STEM CELLS®

EMBRYONIC STEM CELLS/INDUCED PLURIPOTENT STEM CELLS

SMAD4 Is Essential for Human Cardiac Mesodermal Precursor Cell Formation

JIEJIA XU,[a] PETER J. GRUBER,[b] KENNETH R. CHIEN[a,c]

Key Words. *SMAD4* • Mesoderm • Heart • Human embryonic stem cells

ABSTRACT

Understanding stage-specific molecular mechanisms of human cardiomyocyte (CM) progenitor formation and subsequent differentiation are critical to identify pathways that might lead to congenital cardiovascular defects and malformations. In particular, gene mutations in the transforming growth factor (TGF)β superfamily signaling pathways can cause human congenital heart defects, and murine loss of function studies of a central component in this pathway, *Smad4*, leads to early embryonic lethality. To define the role of *SMAD4* at the earliest stages of human cardiogenesis, we generated *SMAD4* mutant human embryonic stem cells (hESCs). Herein, we show that the loss of *SMAD4* has no effect on hESC self-renewal, or neuroectoderm formation, but is essential for the formation of cardiac mesoderm, with a subsequent complete loss of CM formation during human ES cell cardiogenesis. Via transcriptional profiling, we show that *SMAD4* mutant cell lines fail to generate cardiac mesodermal precursors, clarifying a role of NODAL/SMAD4 signaling in cardiac mesodermal precursor formation via enhancing the expression of primitive streak genes. Since *SMAD4* relative pathways have been linked to congenital malformations, it will become of interest to determine whether these may due, in part, to defective cell fate decision during cardiac mesodermal precursor formation. STEM CELLS *2019;37:216–225*

SIGNIFICANCE STATEMENT

This article defines the role of *SMAD4* at the earliest stages of human cardiogenesis. It will also be of considerable interest to scientists interested in human stem cell models of development and disease, physician-scientists interested in congenital heart defects, and basic scientists interested in the transforming growth factor-β signaling pathways.

[a]Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden; [b]Department of Surgery, Yale University School of Medicine, New Haven, Connecticut, USA; [c]Department of Medicine, Karolinska Institutet, Huddinge, Sweden

Correspondence: Kenneth R. Chien, M.D., Ph.D., Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm SE-17177, Sweden. Telephone: +46 70 785 9365; e-mail: kenneth.chien@ki.se

Received April 5, 2018; accepted for publication October 9, 2018; first published online in STEM CELLS *EXPRESS* October 30, 2018.

http://dx.doi.org/10.1002/stem.2943

INTRODUCTION

The transforming growth factor (TGF)β superfamily signaling pathways have been implicated in a diverse array of biological processes, including cell proliferation, differentiation, morphogenesis, tissue homeostasis, and regeneration. The molecular framework for the TGFβ superfamily signaling pathways is well-described [1] with more than 30 ligands, 4 type-II, and 7 type-I receptors. SMAD proteins serve as intracellular mediators and SMAD family member participation in these pathways can be broadly subdivided into two groups: bone morphogenetic protein (BMPs) and TGFβ/Activin/NODAL. BMP signals phosphorylate SMAD1, SMAD5, and SMAD8, while SMAD2 and SMAD3 are phosphorylated in response to activation of the TGFβ, Activin, and NODAL signals. Phosphorylation of the R-SMADs (SMAD1/5/8 and SMAD2/3) permits their association with the common-mediator (co-Smad), SMAD4, resulting in nuclear translocation and formation of higher order transcriptional complexes. Thus, SMAD4 plays a central role in TGFβ superfamily signaling.

Both BMP and TGFβ/Activin/NODAL signaling pathways play an important role in human embryonic stem cell (hESC) self-renewal and differentiation. Mouse embryonic stem cells are routinely cultured in medium containing GSK3β and Mek1/2 inhibitors, which can, but not required to, include leukemia inhibitory factor. hESCs use distinct signaling pathways to support their self-renewal. The routine hPSC culture medium contains two essential growth factors: basic fibroblast growth factor (FGF2), TGFβ, or NODAL. FGF2 promotes hESC self-renewal by repressing BMP signaling pathways [2]. TGFβ or NODAL can increase the expression of the pluripotency gene NANOG and leads to consistent long-term culture stability of hESCs [3]. Inhibition of the TGFβ/Activin/NODAL receptors with SB431542, an inhibitor of Activin A receptor-like kinase ALK5 and its relatives ALK4 and ALK7, results in rapid differentiation of hESCs [2, 4–7].

BMP and Activin, on the other hand, can be used to induce differentiation of hPSCs. High-dose Activin A is used to induce mesendoderm differentiation [8], while the combination of BMP and Activin A is used to induce cardiomyocyte (CM) differentiation [9]. Although CM differentiation can be achieved by temporal manipulating WNT signaling pathways [10], Activin A and BMP signaling pathways are still essential. It has been shown that SB431542 can completely block CM differentiation [10]. And BMP inhibitors, such as DMH1, LDN193189 can partially block CM differentiation [11].

Loss-of-function studies in mice have revealed important roles of the TGFβ superfamily signaling pathways in early embryogenesis [12] and later embryonic heart development [13–15]. Previous studies have shown that the loss of function of *Smad4* in murine models leads to embryonic lethality before E7.5 with reduced growth and a secondary, noncell autonomous defect in gastrulation [16, 17]. Murine *Smad4* deficiency results in partial mesoderm formation, with abnormal cardiac development [18, 19]. Together, these studies demonstrated that the Smad4-null mice epiblast still maintain the capacity for CM differentiation.

In the human setting, in vitro CM differentiation recapitulates the in vivo regulatory pathways that control the establishment of the corresponding lineage in the early embryo [20, 21]. To define whether these studies result in a corresponding effect on human cardiogenesis, we used the CRISPR-Cas9 system to engineer *SMAD4* mutated hESCs, allowing the evaluation of the direct cell autonomous role of *SMAD4* in both hESC self-renewal and subsequent CM differentiation. First, we demonstrate that *SMAD4* is not required for long-term maintenance of hESCs. Second, we show that *SMAD4* is essential for cardiac mesodermal precursor cell formation, and reveal the distinct roles of WNT/β-catenin and NODAL/SMAD4 signaling pathways at the earliest step of human cardiogenesis.

Materials and Methods

Maintenance of hESCs

Human pluripotent stem cells (ES03, NIH code: HES3) were obtained from WiCell and maintained on Vitronectin (Thermo Fisher Scientific, USA) coated plates in Essential 8 (Thermo Fisher Scientific). And Versene (Thermo Fisher Scientific) was used to dissociate the hESCs for passage.

Cardiac Differentiation of hESCs Via the GiWi Method

After the hESCs achieved confluence on a Vitronectin-coated surface, the cells were dissociated with Versene (Thermo Fisher Scientific) at 37°C for 5 minutes and then seeded onto a Matrigel-coated cell culture dish at a density of 100,000 cells per cm^2 in Essential 8 supplemented with 5 μM Y-27632 (Sigma–Aldrich, St. Louis, MO; day −2) for 24 hours. Cells were then cultured in Essential 8 without Y-27632, and changed daily. Cardiac differentiation of hESCs was performed according to previously published GiWi methods [10]. Briefly, at day 0, cells were treated with 12 μM CHIR99021 (Sigma–Aldrich) for 24 hours in RPMI/B-27-ins medium (RPMI medium with B27 supplement minus insulin), followed by a change with RPMI/B27-ins medium at day 1. At day 3, half of the medium was changed to the RPMI/B27-ins medium containing 5 μM IWP2 (Sigma–Aldrich), which was then removed during the medium change on day 5. Cells were then cultured in RPMI/B27 medium (RPMI medium with B27 supplement) and changed every 2 days.

