

# A Molecular Signature for Purified Definitive Endoderm Guides Differentiation and Isolation of Endoderm from Mouse and Human Embryonic Stem Cells

Pei Wang,<sup>1</sup> Kristen D. McKnight,<sup>1</sup> David J. Wong,<sup>2</sup> Ryan T. Rodriguez,<sup>1</sup> Takuya Sugiyama,<sup>1</sup> Xueying Gu,<sup>1</sup> Amar Ghodasara,<sup>1</sup> Kun Qu,<sup>2</sup> Howard Y. Chang,<sup>2,3</sup> and Seung K. Kim<sup>1,3,4</sup>

Embryonic definitive endoderm (DE) generates the epithelial compartment of vital organs such as liver, pancreas, and intestine. However, purification of DE in mammals has not been achieved, limiting the molecular “definition” of endoderm, and hindering our understanding of DE development and attempts to produce endoderm from sources such as embryonic stem (ES) cells. Here, we describe purification of mouse DE using fluorescence-activated cell sorting (FACS) and mice harboring a transgene encoding enhanced green fluorescent protein (eGFP) inserted into the *Sox17* locus, which is expressed in the embryonic endoderm. Comparison of patterns of signaling pathway activation in native mouse DE and endoderm-like cells generated from ES cells produced novel culture modifications that generated *Sox17*-eGFP<sup>+</sup> progeny whose gene expression resembled DE more closely than achieved with standard methods. These studies also produced new FACS methods for purifying DE from nontransgenic mice and mouse ES cell cultures. Parallel studies of a new human *SOX17*-eGFP ES cell line allowed analysis of endoderm differentiation in vitro, leading to culture modifications that enhanced expression of an endoderm-like signature. This work should accelerate our understanding of mechanisms regulating DE development in mice and humans, and guide further use of ES cells for tissue replacement.

## Introduction

THE DEFINITIVE ENDODERM (DE) is 1 of the 3 germ layers in mammalian embryos and is the progenitor for the functioning epithelial component of all organs comprising the gastrointestinal (GI) and respiratory tracts, including lungs, liver, and pancreas. Delineation of gene expression patterns and signaling pathways that distinguish DE from other germ layers or extraembryonic tissues should provide fundamental insights about mechanisms controlling internal organ development [1]. However, previous studies of embryonic mice reported only partial purification of DE from contaminating visceral endoderm (VE), which generates extraembryonic tissues such as yolk sac; thus, gene expression profiles reported for mouse endoderm isolated by microdissection [2] or by fluorescence-activated cell sorting (FACS) from transgenic mice [3,4] had some overlap, but were largely distinct from one another. Thus, characterization of a distinct molecular signature for mouse DE remains incomplete.

The goal of GI and respiratory organ replacement or regeneration has promoted prolonged, intensive efforts to guide development of renewable cell sources such as em-

embryonic stem (ES) cells toward an endodermal fate [5–8]. In mice and other animals, endoderm formation, patterning, and differentiation is the culmination of a dynamic, complex series of cell fate decisions and morphogenetic movements orchestrated by Wnt, Nodal/Activin, bone morphogenetic protein (BMP), fibroblast growth factor (FGF), and retinoic acid (RA) signaling [9–11]. However, exposure of mouse and human ES cells to a combination of purified Wnt and Activin A, which only recapitulates a subset of the signals that regulate endoderm development, is a common method for producing heterogeneous cultures that include progeny with endoderm-like properties [7,12]. Recently, small molecule screens have identified individual index compounds that also produced endoderm-like cells on exposure to mouse ES cultures [4]. A gene expression profile of native DE would be necessary for assessing these ES cell-derived products, and could suggest how to modify ES cell culture conditions to enhance their molecular similarity to native DE. However, due to an inability to isolate DE, this fundamental comparison of native DE and endoderm-like ES cell progeny has not been reported.

*Sox17* encodes a transcription factor expressed in definitive and VE that controls development of these tissues

<sup>1</sup>Department of Developmental Biology, and <sup>2</sup>Program in Epithelial Biology Stanford University School of Medicine, Stanford, California.

<sup>3</sup>Howard Hughes Medical Institute, Stanford, California.

<sup>4</sup>Department of Medicine (Oncology), Stanford University School of Medicine, Stanford, California.

[13,14]. Thus, while useful for discriminating endoderm from other germ layers, *Sox17* expression alone is not sufficient to distinguish visceral and DE. Recent studies used transgenic mice expressing a modified yellow fluorescent protein [3] or dsRed protein [4] from the *Sox17* locus to isolate endoderm by FACS, but separation of definitive and VE was not achieved. Here, we describe FACS purification of definitive and VE from embryonic mice, and report gene expression profiling of native mouse DE. Based on this gene expression profile, we developed a strategy to isolate and separate visceral and DE cells from nongenetically modified mouse embryos, as well as ES cells. Gene expression of purified native mouse DE was not fully recapitulated in endoderm-like progeny derived from mouse ES cells *in vitro*. However, modification of culture conditions resulted in endoderm-like progeny produced from mouse and human ES cell cultures that more closely resembled native DE.

## Materials and Methods

### Generation of *Sox17-eGFP knock-in ES cells and mice*

The targeting construct shown in Supplementary Fig. S1A (Supplementary Data are available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)) was used to generate mouse ES cells expressing eGFP from the endogenous *Sox17* locus. The left and right homologous arms (1.8kb and 7.2kb, respectively) were prepared by polymerase chain reaction (PCR) of genomic DNA from bacterial artificial chromosomes (Oakland Children's Hospital). The eGFP transgene was inserted in-frame directly after the endogenous *Sox17* start codon. The neomycin resistance (*neo*) gene was inserted after the eGFP cassette. The thymidine kinase (*TK*) gene served as a marker to select against the integration of vector sequences. After linearization, the construct was electroporated into R1 mouse ES cells using previously described methods [15]. After 8 days' selection with the drugs G418 and gancyclovir, ES cell colonies were picked and expanded. The correctly sized product (2.2kb) after PCR genotyping with appropriate primers indicated correct targeting at the *Sox17* locus. The genotype was then confirmed by Southern blotting using probes that detect sequences outside of the homology arms (Supplementary Fig. S1B): the wild-type allele produced a 10kb product, while the mutant allele produced the expected 5kb product. Three independent lines of correctly targeted mouse ES cells were injected into blastocysts of C57BL/6 mice at the Transgenic Research Center at Stanford School of Medicine. Germ line transmission was confirmed by both Southern blot and PCR. The primers used for genotyping were S17-F, CGCTCAGCTTACGAGTTCC, eEGFP-R AA GTCGTGCTGCTTCATGTG, with a product size of 385bp. The transgenic mice were back-crossed with FVB strain and maintained on a FVB background. FVB mice were purchased from Charles River Laboratories. All animal studies were performed in accordance with Stanford University Animal Care and Use guidelines.

### Endoderm dispersion and cell sorting

E7.5 and E8.25 mouse embryos were dissected from pregnant mice in Dulbecco's phosphate-buffered saline (PBS) followed by exposure to 0.05% trypsin/ethylenediaminete-

traacetic acid (Invitrogen) for 5 min at 37°C. Trypsin was neutralized with FACS buffer [PBS, 2 mM ethylene glycol tetraacetic acid (EGTA), and 2% fetal bovine serum (FBS)]. Cells were triturated and treated for 15 min in a blocking solution composed of FACS buffer containing 300 ng/mL rat IgG (Jackson ImmunoResearch). The following primary antibodies were used in this study: biotin anti-CXCR4 (2B11, 1:100; BD Biosciences), PB-anti-CD24 (M1/69, 1:100; Biolegend), APC-Cy7-anti-CD38 (clone10, 1:100; Biolegend), APC-EpCAM (Biolegend), and PE anti-CD55 (RIKO-5, 1:100; BD Biosciences). Streptavidin-Qdot605 (1:200; Invitrogen) was used to detect biotinylated antibodies. Propidium iodide was used to exclude dead cells. Both mouse and human ES cells and their progeny were analyzed and sorted using an FACSaria (BD Biosciences). FACS data were analyzed using FlowJo software (Tree Star).

### Mouse and human ES cell culture and differentiation

Undifferentiated mouse ES cells were cultured on a feeder layer of irradiated mouse embryonic fibroblasts (MEF) in M15 medium containing knockout-Dulbecco's modified Eagle's medium (DMEM), 15% FBS (HyClone), 100  $\mu$ M  $\beta$ -mercaptoethanol, 2 mM L-glutamine, 100 mM nonessential amino acids (GIBCO/BRL), and 1,000 units/mL leukemia inhibitory factor (Chemicon). Culture medium was changed daily. For differentiation, cells were permitted to attach to plastic culture plates for 1 h to eliminate MEFs. Nonadherent cells were harvested and plated at a density of  $1 \times 10^4$ /cm<sup>2</sup> onto fresh gelatin-coated culture plates in M15 medium. 24 h later, the culture medium was changed to a 1:1 mixture of serum-free expansion medium (SFEM) and Iscove's modified Dulbecco's medium (IMDM) (Invitrogen) supplemented with Wnt (25 ng/mL final concentration) and Activin A (50 ng/mL final concentration) (R&D) and maintained for 4 days. This was followed by 2 days culture in SFEM/IMDM, Activin A (50 ng/mL) and 0.2% FBS. For FGF and Noggin (FN) and FGF, Noggin and retinoic acid (FNRA) differentiation, cells were placed in media containing SFEM/IMDM, 2% FBS, FGF1 (5 ng/mL), FGF4 (25 ng/mL), and Noggin (25 ng/mL) (R&D), with or without 0.5  $\mu$ M all trans-RA (Sigma) and maintained for 3 days.

Construction of human embryonic stem cell line hS17 is described elsewhere [16]. hS17 cells were grown on irradiated CF1 mouse embryonic fibroblast cells in DMEM/F12 supplemented with 20% (vol/vol) knockout serum replacement,  $\beta$ FGF (8 ng/mL, Peprotech), 3 mM L-glutamine, 0.1 mM nonessential amino acids, and 0.1 mM  $\beta$ -mercaptoethanol (Invitrogen). To differentiate human ES (hES) with Wnt and activin (WA)-treatment, 90% confluent hES cells were cultured in RPMI medium (Invitrogen) with 25 ng/mL Wnt3a [Gift from R. Nusse lab, Stanford University and Howard Hughes Medical Institute (HHMI)] and 100 ng/mL Activin A (R&D Systems) for 1 day. The medium was then changed to RPMI with 0.2% FBS and 100 ng/mL Activin A for 2 days. To further differentiate the cells with FN or FNRA media, the cell medium was changed to RPMI containing 2% FBS, 5 ng/mL human FGF1, 25 ng/mL human FGF4 (R&D Systems), with (FNRA) or without (FN) 0.5  $\mu$ M RA (Sigma) for 3 days. For induction of *Liv2*<sup>+</sup> cells, the media was changed to DMEM containing 1% B27 supplement (Invitrogen), 50 ng/mL FGF10 (R&D), 0.25  $\mu$ M

KAAD-cyclopamine (Toronto Research Chemicals), and 2  $\mu$ M all trans-RA (Sigma) for 3 days, before harvest of cells.

### Immunostaining and western blot analysis

Mouse E7.5 and E8.25 embryos were dissected, fixed with 4% paraformaldehyde in PBS for 1 h at 4°C, equilibrated with 30% sucrose in PBS, and cryo-embedded. Sections were stained with antibodies against Sox17 (AF1924, R&D; gift from Dr. James Wells, University of Cincinnati) and eGFP (A-11122, Molecular Probes). Sox17 was visualized using a TSA kit (Molecular Probes). Images were obtained by using a Zeiss confocal laser scanning microscope.

