Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing

Fernando H. Biase, Xiaoyi Cao, and Sheng Zhong

Department of Bioengineering, University of California San Diego, La Jolla, California 92093, USA

It remains an open question when and how the first cell fate decision is made in mammals. Using deep single-cell RNA-seq of matched sister blastomeres, we report highly reproducible inter-blastomere differences among 10 2-cell and five 4-cell mouse embryos. Inter-blastomere gene expression differences dominated between-embryo differences and noise, and were sufficient to cluster sister blastomeres into distinct groups. Dozens of protein-coding genes exhibited reproducible bimodal expression in sister blastomeres, which cannot be explained by random fluctuations. The protein expression of one gene out of four of these bimodal genes tested, *Gadd45a*, exhibited clear inter-blastomeric contrasts. We traced some of the bimodal mRNA expressions to embryonic genome activation, and others to blastomere-specific RNA depletion. Inter-blastomere differences created coexpression gene networks that were much stronger and larger than those that can possibly be created by random noise. The highly correlated gene pairs at the 4-cell stage overlapped with those showing the same directions of differential expression between inner cell mass (ICM) and trophectoderm (TE). These data substantiate the hypothesis of inter-blastomere differences in 2- and 4-cell mouse embryos, and associate these differences with ICM/TE differences.

[Supplemental material is available for this article.]

An important question in cell biology is how cells break the symmetry during mitotic divisions. During mammalian preimplantation embryonic development, the embryo has to decide how to set apart the first two cell populations (Johnson and Ziomek 1981; Fleming and Johnson 1988). It remains an open question when and how the first cell fate decision is made. Cell fate-associated inter-blastomere differences in transcript and protein concentrations were reported as early as the 8–16 cell stage (Avilion et al. 2003; Plusa et al. 2005a; Zhang et al. 2006; Yagi et al. 2007; Jedrusik et al. 2008; Nishioka et al. 2008; Guo et al. 2010). However, it is not clear whether these are the earliest differences. Even the basic physical principle behind the earliest cell fate decision is debatable. Two hypotheses have been proposed (Zernicka-Goetz et al. 2009).

The "equivalence hypothesis" suggests that the individual blastomeres in 2-cell and 4-cell embryos are homogeneous. Any quantitative differences between the blastomeres are either irrelevant to cell fate (Alarcon and Marikawa 2005; Motosugi et al. 2005; Kurotaki et al. 2007) or produced as random noise. This hypothesis is theoretically compatible with the lineage specification at the blastocyst stage, under the assumption that the gene regulatory networks are sufficient to converge the initial random concentrations of the gene products into a finite number of equilibrium concentrations (Raj and van Oudenaarden 2008; Huang 2010; Johnston and Desplan 2014). This theory implies that noise plays a central role in cell fate decision. The opposing "asymmetric hypothesis" suggests that the individual blastomeres in early embryos (2-cell or 4-cell stage embryos) are not equivalent. There are reproducible quantitative differences of molecules, modifications, external clues, or orientations between the blastomeres (inter-blastomere differences), which may influence the blastomere toward the specification of either cell lineage (Plusa et al. 2005b; Torres-Padilla et al. 2007; Bischoff et al.

2008; Plachta et al. 2011). Here, we report reproducible inter-blastomere differences within each 2-cell or 4-cell embryo from single-cell RNA sequencing (RNA-seq) data.

Results

Generation and visualization of single-cell RNA-seq

We generated 1857 million SMART-seq reads (Ramskold et al. 2012) from 49 single cells composed of nine 1-cell (zygote), 10 mid-stage 2-cell, and five 4-cell embryos (Fig. 1A). The generation of data from every sister blastomere turned out indispensable for testing the hypotheses above. We deeply sequenced each library to produce on average 37.9 million 100-base long reads (Fig. 1B), which covered nearly the entire length of the transcripts (Supplemental Fig. S1). After removing clonal reads, we detected transcripts of 9432 protein-coding genes and observed a larger separation of the cells between embryonic stages than within any stage (Fig. 1C). We then generated RNA-seq data from 4 inner cell mass (ICM) and three trophectoderm (TE) samples (Fig. 1B). Based on the UCSC Genome Browser (Kent et al. 2002), we built a single-cell transcriptome browser to visualize and compare these data: http://singlecell.cepbrowser.org (Supplemental Fig. S2).

