Transcriptional wiring for establishing cell lineage specification at the blastocyst stage in cattle

Running title: Transcriptomic profile of bovine blastocysts

Summary sentence: The first cell segregation into inner cell mass and trophectoderm lineages in bovine blastocysts is maintained in a bovine-specific manner, unlike in mouse blastocysts.

Hiroaki Nagatomo3, Shinjiro Kagawa3, Yasunori Kishi4, Tetsuya Takuma5, Ayari Sada3, Ken-ichi Yamanaka6, Yasuyuki Abe7, Yasuhiko Wada6, Masashi Takahashi3, Tomohiro Kono4, and Manabu Kawahara2,3

3Laboratory of Animal Breeding and Reproduction, Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan, 4Department of BioScience, Tokyo University of Agriculture, Setagaya-ku, Tokyo 156-8502, Japan, 5Saga Prefectural Livestock Experiment Station, Takeo, Saga 849-2315, Japan

6Faculty of Agriculture, Saga University, 1 Honjo, Saga 840-8502, Japan, 7Department of Biochemical Engineering, Graduate School of Science and Engineering, Yamagata University, Yonezawa 992-8510, Japan

Key words: bovine, blastocyst, cell lineage specification, gene expression

Correspondence:
Dr. Manabu Kawahara, Ph.D.
Laboratory of Animal Breeding and Reproduction, Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan
Tel. & Fax: +81 11 706 2541
E-mail address: k-hara@anim.agr.hokudai.ac.jp

Grant support:
This work was supported by Grant-in-Aid for Research on Priority Area to Young Scientists (B) to M. K. (No. 24780265) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.
ABSTRACT

Mice and cattle use distinct pathways for the first cell segregation into inner cell mass (ICM) and trophoderm (TE) lineages at the blastocyst stage. However, limited knowledge is available regarding reliable transcriptional networks that orchestrate complex developmental processes at this stage in nonrodent species. In order to elucidate the site-dominant transcriptomic properties of bovine blastocysts, we separated cell samples into the ICM and TE using both mechanical and chemical methods and performed in silico prescreening for candidate genes that were site-dominantly expressed in bovine blastocysts. We further performed quantitative RT-PCR and in situ hybridization using the site-specific cell samples. As a result, we identified 7 ICM-dominant genes and 5 TE-dominant genes not yet found in earlier studies. Our findings provide novel insights into the mechanism of cell-fate specification in the pre-implantation bovine embryo.
INTRODUCTION

In mammalian development, the first cell fate decisions are clearly specified during blastocyst formation, when the trophectoderm (TE) surrounds the inner cell mass (ICM), forming a fluid-filled cavity named the blastocoel, which is easily recognized. The ICM gives rise to the embryo proper and other extra-embryonic tissues. Meanwhile, the TE becomes the embryonic part of the placenta. In the mouse embryo, the TE can be classified as 2 TE types: 1) the cells that surround the ICM directly, or the polar trophectoderm (pTE); and 2) the cells that cover the blastocoelic cavity and are not in close contact with the ICM, or the mural trophectoderm (mTE). Differential transcriptional regulation between blastomeres toward either the ICM or TE lineages occurs until the blastocyst stage. Specific transcription factors, i.e., Oct4, Cdx2, Nanog, Sox2, Tead4, and various other types of GATA-family transcription factors, control the cell specification and maintenance of the ICM and TE in the mouse blastocyst [1-6]. Among these, the best-known transcription factors are Oct4 and Cdx2. Oct4 is required for maintenance of ICM cell pluripotency, for which transcription is restricted to the ICM after initiation of blastocyst formation [1]. Embryos harboring homozygous Oct4 mutants can develop to the blastocyst stage, but the ICM cells lack pluripotency, and most blastomeres dominantly differentiate into the extra-embryonic trophoblast lineages [7]. Cdx2 is also known as a TE-specific transcription factor, and genetic deletion of Cdx2 resulted in failure to maintain TE cells [5]. Oct4 and Cdx2 mutually inhibit gene expression, resulting in the site-specific transcription of target genes in blastocysts [8].

In the mouse blastocyst, molecular interactions between essential site-specific transcription factors and ICM/TE positional information has been solved on the basis of numerous genetic approaches [4, 5, 7, 9-15]. However, these mechanisms through which key transcription factors control and maintain the first segregation of the ICM and TE lineages may be exclusive to a limited number of species with early implantation around day 5.5 postcoitus, such as mice. Actually, this rodent model does not necessarily represent all mammalian species. As reported by Berg et al., in cattle, OCT4 transcripts are never restricted to the ICM of bovine blastocysts, and additionally, the effects of the transcription factor Tcfap2 on Oct4
repression in mice was distinct from that in cattle [16]. Therefore, the circumstances surrounding the first segregation into ICM/TE cell lineages in the blastocyst stage are unclear, particularly in nonrodent species. This may be due to the overwhelming lack of reliable transcriptional information on bovine blastocysts as compared to mouse blastocysts. Recently, with the development of transcriptome analyses such as microarrays and next-generation sequencing, global gene expression studies using bovine blastocysts have been increasing gradually [17, 18]. However, while there may be a great deal of information to be gleaned from the transcriptomic data, the meaning of this information is not known because the emerging information has not been experimentally tested in order to fully interpret the transcriptomic analyses, for example, by quantitative RT-PCR of separated ICM and TE cell samples and in situ hybridization for the prediction of precise spatiotemporal expression patterns. Moreover, the particularity of bovine embryos is reflected in the lack of precedent for successful establishment of embryonic stem (ES) cells, trophoblast stem (TS) cells, and even induced pluripotent stem (iPS) cells in cattle, potentially due to insufficient information on reliable marker genes in the ICM and TE cell lineages in cattle.

