regulatory element and would likely have functioned better with a 3' insertion.

4.4 Identifying *CDX4*+ mesoderm with scRNAseq

With the 3xmCherry-*CDX4* reporter disrupting the locus and not displaying any mCherry expression at expected times during the differentiation, other methods were needed to be able to address the potential lineages arising from *CDX4*+ mesoderm. Our lab had recently performed single cell transcriptomic analysis on our bulk day 3 cultures under definitive (CHIR99021+SB431542) and primitive (IWP2) hematopoietic conditions. This powerful dataset would allow for precise transcriptional analysis of *CDX4*+ cells in the CHIR+SB (CSB) dataset and could be used to search for potential surface markers that could be used to isolate *CDX4* expressing mesoderm. After the cells were processed using the 10X Chromium system, the

sample sets were sequenced, and the reads aligned and counted using the 10X CellRanger package. With >6,000 cells identified in each dataset after removing low gene count and normalized to percent mitochondrial gene expression, the CSB and IWP2 datasets were integrated using the Seurat software package[136], and a UMAP generated to place similar cells together in two dimensional space. The Seurat clustering algorithm determined that there were 11 distinct clusters of cell types within the integrated UMAP (**Figure 4.2A**), with variable contribution from either dataset to each cluster.

In order to determine which cell types were represented within each of the clusters within the dataset, the expression of several germ layer markers was assessed: *NANOG* and *SOX2* for pluripotent/primitive streak cells, *KRT7*, *DLX5*, and *TFAP2A* for ectoderm, *HNF1B*, *SOX17*, and *FOXA2* for endoderm, and *KDR/T* for mesoderm (**Figure 4.2B**). This analysis revealed there were 5 distinct mesodermal clusters, one of which was composed of *GYPA/B/E* (cluster 5), which was comprised mainly of cells from the IWP dataset, suggesting these cells are the primitive hemogenic mesoderm. Of the remaining four mesoderm clusters, *CDX4* expression was predominately found within clusters 1 and 2, with some expression found within one of the ectoderm clusters (**Figure 4.2C**). Having identified the mesodermal cell clusters that are *CDX4+*, gene expression analysis could be performed on these clusters using only the cell from the CSB dataset, enriching for the possibility of identifying definitive hemogenic mesoderm. When cluster 1 cells were compared to the rest of the cells in the CSB dataset, differentiation gene analysis revealed a number of *HOXA* genes, as well *CDX1/2* (**Table 4.1**). Interestingly, one of the top 30 genes identified was *CD1D*, a non-canonical MHC receptor found on antigen presenting cells that binds to various lipid/glycolipid ligands[168]. It is unclear why *CD1D* would be expressed in developing hemogenic mesoderm, as there would be no immune cells in the

developing embryo at an equivalent timepoint. However, the expression of *CD1D* was highly correlated with *CDX4* (**Figure 4.2D**), suggesting that CD1d antibodies could be used to help purify *CDX4*+ cells from mesoderm via FACS. Additionally, when all KDR+ cells were examined for *CD1D* expression, it was found that *CDX4*+*HOXA*+ cells had the highest proportion of *CD1D*+ cells, over the only *CDX4*+ or only *HOXA*+ cells alone (**Figure 4.2E**). While there are *CD1D*+ cells that are not *CDX4*+ and vice versa, these data suggest that FACS isolation of CD1d+ cells could be used to enrich for *CDX4*+ cells.

4.5 Phenotypic characterization of CD1d+ mesoderm

To test this new potential marker of *CDX4* expressing mesoderm, day 3 WT hPSCs differentiated under definitive (CSB) and primitive (IWP2) hemogenic conditions were harvest for flow cytometry. KDR+CD235a- mesoderm was assess for CD1d expression, where approximately 60% of the cells were positive for CD1d gated on FMO controls (**Figure 4.3A**). Despite low levels of *CD1D*+ cells in the IWP2 scRNAseq dataset, KDR+CD235a+ hemogenic mesoderm also appeared to have some CD1d+ cells, though fewer than KDR+CD235a-. To confirm the hypothesis that CD1d expression would be surrogate marker for *CDX4*, KDR+CD235a-CD1d-/+ (CSB) and KDR+CD235a+CD1d-/+ (IWP2) cells were FACS isolated and qPCR performed. *CDX4* was highly enriched (~20 fold) in KDR+CD1d+ mesoderm over KDR+CD1d- (**Figure 4.3B**), validating the previous observations in the scRNAseq dataset. The CSB derived KDR+CD235a-CD1d+ sorted cells were also confirmed as expressing *CD1D* at the transcript level via qPCR, but interestingly the IWP2 derived KDR+CD235a+CD1d+ cells did not have a significant difference in *CD1D* expression over the KDR+CD235a+CD1d- cells. This suggests either that *CD1D* is being downregulated in this population and only residual protein is being detected or that the CD1d+ cells are a potential artifact of flow cytometry. The

KDR+CD235a-CD1d+ cells also had increased expression compared to the KDR+CD235a-CD1d- cells for *HOXA7*, *HOXA9*, and *CDX1*, again confirming the informatics analysis that suggested *CD1D*+ cells would be enriched for *CDX/HOXA* expression (**Figure 4.3C**).

4.6 Functional characterization of CD1d+ mesoderm

Having demonstrated that CD1d+ mesoderm strongly enriches for *CDX4*+ mesoderm, it is now possible to ask whether *CDX4*+ mesoderm gives rise to hematopoietic lineages by using CD1d as a surrogate surface marker. CD1d-/+ cells were FACS isolated on day 3 of differentiation from KDR+CD235a- (CSB) definitive KDR+CD235a+ (IWP) primitive hemogenic mesoderm, followed by reaggregation and continued culture until day 8. Flow cytometry on day 8 of these four populations revealed a production of mainly CD43+ hematopoietic progenitors from KDR+CD235a+ mesoderm and the production of CD34+CD43- endothelium from KDR+CD235a- mesoderm, in line with previous observations of these populations[85] (**Figure 4.4A**). To assess which of these populations could give rise to lymphoid cells, indicative of definitive hematopoietic potential, either CD34+CD43+ progenitors (IWP2) or CD34+CD43- progenitors (CSB) were FACS isolated and cultured with OP9-DL4 stroma under T-cell promoting conditions. After 22 additional days in culture, flow cytometry of each respective population revealed that only the KDR+CD235a-CD1d+ mesoderm was ultimately capable of giving rise to CD45+CD56-CD4/8+ T-cells. Interestingly, some CD56+ cells were generated in OP9-DL4 culture from the KDR+CD235a+CD1d+ culture. This suggests that perhaps this population is capable of giving rise to CD56 natural killer (NK) cells, and it has been previously demonstrated that CD34+CD43+ WNT independent primitive hematopoietic progenitors can give rise to NK cells[34].

Having demonstrated that lymphoid potential was restricted to KDR+CD235a-CD1d+ mesoderm, myeloid and erythroid lineages also needed to be assessed, to determine whether KDR+CD235a-CD1d+ is exclusive for definitive hemogenic mesoderm. Additionally, it was unclear if CD1d was a functional marker for segregating any types of primitive hemogenic mesoderm. CD34+CD43+ progenitors were FACS isolated on day 8 from reaggregated KDR+CD235a+CD1d-/+ mesoderm and plated into hematopoietic methylcellulose media and colonies assessed 8-10 days later. There was no significant difference in myeloid or erythroid colony forming units between KDR+CD235a+CD1d- and KDR+CD235a+CD1d+ mesoderm, with only a slightly larger amount of granulocyte forming colonies (**Figure 4.4B**). When KDR+CD235a-CD1d-/+ derived CD34+CD43- endothelium was FACS isolated on day 8 and placed into EHT culture for an additional 9 days[72,87], the CD1d+ derived endothelium produced significantly more floating hematopoietic progenitors compared the CD1d- derived endothelium, indicating that more cells were likely undergoing the EHT. When the EHT cultures were harvested and plated into hematopoietic methylcellulose media and colonies counted after 12-14 days, the majority of erythroid and myeloid colonies were restricted to the CD1d+ mesoderm, with 3/6 replicates being completely absence of any hematopoietic colonies. Taken together, these data demonstrate that CD1d is a marker of definitive hemogenic mesoderm, as KDR+CD235a-CD1d+ mesoderm contains nearly all definitive lymphoid, myeloid, and erythroid potential in a high WNT, low TGFβ culture.

4.7 Discussion

Despite the failure of the 3xmCherry-*CDX4* reporter to effectively label *CDX4+* mesoderm and disruption of endogenous expression, scRNAseq allowed for the identification of a correlate marker, CD1d. The isolation of CD1d+ mesoderm enriched not only for *CDX4*/*HOXA*

expressing cells, but also nearly all definitive hematopoietic potential, representing a novel marker for definitive hemogenic mesoderm. It still remains unclear why CD1d is being expressed in this particular developmental population and why it is correlated with hemogenic mesoderm. Given that CD1d responds not only to foreign glycolipids from bacteria and other organism, but also to endogenously produced glycolipids[169], it is possible that it is being utilized in some regulatory manner in these cell types. There is some evidence that CD1d regulates HSC potential and function, albeit in a non-cell autonomous manner[170]; perhaps CD1d also plays a role in the endothelial to hematopoietic transition. Treatment with α-galactosylceramide or other CD1d ligand could test these possibilities[169], perhaps leading to an inflammatory response in these cell types or some other, poorly understood effect.

In addition to understanding why CD1d is expressed in these cell types, there is also the question of how it becomes expressed in these cell types. Cursory experiments using the *CDX4*$^{-/y}$ KO line suggested that CD1d expression at day 3 was not affected by loss of *CDX4*. Additionally, data from paired ChIP/RNAseq of mesoderm isolated from the *CDX4*$^{-/y}$, *iCDX4* rescue line revealed no CDX4 binding peaks near the *CD1D* locus, nor any changes in the gene expression as a result of *CDX4* induction via DOX treatment. This would suggest that CD1d is simply a correlative mesodermal marker and is not directly regulated by *CDX4*. Analysis of expression in mouse scRNAseq gastrulation datasets revealed the expression of *Cd1d1* within some mesodermal cells, suggesting that its expression is also found in *in vivo* and within other species at a correlate timepoint (E7.25, E7.5)[171]. In this dataset, *Cd1d1*+ cells did not overlap significantly with *Cdx4* expressing cells, perhaps suggesting species differences, though whether CD1d protein is detectable in these populations or the lineages that *Cd1d1* expressing mesoderm in the mouse gives rise to is not known. Given that CD1d expression was lower in IWP2 treated

KDR+CD235a+ primitive hemogenic mesoderm, it is possible that the WNT agonist CHIR99021 that is used to specify definitive hemogenic mesoderm is leading to this increase in CD1d. The TGFβ inhibitor SB431542 was also used to aid specification of *CDX/HOXA+* hemogenic mesoderm as demonstrated by others[118], which could also leading to an increase in CD1d expression, as TGFβ signaling has been shown to inhibit CD1d expression[172]. Perhaps the inhibition of TGFβ is leading to the expression of CD1d and by testing varying concentrations of SB431542 or adding the Nodal ligand activin A, which is also inhibited by SB431542, could lead to some insight on what regulates CD1d expression. Regardless, the identification of this novel marker allows for greater purification of *CDX4+* definitive hemogenic mesoderm and CD1d can be used as a selection marker to purify mesoderm and allow for greater interrogation of further lineage questions.

