Prospective isolation and global gene expression analysis of definitive and visceral endoderm

Richard I. Sherwood [a,*], Cristian Jitianu [b,c], Ondine Cleaver [a,1], David A. Shaywitz [a], Julia O. Lamenzo [a,b], Alice E. Chen [a], Todd R. Golub [b,d,e], Douglas A. Melton [a,e,*]

[a] *Department of Molecular and Cellular Biology, Harvard Stem Cell Institute, Harvard University, 7 Divinity Ave, Cambridge, MA 02138, USA*
[b] *Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA*
[c] *Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA*
[d] *Department of Pediatric Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA 02115, USA*
[e] *Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA*

Abstract

In spite of the therapeutic importance of endoderm derivatives such as the pancreas, liver, lung, and intestine, there are few molecular markers specific for early endoderm. In order to identify endoderm-specific genes as well as to define transcriptional differences between definitive and visceral endoderm, we performed microarray analysis on E8.25 definitive and visceral endoderm. We have developed an early endoderm gene expression signature, and clarified the transcriptional similarities and differences between definitive and visceral endoderm. Additionally, we have developed methods for flow cytometric isolation of definitive and visceral endoderm. These results shed light on the mechanism of endoderm formation and should facilitate investigation of endoderm formation from embryonic stem cells.
© 2007 Elsevier Inc. All rights reserved.

Introduction

During early vertebrate development, formation of the three germ layers (ectoderm, mesoderm, and endoderm) is one of the first major events in specifying cell fates in the embryo. Since embryonic stem (ES) cells are derived from and exhibit gene expression and functional properties characteristic of pluripotent embryonic cells, it is widely believed that directed differentiation of ES cells into specific cell types for therapeutic purposes will necessarily begin by inducing ES cells to form germ layer intermediates.

Endoderm gives rise to pancreatic, hepatic, lung, intestinal and other therapeutically relevant cell types, yet early endoderm development is not well understood. Fate mapping studies of

cultured mouse embryos (Lawson et al., 1986, 1991; Lawson and Pedersen, 1987) have revealed that definitive endoderm begins to form at embryonic days 6–6.5 (E6–E6.5) and that by the end of gastrulation (E7.5), some labeled cells only give rise to endodermal derivatives. Heterotopic transplantation studies have been carried out in zebrafish (David and Rosa, 2001) and in *Xenopus* (Heasman et al., 1984; Wylie et al., 1987), and these results demonstrate that by mid-to-late gastrulation, cells are determined to give rise to endoderm. Whether mice and humans have a similar time course for endoderm determination remains to be determined.

It is not known whether the initial definitive endoderm cells are multipotent. Fate mapping studies (Lawson et al., 1991; Tremblay and Zaret, 2005) suggest that the first endoderm cells that migrate through the primitive streak at E6.5 are fated to become liver, ventral pancreas, lungs and stomach; later cells become esophagus, stomach, dorsal pancreas, and duodenum; and the last cells out become intestine. Several early endodermal transcription factors, including Otx2, Hesx1, Hex, and Cdx2, are regionally expressed prior to the time that organ

---

* Corresponding authors.
  E-mail addresses: rsherw@fas.harvard.edu (R.I. Sherwood),
dmelton@harvard.edu (D.A. Melton).
  [1] Department of Molecular Biology, University of Texas–Southwestern
Medical Center, Dallas, TX 75390, USA.

specific genes are activated, at ∼E8.5 (Wells and Melton, 1999). However, co-culture experiments show that the endoderm in these regions is not fully committed at early stages. For example, E7.5 mouse anterior endoderm co-cultured with posterior mesoderm expresses markers of posterior endoderm (Wells and Melton, 2000), posterior transformations can occur after heterotopic transplantation of anterior chick endoderm from 12–14 somite-stage embryos (equivalent to ∼E8.5) (Kumar et al., 2003), and ventral foregut endoderm explants from mouse 2- to 6-somite-stage embryos (∼E8.25) can activate genes characteristic of the pancreas, liver or lung depending on the mesodermal tissues with which they are co-cultured (Deutsch et al., 2001; Serls et al., 2005). Thus, while early endoderm is regionalized in transcription factor expression as it migrates through the primitive streak, some of these endoderm cells appear to be capable of multilineage differentiation.

A complication in the study of endoderm is that mammals possess extraembryonic endoderm. Extraembryonic endoderm arises at the blastocyst stage and eventually forms two subpopulations: visceral endoderm, the chief metabolic component of the visceral yolk sac, and parietal endoderm, which secretes Reichert's membrane and contributes to the transient parietal yolk sac. Extraembryonic endoderm cells share the expression of many genes with definitive endoderm (cells that give rise to the endodermal organs), including the often analyzed transcription factors Sox17 (Kanai-Azuma et al., 2002), FoxA1 and FoxA2 (Belo et al., 1997; Sasaki and Hogan, 1993). In spite of their genetic similarity, lineage tracing evidence combined with transplantation experiments suggest that, as early as E3.5, cells are determined either to give rise to extraembryonic endoderm or to give rise to the precursors of all embryonic cell types, the pluripotent epiblast cells (Chazaud et al., 2006; Kunath et al., 2005). The genetic pathways that regulate cell fate determination of extraembryonic endoderm, as well as genes that can serve as markers to distinguish definitive and extraembryonic endoderm, are not understood.

Advances have been made in deriving endoderm from ES cells, yet a better understanding of definitive and extraembryonic endoderm is necessary for the field to progress. Recently, several groups have reported differentiation of mouse or human ES cells into definitive endoderm (D'Amour et al., 2005; Kubo et al., 2004; Tada et al., 2005; Yasunaga et al., 2005); however, these reports rely on the expression of a relatively small number of marker genes. In order to better understand the distinction between definitive and extraembryonic endoderm and to enable recognition of these populations from differentiating ES cells, we undertook gene expression profiling of E8.25 definitive and extraembryonic endoderm as well as non-endodermal tissues at this developmental stage. Previously, gene profiling of endoderm was performed exclusively using E7.5 definitive endoderm and mesectoderm (Gu et al., 2004). This did not permit a comparative analysis between definitive and extraembryonic endoderm.

Here we also report on methods that allow for prospective isolation of mouse definitive and extraembryonic endoderm. Using these isolation techniques, we perform global gene expression profiling on E8.25 definitive and extraembryonic endoderm and define a gene expression signature for each tissue. These results should facilitate the recognition, genetic analysis, and testing of the developmental potential of embryo-derived and ES cell-derived endoderm.

Materials and methods

Antibodies and immunostaining

The following primary antibodies were used: G8.8 (anti-EpCAM, Developmental Studies Hybridoma Bank, Iowa City, IA); FE-J1 (Developmental Studies Hybridoma Bank, Iowa City, IA); biotinylated SBA (Vector, Burlingame, CA); biotinylated DBA (Vector, Burlingame, CA); APA5 (anti-Pdgfra, ebioscience, San Diego, CA); Sox1 (Santa Cruz Biotechnology, Santa Cruz, CA); Brachyury (Santa Cruz Biotechnology, Santa Cruz, CA); biotin anti-CD106 (Vcam-1, BD Biosciences, San Jose, CA); PE anti-Flk1 (VEGFR2, BD Biosciences, San Jose, CA); MC-813-70 (anti-SSEA-4, DSHB, Iowa City, IA); Sox7 (Santa Cruz Biotechnology, Santa Cruz, CA); Cytokeratin-7 (Santa Cruz Biotechnology, Santa Cruz, CA); Annexin IV (Santa Cruz Biotechnology, Santa Cruz, CA); Sox17 (R+D Systems, Minneapolis, MN); YN1/1.7.4 (anti-Icam-1, ebioscience, San Diego, CA), Ceacam-1 (Santa Cruz Biotechnology, Santa Cruz, CA); Claudin-8 (Aviva Biosciences, San Diego CA), Desmoglein-2 (R&D Systems, Minneapolis, MN), FITC anti-Dpp4 (BD Biosciences, San Jose, CA), biotin anti-CD38 (ebioscience, San Diego, CA), biotin anti-Tim-2 (ebioscience, San Diego, CA). Rhodamine Red-X, APC, PE, PE-Cy7, and FITC-conjugated secondary antibodies (Jackson ImmunoResearch, West Grove, PA) were used.

For immunostaining, whole-mount embryos or sections were blocked with 20% donkey serum and stained with primary antibody overnight at 4 °C. For immunofluorescence, secondary antibody was added for 1 h at 25 °C, and for immunohistochemistry, Vectastain EliteABC followed by DAB (Vector Labs, Burlingame, CA) was used. Nuclei were visualized with Hoechst 33342 (Invitrogen, Carlsbad, CA). Photographs were taking using an LSM 510Meta confocal microscope (Carl Zeiss, Inc., Germany) or a dissecting microscope (Leica Microsystems, Germany).

Transgene construction and generation of Sox17-Venus transgenic mice

Recombineered bacterial artificial chromosomes (BACs) were generated (Yang and Seed, 2003). In brief, a kanamycin/neomycin-resistant template vector was engineered encoding the yellow fluorescent protein Venus (Nagai et al., 2002) (kindly provided by A. Miyawaki), followed by an SV40 poly A site. Oligonucleotides were designed to amplify the template vector as well as approximately 80 nucleotides of genomic sequence immediately preceding the initial ATG of Sox17 (5′ oligo) and immediately following the 10th nucleotide of Sox17 (3′ oligo) such that, upon recombination with a BAC, the first 10 nucleotides of Sox17 coding region (all in exon 4) would be replaced by the Venus-encoding gene and antibiotic resistance cassette. The resulting amplicon was electroporated into induced bacteria harboring both a Sox17-containing BAC (resistant to chloromycin) as well as 300 arabinose-inducible recombinase components. Colonies resistant to kanamycin and chloramphenicol were obtained, and successful recombination was confirmed by PCR. BAC DNA was subsequently electroporated into mouse embryonic stem cells, and neomycin-resistant colonies were obtained. Following PCR confirmation of the BAC DNA, transgenic ES cells were injected into blastocysts, and founder mice ultimately identified. Primers are listed in Supplementary Table 1.

Tissue isolation and cell sorting

For all experiments, outbred ICR mice were used for breeding, and embryos were considered to be E0.5 at noon of the day the plug was detected.

For manual endoderm isolation, E8.0–E8.25 definitive endoderm medial to somites, spanning from anterior intestinal portal to unsegmented mesoderm, was

isolated by manual dissection with a tungsten needle in PBS+0.1% dispase. Endoderms were pooled in groups of five and immediately transferred to extraction buffer (XB) of the Arcturus RiboAmp kit for cRNA amplification.