Here, we determined which genes were expressed in a site-dominant manner in the ICM and TE by completely separating ICM and TE cell samples at the expanded blastocyst stage by using cattle as our nonrodent model. We validated the data reliability by quantitative RT-PCR using divided cell samples and whole-mount in situ hybridization for individual candidate genes. This approach can not only determine newfound ICM/TE lineage predominant genes, but also provide novel insights into whether the positional differences in the TE between polar and mural sides in blastocysts could affect transcription in bovine blastocysts. Our findings provided clear evidence that the ICM and TE lineages in bovine blastocysts were maintained in a bovine-specific manner, unlike in mouse blastocysts, and will yield a deeper understanding of the properties of nonrodent mammalian blastocysts. Thus, this study, based on a nonrodent genetic model, may enable further characterization of diversified developmental programs during early embryonic stages, including blastocyst formation, in mammals.
MATERIALS AND METHODS

Bovine in vitro fertilization (IVF) and embryo culture

Bovine oocyte retrieval, in vitro oocyte maturation, fertilization, and subsequent in vitro bovine embryo culture were performed by the modification of methods described in a previous study[19]. Briefly, cumulus-oocyte complexes (COCs) were aspirated from 3- to 8-mm follicles of slaughterhouse-derived ovaries. COCs with an intact cumulus oophorus were matured in TCM-199 (Gibco, Grand Island, NY, USA) containing 10 μM cysteamine (Sigma Aldrich, St. Louis, MO, USA), 10% (v/v) fetal bovine serum (FBS; PAA Laboratories, QLD, Australia), 0.5 mg/mL FSH (Kyoritsu Seiyaku Corp., Tokyo, Japan), 100 U/mL penicillin (Nacalai Tesque, Inc, Kyoto, Japan), and 100 U/mL streptomycin (Nacalai Tesque, Inc) in each 100-μL droplet covered with liquid paraffin (Nacalai Tesque) and were cultured at 38.5°C in a humidified atmosphere of 5% CO₂ and air for 20–22 hours. In vitro matured oocytes were transferred to Brackett and Oliphant (BO) medium [20] containing 2.5 mM theophylline (Wako Pure Chemical Industries, Ltd., Osaka, Japan). Subsequently, frozen-thawed semen was centrifuged at 600 × g for 8 minutes in BO medium, and the spermatozoa were added to the COCs at a final concentration of 5 × 10⁶ cells/mL. After 18 hours of incubation, presumptive zygotes were denuded and cultured in synthetic oviduct fluid (SOF) medium supplemented with 10 μg/mL insulin (Sigma Aldrich), 1 mg/mL polyvinyl alcohol (Sigma Aldrich), and 10 μM cysteamine (Sigma Aldrich) at 38.5°C in a humidified atmosphere of 5% CO₂. The embryos were cultured until expanded blastocysts at day 8 were harvested and used for subsequent experiments.

Isolation of ICM and TE from bovine blastocysts

We prepared 4 types of cell samples from expanded blastocysts through both mechanical and chemical methods: whole blastocysts (Whole), ICM and polar TE (ICM + pTE), mural TE (mTE), and pure ICM (pICM). In order to mechanically isolate the ICM + pTE and mTE from blastocysts, blastocysts were dissected into 2 types of cells, the ICM + pTE and mTE, using a micromanipulator equipped with a microsurgical blade (Feather, Osaka, Japan) under an inverted microscope (Olympus, Tokyo, Japan),
resulting in removal of the zona pellucida from dissected embryos. Meanwhile, chemical isolation of pICM was performed by modification of the detergent procedure for differential staining as reported by Thouas et al. [21]. In brief, the zona pellucida was removed, and the blastocysts were mildly treated with detergent solution containing 0.2% (v/v) Triton X-100 in PBS at room temperature for approximately 2 minutes in order to remove trophoblasts from whole blastocysts and collect pure cell samples consisting of ICMs. After detergent treatment, trophoblasts were gently detached from the blastocysts by pipetting. After removal of trophoblasts, isolated pICMs were used for subsequent experiments.

**Immunostaining of pre-implantation embryos**

Immunofluorescent staining of whole blastocysts and dissected cell samples (ICM + pTE, mTE, and pICM) was carried out by following standard protocols. The samples were fixed in 4% (w/v) paraformaldehyde in PBS for 60 minutes and then permeabilized with 0.2% (v/v) Triton-X (Sigma Aldrich) in PBS for 60 minutes at room temperature. Subsequently, the samples were washed 3 times in washing solution (0.1% [v/v] Triton-X, 0.3% [w/v] BSA in PBS), blocked with blocking solution (0.5% [w/v] BSA, 1% [w/v] skim milk in PBT) for 60 minutes, and incubated with primary anti-CDX2 antibodies (mu392uc, mouse monoclonal; 1:300; BioGenex, San Ramon, CA, USA]) diluted in blocking solution at 37°C overnight. After washing 3 times in washing solution for 30 minutes, the samples were incubated with the following secondary antibodies [Alexa Fluor 488 goat anti-mouse IgG (A11001; 1:400; Invitrogen, Eugene, OR, USA]) diluted in blocking solution for 30 minutes at room temperature. Nuclei were visualized by staining with 25 μg/mL Hoechst33342 (Sigma) and mounted before observation. Fluorescent signals were visualized using an epifluorescence microscope (Nikon, Tokyo, Japan).