Cardiac Differentiation from hESC with Growth Factor Based Method

After the hESCs achieved confluence on the Vitronectin-coated surface, the cells were dissociated with Versene (Thermo Fisher Scientific) at 37°C for 5 minutes and then seeded onto a Matrigel-coated cell culture dish at 100,000 cells per cm^2 in Essential 8 supplemented with 5 μM Y-27632 (Sigma–Aldrich; day −2) for 24 hours. Cells were then cultured in Essential 8 without Y-27632, and changed daily. Cardiac differentiation of hESCs was performed according to previously published growth factor based methods [10, 22–24]. Briefly, at day 0, cells were treated with 100 ng/ml Activin A (R&D Systems Inc., Minneapolis, MN) in RPMI/B27-insulin. After 24 hours, the medium was changed to RPMI/B27-insulin supplemented with 5 ng/ml BMP4 (R&D Systems Inc.) for the next 4 days. At day 5, the medium was changed to RPMI/B27 and medium was changed every 2 or 3 days.

Plasmid Constructs

CRISPR-Cas9. *PiggyBac* CRISPR-Cas9 (PB-CRISPR) was constructed by cloning the sgRNA and Cas9 expression cascade from pLentiCRISPR [25] into a *PiggyBac* vector modified from PB-OSKML [26]. We designed two gRNAs targeting exon 3 of *SMAD4*, smad4-g1: TGATCTATGCCCGTCTCTGG and smad4-g2: AGACGGGCATAGATCACATG. The two gRNAs were cloned into PB-CRISPR similar to pLentiCRISPR vector, termed pPB-CRISPR-SMAD4-g1 and pPB-CRISPR-SMAD4-g2.

PiggyBac UbC-Smad4V5. V5-tagged *SMAD4* cDNA and GFP-Blasticidin chimeras were synthesized using IDT's Gene Block, and the fragment's cloned into modified PB-UbC vector. P2A sequence linked the SMAD4v5 with a chimera protein GFPBsd, which functions as both GFP and Blasticidin resistance. The whole cascades were integrated into the genome using *PiggyBac* System [26].

Cell Line Generation

SMAD4 Mutant hESC Lines. Three plasmids, pPB-CRISPR-SMAD4-g1, pPB-CRISPR-SMAD4-g2 and pCyL43, were used for genome editing in hESCs. Two million hESCs, 1 μg pCyL43 and 4 μg pPB-CRISPR-SMAD4-g1, pPB-CRISPR-SMAD4-g2 DNA were mixed in 100 μl nucleofection solution and then nucleofected with program B-16 using a Nucleofector 2b device (Lonza, Germany). After 2 weeks of 0.5 μg/ml Puromycin selection, cells were singularized and sparsely seeded to generate single cell derived clones.

SMAD4 *Re-Expression hESC Lines.* Two plamsids, pPB-UbC-SMAD4V5 and pCyL43, were used to generate *SMAD4* overexpression/rescue hESCs. Two million hESCs, 1 μg pCyL43 and 4 μg pPB-UbC-SMAD4V5 DNA were mixed in 100 μl nucleofection solution and then nucleofected with program B-16 using a Nucleofector 2b device (Lonza). After 2 weeks of 5 μg/ml Blasticidin selection, cells were singularized and sparsely seeded to generate single cell derived clones. Colonies with homogeneous GFP expression were picked for final rescue experiments.

Sequencing

Genomic DNA was extracted from putative *SMAD4* mutated hESC clone cells using Quick-gRNA MiniPrep kit (Zymo Research, D3007). The targeted genome region was amplified with the following primer pairs, forward: GACAAGGGTCTTGCC ATGTT and reverse: TAAAGTCGCGGGCTATCTTC. The amplified PCR fragment was inserted into pCR4TOPO vector via the TOPO-TA cloning system (Thermo Fisher Scientific). Plasmids from individual *Escherichia coli* clones were Sanger Sequencing using a T7 primer.

Antibodies

POU5F1 (sc-5279, Santa Cruz Biotechnology Inc., Santa Cruz, CA); NANOG (4903S, Cell Signaling Technology, Beverly, MA); SOX2 (3579, Cell Signaling); PAX6 (Developmental Studies Hybridoma Bank [DSHB], Iowa City, IA); TBXT (AF2085, R&D Systems Inc.); MESP1 (ab77013, Abcam, Cambridge, U.K.); ISL1 (39.4D5, DSHB); NKX2-5 (sc-14033, Santa Cruz Biotechnology Inc.); TNNT2 (MS-295-P1, Thermo Fisher Scientific); V5 (R960-25, Thermo Fisher Scientific); SMAD4 (ab110175, Abcam; sc-7966, Santa Cruz Biotechnology Inc.); GAPDH (8884S, Cell Signaling); CXCR4 (ab1670, Abcam); APLNR (702069, Thermo Fisher Scientific); CD13 (ab7417, Abcam).

Flow Cytometry

Cells were dissociated into single cells with Accutase (Thermo Fisher Scientific) for 10 minutes, and subsequently fixed with 1% paraformaldehyde for 20 minutes at room temperature and stained with primary and secondary antibodies in phosphate buffered solution (PBS) with 0.2% Triton X-100% and 0.5% bovine serum albumin (BSA). Data were collected on a FACSCaliber or FASCanto flow cytometer (Becton, Dickinson and Company, Franklin Lakes, NJ) and analyzed using FlowJo. Fluorescence-activated cell sorter (FACS) gating was based on the corresponding isotype or secondary only antibody control.

Immunostaining

Cells were fixed with 4% paraformaldehyde for 15 minutes at room temperature and stained with primary and secondary antibodies in PBS plus 0.4% Triton X-100% and 5% nonfat dry milk. Nuclei were stained with DAPI. A Zeiss LSM 700 confocal microscope was used for imaging.

Western Blot Analysis

Cells were lysed in M-PER Mammalian Protein Extraction Reagent (Thermo Fisher Scientific) in the presence of Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific). Proteins were separated by NuPAGE 4%–12% Bis–Tris Protein Gels (Thermo Fisher Scientific) and transferred to a nitrocellulose membrane with Trans-Blot Turbo Transfer system (Bio-Rad, Hercules, CA). After blocking with 5% nonfat dry milk in TBST, the membrane was incubated with primary antibody overnight at 4°C. The membrane was then washed, incubated with an anti-rabbit peroxidase-conjugated secondary antibody at room temperature for 1 hour, and developed using SuperSignal West Femto Maximum sensitivity substrate (Thermo Fisher Scientific).

RNA Sequencing

Total RNA was isolated with Direct-Zol RNA Kits (Zymo Research, USA). RNA quality was checked using the BioAnalyzer 2100 (Agilent Technologies, Palo Alto, CA). Library preparation type was Illumina TruSeq Stranded mRNA, Poly-A selection. Samples were sequenced on an Illumina HiSeq2500 with a 2×126 setup using HiSeq SBS Kit v4 chemistry. Basic data processing was performed in the SciLifeLab facility. Briefly, the FASTQ sequence files were mapped to Human Genome, GRCh37 with STAR aligner [27]. The outputted BAM alignment files were used for further analysis described below.

