Supplemental Information

Targeting *SOX17* in Human Embryonic Stem Cells

Creates Unique Strategies for Isolating

and Analyzing Developing Endoderm

Pei Wang, Ryan T. Rodriguez, Jing Wang, Amar Ghodasara, and Seung K. Kim

A

5' internal probe

H9 WT   KI

WT

Mu

Digestion:
BamHI
WT : 10Kb
Mu : 5kb

3' external probe

H9 WT   KI

WT

Mu

Digestion:
ScaI + XhoI
WT : 22kb
Mu : 12kb

B

hS17

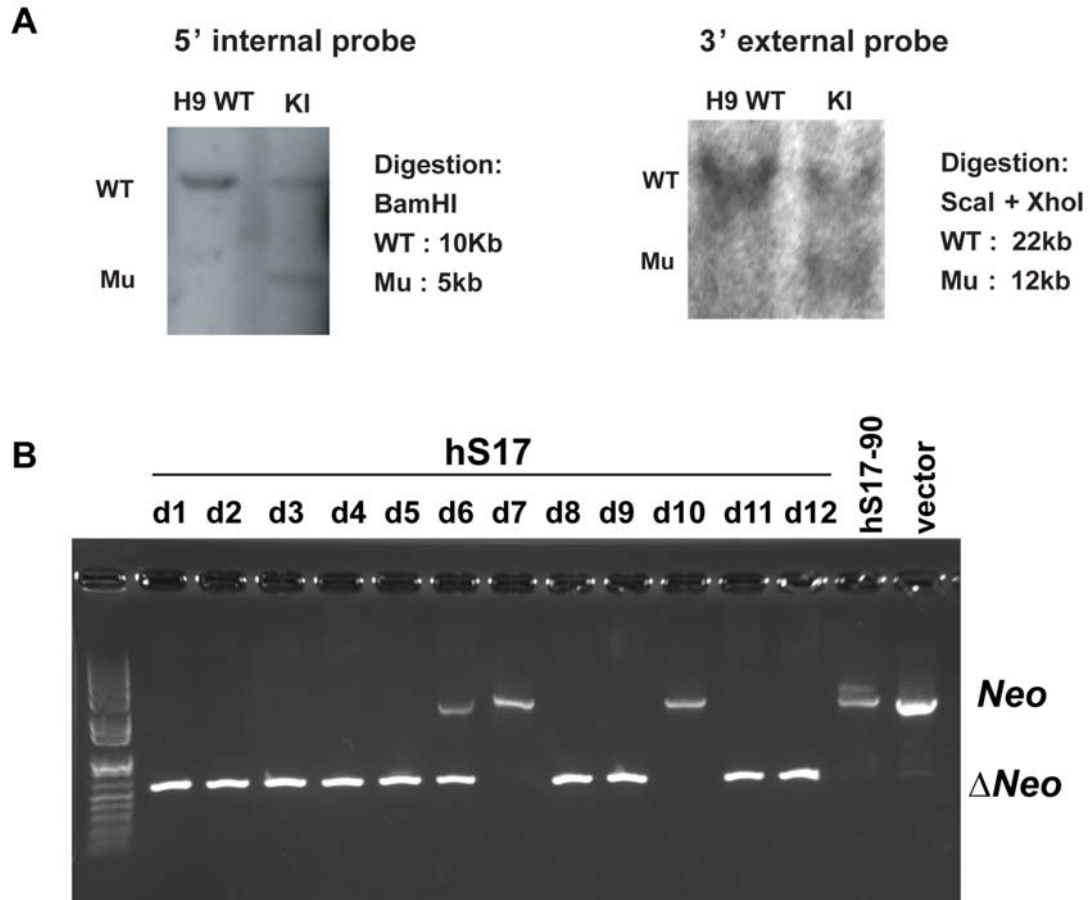d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12   hS17-90   vector