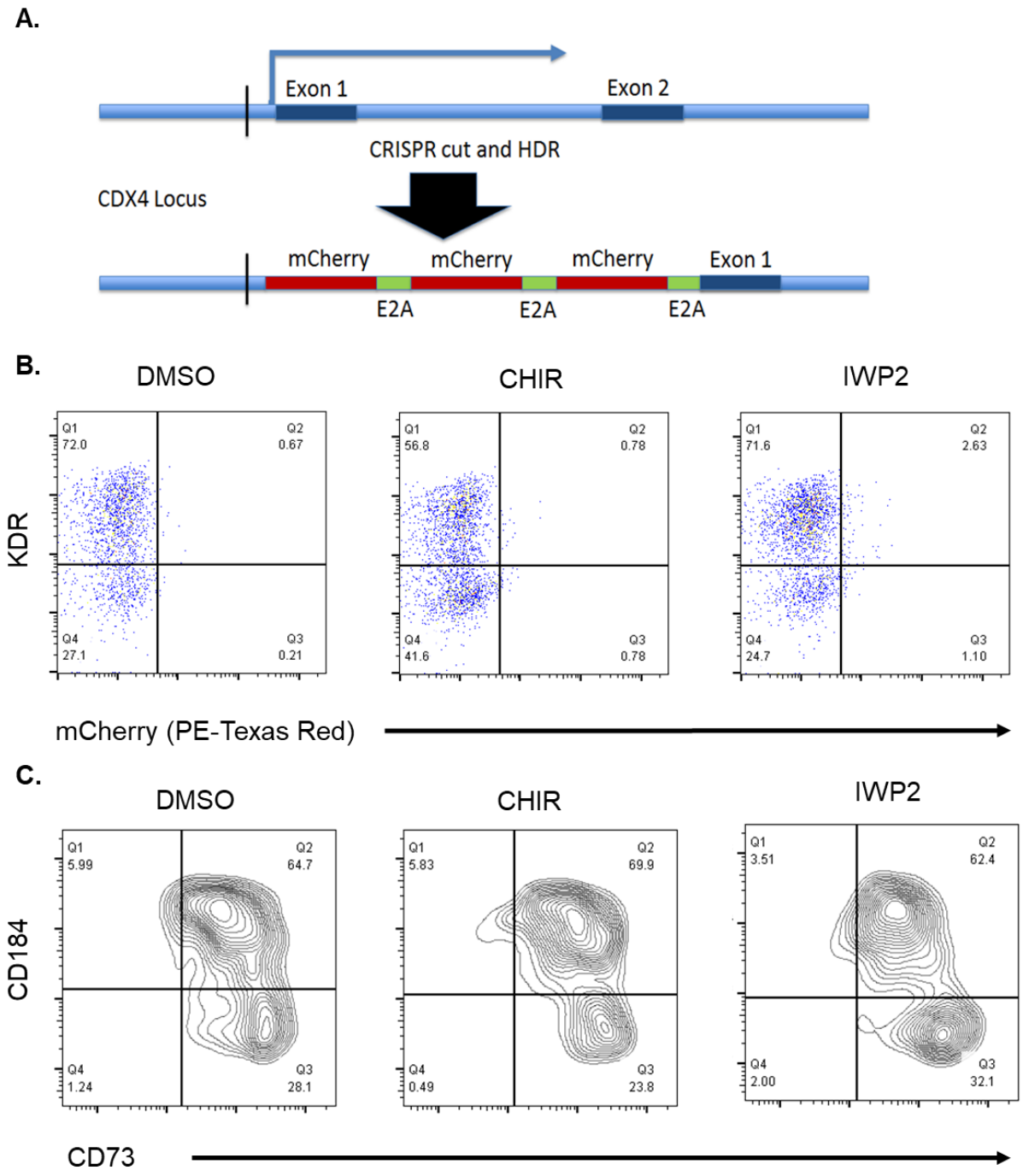


Figure 4.1: Design and test of an in-frame *CDX4* reporter line

A. A schematic of the *CDX4* locus, detailing the design of a triple mCherry (3xmCherry) florescent construct to be placed at the 5' end of exon 1 of *CDX4*, through homologous recombination after cleavage with CRISPR Cas9. **B.** The *CDX4* 3xmCherry hPSC line was

differentiated to day 3 mesoderm with BMP4/bFGF/activin A, as well as to definitive hemogenic mesoderm via CHIR99021 (CHIR) treatment, or primitive hemogenic mesoderm with IWP2 treatment. Flow cytometry was performed to assess for the formation of KDR+ mesoderm, as well as to assess for *CDX4* dependent mCherry expression. **C.** *CDX4* 3xmCherry hPSCs were differentiated to day 8 hemogenic endothelium stage under established conditions and flow cytometry performed. Shown gated on CD34+ cells, looking at endothelial markers CD184 and CD73.

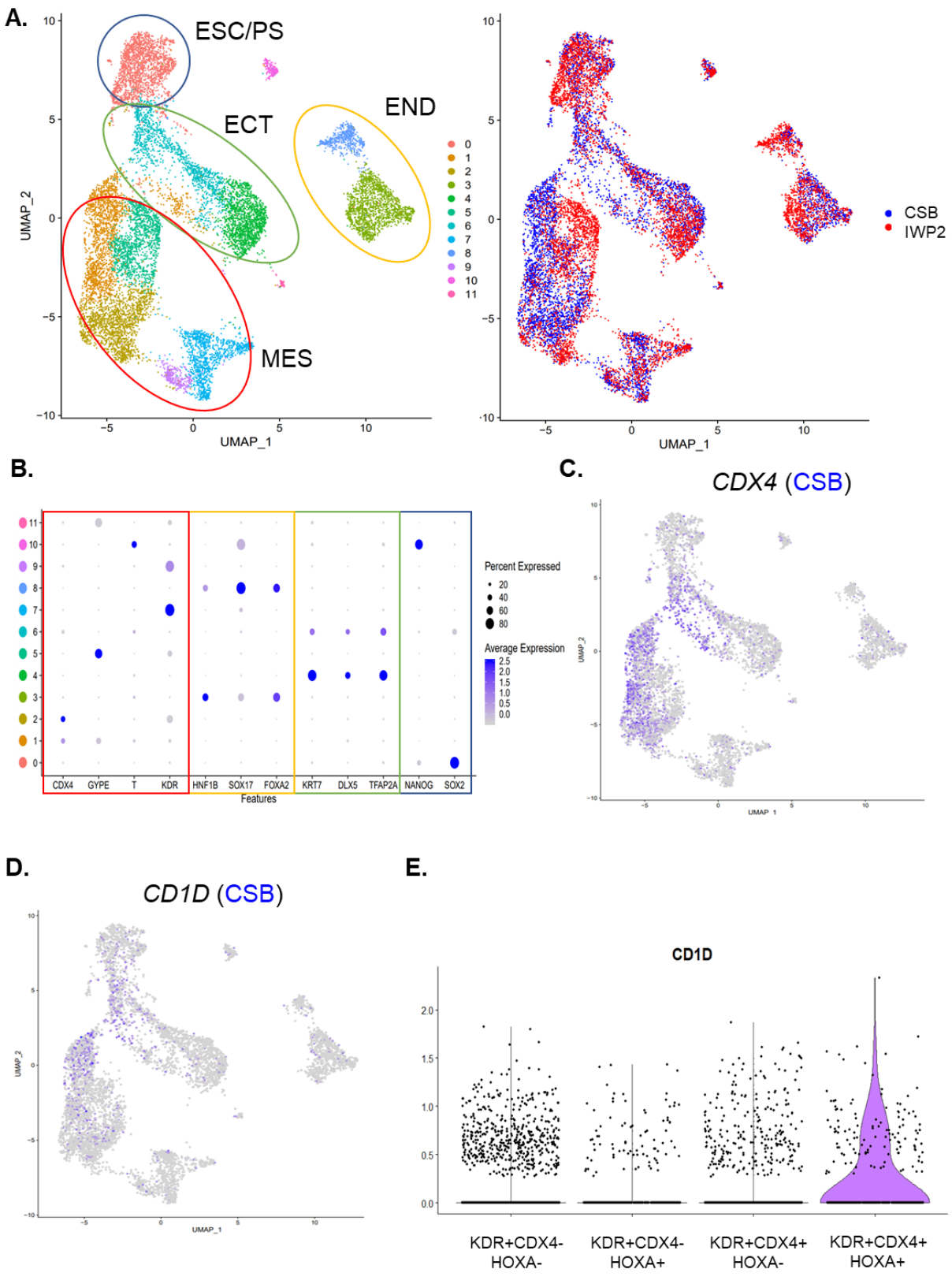


Figure 4.2: scRNAseq of allows for characterization of *CDX4+* mesoderm and reveals *CD1D* as a potential surface marker

A. scRNAseq was performed on bulk culture treated with CHIR99021 and SB431542 (CSB) or IWP2 at day 3, after the specification of mesoderm to better assess the culture heterogeneity and to find hemogenic mesoderm. Briefly, the cells were dissociated via trypsin and fixed in methanol, before being processed via 10X Genomics Chromimum 3' kit, capturing >6000 cells per sample set. The libraries were sequencing via an Illumina HiSeq 3000 and processed via the Cell Ranger pipeline. The Seurat v3 R package was used to integrate the cells from the CSB and IWP2 datasets and organize them into two-dimensional space via UMAP. Clustering of the cells was then performed using the Seurat algorithm, resulting in 11 algorithmically distinct clusters denoted by different colors. On the right, the contribution of each sample (CSB in blue or IWP2 in red) is shown to each region of the UMAP. Presumptive identity of each cluster is indicated by colored circle (pluripotent/primitive streak - blue, ectoderm - green, endoderm - yellow, and mesoderm - red). **B.** Expression of various pluripotent/primitive streak (blue), ectoderm (green), endoderm (yellow) and mesoderm (red) genes was assess via dot plot, where the expression of each gene is shown for a particular cluster. Size indicates relative numbers of cells expressing a particular gene in the given cluster, blue color intensity indicates average relative level of expression among cells in the cluster. **C.** Expression of *CDX4* shown on the UMAP (within only the CSB cells), where darker blue indicates higher expression. **D.** Expression of *CD1D* shown on the UMAP (within only the CSB cells), where darker blue indicates higher expression. **E.** Violin plot of the *CD1D* expression of different KDR+ cell types with or without expression of *CDX4* or *HOXA* genes. Dots indicate individual cells, and the colored outline indicates quantiles based on the total number of cells.

