Targeting *SOX17* in human embryonic stem cells creates unique strategies for isolating and analyzing developing endoderm

Pei Wang[1], **Ryan T. Rodriguez**[1], **Jing Wang**[1], **Amar Ghodasara**[1], and **Seung K. Kim**[1,2,3,4]

[1]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA

[2]Howard Hughes Medical Institute

[3]Department of Medicine (Oncology Division), Stanford University School of Medicine, Stanford, CA

Abstract

Investigating development of inaccessible human tissues like embryonic endoderm with embryonic stem cell (ESC) has been hindered by a lack of methods for marking and isolating endodermal cells, and tracing fates of their progeny toward differentiated lineages. Using homologous recombination in human ESC, we inserted an enhanced green fluorescent protein (*eGFP*) transgene into a locus encoding a postulated marker of human endoderm, *SOX17*, permitting purification of SOX17+ hESC progeny by fluorescence activated cell sorting (FACS). Microarray studies revealed a unique gene expression profile of human SOX17+ cells including endodermal marker enrichment, and unveiled specific cell surface protein combinations that permitted FACS-based isolation of primitive gut tube endodermal cells produced from unmodified human ESCs and from induced pluripotent stem cells (iPSC). FACS-isolated SOX17+ endodermal cells differentiated to progeny expressing markers of liver, pancreas, and intestinal epithelium, providing unprecedented evidence that human gastrointestinal lineages derive from SOX17+ cells. Thus, prospective isolation, lineage tracing, and developmental studies of hESCs described here have revealed fundamental aspects of human endodermal biology.

INTRODUCTION

Many internal organs are derived from definitive endoderm, one of the three principal germ layers in the developing embryo. However, much remains to be learned about the cellular and molecular basis for definitive endoderm development. While forming, definitive endoderm is incorporated by morphogenetic movements into a primitive tubular anlage ('primitive gut tube' stage) that is subsequently patterned to form the functional epithelial compartment of multiple internal organs including liver, intestines, lungs and pancreas. Prospective cell isolation followed by evaluation of developmental potential is a paradigmatic strategy for understanding cellular and molecular properties of multipotent cells like endoderm (McKnight et al., 2010). Recently, transgenic cell marking and flow

[4]Corresponding author Howard Hughes Medical Institute Department of Developmental Biology B300 Beckman Center 279 Campus Drive, Stanford CA, 94305-5329 seungkim@stanford.edu T: 650-723-6230 F: 650-725-7739.

Author Contributions P.W. and S.K.K conceived the project, and designed the experiments, P.W., R.T.R., J.W. and A.G. performed the experiments; P.W. and R.T.R. performed the data analysis; P.W., R.T.R. and S.K.K. wrote the manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

cytometry have been successfully used to isolate mouse native definitive endoderm and ES cell-derived endoderm (Morrison et al., 2008; Sherwood et al., 2007; Yasunaga et al., 2005). By contrast, isolation and assessment of hESC developing toward endodermal fates has been limited, in part by the absence of methods to purify endoderm-like cells and their progeny. Cells with endoderm-like properties have been produced from hESCs by attempting to recapitulate endogenous signaling programs (D'Amour et al., 2005; D'Amour et al., 2006) or through the screening of chemical libraries (Borowiak et al., 2009). However, these methods produce heterogeneous cell mixtures, limiting evaluation and studies of endoderm-like cells. For example, endoderm-like cells in these studies maintain expression of the pluripotency marker *NANOG*. Indeed, persistent ACTIVIN/NODAL signaling used to induce endoderm development in prior hESC studies may block progression toward an endodermal lineage (Vallier et al., 2009). Thus, the detailed molecular signature of hESC-derived endoderm has remained elusive.

Sox17 is a HMG box transcription factor required for definitive endoderm development in mice (Kanai-Azuma et al., 2002), and recent studies in mice show that Sox17[+] progeny comprise the functional epithelium of organs like lung, liver, pancreas and intestines (Kanai-Azuma et al., 2002; Shivdasani, 2002; Spence et al., 2009; Tam et al., 2003). Indelible cell "marking" using conditional genetics is one method for lineage tracing and has been used to show that pancreatic and intestinal epithelial cells derive from embryonic Sox17[+] endoderm in mouse (Engert et al., 2009). Another method for lineage tracing requires unambiguous cell purification followed by assessment of cell fates. Expression of SOX17 has been detected in hESC cultures treated with high doses Activin A, which is similar to mouse ESCs differentiating toward endodermal fates (D'Amour et al., 2005; Yasunaga et al., 2005), but purification of SOX17[+] cells and fate tracing has not been previously reported. Thus, unlike in studies of embryonic mice or mouse ESCs, it remains unknown if hESC-derived endodermal progeny resembling liver, pancreas or intestinal cells are produced from SOX17[+] cells. To address these fundamental questions, we used homologous recombination to target the *eGFP* reporter gene to the endogenous *SOX17* locus in hESC. This unique hESC line permitted FACS purification and assessment of the developmental potential of hESC-derived SOX17[+] cells. Microarray-based profiling of purified SOX17[+] cells revealed unique gene expression patterns, including new cell surface markers for isolating endoderm, derived from human ESC and iPSC. Additionally, FACS-isolated SOX17[+] primitive gut tube endodermal cells differentiated both *in vitro* and *in vivo* to progeny expressing markers of liver, pancreas, and intestinal epithelium, providing strong evidence that human gastrointestinal lineages derive from SOX17[+] cells. Thus, our studies have generated new platforms for revealing fundamental developmental and molecular properties of human endoderm and organs.

RESULTS

Creation of a *SOX17-eGFP* reporter hESC line by homologous recombination

To target the *eGFP* reporter trangene into the human *SOX17* locus, we built a DNA construct (Figure 1A) using homologous recombination-mediated genetic engineering ('recombineering': Copeland et al., 2001). The *eGFP* transgene was inserted directly after the start codon in exon 1 of *SOX17*. A *loxP-PGK-neoR-loxP* cassette was inserted as a positive selection marker, while the dipheria toxin (DT) gene and herpes simplex virus thymidine kinase (HSV-TK) gene on the construct 'arms' were used to select against random integration. Following electroporation and screening of 200 clones resistant to G418, one homologous recombination event was observed and correct targeting in this cell line (hS17-90) was confirmed by Southern blotting (Figure S1A). To minimize the interference by the neoR cassette on the expression of neighboring genes (Davis et al., 2008), we used Adeno-CRE-GFP virus to excise the loxP flanked PGK-neo cassette (Figure

S1B). After infection, we purified 5000 GFP[+] cells by FACS and, at clonal dilution, cultured these for 12 days to allow colony formation. Cells derived from 9 independent clones (hS17-d1-d5, d8, d9, d11 and d12) were treated with G418, and all died within 5 days. This result was consistent with PCR genotyping analysis revealing excision of the neoR cassette in these clones (Figure S1B). Undifferentiated hS17-d2 colonies had characteristic hESC morphology, and did not express eGFP (Figure 1B) or SOX17 (data not shown). To assess developmental regulation of the *SOX17-eGFP* marker, we used an established protocol (D'Amour et al., 2006; hereafter called 'Protocol D') to induce endoderm differentiation and *SOX17* expression in hS17 hESCs. We exposed hESCs to Activin A and Wnt3a, which resulted in the development of progeny resembling definitive endoderm ('stage 1′). Subsequent media changes, including Activin A removal in 'stage 2′ led to the development of cells resembling primitive gut tube-stage endoderm, then posterior foregut epithelium ('stage 3′; Figure 1B). With this protocol, the number of SOX17-eGFP[+] cells peaked at stage 1, then declined in stages 2 and 3 (Figure 1C). This temporal pattern of eGFP expression was consistent with a previous report that *SOX17* mRNA and protein expression peaked at stage 1, and was reduced by stage 3 (D'Amour et al., 2006). Immunostaining of clones hS17-d1 to d5 showed that 100% eGFP[+] cells were SOX17[+], $90 \pm 5\%$ (n=5) of SOX17[+] cells were eGFP[+] at both stage 1 and stage 2 (Figure 1D). Collectively, these results showed that the *SOX17-eGFP* marker can be used to monitor *SOX17* expression and to identify SOX17[+] hESC progeny in hS17-d1 to d5 cells.

SOX17-eGFP marks the endodermal progeny of differentiating hESCs

To investigate if *SOX17-eGFP* expression in hS17-d2 could be used to monitor endoderm differentiation and to purify endodermal progeny, we performed flow cytometry-based studies of hS17-d2 during endodermal differentiation using Protocol D. At stage 1, $73 \pm 4.5\%$ of cells were eGFP[+] (Figure 2A; n=5). $39 \pm 14\%$ (n=5) of cells were eGFP[+] at stage 2, and only $8 \pm 1.5\%$ (n=4) of cells were eGFP[+] by stage 3 (Figure 2A). To assess gene expression, we first measured mRNA levels of marker genes in FACS-purified SOX17-eGFP[+] and SOX17-eGFP[−] cells by Quantitative Reverse Transcriptase-Polymerase Chain Reaction (Q-PCR) at stages 1-2. Expression of the pluripotency marker *OCT4*, mesoderm marker *BRACHYURY* (*BRA*), and ectoderm marker *ZIC1* was relatively high in eGFP[−] cells, but low or undetectable in eGFP[+] cells (Figure 2B). Likewise, levels of Tra-1-60 and SSEA4, markers of undifferentiated hESCs, were higher in eGFP[−] cells than in eGFP[+] cells at these stages (Figure S2). Expression of the pluripotency marker *NANOG* in eGFP[+] cells was comparable in stage 1 to levels in undifferentiated hESCs and did not decline until stage 2, following Activin withdrawal (Figure 2B), consistent with findings by others (Vallier et al., 2009). Expression of mesendoderm markers *GSC* and *MIXL1*, and endoderm markers *SOX17, FOXA1* and *FOXA2* was highly enriched in eGFP[+] cells by stage 1, and increased 200 to 8000-fold compared to hESCs (Figure 2B). The expression of *GSC, MIXL1* and *SOX17* was reduced at stage 2, while the expression of primitive gut endoderm markers *HNF1β* and *HNF4α* was highly induced at stage 2. Immunostaining confirmed that stage 2 cells expressing HNF1β or HNF4α protein also express eGFP protein (Figure 2C), consistent with the conclusion that hSOX17-eGFP[+] cells at this stage resemble primitive gut endoderm. By contrast, AFP expression was not detected, and SOX7 mRNA levels were at the limit of detection by RT-PCR in undifferentiated hESCs and in eGFP[+] cells at stage 1-2 (Figure 2B), indicating hS17-eGFP[+] cells were not visceral endoderm. The cell surface markers EpCAM and E-cadherin found expressed in mouse ESC-derived endoderm (Sherwood et al., 2007; Yasunaga et al., 2005) and we analyzed expression of these markers in the hS17 reporter line by flow cytometry. As expected, EpCAM and E-Cadherin were expressed in hS17-eGFP[+] endoderm cells. However, both markers were also expressed in undifferentiated hESCs, precluding their use in isolation of hESC derived endoderm (Figure S2). We tested the expression of mesoderm markers PDGFRα and TIE2 using flow

cytometry and found that these markers were not expressed in SOX17-eGFP$^+$ (Figure S2). Thus, SOX17-eGFP expression in differentiating hS17-d2 cells permitted FACS purification of SOX17$^+$ cells resembling early definitive endoderm at stage 1, and primitive gut tube endoderm at stage 2.