For western blots, sorted cells were lysed in standard sodium dodecyl sulfate (SDS) buffer; proteins were resolved on SDS-polyacrylamide gel electrophoresis and transferred to polyvinylidene fluoride membranes (Amersham Pharmacia) for immunoblotting with specific antibodies, including rabbit monoclonal anti-phospho-p44/42 MAPK (Erk1/2) (1:1,000; Cell Signaling), Goat anti-human SOX17 (1:1,000; R&D), and mouse monoclonal anti- $\beta$ -actin (1:4,000; Sigma). Signals were visualized using electrochemiluminescence detection (Amersham Pharmacia) on Kodak film after further incubation with horseradish peroxidase-conjugated secondary antibodies.

### Whole mount *in situ* hybridization

Embryos obtained from intercrosses were dissected in Dulbecco's Phosphate Buffered Saline (Invitrogen) with 0.1% Fetal Bovine Serum (Invitrogen). Staging of embryos was carried out as previously described [17]. Embryos were fixed overnight at 4°C in 4% paraformaldehyde, dehydrated through a graded methanol series, and stored at -20°C. For preparation of probes, PCR products were generated with primers described in Supplementary Table S1, and subcloned into the pCRII-TOPO-Blunt vector (Invitrogen). For synthesis of digoxigenin-UTP labeled mRNA probes, 10  $\mu$ g of template DNA was linearized with 10 Units of an appropriate restriction enzyme and purified. RNA probes were transcribed *in vitro* using 1  $\mu$ g of linearized template and 20 Units of SP6, T3, or T7 RNA polymerase (Roche). RNA probes were purified using G-50 Sephadex columns (Amersham) and stored at -80°C. Whole mount *in situ* hybridization was performed as described in [18]. Embryos were incubated in BM purple alkaline phosphatase substrate (Roche) at 4°C or room temperature. Staining reactions were stopped by washes in PBS with 0.1% Tween-20. Subsequently, embryos were processed through a graded PBT:glycerol series into 80% glycerol. Photomicrographs were obtained with an MZ 16FA microscope (Leica), and a Micropublisher 3.3 RTV camera with Q-Capture Pro software (Q-imaging).

### RNA isolation and quantitative PCR

FACS sorted cells were lysed for extraction of RNA using the Picopure RNA isolation kit (Arcturus Molecular Devices). About 0.5  $\mu$ g of total RNA was used for reverse transcription with a cDNA synthesis kit (Ambion Applied Biosystems). PCR reactions were run in duplicate using 1/20th to 1/60th of the cDNA per reaction using Taqman gene expression assays (Applied Biosystems) in which  $\beta$ -actin was

used as an endogenous control (See Supplementary Table S1 for a list of PCR primers and assay conditions used in this study).

### Microarray analysis

Total mRNA from FACS sorted cells was purified using the Picopure RNA extract kit (Arcturus). cDNA was synthesized and amplified using the Ovation RNA amplification system V2 (NuGEN). About 3.75  $\mu$ g of cDNA was fragmented and biotin labeled using the FL-Ovation cDNA biotin module V2 (NuGEN). Labeled cDNAs were hybridized to Affymetrix Mouse Genome 430\_2 microarrays or Affymetrix Human Genome U133 Plus 2.0 arrays. Microarray data was normalized and summarized by Robust Multi-chip Averaging (RMA) algorithm using GeneSpring GX 10.0 (Agilent). Baseline transformations of the data for the various analyses were carried out by either of the 2 methods: (1) Baseline to median of all samples (Figs. 2C, 3A, and 4A; Fig. 7; Supplementary Figs. S2, S3B); or (2) Baseline to median of mES samples (Fig. 5A; Supplementary Fig. S3A). For statistical analysis, one-way analysis of variance (ANOVA) tests with Bonferroni FWER < 0.05 was used. Signal intensities of probe sets mapping to the same gene were averaged. Multi-class significance analysis of microarrays (SAM) was used to identify the genes to define the centroids. To classify samples, Pearson correlation coefficients of each sample were calculated for each centroid.

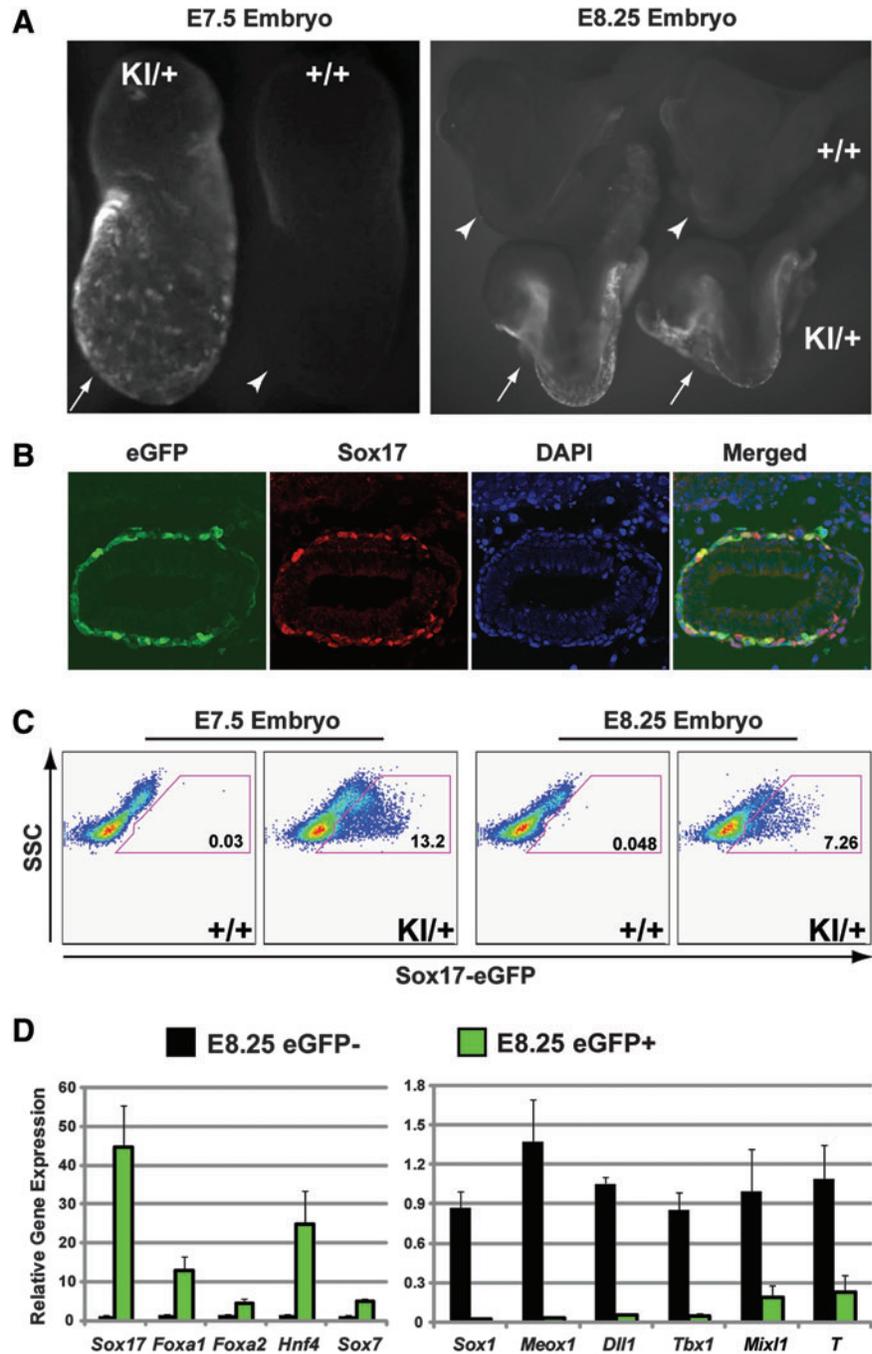
## Results

### Isolation of endoderm from Sox17-eGFP mouse embryos

We used a homologous recombination in mouse ES cells to target an eGFP-encoding transgene into the mouse Sox17 locus (Supplementary Fig. S1A). Sox17-eGFP ES cells were used to generate mice (see Experimental Procedures; Supplementary Fig. S1A). In Sox17-eGFP mice at embryonic day (E) 7.5, eGFP expression was found in the external cell layer of the cup-shaped embryo (Fig. 1A), which is comprised of visceral and DE [19,20]. At E8.25, Sox17-eGFP<sup>+</sup> cells mark and comprise DE in the tubular foregut and hindgut, and the open midgut (Fig. 1A). To verify that expression of the Sox17-eGFP transgene reported endogenous Sox17 expression, we performed immunostaining which revealed that eGFP<sup>+</sup> cells expressed Sox17 protein at E7.5 (Fig. 1B) and E8.25 (data not shown). Sox17-eGFP expression in GI and respiratory organs was decreased thereafter, consistent with previous studies that revealed reduced Sox17 mRNA and protein levels in embryonic mice after E8.5 [21,22]. Collectively, these results indicate that the Sox17-eGFP transgene accurately reported Sox17 expression.

We used flow cytometry to isolate Sox17-eGFP<sup>+</sup> cells at E7.5 and E8.25 after proteolytic dispersion of whole embryos into single-cell suspensions. Consistent with our immunohistologic studies, flow cytometry revealed a distinct population of Sox17-eGFP<sup>+</sup> cells comprising 12.5%  $\pm$  2.5% of total cells at E7.5 and 6.9%  $\pm$  1.1% at E8.25 (Fig. 1C). Quantitative PCR (Q-PCR) revealed enrichment of mRNA encoding the endoderm markers Sox17, FoxA1, FoxA2, HNF4 $\alpha$ , and Sox7 in Sox17-eGFP<sup>+</sup> cells (Fig. 1D). In contrast, mRNAs encoding markers of mesoderm and neuroectoderm, such as

**FIG. 1.** Isolation of endoderm from *Sox17-eGFP* transgenic mouse embryos. **(A)** Whole mount fluorescence of E7.5 and E8.25 *Sox17-eGFP* (KI/+, white arrows) and wild-type (+/+, white arrowheads) embryos (lateral views with anterior to the left). eGFP<sup>+</sup> cells are observed in the outer endoderm layer at E7.5 and throughout the gut tube at E8.25. **(B)** Immunofluorescence images of E7.5 *Sox17-eGFP* embryo sections demonstrating colocalization of eGFP and Sox17 within the endoderm. **(C)** Flow cytometric analyses of dissociated E7.5 and E8.25 *Sox17-eGFP* embryos. eGFP<sup>+</sup> cells comprise 13% and 7% of the E7.5 and E8.25 embryo, respectively. **(D)** Q-PCR analysis of eGFP<sup>-</sup> and eGFP<sup>+</sup> cells isolated by FACS from E8.25 *Sox17-eGFP* embryos. Expression of endoderm markers (*Sox17*, *Foxa1*, *Foxa2*, *Hnf4*, and *Sox7*) are enriched in the eGFP<sup>+</sup> fraction, while ectoderm, mesoderm, and mesoderm markers (*Sox1*, *Meox1*, *Dll1*, *Tbx6*, *Mixl1*, and *T*) are enriched in the eGFP<sup>-</sup> fraction (mean ± SD, *n* = 3). See also Supplementary Fig. S1. Q-PCR, quantitative polymerase chain reaction; FACS, fluorescence-activated cell sorting; eGFP, enhanced green fluorescent protein. Color images available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)