Neo

ΔNeo

Figure S1, related to Figure 1. Characterization of hS17 hESC lines. (A) Southern blot confirmed targeting of the eGFP-encoding transgene to the endogenous *SOX17* locus in hS17-90 hESC. (B) PCR genotyping showed successfully deleted Neo cassette in clone 1, 2, 3, 4, 5, 8, 9, 11 and 12. Primers used here were 5'-GCATCATAATCAGCCATACC-3' and 5'-AAACTGTTCAAGTGGC AGAC-3'.
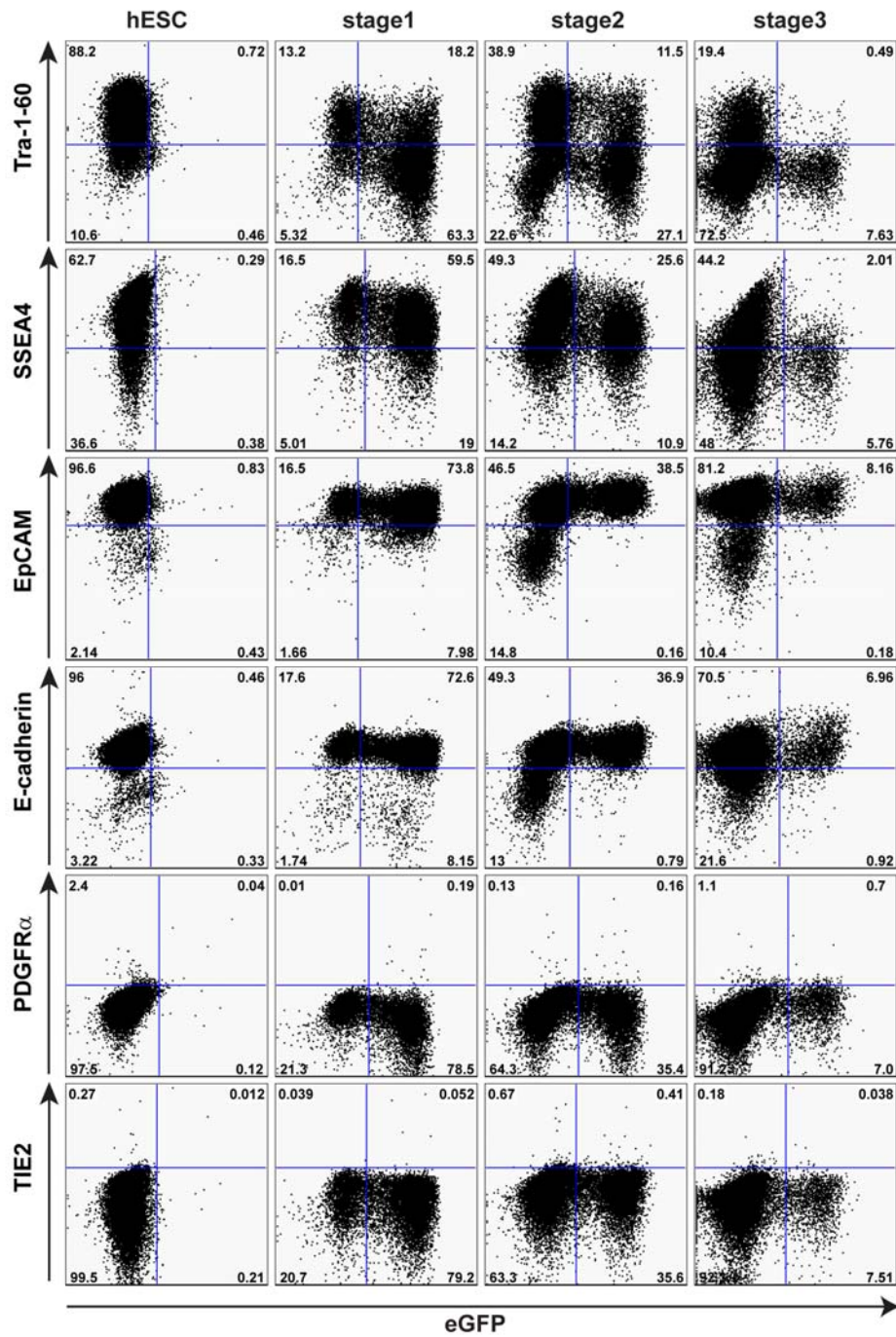
Figure S2, related to Figure 2. Flow cytometry analysis of cell surface markers of hS17-d2 hESC line during in vitro endoderm differentiation. Pluripotent markers Tra-1-60 and SSEA4, were downregulated after hS17-d2 ESC differentiation. EpCAM and E-cadherin were expressed highly in undifferentiated hESCs and SOX17-eGFP⁺ cells. PDGFRα and TIE2 were not induced during endoderm differentiation.