Embryos to be used for flow cytometry were dissected in PBS and, when noted, extraembryonic membranes were peeled from embryonic regions with fine forceps. Embryos were proteolytically dissociated with 0.25% trypsin (Invitrogen, Carlsbad, CA) or with 1 mg/ml papain (US Biological, Swampscott, MA). For live cell sorting, antibody staining was performed in DMEM:F12+2% FBS+ 10 mM EDTA for 20 min on ice. Before flow cytometric sorting, cells were resuspended in staining buffer with calcein blue AM (Invitrogen, Carlsbad, CA). For intracellular flow cytometric analysis, cells were fixed in PBS with 1% paraformaldehyde and 0.01% Tween-20 (Sigma-Aldrich, St. Louis, MO) and stained in PBS with 5% donkey serum and 0.05% Tween-20. Flow cytometric sorting was performed using either a FACSAria (Becton Dickinson, San Jose,

CA) or a MOFLO (Dako Cytomation, Carpinteria, CA), and data were analyzed using DIVA software (Becton Dickinson, San Jose, CA).

cDNA Generation and hybridization to Affymetrix microarrays

Total RNA was isolated using TRIzol (Invitrogen, Carlsbad, CA) or RiboAmp extraction buffer (Arcturus, Mountain View, CA). Target was prepared from ≥50 ng of isolated RNA using the NuGEN WT-Ovation RNA Amplification Kit (NuGen, San Carlos, CA) or RiboAmp RNA Amplification Kit (Arcturus, Mountain View, CA) and hybridized to Affymetrix Mouse Genome 430 2.0 GeneChips. All samples types were prepared as 3–6 biological replicates. Data were analyzed using GenePattern software package (www. broad.mit.edu/cancer/software/genepattern/) and Rosetta Resolver (Rosetta Biosoftware, Seattle, WA).
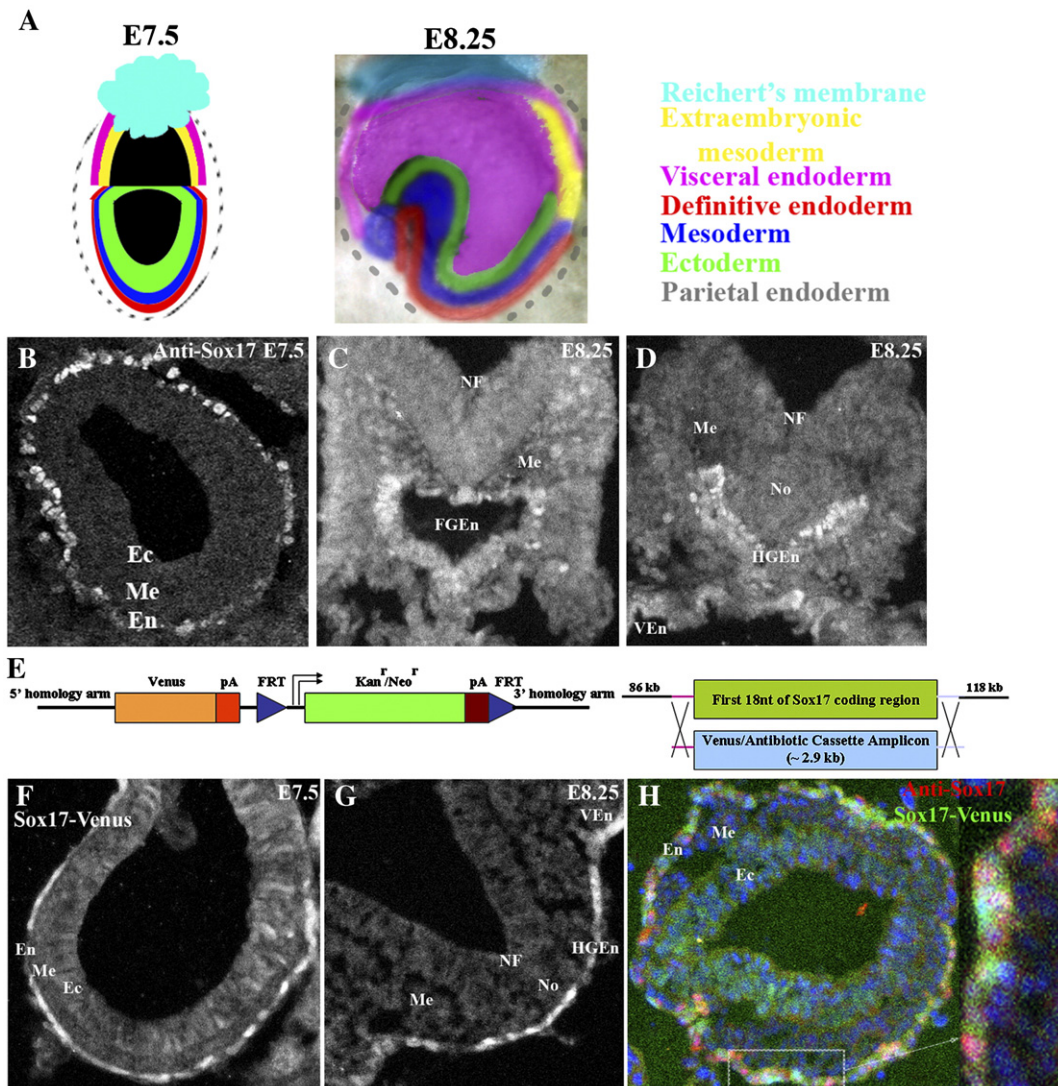

Fig. 1. Construction and endoderm-specific fluorescence of Sox17-Venus reporter mice. (A) Diagrams of the germ layers in developing mouse embryos. At E7.5, endoderm composes a single-cell-thick layer on the outside of the cup-shaped embryo, surrounding mesoderm and ectoderm. The embryo composes the bottom of the cup, and visceral endoderm is above definitive endoderm. At E8.25, the foregut and hindgut endoderm have begun to form a single-cell-thick tube, while the midgut remains a single-cell-thick sheet. Visceral endoderm surrounds the embryo and is contiguous with definitive endoderm. (B–D) Cryosection immunofluorescence images of Sox17 antibody staining in E7.5 transverse section (B) and in E8.25 transverse sections (C: foregut, D: hindgut). Sox17 expression is found in regions known by fate mapping to be endodermal. (E) Schematics of the Sox17-Venus bacterial artificial chromosome (BAC) targeting construct (left) and the region within the BAC at which the construct integrated by homologous recombination. (F–H) Cryosection immunofluorescence images of Venus (white in F–G; green in H) and Sox17 antibody staining (red in H). Venus fluorescence is found in endodermal regions in both E7.5 longitudinal section (F, H) and E8.25 hindgut transverse section (G) and colocalizes with Sox17 protein in E7.5 transverse section (H, close-up of boxed area on right). Ec=ectoderm, Me=mesoderm, En=endoderm, FGEn=foregut endoderm, NF=neural folds, HGEn=hindgut endoderm, No=notochord, VE=visceral endoderm.

In situ hybridization

In situ hybridization was performed according to Wilkinson and Nieto (1993). cDNA was obtained from commercial clone libraries (Open Biosystems, Huntsville, AL) and digoxigenin-labeled probes were generated. E8.25 and E9.5 embryos were fixed overnight in 4% paraformaldehyde, dehydrated in methanol, bleached in hydrogen peroxide and treated with proteinase K, followed by re-fixation. Probe was added at 1 μg/ml overnight at 70 °C. Embryos were washed and labeled overnight with anti-digoxigenin–AP antibody (Roche Applied Science, Indianapolis, IN). Embryos were developed in BM Purple (Roche Applied Science, Indianapolis, IN).

Quantitative PCR

Quantitative PCR was carried out using the CellsDirect RNA isolation system (Invitrogen, Carlsbad, CA) followed by Superscript III-mediated reverse transcription (Invitrogen, Carlsbad, CA) and SYBR Green PCR (Invitrogen, Carlsbad, CA). Flow cytometrically sorted cell populations were pelleted and resuspended in CellsDirect buffer. PCR was performed in a MX3000p light cycler (Stratagene, La Jolla, CA) up to 45 cycles, with 55 °C annealing temperature. Primers used are listed in Supplementary Table 1. Expression values were normalized to beta actin and GAPDH.

Embryonic stem cell culture

Undifferentiated mouse ES cells were maintained on gelatin-coated plates with mouse embryonic fibroblast (MEF) feeders in DMEM (Invitrogen, Carlsbad, CA) supplemented with 15% defined fetal bovine serum (FBS) (HyClone, Logan, UT), 0.1 mM nonessential amino acids (Invitrogen, Carlsbad, CA), Glutamax (Invitrogen, Carlsbad, CA), penicillin/streptomycin (Invitrogen, Carlsbad, CA), 0.55 mM 2-mercaptoethanol (Sigma-Aldrich, St. Louis, MO), 5×10^5 units LIF (Chemicon, Temecula, CA).

Prior to differentiation, ES cells were passaged onto gelatin-coated plates for 45 min to deplete MEFs. MEF-depleted ES cells were then seeded at 10^6 cells/5 ml/well of 6-well low cluster dishes for EB formation or seeded at ~2700 cells/cm² on gelatin-coated dishes for monolayer culture. Cells were allowed to settle overnight and switched to differentiation media the following day following a brief wash in phosphate-buffered saline. Differentiation was carried out for 7 days in DMEM (Invitrogen, Carlsbad, CA) supplemented with 5% defined FBS, 0.1 mM nonessential amino acids (Invitrogen, Carlsbad, CA), Glutamax (Invitrogen, Carlsbad, CA), penicillin/streptomycin (Invitrogen, Carlsbad, CA), 0.55 mM 2-mercaptoethanol (Sigma-Aldrich, St. Louis, MO), and 100 ng/ml recombinant Activin A (R&D Systems, Minneapolis, MN), with fresh media change every other day.

Results

Identification of endoderm by a Sox17-venus reporter mouse

In order to enable isolation of endoderm by methods distinct from physical dissection, we developed a transgenic mouse strain with endoderm-specific fluorescence. We took advantage of previous experimental evidence from *Xenopus* (Hudson et al., 1997) and mouse (Kanai-Azuma et al., 2002) that indicates that the transcription factor Sox17 is specifically expressed in all early endoderm. Fate mapping experiments (Lawson et al., 1991) demonstrate that both definitive and visceral endoderm compose the external cell layer of the embryonic cup at E7.5 (diagrammed in Fig. 1A). At E8.25, definitive endoderm has begun to form a tube of cells in the foregut and hindgut while remaining a sheet of cells in the midgut, and visceral endoderm forms a sheet of cells surrounding the embryo on all sides (diagrammed in Fig. 1A). It was confirmed that Sox17 protein is restricted to these endodermal regions at E7.5 (Fig. 1B) as well as at E8.25 (Figs. 1C, D).

Thus, a transgenic mouse expressing the yellow fluorescent protein variant Venus under the control of Sox17 regulatory elements was created. A bacterial artificial chromosome (BAC) containing the genomic region 86 kb upstream through 111 kb downstream of the *Sox17* gene was engineered with the first 10 nucleotides of Sox17 coding region replaced by a Venus-containing construct (Fig. 1E). This construct was electroporated into ES cells, chimeras were generated by blastocyst injection, and Sox17 transgenic founders were selected. Section immunofluorescence analysis of transgenic embryos at E7.5 and E8.25 revealed fluorescence in definitive and visceral endoderm (Figs. 1F, G), and Venus fluorescence was confirmed to colocalize with Sox17 antibody staining in cryosections (Fig. 1H).

Flow cytometric analysis of E8.25 Sox17-Venus embryos proteolytically dissociated into a single cell suspension revealed a higher fluorescence level in all cells as compared to non-transgenic littermates, due to either ubiquitous low-level expression of Sox17 or transgene leakiness; however, a Venus (hi) subset could be distinguished (Supplemental Fig. 1). The Venus(hi) population comprised between 5 and 15% of total cells at all stages between E7.5 and E10.5 (data not shown), and microarray analysis (see below) confirms that this Venus(hi) population is highly enriched in Sox17 transcript compared to Venus(lo) cells.