**Microarray analysis**

All the 4 types of samples described above were obtained from 15 different blastocysts per replication, namely, 5 embryos for Whole, 5 for both of ICM + pTE and mTE, and 5 for pICM, which were
used for 1 analysis. We performed microarray analysis (n = 3) using small quantities of total RNA from samples according to the methods optimized by Fukuda et al. [22]. Total RNA was extracted from each sample in 11 μL RNase-free water using an RNeasy Micro Kit (Qiagen, Valencia, CA, USA). The Two-Cycle Eukaryotic Target Labeling Kit (Affymetrix, Santa Clara, CA, USA) was used for synthesizing cRNA, starting from 9 μL of total RNA solution. The first cycle of amplification was conducted in a 65-μL reaction mixture. After the quality of the amplified product was verified by Experion capillary electrophoresis (Bio-Rad, Hercules, CA, USA), 10 μg of fragmented cRNA was hybridized to a GeneChip Bovine Genome 430 2.0 Array (Affymetrix) containing 24,027 probe sets. GeneChip Operating Software (GCOS) version 1.3 (Affymetrix) output files were then loaded into GeneSpring v7.3 (Agilent Technologies, Santa Clara, CA) with per-chip normalization to the 50th percentile and per-gene normalization to the median expression level of the control blastocysts. The filtered genes were used for one-way ANOVA with the post-hoc test using Tukey’s honest significant difference test, and the cutoff value used to identify differentially expressed genes in our study was a false discovery rate of 5%. Principal component analysis (PCA) was employed to analyze the gene expression patterns of all the samples. Hierarchical clustering was performed with the Pearson correlation for measurement of similarity and clustering algorithm with average linkage. The genes were analyzed by gene ontology analysis using FatiGO at Babelomics (www.fatigo.org). Three independent experiments were replicated for each type of sample, namely, Whole, ICM + pTE, mTE, and pICM.

**Whole-mount in situ hybridization (WISH)**

All bovine cDNA fragments encoding the following 30 transcripts were derived from PCR amplifications, sequenced, and cloned into pGEM-T Easy vectors (Promega, Madison, WI, USA): OCT4 (POU5F1), NANOG, FGF4, ID1, ID2, NASP, TEX9, IFT122, FYN, HNF4A, IFITM1, IRF1, CIITA, FGF2, FGFR2, KIAA0664, SMARCA, SOX2, CDX2, GATA3, ZFX, GATA2, ELF3, CTGF, SCUBE2, PDGF, ATF3, DLX4, TFAP2A, and IFNT. The primers used for this analysis are shown in Supplemental Table 1. In vitro
transcription with DIG-labeling was carried out using the T7 RNA polymerase for sense and antisense probes (DIG RNA Labeling Mix, Roche Diagnostics, Basel, Switzerland). All examined antisense probes corresponded to GenBank sequences (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The expanded blastocysts at day 8 were fixed and stored in 4% (w/v) paraformaldehyde (Wako) in PBS with 0.1% (v/v) Tween-20 (PBT) at 4°C for 30 minutes to overnight. Fixed embryos were treated with proteinase K (Sigma) and placed in 4% (w/v) paraformaldehyde, 0.2% (v/v) EM-grade glutaraldehyde in PBT at room temperature for 20 minutes. Subsequently, the blastocysts were placed in hybridization buffer (4× SSC [pH 7.0], 50% [v/v] deionized formamide [Wako], 100 μg/mL heparin [Wako], 100 μg/mL salmon sperm DNA [Wako], 2× Denhardt’s solution, and 0.1% [v/v] Tween-20 [Wako]) at 60–70°C for 3–8 hours. After these pretreatments, the blastocysts were subjected to whole-mount in situ hybridization with DIG-labeled riboprobes in hybridization buffer overnight at the same temperature as prehybridization, followed by washing with buffer (50% [v/v] formamide, 2× SSC, 0.1% [v/v] Tween-20) 3 times at the hybridization temperature. According to the manufacturer’s recommended procedure to detect DIG-labeled probes, samples were incubated in NBT/BCIP solution (Roche) until coloration was obtained after incubation with anti-DIG antibodies (Roche). The hybridized embryos were observed in PBS containing 1 mM EDTA and 20% (v/v) glycerol for bright-field photographing under an inverted microscope (AMEX-1200, AMG, Bothell, WA, USA).

**Quantitative RT-PCR analysis**

Total RNA from 5 blastocysts and separated cell samples per replication was isolated using an RNeasy Micro kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. RNA concentration were assessed using NanoDrop (Thermo Fisher Scientific, Wilmington, DE, USA), and thereafter, quantity of the extracted RNA were standardized at a concentration equal to the lowest value among calculated samples per one replication. After standardizing RNA quantity, cDNA was then synthesized using the ReverTra Ace qPCR RT Master Mix (TOYOBO, Osaka, Japan) including a mix of
random (12 mer) and oligo-dT primers in a reaction solution (20 µL) containing total RNA from the cell samples. Finally, we performed quantitative analysis (n = 3) of gene expression using real-time PCR (LightCycler, Roche Applied Science) after preparing the reaction mixtures in Thunderbird Sybr qPCR Mix (TOYOBO). The primers used for the analysis are described in Supplemental Table 1. Each transcript levels were calculated relative to the transcription of the housekeeping gene \textit{H2AFZ} in every sample. We further confirmed expression level of \textit{H2AFZ} among the standardized samples (Supplemental Fig. 2).