Dissecting within-embryo and between-embryo variation

We asked whether there are gene products that can reproducibly mark inter-blastomere differences within every embryo. We decomposed the cell-to-cell variation (sum of squares, SS) of transcript abundance of each gene into between-embryo variation (SSbe) and within-embryo variation (SSwe; SS = SSbe + SSwe) (Fig. 2A), and ordered genes by the proportion of SSwe (SSwe/SS)
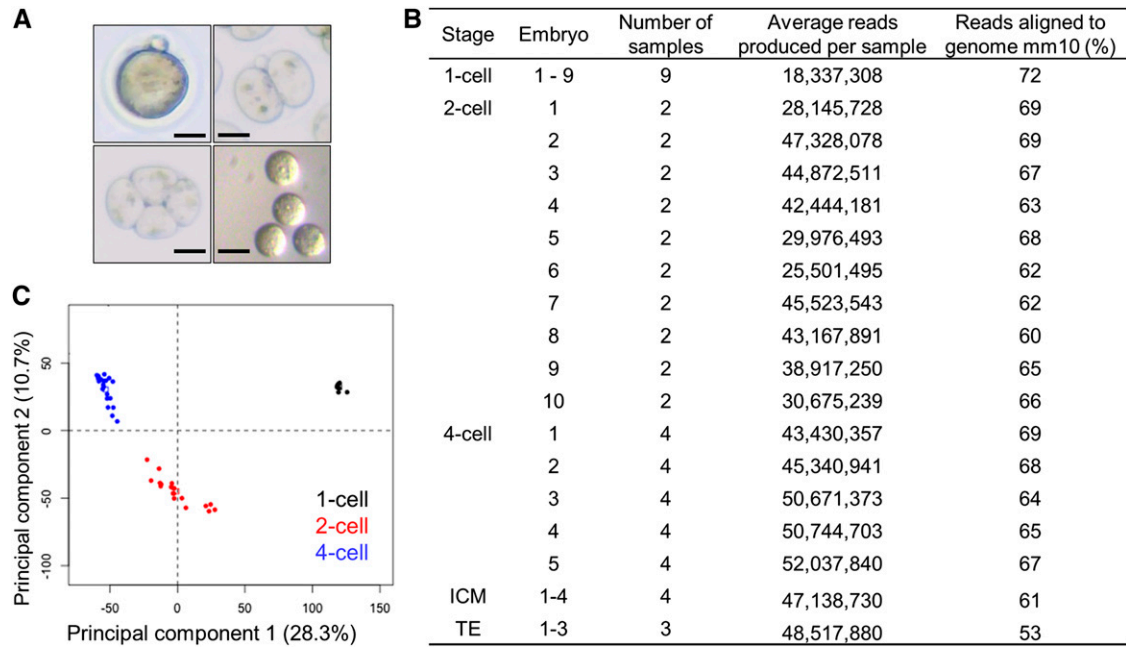
Figure 1. Single-cell SMART-seq data. (*A*) Representative images of 1-cell, 2-cell, and 4-cell embryos; and isolated blastomeres of a 4-cell embryo. (*B*) Summary of sequenced and mapped reads. Each sample in 1-cell, 2-cell, and 4-cell stages was a single cell. Each ICM and TE sample was prepared from a separate embryo. (*C*) Principal components of the 49 single cells. Scale bar, 50 μm.

(Fig. 2B). A nontrivial number of 3467 and 7008 genes exhibited larger SSwe than SSbe at the 2-cell and the 4-cell stages, respectively. At the 2-cell stage, the top ranked genes (largest SSwe/SS) reproducibly presented a large FPKM (>1000, upper quartile normalized) in one blastomere but near zero FPKM in the other blastomere in eight or nine out of the 10 embryos (Fig. 2B). At the 4-cell stage, 590 genes exhibited ninefold or greater SSwe than SSbe (FDR ≤ 0.08) (Fig. 2C). These genes showed three to six orders of magnitude of expression differences between the blastomeres of the same embryo, but exhibited relatively stable expression levels across the embryos, opening the possibility of observing reproducible inter-blastomere differences.