Cell surface markers dynamics during endoderm differentiation

Enrichment of endoderm markers in FACS-isolated SOX17-eGFP$^+$ cells suggested that gene expression studies might identify a cell surface marker 'signature' for human endodermal cells. We compared gene expression patterns of undifferentiated hS17 ESCs, and their FACS-sorted eGFP$^+$ progeny at stage 1 and stage 2 using microarray analysis (see Methods). A total of 13,209 genes showed statistically-significant differential expression among the three samples. To simplify analysis of gene expression changes, we clustered the 13,209 genes into 6 groups using K-Means clustering (K=6; Figure 3A; see Methods) and present analysis of clusters 1-3 here. The expression of genes in 'cluster 1´, including hESC pluripotency markers *OCT4, SOX2, SALL2, THAP11* (*RONIN*), *TERT, TERF1, PIPOX, TOX3*, and *FGFR1*, was reduced in SOX17-eGFP$^+$ cells at stages 1 and 2 (Figure 3A). The expression of cluster 2 genes, including the endoderm markers *SOX17, FOXA1, FOXA2, GATA3, GATA4, GATA6, SHH* and *HNF1β*, was increased in stage 1 and 2 eGFP$^+$ cells; expression of cluster 3 genes, which included primitive gut tube-stage endoderm markers *VTN, KITLG, TRH, HES1, ANXA4*, and *HNF4α* (Hou et al., 2007; Sherwood et al., 2007) was increased in SOX17-eGFP$^+$ cells only at stage 2 (Figure 3A). To identify new cell surface markers for endoderm, we focused on transmembrane proteins whose expression changed in SOX17-eGFP$^+$ cells at stage 1 and 2. The Gene Ontology (GO) term, "integral to plasma membrane" GO:0005887, identified 608 genes encoding adhesion molecules, receptors, transporters and ion channels which were differentially expressed between undifferentiated hESCs and SOX17-eGFP$^+$ endodermal cells (Figure 3B). Hierarchical clustering of these 608 genes revealed dynamic expression during endodermal differentiation of hS17 ESCs (Figure 3B). To examine the function of the 608 genes (see Table S1 for a complete list), we analyzed enriched GO-terms and found that the majority of the 25 most highly-enriched GO-terms were transporters and adhesion molecules (Figure 3B), suggesting hESCs changed their surface properties during endoderm differentiation.

To validate findings from our microarray studies and create methods to purify endodermal progeny from hESC cultures, we used flow cytometry to investigate integral membrane proteins whose expression increased in SOX17-eGFP$^+$ cells. In confirmation of a prior study (D'Amour et al., 2006), we found that the chemokine receptor CXCR4 was expressed in SOX17-eGFP$^+$ cells at stage 1 but reduced at stage 2 (Figure S3A); hence, additional methods were needed for isolation of differentiating stage 2 endodermal progeny. Thus, we obtained 14 commercially-available antibodies for cluster of differentiation (CD) surface molecules which were induced at stage 1 and 2 in our microarray analysis (Figure S3B). We analyzed the expression of these proteins by flow cytometry and identified antibodies to three molecules induced in stage 2 SOX17-eGFP$^+$ cells: CD49e (also known as integrin α5) (Argraves et al., 1986; Friedland et al., 2009), CD141 (thrombomodulin) (Isermann et al., 2003; Wen et al., 1987) and CD238 (Kell blood group antigen) (Lee et al., 1991; Russo et al., 2000). Unlike CXCR4 mRNA levels, mRNA encoding CD49e increased during both stage 1 and 2, and CD141 and CD238 mRNAs increased specifically at stage 2 in SOX17-eGFP$^+$ cells (Figure 4A). Consistent with these data, flow cytometry analysis showed that CD49e expression was upregulated at stage 1 (compared to undifferentiated hS17 hESCs) and by stage 2 only SOX17-eGFP$^+$ cells were CD49e$^{+/high}$ (Figure 4B). Stage 1 hS17 eGFP$^+$ cells had low or no expression of CD141 and CD238: by stage 2, however, eGFP$^+$ cells were CD141$^+$ or CD238$^+$ (Figure 4A, 4B). Moreover, like SOX17-eGFP$^+$ cells, the number of CD49e$^+$, CD141$^+$ or CD238$^+$ cells sharply declined by stage 3 (Figure 4B).

Consistent with these findings, immunostaining showed that SOX17[+] cells from stage 2 coexpressed CD49e and CD141 (Figure 4C). Our data suggested that SOX17-eGFP[+] cells expressed all three surface markers. Thus, our microarray-based analysis identified cell surface markers for isolating endodermal progeny from hESC cultures.

Surface markers for isolation of endodermal progeny of genetically unmodified hESCs

We investigated if our FACS methods could be used in WA09 cells, the genetically unmodified line used to generate hS17. Following treatment of WA09 cells through stage 2 culture conditions, we analyzed the expression of three novel surface markers by flow cytometry. We showed that CD49e[+] cells are 'positive' for both CD141 and CD238, and that CD49e[low] cells are 'negative' for both CD141 and CD238 (Figure 4D). After sorting stage 2 WA09 cells with these surface markers we analyzed gene expression by Q-PCR. Expression of endoderm markers including *SOX17, FOXA1, FOXA2, HNF1β, HNF4α, KITL, SHH* and *HB9* was highly enriched in CD49e[+] CD141[+] CD238[+] cells compared to CD49e[−/low] CD141[−] CD238[−] cells (Figure 4E). By contrast, mRNAs encoding markers of pluripotency, mesoderm and visceral endoderm, including *OCT4, NANOG, MEOX1*, and *SOX7*, were expressed at higher levels in CD49e[−/low] CD141 [−] CD238[−] cells compared with CD49e[+] CD141[+] CD238[+] cells (Figure 4E). We obtained similar gene expression patterns using two additional hESC lines, WA01 and LSJ-1 (data not shown, Chiao et al., 2008; Thomson et al., 1998).

To assess the relevance of our findings to human iPSC, we exposed the HiPSC-5 line (Byers et al, submitted; see Methods) to conditions that induce endoderm development by hESC, and measured gene expression by Q-PCR. The expression of marker genes in HiPSC-5 progeny was similar to those in hESC progeny at stage 1 and 2 (D'Amour et al, 2006; Figure S4A and S4B). We next used flow cytometry to analyze cell surface markers during endodermal differentiation of HiPSC-5 cells. CXCR4 expression in HiPSC-5 progeny peaked at stage 1 (Figure S4C), in agreement with gene expression data (Figure S4A). CD49e, CD141 and CD238 were expressed in HiPSC-5 progeny at stage 2, similar to our findings with hESCs. We then sorted stage 2 iPSC with antibodies to CD49e, CD141 and CD238 by FACS, and analyzed gene expression by Q-PCR. Expression of endoderm markers was highly enriched in CD49e[+] CD141[+] CD238[+] cells compared to CD49e[−] CD141[−] CD238[−] cells. By contrast, mRNAs encoding markers of pluripotency, mesoderm and visceral endoderm were expressed at higher levels in CD49e[−] CD141[−] CD238[−] cells compared to CD49e[+] CD141[+] CD238[+] cells (Figure S4D). Thus, our studies of hS17 ES cells have created new tools for analyzing endodermal development in iPSC and unique FACS methods for isolating iPSC progeny resembling primitive gut endoderm.

In organs of the gastrointestinal and respiratory tracts, the epithelial component derives from definitive endoderm. To test the relevance of our findings to native definitive endoderm progeny in humans, we used FACS to assess expression of CD49e in fetal human pancreas from 10-12 weeks gestation. CD49e-based sorting fractionated cells into two populations, CD49e[+] and CD49[−] (Figure S4E); in our study we also used CD31-based sorting to exclude endothelial cells (Albelda et al., 1991). We then assessed gene expression in fractionated CD49e[+] and CD49e[−] cells. Genes expressed in pancreatic epithelium (Lyttle et al., 2008; Sarkar et al., 2008), including *SOX9, PDX1, NGN3, CHROMOGRANIN A, INSULIN* and *AMYLASE*, were enriched in CD49e[+] CD31[−] cells. By contrast, mRNAs enriched in mesenchyme (Hoffman et al., 2008; Zhou et al., 2007), like *VIMENTIN, OSR2* and *AKAP12*, were highly enriched in pancreatic CD49e[−] CD31[−] cells (Figure S4F). Thus, studies of endoderm differentiation in hS17 ESCs have identified unique methods for FACS separation of native human endoderm progeny.