Prospective isolation of endoderm

We used the endoderm-enriched fluorescence of Sox17-Venus transgenic mice to distinguish live early endoderm cells

---

Fig. 2. Flow cytometric prospective isolation of endoderm. (A–B) Flow cytometric analysis of live, dissected and dissociated Sox17-Venus transgenic embryonic cells. In E8.25 (A) and E9.5 (B) embryonic regions, cells that have highest expression of the EpCAM antigen (*Y*-axis) display highest Sox17-Venus fluorescence (*X*-axis). (C–E) Cryosection immunofluorescence images of EpCAM expression. In E7.5 transverse section (C), EpCAM is expressed in endoderm (En) as well as ectoderm (Ec) but is absent from mesoderm (Me). In E8.25 transverse section (D), EpCAM is expressed strongly in endoderm of the foregut (FGEn), hindgut (HGEn) and extraembryonic (VE) regions, weakly in ectoderm (NF) and notochord (No) and is absent in mesoderm (Me). In E9.5 sagittal section (E), EpCAM (red) is expressed exclusively in the endodermally derived pharynx (PhEn), pancreas (PaEn) and intestine (InEn). (F) Flow cytometric analysis of live, dissociated embryonic cells stained with soybean agglutinin (SBA, *X*-axis) and EpCAM (*Y*-axis). SBA stains a subset of EpCAM(hi) cells. (G) Cryosection immunofluorescence analysis of SBA in E7.5 transverse section. SBA is restricted to the endoderm. (H–J) Combined extracellular (H, I: EpCAM, *Y*-axis; J: SBA, *Y*-axis) and intracellular (Sox1, *X*-axis) flow cytometric analysis. (H) In E8.25 embryonic regions, cells expressing the neural-specific transcription factor Sox1 are exclusively found in the EpCAM(lo/−) population. (I–J) Sox1 expression in E7.5 embryonic regions is contained within EpCAM(hi) cells, but it is confined to SBA(−) cells. (K) Whole-mount immunohistochemistry image of Pdgfra expression at E8.25. Staining is detected in mesodermally derived head mesenchyme (HMe) and somites (So). (L) Flow cytometric analysis of live, dissociated E8.25 embryonic cells stained with Pdgfra (*X*-axis) and EpCAM (*Y*-axis). Pdgfra expression is seen in the 40% of cells expressing the lowest levels of EpCAM. Ec=ectoderm, Me=mesoderm, En=endoderm, FGEn=foregut endoderm, NF=neural folds, HGEn=hindgut endoderm, No=notochord, VE=visceral endoderm, HMe=head mesenchyme, So=somitic mesoderm.
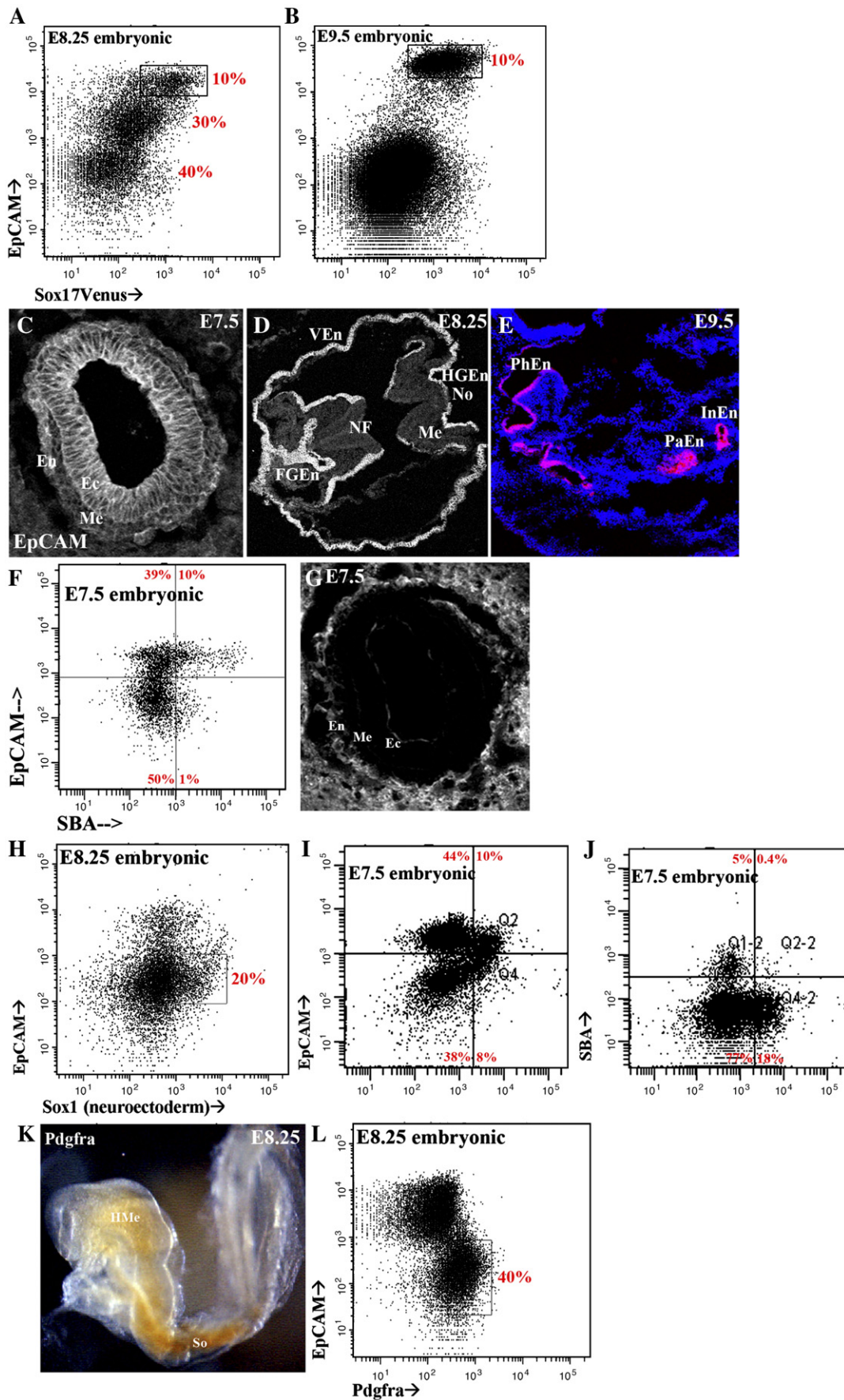
and to identify antibodies and lectins that bind preferentially to these cells. Antibodies and lectins that recognize cell surface epitopes were screened by flow cytometry for enriched binding to the Sox17-Venus(hi) definitive endoderm population in embryos stripped of extraembryonic cells and proteolytically dissociated into a single cell suspension. A monoclonal antibody, G8.8, that recognizes the cell adhesion protein EpCAM (Farr et al., 1991), was found to bind preferentially to Venus(hi) cells at E8.25 (Fig. 2A) and to bind exclusively to Venus(hi) cells at E9.5 (Fig. 2B). At early gastrulation stages such as E6.5 and E7.5, EpCAM displays broader flow cytometric expression, and EpCAM is expressed highly in embryonic stem cells, which resemble blastula-stage epiblast (data not shown).

The endoderm-enriched binding of EpCAM at E8.25 and E9.5 but not at E7.5 was confirmed by section immunofluorescence analysis. EpCAM expression is strong in both ectoderm and endoderm at E7.5 (Fig. 2C) 4 and becomes weaker in neuroectoderm by E8.25 (Fig. 2D), and EpCAM expression is exclusive to endoderm at E9.5 (Fig. 2E). EpCAM remains strongly expressed in most endodermal cells until at least E12.5 (data not shown). Thus, EpCAM allows prospective isolation of endoderm at and after E8.25.

To permit prospective isolation of endoderm at stages earlier than E8.25, dissociated embryonic regions of E7.5 embryos were further screened by flow cytometry for antibodies or lectins that recognize a subset of EpCAM(hi) cells. The lectin soybean agglutinin (SBA) recognizes 10% of all E7.5 cells,
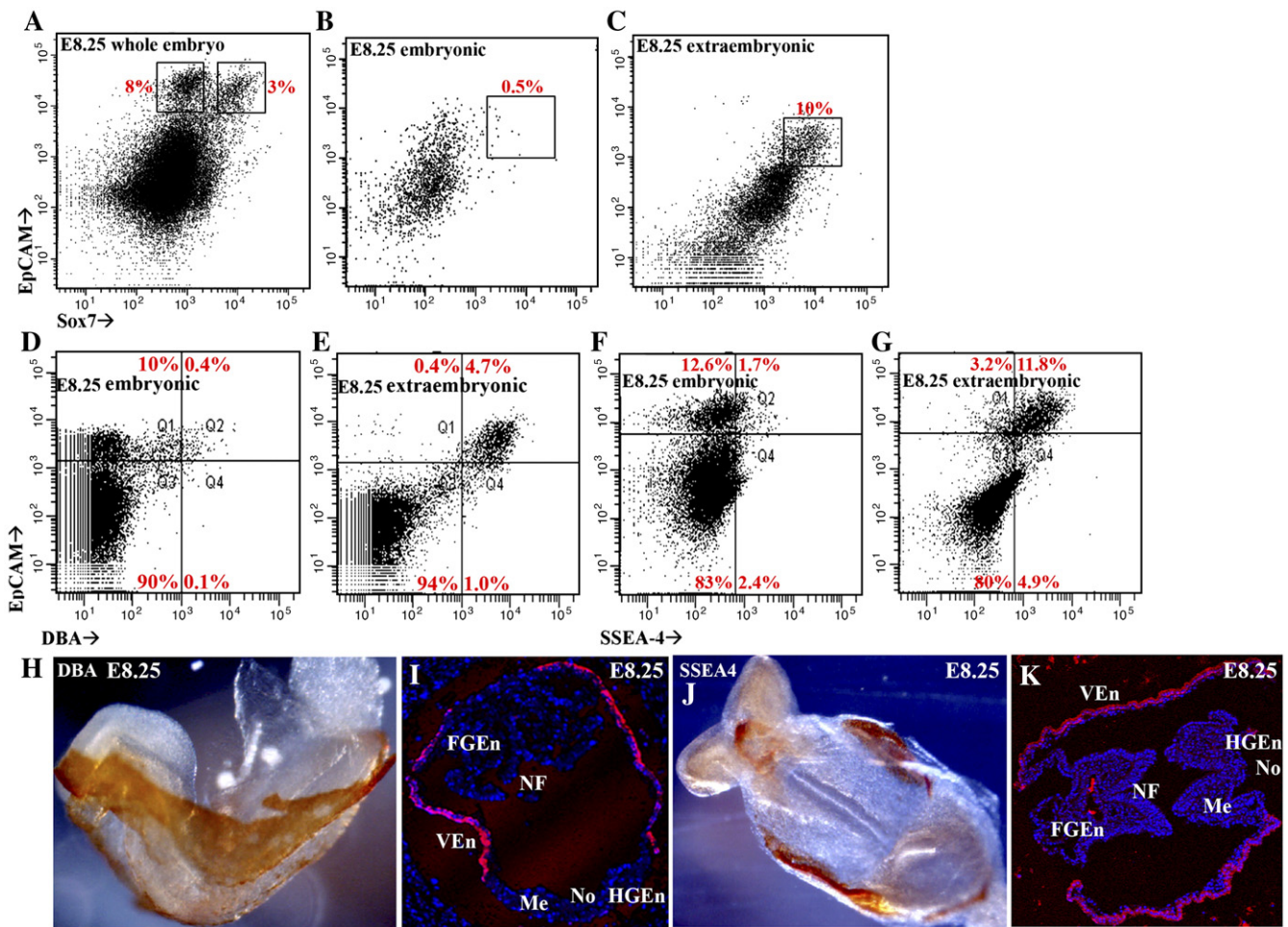

Fig. 3. Flow cytometric distinction of definitive and visceral endoderm. (A–C) Combined extracellular (EpCAM) and intracellular (Sox7) flow cytometric analysis of E8.25 embryos. Sox7, a known extraembryonic endoderm-specific transcription factor, stains a subset of EpCAM(hi) cells in whole embryos (A). When embryonic ad extraembryonic regions (which contain visceral endoderm) are dissected apart from each other, Sox7 reactivity is present in nearly all extraembryonic EpCAM(hi) cells (C) yet only a small percentage of embryonic EpCAM(hi) cells (B). (D–E) Flow cytometric analysis of live, dissected cells stained with dolichos biflorus agglutinin (DBA, X-axis) and EpCAM (Y-axis). Dissected E8.25 embryonic cells (D) do not express DBA, whereas nearly all dissected E8.25 extraembryonic EpCAM(hi) cells (E) express DBA. (F–G) Flow cytometric analysis of live, dissected cells stained with SSEA-4 (X-axis) and EpCAM (Y-axis). Dissected E8.25 embryonic cells (F) do not express SSEA-4, whereas nearly all dissected E8.5 extraembryonic EpCAM(hi) cells (G) express SSEA-4. (H) E8.25 whole-mount immunohistochemistry image demonstrating DBA expression in visceral endoderm. (I) Section immunofluorescence image demonstrating DBA expression in visceral endoderm (VE) but not in definitive foregut endoderm (FGEn) or hindgut endoderm (HGEn). (I) E8.25 whole-mount immunohistochemistry image demonstrating SSEA-4 expression in visceral endoderm. (J) Section immunofluorescence image demonstrating SSEA-4 expression in visceral endoderm (VE) but not in definitive foregut endoderm (FGEn) or hindgut endoderm (HGEn). Me=mesoderm, FGEn=foregut endoderm, NF=neural folds, HGEn=hindgut endoderm, No=notochord, VE=visceral endoderm.