A

hESC	stage1	stage2	CD	Symbol
			CD184	CXCR4
			CD49d	ITGA4
			CD48	CD48
			CD49e	ITGA5
			CD141	THBD
			CD238	KEL
			CD99	CD99
			CD325	CDH2
			CD51	ITGAV
			CD130	IL6ST
			CDH12	CDH12
			CD88	C5AR1
			CD252	TNFSF4
			CD249	ENPEP
			CD163	CD163

-3   0   3

B

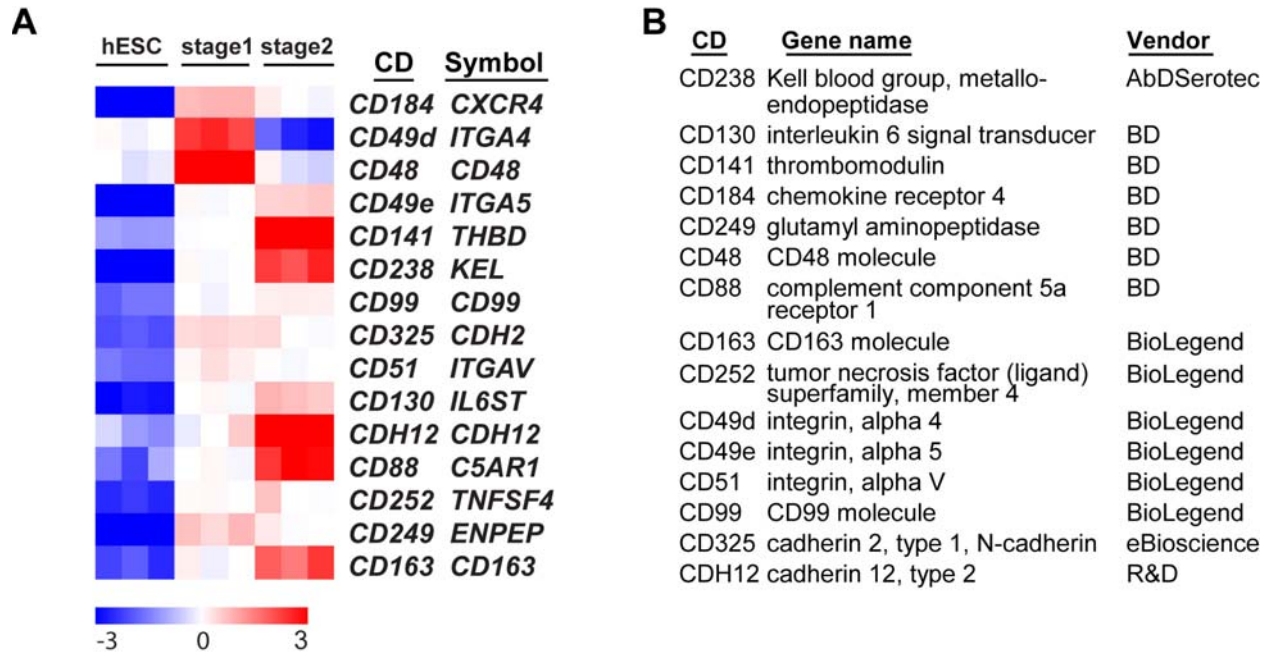| CD	Gene name	Vendor
CD238	Kell blood group, metallo-endopeptidase	AbDSerotec
CD130	interleukin 6 signal transducer	BD
CD141	thrombomodulin	BD
CD184	chemokine receptor 4	BD
CD249	glutamyl aminopeptidase	BD
CD48	CD48 molecule	BD
CD88	complement component 5a receptor 1	BD
CD163	CD163 molecule	BioLegend
CD252	tumor necrosis factor (ligand) superfamily, member 4	BioLegend
CD49d	integrin, alpha 4	BioLegend
CD49e	integrin, alpha 5	BioLegend
CD51	integrin, alpha V	BioLegend
CD99	CD99 molecule	BioLegend
CD325	cadherin 2, type 1, N-cadherin	eBioscience
CDH12	cadherin 12, type 2	R&D