Developmental potential of primitive gut endoderm derived from hESCs

While gene expression analysis suggested that CD49e$^+$ CD141$^+$ CD238$^+$ cells are endodermal progeny, it was unclear if these cells retain latent potential to differentiate to ectoderm or mesoderm. Thus, we used established methods for promoting hESC differentiation toward neuroectodermal or mesodermal fates to assess the potential of CD49e$^+$CD141$^+$CD238$^+$ and CD49e$^-$CD141$^-$CD238$^-$ stage 2 cells (Figure 5A). For neuroectoderm differentiation, we cultured hS17 ESC with Noggin and SB431542 to inhibit SMAD (Chambers et al., 2009), followed by exposure to SHH (Figure 5A), and assessed gene expression at 5 and 9 days. As expected, neuroectoderm differentiation conditions induced neuroectoderm specific markers *ZIC1, SOX1* and *PAX6* from hS17 hESCs (Figure 5B, 'NE'). Expression of native *SOX17* or *SOX17*-eGFP was not observed (Figure S5A). There was little expression of *ZIC1, SOX1* and *PAX6* in undifferentiated hS17 cells (Figure 5B, 'hESC') or in unsorted stage 2 cells derived from Protocol D (Figure 5B, 'UN'). Expression of these markers was high in CD49e$^-$CD141$^-$CD238$^-$ cells exposed to neuroectodermal differentiation conditions (Figure 5B, 'Neg'), and similar to levels found in neuroectoderm derived from hESCs. By contrast, we did not detect the expression of *ZIC1, SOX1* and *PAX6* in CD49e$^+$ CD141$^+$ CD238$^+$ cells exposed to neuroectodermal differentiation conditions (Figure 5B, 'Pos'). Thus, under these conditions, CD49e$^+$ CD141$^+$ CD238$^+$ cells had diminished neuroectodermal potential.

Based on a similar experimental logic, we next tested if CD49e$^+$ CD141$^+$ CD238$^+$ cells retain the potential to form mesodermal progeny. We used a method for inducing mesoderm formation (Nostro et al., 2008; Yang et al., 2008) based on hESC exposure to a sequence of BMP4, bFGF and Wnt3a for 5 days followed by vascular endothelial growth factor (VEGF) and Dickkopf homolog 1 (DKK1) for another 4 days (Figure 5A). Consistent with these reports, undifferentiated hS17 cells did not express mesodermal markers like *MIXL* or *BRA*, but expressed these markers after exposure to mesoderm promoting factors (Figure 5C, 'hESC' versus 'ME'). Expression of *SOX17*-eGFP was observed in few cells (Figure S5A). Expression of *MXL1* and *BRA* was high in CD49e$^-$CD141$^-$CD238$^-$ cells exposed to mesoderm differentiation conditions, and similar to levels found during mesoderm derived from hESCs (Figure 5C 'Neg'). By contrast, we did not detect the expression of *MIXL1* or *BRA* in CD49e$^+$ CD141$^+$ CD238$^+$ cells exposed to mesodermal differentiation conditions. The expression of Tra-1-60, EpCAM and E-CADHERIN was reduced in cells produced with methods enhancing neuroectoderm or mesoderm differentiation (Figure S5C). However, the endoderm markers CXCR4, CD49e, CD141 and CD238 were not induced in 9 days culture (Figure S5C). Collectively, these studies suggested that both mesodermal and neuroectodermal potential were reduced in isolated CD49e$^+$ CD141$^+$ CD238$^+$ cells.

Assessment of endodermal lineage potential of purified hESC progeny

We hypothesized that CD49e$^+$ CD141$^+$ CD238$^+$ cells resembling primitive gut endoderm and purified at stage 2 in protocol D might develop toward fates resembling differentiated epithelial cells found in the intestines, pancreas and liver. To test this hypothesis, we used methods for generating cells resembling gastrointestinal cells, pancreatic progenitors and hepatocytes (D'Amour et al 2006, 'protocol D'; Mfopou et al. 2010, hereafter 'protocol M'; Figure 6A). For both methods, we compared gene expression at late stages of differentiation by Q-PCR ('stages 3 and 4′') and immunocytochemistry ('stages 4′') in sorted CD49e$^+$ CD141$^+$ CD238$^+$ cells and CD49e$^-$ CD141$^-$ CD238$^-$ cells, and unsorted control cells (Figures 6B, 6C). Expression of *HNF1β* and *HNF4α*, markers of both pancreas and liver development, was enriched in CD49e$^+$ CD141$^+$ CD238$^+$ progeny produced with both protocols (Figure 6B). Likewise expression of the pancreatic markers *PDX1*, the hepatic markers *AFP, ALBUMIN* and *CEBPα*, and intestinal marker *CDX2* was also enriched in CD49e$^+$ CD141$^+$ CD238$^+$ progeny (Figure 6B). These findings are consistent with studies

revealing significant hepatic differentiation resulting from Protocol D (Mfopou et al 2010). By contrast, the mesodermal markers *KDR* (Yamaguchi et al., 1993) and *MEOX1*, and the visceral endoderm marker *SOX7* were enriched in CD49e⁻ CD141⁻ CD238⁻ progeny (Figure 6B). Immunostaining of stage 4 cells (Figures 6C, Figures S6A, S6B and S6C) confirmed enriched development of progeny expressing PDX1, ALBUMIN, AFP, CDX2, HNF1β and HNF4α from CD49e⁺ CD141⁺ CD238⁺ cells. By comparison, expression of these markers by progeny of CD49e⁻ CD141⁻ CD238⁻ cells or unsorted cells was reduced or undetectable (Figures 6C, Figures S6A, S6B and S6C). Thus, CD49e⁺ CD141⁺ CD238⁺ cells generated progeny in vitro that resembled differentiated endodermal progeny.

To determine if SOX17⁺ cells also give rise to differentiated endodermal progeny *in vivo*, we transplanted CD49e⁺ CD141⁺ CD238⁺ cells (see Methods) and compared outcomes to those after transplantation of CD49e⁻ CD141⁻ CD238⁻ cells or unsorted cells from stage 2. Transplanted CD49e⁺ CD141⁺ CD238⁺ cells failed to survive in a renal capsule transplantation site (0/6 grafts; Figure S7A and data not shown). By contrast transplantation of CD49e⁻ CD141⁻ CD238⁻ cells led to formation of tumors with features of teratomas in 5/6 grafts (data not shown). Likewise transplantation of unsorted cells from stage 2 produced teratomas. Since adjacent mesoderm supports embryonic endoderm development, we reasoned that signals from co-transplanted mesoderm might support survival and development of purified CD49e⁺ CD141⁺ CD238⁺ cells. Consistent with this view, we observed enhanced survival and growth of CD49e⁺ CD141⁺ CD238⁺ cells transplanted with mouse embryonic fibroblasts (MEFs; Figure S7A). Hematoxylin-eosin staining and immunohistology with antibodies that identify human nuclear antigen (HuNu) revealed the development of tissues with characteristic features of epithelial cells in these grafts after 3 weeks (Figure 7A). HuNu antibody did not stain adjacent mouse cell nuclei (Figure 7). Immunostaining revealed that CD49e⁺ CD141⁺ CD238⁺ progeny formed structures comprised of laminar epithelium surrounding a central lumen, with component epithelium expressing multiple markers of differentiated gastrointestinal organs including EpCAM, E-cadherin, PDX1, SOX9, CHROMOGRANIN A (ChrA), CDX2, ALBUMIN and FOXA2 (Figure 7B-G, Figure S7B). We did not detect expression of the neuronal markers TUJ1 and SOX2, and the mesodermal marker VIMENTIN in CD49e⁺ CD141⁺ CD238⁺ progeny (Figure 7D, 7G, Figure S7B). By contrast, similar analysis revealed that CD49e⁻ CD141⁻ CD238⁻ cells co-transplanted with MEFs produced tumors resembling teratomas (Figure 7H). Engrafted progeny of unsorted or CD49e⁻ CD141⁻ CD238⁻ cells co-transplanted with MEFs contained human cells expressing TUJ1 and SOX2, VIMENTIN, and endodermal markers like E-cadherin, HNF4a, CDX2 and FOXA2 (Figure 7I-N, Figure S7C). By contrast, we did not detect PDX1 or CHROMOGRANIN A in these teratomas (Figure 7I, 7J). Thus, CD49e⁺ CD141⁺ CD238⁺ cells generated progeny in vivo that resembled differentiated endodermal progeny.

DISCUSSION

The epithelial compartment of many vital organs such as the liver, lungs, pancreas and intestines is derived from endoderm. Prior studies have reported methods for hESC differentiation into cell mixtures containing a subset of progeny resembling immature hepatic or pancreatic cell subsets (D'Amour et al., 2006; Duan et al., 2007; Kroon et al., 2008). While a transient endoderm-like stage of differentiation was reported in these studies, genetic or cell purification-based lineage tracing methods including (1) indelible cell marking and lineage analysis, (2) cell ablation to eliminate progeny, or (3) prospective cell purification with analysis of progeny was not used. Thus, the lineage of endoderm-like cells derived from these prior studies was unclear. Here, we used prospective cell purification followed by in vitro and in vivo assessment of cell potential to demonstrate unambiguously, for the first time to our knowledge, that human SOX17+ progeny derived from ES cells

represent definitive endoderm. Our cell marking and purification approach represents a practical general solution for demonstrating lineage relationships in cultured human ES cell progeny.

To monitor endoderm development in hESC lines, we used homologous recombination to insert the *eGFP* reporter transgene into the *SOX17* locus. Homologous gene targeting in hESCs has been described in only several prior studies (reviewed by Zeng and Rao, 2008). However, we reasoned that this strategy would maximize preservation of crucial *cis*- and *trans*-regulatory mechanisms controlling *SOX17* expression in endoderm, thereby offsetting the inherent difficulties of homologous gene targeting in hESCs. *SOX17*-eGFP expression was well matched with expression of endogenous *SOX17* mRNA and protein in our hS17 cell line. Monitoring endodermal differentiation using the hS17 reporter hESC line allowed unique opportunities for isolation and analysis of developing human endodermal cells, and should enhance efforts to screen endogenous factors or compound libraries for activities that promote endodermal development (Borowiak et al., 2009). *SOX17* also has established roles in developing mouse oligodendrocytes and hematopoietic stem cells (Kim et al., 2007; Sohn et al., 2006). Thus, we speculate that the hS17 lines described here may also be useful for studies of human neuronal and blood lineages.

While earlier studies had described successful differentiation of endodermal cells in cultured mouse and human ESCs, genome-scale gene expression analysis of FACS-isolated endoderm from hESCs has not been previously reported, to our knowledge. With microarrays we interrogated the expression profile of hS17-derived endodermal progeny. From this unique data set we identified a panel of novel surface antigens that permitted the isolation of SOX17$^+$ endodermal cells. This new surface antigen profile also permitted FACS isolation of progeny resembling primitive gut tube-stage endoderm derived from multiple genetically unmodified hESCs and from human iPSCs. Thus, our work should facilitate studies of endodermal development and disease modeling using hESCs and iPSCs from diverse genetic backgrounds. Identification of a surface antigen profile for human endoderm should be useful for monitoring the quality of endodermal cells and endodermal progeny derived from multiple pluripotent cell sources.