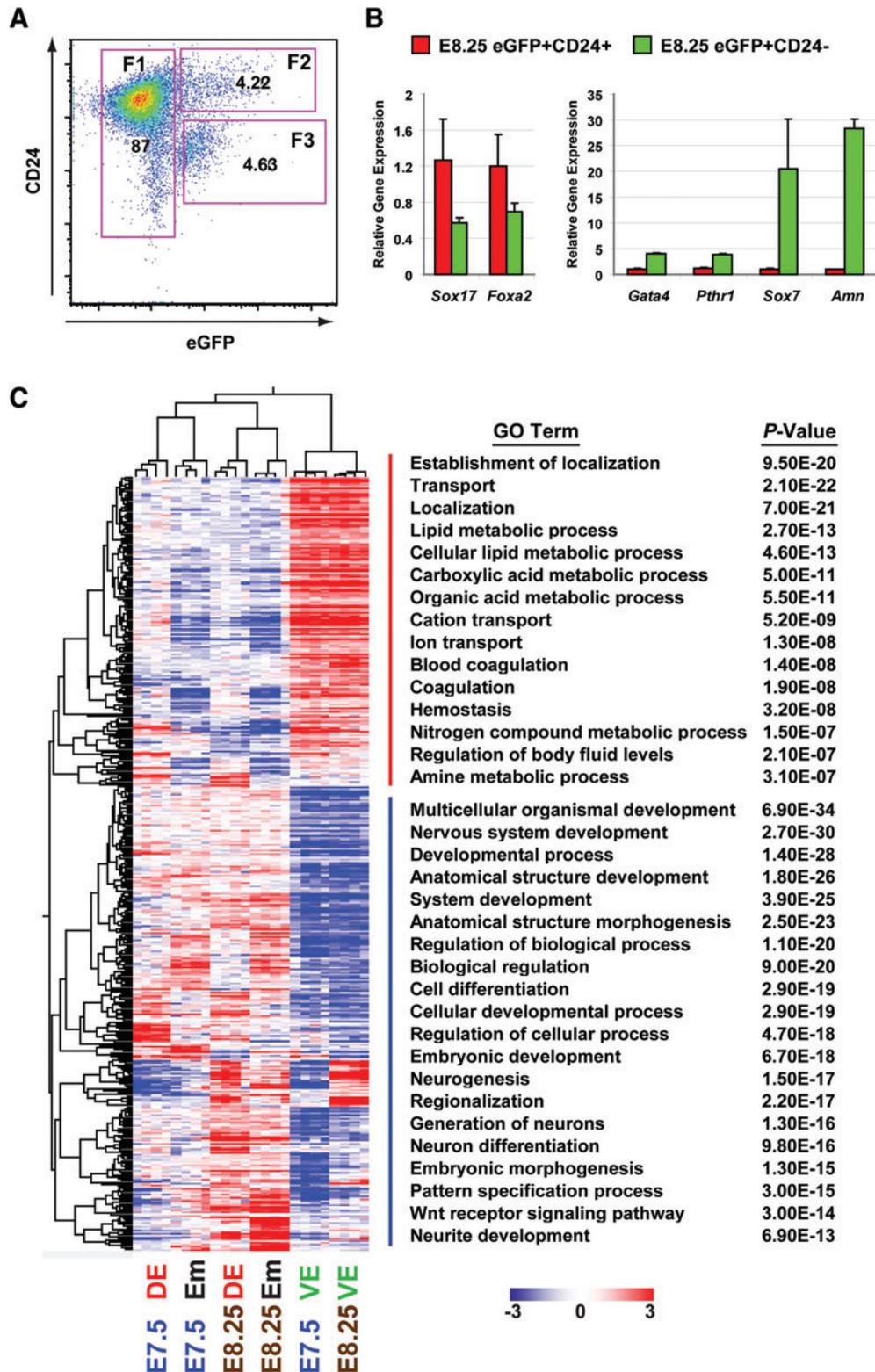


*Sox1*, *Meox1*, *Dll1*, and *Tbx6* [23,24], were depleted from *Sox17-eGFP*<sup>+</sup> cells, but enriched in the eGFP<sup>-</sup> cell subset (Fig. 1D). mRNAs encoding the mesendodermal markers *Brachyury* (*T*) [25] and *Mixl1* [26] were expressed at low levels in endoderm at E8.25 (Fig. 1D). These results suggest that FACS of *Sox17-eGFP*<sup>+</sup> cells permitted separation and isolation of endoderm from mesoderm and ectoderm at E7.5 and E8.25.

#### Separation of visceral and DE in *Sox17-eGFP*<sup>+</sup> cells

Previous studies have reported that *Sox17* is expressed in both DE and VE [13]. To identify methods for separating DE and VE in the *Sox17-eGFP*<sup>+</sup> population, we initially used an antibody that recognized epithelial cell adhesion molecule

(EpCAM; Supplemental Fig. 1D), a marker that permits isolation of endoderm at E8.25 and later stages, but not at E7.5 [3]. A subset of *Sox17-GFP*<sup>+</sup> cells from manually isolated E8.25 extraembryonic tissue was EpCAM negative using flow cytometry (data not shown). We sorted *Sox17-eGFP*<sup>+</sup>EpCAM<sup>+</sup> cells and *Sox17-eGFP*<sup>+</sup>EpCAM<sup>-</sup> cells at E8.25 and performed a microarray analysis comparing the gene expression profiles of these 2 populations. We identified mRNA encoding CD24 (also known as heat-stable antigen) [27,28] as a marker enriched in *Sox17-eGFP*<sup>+</sup>EpCAM<sup>+</sup> cells at E8.25. Subsequently, flow cytometry analysis revealed that *Sox17-eGFP*<sup>+</sup> cells at E7.5 and E8.25 were separated by CD24 sorting into 2 populations, *Sox17-eGFP*<sup>+</sup> CD24<sup>+</sup> and *Sox17-eGFP*<sup>+</sup> CD24<sup>lo/-</sup> (Fig. 2A). Microdissection to separate



**FIG. 2.** CD24 expression separates Sox17-eGFP+ endoderm to definitive and visceral population. **(A)** FACS analysis of CD24 and eGFP expression in E8.25 *Sox17-eGFP* embryos. CD24 separates eGFP+ cells into 2 fractions—eGFP+CD24<sup>hi</sup> (Fraction 2) and eGFP+CD24<sup>lo/-</sup> (Fraction 3). **(B)** Q-PCR analysis of eGFP+CD24<sup>hi</sup> and eGFP+CD24<sup>lo/-</sup> fractions. The endoderm markers *Sox17* and *Foxa2* are expressed in both eGFP+CD24<sup>hi</sup> and eGFP+CD24<sup>lo/-</sup> fractions; however, their expression is enriched in the eGFP+CD24<sup>hi</sup> fraction. Visceral endoderm specific markers (*Gata4*, *Pthr1*, *Sox7*, and *Amn*) are enriched in the eGFP+CD24<sup>lo/-</sup> fraction (mean±SD, n=3). **(C)** Hierarchical clustering and GO term analysis of 3870 probe sets differentially expressed between the definitive endoderm (DE), visceral endoderm (VE), and nonendodermal tissues (Em) at E7.5 and E8.5. Color scale indicates normalized expression values. GO, gene ontology. Color images available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)

embryonic from extraembryonic tissues followed by flow cytometry analysis revealed Sox17-eGFP<sup>+</sup> CD24<sup>+</sup> cells were from embryonic tissues, while Sox17-eGFP<sup>+</sup> CD24<sup>lo/-</sup> cells were from extraembryonic tissues. As predicted by the analysis of manually separated extra- and embryonic tissue, Q-PCR analysis demonstrated that both Sox17-eGFP<sup>+</sup> CD24<sup>+</sup> and Sox17-eGFP<sup>+</sup> CD24<sup>lo/-</sup> cells expressed endoderm markers, including *Sox17* and *FoxA2*. However, expression of established VE markers, including *Gata4*, *Pthr1*, *Sox7*, and *Amn*, was enriched in Sox17-eGFP<sup>+</sup> CD24<sup>lo/-</sup> cells compared with Sox17-eGFP<sup>+</sup> CD24<sup>+</sup> cells (Fig. 2B). These data suggested that DE was enriched in the Sox17-eGFP<sup>+</sup> CD24<sup>+</sup> fraction, while VE was enriched in the Sox17-eGFP<sup>+</sup> CD24<sup>lo/-</sup> fraction. Thus, we concluded that CD24 combined with Sox17-eGFP permitted isolation and separation of native DE and VE from tissues comprising the rest of the embryo.

### Analysis of visceral and DE gene expression

To obtain genomic-scale expression profiles of definitive and VE, we sorted Sox17-eGFP<sup>+</sup> CD24<sup>+</sup> (F2, DE) and Sox17-eGFP<sup>+</sup> CD24<sup>lo/-</sup> (F3, VE) cells, as well as the nonendodermal cells (F1, Em, Fig. 2A) from E7.5 and E8.25 embryos using flow cytometry (see Methods). We identified 3870 probe sets differentially expressed between DE, VE, and nonendodermal tissues (E7.5 DE, E7.5 VE, E7.5 Em, E8.25 DE, E8.25 VE and E8.25 Em) using one-way ANOVA (Bonferroni FWER < 0.05). Hierarchical gene clustering of these differentially expressed probe sets revealed that VE at E7.5 and E8.25 clustered together and was distinct from the DE and Em populations (Fig. 2C). In contrast, the DE from these 2 stages did not cluster with each other; instead, E7.5 DE clustered closer to the E7.5 nonendodermal tissue signature, and E8.25 DE clustered to the E8.25 nonendodermal tissue signature (Fig. 2C), suggesting that many genes expressed in embryonic tissues are co-regulated during and after gastrulation. Gene ontology (GO) term analysis of the gene clusters is highly expressed in VE unveiled enrichment of GO terms, including transport and metabolic processes, which are consistent with established roles for VE and its derivatives in nutrient, gas, and waste metabolite exchange, and embryonic hematopoiesis. In contrast, clusters of genes encoding known developmental regulators were expressed at low levels in VE and high levels in embryonic tissues (Fig. 2C), reflecting the rapid growth and differentiation of mid-gestational embryos [10,29].

To identify genes differentially expressed between definitive and VE at E7.5 and E8.25, we analyzed our microarray data using significance SAM (False Discovery Rate, FDR=0). We identified several classes of genes whose expression was enriched in (1) DE at E7.5 or E8.25; (2) DE at both E7.5 and E8.25; (3) both definitive and VE and (4) VE only (see Supplementary Table S2). Expression of established markers of mouse DE [2,3,18,30,31] was enriched in CD24<sup>+</sup> Sox17-eGFP<sup>+</sup> cells, demonstrating the strength of our FACS purification scheme. For example, we detected enriched expression of *FoxA1*, *Trh*, *Cldn4*, *Cldn8*, *Hhex*, and *Kitl* in CD24<sup>+</sup> Sox17-eGFP<sup>+</sup> cells (Fig. 3A). Likewise, we detected enriched expression of known markers of VE in CD24<sup>lo/-</sup> Sox17-eGFP<sup>+</sup> cells, including *Sox7*, *Amn*, *Afp*, *Sox7*, *Cited1*, *Dab2*, *IHH*, and *Hnf4x* (Fig. 3A; [2,3,13]). We also identified genes whose relative expression distinguished DE between E7.5 and E8.25 (*Cer1*, *Gsc*, *Pax9*, *Shh*, *Nepn*,

*Pyy*, and *Mnx1* in Fig. 3A). Thus, our microarray data of purified DE and VE reflected the expression pattern of known endoderm markers.

To determine whether our data revealed new markers of DE, we selected a subset of genes (*Sorcs2*, *Krt19*, *Gfpt2*, *Eppk1*, *Nedd9*, *Vtn*, *Plat*, and *9130213B05Rik*) enriched in FACS-isolated E7.5 and/or E8.25 cells, and used in situ hybridization to assess their expression in early embryos. At E8.25, of the 8 genes analyzed by in situ hybridization, 5 (*Sorcs2*, *Krt19*, *Gfpt2*, *Eppk1*, and *Nedd9*) were detected in the DE. While *Vtn* expression was detected in posterior notochord, we did not consistently detect *Vtn* expression in DE by in situ hybridization due to low levels of expression (Fig. 3B). Expression of *Eppk1*, *Gfpt2*, *Sorcs2*, *Nedd9*, and *Krt19* was found localized to a subset of fore-, mid-, and/or hindgut endoderm at E8.25 (Fig. 3B). We assessed *Gfpt2* and *Krt19* expression at E7.5 and found that both were also expressed in DE (Supplementary Fig. S2D). Two other genes (*Plat*, *9130213B05Rik*) were not detected by in situ hybridization at E7.5 or E8.25, consistent with our inability to detect mRNA encoded by these loci with RT-PCR at these stages (data not shown). These findings support the ability of CD24 to separate Sox17-eGFP<sup>+</sup> population into DE and VE, and demonstrate that molecular profiles generated here reflect gene expression in DE and VE.