RNA-Seq Analysis

The expression count table was produced by feeding FeatureCounts [28] with the BAM files produced in the previous step and annotation file gencode.v27lift37 from GENCODE. Differential gene expression analysis was performed using the R package edgeR [29]. Gene expression pattern analysis was performed with *k*-mean function in R. Analysis and graphs were produce with custom scripts available by request.

Results

SMAD4 Is Not Required for hESC Self-Renewal

hESCs can be cultured long-term with the maintenance of pluripotency using a chemically defined E8 medium, but in the absence of TGFβ or NODAL, hESCs can become unstable and will differentiate after long-term passage [3]. Since *SMAD4* is an irreplaceable component in these signaling pathways, *SMAD4* mutant hESCs might interrupt TGFβ/NODAL signaling, leading to instability following long-term passage. Despite the clear requirement of TGFβ/NODAL signaling in the maintenance of pluripotency, the requirement of *SMAD4* is still unclear.

Accordingly, we used the CRISPR-Cas9 system to generate six *SMAD4* mutated hESC clonal cell lines, with further analysis of two independent lines. The targeted region was sequenced, confirming that the two alleles of clone 1 and 2 had deletions in the targeted site (Fig. 1A). At the protein level, both Western blot and immunostaining confirmed the *SMAD4* mutant cannot produce stable SMAD4 proteins (Fig. 1B, 1C). The success of generating *SMAD4* mutant hESC clones demonstrated their self-renewing capacity. *SMAD4* mutants can be expanded and passaged in E8 medium without morphological defects in standard culture up to 6 months (Fig. 1B). Furthermore, all of the *SMAD4* mutant clones expressed the pluripotency markers: POU5F1, NANOG, and SOX2 (Fig. 1C and Supporting Information Fig. S4A).

The E8 medium we used, was based on FGF2 and TGFβ. Since FGF2 and NODAL can also support pluripotency, we further tested the alternative medium. We added 100 ng/ml FGF2 to E6, which is E8 without FGF2 and TGFβ, as a basal medium, which we designated as medium E7. In total, four media were tested, E7, E7 with 100 ng/ml NODAL, E7 with 2 ng/ml TGFβ, and E7 with 2 ng/ml Activin A. Since the *SMAD4* mutants cannot respond to NODAL, TGFβ, or Activin A, they can be maintained in all four media, while wild-type ES03 differentiated in E7 and Activin A after a few days in culture (Supporting Information Fig. S3). This result further demonstrates that SMAD4 is not required for hESC self-renewal.

We also used RNA-Seq to compare wild-type and *SMAD4* mutant hESCs. In *SMAD4* mutants, splicing appeared normal (Supporting Information Fig. S1A, S1B), but the DNA-binding domain was disrupted (Supporting Information Fig. S1C).

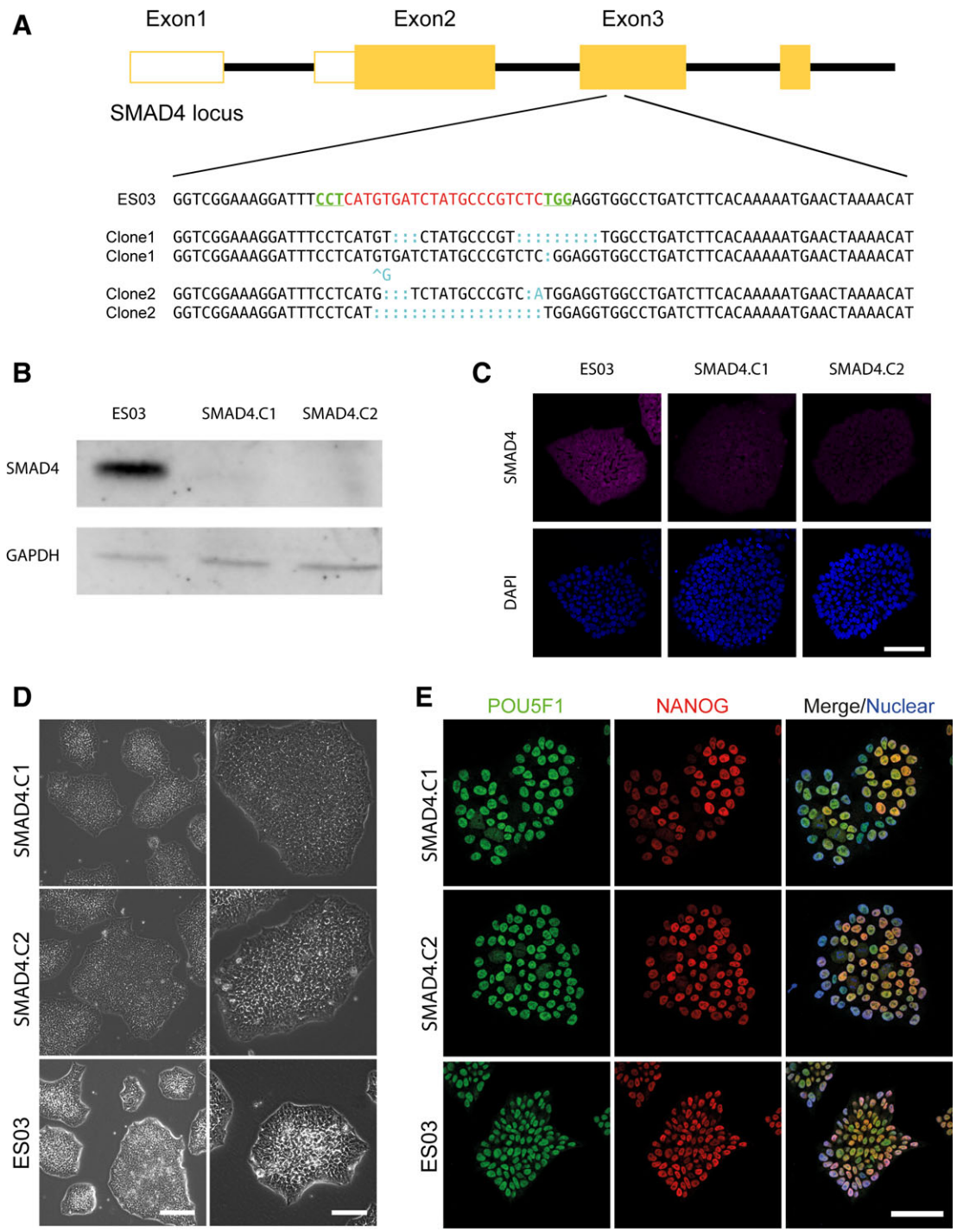

Figure 1. *SMAD4* is not required for the maintenance of human embryonic stem cells (hESCs). **(A):** Mutated sequences in the *SMAD4* loci. Clone 1, two patterns: 12 deletion/1 insertion-1 deletion; clone 2, two patterns: 4 deletion/18 deletion. **(B):** Western blot demonstrates absence of detectable SMAD4 protein in mutant clones SMAD4.C1 and SMAD4.C2. **(C):** Immunostaining demonstrates absence of detectable SMAD4 protein in mutant clones SMAD4.C1 and SMAD4.C2 (scale bar = 100 μm). **(D):** Bright field images demonstrate normal hESC morphology of *SMAD4* mutant clones compared with wild-type hESC (scale bar: left = 400 μm; right = 200 μm). **(E):** Immunostaining images demonstrate both SMAD4 mutant and the wild-type hESCs express the pluripotency markers POU5F1 in green and NANOG in red (scale bar = 100 μm).