Table 1
Endoderm-enriched transcription factors

All endoderm enriched	Definitive endoderm enriched (>2-fold)	Visceral endoderm enriched (>3-fold)
5730467H21Rik	Dlx3	Cited1
Foxa1	Dlx5	Foxa3
Ripk4	Foxg1	Hnf4A
Sox17	Gata3	Irf6
	Idb4	Nfatc2
	Pax9	Npas2
	Sp6	Tcf2
	Trp63	Vdr

Transcripts characterized by gene ontology as possessing transcription factor activity that are enriched >2-fold in definitive endoderm and/or >3-fold in extraembryonic endoderm than in all other microarray sample groups are listed according to their expression pattern.

virtually all EpCAM(hi) ectoderm/endoderm cells (Fig. 2F). Flow cytometric analysis at later stages reveals that SBA colocalizes with EpCAM(hi) endoderm at E8.25 and E9.5 as well (Supplemental Fig. 2). Section immunofluorescence demonstrates that SBA is exclusively expressed in endoderm at E7.5 and E8.25 (Fig. 2G and Supplemental Fig. 2). Interestingly, the monoclonal antibody FE-J1 (Fenderson et al., 1984) has a similar flow cytometric and immunofluorescence profile at E7.5 and E8.25 but is expressed by very few cells in the E9.5 embryo (Supplemental Fig. 2).

As further proof that EpCAM at E8.25 and SBA at E7.5 allow flow cytometric isolation of endoderm, their expression was compared to markers of neuroectoderm and mesoderm. Intracellular flow cytometric analysis of fixed E8.25 embryonic cell suspensions confirms that the neuroectodermal marker Sox1 (Wood and Episkopou, 1999) is expressed in a subset of EpCAM(lo/−) cells but not in EpCAM(hi) endoderm (Fig. 2H). Unfortunately, the fixation required for intracellular flow cytometric analysis prevents resolution of EpCAM(lo) and EpCAM(−) subsets. At E7.5, Sox1-expressing cells comprise a subset of EpCAM(hi) cells but are mutually exclusive with SBA-expressing endoderm (Figs. 2I, J). Similarly, flow cytometric co-staining of live, E8.25 embryonic cell suspensions with EpCAM and Pdgf receptor alpha (Pdgfra), which is expressed in all E7.5–E8.25 non-notochordal mesoderm (Orr-Urtreger et al., 1992) demonstrates that Pdgfra is exclusively and uniformly expressed in the EpCAM(−) fraction (Fig. 2L). Not only does this finding lend support to the idea that EpCAM is expressed most strongly in endoderm at E8.25, it gives insight into the nature of EpCAM(lo) cells (see Fig. 2A). The E8.25 EpCAM section immunofluorescence analysis (Fig. 2C) shows dim expression in neural cells yet no expression in non-notochordal mesodermal cells. The flow cytometric co-staining of EpCAM and Pdgfra (Fig. 2L) shows that all EpCAM(−) cells express the mesodermal Pdgfra. Thus, it is evident that expression of EpCAM alone distinguishes non-notochordal mesoderm as EpCAM(−), neuroectoderm and notochord as EpCAM(lo) and endoderm as EpCAM(hi) at E8.25. Thus, using EpCAM at E8.25 and later or SBA or FE-JI at E7.5, definitive endoderm can be prospectively distinguished from other germ layer derivatives.

Prospective distinction of definitive and visceral endoderm

The analysis above demonstrates that, in embryos dissected away from extraembryonic components, definitive endoderm can be distinguished from the other germ layers based on cell surface epitopes. However, as previously stated, mice possess extraembryonic endoderm as well as definitive endoderm, and these tissues co-express many genes including Sox17.

Thus, we sought to find epitopes that allow for distinction of live definitive and extraembryonic endoderm. We focused on distinguishing definitive and visceral endoderm and not parietal endoderm because visceral endoderm is morphologically more similar to definitive endoderm. To confirm that both definitive and visceral endoderm can be identified flow cytometrically by EpCAM expression, as expected by immunofluorescence (Fig. 2C), dissociated E8.25 whole embryo, dissected embryonic tissues, or dissociated extraembryonic tissues were co-stained with EpCAM and the extraembryonic endoderm-specific transcription factor Sox7 (Kanai-Azuma et al., 2002). In the whole embryo, Sox7 was expressed in a subset of EpCAM(hi) cells (Fig. 3A), which comprised few embryonic EpCAM(hi) cells (Fig. 3B) and nearly all extraembryonic EpCAM(hi) cells (Fig. 3C). Thus, EpCAM marks both definitive and visceral endoderm.

Next, antibody and lectin screening was performed to identify epitopes expressed at different levels in definitive and visceral endoderm. This analysis revealed that dolichos biflorus lectin (DBA) and SSEA-4 were both flow cytometric markers of visceral endoderm. DBA and SSEA-4 expression were, similarly to Sox7, not found in embryonic EpCAM(hi) cells but found at high levels in extraembryonic EpCAM(hi) cells (Figs. 3D–G). The flow cytometric profiles of DBA and SSEA-4 were confirmed by whole-mount immunohistochemistry and

Table 2
Non-endoderm-enriched transcription factors that distinguish definitive and visceral endoderm

Definitive endoderm enriched (>3-fold)			Visceral endoderm enriched (>3-fold)	
Arnt2	Hoxb9	Pax6	Asb8	Nfatc1
Cdx4	Hoxc4	Pax8	Atp6v0a1	Nfe2l2
Cutl2	Hoxc8	Pbx1	Ehf	Nfla
Dlx2	Hoxd1	Pknox2	Elfl	Pcbd
Dmrta1	Hoxd8	Rfx3	Foxf1A	Ppargc1a
Evx1	Hoxd9	Six1	Foxq1	Runx1
Foxc1	Irx2	Six3	Gata4	Sec14l2
Foxc2	Irx3	Sox11	Gata6	Sox7
Gli3	Irx5	Sox21	Hoxb8	Stat5a
Hey2	Isl1	Sox9	Ipf1	Stat5b
Hoxa1	Meox1	Ssbp2	Irf2bp2	Tcf19
Hoxa3	Morf4lp1	T	Irf6	Tfec
Hoxa9	Mrg1	Tcfap2a	Lass2	Twist1
Hoxb1	Nkx1-2	Tcfap2b	Lhx1	
Hoxb2	Pax1	Tpbg		
Hoxb3	Pax3	Zhx2		

Transcripts characterized by gene ontology as possessing transcription factor activity that are enriched or depleted >3-fold in definitive endoderm as compared to extraembryonic endoderm but that are additionally expressed in other microarray sample groups are listed according to their expression pattern.

section immunofluorescence at E8.25, and they were both largely confined to visceral endoderm (Figs. 3H–K). Thus, from E8.25 whole embryos, definitive endoderm can be prospectively isolated as EpCAM(hi)DBA(−)SSEA-4(−) and visceral endoderm as EpCAM(hi)DBA(+)SSEA-4(+).

Microarray analysis of E8.25 tissues

The ability to isolate definitive and visceral endoderm as well as mesoderm and neuroectoderm/notochord permitted global gene expression analysis of these tissues. Thus, microarrays were performed on E8.25 (2–8 somites) embryonic and extraembryonic tissues using three different isolation techniques. In the first approach, E8.25 midgut and hindgut definitive endoderm was micro-dissected away from mesectoderm. In the second approach, whole embryo (definitive and visceral) Sox17-Venus(hi) endoderm and Sox17-Venus(lo) mesectoderm were flow cytometrically isolated. In the third approach, dissected and sorted extraembryonic EpCAM(hi) DBA(+) visceral endoderm, extraembryonic EpCAM(−) DBA(−) yolk sac mesoderm, embryonic EpCAM(hi)DBA(−) definitive endoderm, embryonic EpCAM(lo)DBA(−) neuroectoderm/notochord, and embryonic EpCAM(−)DBA(−) mesoderm were isolated from E8.25 embryos. Undifferentiated ES cells were also profiled as a control. In all cases, 3–6 biological replicates of amplified cDNA were hybridized to Affymetrix 430_2 microarrays.

In order to confirm the validity of the microarray results, microarray foldchange values of well-characterized marker genes were analyzed. Genes known to be expressed in all endoderm [Sox17 (Kanai-Azuma et al., 2002), FoxA1 (Ang et al., 1993)], visceral endoderm [Hnf4a (Duncan et al., 1994) and Amn (Kalantry et al., 2001)], definitive endoderm and notochord (Shh (Echelard et al., 1993)), notochord (T (Wilkinson et al., 1990)), neuroectoderm [Sox1 (Wood and Episkopou, 1999) and Pax6 (Walther and Gruss, 1991)], mesoderm [Pdgfra (Orr-Urtreger et al., 1992), Meox1 (Candia et al., 1992), and Tbx5 (Chapman et al., 1996)], and yolk sac mesodermal hematopoietic progenitors [Gata1 and Tal1 (Silver and Palis, 1997)] were upregulated in the expected populations (Supplemental Fig. 3). Also of note, CXCR4, a gene utilized to isolate endoderm from differentiating mouse and human ES cells (D'Amour et al., 2005; Yasunaga et al., 2005), is expressed equally strongly in definitive endoderm, neuroectoderm/noto-

chord and mesoderm although expression is significantly weaker in extraembryonic tissues (Supplemental Fig. 4), suggesting that its utility derives primarily in distinguishing embryonic and extraembryonic tissues and not in recognizing endoderm. Other genes used as markers for endoderm in differentiated ES cells such as MixL1, Gsc, T, Cer1, and Hhex [(D'Amour et al., 2005; Kubo et al., 2004; Tada et al., 2005; Yasunaga et al., 2005)] were not found to be expressed at higher levels in E8.25 definitive endoderm than in non-endodermal tissues (Supplemental Table 1).

The profiling of definitive and visceral endoderm afforded an opportunity to analyze the transcriptional regulatory similarities and differences between these tissues. From a gene ontology (GO) search, only four transcription factors were expressed >2-fold higher in definitive endoderm and >3-fold higher in visceral endoderm than any other comparison tissue (Table 1). For these four transcription factors, Sox17 is known to be essential for endoderm formation and maintenance (Kanai-Azuma et al., 2002), FoxA1 is a member of the HNF3 family that is also vital in endoderm formation and maintenance (Dufort et al., 1998), Ripk4 is not essential for endoderm formation but is necessary for proper morphogenesis of the oral, nasal, anal, and esophageal endoderm (Holland et al., 2002), and the role of 5730467H21Rik is unknown (expression analysis confirming endoderm-specific expression of Ripk4 and 5730467H21Rik is presented below).