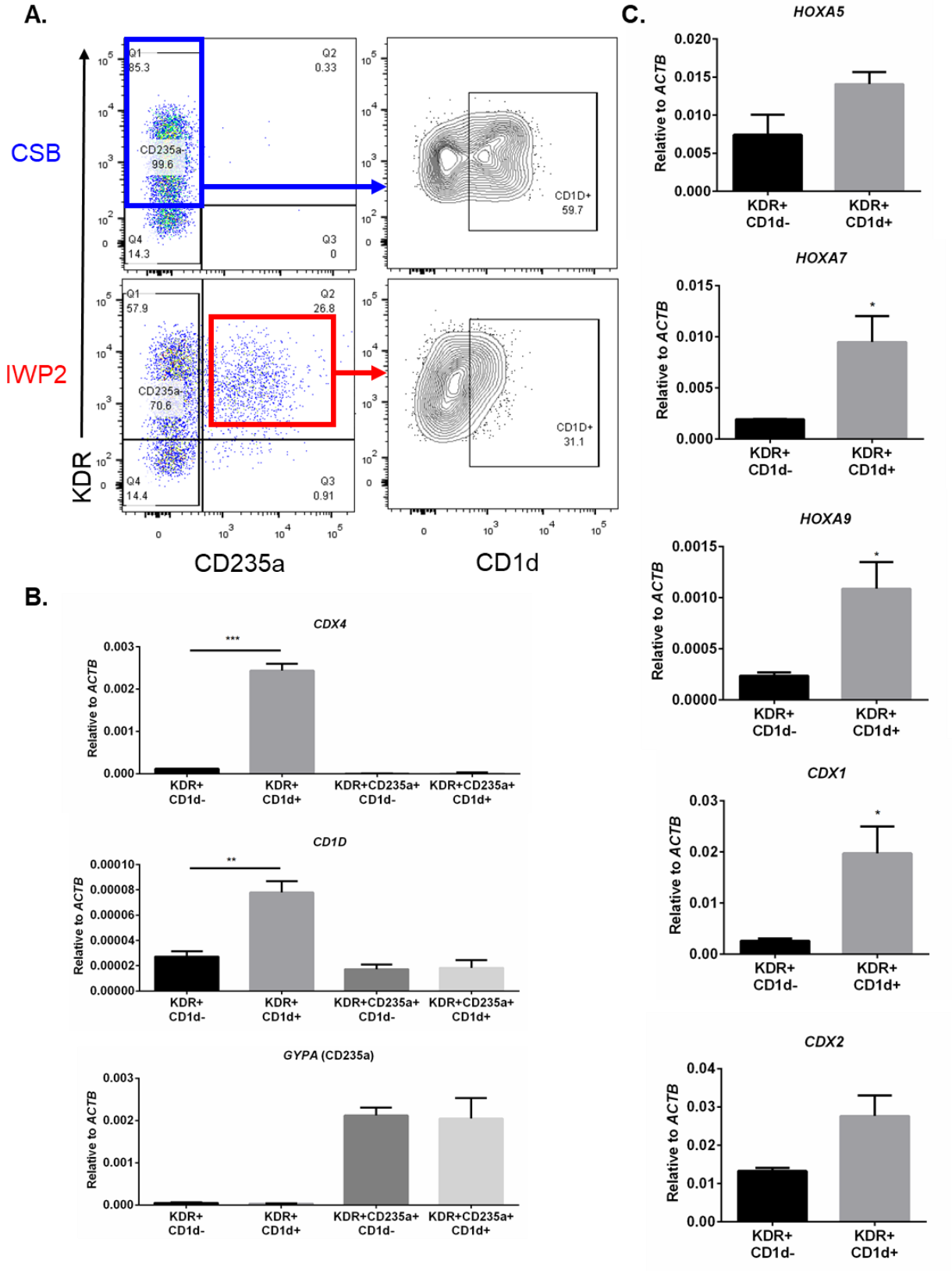


Figure 4.3: Phenotypic characterization of CD1d expressing cells within day 3 mesoderm

A. Representative flow cytometry of both CHIR99021/SB431542 (CSB) and IWP2 treated day 3 cultures for KDR, CD235a, and CD1d. KDR+CD235a- definitive hemogenic mesoderm and KDR+CD235a+ primitive hemogenic mesoderm were then assessed for the expression of CD1d in the plots on the right. **B.** KDR+CD235a-CD1d-/+ and KDR+CD235a+CD1d-/+ populations were isolated via FACS for RNA, in order to make cDNA and perform RT-qPCR for *CDX4, CD1D,* and *GYPA*. n = 3, ** $p < 0.01$, *** $p < 0.001$ via students t test. **C.** Expression of *CDX* and *HOXA* genes via qPCR in KDR+CD235a-CD1d-/+ FACS isolated cells. n =3, * is $p < 0.05$ via students t test, error bars SEM.

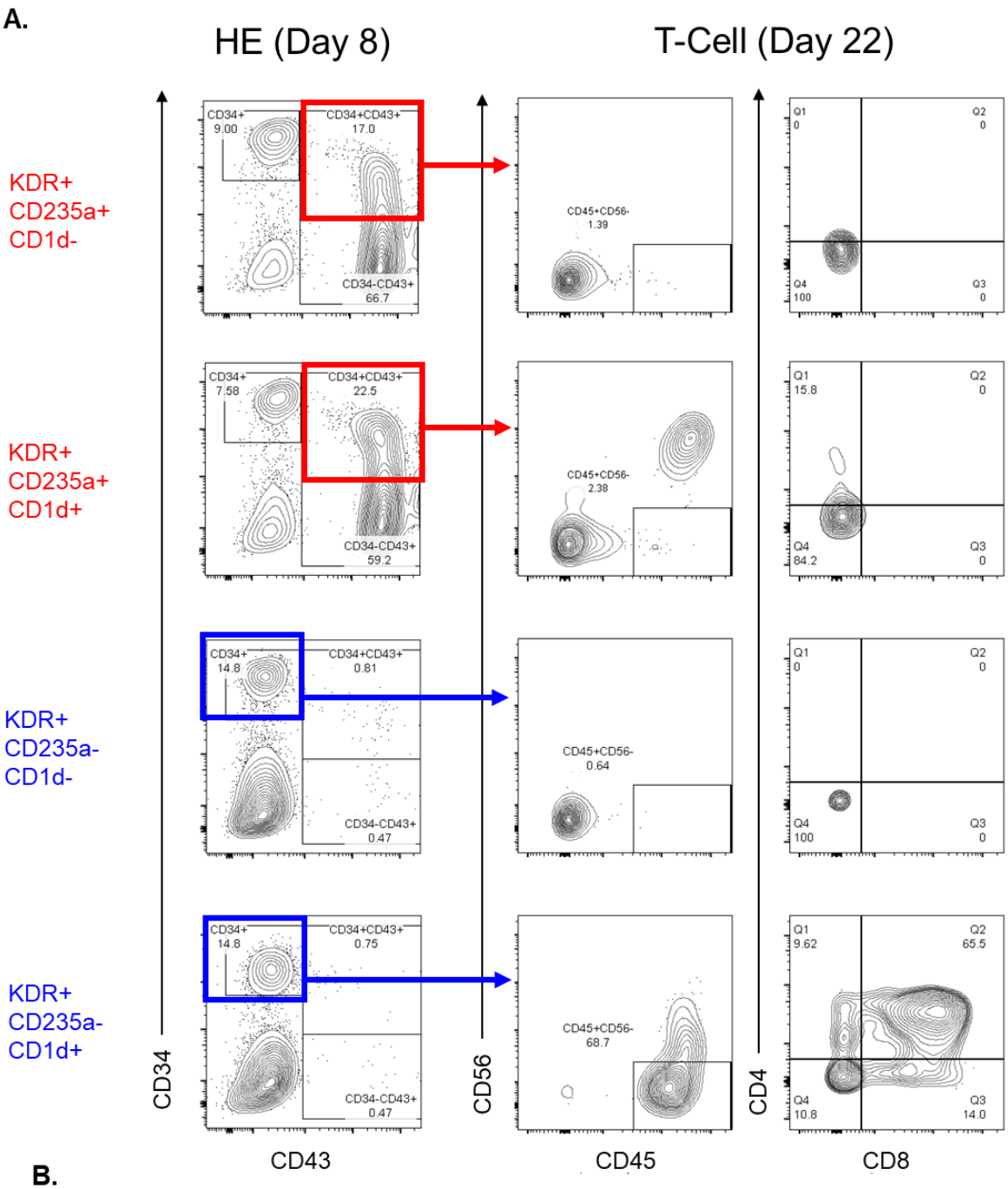


Figure 4.4: Functional characterization of CD1d+ mesoderm

A. Representative flow cytometry plots of CHIR99021/SB431542 (blue) and IWP2 (red) differentiated hPSCs, sorted on day 3 for KDR+CD235a-/+CD1d-/+ and reaggregated for an additional 5 days and harvested again for FACS on day 8. IWP2 treated samples were then FACS isolated for CD34+CD43+ and CSB samples for CD34+CD43- progenitors and cultured on OP9-DL4 stroma under T-Cell promoting conditions to test for lymphoid potential after 22 days of culture. Cells were then harvested for flow cytometry, CD45+CD56- (middle flow plots) progenitors were then gated to exclude natural killer cells and assessed for CD4/CD8 T-cells (right flow plots). **B.** Day 3 sorted KDR+CD235a+CD1d-/+ (IWP2) or KDR+CD235a-CD1d-/+ (CSB) hemogenic mesoderm was isolated and reaggregated for an additional 5 days, flowed by a second FACS isolation of CD34+CD43+ (IWP2) or CD34+CD43- (CSB) progenitors, which were seeded into hematopoietic methylcellulose media for colony forming assays (after additional EHT culture for CSB CD34+). The numbers of colonies were then assessed after 8-10 days (IWP2) or 10-12 days (CSB) and the numbers were counted of burst forming units erythroid (BFU-E) and colony forming units (CFU) of erythroid (E), granulocyte (G), myeloid (M), and mixed granulocyte/myeloid (GM). IWP2 n=4, CSB n =6, error bars SEM.

Table 4.1: *CD1D* identified as a DEG in UMAP cluster 1

Cluster 1 from the Seurat generated UMAP projection of scRNAseq of hemogenic mesoderm was identified as having mesodermal markers such as KDR, as well as a large number of CDX4+ cells. Taking only the CSB derived cells from this cluster (further enriching for CDX4+ cells) and then performing DEG analysis in comparison to the rest of the CSB culture revealed a multitude of DEG's, including several HOXA and CDX genes (highlighted in red). Displayed is the top 30 differentially expressed genes ordered by p_{adj} value, CD1D highlighted in green.