The heterogeneity of differentiating hESC cultures is well described (Carpenter et al., 2009) and eradicating tumor-forming potential in these cultures has been challenging (Matveyenko et al., 2010; Schriebl et al., 2009). Therefore, another practical use of FACS-based isolation methods is the reduction or elimination of hazards inherent to complex hESC cultures, like tumor formation by engrafted cells. FACS-isolated CD49e$^-$ CD141$^-$ CD238$^-$ cells contained tumorigenic cells in xenotransplant assays, and *in vitro* differentiation of these cells revealed enriched potential for neuroectoderm and mesodermal development. By contrast, assays of FACS-isolated CD49e$^+$ CD141$^+$ CD238$^+$ cells in culture-based or xenograft studies revealed their capacity to form endodermal progeny but not neuroectoderm and mesoderm. Previous studies have not resolved cellular heterogeneity at the later "stages" of *in vitro* endodermal differentiation. To our knowledge, direct isolation of enriched endodermal cell subsets followed by continuation of development in vitro or in vivo has not been previously reported. Our methods could facilitate identification of lineage committed progenitor cells – a transient population of cells derived from endoderm whose mature progeny later comprise the physiologically-functional epithelium of gastrointestinal and respiratory organs. Although the detailed methods to generate different mature lineages from human endoderm requires further studies, our work establishes a new starting point for understanding endodermal organ development and reconstituting organogenesis *in vitro* or with transplant-based *in vivo* strategies.

METHODS

Human ES cell culture and targeting

The human embryonic stem cell lines H9, H1 (WiCell) and LSJ-1 were grown on irradiated CD1 mouse embryonic feeder cells in DMEM/F12 supplemented with 20% (vol/vol) knockout serum replacement, βFGF (8 ng/ml, Peprotech), 3 mM L-glutamine, 0.1 mM nonessential amino acids and 0.1 mM β-mercaptoethanol (Invitrogen) (hESC media). The gene-targeting vector was constructed by replacing the coding region of the *SOX17* gene with a cassette containing Enhanced Green Fluorescent Protein (*eGFP*)+SV40 polyA and a *neo* gene under the control of the *tk* promoter. This cassette was flanked on both the 5′ and 3′ directions by a 12 kb homologous arm. Non-isogenic homologous DNA was obtained from BAC clone RP11-359F18 (http://bacpac.chori.org/) using recombineering (recombination-mediated genetic engineering) method. Two negative selection cassettes were cloned onto each side of the homologous arms (Figure 1A). The targeting construct was electroporated into H9 human ES cells and drug selection was started two days later. After 12 days, surviving colonies were manually dissected, expanded and analyzed individually by PCR and Southern blot. Karyotype analysis was performed by CellLine Genetics. To delete the PGK-neo cassette, we infected hS17-90 hESC with Ad-cre-GFP (Vector Biolabs). After 24 hours of infection, we FACS sorted 5000 GFP positive cells and plated for single colonies. After picking and expanding the 12 single colonies, one set of cells were treated with 50ug/ml G418, and 9 clones (hS17-d1-d5, d8, d9, d11 and d12) all died within 5 days. We identified correct excision events by PCR using primers outside of two loxP sites.

Human ESC and iPSC endoderm differentiation

To differentiate hESC toward an endodermlike fate, we used two protocols. In protocol D (D'Amour et al, 2006), at stage 1, 90% confluent hES cells were cultured in RPMI (Invitrogen) with 25ng/ml Wnt3a and 100ng/ml Activin A for one day. The medium was then changed to RPMI with 0.2% FBS and 100ng/ml Activin A for two days. To further differentiate to primitive gut endoderm (stage 2), the medium was changed to RPMI with 2% FBS, 50 ng/ml human FGF10 and 0.25μM KAAD-cyclopamine (Toronto Research Chemicals) for 3 days. For differentiation to stage 3, the medium was changed to DMEM with 1× B27 (Invitrogen), 50 ng/ml human FGF10 and 0.25μM KAAD-cyclopamine (Toronto Research Chemicals), 2μM retinoic acid (Sigma) for 3 days; to stage 4, the medium was changed to DMEM with 1× B27 (Invitrogen) for 3 days; to stage 5, the medium was changed to CMRL with 1× B27 (Invitrogen) for 3 days. In protocol M (Mfopou et al., 2010), starting at stage 3, the media was changed to DMEM with 1× B27, 50 ng/ml human Noggin (PeproTech) and 0.25μM KAAD-cyclopamine, 2μM retinoic acid for 3 days; to stage 4, the medium was changed to DMEM with 1× B27 (Invitrogen), 50ng/ml human FGF10 for 3 days; to stage 5, the medium was changed to CMRL with 1× B27 (Invitrogen) for 3 days.

Human ESC and iPSC neuroectoderm and mesoderm differentiation

To differentiate to neuroectoderm lineage, we used the first two stages of Chambers et al.'s protocol, The initial differentiation SN media contained hESC media minus bFGF supplement with 20% knockout serum replacement, 10 mM TGF-b inhibitor SB431542 (Tocris) and 100 ng/ml Noggin for 5 days. Then we changed to SHH media: hESC media minus bFGF, adding 1× N2 suppliment (Invitrogen), 100 ng/ml Noggin and 200ng/ml SHH for 4 days. To differentiate to mesoderm lineage, we adapted the first two stages of Yang et al protocol. In brief, 90 % confluent hESC were cultured in StemPro34 media (Invitrogen), containing 2mM glutamine, 4×10^{-4}M monothioglycerol (MTG), 50μg/ml ascorbic acid (Sigma), with growth factors, 3ng/ml Activin A, 10.5ng/ml BMP4, and 5ng/ml βFGF for 5

days. We changed growth factor to 0.5ng/ml BMP4, 150ng/ml DKK1, and 10ng/ml VEGF for another 4 days. All factors were purchased from R&D Systems unless indicated. Cultures were maintained in a 5% CO_2/5% O_2/90% N_2 environment.

Fluorescence-activated cell sorting and cell replating

Cells were dissociated using 0.05% trypsin/EDTA (Invitrogen) for 5 min at 37°C and triturated. Trypsin was neutralized with FACS buffer (PBS, 10 mM EGTA, and 2% FBS). Cells were pelleted and resuspended in FACS buffer then labeled with antibodies (Table S2).. Aqua fluorescent reactive dye (Invitrogen) was used to identify and remove dead cells. Cells were analyzed and sorted using a FACSAria (BD Biosciences). FACS data were analyzed using the FlowJo software from Tree Star. For replating experiments, stage 2 primitive gut endoderm cells were dissociated to single cells with trypsin and neutralized with FACS buffer (Hanks Balanced salt solution (HBSS), 2% FBS and 10nM Rock inhibitor Y27632 (Calbiochem)), staining with CD49e, CD141 and CD238. The cells were sorted into 0.5 ml FBS with 4ul of 10mM Y27632. Sorted cells were replated in stage 2 primitive gut endoderm media with 10 nM Y27632 for overnight and changed to next differentiation media.

Immunofluorescence

Cultures were fixed for 15 min at RT in 2% (wt/vol) paraformaldehyde in PBS. Cultures and tissue sections were washed several times in PBS and blocked with 5% (vol/vol) normal donkey serum (Jackson ImmunoResearch Laboratories) for 30 min in PBST (PBS/0.1% (wt/vol) Triton X-100). Primary antibodies were incubated for 16 h at 4 °C and then washed three times with PBST at RT. Secondary antibodies were incubated at RT for 1 h. Antibodies used are listed in Table S2. Photomicrographs were obtained with a Axio Imager M1 microscope (Zeiss), and a AxioCam MRC5 and MRM cameras with Axiovision 4.7 software (Zeiss).

Real-time quantitative PCR

Total RNA was isolated from FACS sorted cells using the PicoPure RNA isolation kit (Arcturus) and was used to perform reverse transcription with a cDNA synthesis kit from Ambion. PCR reactions were run in duplicate with 1/40th of the cDNA per reaction using Taqman gene expression assays (Applied Biosystems) in which β-actin was used as an endogenous control. The analyses were done on three indepent experiments. The assay list was shown in Table S2.

Microarray analysis

Three replicates for each sample were collected by FACS sorting for RNA isolation using the Picopure RNA extract kit (Arcturus). cDNA was synthesized and amplified using the Ovation RNA amplification system V2 (NuGEN). 3.75 ug of cDNA was fragmented and biotin labeled using the FL-Ovation cDNA biotin module V2 (NuGEN). Labeled cDNAs were hybridized to Affymetrix Mouse Human Genome U133 Plus 2.0 arrays. Microarray data was normalized and summarized by the Robust Multi-chip Averaging (RMA) algorithm using GeneSpring GX 11 (Agilent). Baseline transformation of the data was carried out by Baseline to median of samples and one-way analysis of variance (ANOVA) test was used (Benjamini Hochberg, False Discovery Rate<0.05). We averaged the signal intensities of probe sets mapping to the same gene. Microarray data has been deposited in the GEO database (GSE26862)

Kidney capsule transplantation

1×10^6 FACS sorted cells of hESC–derived endoderm were cultured on stage 2 media overnight and scraped from plate, then mixed with Matrigel (BD) 1:1 in volume. 5×10^5 mouse embryonic fibroblasts were added as indicated. 5 μl of cell mixture were implanted into the left kidney capsule of ketamine/xylazine anesthetized severe combined immunodeficiency (SCID) mice. Transplanted kidneys were harvested after 3-6 weeks, then fixed in 4% PFA for 16 hours. The transplants were cut into two; half was imbedded in OCT for frozen section and the other half in paraffin. Following microtomy, tissue sections were stained with hematoxylin and eosin or incubated with antibodies for immunofluorescence (Table S2).