Figure S3, related to Figure 3. Gene expression analysis of hESCs and endoderm derivatives.
(A) 15 surface proteins were highly expressed in hESC-derived stage 1 and stage 2 endoderm.
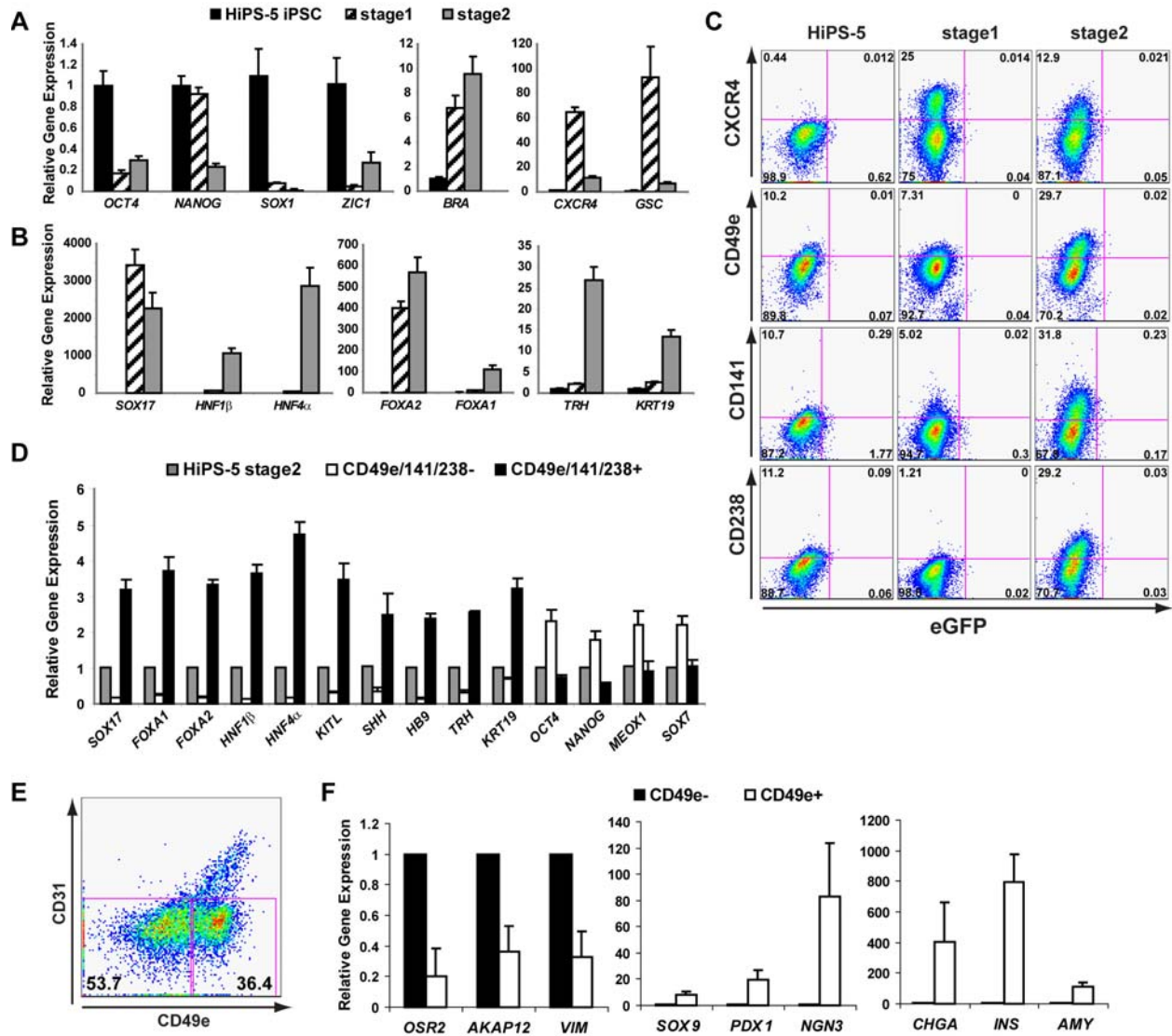(B) Vendors for antibodies to 15 cell surface proteins.