Mutation in one allele of clone 2 introduced a premature stop codon, whereas the two alleles in clone 1 and the other allele in clone 2 might produce mutated proteins. All mutations were found in the link between the β-hairpin and the L4 double loop within the DNA-binding domain of the MH1 domain [30].

Differential gene expression analysis of the wild-type and *SMAD4* mutant cell lines revealed minimal differences (Supporting Information Fig. S1D and Table S1). Pluripotency genes, including *POU5F1*, *NANOG*, *SOX2*, were not differentially expressed (Supporting Information Table S1). In contrast,

TGFβ signaling and target genes including *LEFTY1*, *LEFTY2*, and *CER1* (Supporting Information Table S1), were down regulated as expected, supporting *SMAD4* loss-of-function. Taken together, these data demonstrate that *SMAD4* is not essential for hESC self-renewal.

SMAD4 Is Essential for CM Differentiation

In early differentiation protocols, BMP4 and Activin A were used to differentiate hESCs to cardiomyocytes [9]. Both BMP4 and Activin A require *SMAD4* for proper function. Although some BMPs or Activin A inhibitors have been shown to block or reduce the differentiation process [10, 11], direct evidence is lacking to show that *SMAD4* is essential for cardiac mesodermal formation and/or downstream CM differentiation.

To determine if *SMAD4* is essential for human CM differentiation, we compared wild type (WT) and *SMAD4* mutant hESCs in a CM differentiation assay using the GiWi protocol [10]. Though WT differentiated hESCs produced robust beating CMs, after 14 days, *SMAD4* mutants produced no mature beating cells (n = 4). Furthermore, we tested both WT and mutant *SMAD4* hESC GiWi treated cells at day 14 for the expression of TNNT2, a marker of CM sarcomere formation, via flow cytometry. Similar to the functional beating CM assay, more than 80% of WT hESCs showed expression of TNNT2, while both *SMAD4* mutant clones produced no detectable TNNT2+ cells (Fig. 2A and Supporting Information Fig. S3). To our knowledge this result is the first direct evidence that *SMAD4* is absolutely essential for human CM differentiation from hESCs.

To prove that the failure of CM differentiation is due to the loss of *SMAD4*, we re-expressed *SMAD4* in the mutant cells using a human Ubiquitin C promoter (UbC) driving the expression of a C-terminal V5 tagged SMAD4. We examined transfected hESC cells by immunostaining of POU5F1, SOX2, and V5, marking expression of tagged SMAD4. *SMAD4* mutants stained positive for POU5F1, SOX2, but not V5, while the rescued mutants stained positive for V5, indicating the expression of the tagged SMAD4 (Supporting Information Fig. S4A). Following differentiation using the GiWi protocol, both rescued clones had only a few beating clusters and the flow cytometry showed 1% of TNNT2+ cells (Supporting Information Fig. S5). This is in contrast to the nonrescued mutant clones that produced no detectable CMs. We hypothesized that the low efficiency was due to heterogeneity of the rescue pools suppression of differentiation through a noncell autonomous mechanism. To address this possibility, we generated single cell derived clones from the rescue pools. The single cell derived clones express SMAD4v5 homogeneously at the hESC stage and the differentiation efficiency increased to 36%, statistically equivalent to the WT (p = .266; Fig. 2C, 2D and Supporting Information Fig. S6). These data strongly support that the failure of CM differentiation in the CRISPR-induced *SMAD4* mutants is due to the mutation in *SMAD4*, and cannot be ascribed to off-target effects.

To further demonstrate the requirement of *SMAD4* in human CM differentiation from hESC is independent of the differentiation protocol, we used an Activin A/BMP4 protocol adopted from previous studies [10, 22–24] to differentiate the *SMAD4* mutated hESCs into CM. After 14 days of differentiation, the wild-type ES03 yielded 40% TNNT2+ CMs, while the mutants produced no detectable CMs (Fig. 2D, 2E). Taken together, these results support the conclusion that *SMAD4* is essential for human CM differentiation from hESCs.

RNA-Seq Analysis Reveals the Distinct Roles of WNT/β-Catenin and NODAL/SMAD4 Signaling Pathways

As established in a previous study in hESCs [31], in vitro cardiogenesis occurs via sequential distinct developmental stages, corresponding to the parental embryonic stem cells (day 0), primitive streak/early mesoderm (day 1), cardiac mesoderm (day 3), and cardiac progenitor/early CM stages (day 6). To understand the downstream effect of *SMAD4* mutations in this sequential process of human in vitro cardiogenesis, we used RNA-Seq to comprehensively compare the transcriptional profiles between the WT and the *SMAD4* mutants at these stages. The relationship between the *SMAD4* mutant and the WT cell lines is displayed in the multidimension scaling plot (Fig. 3A). During differentiation, the number of differentiated expressed genes increased beginning with hESCs, 12 up, 8 down and by day 6 of differentiation, 223 up, 573 down (log fold change, $\log_2$ FC > 3, p-value $<10^{-5}$; Fig. 3B).

We subsequently compared the expression of ligands from three major signaling pathways (TGFβ superfamily, WNT, and FGF) between *SMAD4* mutants and wild-type differentiating hESC (Fig. 3C). At day1 after induction, the signaling cues for initiating the differentiation program were formed in WT cells, including *DKK4*, *DKK1*, *FGF4*, *FGF19*, *FGF13, FST*, *LEFFTY1*, *NODAL*, and *WNT5B.* In contrast, *SMAD4* mutants failed to induce *DKK1*, *DKK4*, *FGF4*, and *FGF13*, suggesting these might be the downstream targets of SMAD4. It was notable that NODAL was one of the very few TGFβ pathway ligands expressed (Fig. 3C). Furthermore, *NODAL* was expressed in both wild-type and *SMAD4* mutants, suggesting its expression was induced through WNT/β-catenin signaling. This result supports the notion that in the absence of SMAD4, hESCs cannot respond to NODAL, NODAL/SMAD4 downstream targets remain inactive, and fail to induce the molecular program necessary for the earliest stages of cardiogenesis.

To further study the initial formation of cardiac mesoderm, we examined the effects of WNT/β-catenin activation at day 1. As the first step of GiWi protocol, high-dose GSK3β inhibitor (CH99021, 12 μM) activated the primitive streak (PS) genes: *T* (*TBXT*), *EOMES*, *MIXL1*, and *GSC* (Fig. 3D). The expression of PS genes in *SMAD4* mutant cells was clearly weaker than wild-type cells (Fig. 3D and Supporting Information Fig. S7B, S7C), suggesting the initial expression of PS genes might be dependent on WNT/β-catenin, but the level and duration was regulated by NODAL/SMAD4. Consistent with a previous study [32], we noted the proneuroectoderm genes: *GBX2*, *AXIN2* were also induced (Fig. 3D). These results untangle the roles of WNT/β-catenin and NODAL/SMAD4 signaling pathways in human mesoderm induction. With WNT stimulation, primitive streak (PS) genes are expressed, but relatively weakly, and it appears that the NODAL/SMAD4 pathway enhances PS gene expression and takes the activated cells into formation of mesodermal lineages.