\textit{Statistical analysis}

Data on mRNA expression determined by quantitative RT-PCR were analyzed using one-way analysis of variance (ANOVA) and Fisher’s protected least significant difference test by using the statistical analysis software Statview (Abacus Concepts, Inc., Berkeley, CA, USA). A \( p \)-value of less than 0.05 was considered significant.
RESULTS

Transcriptomic analysis using site-specific cell samples in bovine blastocysts

In order to elucidate the site-specific transcriptomic properties of bovine blastocysts, we prepared 4 types of cell samples from expanded blastocysts: Whole, ICM + pTE, mTE, and pICM (Fig. 1A and Supplemental Fig. 1A). We first counted trophoblasts and all cells per each collected cell sample by CDX2 immunostaining and Hoechst staining to clarify the ratio of ICM to TE cells (Fig. 1B and C). The proportion of TE cells in Whole was 65.8%, whereas that in ICM + pTE decreased to 43.3%. All examined cells in mTE and pICM were CDX2-positive and CDX2-negative, respectively, showing that mTE and pICM samples were separated into trophoblasts and ICM cells, respectively. These results demonstrated that mechanical dissection by micromanipulation allowed us to collect only trophoblasts (mTE), but in order to acquire unmixed origin ICM cell, the mechanical method was unsuitable. Therefore, we further prepared a genuine ICM sample, the pICM, derived from blastocysts whose trophoblasts were chemically treated with a surfactant, and regarded ICM + pTE as cell samples showing intermediate transcriptional patterns between pICM and mTE.

Next, we carried out global gene expression analysis for the above 4 types of cells using a microarray method. On the basis of the obtained expression data, we performed hierarchical clustering using GeneSpring GX7.3 software and constructed a dendrogram for all samples (Fig. 1D). As expected, the expression patterns for the 4 types of samples clustered in a site-specific manner; in particular, pICM exhibited a gene expression pattern that differed substantially from the other 3 groups. The expression patterns of ICM + pTE, Whole, and mTE clustered in a similar manner. The validity of the hierarchical clustering was also consistent with PCA (x-axis, PCA component 1: 48.67% variance; y-axis, PCA component 2: 21.85% variance; Fig. 1E). PCA revealed that each expression pattern was definitely divided into 4 categories according to cell origin. In the x-y plane, the global gene expression pattern of pICM was the most different from the other 3 groups. Among all the groups, pICM and mTE were mutually distant, whereas Whole was located at the middle of mTE and ICM + pTE. Thus, these results suggested that our
dataset for each embryonic cell sample allowed comparisons of site-specific transcriptional mechanisms associated with differentiation processes of diverse cell lineages.

Third, we performed ANOVA with the post-hoc test at a false discovery rate of 5% after GeneSpring normalization. The result showed that 3600 genes showed significantly differential expression between pICM and mTE (Fig. 1F). When we compared ICM + pTE with mTE, 1499 genes were significantly differentially expressed (Fig. 1F). In order to understand the biological roles of these differentially expressed genes, we used FatiGO at Babelomics (www.fatigo.org) for ontological comparison of the probe sets segregated using the post-hoc test (Supplemental Fig. 3 and 4) [23]. The ontological comparison provides controlled vocabularies for the description of the molecular function, cellular component, and biological process of analyzed genes and assists with our understanding of large-scale experiments such as microarrays as functional profiling tools. In the molecular function category, 3 terms were significantly different between pICM and mTE, but there was no significant difference between ICM + pTE and mTE (Supplemental Figs 3A and 4A). In the cellular component category, 39 terms were significantly different between ICM and mTE, and 16 terms were significantly different between ICM + pTE and mTE (Supplemental Figs 3B and 4B). Within the biological process category, 17 terms were significantly different between pICM and mTE, and 3 terms were significantly different between ICM + pTE and mTE (Supplemental Figs 3C and 4C). Furthermore, when we evaluated each term in ICM-upregulated genes individually, we came across a lot of terms related to cellular transcriptional activity, such as helicase activity (GO:0004386), ribosome (GO:0005840), ribonucleoprotein complex (GO:0030529), nucleus (GO:0005634), nucleolus (GO:0005730), nuclear part (GO:0044428), chromosome (GO:0005694), translation (GO:0006412), and RNA processing (GO:0006396) (Supplemental Fig. 3A, B, and C). In contrast, in ICM-downregulated genes and TE-upregulated genes, most terms were related to cytoskeletal dynamics, membrane function, and substance transport, e.g., actin binding (GO:0003779), integral to membrane (GO:0016021), intrinsic to membrane (GO:0031224), membrane part (GO:0044425), plasma membrane (GO:0005886), organelle membrane (GO:0031090), cytoskeleton (GO:0005856), actin
cytoskeleton (GO:0015629), actin filament (GO:0005884), endocytosis (GO:0006897), ion transport (GO:0006811), and actin cytoskeleton organization (GO:0030036) (Supplemental Fig. 3A, B, and C). These results appeared to reflect transcriptional activity that dictates whether bovine embryonic cells become the ICM or TE.

In silico prescreening for candidate genes that were site-specifically expressed in bovine blastocysts

In order to extract candidate genes that were predominantly expressed in the ICM from microarray results consisting of an enormous number of genes, we prepared 2 lists of genes that were differentially expressed in ICM + pTE and pICM as compared with mTE (Supplemental Fig. 1B). Of the transcripts that were significantly altered in ICM + pTE and pICM (Fig. 1F), 191 and 472 genes, respectively, were upregulated by at least 3-fold. Using these genes, we constructed a Venn diagram with the Gene Spring software, and 148 genes that showed upregulation in both ICM + pTE and pICM were considered to be predominantly expressed in the ICM relative to the TE (ICM-gene list, Supplemental Fig. 1B and Supplemental Table 2). On the other hand, to identify other candidate genes that were predominantly expressed in the TE, we also prepared 2 lists of genes that were significantly upregulated in mTE and ICM + pTE as compared with pICM by at least 3-fold (Supplemental Fig. 1C). Using these 799 and 295 genes, as well as the genes predominantly expressed in pICM, a Venn diagram was constructed (Supplemental Fig. 1C). The 292 genes that were upregulated in both gene lists were extracted as genes specifically expressed in the TE relative to pICM (TE-gene list, Supplemental Fig. 1C and Supplemental Table 3). We further investigated whether there were transcriptional differences between pTE and mTE in bovine blastocysts. When genes showing pTE-dominant expression were explored as described in Supplemental Fig. 1C, only 3 genes were upregulated in pTE by at least 3-fold: \( CCL26, ASGR1, \) and \( LOC100295019 \) (Supplemental Table 4). However, the biological functions of these genes in pre-implantation embryos are unknown. Thus, these results revealed that the positional effects in the mural and polar trophectoderm of bovine blastocysts were unaccompanied by transcriptional and functional differences.
Array data validation by quantitative real-time PCR