Figure S4, related to Figure 4. iPSC line HiPSC-5 differentiated towards endodermal lineage was similar to hESCs. (A). Pluripotent markers were downregulated during differentiation and mesodermal markers were induced in stages 1 and 2 (mean ± SEM, n=3). (B). Endodermal markers were greatly induced at stages 1 and 2 (mean ± SEM, n=3). (C). Flow cytometry analysis of CXCR4, CD49e, CD141 and CD238 during HiPSC-5 differentiation. (D). Gene expression analysis showed that endoderm markers were highly enriched in CD49e/CD141/CD238 positive cells; the expression of *OCT4*, *NANOG*, *MEOX1* and *SOX7* were higher in CD49e/CD141/CD238 negative cells (mean ± SEM, n=3), similar to hESCs. (E) FACS isolation of endoderm derivative from human fetal pancreas. Human fetal pancreas was separated into two fractions, CD49e$^+$ and CD49e$^-$ using a CD49e antibody. (F) Mesenchymal markers were expressed higher in CD49e$^-$ cells, while the pancreatic marker genes were enriched in CD49e$^+$ cells (mean ± SEM, n=3).
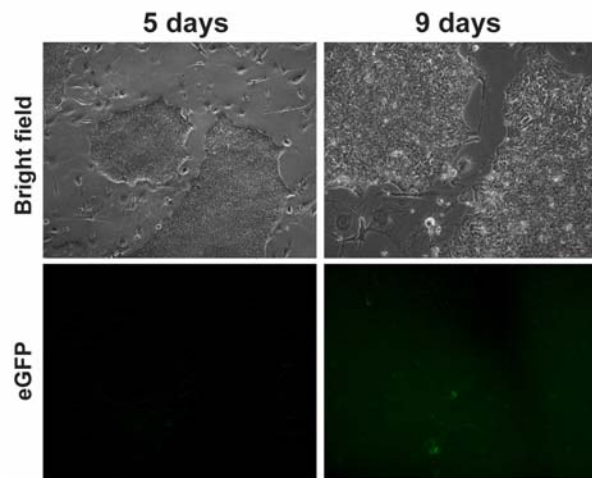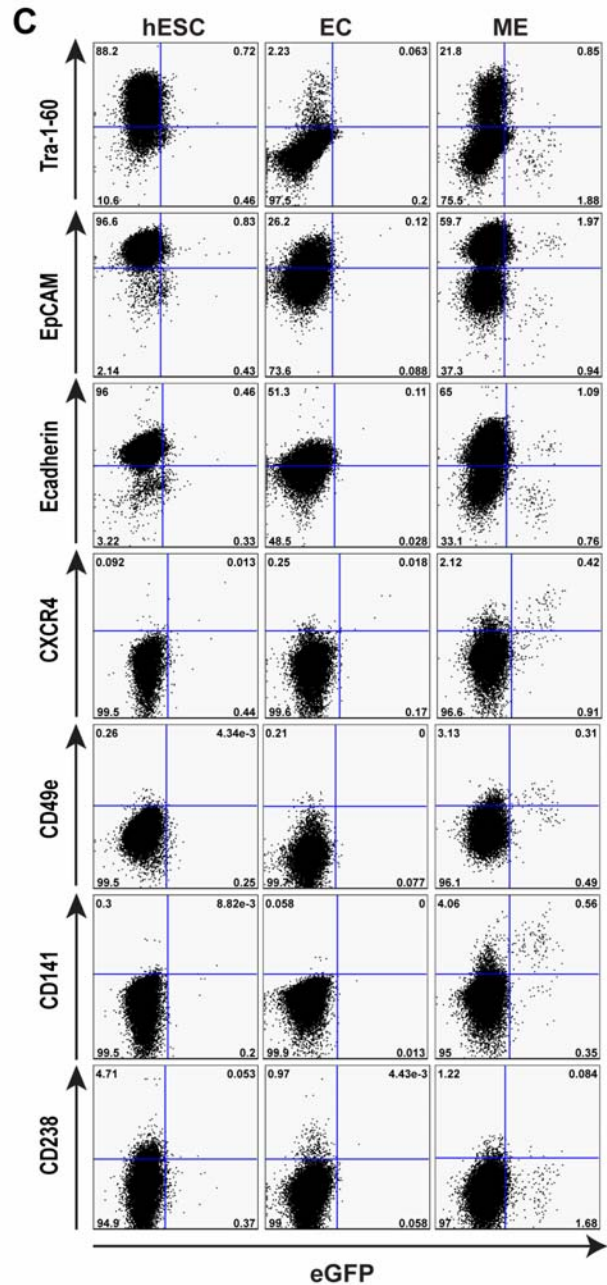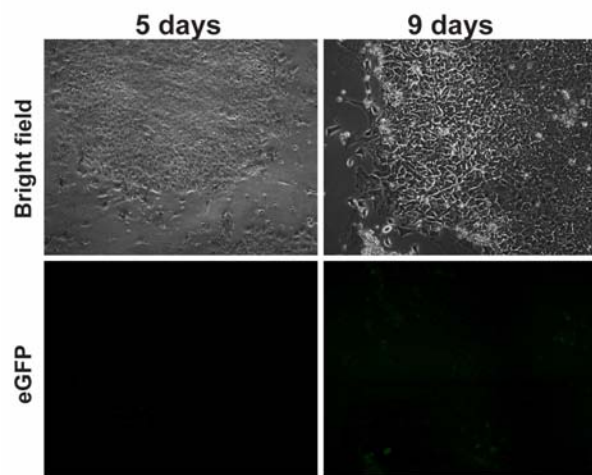
Figure S5, related to Figure 5. Analysis of the differentiation potential of stage 2 cells using neuroectoderm and mesoderm protocols. (A). hESCs did not express eGFP when differentiated using neuroectoderm protocol. (B). hESCs did not express eGFP when differentiated using mesoderm protocol. (C). Flow cytometry analysis of hESCs differentiated for 9 days using neuroectoderm protocol (NE) or mesoderm protocol (ME). CXCR4, CD49e, CD141 and CD238 were negative/low in NE and ME.
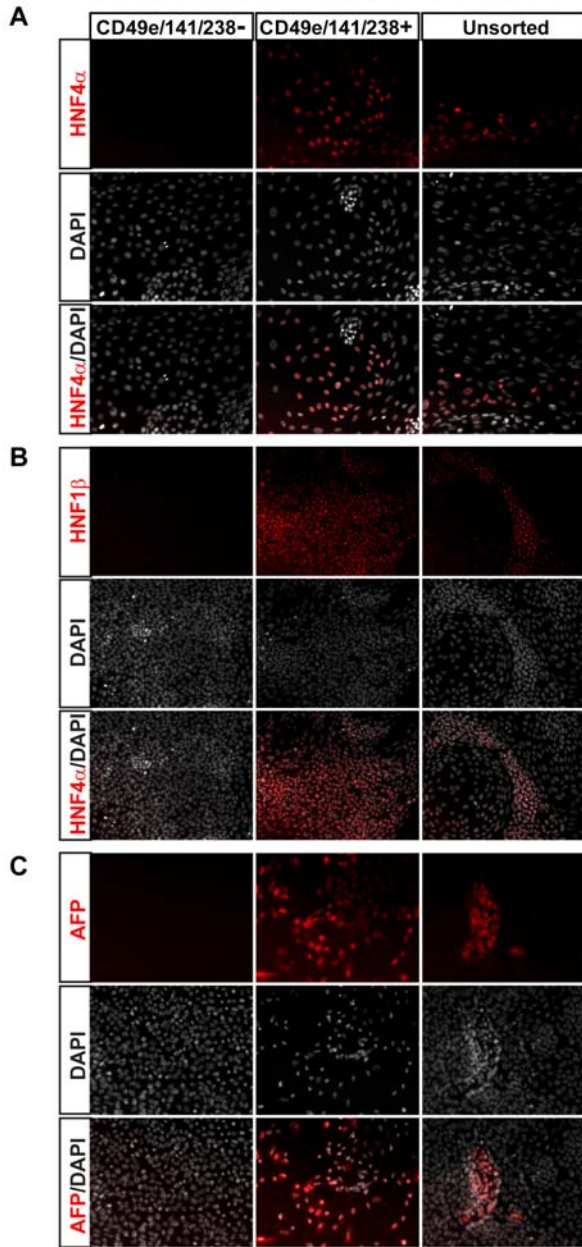