Top 30 Differentially Expressed genes (by significance)		
HOXB-AS3	HOXB3	HOXB1
WNT5A	TBX6	PTCH1
CDX2	HES7	MSX1
MESP1	HOXB9	ABLIM1
DLL3	MEST	COL6A1
CDX1	S100A10	OSR1
WNT5B	SALL1	FABP5
HOTAIRM1	MCOLN3	HOXA3
CDX4	HOXA1	CD1D
HOXA-AS3	RBP1	URAD

Chapter 5: Discussion

5.1 *CDX4* in hematopoietic development

True functional assessment via engraftment studies in immunocompromised mice of mesoderm induced *iCDX4* hematopoietic progenitors was not assessed in this work, questioning whether *CDX4* can be truly termed as a 'regulator' of definitive hematopoiesis. While T cell potential was found to be reduced in frequency as a result of loss of *CDX4* and restricted to a KDR+CD1d+*CDX4*+ mesodermal progenitors, the presence of lymphoid potential is not sufficient to characterize these as definitive hematopoietic progenitors. The existence of lineages such as the LMPP[53] in mouse development that give rise to lymphoid progenitors independent of the HSC suggests that this metric alone cannot be used alone to denote definitive, HSC dependent hematopoiesis. Speculating on whether enforced mesodermal *CDX4* expression would result in the production of an engraftable progenitor, it is probable that *CDX4* expression alone would be insufficient to specify a true HSC for several reasons. The first being that mesodermal *CDX4* expression only lead to comparable levels of progenitors as mesodermal WNT agonism when measured by colonies in methylcellulose, suggesting that WNT agonism alone is able to activate *CDX4* with the same functional consequences as overexpression. Secondly, $CDX4^{y/-}$ did not lead to a complete loss of erythro-myleloid-lymphoid bearing HE, suggesting perhaps *CDX4* simply influences the amount of HE specified, not necessarily its lineage potential or ability to engraft. Thirdly, other transgenic approaches to producing an HSC *in vitro* have need a minimum of seven transcription factors activated post EHT in order to produce an engraftable progenitor[173]. While three of these factors were *HOXA* genes and appear to activated by *CDX4* at an early stage after mesoderm specification, overexpression of *ERG*, *LCOR*, *RUNX1*, and *SPI1*,

were also required to produce an engraftable progenitor. In this work, none of these genes are activated by mesodermal *CDX4* expression at any measured stage, suggesting that there are other regulatory signals that are likely needed outside of *CDX4* that lead to the upregulation of these factors and ultimately specification of an engraftable progenitor. For similar reasons, enrichment of *CDX4+* mesoderm using FACS of KDR+CD1d+ cell is also unlikely to result in the specification of an engraftable progenitor

Despite these caveats, this work demonstrates *CDX4* appears to regulate the development of human definitive-like hemogenic mesoderm through genetic studies. Important questions are still raised, however, when examining the literature surrounding *CDX4* in other model systems. In the zebrafish, *cdx4* KO appears to affect all waves of hematopoiesis[112,130], as the area known as the interior cell mass (ICM), where primitive hematopoiesis occurs in the zebrafish[174], also displayed deficits in hematopoiesis. Additionally, in mouse embryonic stem cells models, *Cdx4* induction appeared to positively regulate primitive hematopoiesis[84,114] and KO of *Cdx4* in E8.5 mouse embryos lead to a deficit in primitive hematopoietic progenitors[113]. This directly contrasts the results from our differentiation model system, where loss of *CDX4* lead to an expansion in primitive hematopoiesis and a deficit in definitive.

What could explain these contrasting results? Perhaps some of the explanation could also lie in species differences between *CDX* and *HOX* gene expression; zebrafish have only 2, instead of 3 cdx genes: cdx4 and cdx1a. Although it is broadly assumed that CDX genes are redundant in their functions[106,107,130], but perhaps the absence of one CDX gene in the zebrafish could explain the differences in their regulation and expression. As for the mouse, it is less clear what could be causing these differences in CDX dependent regulation of hematopoiesis. Some of these differences could be a result of timing of the stages of hematopoietic development, which is

faster in mouse pluripotent stem cell differentiations compared to hPSCs. Additionally, the aforementioned functional analysis of *Cdx4* KO mouse yolk sacs on embryonic day 8.5[113] would also capture the erythro-myeloid progenitors (EMPS) at this particular time point[48,120], which could suggest that the EMP in the mouse might be dependent on *Cdx4*. The primitive program has also showed to be dependent on WNT activation in mouse pluripotent stem cell differentiations[86], whereas the opposite is true in hPSCs, as the primitive program is independent of WNT signaling[85]. Interestingly, in other studies there was no overt hematopoietic phenotype in *Cdx4* KO mice, which appeared normal and healthy despite apparent reductions in yolk sac hematopoiesis[175].

Although not capable of an HSC, the EMP in the mouse gives rise to more mature erythrocytes and granulocytes, as well as natural killer cells[34] and has sometimes been termed as a "definitive" hematopoietic progenitor with more in common with the HSC than other primitive yolk sac progenitors[176]. Perhaps the *CDX4* dependent population in our hPSC differentiation model could resemble a human EMP, though the mouse EMP does not have lymphoid/T-Cell potential (excepting natural killers), whereas our *CDX4* dependent definitive progenitors are capable of robust T-cell production on OP9-DL4 stroma. There is also *in vivo* evidence that the human aorta gonad mesonephros (AGM) region where the HSC emerges is highly enriched for *HOXA* genes[118], again supporting the idea that *CDX4/HOXA* expression is correlative of a definitive hematopoietic progenitor. Observations from our lab suggest that KDR+CD184-*CDX4*+ mesoderm gives rise to hematopoietic progenitors independent of retinoic acid signaling, which the HSC is known to dependent on[147], but it has not been assessed whether *CDX4* KO would affect the specification of KDR+CD184+$CDX4_{lo}$ retinoic acid dependent hematopoiesis. Another possibility is that *CDX4* dependent hematopoiesis could represent a definitive, intra-

embryonic hemogenic endothelium that is incapable of giving rise to an HSC; recent single cell transcriptomics of early human embryos identified two transcriptional distinct waves of hemogenic endothelial cells in the embryo proper, one that appeared primed for HSC specification and one not[71]. Whether the EMP or human equivalent of the EMP is dependent on *CDX4*, or the *CDX4* dependent HE in our system represent an 'early' hemogenic endothelium that does not produce an HSC, this work advances the intricate and complicated field of embryonic hematopoietic development.

Now that *CDX4+* mesoderm can be isolated effectively through the use of CD1d as a correlative marker, further questions on what lineages in our system arise from *CDX4+* cells. It appears based on our informatics data that the CD184-, *CDX4+*, *CYP26A1+*, retinoic independent mesoderm also expresses *CD1D*. This would indicate that *CDX4+* mesoderm is likely giving rise to retinoic acid independent hematopoiesis, although *CDX4* and CD184 are not mutually exclusive. The ability of KDR+CD235a-CD1d-/+ mesoderm to give rise to retinoic dependent hematopoiesis has not been tested, though it is likely that KDR+CD1d- mesoderm might represent the CD184+*ALDH1A2+CYP26A1-* retinoic acid dependent hemogenic mesoderm. This hypothesis is further supported by the fact that exogenous *CDX4* results in the expression of *CYP26A1*, which is expressed under normal, WT conditions in the KDR+CD184- mesoderm. Additionally, the *HOXA* signature of CD184- mesoderm was recapitulated in the CD184+ mesoderm when *CDX4* expression was enforced, suggesting that exogenous *CDX4* might convert CD184+ retinoic acid dependent mesoderm to a CD184- like retinoic acid independent mesoderm. The summarized model of these observations and hypothesizes are displayed in **Figure 5.1**.

It is still unclear how *CDX4* is leading to the repression of the primitive program from the paired ChIP/RNAseq experiments. Crosstalk between *CDX* and WNT signaling has been observed in different contexts[84,108,109], and it could be possible that *CDX4* induction is leading to an upregulation of WNT signaling, which could be repressing the primitive program as has been previously observed[85]. Looking at genes that were targeted by *CDX4*, both *WNT5A* and *WNT5B* were induced as a result of enforced *CDX4* expression, though these ligands are known to stimulate the non-canonical or β-catenin independent WNT signaling[177]. Given that low levels of β-catenin are needed to produce primitive hemogenic mesoderm[85], the induction of *CDX4* leading to more non-canonical WNT and further suppression of β-catenin does not support a WNT-*CDX4* based repression of primitive hematopoiesis. Changes in *FZD6, 7,* and *8* were also observed in CD235a+ primitive hemogenic mesoderm after induction of *CDX4*, further complicating the possible interactions, making it difficult to formulate a hypothesis of a precise mechanism for the WNT related suppression of primitive hematopoiesis. Further study will be needed to tease out the precise nature of this relationship, perhaps through the use of different WNT ligands or genetic studies with KO's of various WNT receptors.

With the knowledge that CD184 and CD1d are markers segregating (potentially) very different types of definitive hemogenic mesoderm, future studies can tease out the precise lineage contributions of these mesodermal populations to *CDX4*-dependent/independent and retinoic acid-dependent/independent hematopoiesis. The genetic tools also provided as a result of this work, including the *CDX4*$^{-/y}$, *iCDX4* rescue line will also allow for testing the dependency of retinol responsive hematopoiesis on *CDX4*. The link between *CDX4* and medial *HOXA* genes is also strongly suggested in the work, though the precise population dependent relationships will need to be teased out using genetic complementation studies. These advances will also provide a

foundation to further improve the hematopoietic differentiation of hPSCs, providing a more pure and better defined hemogenic mesoderm. Knowledge of the role of *CDX4* also provides an indicator of definitive hematopoietic potential and can be leveraged to find culture conditions that increase its expression and perhaps, eventually, the HSC.

5.2 *CDX4* in cardiac development

The discovery that *CDX4* is negatively regulating the expression of *TBX20* directly lead to the hypothesis that *CDX4* expression represses cardiomyocyte specification, which was demonstrated functionally using the *CDX4*$^{-/y}$*, iCDX4* rescue line. Although similar observations had been found in mouse embryonic stem cells and zebrafish[159], as well as in mouse embryos[158], the mechanism of how *CDX* gene expression lead to a repression of the cardiac program was not clear. It is possible that CDX4 protein binding to an enhancer or regulatory region could recruit other factors that lead to a suppression of genes such as *TBX20*, through inactivation of these regions. While *CDX* genes were found to interact with chromatin remodelers such as SWI/SNF[154], it is unclear if CDX4 working in a similar manner, as SWI/SNF generally opens chromatin, whereas in this work it appears to be repressive. Additional experiments will be necessary to show a non-canonical, repressive function of *CDX4* outside of its known function as a transcription factor. DNA Binding motif analyzes of these CDX4-bound regions could reveal partners that could be acting cooperative with *CDX4*, though it is possible that CDX4 protein might be the only member of a putative complex that directly binds to DNA. To address this possibility and to more directly discover binding partners, lysates from differentiated mesoderm could be probed with FLAG-CDX4 protein and mass spectrometry performed to find enriched these partners. Without these analyzes identifying possible repressive binding partner for *CDX4*, there still remains the possibility that *CDX4* binding near *TBX20* is simply a coincidence and the

transcriptional repression seen is resulting from unknown indirect regulation from *CDX4* acting on another gene or genes.