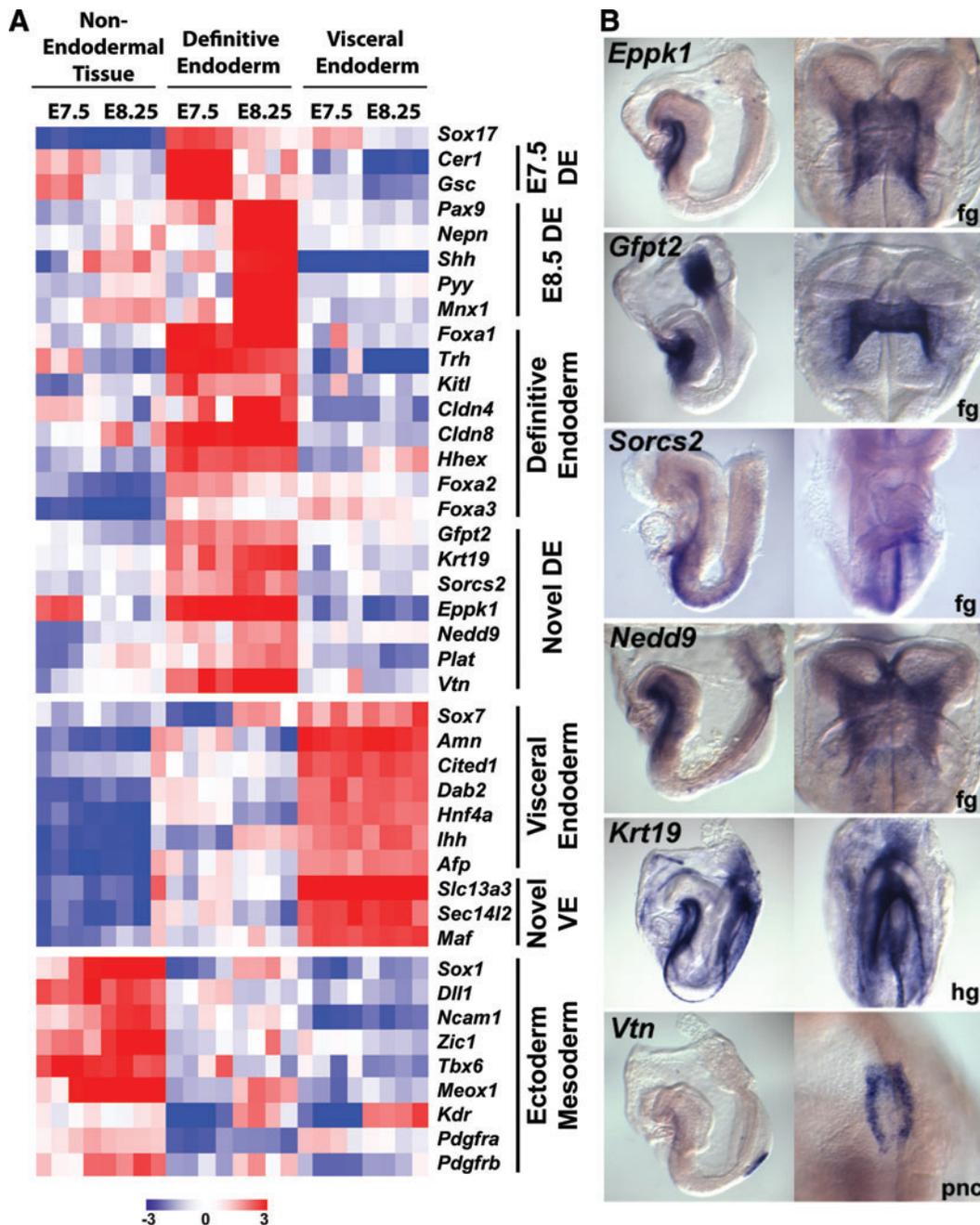
### Identification of new markers for purifying mouse definitive and VE

We used our expression analysis to identify surface epitopes that would permit FACS isolation of DE and VE without reliance on *Sox17-eGFP* marking. We identified surface proteins for which antibodies were available in our microarray data of CD24<sup>+</sup> Sox17-eGFP<sup>+</sup> DE and CD24<sup>lo/-</sup> Sox17-eGFP<sup>+</sup> VE (Fig. 4A). Supporting our findings from flow cytometry, the expression of *CD24* mRNA was enriched in DE and in total embryonic cells, but low in VE (Fig. 4A). In VE, we observed enriched expression of *CD38*, consistent with previous reports [3,32]. Our analysis showed that expression of *CD55*, encoding the transmembrane glycoprotein known as complement decay accelerating factor (DAF; [32]), was enriched in both DE and VE compared with total cells (Fig. 4A). We next used flow cytometry to test whether combinations of antibodies recognizing CD24, CD38, and CD55 would permit isolation of DE and VE from genetically unmodified mice. First, using manual dissection, we separated Sox17-eGFP<sup>+</sup> DE (marked red, Fig. 4B) from VE (marked green, Fig. 4B). Consistent with our microarray data, EpCAM and CD55 antibodies labeled both Sox17-eGFP<sup>+</sup> DE and VE (Fig. 4B). By contrast, only Sox17-eGFP<sup>+</sup> VE cells were CD38<sup>+</sup> and CD24<sup>lo/-</sup>, while Sox17-eGFP<sup>+</sup> DE cells were CD38<sup>-</sup> and CD24<sup>+</sup> (Fig. 4B). *Cxcr4* is expressed in ESC-derived endoderm [8]; so, we examined *Cxcr4* expression in mouse embryos. The expression of *Cxcr4* was high in DE compared with VE (Fig. 4A), but due to auto-fluorescence of VE cells, we were unable to separate DE from VE with flow cytometry using the *Cxcr4* antibody (Fig. 4B). Collectively, our analysis confirmed that EpCAM, CD24, CD38, and CD55 can be used to separate DE and VE from the rest of the embryo.

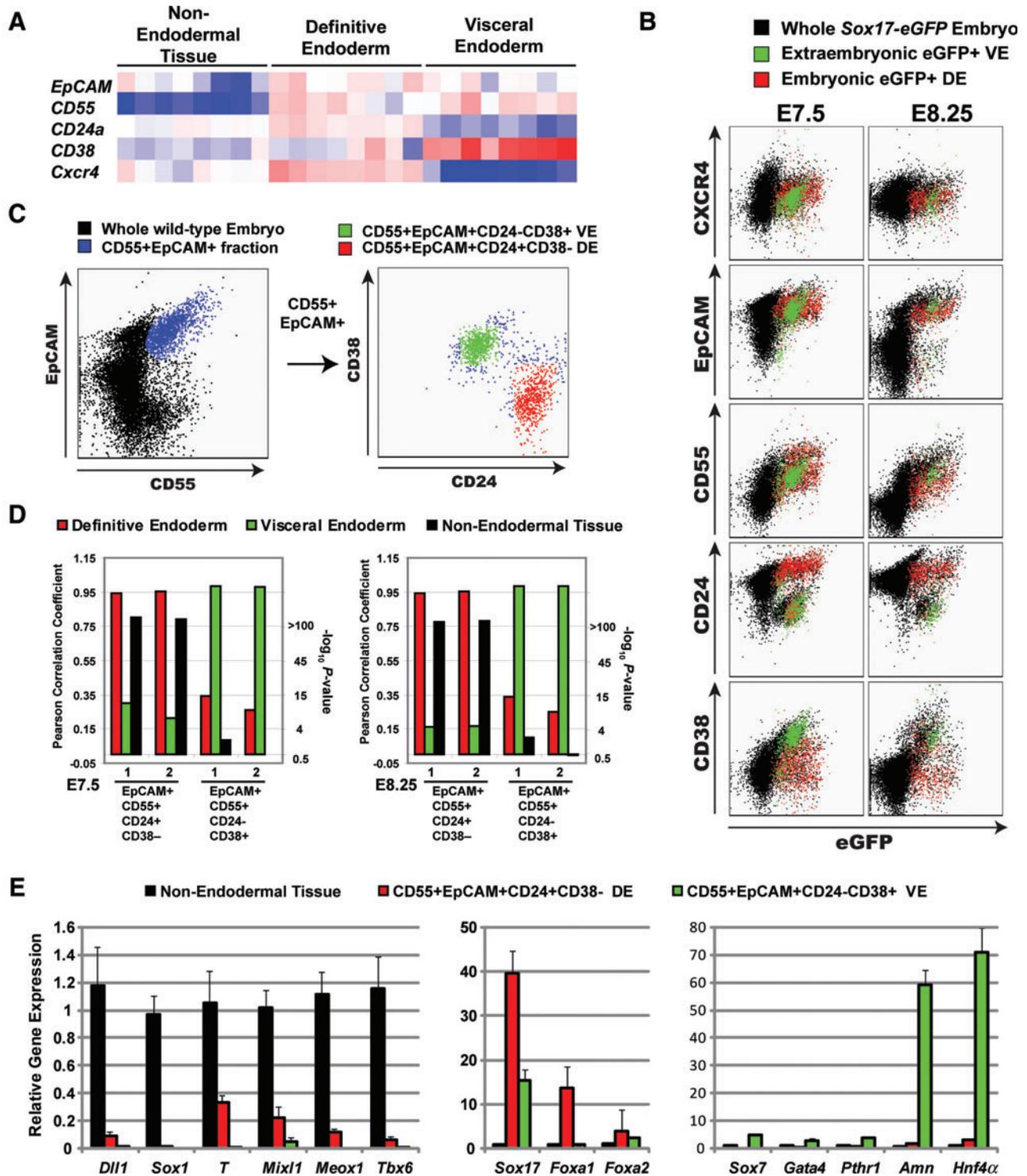
Based on these data, we hypothesized that the EpCAM<sup>+</sup> CD55<sup>+</sup> CD24<sup>+</sup> CD38<sup>-</sup> fraction represents DE, while the EpCAM<sup>+</sup> CD55<sup>+</sup> CD24<sup>lo/-</sup> CD38<sup>+</sup> fraction represents VE. To

test this hypothesis, we assessed whether combining antibodies against these cell surface markers allowed FACS isolation of DE and VE from wild-type mouse embryos at E7.5 and E8.25. As shown in Fig. 4C, we sorted EpCAM<sup>+</sup> CD55<sup>+</sup> CD24<sup>+</sup> CD38<sup>-</sup> cells (red) and EpCAM<sup>+</sup> CD55<sup>+</sup> CD24<sup>lo/-</sup> CD38<sup>+</sup> cells (green). To identify sorted cells, we performed microarray analysis of gene expression in these purified populations, and compared these data to gene expression in control CD24<sup>+</sup> Sox17-eGFP<sup>+</sup> and CD24<sup>lo/-</sup> Sox17-eGFP<sup>+</sup>, which represent DE and VE, respectively. To

assess the similarity of gene expression profiles, we performed Pearson correlation analysis to centroids for DE, VE, and nonendodermal tissue (see Methods, and Supplementary Table S3). As shown in Fig. 4D, EpCAM<sup>+</sup> CD55<sup>+</sup> CD24<sup>+</sup> CD38<sup>-</sup> cells were highly correlated with CD24<sup>+</sup> Sox17-eGFP<sup>+</sup> DE cells (average Pearson=0.95, P=0), and EpCAM<sup>+</sup> CD55<sup>+</sup> CD24<sup>lo/-</sup> CD38<sup>+</sup> cells with CD24<sup>lo/-</sup> Sox17-eGFP<sup>+</sup> VE (average Pearson=0.98, P=0) (Fig. 4D). To confirm these microarray results, we used Q-PCR to analyze the expression of marker genes in cells at E8.25 isolated by FACS with the