Human pancreatic tissue acquisition and cell dispersion

Human fetal pancreata (n=3) were obtained from the Birth Defects Research Laboratory, Univeristy of Washington (Seattle, WA). Informed consent was obtained for tissue donation as well as approval from institutional review boards. Staged fetal pancreata between 10 and 12 weeks were processed within 24 hours of receipt, minced, washed in PBS supplemented with 0.02% EDTA and 0.1% bovine serum albumin, and then exposed to 1.5 mg/ml Liberase H1 or TL enzyme mix (Roche, Indianapolis, IN) for 10 minutes in a shaking water bath at 37°C. The enzyme solution was neutralized with 1% HEPES buffer in HBSS. Cells were then subjected to antibody staining and analysis as above.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Albelda SM, Muller WA, Buck CA, Newman PJ. Molecular and cellular properties of PECAM-1 (endoCAM/CD31): a novel vascular cell-cell adhesion molecule. J Cell Biol. 1991; 114:1059–1068. [PubMed: 1874786]

Argraves WS, Pytela R, Suzuki S, Millan JL, Pierschbacher MD, Ruoslahti E. cDNA sequences from the alpha subunit of the fibronectin receptor predict a transmembrane domain and a short cytoplasmic peptide. J Biol Chem. 1986; 261:12922–12924. [PubMed: 2944883]

Borowiak M, Maehr R, Chen S, Chen AE, Tang W, Fox JL, Schreiber SL, Melton DA. Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. Cell Stem Cell. 2009; 4:348–358. [PubMed: 19341624]

Carpenter MK, Frey-Vasconcells J, Rao MS. Developing safe therapies from human pluripotent stem cells. Nat Biotechnol. 2009; 27:606–613. [PubMed: 19587662]

Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol. 2009; 27:275–280. [PubMed: 19252484]

Chiao E, Kmet M, Behr B, Baker J. Derivation of human embryonic stem cells in standard and chemically defined conditions. Methods Cell Biol. 2008; 86:1–14. [PubMed: 18442641]

Copeland NG, Jenkins NA, Court DL. Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet. 2001; 2:769–779. [PubMed: 11584293]

D'Amour KA, Agulnick AD, Eliazer S, Kelly OG, Kroon E, Baetge EE. Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat Biotechnol. 2005; 23:1534–1541. [PubMed: 16258519]

D'Amour KA, Bang AG, Eliazer S, Kelly OG, Agulnick AD, Smart NG, Moorman MA, Kroon E, Carpenter MK, Baetge EE. Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. Nat Biotechnol. 2006; 24:1392–1401. [PubMed: 17053790]

Davis RP, Costa M, Grandela C, Holland AM, Hatzistavrou T, Micallef SJ, Li X, Goulburn AL, Azzola L, Elefanty AG, et al. A protocol for removal of antibiotic resistance cassettes from human embryonic stem cells genetically modified by homologous recombination or transgenesis. Nat Protoc. 2008; 3:1550–1558. [PubMed: 18802436]

Duan Y, Catana A, Meng Y, Yamamoto N, He S, Gupta S, Gambhir SS, Zern MA. Differentiation and enrichment of hepatocyte-like cells from human embryonic stem cells in vitro and in vivo. Stem Cells. 2007; 25:3058–3068. [PubMed: 17885076]

Engert S, Liao WP, Burtscher I, Lickert H. Sox17-2A-iCre: a knock-in mouse line expressing Cre recombinase in endoderm and vascular endothelial cells. Genesis. 2009; 47:603–610. [PubMed: 19548312]

Friedland JC, Lee MH, Boettiger D. Mechanically activated integrin switch controls alpha5beta1 function. Science. 2009; 323:642–644. [PubMed: 19179533]

Hoffman BG, Zavaglia B, Witzsche J, de Algara T. Ruiz, Beach M, Hoodless PA, Jones SJ, Marra MA, Helgason CD. Identification of transcripts with enriched expression in the developing and adult pancreas. Genome Biol. 2008; 9:R99. [PubMed: 18554416]

Hou J, Charters AM, Lee SC, Zhao Y, Wu MK, Jones SJ, Marra MA, Hoodless PA. A systematic screen for genes expressed in definitive endoderm by Serial Analysis of Gene Expression (SAGE). BMC Dev Biol. 2007; 7:92. [PubMed: 17683524]

Isermann B, Sood R, Pawlinski R, Zogg M, Kalloway S, Degen JL, Mackman N, Weiler H. The thrombomodulin-protein C system is essential for the maintenance of pregnancy. Nat Med. 2003; 9:331–337. [PubMed: 12579195]

Kanai-Azuma M, Kanai Y, Gad JM, Tajima Y, Taya C, Kurohmaru M, Sanai Y, Yonekawa H, Yazaki K, Tam PP, et al. Depletion of definitive gut endoderm in Sox17-null mutant mice. Development. 2002; 129:2367–2379. [PubMed: 11973269]

Kim I, Saunders TL, Morrison SJ. Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells. Cell. 2007; 130:470–483. [PubMed: 17655922]

Kroon E, Martinson LA, Kadoya K, Bang AG, Kelly OG, Eliazer S, Young H, Richardson M, Smart NG, Cunningham J, et al. Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. Nat Biotechnol. 2008; 26:443–452. [PubMed: 18288110]

Lee S, Zambas ED, Marsh WL, Redman CM. Molecular cloning and primary structure of Kell blood group protein. Proc Natl Acad Sci U S A. 1991; 88:6353–6357. [PubMed: 1712490]

Lyttle BM, Li J, Krishnamurthy M, Fellows F, Wheeler MB, Goodyer CG, Wang R. Transcription factor expression in the developing human fetal endocrine pancreas. Diabetologia. 2008; 51:1169–1180. [PubMed: 18491072]

Matveyenko AV, Georgia S, Bhushan A, Butler PC. Inconsistent formation and nonfunction of insulin-positive cells from pancreatic endoderm derived from human embryonic stem cells in athymic nude rats. Am J Physiol Endocrinol Metab. 2010; 299:E713–720. [PubMed: 20587750]

McKnight KD, Wang P, Kim SK. Deconstructing pancreas development to reconstruct human islets from pluripotent stem cells. Cell Stem Cell. 2010; 6:300–308. [PubMed: 20362535]

Mfopou JK, Chen B, Mateizel I, Sermon K, Bouwens L. Noggin, retinoids, and fibroblast growth factor regulate hepatic or pancreatic fate of human embryonic stem cells. Gastroenterology. 138:2233–2245. 2245, e2231–2214. [PubMed: 20206178]

Mfopou JK, Chen B, Mateizel I, Sermon K, Bouwens L. Noggin, retinoids, and fibroblast growth factor regulate hepatic or pancreatic fate of human embryonic stem cells. Gastroenterology. 2010; 138:2233–2245. 2245, e2231–2214. [PubMed: 20206178]

Morrison GM, Oikonomopoulou I, Migueles RP, Soneji S, Livigni A, Enver T, Brickman JM. Anterior definitive endoderm from ESCs reveals a role for FGF signaling. Cell Stem Cell. 2008; 3:402–415. [PubMed: 18940732]

Nostro MC, Cheng X, Keller GM, Gadue P. Wnt, activin, and BMP signaling regulate distinct stages in the developmental pathway from embryonic stem cells to blood. Cell Stem Cell. 2008; 2:60–71. [PubMed: 18371422]

Russo D, Wu X, Redman CM, Lee S. Expression of Kell blood group protein in nonerythroid tissues. Blood. 2000; 96:340–346. [PubMed: 10891471]

Sarkar SA, Kobberup S, Wong R, Lopez AD, Quayum N, Still T, Kutchma A, Jensen JN, Gianani R, Beattie GM, et al. Global gene expression profiling and histochemical analysis of the developing human fetal pancreas. Diabetologia. 2008; 51:285–297. [PubMed: 18094957]

Schriebl K, Lim S, Choo A, Tscheliessnig A, Jungbauer A. Stem cell separation: A bottleneck in stem cell therapy. Biotechnol J. 2009

Sherwood RI, Jitianu C, Cleaver O, Shaywitz DA, Lamenzo JO, Chen AE, Golub TR, Melton DA. Prospective isolation and global gene expression analysis of definitive and visceral endoderm. Dev Biol. 2007; 304:541–555. [PubMed: 17328885]

Shivdasani RA. Molecular regulation of vertebrate early endoderm development. Dev Biol. 2002; 249:191–203. [PubMed: 12221001]

Sohn J, Natale J, Chew LJ, Belachew S, Cheng Y, Aguirre A, Lytle J, Nait-Oumesmar B, Kerninon C, Kanai-Azuma M, et al. Identification of Sox17 as a transcription factor that regulates oligodendrocyte development. J Neurosci. 2006; 26:9722–9735. [PubMed: 16988043]

Spence JR, Lange AW, Lin SC, Kaestner KH, Lowy AM, Kim I, Whitsett JA, Wells JM. Sox17 regulates organ lineage segregation of ventral foregut progenitor cells. Dev Cell. 2009; 17:62–74. [PubMed: 19619492]

Tam PP, Kanai-Azuma M, Kanai Y. Early endoderm development in vertebrates: lineage differentiation and morphogenetic function. Curr Opin Genet Dev. 2003; 13:393–400. [PubMed: 12888013]

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. Science. 1998; 282:1145–1147. [PubMed: 9804556]

Vallier L, Mendjan S, Brown S, Chng Z, Teo A, Smithers LE, Trotter MW, Cho CH, Martinez A, Rugg-Gunn P, et al. Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development. 2009; 136:1339–1349. [PubMed: 19279133]

Wen DZ, Dittman WA, Ye RD, Deaven LL, Majerus PW, Sadler JE. Human thrombomodulin: complete cDNA sequence and chromosome localization of the gene. Biochemistry. 1987; 26:4350–4357. [PubMed: 2822087]

Yamaguchi TP, Dumont DJ, Conlon RA, Breitman ML, Rossant J. flk-1, an flt-related receptor tyrosine kinase is an early marker for endothelial cell precursors. Development. 1993; 118:489–498. [PubMed: 8223275]

Yang L, Soonpaa MH, Adler ED, Roepke TK, Kattman SJ, Kennedy M, Henckaerts E, Bonham K, Abbott GW, Linden RM, et al. Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. Nature. 2008; 453:524–528. [PubMed: 18432194]

Yasunaga M, Tada S, Torikai-Nishikawa S, Nakano Y, Okada M, Jakt LM, Nishikawa S, Chiba T, Era T. Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. Nat Biotechnol. 2005; 23:1542–1550. [PubMed: 16311587]

Zeng X, Rao MS. Controlled genetic modification of stem cells for developing drug discovery tools and novel therapeutic applications. Curr Opin Mol Ther. 2008; 10:207–213. [PubMed: 18535927]