Figure S6, related to Figure 6. Analysis of endoderm potential of FACS isolated endoderm derived from hESCs using D'Amour protocol from stage 2 to stage 4. Immunostaining of (A). HNF1β, (B). HNF4α and (C). AFP on stage 4 cells differentiated from unsorted and sorted stage 2 cells. All markers had no expression in CD49e/CD141/CD238 negative cells but were expressed in CD49e/CD141/CD238 positive cells.
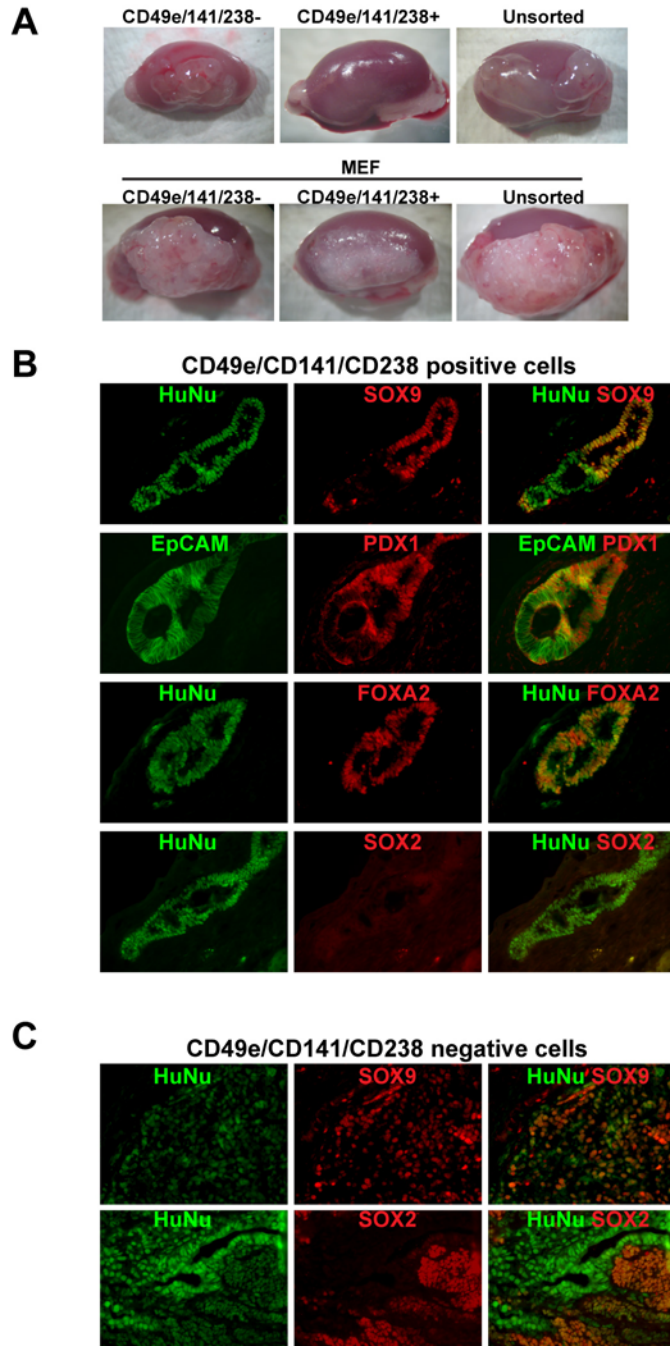
Figure S7, related to Figure 7. Transplantation-based analysis of developmental potential in isolated hESC derived endoderm. (A). Engraftments of transplantations of unsorted and sorted cells with or without MEF showed different growth. (B) CD49/141/238 positive cells co-transplanted with MEF differentiated into tissues that expressed markers of endodermal lineage. (C) CD49/141/238 negative cells co-transplanted with MEF differentiated into teratoma that expressed markers of ectoderm, mesoderm and endoderm lineages.