SMAD4 Is Required for the Formation of Human Cardiac Mesodermal Precursor Cells

Consistent with reduced expression of PS genes, there was a failure of mesendoderm induction, attenuating the expression of *CXCR4, HAND1, MESP1*, *SOX17*, and *TBX6* (Fig. 3D and Supporting Information Fig. S7A). However, at the protein level, mesendoderm markers, such as APLNR, ANPEP (CD13), CXCR4,

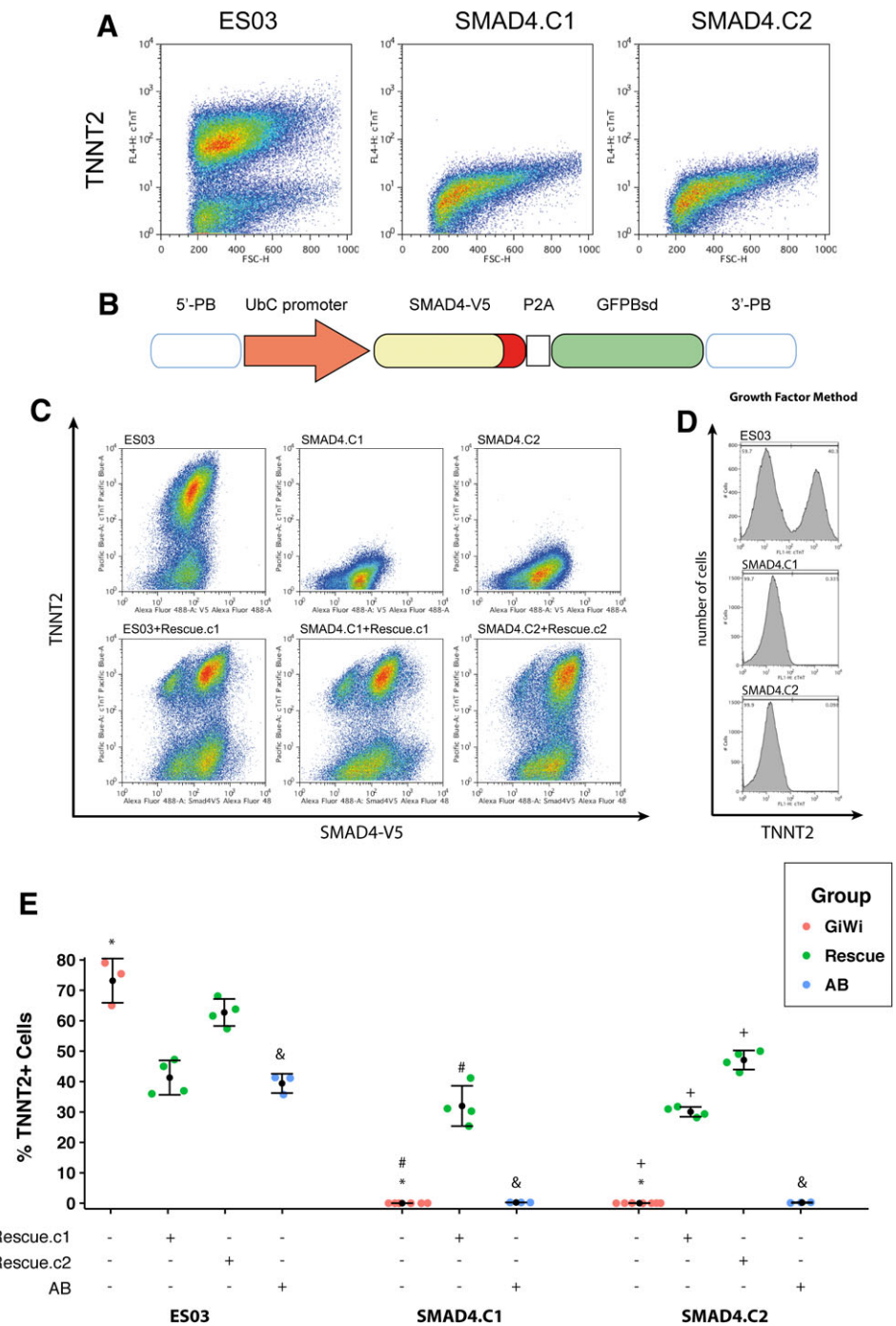

Figure 2. *SMAD4* is required for human cardiomyocyte differentiation. **(A):** Flow cytometry analysis demonstrates *SMAD4* mutant human embryonic stem cells (hESCs) fail to form TNNT2+ cardiomyocytes. Compare left panel hESC with middle and right panel mutant SMAD4 clones C1 and C2, respectively. **(B):** SMAD4 rescue construct used to overexpress/rescue SMAD4 in wild-type and the mutant clones contains a UbC promoter driving V5-tagged *SMAD4*. **(C):** Flow cytometry analysis demonstrates rescued clones recue cardiac differentiation to became TNNT2+ cardiomyocytes. **(D):** Flow cytometry analysis demonstrates *SMAD4* mutated hESCs failed to become TNNT2+ cardiomyocyte using growth factor-based differentiation method. **(E):** Summary graph of TNNT2 flow cytometry of SMAD4 WT, mutant clones, rescue clones, and growth factor-based differentiation method. *, $p < .0001$; compared with ES03 ($n \geq 3$). #, $p < .0001$; compared with SMAD4.C1 ($n = 4$). +, $p < .0001$; compared with SMAD4.C2 ($n = 4$). ES03 + Rescue.C2 with similar level SMAD4v5 expression as the mutant rescue clones. Ns, $p = .2664$ ($n = 4$). &, $p < .0001$; compared with ES03.AB ($n = 3$). AB indicates Activin a/BMP4 differentiation protocol.

were expressed in day 2 cells (Supporting Information Fig. S8). This suggests that these genes were stimulated by WNT/β-catenin, but required SMAD4 relative signaling pathways to correctly direct cells into the mesendoderm cell fate. We also used immunostaining to verify that the *SMAD4* mutants failed to form MESP1+ cardiac mesodermal precursor cells. Compared with the wild-type cells, the expression of MESP1 in the *SMAD4* mutants was much lower. Although the cells were still in the early stage, cardiac progenitor marker ISL1 and early CM marker NKX2-5 also were expressed at a very low level in