Evidence of a common origins and opposing, intertwining genetic regulation of cardiac lineages and the primitive hemangioblast has been hinted at in the literature previously. Exogenous activation of notch signaling in mouse pluripotent stem cell differentiations was able to respecify hemangioblast containing populations to cardiac cells[95]. Additionally, loss of *Scl* in mouse embryos lead to the exogenous production of beating cardiomyocytes from yolk sac endothelium[96]. In the ChIP/RNAseq of KDR+CD235a+ mesoderm, there were no changes in *TAL1(SCL)* or in components of the notch signaling pathway between KO of *CDX4* and *CDX4* induction, which is somewhat surprising given that *cdx4* was found to bind and activate *scl* in zebrafish[131]. It is possible that changes in these genes/pathways might become apparent in later stages in of the differentiation post mesoderm formation and that *CDX4* might result in indirect activation of notch or *TAL1*. It is also unclear if the increase in cardiomyocytes seen as a result of loss of *CDX4* represents the conversion of progenitors to the cardiac lineage, or if there if there is simply an expansion of cardiomyocyte progenitors as a result of the formation of less hemogenic mesoderm.

The observation that ventricular cardiac progenitors arise from a CD235a+ mesoderm[134], low in *CDX4* expression, strongly suggests that *CDX4* is playing an important gate keeper role in cardiac specification, potentially by promoting hematopoietic lineages over cardiac. Although it has not been proven functionally, it is likely that day 3 KDR+CD235a+ progenitors are the source of ventricular cardiomyocytes as well as primitive hemogenic mesoderm. Whether CD235a markers both hemogenic and cardiogenic mesoderm or whether KDR+CD235a+ mesoderm represents a dually cardiac/hematopoietic capable population is not precisely clear.

Additionally, the knowledge that a retinol responsive mesoderm population that gives rise to atrial cardiomyocytes[134] could indicate that atrial cardiac progenitors might resided with KDR+CD184+ mesoderm. Given that this population is low in *CDX4* and *CYP26A1* expression, while high in *ALDH1A2*, and clearly responsive to retinol in a hematopoietic context, cardiac promoting conditions might reveal that CD184+ is also capable of atrial cardiomyocytes. Observations in zebrafish suggest that effects on cardiac development due to loss of *cdx* genes is dependent on concomitant inhibition of retinoic acid signaling[159]. This hypothesis would also suggest that not only is *CDX4* repressing ventricular cardiomyocytes through direct repression of *TBX20* and indirect repression of other factors (perhaps through *CDX2*), but *CDX4* might also be repressing atrial cardiac progenitors through induction of *CYP26A1*.

These predications could represent a more in-depth understanding of which types of mesoderm both hematopoietic and cardiac progenitors come from, summarized in **Figure 1**. If true, this greater model of what mesoderm gives rise to which progenitor could have great implications for hematopoietic and cardiac differentiation protocols. Ensuring the correct types of mesoderm are specified at outset of a differentiation has a large impact of the efficiency and numbers of the target cell type. Reducing the expression of *CDX4* during mesoderm formation could possibly be used to enhance the differentiation of ventricular cardiomyocyte differentiations, and possibly atrial cardiomyocyte differentiations as well. Given that the treatment of the WNT inhibitor IWP2 from days 3-5 is critical for cardiomyocyte differentitation[178], it is possibly that this leads to low *CDX* gene expression and thus prevents *CDX* dependent inhibition of cardiomyocyte transcription factors. Further experimentation with FACS isolation of these different mesodermal populations, as well as under various will be needed to determine if these

observations will be supported by the data and applicable to improving existing cardiomyocyte differentiation protocols.

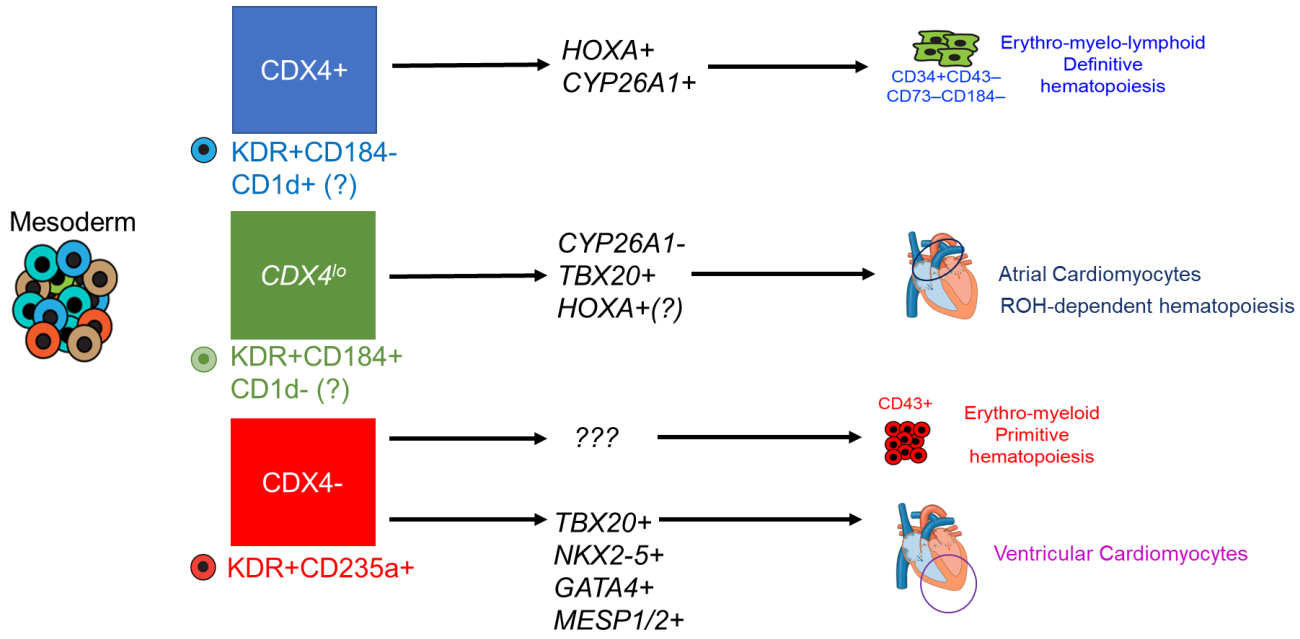


Figure 5.1 Model of early mesodermal development in hPSCs

Within mesoderm under high WNT conditions, KDR+CD184-(CD1d+) ROH independent hemogenic mesoderm is specified, having a *CDX4+HOXA+CYP26A1+* gene signature, ultimately giving rise to erythro-myelo-lymphoid definitive hemogenic endothelium. High WNT conditions also give rise to KDR+CD184+(CD1d-) ROH dependent hemogenic mesoderm, with lower levels of *CDX4* expression, allowing for expression of *TBX20* and leading to low *CYP26A1* expression. This population likely also gives rise to atrial cardiomyocytes, in addition to ROH-dependent erythron-myelo-lymphoid hematopoiesis. Lastly, under low WNT conditions a KDR+CD235a+ that is *CDX4-* is produced. This lack of *CDX4* expression allows for increased *TBX20* expression, as well as induction of other cardiac related transcription factors, leading to the specification of ventricular cardiomyocytes. This population gives rise to primitive hematopoietic progenitors through an unknown, but low WNT dependent mechanism.

References

1 Till, J. E. & McCulloch, E. A. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiation research* **14**, 213-222 (1961).

2 Becker, A. J., McCulloch, E. A. & Till, J. E. Cytological Demonstration of the Clonal Nature of Spleen Colonies Derived from Transplanted Mouse Marrow Cells. *Nature* **197**, 452-454, doi:10.1038/197452a0 (1963).

3 Siminovitch, L., McCulloch, E. A. & Till, J. E. The distribution of colony-forming cells among spleen colonies. *Journal of Cellular Comparative Physiology* **62**, 327-336 (1963).

4 Wu, A. M., Till, J. E., Siminovitch, L. & McCulloch, E. A. Cytological evidence for a relationship between normal hematopoietic colony-forming cells and cells of the lymphoid system. *J Exp Med* **127**, 455-464 (1968).

5 Berman, J. W. & Basch, R. S. Thy-1 antigen expression by murine hematopoietic precursor cells. *Experimental hematology* **13**, 1152-1156 (1985).

6 Spangrude, G., Heimfeld, S. & Weissman, I. Purification and characterization of mouse hematopoietic stem cells. *Science (New York, N.Y.)* **241**, 58-62, doi:10.1126/science.2898810 (1988).

7 Ikuta, K. & Weissman, I. L. Evidence that hematopoietic stem cells express mouse c-kit but do not depend on steel factor for their generation. *Proc Natl Acad Sci U S A* **89**, 1502-1506, doi:10.1073/pnas.89.4.1502 (1992).

8 Krause, D. S. *et al.* Multi-organ, multi-lineage engraftment by a single bone marrow-derived stem cell. *Cell* **105**, 369-377 (2001).

9 Oguro, H., Ding, L. & Morrison, S. J. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell* **13**, 102-116, doi:10.1016/j.stem.2013.05.014 (2013).

10 Kiel, M. J. *et al.* SLAM Family Receptors Distinguish Hematopoietic Stem and Progenitor Cells and Reveal Endothelial Niches for Stem Cells. *Cell* **121**, 1109-1121, doi:https://doi.org/10.1016/j.cell.2005.05.026 (2005).

11 Ivanovs, A., Rybtsov, S., Anderson, Richard A., Turner, Marc L. & Medvinsky, A. Identification of the Niche and Phenotype of the First Human Hematopoietic Stem Cells. *Stem Cell Reports* **2**, 449-456, doi:https://doi.org/10.1016/j.stemcr.2014.02.004 (2014).

12 Sintes, J., Romero, X., Marin, P., Terhorst, C. & Engel, P. Differential expression of CD150 (SLAM) family receptors by human hematopoietic stem and progenitor cells. *Experimental hematology* **36**, 1199-1204, doi:10.1016/j.exphem.2008.03.015 (2008).

13 Thomas, E. D., Lochte, H. L., Cannon, J. H., Sahler, O. D. & Ferrebee, J. W. Supralethal whole body irradiation and isologous marrow transplantation in man. *The Journal of Clinical Investigation* **38**, 1709-1716 (1959).

14 Apperley, J. *et al.* Bone marrow transplantation for patients with chronic myeloid leukaemia: T-cell depletion with Campath-1 reduces the incidence of graft-versus-host disease but may increase the risk of leukaemic relapse. *Bone marrow transplantation* **1**, 53-66 (1986).