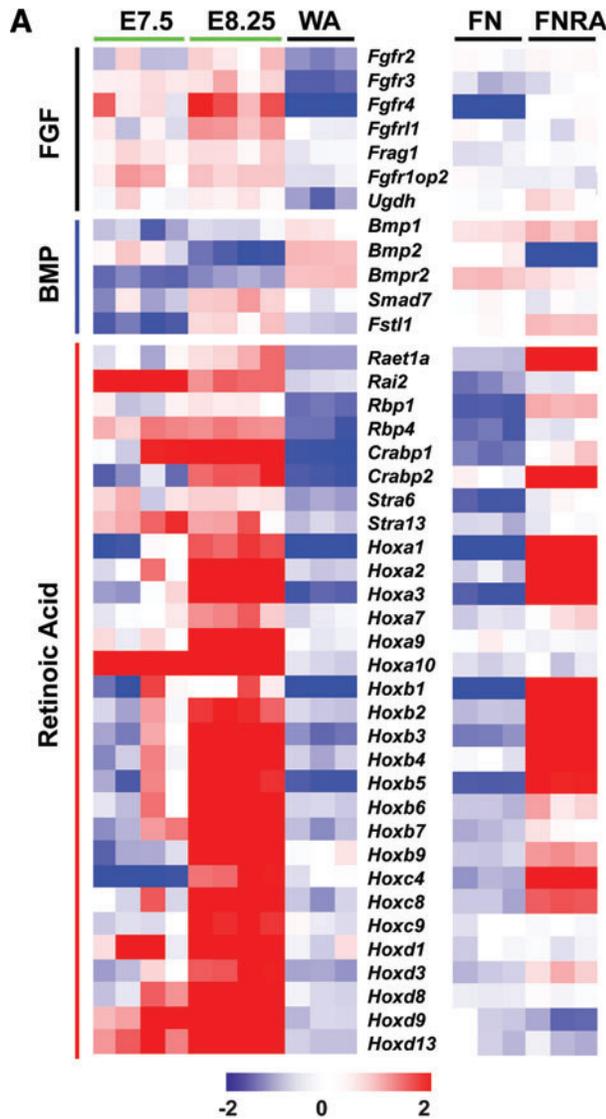
**FIG. 3.** Microarray and in situ hybridization analysis of endoderm markers in DE and VE isolated with Sox17-eGFP and CD24. **(A)** Expression levels of established and novel endoderm markers in Sox17-eGFP<sup>+</sup>CD24<sup>+</sup> DE and Sox17-eGFP<sup>+</sup>CD24<sup>-</sup> VE at E7.5 and E8.25. Color scale indicates normalized expression values. **(B)** Whole-mount in situ hybridization validation of the predicted novel definitive endoderm markers *Eppk1*, *Gfpt2*, *Sorcs2*, *Nedd9*, *Krt19*, and *Vtn*. Fg, foregut; hg, hindgut; pnc, posterior notochord. See also Supplementary Fig. S2 and Supplementary Table S2. Color images available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)



**FIG. 4.** Identification of a cell surface marker profile of native mouse definitive and visceral endoderm. **(A)** RNA expression levels of *EpCAM*, *CD55*, *CD24*, *CD38*, and *CXCR4* in endoderm and nonendodermal tissues isolated using *Sox17-GFP* and *CD24*. **(B)** FACS analysis of manually dissected E7.5 and E8.25 *Sox17-eGFP*<sup>+</sup> embryos. Embryonic *eGFP*<sup>+</sup> definitive endoderm (*red*) is *EpCAM*<sup>+</sup>*CD55*<sup>+</sup>*CD24*<sup>+</sup>*CD38*<sup>-</sup> and extraembryonic *eGFP*<sup>+</sup> visceral endoderm (*green*) is *EpCAM*<sup>+</sup>*CD55*<sup>+</sup>*CD24*<sup>lo/-</sup>*CD38*<sup>+</sup>. **(C)** FACS isolation of definitive and visceral endoderm from E8.25 wild-type embryos using *EpCAM*, *CD55*, *CD24*, and *CD38*. **(D)** Pearson correlation coefficient analysis showed that gene expression of DE and VE isolated from wild-type embryos by surface marker expression correlated highly with the centroids of DE and VE from *Sox17-eGFP*<sup>+</sup> embryos respectively. Note that the statistical significance of the degree of matching in gene expression signatures, as measured by Pearson correlations (*left* Y-axis) scales exponentially (*right* Y-axis, *P* values by *t*-test). **(E)** Q-PCR analysis of *EpCAM*<sup>+</sup>*CD55*<sup>+</sup>*CD24*<sup>+</sup>*CD38*<sup>-</sup> (DE, *red*), *EpCAM*<sup>+</sup>*CD55*<sup>+</sup>*CD24*<sup>lo/-</sup>*CD38*<sup>+</sup> (VE, *green*) and non-*EpCAM*<sup>+</sup>*CD55*<sup>+</sup> (nonendoderm tissue, *black*) cells isolated by FACS from E8.25 wild-type embryos. Expression of endoderm markers (*Sox17*, *Foxa1*, and *Foxa2*) were enriched in the DE and VE fractions, while ectoderm, mesoderm, and mesendoderm markers (*Sox1*, *Meox1*, *Dll1*, *Tbx6*, *Mixl1*, and *T*) were enriched in the nonendoderm tissue fraction. Visceral endoderm specific markers (*Gata4*, *Pthr1*, *Sox7*, *Amn*, and *HNF4 $\alpha$* ) were enriched in the *EpCAM*<sup>+</sup>*CD55*<sup>+</sup>*CD24*<sup>lo/-</sup>*CD38*<sup>+</sup> (VE, *green*) fraction (mean  $\pm$  SD, *n* = 3). See also Supplementary Table S3. *EpCAM*, epithelial cell adhesion molecule. Color images available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)

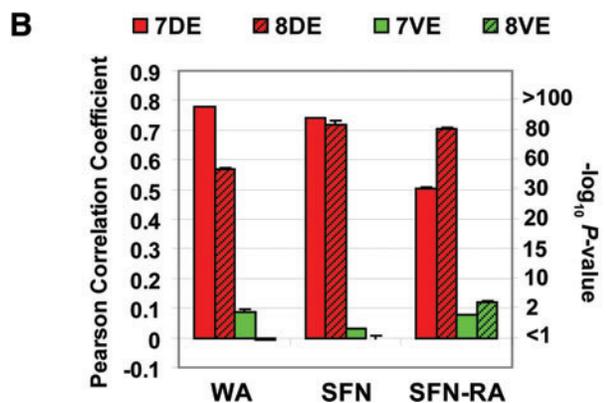
surface markers EpCAM, CD55, CD24, and CD38. The expression of *Dll1*, *Sox1*, *T*, *Mixl1*, *Meox1*, and *Tbx6* was high in nonendodermal cells (Fig. 4E). This expression pattern was similar to that measured for purified Sox17-eGFP<sup>-</sup> cells at E8.25 (Fig. 1D). The expression of *Sox17* and *FoxA2*, markers common to both native DE and VE, was high in both EpCAM<sup>+</sup> CD55<sup>+</sup> CD24<sup>+</sup> CD38<sup>-</sup> DE and EpCAM<sup>+</sup> CD55<sup>+</sup>

CD24<sup>lo/-</sup> CD38<sup>+</sup> VE cell populations. However, the expression of the VE markers *Gata4*, *Pthr1*, *Sox7*, *Amn*, and *HNF4α* was enriched in EpCAM<sup>+</sup> CD55<sup>+</sup> CD24<sup>lo/-</sup> CD38<sup>+</sup> VE cells compared to EpCAM<sup>+</sup> CD55<sup>+</sup> CD24<sup>+</sup> CD38<sup>-</sup> DE. Collectively, these data are in good agreement with our expression data from cells sorted with Sox17-eGFP and CD24 (Fig. 2B). Thus, our studies have identified unique methods for purifying definitive and VE at E7.5 and E8.25 from wild-type mice.