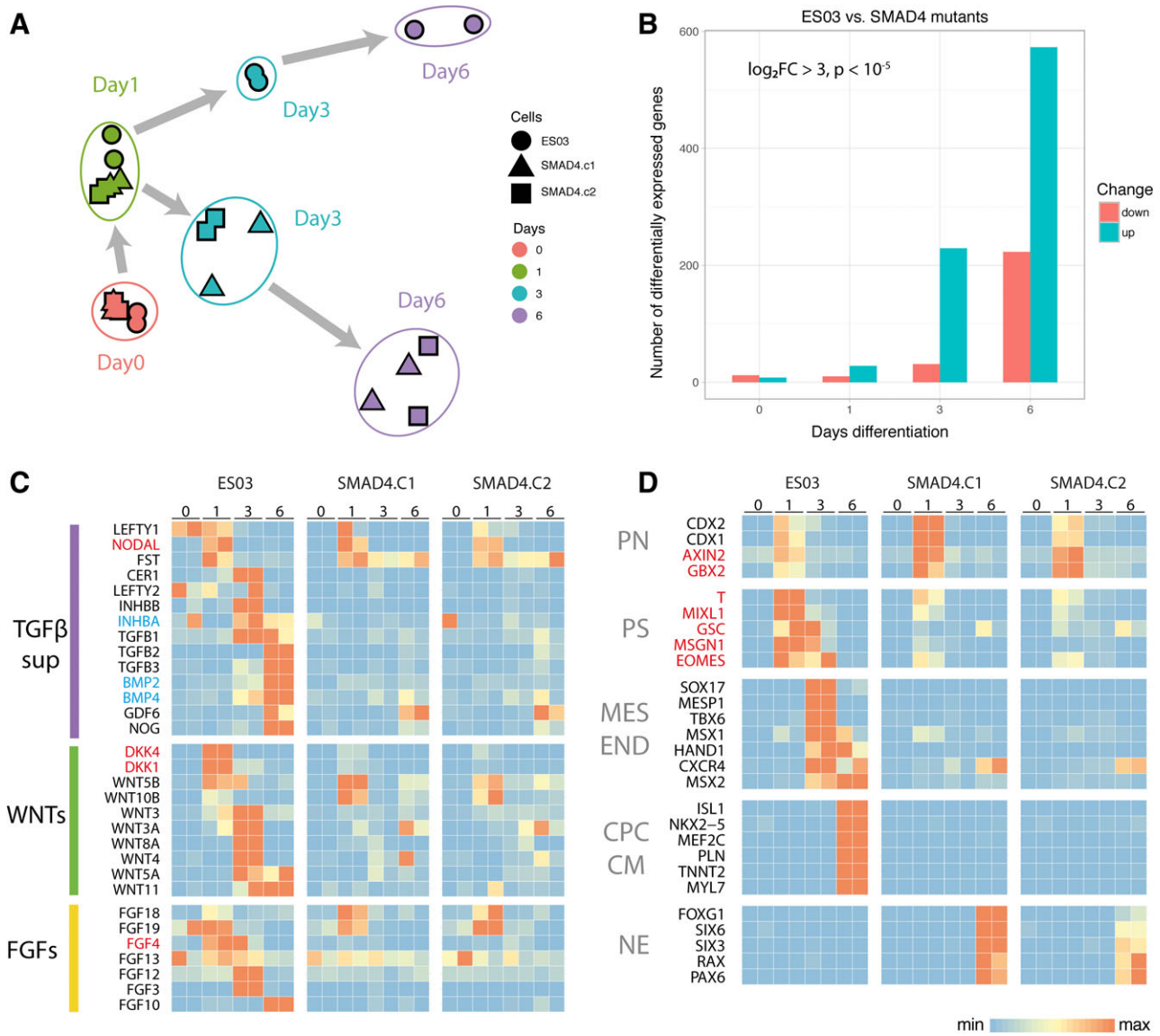

Figure 3. RNA-Seq reveals stage-specific difference between the wild-type and *SMAD4* mutants. **(A):** MDS plot demonstrates the similarities in transcriptional profiles between samples. **(B):** Increasing numbers of differentially expressed genes comparing SMAD4 wild-type and mutants at 0, 1, 3, and 6 days. **(C):** Heat map demonstrates the dynamic expression of ligands from WNT, TGFβ superfamily, and FGF signaling pathways. **(D):** Heat map demonstrates representative marker genes expression. Abbreviations: PN, proneuroectoderm; PS, primitive streak; MESEND, mesendoderm; CPC, cardiac progenitor cells; CM, cardiomyocyte; NE, neuroectoderm; MDS, multidimensional scaling.

the wild-type. However, in the *SMAD4* mutants, ISL1 and NKX2-5 were at negligible levels below the level of detection. Taken together, these data suggest the *SMAD4* mutants fail to form cardiac mesoderm precursor cells.

By day 6, wild-type differentiating hESCs became *ISL1*+ cardiac progenitors [33] and began to express early CM markers: *NKX2-5*, *TNNT2*, *MYL7*, and *PLN*. At the same stage of differentiation, *SMAD4* mutant cells did not express any of these cardiac genes. In contrast, *SMAD4* mutant cells expressed *FOXG1, PAX6*, *RAX*, and *SIX6*, consistent with a neuroectoderm fate [34] (Fig. 3D). We also confirmed the expression of *PAX6* in the mutant cells, but not in wild-type, by immunostaining (Fig. 4B). Taken together, these results demonstrate that the *NODAL*/*SMAD4* pathway plays an important role in the early cell fate decision between mesendoderm and neuroectoderm in human cells, and it essential for the formation of cardiac mesoderm precursor cells.

Discussion

SMAD4 and hESC Self-Renewal

Early studies have demonstrated the essential roles for TGFβ and NODAL signaling play in hESC self-renewal [2, 4–6]. SMAD4 plays a critical role in the modulation of TGFβ signaling, suggesting it would be important for hESC self-renewal. Avery et al. first challenged this idea showing that *SMAD4* knockdown in hESCs prevents the rapid differentiation normally induced by the TGFβ inhibitor SB431542. Furthermore, Avery observed a loss of the stability of hESC cultures and differentiation toward neural lineage. Thus, their result demonstrated that SMAD4 was not required for hESC self-renewal in the short-term. Our results agree with this short-term assessment. However, while Avery's SMAD4 KD cells were unstable in the long-term, our SMAD4 mutant cells are stable in the