Inter-blastomere differences dominate embryo differences and noise

How does the genome-wide signal strength of inter-blastomere differences compare to that of between-embryo differences and to the noise in gene expression? We did a series of clustering analyses of the blastomeres, starting from using all detected genes (coefficient of variation [CV] ≥ 0) and gradually restricting to using the genes with large overall cell-to-cell variation (CV > 0.25, 0.35, ..., 0.85). Because CV represented the total variation rather than the within-embryo variation, this criterion would not favor any specific clustering outcome. However, the higher the CV filter, the more blastomeres of the same embryo were separated into different clusters (Supplemental Fig. S3). At CV > 0.65, the 288 genes left clustered the 2-cell stage blastomeres into two distinct groups, composed of nine (blue, Fig. 2D) and 11 cells (red), respectively. The nine cells of the blue cluster came from nine distinct embryos, (*P*-value = 6 × 10^{-5}, hypergeometric test). Similarly, at CV > 0.70, every 4-cell embryo contributed blastomeres to the two furthest clusters (Fig. 2E). This unsupervised analysis suggests that the signal of within-embryo difference is strong as compared to between-embryo differences

and random noise. Essentially most genes or gene sets with large total variation can distinguish sister blastomeres.

Bimodality of gene expression in sister blastomeres

It remains unclear whether any mammalian developmental processes or stem cells have bimodal gene expression (Shalek et al. 2013). A total of 138 and 205 genes showed bimodal expression patterns at the 2-cell and the 4-cell stages, respectively (FDR < 0.05, Dip Test of Unimodality [Hartigan and Hartigan 1985]). We then asked whether any genes were specifically bimodal among the sister blastomeres of the same embryo. After removing between-embryo variation, we found 12 and 13 genes with within-embryo bimodality at the 2-cell and the 4-cell stages, respectively (FDR < 0.05, Dip Test) (Fig. 3A,B; Supplemental Tables S1, S2). The majority of these genes had near zero FPKM in some blastomeres but $10^3 - 10^6$ FPKM (upper quartile normalized) in the sister blastomeres in nearly every embryo. The reproducible bimodal gene expression between the sister blastomeres is incompatible with the equivalence hypothesis.

We investigated whether the bimodality of one gene was predictive of another. At the 2-cell stage, the bimodal mRNAs of *Ap2s1* and *Rsph3a* showed opposite directions in their inter-blastomere differences in nine out of 10 embryos (FDR < 6.7 × 10^{-4}) (Fig. 3C). At the 4-cell stage, two pairs of bimodal genes, *Camk1d*-*Acox3* and *Gpr124*-*Csf2rb2*, exhibited strong correlations (FDR < 0.01) (Fig. 3D). The correlation and anticorrelation of bimodal gene expression are unexpected from random fluctuations of gene expression.