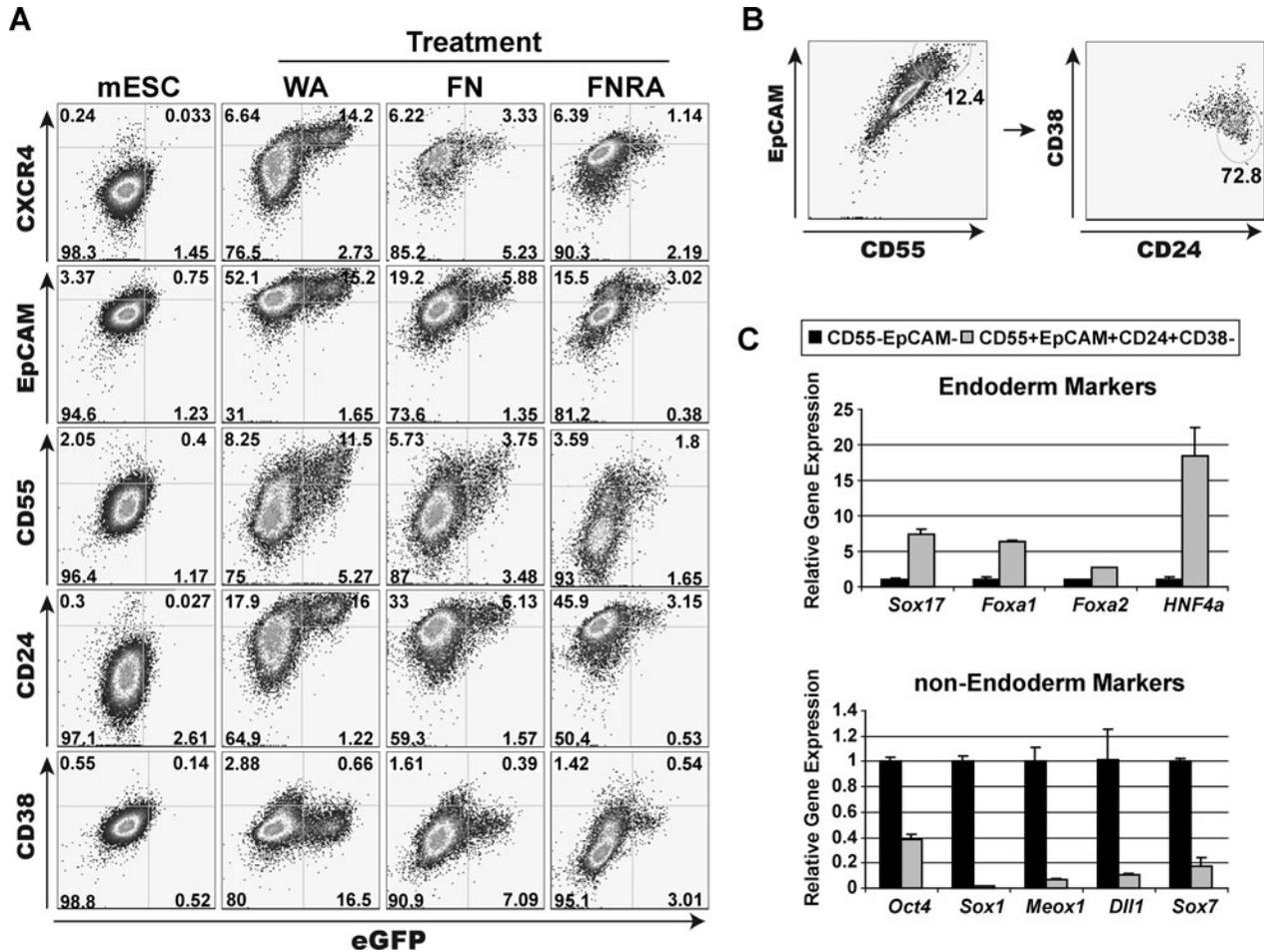


*Endoderm development from ES cells directed by RA, FGF, and BMP signaling*

Earlier, the identification of cells resembling endoderm produced in vitro from pluripotent cells was based on analysis of a subset of endodermal markers. Our identification of a detailed and comprehensive molecular signature for native mouse definitive and VE provided a unique tool to interrogate the gene expression profile of endoderm-like cells derived from ES cell cultures. Here, we used established methods [7,8] involving Wnt and Nodal signaling to produce endoderm-like cells from *Sox17-eGFP* knock-in mESCs for comparison to native DE. Consistent with its expression in native endoderm, *Sox17* expression was induced on differentiation of *Sox17-eGFP* mESC in both aggregation and monolayer cultures, as indicated by the expression of eGFP. *eGFP* expression was not observed in undifferentiated *Sox17-eGFP* ES cells (Supplementary Fig. S1C). Moreover, we observed co-expression of eGFP and *Sox17* protein detected by immunostaining (data not shown), indicating the *Sox17-eGFP* knock-in mESCs faithfully reported *Sox17* expression during in vitro differentiation. To assess the similarity of these ES-derived endoderm-like progeny to native DE, we performed microarray analysis of *Sox17-eGFP*<sup>+</sup> ES cell progeny from Wnt3a and Activin A treatment (hereafter, “WA”). We observed a gene expression profile of mESC progeny after WA treatment that was most similar to native E7.5 DE, with lesser similarity to E8.25 DE and to VE based on Pearson correlation coefficients to centroids (Supplementary Table S4) derived from the native endoderm (Fig. 5A,B). We hypothesized that differences in signaling pathway(s) activation might underlie the distinct gene expression profiles of native endoderm and endoderm-like cells from ES



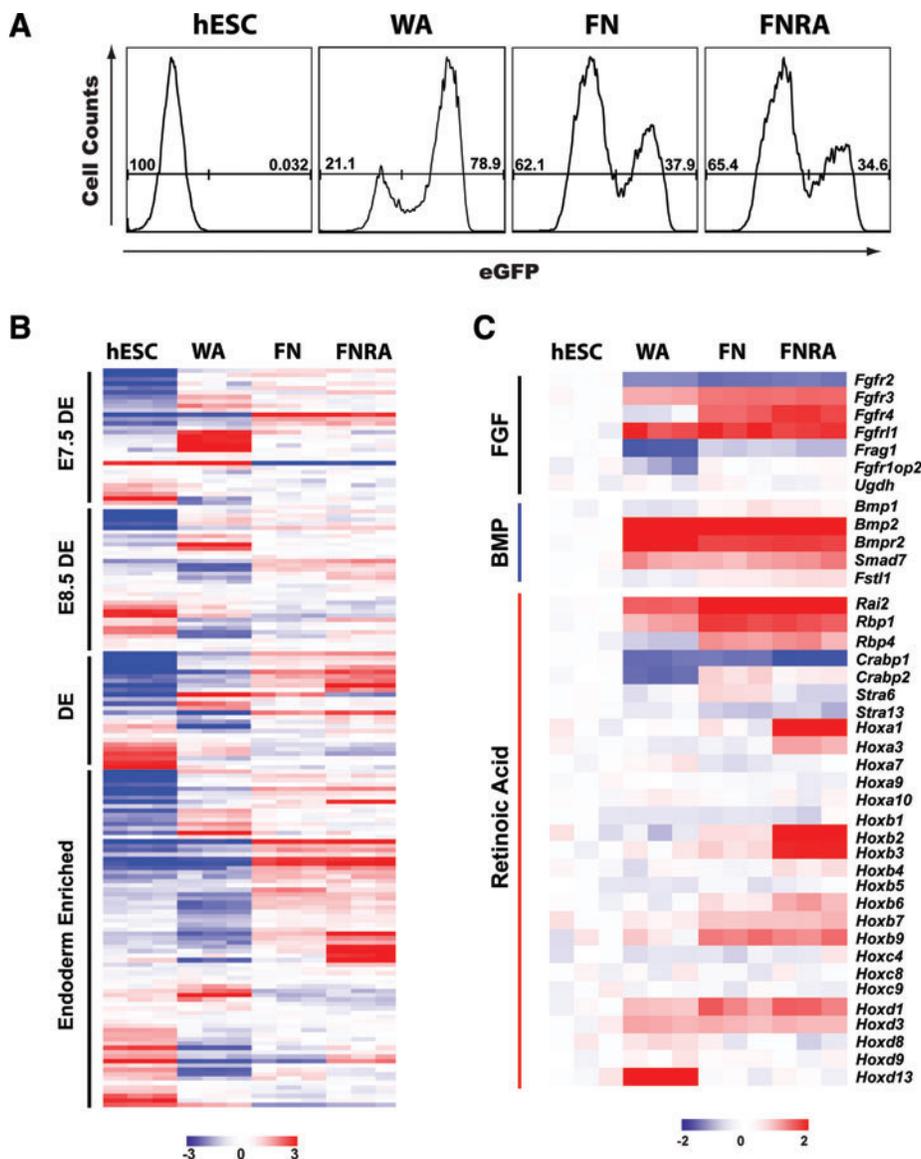
**FIG. 5.** Modification of mouse ES cell culture conditions based on patterns of signaling pathway activation in E8.25 DE produces cell progeny more closely resembling native definitive endoderm. **(A)** Expression profile of components and targets of the BMP, FGF, and RA signaling pathways in E7.5 and E8.25 native mouse definitive endoderm (*Sox17-eGFP*<sup>+</sup> CD24<sup>+</sup>) and mESC-derived endoderm differentiated after WA, FN, or FNRA treatment. Color scale indicates normalized expression values. **(B)** The centroid of WA-treated ESC-derived DE correlates highly with the centroid of DE of E7.5 *Sox17-eGFP*<sup>+</sup> embryos, the centroid of FN-treated ESC-derived DE correlates highly with the centroids of DE of both E7.5 and E8.25 *Sox17-eGFP*<sup>+</sup> embryos, and the centroid of FNRA-treated ESC-derived DE correlates highly with the centroid of DE of E8.25 *Sox17-eGFP*<sup>+</sup> embryos. (right axis, *P* values by *t*-test). See also Supplementary Fig. S3. ES, embryonic stem; BMP, bone morphogenetic protein; FGF fibroblast growth factor; RA, retinoic acid; FN, fibronectin; WA, Wnt and activin; FNRA, F6F, Noggin and retinoic acid. Color images available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)



**FIG. 6.** The cell surface marker profile of native definitive endoderm allows FACS isolation of mouse ES cell progeny resembling definitive endoderm. **(A)** Flow cytometric analysis of the cell surface markers CXCR4, EpCAM, CD55, CD24, and CD38 in undifferentiated *Sox17-eGFP* mouse embryonic stem cells, and *Sox17-eGFP* mESCs after differentiation with WA, FN, or FNRA. **(B)** Flow cytometric analysis of wild-type R1 mESCs using antibodies for EpCAM, CD55, CD24, and CD38 after FNRA treatment. **(C)** Q-PCR analysis of isolated CD55<sup>-</sup>EpCAM<sup>-</sup> and EpCAM<sup>+</sup>CD55<sup>+</sup>CD24<sup>+</sup>CD38<sup>-</sup> cells. Expression of endoderm markers (*Sox17*, *Foxa1*, *Foxa2*, and *Hnf4*) was enriched in the EpCAM<sup>+</sup>CD55<sup>+</sup>CD24<sup>+</sup>CD38<sup>-</sup> fraction, while expression of pluripotency marker (*Oct4*), ectoderm marker (*Sox1*), and mesoderm markers (*Meox1* and *Tbx6*) was enriched in the CD55<sup>-</sup>EpCAM<sup>-</sup> fraction (mean  $\pm$  SD,  $n=3$ ). See also Supplementary Table S4.

cultures. Statistical and hierarchical clustering analysis revealed that 3123 probe sets were differentially expressed between native DE and *Sox17-eGFP*<sup>+</sup> progeny derived from ES cell differentiation cultures (Supplementary Fig. S3A). Several signaling pathways such as BMP, FGF, and RA have been suggested as playing important roles in endoderm development [33,34]. Hence, we wanted to examine whether these pathways were properly regulated. Compared with native E8.25 DE, we found that *Sox17-eGFP*<sup>+</sup> progeny derived from WA treated cultures failed to induce components and targets of RA and FGF signaling (Fig. 5A). This analysis further suggested that BMP signaling activation in ES cell-derived *Sox17-eGFP*<sup>+</sup> progeny was higher than in native DE (Fig. 5A). To test whether stimulation of RA and FGF signaling, and inhibition of BMP signaling might refine the development of endoderm-like cells from ES cell cultures, we systematically added RA, FGF1 and FGF4, and Noggin (an extracellular inhibitor of BMP signaling) to WA-treated ES cells, and compared gene expression profiles with the native

endoderm signatures using microarrays. Exposure of WA cultures to FGF and Noggin ("FN") or to FGF, Noggin and RA ("FNRA") increased the Pearson correlation value of these *Sox17-eGFP*<sup>+</sup> progeny compared with native E8.25 DE (Fig. 5B). Phospho-ERK is an established downstream effector of FGF signaling, and consistent with our microarray analysis, we found that phospho-ERK levels were relatively increased in western blots of protein from purified E8.25 DE and FNRA cultures compared with WA-treated mES cells (Supplementary Fig. S3B). *Hox* genes, established targets of the RA signaling pathway, were expressed in E8.25 native DE, but not in WA- or FN-induced endoderm. RA supplementation in FNRA cultures led to increased *Hox* gene expression (Fig. 5A, "FNRA"). Thus, gene expression profiling led to specific modifications that fine-tuned the production of DE-like cells from mouse ES cells. These studies also revealed that ES cell cultures are competent to respond to endogenous signals regulating native endoderm development.



**FIG. 7.** Induction of hSOX17-eGFP<sup>+</sup> endoderm cells from human ES cells. **(A)** Flow cytometric analysis of SOX17-eGFP<sup>+</sup> human embryonic stem cells after treatment with WA, FN, and FNRA. 30%–40% of cells became eGFP<sup>+</sup> after differentiation. **(B)** Expression profile of genes enriched in E7.5 DE, E8.25 DE, DE, and endoderm (DE+VE) of Sox17-eGFP<sup>+</sup> mouse embryos in undifferentiated human embryonic stem cells and hESCs after differentiation with WA, FN, or FNRA. Treatment with FN or FNRA induces expression of genes enriched in E8.25 DE, DE, and endoderm. **(C)** Expression profile of components and targets of the BMP, FGF, and RA signaling pathways in hESC-derived endoderm differentiated after WA, FN, or FNRA treatment. Color scale indicates normalized expression values. See also Supplementary Fig. S4. Color images available online at [www.liebertonline.com/scd](http://www.liebertonline.com/scd)

*FACS isolation of endoderm-like cells derived from mouse ES cells*

To identify methods for purifying DE-like cells derived from ES cultures, we exposed the *Sox17-eGFP* mouse ES cell line to Wnt and Activin A to induce endoderm-like differentiation, then analyzed the induction of surface markers found on native mouse embryonic endoderm. Flow cytometry revealed that undifferentiated *Sox17-eGFP* ES cells were negative or low for EpCAM, CD55, CD24, or CD38 (Fig. 6A), but WA treatment led to differentiation of Sox17-eGFP<sup>+</sup> Cxcr4<sup>+</sup> cells, Sox17-eGFP<sup>+</sup> EpCAM<sup>+</sup> cells, Sox17-eGFP<sup>+</sup> CD55<sup>+</sup> cells, and Sox17-eGFP<sup>+</sup> CD24<sup>+</sup> cells (Fig. 6A). In contrast, all Sox17-eGFP<sup>+</sup> cells remained CD38<sup>-</sup>. The FN and FNRA conditions resulted in reduced Cxcr4 expression by Sox17-eGFP<sup>+</sup> cells. However, the expression of EpCAM, CD55, CD24, and CD38 of Sox17-eGFP<sup>+</sup> cells was similar in WA, FN, and FNRA cultures, although the proportion of the Sox17-eGFP<sup>+</sup> was reduced (Fig. 6A). Thus, ES cell cultures produced progeny with several features of DE, including expression of cell surface markers present on native DE.