Table S1. Q-PCR assays and antibodies used in the study.

Gene	ABI Assay	Antibody	Catalog #	Vendor
ACTIN	Hs99999903_m1	MOUSE ANTI HUMAN CD238	MCA1987T	AbDSerotec
AFP	Hs00173490_m1	CD184 (CXCR4, Fusin)-biotin	551968	BD
AKAP12	Hs00374507_m1	CD141	559780	BD
ALB	Hs00609411_m1	CD141-PE	559781	BD
AMY2B	Hs00949916_m1	Alexa Fluor® 647 anti-mouse CD31	102415	biolegend
BRA	Hs00610080_m1	APC anti-human CD326 (EpCAM)	324207	biolegend
CDX2	Hs00230919_m1	APC anti-human CD184 (CXCR4)	306509	biolegend
CEBPa	Hs00269972_s1	PE anti-human SSEA-4	330405	biolegend
CHGA	Hs00154441_m1	Alexa Fluor® 647 Anti-human CD31 Antibody	303111	biolegend
FOXA1	Hs00270129_m1	Biotin Anti-human CD49e Antibody	328005	biolegend
FOXA2	Hs00232764_m1	Biotin Anti-human TRA-1-60-R Antibody	330603	biolegend
GSC	Hs00418279_m1	Biotin Anti-human TRA-1-81 Antibody	330703	biolegend
HB9	Hs00232128_m1	Biotin Anti-mouse CD326 (Ep-CAM) Antibody	118204	biolegend
HNF1b	Hs01001602_m1	Streptavidin-PE/Cy7	405206	biolegend
HNF4a	Hs01023298_m1	PE Anti-human CD202b (Tie2/Tek) Antibody	334205	biolegend
INS	Hs00355773_m1	APC Anti-human CD324 (E-Cadherin) Antibody	324107	biolegend
ITGA5 (CD49e)	Hs01547673_m1	APC Anti-human CD49e Antibody	328011	bolegend
KDR	Hs00176676_m1	Biotin Anti-human CD140a (PDGFRα) Antibody	323503	bolegend
KEL (CD238)	Hs00166270_m1	Human Kell Affinity Purified Ab, Goat IgG	AF1914	R&D
KITLG	Hs00241497_m1	Anti-human SOX17 Antibody	AF1924	R&D
KRT19	Hs00761767_s1	Qdot® 605 streptavidin conjugate	Q10101MP	invitrogen
MEOX1	Hs00244943_m1	Alexa Fluor® 633 donkey anti-goat IgG (H+L)	SKU# A-21082	invitrogen
MIXl1	Hs00430824_g1	Donkey Anti-Goat IgG (bio)	705-065-147	Jackson Immuno
NANOG	Hs02387400_g1	Purified Anti-human CD49e Antibody	328002	biolegend
NEUROG3	Hs00360700_g1	Rabbit Monoclonal, Clone: EP700Y, E-cadherin	246R-14	Cell marque
OCT4	Hs01895061_u1	Mouse Monoclonal, Clone: V9, VIMENTIN	347M-14	Cell marque
OSR2	Hs00369588_m1	Rabbit Monoclonal, Clone: EPR2764Y, CDX2	235R-14	Cell marque
PAX6	Hs00240871_m1	Neuronal Class III β-Tubulin (TUJ1) Antibody	MMS-435P-100	Convance
PDX1	Hs00236830_m1	Chromagranin A	20085	ImmunoStar
SHH	Hs00179843_m1	Anti-GFP, rabbit IgG fraction	A-11122	invitrogen
SOX1	Hs00534426_s1	Chicken anti-GFP	ab13970	Abcam
SOX17	Hs00751752_s1	Rabbit anti-FOXA2	AB4125	Millipore
SOX7	Hs00846731_s1	Human nuclear antigen	MAB1281	Millipore
SOX9	Hs00165814_m1	Goat anti-HNF1b	SC-7411	santa cruz
THBD (CD141)	Hs00264920_s1	Rabbit anti-HNF4a	SC-8987	santa cruz
TRH	Hs00175078_m1	goat anti-SOX2	SC-17320	santa cruz
VIMENTIN	Hs00958116_m1	Monoclonal Anti-α-Fetoprotein (AFP) antibody	A8452-.2ML	sigma
ZIC1	Hs00602749_m1	Monoclonal Anti-Albumin antibody	A6684-.2ML	sigma
		Rabbit anti-PDX1		Gift of Edlund lab Umea University