In addition to these four pan-endodermal transcription factors, transcription factors expressed exclusively in either definitive or visceral endoderm were analyzed. The 10 transcription factors expressed >2-fold higher in definitive endoderm than in any other tissue (Table 1) all have known expression patterns at E8.25 and are involved in components of pharyngeal endoderm organogenesis—none of these genes is pan-endodermal. Four of the eight transcription factors expressed >3-fold higher in visceral endoderm than in all other tissues (Irf5, Nfatc2, Npas2, Vdr; Table 1) are previously uncharacterized. Of the characterized genes, Hnf4a (Chen et al., 1994) and Tcf2 (Barbacci et al., 1999; Coffinier et al., 1999) are necessary for visceral endoderm formation, Cited1 is necessary for proper extraembryonic development (Rodriguez et al., 2004), and FoxA3 is a member of the HNF3 family, which is necessary for visceral endoderm formation (Ang and Rossant, 1994).

A further comparison on transcription factors not exclusive to endoderm was performed, detecting transcriptional differences

Fig. 4. Confirmation of microarray-enriched visceral endoderm-specific cell surface proteins and models of endoderm identification and specification. (A–I) Analysis of visceral endoderm-enriched cell surface proteins. Flow cytometric analysis of live, dissected embryonic (A, D, G) and extraembryonic (B, E, H) cells stained with EpCAM (Y-axis) and Dpp4 (A–B), CD38 (D–E), or Ceacam1 (G–H) on the X-axis. All antibodies display specific staining of extraembryonic (visceral) endoderm. Section immunofluorescence images demonstrating expression of Dpp4 (C), CD38 (F) and Ceacam1 (I) in visceral endoderm (VE) but not in definitive foregut endoderm (FGEn) or hindgut endoderm (HGEn). (G–L) Whole-mount in situ hybridization analysis of Ripk4 (J–L) and Rbm35a (M–O). Both probes display endoderm-specific staining in E8.25 foregut (J, M), E8.25 hindgut (K, N) and E9.5 gut (L, O). (P) Diagram displaying cell surface epitope expression in ES cells (black), visceral endoderm (blue), definitive endoderm (red) and mesoderm/neuroectoderm (green). Using combinations of these markers, all of the listed cell types can be distinguished in embryos and some in cultures of differentiating ES cells. (Q) Diagram depicting the implications of the microarrays of E8.25 tissues on understanding early endoderm specification. The microarrays reveal two classes of transcriptional regulatory groups, an "endoderm identity" group containing of Sox17, FoxA1, Ripk4, and 5730467H21Rik, and an "extraembryonic endoderm cassette" containing Hnf4a, Tcf2, Cited1, Gata4, Gata6, and others. Specification of both extraembryonic and definitive endoderm requires induction of the "endoderm identity" group, yet specification of extraembryonic endoderm but not definitive endoderm specification requires induction of the "extraembryonic endoderm cassette." Details in text. Me=mesoderm, FGEn=foregut endoderm, NF=neural folds, HGEn=hindgut endoderm, No=notochord, VE=visceral endoderm.
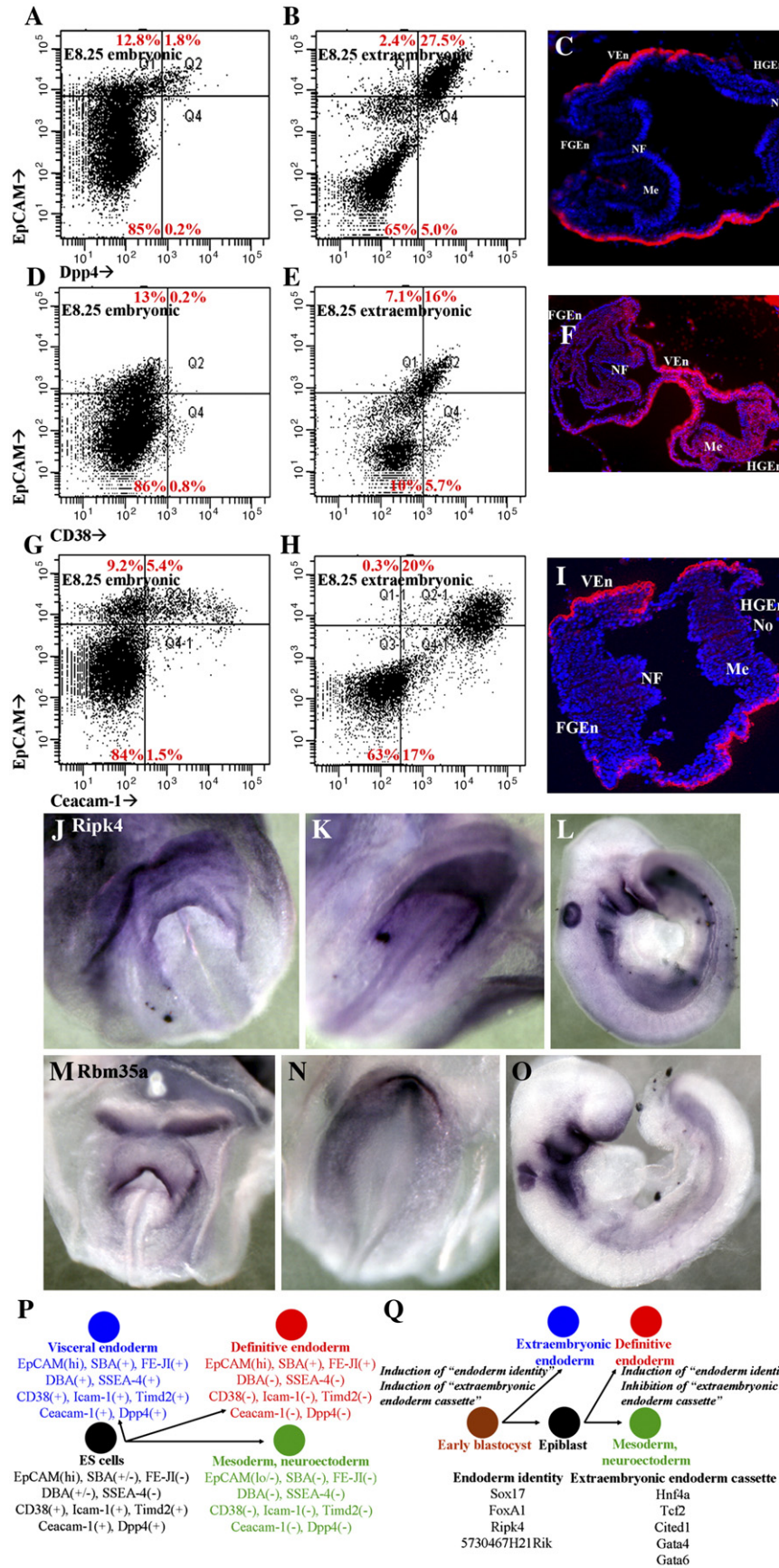
between definitive and visceral endoderm. Transcription factors expressed >3-fold higher in definitive than visceral endoderm but also expressed in non-endodermal tissues included 12 homeobox family transcription factors (HoxA1, HoxA3, HoxA9, HoxB1, HoxB2, HoxB3, HoxB9, HoxC4, HoxC8, HoxD1, HoxD8, HoxD9) and members of other homeobox families involved in embryonic patterning such as paired (Pax1, Pax3, Pax6, Pax8, Pax9), distal-less (Dlx2), Iroquois related (Irx2, Irx3, Irx5), sine oculis-related (Six1, Six3) and SRY-box containing (Sox9, Sox11, Sox21) (Table 2). Transcription factors expressed >3-fold higher in visceral than definitive endoderm but also expressed in non-endodermal tissues included only one homeobox transcription factor (HoxB8) but contained transcription factors such as Gata4, Gata6, Lhx1, and Sox7 that are implicated in endoderm formation in lower vertebrates (Table 2).

An additional GO search was performed on plasma membrane proteins, expanding the list of cell surface proteins allowing prospective isolation of visceral endoderm. Plasma membrane-localized genes upregulated >3-fold in visceral endoderm as compared to all other tissues were identified by GO search. Five candidates from this *in silico* analysis, Icam1, Ceacam1, Dpp4, CD38 and Timd2, were confirmed to be visceral endoderm-enriched at E8.25 by flow cytometry (Figs. 4A–B, D–E, G–H, data not shown) and section immunofluorescence (Figs. 4C, F, I, data not shown). Notably, Icam1, CD38, and Timd2 are also expressed in undifferentiated ES cells, whereas Ceacam1 and Dpp4 are not (data not shown). This analysis expands the list of cell surface epitopes that prospectively distinguish visceral endoderm from all other cells in the E8.25 embryo.

A pan-endodermal signature

A more global analysis of the data was performed to identify endoderm-enriched transcripts and to devise a signature of genes that characterize early endoderm. Since the transcriptional analysis suggested that the vital definitive endoderm transcription factors are also expressed in visceral endoderm, the signature was designed to find genes overexpressed in definitive endoderm versus non-endodermal tissues and not to exclude visceral endoderm expression. Thirty-one genes that were enriched >2-fold in dissected definitive endoderm, in sorted Sox17Venus(hi) endoderm and in embryonic EpCAM(hi)DBA(−) definitive endoderm as compared to all other non-endodermal tissues were chosen for further examination as "endoderm signature" genes (Table 3). Three such genes, *Sox17* (Kanai-Azuma et al., 2002), *FoxA1* (Ang et al., 1993; Monaghan et al., 1993), and *Embigin* (Huang et al., 1990; Sousa-Nunes et al., 2003) have already been shown to have exclusive endodermal expression at E7.5–E9.5, and a fourth, *Tacstd1* is the epitope recognized by EpCAM and has been confirmed by this work to be upregulated in endoderm.

The remaining 27 genes were analyzed for their endoderm specificity at multiple developmental stages. Quantitative PCR (qPCR) was performed on flow cytometrically isolated definitive endoderm and mesectoderm at E7.5, E8.25 and

Table 3
Characterization of "endoderm signature" candidate genes

Endoderm signature genes					
	mRNA/Protein localization verified as endoderm-enriched?	Expressed >2-fold higher in embryonic EpCAM(hi)DBA(−) vs. ES cells?	qPCR confirmation of definitive endoderm enrichment over mesectoderm		
			E7.5	E8.25	E9.5
Previously confirmed endoderm-enriched genes					
Emb	Yes (Sousa-Nounes et al. 2003)	Yes	ND	ND	ND
FoxA1	Yes (Monaghan et al. 1993)	Yes	ND	ND	ND
Sox17	Yes (Kanai-Azuma et al. 2002)	Yes	ND	ND	ND
Tacstd1	Yes (AB)	No	ND	ND	ND
Novel E7.5–E9.5 "endoderm signature" genes					
5730521E12RIK	Yes (ISH)	Yes	Yes	Yes	Yes
AnxA4	Yes (AB)	No	Yes	Yes	Yes
Bnipl	Yes (ISH)	No	Yes	Yes	Yes
Cacna1b	Yes (ISH)	Yes	Yes	Yes	Yes
Cdcp1	Yes (ISH)	Yes	Yes	Yes	Yes
Cldn8	Yes (AB)	Yes	Yes	Yes	Yes
Clic6	Yes (ISH)	Yes	Yes	Yes	Yes
Dsg2	Yes (AB)	No	Yes	Yes	Yes
Krt2-7	Yes (AB)	Yes	Yes	Yes	Yes
Npnt	Yes (ISH)	Yes	Yes	Yes	Yes
Rab15	Yes (ISH)	Yes	Yes	Yes	Yes
Rbm35a	Yes (ISH)	No	Yes	Yes	Yes
Ripk4	Yes (ISH)	Yes	Yes	Yes	Yes
Sh3gl2	Yes (ISH)	No	Yes	Yes	Yes
Spink3	No (untested)	Yes	Yes	Yes	Yes
St14	Yes (ISH)	No	Yes	Yes	Yes
Tmem30b	No (no working probe)	Yes	Yes	Yes	Yes
Tmprss2	Yes (ISH)	Yes	Yes	Yes	Yes
E8.25 endoderm-enriched genes not enriched at all early embryonic timepoints					
5730467H21RIK	Yes (ISH)	Yes	No	Yes	Yes
Crb3	No (no working probe)	No	Yes	Yes	No
Dpp4	No (untested)	Yes	Yes	No	Yes
Ell3	Yes (ISH)	No	No	Yes	Yes
Gprc5c	Yes (ISH)	Yes	No	No	Yes
Prss8	No (untested)	Yes	No	No	Yes
Genes not enriched in E8.25 endoderm					
Igfbp5	No (ISH)	Yes	ND	ND	ND
Spint1	No (ISH)	No	ND	ND	ND
Tmprss13	No (untested)	Yes	No	No	No