Based on these results, we asked whether flow cytometry based on EpCAM, CD55, CD24, and CD38 could be used to isolate endoderm derivatives from genetically unmodified ES cells. We sorted EpCAM<sup>+</sup> CD55<sup>+</sup> CD24<sup>+</sup> CD38<sup>-</sup> cells derived from FNRA treatment of the R1 mouse ESCs, the parental cell line used to construct the *Sox17-eGFP* line (Fig. 6B). Q-PCR revealed enrichment of mRNA encoding the DE markers *Sox17*, *FoxA1*, *FoxA2*, and *HNF4a* in R1-derived EpCAM<sup>+</sup> CD55<sup>+</sup> CD24<sup>+</sup> CD38<sup>-</sup> cells from FNRA cultures (Fig. 6C). In contrast, mRNAs encoding markers of undifferentiated ES cells, neuroectoderm, mesoderm, and VE (*Oct4*, *Sox1*, *Meox1*, *Dll1*, and *Sox7*) were depleted from EpCAM<sup>+</sup> CD55<sup>+</sup> CD24<sup>+</sup> CD38<sup>-</sup> cells but enriched in the EpCAM<sup>-</sup> CD55<sup>low/-</sup> cell subset (Fig. 6B,C). Thus, we have identified FACS methods that enrich for a population of DE-like cells derived from genetically unmodified ES cells.

*Differentiation of human ES cells towards DE*

To test whether the modifications used in mouse ES cell differentiation cultures might similarly refine endoderm

development from human ES cells, we exposed *hSOX17-eGFP* cells (abbreviated "*hS17-eGFP*"), which harbor an *eGFP* transgene insertion at the *SOX17* locus achieved through homologous targeting [16], to culture modifications including RA, FGF1, FGF4, and Noggin addition followed by microarray analysis. In contrast to undifferentiated *hS17* cells, which are *eGFP*<sup>-</sup>, flow cytometry analysis showed that *hS17-eGFP*<sup>+</sup> cells developed in all 3 culture media tested, with FN and FNRA conditions yielding slightly reduced numbers of *hS17-eGFP*<sup>+</sup> cells compared with WA-treated cells (Fig. 7A). *hS17-eGFP*<sup>+</sup> replicates for each sample were collected for RNA isolation, probe synthesis, and hybridization to the Affymetrix Human Genome U133 Plus 2.0 arrays and analysis with GeneSpring software (Supplementary Fig. 4A). We identified human orthologs of mouse genes that changed on modification of culture media in mESC (Fig. 5A) and the mouse endoderm-enriched gene modules (Supplementary Table S2), then analyzed expression of these genes in undifferentiated hES cells and *hS17-eGFP*<sup>+</sup> endoderm-like cells. The human orthologs of mRNAs enriched in E7.5 mouse DE had an increased expression in *hS17-eGFP*<sup>+</sup> cells after exposure to WA. As anticipated, the expression of human mRNAs orthologous to those enriched in E8.25 mouse DE and endoderm modules was increased in *hS17-eGFP*<sup>+</sup> cells resulting from FN and FNRA cultures, compared with WA cultures (Fig. 7B,C). Thus, FN and FNRA culture modifications led to similar changes of gene expression by mouse and human ES cells.

We next examined the expression of endodermal cell surface markers in mouse and human ES cell-derived endoderm [16]. Microarray analysis showed that, similar to mouse ES cells, human *hS17-eGFP*<sup>+</sup> cell expression of *CXCR4* was increased in WA and FN cultures, while the expression of *CD38* was unchanged. However, in contrast to mouse *eGFP*<sup>+</sup> progeny, the expression of *CD55*, *CD24*, and *EpCAM* was decreased or unchanged in human *hS17-eGFP*<sup>+</sup> cells (Supplementary Fig. S4B). By contrast, the human primitive gut specific markers *CD238*, *CD141*, and *CD49e* were not induced in mouse ESC derived endoderm (Supplementary Fig. S4C). These data reveal specific differences in endoderm-like cells derived from human or mouse ES cultures. Thus, surface antigens marking endoderm may differ in humans and mice, reminiscent of findings from human and mouse hematopoietic stem cells [35,36], and embryonic stem cells [37].

## Discussion

Studies described here advance our understanding of the molecular properties of mouse endoderm, leading to practical innovations for marking, generating, purifying, and assessing both native and in vitro-derived DE. While previous studies had reported gene expression profiles of mouse endoderm partially purified by microdissection or FACS [2,3], separation of definitive from VE was not previously achieved, thereby precluding discovery of a specific gene expression profile of DE. Thus, our use of classical FACS- and microarray-based methods to purify and analyze native DE from other primary germ layers and extraembryonic mouse cells generated a unique "molecular signature" of this tissue that will aid studies of the development of this germ layer and its derivatives both in vivo and in vitro. These

signatures proved heuristic in suggesting signaling pathway modifications that fine-tuned the development of *Sox17-eGFP*<sup>+</sup> ES cell progeny toward DE, while reducing its similarity to E7.5 DE. Thus, our study provides a paradigm for using the rigor of genomic-scale expression profiling of native endoderm to guide ES cell differentiation toward endodermal fates, a strategy not used in previous studies.

One impediment to investigating mouse endoderm biology has been a relative paucity of molecular probes for fundamental aspects of endoderm development, such as axial patterning, tissue specification, and differentiation. In this study, we generated highly specific gene expression profiles of DE and VE, which identified several endodermal markers we characterized and authenticated using flow cytometry and in situ hybridization. To generate endoderm purification strategies independent of transgene marking or mouse genotype, we focused attention on markers predicted by microarray analysis to be cell surface proteins expressed in endoderm, such as *CD24*, *CD55*, and *CD38*. Thus, our flow cytometry studies showing that native DE at E7.5 and E8.25 is *EpCAM*<sup>+</sup> *CD55*<sup>+</sup> *CD38*<sup>-</sup> *CD24*<sup>+</sup> provide a unique strategy for purifying endoderm from mice in a variety of genetic conditions. Combined with other methods [38], these findings should facilitate studies of endoderm formation.

DE is patterned along all 3 major embryonic axes, most strikingly the anterior-posterior axis, and DE isolated by FACS based on *CD24* or *CD55* was enriched for a variety of markers expressed throughout the endoderm axis, such as *Gfpt2* and *Eppk1* in foregut, and *Krt19* in hindgut. Thus, results reported here should facilitate region specific DE cell labeling and purification, allowing further developmental studies of mechanisms underlying foregut, midgut, and hindgut endoderm regionalization, a highly dynamic process [10]. Our studies also reveal methods to isolate and investigate embryonic VE, a vital source of extraembryonic tissues such as the yolk sac, where embryonic hematopoiesis initiates.

Previous studies have reported differentiation of endoderm-like cells from mouse and human embryonic stem cells after treatment with growth factors or small molecules [4,6–9,39–41]. These differentiated cell populations were found to be heterogeneous, containing undifferentiated cells, ectoderm and mesodermal derivatives in addition to endoderm-like cells. To address this, transgenic mouse ES cell lines expressing fluorescent protein markers from loci such as *Gsc*, *FoxA2*, *Hex*, and *Sox17* were used to isolate endoderm from these cell mixtures by FACS, but only 1 previous report described attempts to purify DE from genetically unmodified mouse ES cells [7]. In that study, antibodies to E-cadherin and *Cxcr4* permitted FACS isolation of mesendoderm-like progeny of mES cells, but did not permit FACS purification of native mouse DE, largely because E-cadherin and *Cxcr4* are expressed in ectoderm and mesoderm, respectively [42,43]. Here, we identified new combinations of antibodies recognizing cell surface markers for FACS-based purification of native definitive and VE, and of endoderm-like progeny from ES cell lines. Thus, our studies of endodermal development produced unique methods to purify relevant cells from heterogeneous in vitro cultures or embryonic tissues, which should accelerate understanding of mechanisms regulating endodermal development and organogenesis.

Our findings correlate with a subset of findings by Hoodless and colleagues [2] that coupled mouse embryonic

microdissection with Serial Analysis of Gene Expression to identify markers of mouse endoderm (Supplementary Fig. S2). For example, similar to an earlier study, we also found enrichment of *Neprn*, *Pyy*, and *Trh* in Sox17-eGFP<sup>+</sup> CD24<sup>+</sup> DE and of *Apoc2*, *Lgals2*, *Tdh*, *Pla2g12b*, and *Cubilin* in Sox17-eGFP<sup>+</sup>CD24<sup>lo/-</sup> VE. This correlation supports the use of CD24 as a marker that can distinguish between DE and VE in the mouse embryo. We also identified additional DE markers not described in Hou et al. (2007), including *Eppk1*, *Gfpt2*, *Sorcs2*, *Nedd9*, and *Krt19*. Our gene expression profiling also corroborates a subset of data reported by Sherwood et al. (2007), who used a FACS-based strategy to isolate endoderm from embryonic mice harboring a transgene encoding *Sox17 cis*-regulatory elements that drove expression of a modified yellow fluorescent protein in all 3 primary embryonic germ layers (Supplementary Fig. S2). However, unlike that previous study, our gene expression profiling clearly distinguishes DE and VE, allowing for several innovations, including separation and purification of genetically unmodified DE and VE.

Our understanding of signaling pathways that regulate endoderm development has grown in recent years [10,44], but much remains to be learned about the cell interactions and signaling pathways that govern endodermal differentiation, which occurs in the context of complex morphogenetic movement of the primary germ layers. Experiments in frogs, fish, and mice provide strong evidence that Wnt and Nodal signaling regulate early endoderm development [10,45], and addition of purified Wnt and Activin A to mouse and human ES cell cultures is a “standard” strategy for inducing differentiation of endoderm-like progeny [46,47]. However, results described here demonstrate substantial differences between the gene expression profile of purified native Sox17<sup>+</sup> DE and that of Sox17<sup>+</sup> endoderm-like progeny derived from Wnt and Activin A treated ES cells, including pathways regulated by FGFs, BMP, and RA. For the FGF pathway, these data support conclusions from recent studies on FGF regulation of mouse DE development [9]. Collectively, our findings may prove useful for verifying the quality of endoderm-like cells produced from multipotent sources such as ES or induced pluripotent stem (iPS) cells. Moreover, our gene expression findings motivated us to modify FGF, Noggin, and RA signaling in mouse ES cell cultures, resulting in an increased correlation of the gene expression patterns between ES-derived endoderm-like cells and native DE. Thus, our studies have proved heuristic for optimizing derivation and purification of ES cell progeny with features of DE. BMP and FGF signaling are also thought to direct axial specialization of native endoderm; thus, endodermal development from ES progeny with modulators of these signaling pathways may reflect both refined formation of DE-like cells and recapitulation of axial patterning in our cultures. We postulate that the remaining gap between native and ES cell-derived DE reflects the absence of additional signaling interactions in ES cell cultures, including the complex, dynamic reciprocal signaling known to occur between DE and adjacent mesoderm. We further speculate that our results should facilitate small molecule screens [4] to identify index compounds that promote endoderm differentiation or maturation.