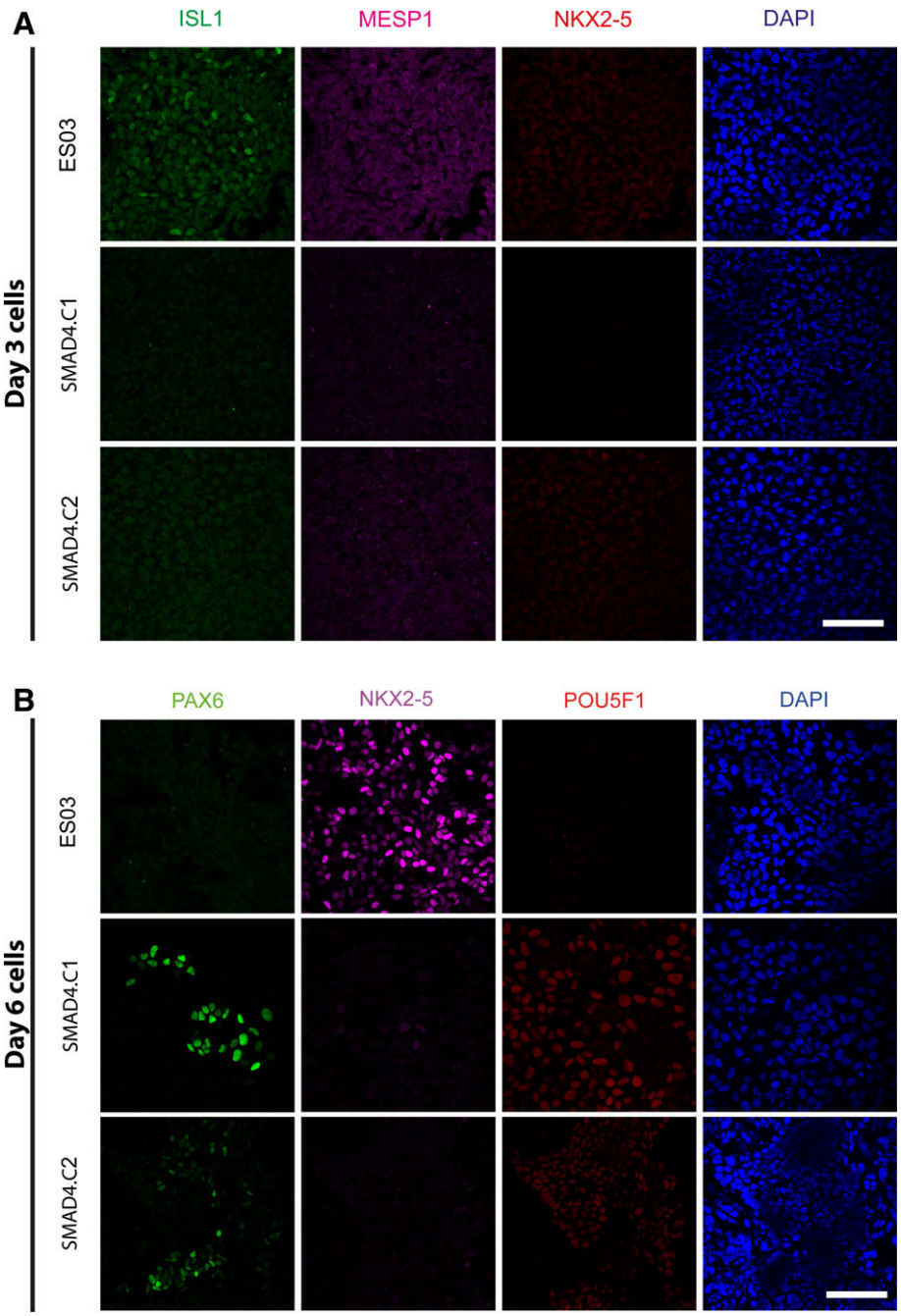

Figure 4. *SMAD4* mutant fail to form cardiac mesoderm precursor cells and defaults to neuroectoderm fate under conditions for cardiac differentiation. **(A):** Immunostaining demonstrates expression of the cardiac progenitor marker ISL1, cardiac mesoderm marker MESP1 and cardiomyocyte marker NKX2-5 after 3 days cardiac differentiation. Scale bar = 100 μm. **(B):** Immunostaining demonstrates expression of the pluripotent marker POU5F1, cardiac marker NKX2-5, and neuroectoderm marker PAX6 after 6 days cardiac differentiation. Scale bar = 100 μm.

long-term. There are several possible explanations for these differences. First are differences in culture conditions- we used a chemically defined media, Essential 8, whereas, Avery used a Dulbecco's modified Eagle's medium supplemented with knockout serum replacement, nonessential amino acids, L-glutamine, and βFGF. Another possible explanation for the discrepancy between Avery's work and ours is the techniques used to eliminate *SMAD4* expression. Avery used a knockdown approach that may have contained residual, functional SMAD4. In contrast, we inactivated *SMAD4* using genetic techniques. Our *SMAD4* mutants demonstrate a loss of both SMAD4 protein and SMAD4 function. Thus, in the current study, we demonstrate *SMAD4* is dispensable for hESC long-term self-renewal.

SMAD4 and Human Cardiac Mesodermal Formation and CM Differentiation

One of the central advantages of human ES cell models of cardiogenesis lies in the ability to study the role of specific signaling pathways at the earliest stages of mesodermal formation. Herein, we document the sequential expression of primitive streak marker *TBXT* at day 1, cardiac mesoderm marker *MESP1* at day 3, and cardiac progenitor marker, *ISL1*, at day 6.

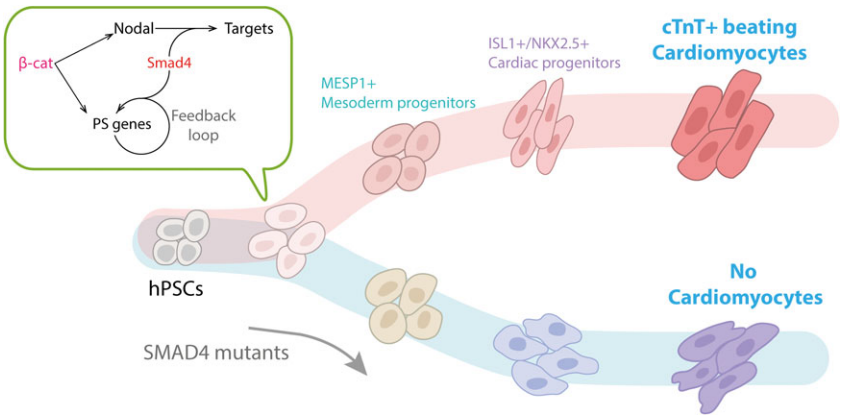

Figure 5. Model summarizing the essential roles of *SMAD4* in human cardiomyocyte differentiation. SMAD4 is essential for human cardiac mesodermal precursor cell formation.

However, in *SMAD4* mutants, *TBXT* was expressed at a lower level than in wild-type hESCs, while the expression of *MESP1* and *ISL1* were absent in *SMAD4* mutants. Mechanistically, our RNA-Seq results suggest that the differentiation programs diverge at the PS stage, with lower expression level of PS genes, including *TBXT*, *EOMES*, and *MIXL1.* This suggests that the trigger of these genes was present, but the right positive feedback or the necessary enhancer binding was missing. Previous studies [32, 35] showed that β-catenin can trigger the expression of PS genes. But without the feedback from NODAL/SMAD4 signaling pathway, these genes could not reach the levels that are required to drive the downstream signaling cascades. When the first wave of signal cubes fades away, the cells just drop back to their default neuroectoderm fate (Fig. 5). Thus, the enhancement of PS gene expression from NODAL/SMAD4 signaling pathway is essential for human cardiac mesodermal formation and subsequence CM differentiation.

Distinct Role of SMAD4 in Human Versus Murine Models of Cardiogenesis

Homozygous *Smad4* null mice exhibit growth retardation, develop abnormal visceral endoderm, and fail to form mesoderm [16, 17]. Aggregation of mutant *Smad4* mouse embryonic stem cells (mESCs) with wild-type tetraploid morulae rescues the gastrulation defect [17], suggesting that the failure to gastrulate and form mesoderm is secondary and noncell autonomous. Inactivation of *Smad4* at the epiblast stage with *Sox2*Cre permits gastrulation and mesoderm formation. The expression of cardiac differentiation stage markers *Tbxt*, *Mesp1*, *Nkx2-5*, and *Actc1* can be detected at corresponding stages [18, 19]. In the absence of *Smad4*, the cardiac development of the heart is severely compromised with multiple abnormal structures, though in some mutant embryos, a primitive, and unlooped heart tube is observed [19]. This indicates that in mice, cardiac lineage specification occurs in the absence of *Smad4.* In contrast, our current study demonstrates that human *SMAD4* mutants fail to form either cardiac mesoderm or cardiomyocytes, reflecting the differential requirements of SMAD4 in early cardiogenesis between human and mice.

Conclusion

Existing human genetic data supports our in vitro findings. In humans, recessive *SMAD4* somatic mutations are strongly correlated with the development of pancreatic and other cancers [36, 37]. Germline haploinsufficiency of *SMAD4* is associated with the autosomal dominant syndrome juvenile polyposis/hereditary hemorrhagic telangiectasia [38, 39]. In contrast, heterozygous gain of function *SMAD4* mutations are associated with Myhre Syndrome that includes cardiac structural defects [40]. Furthermore, a subtle *SMAD4* mutation (rs281875322) was reported in a patient with aortic arch hypoplasia and a ventricular septal defect [41]. To date, 539 *SMAD4* variants have been reported in ClinVar [42], yet no homozygous mutations have been reported in humans. The lack of any existing report of a human with homozygous *SMAD4* mutations supporting *SMAD4*'s essential role in human development. The current studies suggest a close examination of other known or novel genetic variants for congenital cardiac malformations may also be found at the earliest stages of cardiac mesodermal formation in human pluripotent stem cell model systems.