Summary of microarray expression, expression pattern at E8.25, and quantitative PCR (qPCR) expression data for 31 genes enriched in microarrays of E8.25 endodermal cells as compared to those of non-endodermal cells. Refer to text for explanation of microarray experimental design. Genes are categorized by their spatial expression as determined by in situ hybridization or antibody staining as well as by their comparative qPCR expression in sorted definitive endoderm vs. mesectoderm cells at E7.5, E8.25 and E9.5. Eighteen genes (*5730521E12Rik, AnxA4, Bnipl, Cacna1b, Cdcp1, Cldn8, Clic6, Dsg2, Krt2-7, Npnt, Rab15, Rbm35a, Ripk4, Sh3gl2, St14,* and *Tmprss2*) were newly found to be endoderm enriched by all criteria tested.

E9.5, and either in situ hybridization or antibody staining was performed at E8.25 and E9.5. It was found that 16 of these genes (*5730521E12Rik*, *AnxA4*, *Bnipl*, *Cacna1b*, *Cdcp1*, *Cldn8*, *Clic6*, *Dsg2*, *Krt2-7*, *Npnt*, *Rab15*, *Rbm35a*, *Ripk4*, *Sh3gl2*, *St14*, and *Tmprss2*) were highly enriched in endoderm by in situ hybridization or antibody staining at E8.25 and E9.5 and were significantly enriched in definitive endoderm as compared to mesectoderm by qPCR at E7.5, E8.25 and E9.5 (Figs. 4J–O, Supplemental Fig. 5, summarized in Table 3). An additional two genes (*Spink3* and *Tmem30b*) were confirmed by qPCR at all stages but were unable to be detected by in situ hybridization (Table 3). The expression patterns reveal that all of these genes are expressed in both definitive and visceral endoderm. Of the other nine endoderm signature candidates, six (5730467H21Rik, Crb3, Dpp4, Ell3, Gprc5c, Prss8) were enriched in endoderm by qPCR at some but not all developmental stages tested, two (Igfbp5 and Spint1) were determined by in situ hybridization to be expressed strongly not only in endodermal regions but in other germ layers (Supplemental Fig. 5), and one (Tmprss13) was not enriched in endoderm by qPCR at any stage (Table 3). Thus, we have identified a set of 22 genes, only three of which were previously known to be endoderm-enriched, which are expressed preferentially in definitive and visceral endoderm throughout early development and can collectively be used as an "endoderm signature."

Application in differentiating embryonic stem cells

To determine the utility of the cell surface proteins and marker genes identified in this work for studies on ES cell differentiation, mouse ES cells were induced to differentiate in monolayer culture or as embryoid bodies (EBs) and were assessed after 7 days of differentiation. Some samples were treated with activin A to induce endoderm formation (Kubo et al., 2004; Yasunaga et al., 2005). Undifferentiated and differentiated samples were stained with antibodies against EpCAM, CD38, and Dpp4, which in E8.25 embryos distinguish definitive endoderm (EpCAM$^+$CD38$^-$Dpp4$^-$), visceral endoderm (EpCAM$^+$CD38$^+$Dpp4$^+$), and mesectoderm (EpCAM$^-$CD38$^-$Dpp4$^-$) and should also allow distinction of undifferentiated ES cells (EpCAM$^+$CD38$^+$Dpp4$^-$) (see Fig. 4P).

As expected, >97% of undifferentiated ES cells were EpCAM$^+$, CD38$^+$, and Dpp4$^-$ and induction of differentiation revealed populations with the surface marker expression profile characteristic of definitive endoderm, visceral endoderm and mesectoderm (Fig. 5A). EpCAM$^+$Dpp4$^+$ cells, presumed to be visceral endoderm, were nearly absent in cultures differentiated in monolayer, whereas EBs possessed ∼4% EpCAM$^+$Dpp4$^+$ cells, a percentage that was significantly reduced by activin treatment (Fig. 5A). In both monolayer and EB differentiation, between 10 and 30% of cells became EpCAM$^+$CD38$^-$, presumed to be definitive endoderm. Activin increased the percentage of these cells in EBs but not monolayer culture (Fig. 5A).

To further investigate endoderm induction from ES cells, qPCR analysis was performed on differentiating ES cells using six pan-endodermal genes (*Cdcp1*, *FoxA1*, *Npnt*, *Rab15*, *Ripk4*, *Tmprss2*) as well as six genes specifically expressed in visceral endoderm in the microarray analysis (*Afp*, *Amn*, *Hnf4a*, *Npas2*, *Slc13a4*, and *Tcf2*). Induction of differentiation in monolayer and EBs led to a strong induction of most pan-endodermal genes (6/6 genes in monolayer, 5/6 genes in EB; Supplemental Fig. 6). Activin treatment enhanced the induction of most of these pan-endodermal genes (5/6 genes for monolayer and EB; Fig. 5B), although this enhancement was modest in degree (1.9-fold in monolayer, 1.6-fold in EB; Fig. 5B). Thus, the "endoderm signature" genes do serve as useful markers of ES cell-derived endoderm induction. Induction of visceral endoderm transcripts, however, was dependent on culture mode: in monolayer differentiation, visceral endoderm gene expression was significantly reduced, whereas such expression was significantly increased in EB differentiation, although activin prevented this increase (Fig. 5B).

To determine whether the cell surface markers utilized accurately reflect production of definitive and visceral endoderm, a statistical correlation analysis was performed. The percentage of cells in each culture condition possessing a visceral endoderm surface marker profile (EpCAM$^+$CD38$^+$Dpp4$^+$) or a definitive/visceral endoderm surface marker profile (EpCAM$^+$CD38$^-$Dpp4$^-$ or EpCAM$^+$CD38$^+$Dpp4$^+$) was plotted against the average level of expression of visceral endoderm or pan-endodermal genes (Figs. 5C–D). The results strikingly showed a nearly perfect correlation (0.988 correlation coefficient) between visceral endoderm cell surface marker expression and gene expression (Fig. 5C), while they revealed weak correlation (0.389 correlation coefficient) between all endoderm cell surface marker expression and gene expression (Fig. 5D). These results imply that the cell surface markers discovered to be uniquely expressed in E8.25 visceral endoderm serve as an accurate proxy for ES cell visceral endoderm induction; however, additional markers will need to be identified to allow for pure prospective distinction of definitive endoderm from ES cells.

Discussion

The potential therapeutic application of ES cell-derived cell types and organs has received much attention in recent years. While random differentiation of ES cells in EBs or in teratomas yields a variety of differentiated cell types, this approach has thus far failed to allow efficient generation of specific cell types. On the other hand, recapitulation of developmental signaling and transition through developmentally relevant cell type intermediates has been shown to permit efficient generation of mature, functional cell types (Wichterle et al., 2002). Whether ES cells have the capacity to differentiate into cell types that closely resemble early embryonic populations such as definitive endoderm is unknown in part because the embryonic equivalents have not been sufficiently characterized. Thus, it was the goal of this report to define the transcriptional pattern of early definitive and visceral endoderm and utilize this information to provide a signature with which to identify endoderm derived from ES cells.

In comparing transcriptional profiles of E8.25 definitive and visceral endoderm, the most striking observation is the shared transcriptional regulation of definitive and visceral endoderm. Both endoderm samples share expression of the key endoderm regulators Sox17 and FoxA1 as well as of the newly characterized Ripk4 and 5730467H21Rik. However, whereas the transcription factors expressed in E8.25 visceral but not definitive endoderm, including Hnf4a, Tcf2, Cited1, Gata4, and Gata6 are crucial for its formation and metabolic function (Bielinska et al., 1999), the transcription factors
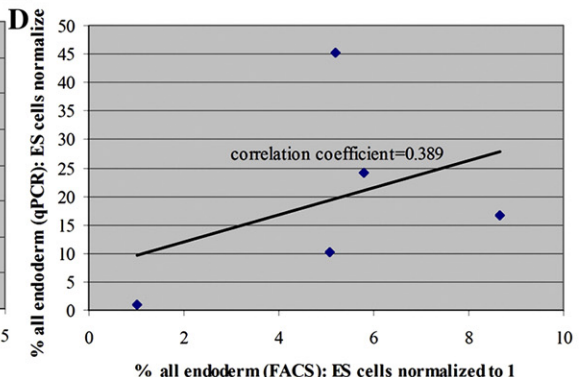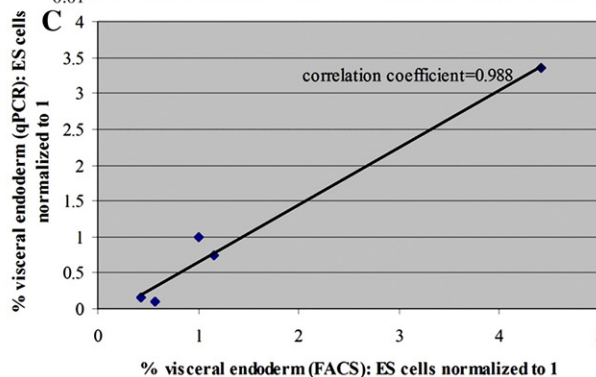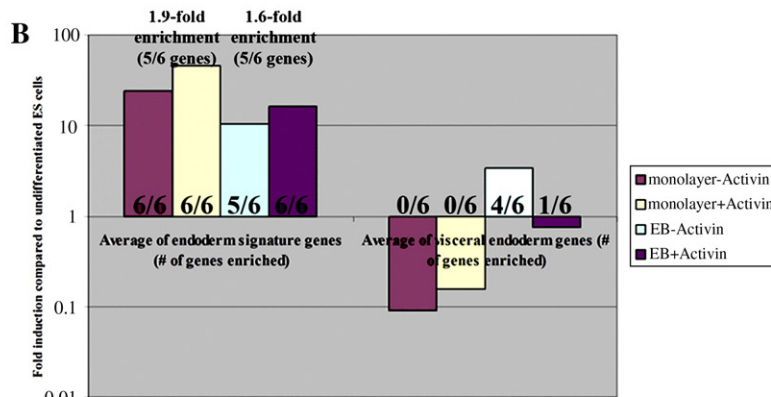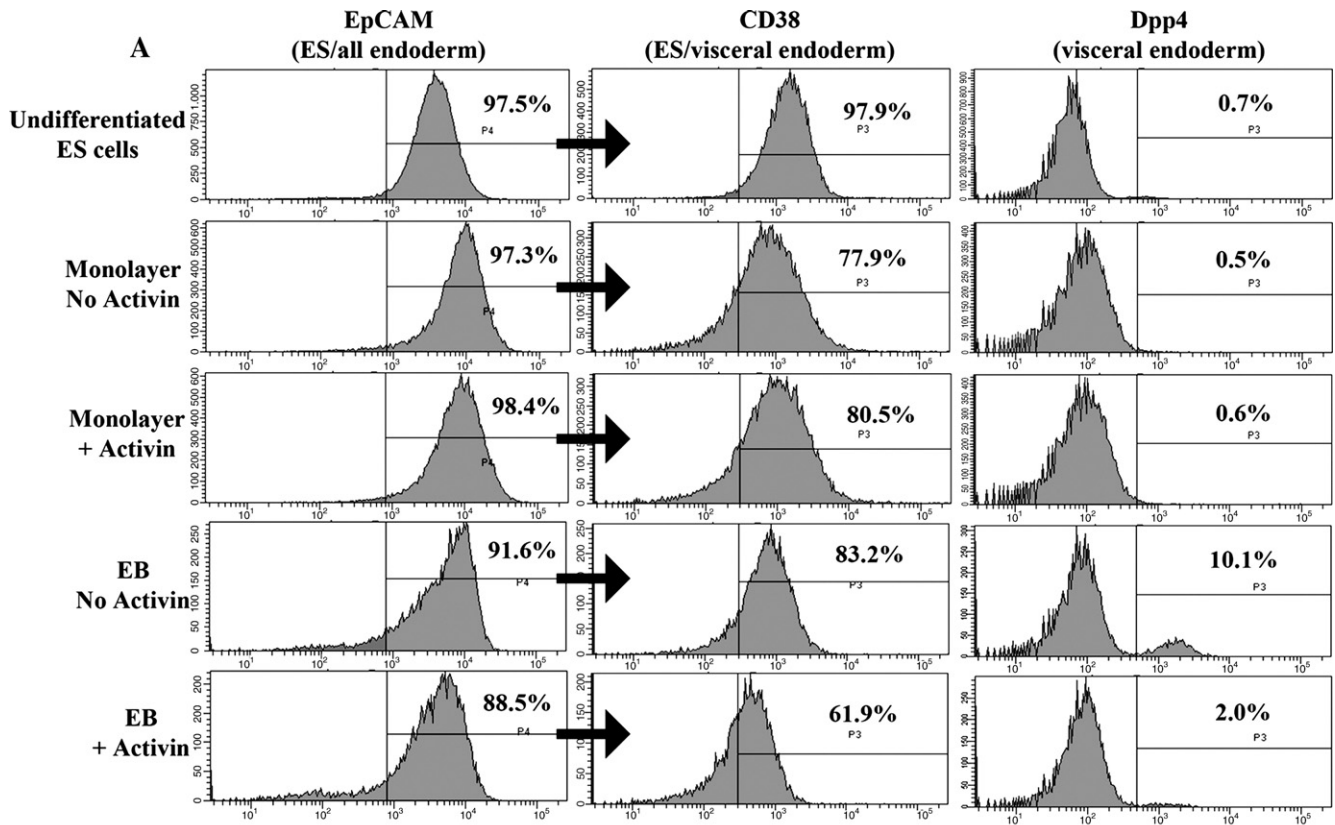
expressed in E8.25 definitive but not visceral endoderm have been demonstrated to act in the endoderm organogenesis process, and a role for such transcription factors in formation of definitive endoderm has not yet been demonstrated.