15 Armand, P. & Antin, J. H. Allogeneic stem cell transplantation for aplastic anemia. *Biology of Blood Marrow Transplantation* **13**, 505-516 (2007).

16 Kamani, N. R. *et al.* Unrelated donor cord blood transplantation for children with severe sickle cell disease: results of one cohort from the phase II study from the Blood and Marrow Transplant Clinical Trials Network (BMT CTN). *Biology of Blood Marrow Transplantation* **18**, 1265-1272 (2012).

17 Gaziev, J. *et al.* Bone marrow transplantation for thalassemia from alternative related donors: improved outcomes with a new approach. *Journal of the American Society of Hematology* **122**, 2751-2756 (2013).

18 Goulmy, E. *et al.* Mismatches of minor histocompatibility antigens between HLA-identical donors and recipients and the development of graft-versus-host disease after bone marrow transplantation. *New England Journal of Medicine* **334**, 281-285 (1996).

19 Buhler, S. & Sanchez-Mazas, A. HLA DNA sequence variation among human populations: molecular signatures of demographic and selective events. *PloS one* **6** (2011).

20 Wiener, L. S. *et al.* Sibling stem cell donor experiences at a single institution. *Psycho-Oncology* **17**, 304-307 (2008).

21 Kharfan-Dabaja, M. A. *et al.* Comparative efficacy of tandem autologous versus autologous followed by allogeneic hematopoietic cell transplantation in patients with newly diagnosed multiple myeloma: a systematic review and meta-analysis of randomized controlled trials. *Journal of Hematology Oncology* **6**, 2 (2013).

22 Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156 (1981).

23 Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255-256 (1984).

24 Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science (New York, N.Y.)* **282**, 1145-1147 (1998).

25 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).

26 Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proceedings of the Japan Academy, Series B* **85**, 348-362 (2009).

27 Murry, C. E. & Keller, G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* **132**, 661-680 (2008).

28 Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J. C. & Keller, G. A common precursor for hematopoietic and endothelial cells. *Development (Cambridge, England)* **125**, 725-732 (1998).

29 Kennedy, M. *et al.* T lymphocyte potential marks the emergence of definitive hematopoietic progenitors in human pluripotent stem cell differentiation cultures. *Cell reports* **2**, 1722-1735, doi:10.1016/j.celrep.2012.11.003 (2012).

30 Sugimura, R. *et al.* Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature* **545**, 432-438 (2017).

31 Lis, R. *et al.* Conversion of adult endothelium to immunocompetent haematopoietic stem cells. *Nature* **545**, 439-445 (2017).

32 Ditadi, A., Sturgeon, C. M. & Keller, G. A view of human haematopoietic development from the Petri dish. *Nature Reviews Molecular Cell Biology* **18**, 56-67 (2017).

33 Dzierzak, E. & Speck, N. A. Of lineage and legacy: the development of mammalian hematopoietic stem cells. *Nature immunology* **9**, 129-136 (2008).

34 Dege, C. *et al.* Potently Cytotoxic Natural Killer Cells Initially Emerge from Erythro-Myeloid Progenitors during Mammalian Development. *Developmental cell* (2020).

35 Sabin, F. R. Studies on the origin of blood-vessels and of red blood-corpuscles as seen in the living blastoderm of chicks during the second day of incubation. *Contrib. Embryol Carneg Inst.* **9**, 214-262 (1920).

36 Murray, P. D. F. r. The development in vitro of the blood of the early chick embryo. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* **111**, 497-521 (1932).

37 Huber, T. L., Kouskoff, V., Fehling, H. J., Palis, J. & Keller, G. Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. *Nature* **432**, 625-630 (2004).

38 Lugus, J. J., Park, C., Ma, Y. D. & Choi, K. Both primitive and definitive blood cells are derived from Flk-1+ mesoderm. *Blood* **113**, 563-566, doi:10.1182/blood-2008-06-162750 (2009).

39 Wareing, S. *et al.* The Flk1-Cre-Mediated Deletion of ETV2 Defines Its Narrow Temporal Requirement During Embryonic Hematopoietic Development. *Stem cells (Dayton, Ohio)* **30**, 1521-1531, doi:10.1002/stem.1115 (2012).

40 Chung, Y. S. *et al.* Lineage analysis of the hemangioblast as defined by FLK1 and SCL expression. *Development (Cambridge, England)* **129**, 5511-5520, doi:10.1242/dev.00149 (2002).

41 Ueno, H. & Weissman, I. L. Clonal Analysis of Mouse Development Reveals a Polyclonal Origin for Yolk Sac Blood Islands. *Developmental cell* **11**, 519-533, doi:https://doi.org/10.1016/j.devcel.2006.08.001 (2006).

42 Lancrin, C. *et al.* The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature* **457**, 892-895, doi:10.1038/nature07679 (2009).

43 Tavian, M., Robin, C., Coulombel, L. & Péault, B. The human embryo, but not its yolk sac, generates lympho-myeloid stem cells: mapping multipotent hematopoietic cell fate in intraembryonic mesoderm. *Immunity* **15**, 487-495 (2001).

44 Peschle, C. *et al.* Haemoglobin switching in human embryos: asynchrony of ζ→ α and ε→ γ-globin switches in primitive and definitive erythropoietic lineage. *Nature* **313**, 235-238 (1985).

45 Tober, J. *et al.* The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. *Blood* **109**, 1433-1441 (2007).

46 Hoeffel, G. & Ginhoux, F. Ontogeny of tissue-resident macrophages. *Frontiers in Immunology* **6**, 486 (2015).

47 Ginhoux, F. *et al.* Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science (New York, N.Y.)* **330**, 841-845 (2010).

48 Palis, J. Primitive and definitive erythropoiesis in mammals. *Frontiers in physiology* **5**, 3 (2014).

49 Ivanovs, A. *et al.* Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. *J Exp Med* **208**, 2417-2427, doi:10.1084/jem.20111688 (2011).

50 Medvinsky, A., Rybtsov, S. & Taoudi, S. Embryonic origin of the adult hematopoietic system: advances and questions. *Development (Cambridge, England)* **138**, 1017-1031, doi:10.1242/dev.040998 (2011).

51 Frame, J. M., McGrath, K. E. & Palis, J. Erythro-myeloid progenitors:"definitive" hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells. *Blood Cells, Molecules, Diseases* **51**, 220-225 (2013).

52 Schulz, C. *et al.* A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science (New York, N.Y.)* **336**, 86-90, doi:10.1126/science.1219179 (2012).

53 Böiers, C. *et al.* Lymphomyeloid Contribution of an Immune-Restricted Progenitor Emerging Prior to Definitive Hematopoietic Stem Cells. *Cell Stem Cell* **13**, 535-548, doi:https://doi.org/10.1016/j.stem.2013.08.012 (2013).

54 McGrath, K. E. *et al.* A transient definitive erythroid lineage with unique regulation of the β-globin locus in the mammalian embryo. *Blood* **117**, 4600-4608, doi:10.1182/blood-2010-12-325357 (2011).

55 Peschle, C. *et al.* Embryonic----Fetal Hb switch in humans: studies on erythroid bursts generated by embryonic progenitors from yolk sac and liver. *Proc Natl Acad Sci U S A* **81**, 2416-2420, doi:10.1073/pnas.81.8.2416 (1984).

56 Bertrand, J. Y. *et al.* Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464**, 108-111, doi:10.1038/nature08738 (2010).

57 Kissa, K. & Herbomel, P. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* **464**, 112-115, doi:10.1038/nature08761 (2010).

58 Boisset, J.-C. *et al.* In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature* **464**, 116-120, doi:10.1038/nature08764 (2010).

59 Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. & Speck, N. A. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature* **457**, 887-891, doi:10.1038/nature07619 (2009).

60 North, T. E. *et al.* Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity* **16**, 661-672, doi:10.1016/s1074-7613(02)00296-0 (2002).

61 Morrison, S. J., Hemmati, H. D., Wandycz, A. M. & Weissman, I. L. The purification and characterization of fetal liver hematopoietic stem cells. *Proceedings of the National Academy of Sciences* **92**, 10302-10306, doi:10.1073/pnas.92.22.10302 (1995).

62 Morrison, S. J. & Scadden, D. T. The bone marrow niche for haematopoietic stem cells. *Nature* **505**, 327-334, doi:10.1038/nature12984 (2014).

63 Oberlin, E., Tavian, M., Blazsek, I. & Péault, B. Blood-forming potential of vascular endothelium in the human embryo. *Development (Cambridge, England)* **129**, 4147-4157 (2002).

64 Gekas, C. *et al.* Hematopoietic stem cell development in the placenta. *Int J Dev Biol* **54**, 1089-1098, doi:10.1387/ijdb.103070cg (2010).

65 North, T. *et al.* Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development (Cambridge, England)* **126**, 2563-2575 (1999).

66 Bertrand, J. Y., Cisson, J. L., Stachura, D. L. & Traver, D. Notch signaling distinguishes 2 waves of definitive hematopoiesis in the zebrafish embryo. *Blood* **115**, 2777-2783, doi:10.1182/blood-2009-09-244590 (2010).

67 Hadland, B. K. *et al.* A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development. *Blood* **104**, 3097-3105, doi:10.1182/blood-2004-03-1224 (2004).

68 Burns, C. E., Traver, D., Mayhall, E., Shepard, J. L. & Zon, L. I. Hematopoietic stem cell fate is established by the Notch-Runx pathway. *Genes & development* **19**, 2331-2342, doi:10.1101/gad.1337005 (2005).

69 Iso, T., Hamamori, Y. & Kedes, L. Notch Signaling in Vascular Development. *Development (Cambridge, England)* **23**, 543-553, doi:doi:10.1161/01.ATV.0000060892.81529.8F (2003).

70 Richard, C. *et al.* Endothelio-Mesenchymal Interaction Controls runx1 Expression and Modulates the notch Pathway to Initiate Aortic Hematopoiesis. *Developmental cell* **24**, 600-611, doi:https://doi.org/10.1016/j.devcel.2013.02.011 (2013).

71 Zeng, Y. *et al.* Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. *Cell Research* **29**, 881-894, doi:10.1038/s41422-019-0228-6 (2019).

72 Ditadi, A. & Sturgeon, C. M. Directed differentiation of definitive hemogenic endothelium and hematopoietic progenitors from human pluripotent stem cells. *Methods (San Diego, Calif.)* **101**, 65-72, doi:10.1016/j.ymeth.2015.10.001 (2016).

73 Bonkhofer, F. *et al.* Blood stem cell-forming haemogenic endothelium in zebrafish derives from arterial endothelium. *Nature Communications* **10**, 3577, doi:10.1038/s41467-019-11423-2 (2019).

74 Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *Development (Cambridge, England)* **87**, 27-45 (1985).

75 Wiles, M. V. & Keller, G. Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture. *Development (Cambridge, England)* **111**, 259-267 (1991).