Based on progress from studies with our mouse *Sox17-eGFP* transgenic ES cell line, we used a homologous recombination to build a similar human ES cell line to investigate

the development of human endoderm-like cells [16]. Here, we showed that human SOX17-eGFP<sup>+</sup> cells derived from cultures exposed to Wnt and Activin A can be FACS purified based on eGFP expression to assess endoderm differentiation. For example, similar to our mouse studies, FACS purification of human SOX17-eGFP<sup>+</sup> cells allowed microarray assessment of cells cultured with specific conditions to develop endoderm-like cells. Fine-tuning of an endoderm-like gene expression signature in hSOX17-eGFP<sup>+</sup> cells derived from cultures exposed to unique combinations of growth factors including Wnt, Activin A, FGF, Noggin, and RA provides evidence that this new human ES cell line will be a useful tool for studies of human endoderm differentiation. Moreover, our work suggests that studies of native mouse DE can guide hES cell differentiation in vitro.

### Acknowledgments

The authors thank Drs. Thomas Zwaka, James Thomson, Roel Nusse, Timothy Blauwkamp, Eric Chiao, Julie C. Baker, Wing Hung Wong, and Yanru Tsai of the Stanford University School of Medicine transgenic mouse core facility, and members of the Kim group for materials, advice, and assistance. P.W. was supported by a Larry L. Hillblom Foundation (LLHF) fellowship. K.M. is supported by Juvenile Diabetes Research Foundation (JDRF) fellowship. D.W. was supported by an NIH NIAMS K08 award and a Dermatology Foundation Career Development Award. H.C. was supported by JDRF, NIH BCBC3, the California Institute for Regenerative Medicine (CIRM) and is an early career scientist of the Howard Hughes Medical Institute (HHMI). The Kim group was supported by the Mead Foundation, LLHF, CIRM, and HHMI. S.K. is an investigator of the HHMI.

### Author Disclosure Statement

There are no conflicts of interest to declare.

### References

- McKnight KD, P Wang and SK Kim. (2010). Deconstructing pancreas development to reconstruct human islets from pluripotent stem cells. *Cell Stem Cell* 6:300–308.
- Hou J, AM Charters, SC Lee, Y Zhao, MK Wu, SJ Jones, MA Marra and PA Hoodless. (2007). A systematic screen for genes expressed in definitive endoderm by Serial Analysis of Gene Expression (SAGE). *BMC Dev Biol* 7:92.
- Sherwood RI, C Jitianu, O Cleaver, DA Shaywitz, JO Lamenza, AE Chen, TR Golub and DA Melton. (2007). Prospective isolation and global gene expression analysis of definitive and visceral endoderm. *Dev Biol* 304:541–555.
- Borowiak M, R Maehr, S Chen, AE Chen, W Tang, JL Fox, SL Schreiber and DA Melton. (2009). Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. *Cell Stem Cell* 4:348–358.
- D’Amour KA, AG Bang, S Eliazar, OG Kelly, AD Agulnick, NG Smart, MA Moorman, E Kroon, MK Carpenter and EE Baetge. (2006). Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nat Biotechnol* 24:1392–1401.
- Kubo A, K Shinozaki, JM Shannon, V Kouskoff, M Kennedy, S Woo, HJ Fehling and G Keller. (2004). Development of definitive endoderm from embryonic stem cells in culture. *Development* 131:1651–1662.

7. Yasunaga M, S Tada, S Torikai-Nishikawa, Y Nakano, M Okada, LM Jakt, S Nishikawa, T Chiba, T Era and S Nishikawa. (2005). Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. *Nat Biotechnol* 23:1542–1550.
8. D'Amour KA, AD Agulnick, S Eliazzer, OG Kelly, E Kroon and EE Baetge. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotechnol* 23:1534–1541.
9. Morrison GM, I Oikonomopoulou, RP Migueles, S Soneji, A Livigni, T Enver and JM Brickman. (2008). Anterior definitive endoderm from ESCs reveals a role for FGF signaling. *Cell Stem Cell* 3:402–415.
10. Zorn AM and JM Wells. (2007). Molecular basis of vertebrate endoderm development. *Int Rev Cytol* 259:49–111.
11. Arnold SJ and EJ Robertson. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat Rev Mol Cell Biol* 10:91–103.
12. Kroon E, LA Martinson, K Kadoya, AG Bang, OG Kelly, S Eliazzer, H Young, M Richardson, NG Smart, et al. (2008). Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells *in vivo*. *Nat Biotechnol* 26:443–452.
13. Kanai-Azuma M, Y Kanai, JM Gad, Y Tajima, C Taya, M Kurohmaru, Y Sanai, H Yonekawa, K Yazaki, PP Tam and Y Hayashi. (2002). Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development* 129:2367–2379.
14. Niakan KK, H Ji, R Maehr, SA Vokes, KT Rodolfa, RI Sherwood, M Yamaki, JT Dimos, AE Chen, et al. (2010). Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes Dev* 24:312–326.
15. Ramirez-Solis R, AC Davis and A Bradley. (1993). Gene targeting in embryonic stem cells. *Methods Enzymol* 225:855–878.
16. Wang P, RT Rodriguez, J Wang, A Ghodasara and SK Kim. (2011). Targeting SOX17 in human embryonic stem cells creates unique strategies for isolating and analyzing developing endoderm. *Cell Stem Cell* 8:335–346.
17. Downs KM and T Davies. (1993). Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* 118:1255–1266.
18. McKnight KD, J Hou and PA Hoodless. (2007). Dynamic expression of thyrotropin-releasing hormone in the mouse definitive endoderm. *Dev Dyn* 236:2909–2917.
19. Lawson KA, JJ Meneses and RA Pedersen. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* 113:891–911.
20. Kwon GS, M Viotti and AK Hadjantonakis. (2008). The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. *Dev Cell* 15:509–520.
21. Kim I, TL Saunders and SJ Morrison. (2007). Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells. *Cell* 130:470–483.
22. Spence JR, AW Lange, SC Lin, KH Kaestner, AM Lowy, I Kim, JA Whitsett and JM Wells. (2009). Sox17 regulates organ lineage segregation of ventral foregut progenitor cells. *Dev Cell* 17:62–74.
23. Bettenhausen B, M Hrabe de Angelis, D Simon, JL Guenet and A Gossler. (1995). Transient and restricted expression during mouse embryogenesis of Dll1, a murine gene closely related to *Drosophila* Delta. *Development* 121:2407–2418.
24. Pevny LH, S Sockanathan, M Placzek and R Lovell-Badge. (1998). A role for SOX1 in neural determination. *Development* 125:1967–1978.
25. Kispert A and BG Herrmann. (1994). Immunohistochemical analysis of the Brachyury protein in wild-type and mutant mouse embryos. *Dev Biol* 161:179–193.
26. Hart AH, L Hartley, K Sourris, ES Stadler, R Li, EG Stanley, PP Tam, AG Elefanty and L Robb. (2002). Mixl1 is required for axial mesendoderm morphogenesis and patterning in the murine embryo. *Development* 129:3597–3608.
27. Uchida N, DW Buck, D He, MJ Reitsma, M Masek, TV Phan, AS Tsukamoto, FH Gage and IL Weissman. (2000). Direct isolation of human central nervous system stem cells. *Proc Natl Acad Sci U S A* 97:14720–14725.
28. Cram DS, A McIntosh, L Oxbrow, AM Johnston and HJ DeAizpurua. (1999). Differential mRNA display analysis of two related but functionally distinct rat insulinoma (RIN) cell lines: identification of CD24 and its expression in the developing pancreas. *Differentiation* 64:237–246.
29. Mitiku N and JC Baker. (2007). Genomic analysis of gastrulation and organogenesis in the mouse. *Dev Cell* 13:897–907.
30. Anderson WJ, Q Zhou, V Alcalde, OF Kaneko, LJ Blank, RI Sherwood, JS Guseh, J Rajagopal and DA Melton. (2008). Genetic targeting of the endoderm with claudin-6CreER. *Dev Dyn* 237:504–512.
31. Echelard Y, DJ Epstein, B St-Jacques, L Shen, J Mohler, JA McMahon and AP McMahon. (1993). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* 75:1417–1430.
32. Lin F, Y Fukuoka, A Spicer, R Ohta, N Okada, CL Harris, SN Emancipator and ME Medof. (2001). Tissue distribution of products of the mouse decay-accelerating factor (DAF) genes. Exploitation of a Dafi1 knock-out mouse and site-specific monoclonal antibodies. *Immunology* 104:215–225.
33. Dessimoz J, R Opoka, JJ Kordich, A Grapin-Botton and JM Wells. (2006). FGF signaling is necessary for establishing gut tube domains along the anterior-posterior axis *in vivo*. *Mech Dev* 123:42–55.
34. Ruberte E, P Dolle, P Chambon and G Morriss-Kay. (1991). Retinoic acid receptors and cellular retinoid binding proteins. II. Their differential pattern of transcription during early morphogenesis in mouse embryos. *Development* 111:45–60.
35. Spangrude GJ, S Heimfeld and IL Weissman. (1988). Purification and characterization of mouse hematopoietic stem cells. *Science* 241:58–62.
36. Baum CM, IL Weissman, AS Tsukamoto, AM Buckle and B Peault. (1992). Isolation of a candidate human hematopoietic stem-cell population. *Proc Natl Acad Sci U S A* 89:2804–2808.
37. Ginis I, Y Luo, T Miura, S Thies, R Brandenberger, S Gerrecht-Nir, M Amit, A Hoke, MK Carpenter, J Itskovitz-Eldor and MS Rao. (2004). Differences between human and mouse embryonic stem cells. *Dev Biol* 269:360–380.
38. Gadue P, V Gouon-Evans, X Cheng, E Wandzioch, KS Zaret, M Grompe, PR Streeter and GM Keller. (2009). Generation of monoclonal antibodies specific for cell surface molecules expressed on early mouse endoderm. *Stem Cells* 27:2103–2113.
39. Tada S, T Era, C Furusawa, H Sakurai, S Nishikawa, M Kinoshita, K Nakao and T Chiba. (2005). Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture. *Development* 132:4363–4374.

40. Hansson M, DR Olesen, JM Peterslund, N Engberg, M Kahn, M Winzi, T Klein, P Maddox-Hyttel and P Serup. (2009). A late requirement for Wnt and FGF signaling during activin-induced formation of foregut endoderm from mouse embryonic stem cells. *Dev Biol* 330:286–304.
41. Gadue P, TL Huber, PJ Paddison and GM Keller. (2006). Wnt and TGF-beta signaling are required for the induction of an *in vitro* model of primitive streak formation using embryonic stem cells. *Proc Natl Acad Sci U S A* 103:16806–16811.
42. Shimamura K and M Takeichi. (1992). Local and transient expression of E-cadherin involved in mouse embryonic brain morphogenesis. *Development* 116:1011–1019.
43. McGrath KE, AD Koniski, KM Maltby, JK McGann and J Palis. (1999). Embryonic expression and function of the chemokine SDF-1 and its receptor, CXCR4. *Dev Biol* 213:442–456.
44. Sherwood RI, TY Chen and DA Melton. (2009). Transcriptional dynamics of endodermal organ formation. *Dev Dyn* 238:29–42.
45. Stainier DY. (2005). No organ left behind: tales of gut development and evolution. *Science* 307:1902–1904.
46. Murry CE and G Keller. (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* 132:661–680.
47. Nostro MC, F Sarangi, S Ogawa, A Holtzinger, B Corneo, X Li, SJ Micallef, IH Park, C Basford, et al. (2011). Stage-specific signaling through TGF{beta} family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells. *Development* 138:861–871.

Address correspondence to:

*Prof. Seung K. Kim*

*Department of Developmental Biology*

*Howard Hughes Medical Institute*

*B300 Beckman Center*

*279 Campus Drive*

*Stanford, CA 94305*

*E-mail: seungkim@stanford.edu*

Received for publication July 31, 2011

Accepted after revision January 11, 2012

Prepublished on Liebert Instant Online January 11, 2012