To validate the microarray data by quantitative RT-PCR using site-specific cell samples, i.e., pICM and mTE, we further selected transcripts related to transcription factors, growth factors, transporters, transcription activators, and oncogenes from 148 and 298 genes in the ICM- and TE-gene lists, respectively (Fig. 2). We consequently selected 15 transcripts as candidate genes expressed predominantly in ICMs as follows: NASP, TEX9, IFT122, ID1, ID2, IFITM1, IRF1, CIITA, FYN, HNF4A, KIAA0664, SMARCA2, SOX2, FGF2, and FGFR2. Furthermore, we analyzed expression levels of OCT4, NANO2, and FGF4 by quantitative RT-PCR. The expression levels of all transcripts analyzed by quantitative RT-PCR were significantly higher in pICM than in mTE, except for that of OCT4 (Fig. 2A). Meanwhile, 10 transcripts were also selected as candidate genes expressed predominantly in TEs as follows: ZFX, GATA2, ELF3, CTGF, SCUBE2, PDG, ATF3, DLX4, TFAP2A, and IFNT. In addition, CDX2 and GATA3 were chosen because these genes are well-known key factors for TE cell differentiation in mouse blastocysts (Fig. 2B) [6, 9, 24, 25]. As expected, most examined genes exhibited site-specific expression patterns, i.e., the genes that were selected by our microarray-based findings showed ICM- or TE-dominant expression patterns in a reciprocal manner from quantitative RT-PCR analysis (Fig. 2A and B). OCT4 mRNA did not exhibit significantly higher expression in pICM (Fig. 2A), which was consistent with previous studies using nonrodent blastocysts, such as those derived from humans, cattle, and pigs [25-27]. As expected, the expression patterns of most examined transcripts by quantitative RT-PCR were identical to the results of microarray analysis, indicating that our constructed gene lists, i.e., ICM- and TE-gene lists, dependably included transcripts showing site-specific expression in bovine blastocysts.

Whole-mount in situ hybridizations (WISHs) in bovine blastocysts

Furthermore, we carried out whole-mount in situ hybridizations (WISHs) for 18 and 12 genes that were examined by quantitative RT-PCR to directly address their asymmetric expression. The results of
WISH using antisense probes for the 30 genes are shown in Fig. 3 and were contrasted with signals using sense probes. Among the 18 genes, i.e., the 15 genes selected from the ICM-gene list and 3 additional genes (OCT4, NANOG, and FGF4), all genes except OCT4 showed clear-cut ICM-dominant expression (Fig. 3A). In particular, ID1, ID2, NASP, TEX9, FYN, IFITM1, IRF1, FGFR2, KIAA0664, SMARCA2, and SOX2 specifically localized to the ICM, and HNF4A was detected in the primitive endoderm of the blastocoelic surface of the ICM. OCT4 was distinctly observed in both the ICM and TE. Moreover, GATA2, CTGF, SCUBE2, ATF3, DLX4, TFAP2A, and IFNT, selected from the TE-gene list, and 2 additional genes (CDX2 and GATA3) exhibited more intense signals in the TE than the ICM (Fig. 3B). However, ZFX and PDGF were detected in not only the TE but also in the ICM. Finally, we have summarized these expression patterns of the genes examined in bovine blastocysts in combination with the available knowledge on mouse blastocysts (Tables 1 and 2). If the site-dominancy of the expression of each gene was confirmed in all experiments, namely, microarray analysis, quantitative RT-PCR, and WISH, we regarded the expression pattern as “predominant expression”. In general, WISH data did not provide quantitative evaluation, and therefore, we gave weight to the result of quantitative PCR experiment rather than that of WISH when we assessed the transcriptional dominancy in each gene.
DISCUSSION

Recent studies have provided evidence for noticeable differences in gene expression between mammals at the blastocyst stage [16, 25, 28, 29]. Nevertheless, due to the lack of reliable transcriptional information in ICM and TE lineages in cattle, there has been no choice but to translate knowledge on mouse blastocysts into mechanisms underlining the first cell segregation in bovine blastocysts. Here, we conducted informative studies on bovine embryos to gain a deeper understanding of the developmental processes that establish the segregation of the ICM and the TE in cattle. We performed transcriptomic analyses for site-specific cell samples isolated from bovine blastocysts using a micromanipulator, namely, Whole, ICM + pTE, mTE, and pICM, and thereby found 17 and 12 genes showing ICM- or TE-dominant expression by quantitative RT-PCR using ICM and TE cells isolated from bovine blastocysts. A subset of these genes showed site-dominant expression patterns in either the ICM or TE of bovine blastocysts. Including these genes, we further carried out WISH of 30 transcripts to assess the spatial expression dynamics in blastocysts. Our overall conclusion was that both the ICM and TE in bovine blastocysts exhibited gene expression patterns in a manner distinctly different from those in mouse blastocysts at the molecular level.