Zhou Q, Law AC, Rajagopal J, Anderson WJ, Gray PA, Melton DA. A multipotent progenitor domain guides pancreatic organogenesis. Dev Cell. 2007; 13:103–114. [PubMed: 17609113]
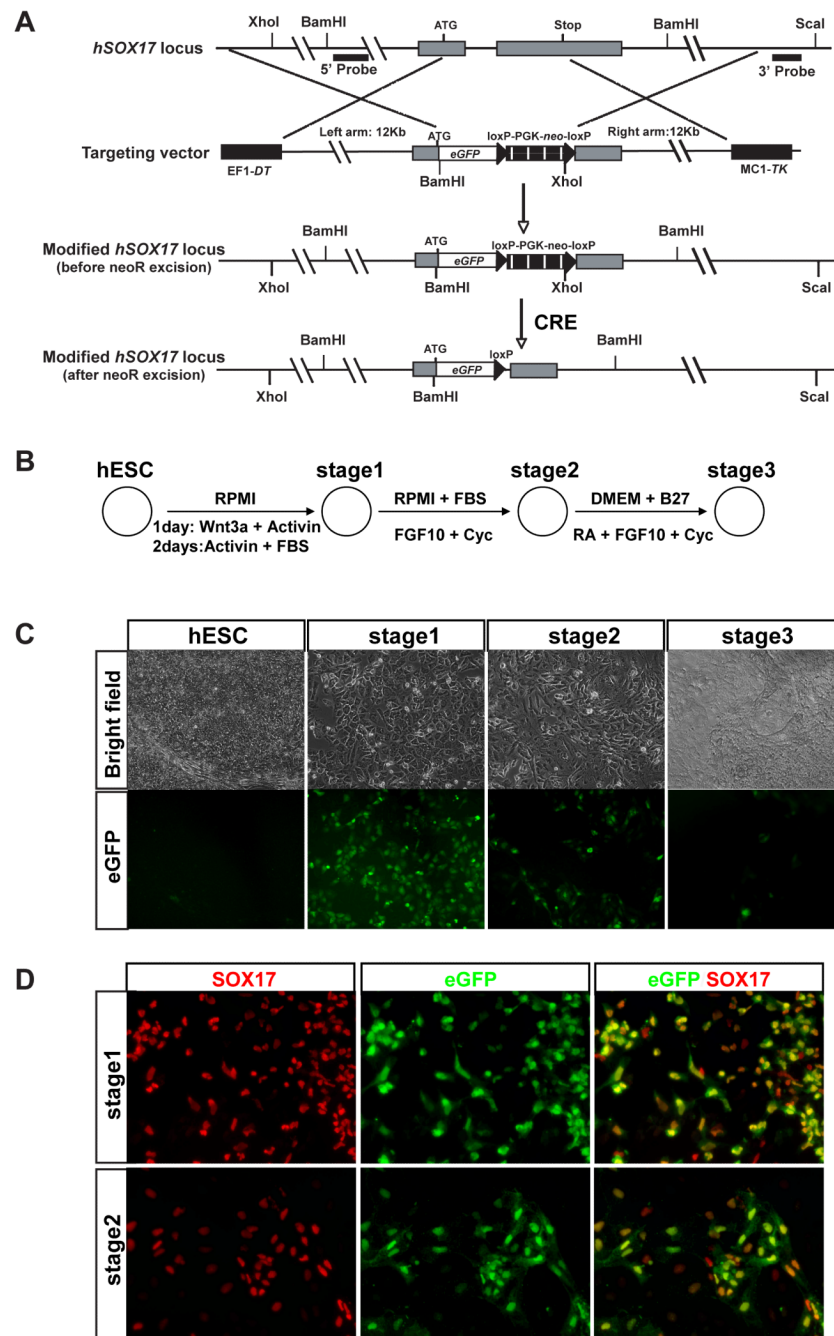
Figure 1.
Targeting of *eGFP* reporter into the human *SOX17* locus. (A) Schematic representation of human *SOX17* targeting strategy. (B) Schematic of endoderm differentiation protocol. (C) eGFP fluorescence visualized throughout differentiation protocol. (D) Immunocytochemistry of eGFP and SOX17 showed colocalization in differentiated hES cells at both stage 1 and 2.
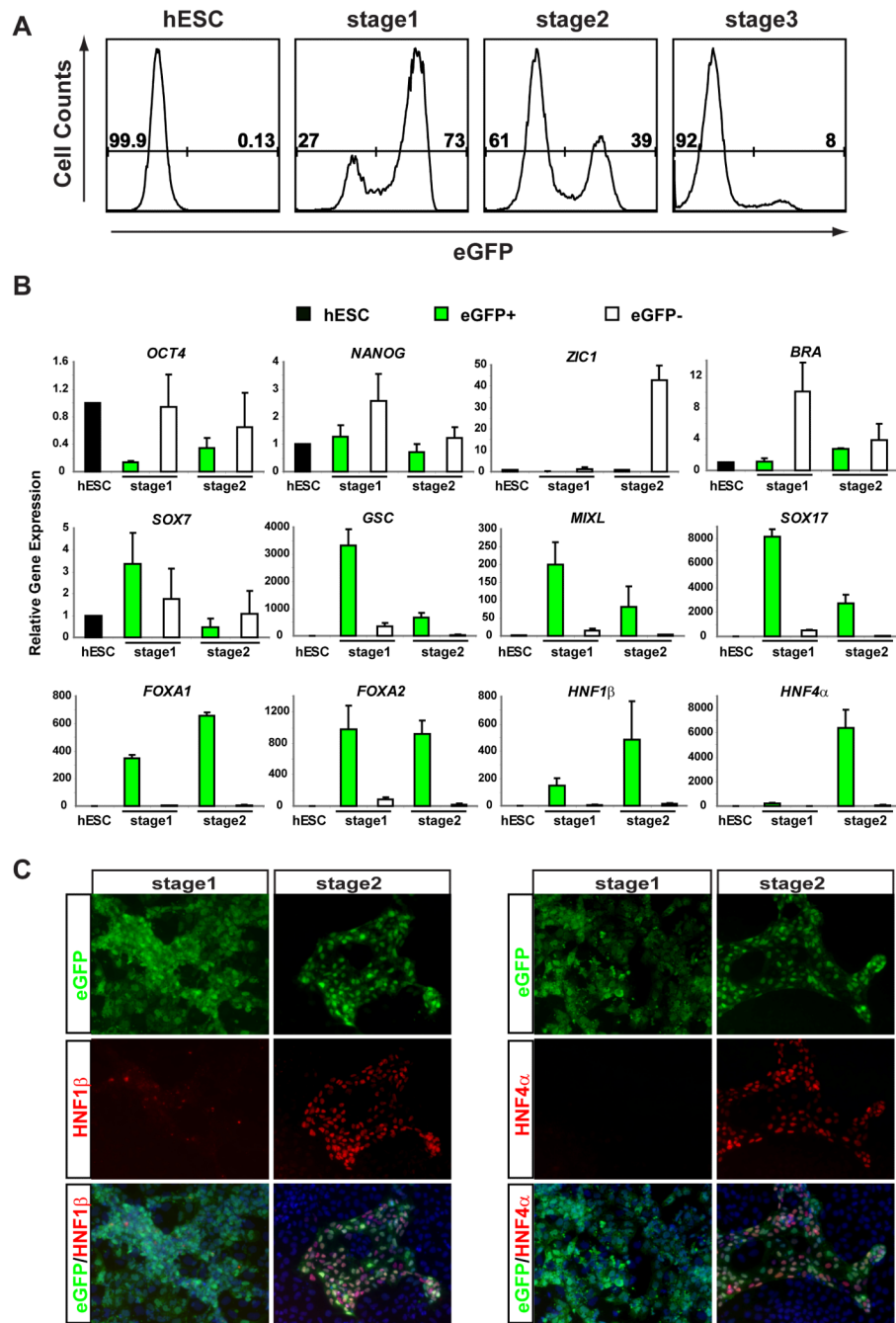
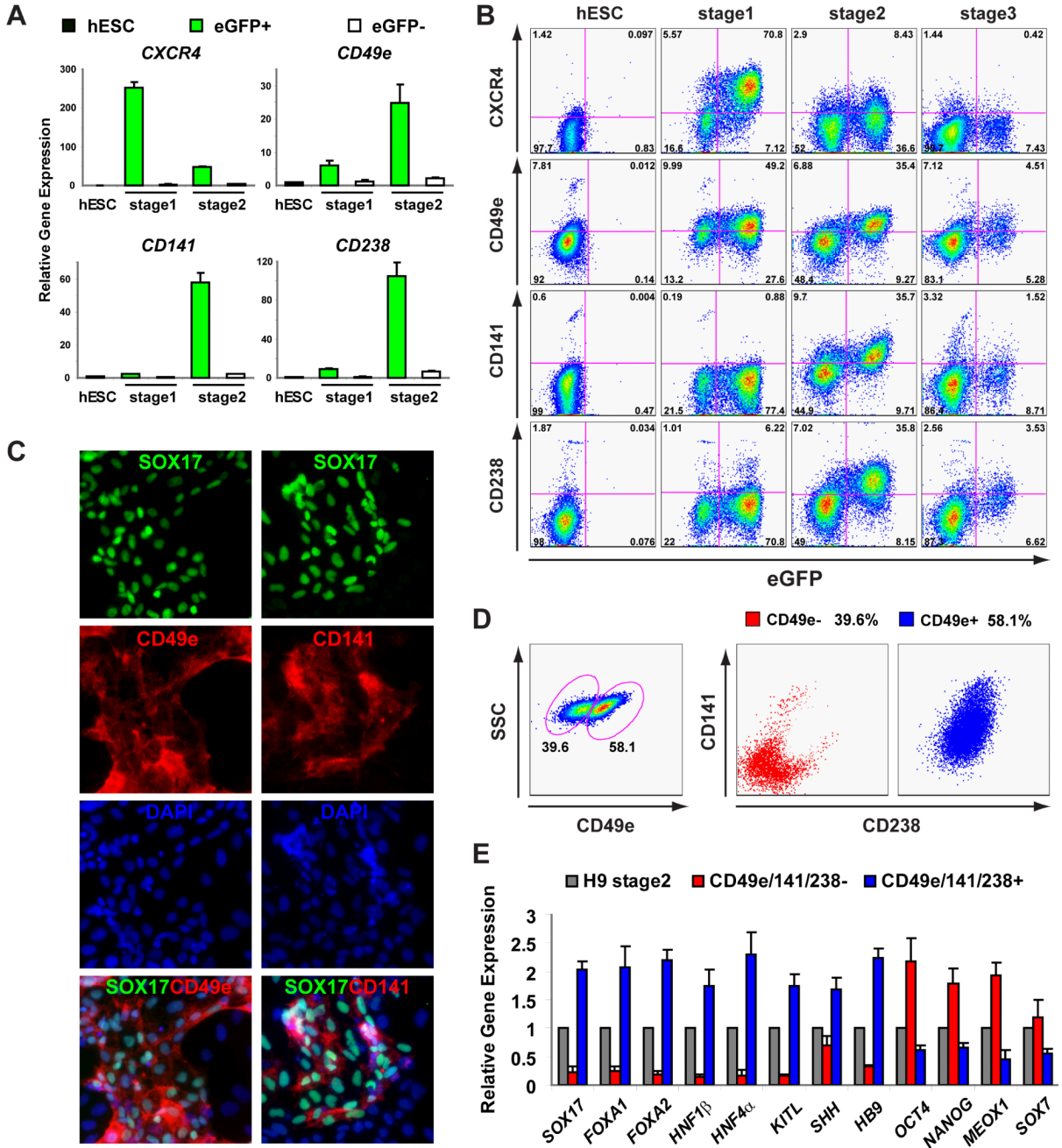**Figure 2.**
Analysis of *in vitro* differentiated SOX17-eGFP hESCs. (A) Flow cytometric analysis of SOX17-eGFP induction during hESC differentiation. (B) Gene expression of sorted eGFP⁺ and eGFP ⁻ cells showed high expression of endoderm specific genes in eGFP⁺ cells and high expression of pluripotent and mesoderm specific genes in eGFP ⁻ cells (mean ± SEM., n=3) at stage 1 and stage 2. The value for hESC was set as 1. (C) HNF1β and HNF4α were not expressed at stage 1 but were coexpressed with eGFP at stage 2 in primitive gut tube endoderm.
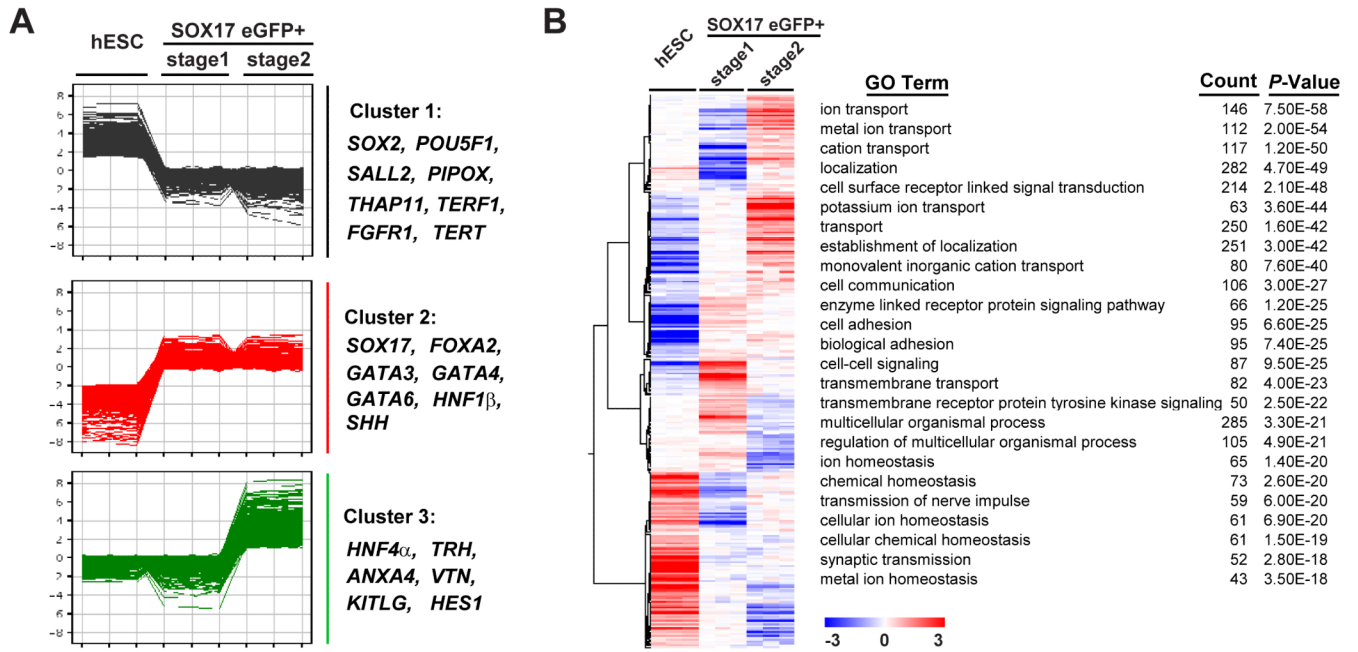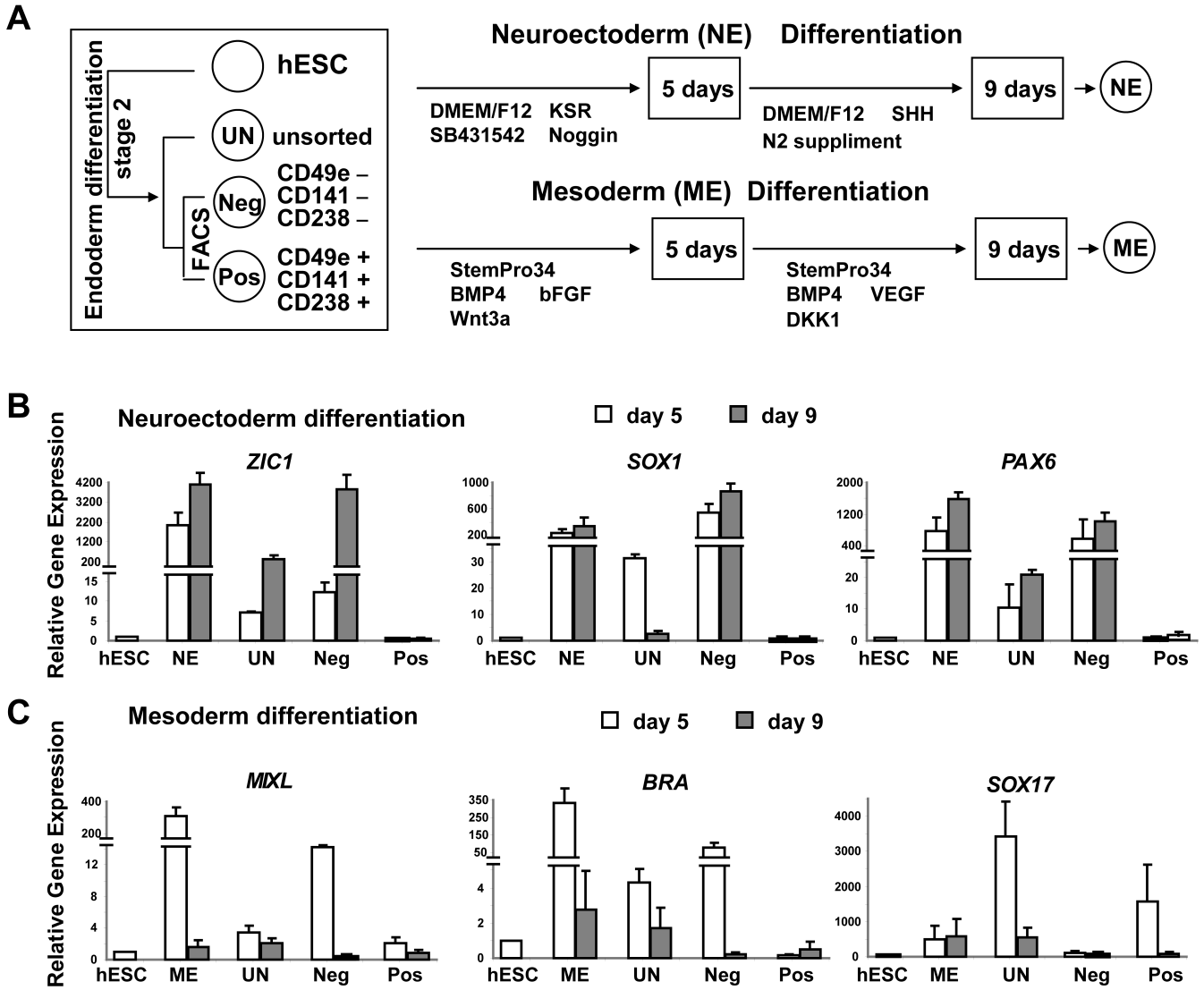