76 Zambidis, E. T., Peault, B., Park, T. S., Bunz, F. & Civin, C. I. Hematopoietic differentiation of human embryonic stem cells progresses through sequential hematoendothelial, primitive, and definitive stages resembling human yolk sac development. *Blood* **106**, 860-870, doi:10.1182/blood-2004-11-4522 (2005).

77 Keller, G., Kennedy, M., Papayannopoulou, T. & Wiles, M. V. Hematopoietic commitment during embryonic stem cell differentiation in culture. *Mol Cell Biol* **13**, 473-486, doi:10.1128/MCB.13.1.473 (1993).

78 Kennedy, M., D'Souza, S. L., Lynch-Kattman, M., Schwantz, S. & Keller, G. Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures. *Blood* **109**, 2679-2687, doi:10.1182/blood-2006-09-047704 (2007).

79 Chadwick, K. *et al.* Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. *Blood* **102**, 906-915, doi:10.1182/blood-2003-03-0832 (2003).

80 Pick, M., Azzola, L., Mossman, A., Stanley, E. G. & Elefanty, A. G. Differentiation of human embryonic stem cells in serum-free medium reveals distinct roles for bone morphogenetic protein 4, vascular endothelial growth factor, stem cell factor, and

fibroblast growth factor 2 in hematopoiesis. *Stem cells (Dayton, Ohio)* **25**, 2206-2214, doi:10.1634/stemcells.2006-0713 (2007).

81 Wang, Y. & Nakayama, N. WNT and BMP signaling are both required for hematopoietic cell development from human ES cells. *Stem cell research* **3**, 113-125, doi:10.1016/j.scr.2009.06.001 (2009).

82 Potocnik, A. J., Nielsen, P. J. & Eichmann, K. In vitro generation of lymphoid precursors from embryonic stem cells. **13**, 5274-5283, doi:10.1002/j.1460-2075.1994.tb06861.x (1994).

83 Timmermans, F. *et al.* Generation of T cells from human embryonic stem cell-derived hematopoietic zones. *Journal of immunology (Baltimore, Md. : 1950)* **182**, 6879-6888, doi:10.4049/jimmunol.0803670 (2009).

84 Lengerke, C. *et al.* BMP and Wnt specify hematopoietic fate by activation of the Cdx-Hox pathway. *Cell Stem Cell* **2**, 72-82, doi:10.1016/j.stem.2007.10.022 (2008).

85 Sturgeon, C. M., Ditadi, A., Awong, G., Kennedy, M. & Keller, G. Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells. *Nature biotechnology* **32**, 554-561, doi:10.1038/nbt.2915 (2014).

86 Nostro, M. C., Cheng, X., Keller, G. M. & Gadue, P. Wnt, Activin, and BMP Signaling Regulate Distinct Stages in the Developmental Pathway from Embryonic Stem Cells to Blood. *Cell Stem Cell* **2**, 60-71, doi:https://doi.org/10.1016/j.stem.2007.10.011 (2008).

87 Ditadi, A. *et al.* Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nature cell biology* **17**, 580-591, doi:10.1038/ncb3161 (2015).

88 Burke, A. C. Development and evolution of the vertebrate mesoderm. *Developmental dynamics : an official publication of the American Association of Anatomists* **236**, 2369-2370, doi:10.1002/dvdy.21290 (2007).

89 Zhang, H. & Bradley, A. Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development. *Development (Cambridge, England)* **122**, 2977-2986 (1996).

90 Sumi, T., Tsuneyoshi, N., Nakatsuji, N. & Suemori, H. Defining early lineage specification of human embryonic stem cells by the orchestrated balance of canonical Wnt/beta-catenin, Activin/Nodal and BMP signaling. *Development (Cambridge, England)* **135**, 2969-2979, doi:10.1242/dev.021121 (2008).

91 Sadlon, T. J., Lewis, I. D. & D'Andrea, R. J. BMP4: its role in development of the hematopoietic system and potential as a hematopoietic growth factor. *Stem cells (Dayton, Ohio)* **22**, 457-474, doi:10.1634/stemcells.22-4-457 (2004).

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

93 Chan, S. S.-K. *et al.* Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. *Cell Stem Cell* **12**, 587-601, doi:10.1016/j.stem.2013.03.004 (2013).

94 Fehling, H. J. *et al.* Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. *Development (Cambridge, England)* **130**, 4217-4227, doi:10.1242/dev.00589 (2003).

95 Chen, V. C., Stull, R., Joo, D., Cheng, X. & Keller, G. Notch signaling respecifies the hemangioblast to a cardiac fate. *Nature Biotechnology* **26**, 1169-1178, doi:10.1038/nbt.1497 (2008).

96 Van Handel, B. *et al.* Scl represses cardiomyogenesis in prospective hemogenic endothelium and endocardium. *Cell* **150**, 590-605, doi:10.1016/j.cell.2012.06.026 (2012).

97 Garber, R., Kuroiwa, A. & Gehring, W. J. Genomic and cDNA clones of the homeotic locus Antennapedia in Drosophila. *EMBO* **2**, 2027-2036 (1983).

98 Carrasco, A. E., McGinnis, W., Gehring, W. J. & De Robertis, E. M. Cloning of an X. laevis gene expressed during early embryogenesis coding for a peptide region homologous to Drosophila homeotic genes. *Cell* **37**, 409-414 (1984).

99 McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. & Gehring, W. J. A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. *Cell* **37**, 403-408 (1984).

100 Lappin, T. R. J., Grier, D. G., Thompson, A. & Halliday, H. L. HOX genes: seductive science, mysterious mechanisms. *Ulster Med J* **75**, 23-31 (2006).

101 Brooke, N. M., Garcia-Fernàndez, J. & Holland, P. W. H. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* **392**, 920-922, doi:10.1038/31933 (1998).

102 Holland, P. W., Booth, H. A. F. & Bruford, E. A. Classification and nomenclature of all human homeobox genes. *BMC Biology* **5**, 47 (2007).

103 Holland, P. W. Evolution of homeobox genes. *Developmental Biology* **2**, 31-45 (2013).

104 Charite, J. *et al.* Transducing positional information to the Hox genes: critical interaction of cdx gene products with position-sensitive regulatory elements. *Development (Cambridge, England)* **125**, 4349-4358 (1998).

105 Neijts, R., Amin, S., van Rooijen, C. & Deschamps, J. Cdx is crucial for the timing mechanism driving colinear Hox activation and defines a trunk segment in the Hox cluster topology. *Developmental Biology* **422**, 146-154, doi:https://doi.org/10.1016/j.ydbio.2016.12.024 (2017).

106 van den Akker, E. *et al.* Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development (Cambridge, England)* **129**, 2181-2193 (2002).

107 Savory, J. G. *et al.* Cdx1 and Cdx2 are functionally equivalent in vertebral patterning. *Dev Biol* **330**, 114-122, doi:10.1016/j.ydbio.2009.03.016 (2009).

108 Keenan, I. D., Sharrard, R. M. & Isaacs, H. V. FGF signal transduction and the regulation of Cdx gene expression. *Dev Biol* **299**, 478-488, doi:10.1016/j.ydbio.2006.08.040 (2006).

109 Mankertz, J. *et al.* Functional crosstalk between Wnt signaling and Cdx-related transcriptional activation in the regulation of the claudin-2 promoter activity. *Biochemical and biophysical research communications* **314**, 1001-1007, doi:10.1016/j.bbrc.2003.12.185 (2004).

110 Kessel, M. & Gruss, P. Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. *Cell* **67**, 89-104, doi:https://doi.org/10.1016/0092-8674(91)90574-I (1991).

111 Grienenberger, A. *et al.* Tgfβ signaling acts on a Hox response element to confer specificity and diversity to Hox protein function. *Development (Cambridge, England)* **130**, 5445-5455, doi:10.1242/dev.00760 (2003).

112 Davidson, A. J. *et al.* cdx4 mutants fail to specify blood progenitors and can be rescued by multiple hox genes. *Nature* **425**, 300-306, doi:10.1038/nature01973 (2003).

113 Wang, Y. *et al.* Cdx gene deficiency compromises embryonic hematopoiesis in the mouse. *Proc Natl Acad Sci U S A* **105**, 7756-7761, doi:10.1073/pnas.0708951105 (2008).

114 McKinney-Freeman, S. L. *et al.* Modulation of murine embryonic stem cell-derived CD41+c-kit+ hematopoietic progenitors by ectopic expression of Cdx genes. *Blood* **111**, 4944-4953, doi:10.1182/blood-2007-11-124644 (2008).

115 Koo, S. *et al.* Cdx4 is dispensable for murine adult hematopoietic stem cells but promotes MLL-AF9-mediated leukemogenesis. *Haematologica* **95**, 1642-1650, doi:10.3324/haematol.2010.023168 (2010).

116 Dou, D. R. *et al.* Medial HOXA genes demarcate haematopoietic stem cell fate during human development. *Nature Cell Biology* **18**, 595-606, doi:10.1038/ncb3354 (2016).

117 Iacovino, M. *et al.* HoxA3 is an apical regulator of haemogenic endothelium. *Nature cell biology* **13**, 72-78, doi:10.1038/ncb2137 (2011).

118 Ng, E. S. *et al.* Differentiation of human embryonic stem cells to HOXA+ hemogenic vasculature that resembles the aorta-gonad-mesonephros. *Nature Biotechnology* **34**, 1168-1179, doi:10.1038/nbt.3702 (2016).

119 Doulatov, S. *et al.* Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. *Cell Stem Cell* **13**, 459-470, doi:10.1016/j.stem.2013.09.002 (2013).

120 Palis, J., Malik, J., McGrath, K. E. & Kingsley, P. D. Primitive erythropoiesis in the mammalian embryo. *Int J Dev Biol* **54**, 1011-1018, doi:10.1387/ijdb.093056jp (2010).

121 de Bruijn, M. F., Speck, N. A., Peeters, M. C. & Dzierzak, E. Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *The EMBO journal* **19**, 2465-2474, doi:10.1093/emboj/19.11.2465 (2000).

122 Sturgeon, C. M. *et al.* Primitive erythropoiesis is regulated by miR-126 via nonhematopoietic Vcam-1+ cells. *Developmental cell* **23**, 45-57, doi:10.1016/j.devcel.2012.05.021 (2012).

123 Sim, X., Cardenas-Diaz, F. L., French, D. L. & Gadue, P. A Doxycycline-Inducible System for Genetic Correction of iPSC Disease Models. *Methods in molecular biology (Clifton, N.J.)* **1353**, 13-23, doi:10.1007/7651_2014_179 (2016).

124 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).

125 La Motte-Mohs, R. N., Herer, E. & Zúñiga-Pflücker, J. C. Induction of T-cell development from human cord blood hematopoietic stem cells by Delta-like 1 in vitro. *Blood* **105**, 1431-1439, doi:10.1182/blood-2004-04-1293 (2005).