As shown in Table 1, we identified 14 genes that were predominantly expressed in the ICM of bovine blastocysts. Moreover, OCT4, NANOG, and FGF4 were also analyzed in the same way, although they were not detected by microarray analysis. Our results showed that OCT4 was not restricted to the ICM and was detected in whole blastocysts, consistent with previous studies [16, 24, 25, 27, 30-32]. Additionally, NANOG and FGF4 also showed more predominant expression in the ICM than that in the TE as determined by quantitative RT-PCR and WISH analyses, similar to observations by other groups [18, 24, 33]. The ICM-dominant expression of HNF4A, SMARCA2, and SOX2 was consistent with the expression of these genes in mouse blastocysts [3, 34, 35], suggesting that the functions of these transcripts in bovine blastocysts were the same as those in mouse blastocysts. Hepatocyte nuclear factor 4 alpha (Hnf4A) is known to be downstream of Gata6 in a transcriptional cascade that regulates differentiation of the primitive
endoderm [36]. As expected, we also found that bovine HNF4A was expressed in the primitive endoderm of the ICM (Fig. 3A). Furthermore, the expression levels of both SMARCA2 and SOX2 were greater in the ICM (Fig. 2A and 3A). The ICM-dominant expression of SMARCA2 and SOX2 in bovine blastocysts might be necessary for maintaining the undifferentiated state of the ICM because these proteins control gene expression during mouse pre-implantation development as chromatin modifiers and transcription factors [35, 37]. In addition, Sox2 acts to preserve developmental potential and is detectable as one of the earliest ICM-specific marker genes in the mouse ICM [38].

Besides these obvious and predictable genes, we found 8 novel ICM-dominant genes in the bovine blastocyst: ID1, ID2, NASP, TEX9, FYN, IFITM1, IRF1, and KIAA0664 (Table 1) [39]. Id family proteins involve many developmental processes, such as murine gastrulation and neurogenesis, in which Id2 is one of the earliest TE-specific marker genes, contrary to Sox2, which exhibits ICM-specific transcription [38]. Interestingly, however, our results revealed the ICM-dominant transcription of both ID1 and ID2 in bovine blastocysts (Fig. 2A and 3A), demonstrating the differences in lineage-restricted transcription factors between mice and cattle. On the other hand, the expression patterns of FGF4 and FGFR2 in bovine blastocysts are similar to those observed in the mouse blastocyst (Table 1) [40-42]. In mice, Fgf4 is restricted to the ICM at the blastocyst stage, which drives trophoblast proliferation and maintains the ICM through the Fgf-receptor response mechanisms [43, 44]. Among the Fgf receptors, Fgfr2 in particular is expressed in both the ICM and TE [41, 45] and is thought to be the main TS-cell-specific receptor [46, 47]. In a rodent placentation model, the interaction between Fgf4 and Fgfr2 in pTE surrounding the ICM was directly essential for subsequent placentation; in parallel, mTE that is nonadjacent to the ICM cannot receive Fgf4 signaling and skips the placentation process [45]. However, this placentation model is not applicable to most mammals, including cattle, because the pTE in nonrodents and higher primates promptly disappears after blastocyst hatching [27]. Nevertheless, we detected predominant FGF4 expression in the ICM and FGFR2 expression occurring predominantly in the ICM. One possibility was that another gene expressed predominantly in the ICM, FGF2, may be associated with this disappearance of pTE in the
bovine conceptus. In humans, FGF2 stimulates cell proliferation and induces migration and apoptosis of tumor cells [48, 49]. Though the direct molecular wiring of FGF2 signaling in cattle has yet to be identified, we detected the predominant expression of FGF2 in the bovine ICM, despite that the transcript levels of FGF2 are extremely low or undetectable in mouse blastocysts [50].

CDX2, GATA3, and IFNT also showed more predominant expression in the TE than that in the ICM as measured by quantitative RT-PCR and WISH analyses, as expected [25, 51, 52]. Moreover, we found 5 novel TE-specific genes in bovine blastocysts: ELF3, CTGF, ATF3, DLX4, and TFAP2A (Table 2).

Dlx4 is a homeobox gene that encodes a family of transcription factors with important roles in patterning and differentiation during embryogenesis [53]. Dlx4 is expressed in the murine and human placentas and in trophoblast cell lines, but little is known about the biological function(s) of this protein during mammalian placentation [54]. The other 4 transcription factors, namely, ELF3, ATF3, DLX4, and TFAP2A, were associated with transforming growth factor (TGF) β signaling activation, regulating a vast array of developmental processes that include maintenance of stem cell pluripotency, cell fate decisions during development, and homeostasis of adult tissue [55-57]. Although the significance of TGFβ signaling in bovine embryos is still unknown, these transcription factors may be involved in differentiation and maintenance of bovine trophoblasts. Interestingly, TGFβ induces connective tissue growth factor (CTGF), which is also a known target of the YAP-TEAD4 signaling pathway that distinguishes TE from ICM in the mouse blastocyst [58-60]. We found that CTGF expression was predominantly observed in the TE of bovine blastocysts, which was consistent with our results demonstrating that ELF3, ATF3, DLX4, and TFAP2A were overexpressed in the TE.

The mammalian blastocyst provides an excellent model system for exploring the relationship between cellular morphogenesis and cell fate. Here, we reported 12 novel site-dominant genes expressed in bovine blastocysts. Our findings further provided a definitive demonstration that ICM- and TE-dominant transcripts accompanied the establishment of cell lineage specification at the blastocyst stage in cattle, a process that was separate from that observed in mice.
FIGURE LEGENDS

Figure 1. Transcriptomic analysis using site-specific cell samples derived from bovine blastocysts. (a) 4 types of cell samples were mechanically and chemically isolated from bovine expanded blastocysts: whole blastocysts (Whole), inner cell mass and polar trophectoderm (ICM + pTE), mural trophectoderm (mTE), and pure inner cell mass (pICM). (b) Immunofluorescence localization of CDX2 protein (green) was determined at the expanded blastocyst stage (middle row). Nuclei were stained with Hoechst33342 (blue, upper row). The inner cells, i.e., the inner cell mass, were identified as CDX2-negative cells, with only nuclei exhibiting Hoechst staining (arrowheads). Note that Whole and ICM + pTE showed CDX2 localization in the nuclei of the TE, and all nuclei of mTE were CDX2-positive. In contrast, all cells in pICM were CDX2 negative. Scale bar: 100 μm. (c) Proportion of cells with or without CDX2 protein expression per each collected cell sample; Whole, ICM + pTE, mTE, and pICM. (d) Principal component analysis of gene expression in all the samples subjected to hierarchical clustering analysis. (e) Hierarchical clustering of 4 types of cell samples, i.e., Whole, ICM + pTE, mTE, and pICM. Colors correspond to the relative RNA abundance of more than 23,000 transcripts. (f) One-way ANOVA post-hoc testing in the 4 types of cell samples derived from bovine blastocysts. Each box shows the number of genes that are statistically different (blue) or similar (orange) in a group-to-group comparison.