Visceral endoderm thus shares an "endoderm identity" gene expression cassette with definitive endoderm but contains an additional cassette of transcription factors that regulates its formation and nutritive function in the embryo (Fig. 4Q). These data lead to the question of the evolution of extraembryonic endoderm. The pan-endodermal Sox17 and Hnf3 factors play crucial roles in endoderm development in frogs and zebrafish, but in these species, these factors act downstream of GATA factors (Shivdasani, 2002). In mice, GATA4 and GATA6 are expressed in E8.25 visceral but not definitive endoderm, suggesting that some components of the transcriptional networks controlling endoderm formation are active in extraembryonic but not definitive endoderm formation. Interestingly, many genes in this "extraembryonic endoderm cassette" are expressed later in development in the liver, intestine and pancreas and regulate metabolism in the mature organism. It is enticing to speculate that this extraembryonic endoderm cassette reflects an evolutionarily ancient metabolic regulation system whose expression has been transferred to extraembryonic endoderm in early mammalian embryos.

The common transcriptional machinery in definitive and visceral endoderm implies a similarity in the mechanism of specification of the two tissues. Even though specification and cell fate commitment of extraembryonic endoderm occurs during the early blastocyst stage (~E3.5; Chazaud et al., 2006; Kunath et al., 2005), and specification of definitive endoderm occurs during gastrulation (~E6.5), it is enticing to consider that common signaling events induce *Sox17* and the *HNF3* genes. These signaling events confer "endoderm identity," and extraembryonic endoderm induction presumably involves additional signaling events that induce its "extraembryonic endoderm cassette."

This line of reasoning implies that selective induction of definitive endoderm from ES cells may require inhibition of visceral endoderm. Thus, factors promoting endoderm formation such as those of the Nodal family (Feldman et al., 1998, Tremblay et al., 2000) should be combined with factors that inhibit induction of the extraembryonic endoderm cassette to specifically induce definitive endoderm. Such extraembryonic endoderm-promoting factors are for the most part unknown, although analysis of mutants has recently suggested an involvement of the FGF signaling pathway (Chazaud et al., 2006).

As far as characterizing the "endoderm identity" cassette, our microarray analysis of E8.25 endoderm isolated by three distinct methods has confirmed a group of 22 genes as reliable markers of E7.5–E9.5 endoderm (Table 3). These genes are expressed in definitive and visceral endoderm, as the current analysis did not reveal a set of genes expressed in all definitive endoderm but not in extraembryonic endoderm. This list of "endoderm signature" genes, as well as the list of visceral endoderm-specific transcription factors and cell surface proteins generated from the microarrays, provides a standard to which ES-derived endoderm-like cells should be compared and will provide an answer to whether current differentiation protocols promote differentiation of ES cells into cells closely resembling definitive endoderm or whether they merely allow for the expression of a few select endodermal genes. Preliminary analysis suggests that activin treatment, currently suggested to promote endoderm formation from ES cells (Kubo et al., 2004; Yasunaga et al., 2005), induces definitive endoderm gene expression but inhibits visceral endoderm gene expression (Supplemental Fig. 6). However, since the present gene expression analysis only examined early embryonic timepoints, determining which of the 22 genes are selectively expressed in endoderm at all stages of development and which genes are expressed more broadly later is important for interpreting expression patterns in differentiating ES cells.

An additional objective of this work was to enable flow cytometric prospective isolation of definitive and visceral endoderm from embryos. Just as prospective isolation of adult hematopoietic cell types paved the way for an impressive array of therapeutic and research advances (reviewed in Shizuru et al., 2005), flow cytometric isolation of embryo- and ES-derived cell populations will be invaluable for evaluating developmental potential and for eventually selecting purified cell types and excluding tumorigenic ES cells for therapeutic applications. The ability to prospectively isolate definitive endoderm, visceral endoderm, mesoderm and other cell types from mouse embryos using the antibodies described in this work (Fig. 4P) opens avenues to investigate cell fate specification and commitment.

Utilizing these surface markers to isolate specific populations from differentiating ES cells is also an exciting prospect. In fact, statistical correlation of visceral endoderm-specific cell surface marker expression and transcript expression suggests that the markers utilized in this study allow for clear distinction of ES cell-derived visceral endoderm. However, this correlation analysis also suggests that the cell surface marker profile that in E8.25 embryos is unique to definitive endoderm does not predict ES cell-derived endoderm

Fig. 5. Analysis of cell surface marker expression and gene expression in differentiating ES cells. Expression of cell surface markers in ES cells differentiated for 7 days either in monolayer or in EB in the presence or absence of activin A, as noted. EpCAM expression (left column) is shown for live cells, and CD38 and Dpp4 expression (middle and right columns, respectively) are shown for live, EpCAM$^+$ cells. qPCR expression averages of six endoderm signature genes (*Cdcp1*, *FoxA1*, *Npnt*, *Rab15*, *Rbm35a*, *Tmprss2*) and six visceral endoderm-specific genes (*Afp*, *Amn*, *Hnf4a*, *Npas2*, *Slc13a4*, *Tcf2*) in ES cells differentiated for 7 days either in monolayer or in EB in the presence or absence of activin A, as noted. qPCR expression for each gene is normalized to expression in undifferentiated ES cells, and the average fold difference and the number of genes significantly upregulated are displayed for each condition and gene set. (C–D) Statistical correlation analysis. For each differentiated ES cell sample, the percentage of cells expressing visceral endoderm (C) or any endoderm (D) cell surface marker profile is normalized to the percentages in undifferentiated ES cells and plotted against the percentage of average qPCR gene induction of visceral endoderm (C) or endoderm signature (D) gene sets as compared to undifferentiated ES cells. A trendline has been added, and the correlation coefficient is noted on the graph.

production. A likely reason for this conflict is that an unexpectedly large percentage of cells retain the surface marker expression profile of undifferentiated ES cells (EpCAM$^+$CD38$^+$Dpp4$^-$) even after 7 days of induced differentiation (Supplemental Fig. 6), even though this marker profile is uncharacteristic of any cells within the E7.5–E9.5 mouse embryo (data not shown). This result suggests that protein expression dynamics may differ in ES cell culture than in embryos, which could be a result of differing cell adhesion mechanisms *in vitro* as compared to *in vivo* or of a lack of tight regulation of gene expression in ES cell culture. Thus, prospective isolation of specific populations of cells from differentiating ES cells will require not only an understanding of marker expression in vivo but also careful comparison of protein dynamics in vitro.

Although isolation of a pure population of ES cell-derived definitive endoderm was not possible using the methods presented in this study, a clear strategy for such isolation comes from the microarray analysis. Among the "endoderm signature" genes are several cell surface proteins specifically enriched in E8.25 definitive and visceral endoderm and absent in ES cells (Cacna1b, Cldn8, Npnt, Tmprss2). We are currently engaged in an effort to produce flow cytometrically functional antibodies to these proteins. These antibodies should be valuable for purification and characterization of endoderm differentiated from mouse and human ES cells.

One additional challenge of ES cell differentiation that this work underscores involves timing. Germ layer tissue isolated from different developmental stages has molecular differences: EpCAM is downregulated in neuroectoderm/notochord gradually from E7.5 to E9.5, FE-J1 is only expressed in endoderm from E7.5 to E8.25, and some genes isolated as endodermally enriched at E8.25 are not enriched in all stages between E7.5 and E9.5. These temporal changes no doubt reflect underlying differences in cells from these developmental stages, as is clearly manifested by the fact that distinct anterior–posterior regions of the endoderm begin to express organ-specific transcripts and proteins at E8.25 8–10 somite stages (Deutsch et al., 2001; Serls et al., 2005) and display pre-patterning prior to this time. In the embryo, unspecified definitive endoderm exists only transiently, and it is unclear whether these cells have an internal clock mechanism that regulates assumption of organ-specific genes or whether early definitive endoderm could be prevented from differentiating if removed from its developmental context. Regardless, attempts to differentiate ES cells into mature cell types through germ layer intermediates must be attuned to issues of timing and stage-specific gene expression, as competence to respond to developmentally appropriate signals is likely to be time-dependent.

This work represents a step forward in understanding the nature of early definitive and visceral endoderm. The markers identified will guide experiments aimed at understanding the commitment of endoderm and its potency as the precursor of all gut-derived organs. The data should also help direct experiments aimed at ES cell differentiation into therapeutically relevant endodermal derivatives.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2007.01.011.