126 Schmitt, T. M. *et al.* Induction of T cell development and establishment of T cell competence from embryonic stem cells differentiated in vitro. *Nature immunology* **5**, 410-417, doi:10.1038/ni1055 (2004).

127 Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *Journal of immunological methods* **347**, 70-78, doi:10.1016/j.jim.2009.06.008 (2009).

128 Pimentel, H., Bray, N. L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods* **14**, 687-690, doi:10.1038/nmeth.4324 (2017).

129 Choi, K. D. *et al.* Identification of the hemogenic endothelial progenitor and its direct precursor in human pluripotent stem cell differentiation cultures. *Cell reports* **2**, 553-567, doi:10.1016/j.celrep.2012.08.002 (2012).

130 Davidson, A. J. & Zon, L. I. The caudal-related homeobox genes cdx1a and cdx4 act redundantly to regulate hox gene expression and the formation of putative hematopoietic stem cells during zebrafish embryogenesis. *Dev Biol* **292**, 506-518, doi:10.1016/j.ydbio.2006.01.003 (2006).

131 Paik, Elizabeth J. *et al.* A Cdx4-Sall4 Regulatory Module Controls the Transition from Mesoderm Formation to Embryonic Hematopoiesis. *Stem Cell Reports* **1**, 425-436, doi:https://doi.org/10.1016/j.stemcr.2013.10.001 (2013).

132 Vodyanik, M. A., Thomson, J. A. & Slukvin, II. Leukosialin (CD43) defines hematopoietic progenitors in human embryonic stem cell differentiation cultures. *Blood* **108**, 2095-2105, doi:10.1182/blood-2006-02-003327 (2006).

133 Ramos-Mejía, V. *et al.* HOXA9 promotes hematopoietic commitment of human embryonic stem cells. *Blood* **124**, 3065-3075, doi:10.1182/blood-2014-03-558825 (2014).

134 Lee, J. H., Protze, S. I., Laksman, Z., Backx, P. H. & Keller, G. M. Human Pluripotent Stem Cell-Derived Atrial and Ventricular Cardiomyocytes Develop from Distinct Mesoderm Populations. *Cell Stem Cell* **21**, 179-194.e174, doi:10.1016/j.stem.2017.07.003 (2017).

135 Alles, J. *et al.* Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biology* **15**, 44, doi:10.1186/s12915-017-0383-5 (2017).

136 Stuart, T. *et al.* Comprehensive integration of single-cell data. **177**, 1888-1902. e1821 (2019).

137 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology* **32**, 381-386, doi:10.1038/nbt.2859 (2014).

138 Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nature Methods* **12**, 963-965, doi:10.1038/nmeth.3542 (2015).

139 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

140 Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research* **41**, e108-e108, doi:10.1093/nar/gkt214 (2013).

141 Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

142 (in revision) Luff, S. A., Creamer, J.P., Dege, C., Scarfo, R., Moris, S., Ditadi, A., and Sturgeon, C.M. Identification of a retinoic acid-dependent definitive hematopoietic progenitor from human pluripotent stem cells. *Nature Biotechnology* (2020).

143 Tam, P. P. L. & Loebel, D. A. F. Gene function in mouse embryogenesis: get set for gastrulation. *Nature Reviews Genetics* **8**, 368-381, doi:10.1038/nrg2084 (2007).

144 Andrews, P. W. Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line in vitro. *Developmental Biology* **103**, 285-293, doi:https://doi.org/10.1016/0012-1606(84)90316-6 (1984).

145 Janesick, A., Wu, S. C. & Blumberg, B. Retinoic acid signaling and neuronal differentiation. *Cellular and Molecular Life Sciences* **72**, 1559-1576, doi:10.1007/s00018-014-1815-9 (2015).

146 Rhinn, M. & Dollé, P. Retinoic acid signalling during development. *Development (Cambridge, England)* **139**, 843-858, doi:10.1242/dev.065938 (2012).

147 Chanda, B., Ditadi, A., Iscove, Norman N. & Keller, G. Retinoic Acid Signaling Is Essential for Embryonic Hematopoietic Stem Cell Development. *Cell* **155**, 215-227, doi:https://doi.org/10.1016/j.cell.2013.08.055 (2013).

148 Creamer, J. P. *et al.* Human definitive hematopoietic specification from pluripotent stem cells is regulated by mesodermal expression of CDX4. *Blood* **129**, 2988-2992, doi:10.1182/blood-2016-11-749382 (2017).

149 Artimo, P. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic acids research* **40**, W597-603, doi:10.1093/nar/gks400 (2012).

150 Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**, W160-W165, doi:10.1093/nar/gkw257 %J Nucleic Acids Research (2016).

151 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28**, 495-501, doi:10.1038/nbt.1630 (2010).

152 Savory, J. G. A., Mansfield, M., St. Louis, C. & Lohnes, D. Cdx4 is a Cdx2 target gene. *Mechanisms of Development* **128**, 41-48, doi:https://doi.org/10.1016/j.mod.2010.09.004 (2011).

153 Lohnes, D. J. B. The Cdx1 homeodomain protein: an integrator of posterior signaling in the mouse. **25**, 971-980 (2003).

154 Nguyen, T. T. *et al.* Cdx2 Regulates Gene Expression through Recruitment of Brg1-associated Switch-Sucrose Non-fermentable (SWI-SNF) Chromatin Remodeling Activity. **292**, 3389-3399, doi:10.1074/jbc.M116.752774 (2017).

155 Zhang, X. *et al.* Transcriptional repression by the BRG1-SWI/SNF complex affects the pluripotency of human embryonic stem cells. *Stem cell reports* **3**, 460-474, doi:10.1016/j.stemcr.2014.07.004 (2014).

156 Singh, M. K. *et al.* Tbx20 is essential for cardiac chamber differentiation and repression of Tbx2. *Development (Cambridge, England)* **132**, 2697-2707, doi:10.1242/dev.01854 (2005).

157 Stennard, F. A. *et al.* Cardiac T-box factor Tbx20 directly interacts with Nkx2-5, GATA4, and GATA5 in regulation of gene expression in the developing heart. *Dev Biol* **262**, 206-224, doi:10.1016/s0012-1606(03)00385-3 (2003).

158 Foley, T. E., Hess, B., Savory, J. G. A., Ringuette, R. & Lohnes, D. Role of Cdx factors in early mesodermal fate decisions. **146**, dev170498, doi:10.1242/dev.170498 %J Development (2019).

159 Lengerke, C. *et al.* Interactions between Cdx genes and retinoic acid modulate early cardiogenesis. *Dev Biol* **354**, 134-142, doi:10.1016/j.ydbio.2011.03.027 (2011).

160 Schott, J.-J. *et al.* Congenital Heart Disease Caused by Mutations in the Transcription Factor NKX2-5. *Science (New York, N.Y.)* **281**, 108-111, doi:10.1126/science.281.5373.108 (1998).

161 Bruneau, B. G. *et al.* A Murine Model of Holt-Oram Syndrome Defines Roles of the T-Box Transcription Factor Tbx5 in Cardiogenesis and Disease. *Cell* **106**, 709-721, doi:https://doi.org/10.1016/S0092-8674(01)00493-7 (2001).

162 Kupzig, S., Parsons, S. F., Curnow, E., Anstee, D. J. & Blair, A. Superior survival of ex vivo cultured human reticulocytes following transfusion into mice. *Haematologica* **102**, 476-483, doi:10.3324/haematol.2016.154443 (2017).

163 Sheikh, F., Lyon, R. C. & Chen, J. Functions of myosin light chain-2 (MYL2) in cardiac muscle and disease. *Gene* **569**, 14-20, doi:https://doi.org/10.1016/j.gene.2015.06.027 (2015).

164 Xu, M. *et al.* CTCF controls HOXA cluster silencing and mediates PRC2-repressive higher-order chromatin structure in NT2/D1 cells. *Mol Cell Biol* **34**, 3867-3879, doi:10.1128/MCB.00567-14 (2014).

165 Ordovás, L. *et al.* Efficient Recombinase-Mediated Cassette Exchange in hPSCs to Study the Hepatocyte Lineage Reveals AAVS1 Locus-Mediated Transgene Inhibition. *Stem Cell Reports* **5**, 918-931, doi:https://doi.org/10.1016/j.stemcr.2015.09.004 (2015).

166 Chen, Z. *et al.* Genetic Engineering of Human Embryonic Stem Cells for Precise Cell Fate Tracing during Human Lineage Development. *Stem Cell Reports* **11**, 1257-1271, doi:https://doi.org/10.1016/j.stemcr.2018.09.014 (2018).

167 Chen, J. Y. *et al.* Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. *Nature* **530**, 223-227, doi:10.1038/nature16943 (2016).

168 Beckman, E. M. *et al.* Recognition of a lipid antigen by CD1-restricted alpha beta+ T cells. *Nature* **372**, 691-694, doi:10.1038/372691a0 (1994).

169 Brutkiewicz, R. R. CD1d Ligands: The Good, the Bad, and the Ugly. *The Journal of Immunology* **177**, 769-775, doi:10.4049/jimmunol.177.2.769 (2006).

170 Broxmeyer, H. E. *et al.* CD1d expression on and regulation of murine hematopoietic stem and progenitor cells. *Blood* **119**, 5731-5741, doi:10.1182/blood-2012-01-404012 (2012).

171 Pijuan-Sala, B. *et al.* A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490-495, doi:10.1038/s41586-019-0933-9 (2019).

172 Ronger-Savle, S. *et al.* TGFbeta inhibits CD1d expression on dendritic cells. *The Journal of investigative dermatology* **124**, 116-118, doi:10.1111/j.0022-202X.2004.23315.x (2005).

173 Sugimura, R. *et al.* Haematopoietic stem and progenitor cells from human pluripotent stem cells. **545**, 432-438 (2017).

174 Davidson, A. J. & Zon, L. I. The 'definitive' (and 'primitive') guide to zebrafish hematopoiesis. *Oncogene* **23**, 7233-7246, doi:10.1038/sj.onc.1207943 (2004).

175 van Nes, J. *et al.* The Cdx4 mutation affects axial development and reveals an essential role of Cdx genes in the ontogenesis of the placental labyrinth in mice. **133**, 419-428, doi:10.1242/dev.02216 %J Development (2006).

176 McGrath, K. E., Fegan, K. H., Frame, J. M., Kingsley, P. D. & Palis, J. Definitive Erythro-Myeloid Progenitors (EMP) Emerge in the Yolk Sac From Hemogenic Endothelium and Share Transcriptional Regulators with Adult Hematopoiesis. *Blood* **118**, 910-910, doi:10.1182/blood.V118.21.910.910 (2011).

177 Kokolus, K. & Nemeth, M. J. Non-canonical Wnt signaling pathways in hematopoiesis. *Immunologic research* **46**, 155-164, doi:10.1007/s12026-009-8116-7 (2010).

178 Lian, X. *et al.* Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proceedings of the National Academy of Sciences* **109**, E1848-E1857, doi:10.1073/pnas.1200250109 (2012).