Figure 2. Quantitative RT-PCR analysis using the site-specific cell samples derived from bovine blastocysts. Examined transcripts were selected from microarray expression profiles. Twenty-six transcripts were selected for quantitative RT-PCR (n = 3) from the in silico gene expression analyses shown in Supplemental Table 1 (ICM-gene list) and Supplemental Table 2 (TE-gene list). Additionally, 4 transcription factors (OCT4, NANOG, CDX2, and GATA3) and 1 growth factor (FGF4) were analyzed. (a) Fifteen transcripts extracted from the ICM-gene list, namely, NASP, TEX9, IFT122, ID1, ID2, IFITM1, IRF1, CIITA, FYN, HNF4A, KIAA0664, SMARCA2, SOX2, FGF2, and FGFR2, in addition to OCT4,
NANOG, and FGF4, that were key genes for the appropriate differentiation into the blastocyst stage in mice.

(b) Ten transcripts were extracted from the TE-gene list, namely, ZFX, GATA2, ELF3, CTGF, SCUBE2, PDGF, ATF3, DLX4, TFAP2A, and IFNT. In addition, CDX2 and GATA3 were also chosen because these transcripts were essential for appropriate TE differentiation in the mouse blastocyst. Values are represented as means ± s.e.m. (indicated by error bars). The asterisk indicates $P < 0.05$.

**Figure 3. Spatiotemporal changes of our screened transcripts in bovine blastocysts.** In situ hybridization analysis was carried out using either DIG-labeled antisense or sense riboprobes that were hybridized to whole-mount bovine blastocysts. All the examined genes corresponded to those analyzed by quantitative RT-PCR, as shown in Fig. 2, i.e., (A) 18 transcripts: OCT4, NANOG, FGF4, NASP, TEX9, IFT122, ID1, ID2, IFITM1, IRF1, CIITA, FYN, HNF4A, KIAA0664, SMARCA2, SOX2, FGF2, and FGFR2 and (B) 12 transcripts: CDX2, GATA3, ZFX, GATA2, ELF3, CTGF, SCUBE2, PDGF, ATF3, DLX4, TFAP2A, and IFNT. Note that 17 transcripts in (A) (excluding OCT4) showed obviously higher signals in the ICM than in the TE. Similarly, the signals of 9 transcripts shown in (B) (excluding ZFX, SCUBE2, and PDGF) were evident in the TE of bovine blastocysts. Scale bar: 50 µm.

**Supplemental Figure 1. Bovine blastocysts analyzed in this study and Venn diagrams for exploring site-specific gene expression.** (A) Bovine expanded blastocysts used in this study, which were prepared by in vitro oocyte maturation, fertilization, and subsequent in vitro embryo culture. (B) Venn diagram constructed by 191 and 472 genes that were upregulated in ICM + pTE and pICM, respectively, by at least 3-fold compared to mTE. (C) Venn diagram constructed by 799 and 295 genes that were upregulated in mTE and ICM + pTE, respectively, by at least 3-fold.
Supplemental Figure 2. Assessment of $H2AFZ$ expression level among the site-specific cell samples.

To evaluate the standardization of mRNA samples for quantitative real-time PCR, we compared amount of $H2AFZ$ transcript in all the site-specific cell samples.

Supplemental Figure 3. Estimation of the biological roles of the differentially expressed genes between pICM and mTE. The top gene ontology (GO) categories were defined on the basis of the probe sets differentially expressed between pICM and mTE. Differences were considered significant at a level of $P < 0.05$. A total of 3, 39, and 17 terms were significantly different in (A) molecular function, (B) cellular component, and (C) biological process categories, respectively.

Supplemental Figure 4. Estimation of the biological roles of the differentially expressed genes between ICM + pTE and mTE. As described in Supplemental Figure 2, we examined genes that were differentially expressed between ICM + pTE and mTE. Differences were considered significant at a level of $P < 0.05$. A total of 16 and 3 terms were significantly different in (A) cellular component and (B) biological process categories, respectively. In the molecular function category, there was no significant difference between ICM + pTE and mTE.
REFERENCES