Figure 3.
Microarray analysis comparing gene expression of undifferentiated hESCs and their *in vitro* derived SOX17-eGFP+ endoderm at stage 1 and stage 2. (A) K-mean clustering showed gene expression patterns of three classes of genes that were changed during differentiation. (B) Hierarchical clustering and GO term analysis of 608 transmembrane proteins differentially expressed between undifferentiated hESCs and eGFP+ cells at stage 1 and 2. Color scale indicates normalized expression values.
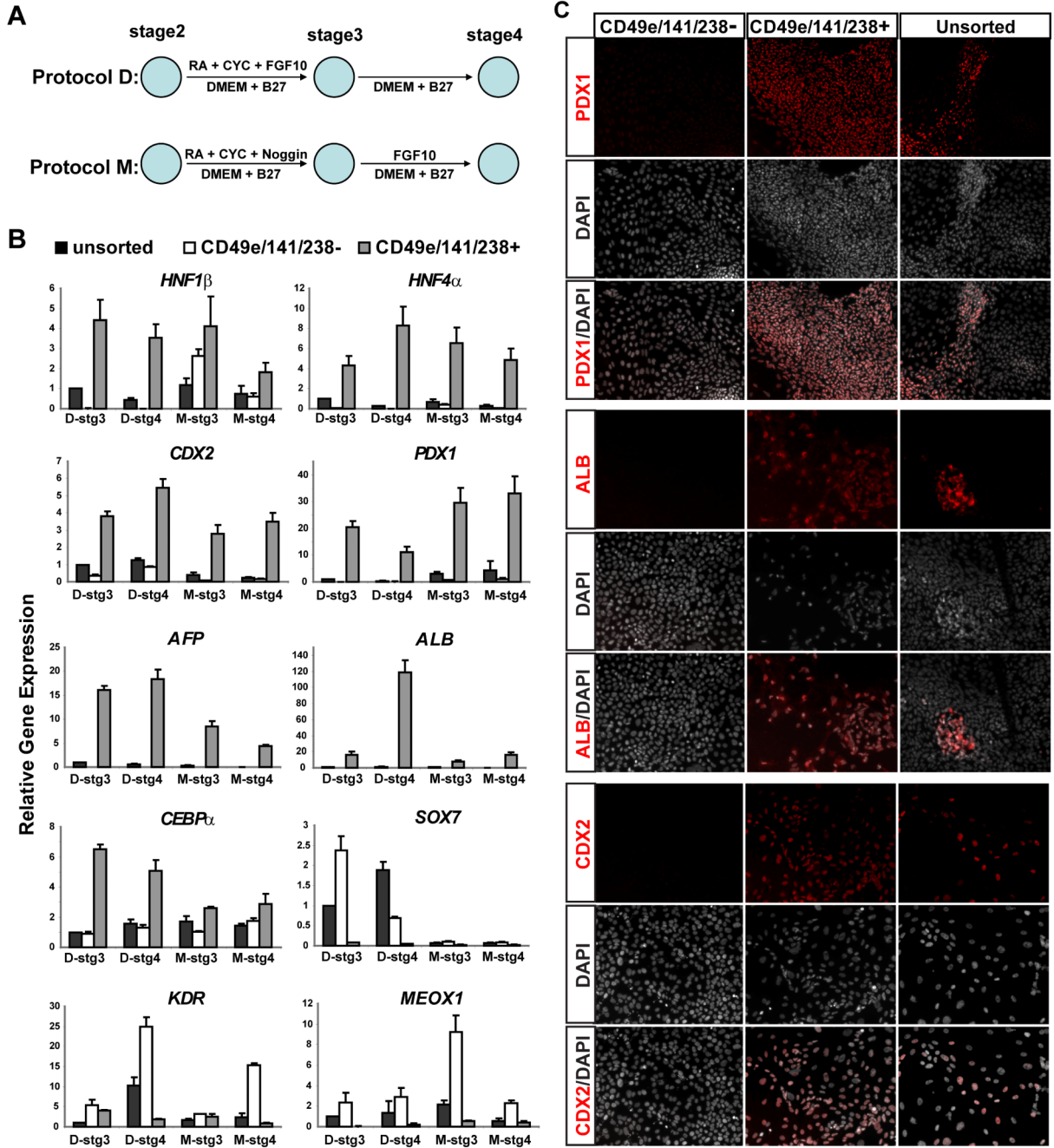
Figure 4.
Novel cell surface markers expressed in hESC derived endoderm. (A) mRNA expression levels of *CXCR4, CD49e, CD141* and *CD238* in hESCs, eGFP+ and eGFP − cells at stage 1 and 2 (mean ± SEM., n=3). (B) Flow cytometric analysis showed the expression of CXCR4, CD49e, CD141 and CD238 changed during hESC differentiation. (C) SOX17 costained with CD49e and CD141 in stage 2 cells by immunocytochemistry. (D) Sorting of CD49e+ CD141+ CD238+ cells by flow cytometry. CD49e^{low} cells were negative for both CD141 and CD238. Similarly CD49e+ cells were positive for both CD141 and CD238. (E) Gene expression analysis showed that endoderm markers were highly enriched in CD49e/CD141/