Acknowledgments

We would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure. pLentiCRISPR was a gift from Feng Zhang at Massachusetts Institute of Technology. We thank Wellcome Trust Sanger Institute for making the *PiggyBac* material available. We also thank Elif Eroglu and Chuen Yan Leung for proofreading. This work was supported by grants from the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

Author Contributions

J.X.: conception and design, collection and assembly of data, data analysis and interpretation, manuscript writing; P.J.G.: data analysis and interpretation, manuscript writing; K.R.C.: conception and design, financial support, data analysis and interpretation, manuscript writing, final approval of manuscript.

Disclosure of Potential Conflicts of Interest

The authors indicated no potential conflicts of interest.

REFERENCES

1 Massague J. TGFbeta signalling in context. Nat Rev Mol Cell Biol 2012;13:616–630.
2 Xu RH, Peck RM, Li DS et al. Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. Nat Methods 2005;2:185–190.
3 Chen G, Gulbranson DR, Hou Z et al. Chemically defined conditions for human iPSC derivation and culture. Nat Methods 2011;8:424–429.
4 Xu RH, Sampsell-Barron TL, Gu F et al. NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. Cell Stem Cell 2008;3:196–206.
5 James D, Levine AJ, Besser D et al. TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development 2005;132:1273–1282.
6 Vallier L, Mendjan S, Brown S et al. Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development 2009;136:1339–1349.
7 Avery S, Zafarana G, Gokhale PJ et al. The role of SMAD4 in human embryonic stem cell self-renewal and stem cell fate. Stem Cells 2010;28:863–873.
8 D,Amour KA, Agulnick AD, Eliazer S et al. Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat Biotechnol 2005;23:1534–1541.
9 Kattman SJ, Witty AD, Gagliardi M et al. Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. Cell Stem Cell 2011;8:228–240.
10 Lian X, Hsiao C, Wilson G et al. Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. Proc Natl Acad Sci USA 2012;109:E1848–E1857.
11 Burridge PW, Matsa E, Shukla P et al. Chemically defined generation of human cardiomyocytes. Nat Methods 2014;11:855–860.
12 Goumans MJ, Mummery C. Functional analysis of the TGFbeta receptor/Smad pathway through gene ablation in mice. Int J Dev Biol 2000;44:253–265.
13 Wang J, Xu N, Feng X et al. Targeted disruption of Smad4 in cardiomyocytes results in cardiac hypertrophy and heart failure. Circ Res 2005;97:821–828.
14 Qi X, Yang G, Yang L et al. Essential role of Smad4 in maintaining cardiomyocyte proliferation during murine embryonic heart development. Dev Biol 2007;311:136–146.
15 Song L, Yan W, Chen X et al. Myocardial smad4 is essential for cardiogenesis in mouse embryos. Circ Res 2007;101:277–285.
16 Yang X, Li C, Xu X et al. The tumor suppressor SMAD4/DPC4 is essential for epiblast proliferation and mesoderm induction in mice. Proc Natl Acad Sci USA 1998;95:3667–3672.
17 Sirard C, de la Pompa JL, Elia A et al. The tumor suppressor gene Smad4/Dpc4 is required for gastrulation and later for anterior development of the mouse embryo. Genes Dev 1998;12:107–119.
18 Costello I, Pimeisl IM, Drager S et al. The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. Nat Cell Biol 2011;13:1084–1091.
19 Chu GC, Dunn NR, Anderson DC et al. Differential requirements for Smad4 in TGFbeta-dependent patterning of the early mouse embryo. Development 2004;131:3501–3512.
20 Murry CE, Keller G. Differentiation of embryonic stem cells to clinically relevant populations: Lessons from embryonic development. Cell 2008;132:661–680.
21 Spater D, Hansson EM, Zangi L et al. How to make a cardiomyocyte. Development 2014;141:4418–4431.
22 Zhang J, Klos M, Wilson GF et al. Extracellular matrix promotes highly efficient cardiac differentiation of human pluripotent stem cells: The matrix sandwich method. Circ Res 2012;111:1125–1136.
23 Birket MJ, Casini S, Kosmidis G et al. PGC-1alpha and reactive oxygen species regulate human embryonic stem cell-derived cardiomyocyte function. Stem Cell Rep 2013;1:560–574.
24 Laflamme MA, Chen KY, Naumova AV et al. Cardiomyocytes derived from human embryonic stem cells in prosurvival factors enhance function of infarcted rat hearts. Nat Biotechnol 2007;25:1015–1024.
25 Shalem O, Sanjana NE, Hartenian E et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 2014;343:84–87.
26 Yusa K, Rad R, Takeda J et al. Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. Nat Methods 2009;6:363–369.
27 Dobin A, Davis CA, Schlesinger F et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.
28 Liao Y, Smyth GK, Shi W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923–930.
29 Robinson MD, McCarthy DJ, Smyth GK. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–140.
30 Jones JB, Kern SE. Functional mapping of the MH1 DNA-binding domain of DPC4/SMAD4. Nucl Acids Res 2000;28:2363–2368.
31 Foo KS, Lehtinen ML, Leung CY et al. Human ISL1(+) ventricular progenitors self-assemble into an in vivo functional heart patch and preserve cardiac function post infarction. Mol Ther 2018;26:1644–1659.
32 Funa NS, Schachter KA, Lerdrup M et al. Beta-catenin regulates primitive streak induction through collaborative interactions with SMAD2/SMAD3 and OCT4. Cell Stem Cell 2015;16:639–652.
33 Bu L, Jiang X, Martin-Puig S et al. Human ISL1 heart progenitors generate diverse multipotent cardiovascular cell lineages. Nature 2009;460:113–117.
34 Meyer JS, Shearer RL, Capowski EE et al. Modeling early retinal development with human embryonic and induced pluripotent stem cells. Proc Natl Acad Sci USA 2009;106:16698–16703.
35 Lian X, Xu J, Bao X et al. Interrogating canonical Wnt signaling pathway in human pluripotent stem cell fate decisions using CRISPR-Cas9. Cell Mol Bioeng 2016;9:325–334.
36 Hahn SA, Hoque AT, Moskaluk CA et al. Homozygous deletion map at 18q21.1 in pancreatic cancer. Cancer Res 1996;56:490–494.
37 Schutte M, Hruban RH, Hedrick L et al. DPC4 gene in various tumor types. Cancer Res 1996;56:2527–2530.
38 Gallione CJ, Repetto GM, Legius E et al. A combined syndrome of juvenile polyposis and hereditary haemorrhagic telangiectasia associated with mutations in MADH4 (SMAD4). Lancet 2004;363:852–859.
39 Howe JR, Roth S, Ringold JC et al. Mutations in the SMAD4/DPC4 gene in juvenile polyposis. Science 1998;280:1086–1088.
40 Myhre SA, Ruvalcaba RH, Graham CB. A new growth deficiency syndrome. Clin Genet 1981;20:1–5.
41 Jin SC, Homsy J, Zaidi S et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. Nat Genet 2017;49:1593–1601.
42 Landrum MJ, Lee JM, Benson M et al. ClinVar: Public archive of interpretations of clinically relevant variants. Nucl Acids Res 2016;44:D862–D868.

See www.StemCells.com for supporting information available online.