References

Ang, S.L., Rossant, J., 1994. HNF-3 beta is essential for node and notochord formation in mouse development. Cell 78, 561–574.

Ang, S.L., Wierda, A., Wong, D., Stevens, K.A., Cascio, S., Rossant, J., Zaret, K.S., 1993. The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. Development 119, 1301–1315.

Barbacci, E., Reber, M., Ott, M.O., Breillat, C., Huetz, F., Cereghini, S., 1999. Variant hepatocyte nuclear factor 1 is required for visceral endoderm specification. Development 126, 4795–4805.

Belo, J.A., Bouwmeester, T., Leyns, L., Kertesz, N., Gallo, M., Follettie, M., De Robertis, E.M., 1997. Cerberus-like is a secreted factor with neutralizing activity expressed in the anterior primitive endoderm of the mouse gastrula. Mech. Dev. 68, 45–57.

Bielinska, M., Narita, N., Wilson, D.B., 1999. Distinct roles for visceral endoderm during embryonic mouse development. Int. J. Dev. Biol. 43, 183–205.

Candia, A.F., Hu, J., Crosby, J., Lalley, P.A., Noden, D., Nadeau, J.H., Wright, C.V., 1992. Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially expressed during early mesodermal patterning in mouse embryos. Development 116, 1123–1136.

Chapman, D.L., Garvey, N., Hancock, S., Alexiou, M., Agulnik, S.I., Gibson-Brown, J.J., Cebra-Thomas, J., Bollag, R.J., Silver, L.M., Papaioannou, V.E., 1996. Expression of the T-box family genes, *Tbx1-Tbx5*, during early mouse development. Dev. Dyn. 206, 379–390.

Chazaud, C., Yamanaka, Y., Pawson, T., Rossant, J., 2006. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Dev. Cell 10, 615–624.

Chen, W.S., Manova, K., Weinstein, D.C., Duncan, S.A., Plump, A.S., Prezioso, V.R., Bachvarova, R.F., Darnell Jr., J.E., 1994. Disruption of the *HNF-4* gene, expressed in visceral endoderm, leads to cell death in embryonic ectoderm and impaired gastrulation of mouse embryos. Genes Dev. 8, 2466–2477.

Coffinier, C., Thepot, D., Babinet, C., Yaniv, M., Barra, J., 1999. Essential role for the homeoprotein vHNF1/HNF1beta in visceral endoderm differentiation. Development 126, 4785–4794.

D'Amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., Baetge, E.E., 2005. Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat. Biotechnol. 23, 1534–1541.

David, N.B., Rosa, F.M., 2001. Cell autonomous commitment to an endodermal fate and behaviour by activation of Nodal signalling. Development 128, 3937–3947.

Deutsch, G., Jung, J., Zheng, M., Lora, J., Zaret, K.S., 2001. A bipotential precursor population for pancreas and liver within the embryonic endoderm. Development 128, 871–881.

Dufort, D., Schwartz, L., Harpal, K., Rossant, J., 1998. The transcription factor HNF3beta is required in visceral endoderm for normal primitive streak morphogenesis. Development 125, 3015–3025.

Duncan, S.A., Manova, K., Chen, W.S., Hoodless, P., Weinstein, D.C., Bachvarova, R.F., Darnell Jr., J.E., 1994. Expression of transcription factor HNF-4 in the extraembryonic endoderm, gut, and nephrogenic tissue of the developing mouse embryo: HNF-4 is a marker for primary endoderm in the implanting blastocyst. Proc. Natl. Acad. Sci. U. S. A. 91, 7598–7602.

Echelard, Y., Epstein, D.J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J.A., McMahon, A.P., 1993. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Cell 75, 1417–1430.

Farr, A., Nelson, A., Truex, J., Hosier, S., 1991. Epithelial heterogeneity in the murine thymus: a cell surface glycoprotein expressed by subcapsular and medullary epithelium. J. Histochem. Cytochem. 39, 645–653.

Feldman, B., Gates, M.A., Egan, E.S., Dougan, S.T., Rennebeck, G., Sirotkin, H.I., Schier, A.F., Talbot, W.S., 1998. Zebrafish organizer development and germ-layer formation require nodal-related signals. Nature 395, 181–185.

Fenderson, B.A., O'Brien, D.A., Millette, C.F., Eddy, E.M., 1984. Stage-specific expression of three cell surface carbohydrate antigens during murine spermatogenesis detected with monoclonal antibodies. Dev. Biol. 103, 117–128.

Gu, G., Wells, J.M., Dombkowski, D., Preffer, F., Aronow, B., Melton, D.A., 2004. Global expression analysis of gene regulatory pathways during endocrine pancreatic development. Development 131, 165–179.

Heasman, J., Wylie, C.C., Hausen, P., Smith, J.C., 1984. Fates and states of determination of single vegetal pole blastomeres of *X. laevis*. Cell 37, 185–194.

Holland, P., Willis, C., Kanaly, S., Glaccum, M., Warren, A., Charrier, K., Murison, J., Derry, J., Virca, G., Bird, T., Peschon, J., 2002. RIP4 is an ankyrin repeat-containing kinase essential for keratinocyte differentiation. Curr. Biol. 12, 1424–1428.

Huang, R.P., Ozawa, M., Kadomatsu, K., Muramatsu, T., 1990. Developmentally regulated expression of embigin, a member of the immunoglobulin superfamily found in embryonal carcinoma cells. Differentiation 45, 76–83.

Hudson, C., Clements, D., Friday, R.V., Stott, D., Woodland, H.R., 1997. Xsox17alpha and-beta mediate endoderm formation in *Xenopus*. Cell, vol. 91, pp. 397–405.

Kalantry, S., Manning, S., Haub, O., Tomihara-Newberger, C., Lee, H.G., Fangman, J., Disteche, C.M., Manova, K., Lacy, E., 2001. The amnionless gene, essential for mouse gastrulation, encodes a visceral-endoderm-specific protein with an extracellular cysteine-rich domain. Nat. Genet. 27, 412–416.

Kanai-Azuma, M., Kanai, Y., Gad, J.M., Tajima, Y., Taya, C., Kurohmaru, M., Sanai, Y., Yonekawa, H., Yazaki, K., Tam, P.P., Hayashi, Y., 2002. Depletion of definitive gut endoderm in Sox17-null mutant mice. Development 129, 2367–2379.

Kubo, A., Shinozaki, K., Shannon, J.M., Kouskoff, V., Kennedy, M., Woo, S., Fehling, H.J., Keller, G., 2004. Development of definitive endoderm from embryonic stem cells in culture. Development 131, 1651–1662.

Kumar, M., Jordan, N., Melton, D., Grapin-Botton, A., 2003. Signals from lateral plate mesoderm instruct endoderm toward a pancreatic fate. Dev. Biol. 259, 109–122.

Kunath, T., Arnaud, D., Uy, G.D., Okamoto, I., Chureau, C., Yamanaka, Y., Heard, E., Gardner, R.L., Avner, P., Rossant, J., 2005. Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. Development 132, 1649–1661.

Lawson, K.A., Pedersen, R.A., 1987. Cell fate, morphogenetic movement and population kinetics of embryonic endoderm at the time of germ layer formation in the mouse. Development 101, 627–652.

Lawson, K.A., Meneses, J.J., Pedersen, R.A., 1986. Cell fate and cell lineage in the endoderm of the presomite mouse embryo, studied with an intracellular tracer. Dev. Biol. 115, 325–339.

Lawson, K.A., Meneses, J.J., Pedersen, R.A., 1991. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. Development 113, 891–911.

Monaghan, A.P., Kaestner, K.H., Grau, E., Schutz, G., 1993. Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. Development 119, 567–578.

Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., Miyawaki, A., 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nat. Biotechnol. 20, 87–90.

Orr-Urtreger, A., Bedford, M.T., Do, M.S., Eisenbach, L., Lonai, P., 1992. Developmental expression of the alpha receptor for platelet-derived growth factor, which is deleted in the embryonic lethal Patch mutation. Development 115, 289–303.

Rodriguez, T.A., Sparrow, D.B., Scott, A.N., Withington, S.L., Preis, J.I., Michalicek, J., Clements, M., Tsang, T.E., Shioda, T., Beddington, R.S., Dunwoodie, S.L., 2004. Cited1 is required in trophoblasts for placental development and for embryo growth and survival. Mol. Cell. Biol. 24, 228–244.

Sasaki, H., Hogan, B.L., 1993. Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. Development 118, 47–59.

Serls, A.E., Doherty, S., Parvatiyar, P., Wells, J.M., Deutsch, G.H., 2005. Different thresholds of fibroblast growth factors pattern the ventral foregut into liver and lung. Development 132, 35–47.

Shivdasani, R.A., 2002. Molecular regulation of vertebrate early endoderm development. Dev. Biol. 249, 191–203.

Shizuru, J.A., Negrin, R.S., Weissman, I.L., 2005. Hematopoietic stem and progenitor cells: clinical and preclinical regeneration of the hematolymphoid system. Annu. Rev. Med. 56, 509–538.

Silver, L., Palis, J., 1997. Initiation of murine embryonic erythropoiesis: a spatial analysis. Blood 89, 1154–1164.

Sousa-Nunes, R., Rana, A.A., Kettleborough, R., Brickman, J.M., Clements, M., Forrest, A., Grimmond, S., Avner, P., Smith, J.C., Dunwoodie, S.L., Beddington, R.S., 2003. Characterizing embryonic gene expression patterns in the mouse using nonredundant sequence-based selection. Genome Res. 13, 2609–2620.

Tada, S., Era, T., Furusawa, C., Sakurai, H., Nishikawa, S., Kinoshita, M., Nakao, K., Chiba, T., Nishikawa, S., 2005. Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture. Development 132, 4363–4374.

Tremblay, K.D., Zaret, K.S., 2005. Distinct populations of endoderm cells converge to generate the embryonic liver bud and ventral foregut tissues. Dev. Biol. 280, 87–99.

Tremblay, K.D., Hoodless, P.A., Bikoff, E.K., Robertson, E.J., 2000. Formation of the definitive endoderm in mouse is a Smad2-dependent process. Development 127, 3079–3090.

Walther, C., Gruss, P., 1991. *Pax-6*, a murine paired box gene, is expressed in the developing CNS. Development 113, 1435–1449.

Wells, J.M., Melton, D.A., 1999. Vertebrate endoderm development. Annu. Rev. Cell Dev. Biol. 15, 393–410.

Wells, J.M., Melton, D.A., 2000. Early mouse endoderm is patterned by soluble factors from adjacent germ layers. Development 127, 1563–1572.

Wichterle, H., Lieberam, I., Porter, J.A., Jessell, T.M., 2002. Directed differentiation of embryonic stem cells into motor neurons. Cell 110, 385–397.

Wilkinson, D.G., Nieto, M.A., 1993. Detection of messenger RNA by in situ hybridization to tissue sections and whole mounts. Methods Enzymol. 225, 361–373.

Wilkinson, D.G., Bhatt, S., Herrmann, B.G., 1990. Expression pattern of the mouse *T* gene and its role in mesoderm formation. Nature 343, 657–659.

Wood, H.B., Episkopou, V., 1999. Comparative expression of the mouse *Sox1*, *Sox2* and *Sox3* genes from pre-gastrulation to early somite stages. Mech. Dev. 86, 197–201.

Wylie, C.C., Snape, A., Heasman, J., Smith, J.C., 1987. Vegetal pole cells and commitment to form endoderm in *Xenopus laevis*. Dev. Biol. 119, 496–502.

Yang, Y., Seed, B., 2003. Site-specific gene targeting in mouse embryonic stem cells with intact bacterial artificial chromosomes. Nat. Biotechnol. 21, 447–451.

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