CD238 positive cells; the expression levels of *OCT4, NANOG, MEOX1* and *SOX7* were higher in CD49e/CD141/CD238 negative cells (mean ± SEM., n=3).

A

B

C

Figure 5.
Analysis of the differentiation potential of stage 2 cells using neuroectoderm and mesoderm protocols. (A) Schematic of neuroectoderm and mesoderm differentiation protocols. hESC: undifferentiated hESC, NE: neuroectoderm differentiated from hESCs, ME: mesoderm differentiated from hESCs, UN: unsorted stage 2 cells, Neg: FACS sorted CD49e/CD141/ CD238 negative cells, and Pos: FACS sorted CD49e/CD141/CD238 positive cells (mean ± SEM., n=3). (B) Gene expression analysis of cells cultured under neuroectoderm differentiation condition for 5, and 9 days. (C) Gene expression analysis of cells cultured under mesoderm differentiation condition for 5, and 9 days (mean ± SEM., n=3).

Figure 6.
Analysis of endoderm potential of FACS isolated endoderm derived from hESCs. (A) Schematic of differentiation protocols from D'Amour (D) and Mfopou (M) from stage 2 to stage 4. (B) Gene expression analysis of stage 2 cells differentiated toward endoderm lineage using two protocols. stg3: stage 3, stg4: stage4 (mean ± SEM., n=3). (C) Immunocytochemistry for PDX1 with Mfopou protocol and for ALB and CDX2 with D'Amour protocol on stage 4 cells differentiated from either CD49e/CD141/CD238– or CD49e/CD141/CD238+ sorted at the end of stage 2 or unsorted stage 2 cells.
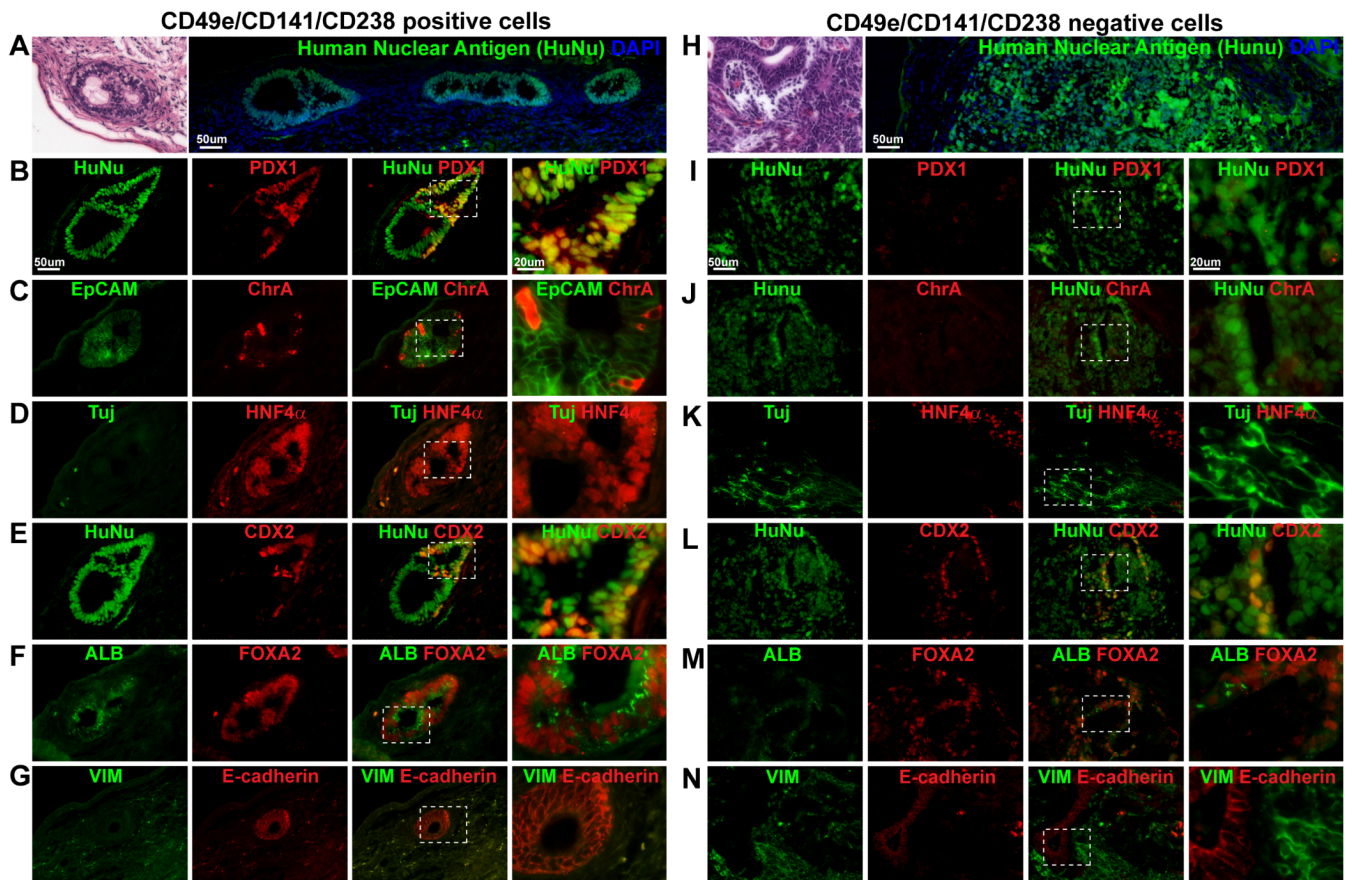
Figure 7.
Transplantation-based analysis of developmental potential in isolated hESC derived endoderm. (A-E) CD49/141/238 positive cells co-transplanted with MEF into murine kidney capsules (n=3) further differentiated in vivo and expressed markers of tissues derived from the endodermal lineage. (A). H&E staining and human nuclear antigen (HuNu) immunostaining showed epithelial structure. Immunostaining of transplant with antibodies to (B).HuNu and PDX1. (C). EpCAM and Chromogranin A (ChrA). (D). Tuj and HNF4α. (E). HuNu and CDX2. (F). ALB and FOXA2. (G). VIM and E-cadherin. (H-N) CD49/141/238 negative cells co-transplanted with MEF differentiated into teratoma (n=3) that expressed markers of ectoderm, mesoderm and endoderm lineages. (H). H&E staining and HuNu immunostaining showed mixed structures. Immunostaining of teratoma with antibodies to (I).HuNu and PDX1. (J). HuNu and Chromogranin A (ChrA). (K). TUJ1 and HNF4α. (L). HuNu and CDX2. (M). ALB and FOXA2. (N). VIM and E-cadherin.