<table>
<thead>
<tr>
<th>gene</th>
<th>Mouse mRNA</th>
<th>Protein</th>
<th>Cattle mRNA</th>
<th>Protein</th>
<th>Microarray</th>
<th>qRT-PCR</th>
<th>WISH</th>
<th>mRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCT4</td>
<td>ICM(^1)</td>
<td>ICM(^2)</td>
<td>ICM and TE(^3)</td>
<td>ICM and TE(^3)</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>non-biased</td>
</tr>
<tr>
<td>NANOG</td>
<td>ICM(^4)</td>
<td>ICM(^4)</td>
<td>ICM(^5)</td>
<td>ICM(^6)</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>predominant in ICM</td>
</tr>
<tr>
<td>FGF4</td>
<td>−</td>
<td>ICM(^7)</td>
<td>ICM(^8)</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>±</td>
<td>predominant in ICM</td>
</tr>
<tr>
<td>ID1</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>predominant in ICM with marginal expression in TE</td>
</tr>
<tr>
<td>ID2</td>
<td>TE(^9)</td>
<td>TE(^9)</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>±</td>
<td>±</td>
<td>predominant in ICM with marginal expression in TE</td>
</tr>
<tr>
<td>NASP</td>
<td>−</td>
<td>ICM and TE(^10)</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>predominant in ICM with marginal expression in TE</td>
</tr>
<tr>
<td>TEX9</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>predominant in ICM</td>
</tr>
<tr>
<td>IFT122</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>increased in ICM</td>
</tr>
<tr>
<td>FYN</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>predominant in ICM</td>
</tr>
<tr>
<td>HNF4A</td>
<td>ICM (PE)(^11)</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>predominant in ICM with marginal expression in TE</td>
</tr>
<tr>
<td>IFITM1</td>
<td>N.D.(^1)</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>predominant in ICM</td>
</tr>
<tr>
<td>IRF1</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>predominant in ICM</td>
</tr>
<tr>
<td>CIITA</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>±</td>
<td>−</td>
<td>increased in ICM</td>
</tr>
<tr>
<td>FGF2</td>
<td>N.D.(^12)</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>predominant in ICM with marginal expression in TE</td>
</tr>
<tr>
<td>FGFR2</td>
<td>TE(^13)</td>
<td>TE(^13) or ICM(^14)</td>
<td>TE and PE(^15)</td>
<td>−</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>increased in ICM with significantly lower level of expression in TE</td>
</tr>
<tr>
<td>KIAA0664</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>predominant in ICM with marginal expression in TE</td>
</tr>
<tr>
<td>SMARCA2</td>
<td>−</td>
<td>ICM(^16)</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>±</td>
<td>±</td>
<td>predominant in ICM</td>
</tr>
<tr>
<td>SOX2</td>
<td>ICM(^17)</td>
<td>ICM(^17)</td>
<td>ICM(^18)</td>
<td>ICM(^18)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>predominant in ICM</td>
</tr>
</tbody>
</table>

ICM: inner cell mass; TE: trophectoderm; PE: primitive endoderm; N.D.: not determined; −: unknown

When all evaluations for each gene were “+”, we considered ICM-dominant expression pattern.

References:
<table>
<thead>
<tr>
<th>gene</th>
<th>Mouse mRNA</th>
<th>Protein</th>
<th>Cattle mRNA</th>
<th>Protein</th>
<th>Microarray</th>
<th>qRT-PCR</th>
<th>WISH</th>
<th>mRNA</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDX2</td>
<td>TE&lt;sup&gt;1&lt;/sup&gt;</td>
<td>TE&lt;sup&gt;2&lt;/sup&gt;</td>
<td>TE&lt;sup&gt;3&lt;/sup&gt;</td>
<td>TE&lt;sup&gt;4&lt;/sup&gt;</td>
<td>*</td>
<td>±</td>
<td>+</td>
<td>predominant in TE with marginal expression in ICM</td>
<td></td>
</tr>
<tr>
<td>GATA3</td>
<td>TE&lt;sup&gt;5&lt;/sup&gt;</td>
<td>TE&lt;sup&gt;6&lt;/sup&gt;</td>
<td>TE&lt;sup&gt;7&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>predominant in TE</td>
</tr>
<tr>
<td>ZFX</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+</td>
<td>±</td>
<td>—</td>
<td>increased in TE</td>
<td></td>
</tr>
<tr>
<td>GATA2</td>
<td>TE&lt;sup&gt;8&lt;/sup&gt;</td>
<td>—</td>
<td>TE&lt;sup&gt;7&lt;/sup&gt;</td>
<td>—</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>predominant in TE</td>
<td></td>
</tr>
<tr>
<td>ELF3</td>
<td>ICM and TE&lt;sup&gt;9&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>predominant in TE</td>
<td></td>
</tr>
<tr>
<td>CTGF</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+</td>
<td>±</td>
<td>±</td>
<td>predominant in TE with marginal expression in ICM</td>
<td></td>
</tr>
<tr>
<td>SCUBE2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>predominant in TE</td>
<td></td>
</tr>
<tr>
<td>PDGF</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+</td>
<td>±</td>
<td>—</td>
<td>increased in TE</td>
<td></td>
</tr>
<tr>
<td>ATF3</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>predominant in TE with marginal expression in ICM</td>
<td></td>
</tr>
<tr>
<td>DLX4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>predominant in TE</td>
<td></td>
</tr>
<tr>
<td>TFAP2A</td>
<td>TE&lt;sup&gt;10&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>+</td>
<td>±</td>
<td>+</td>
<td>predominant in TE with marginal expression in ICM</td>
<td></td>
</tr>
<tr>
<td>IFNγ</td>
<td>—</td>
<td>—</td>
<td>TE&lt;sup&gt;11&lt;/sup&gt;</td>
<td>—</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>predominant in TE</td>
<td></td>
</tr>
</tbody>
</table>

ICM: inner cell mass; TE: trophectoderm; PE: primitive endoderm; N.D.: not determined; —: unknown. When all evaluations for each gene were “+”, we considered ICM-dominant expression pattern. *: The probe for this gene was not present in microarray chip used in this study.

References:
Figure 1

Part A: Schematic diagram showing the dissection and treatment process for samples.

Part B: Representative images of nuclei and CDX2 expression in different samples.

Part C: Bar chart showing the proportion of cells with or without CDX2 protein expression across different samples.

Part D: Heatmap depicting the relative expression level of different samples.

Part E: Scatter plot illustrating the PCA components with data distribution.

Part F: Table summarizing the expression values for different samples.

Figure 1
Figure 2
Figure 3