We further interrogated whether the bimodal distribution of transcript abundance is reflected by protein abundance. We analyzed the protein expression of bimodal genes *Ap2s1* and *Zfp688* at the 2- and 4-cell stages, as well as *Acox3* and *Gadd45a* at the 4-cell stage by immunocytochemistry. The immunofluorescence signals
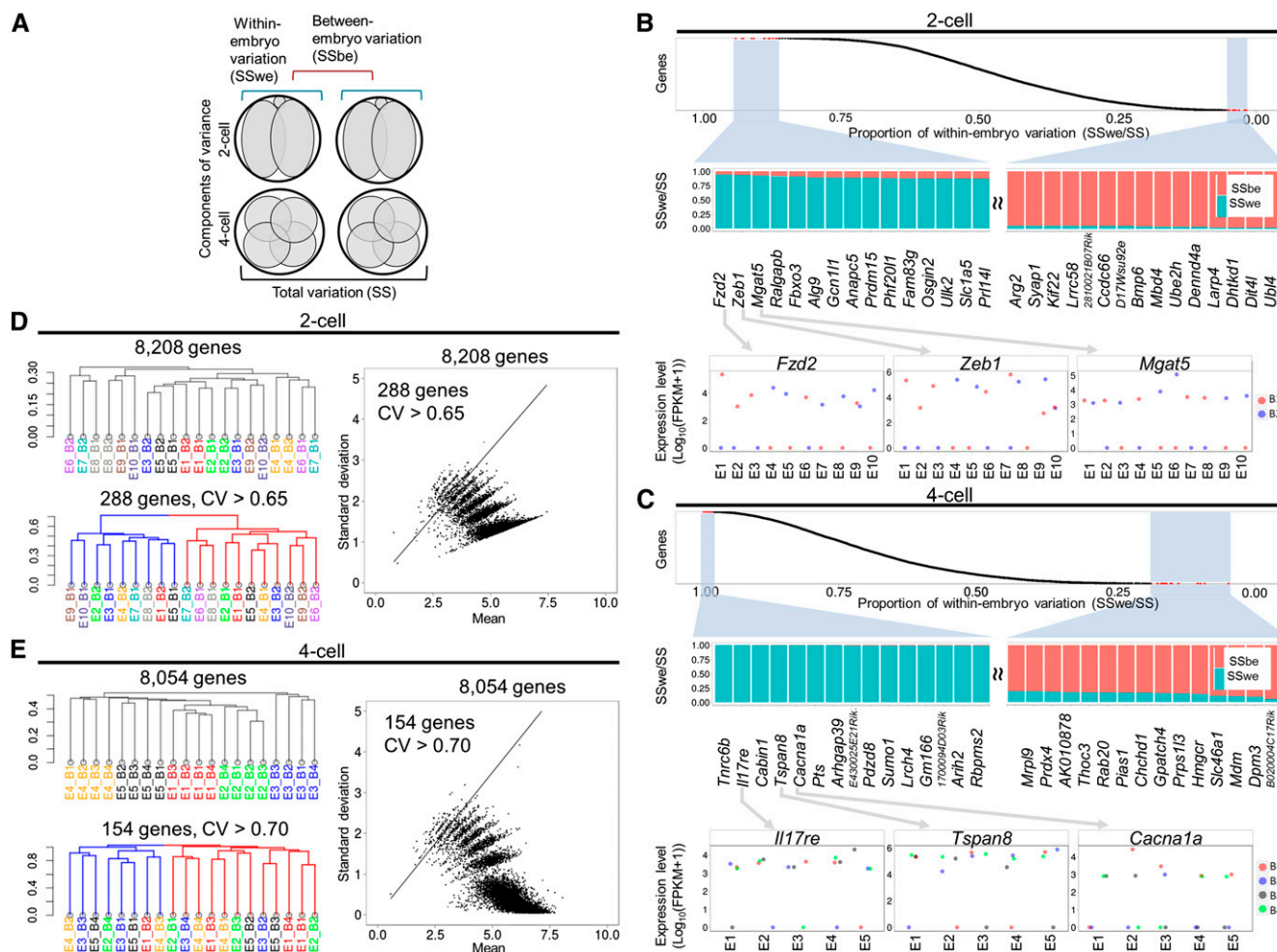
Figure 2. Variance decomposition and unsupervised clustering of single cells. (*A*) The total variation (SS) can be decomposed into the sum of within-embryo variation (SSwe) and between-embryo variation (SSbe). (*B*,*C*) Genes (*y*-axis) are rank-ordered from the largest to the smallest by SSwe/SS (*x*-axis). Expanded views of the genes with the largest and the smallest SSwe/SS are shown *beneath*, followed by the expression levels of selected genes. (E1–E10) Embryos 1 to 10. (B1, B2) Blastomeres 1 and 2. (FPKM) Upper quartile normalized FPKM. (*D*,*E*, *left*) Clustering of the blastomeres using all mRNAs (*upper* dendrogram) and high CV genes (*lower* dendrogram). (Ei-Bj) The jth blastomere of embryo i. (*Right*) A scatter plot of standard deviation (*y*-axis) vs. mean (*x*-axis) of all genes. The CV filter selected for the genes *above* the line.

of alpha tubulin (control), AP2S1, ZFP688, and ACOX3 were comparable between the blastomeres (see Data Access). However, GADD45A exhibited visual and statistical differences in fluorescence intensities between the blastomeres in the majority (five to eight, depending on the method of assessment) of the eight 4-cell embryos scanned (Fig. 4; Supplemental Fig. S4; *P*-value ≤ 0.1 in Supplemental Table S3; see also Data Access). Moreover, each pair of blastomeres originating from the same mother cell, as identified by the spindle apparatus, exhibited similar immunofluorescence signals. In six cases, differences in protein expression were observed between the two pairs of blastomeres defined by the first cleavage. Thus, the distribution of GADD45A proteins resembled the bimodal distribution of *Gadd45a* mRNAs.

Causes of bimodal gene expression

At least three plausible causes could produce transcriptome asymmetry in 2-cell embryos, namely, embryonic genome activation (EGA), polarized cell division, and RNA degradation. We compared the SMART-seq data of the bimodal genes across five mature oocytes

(RNA-seq data from Ramsköld et al. 2012; Xue et al. 2013), zygotes, and the 2-cell embryos. The mRNAs of *Fzd2* and *Zfp688* were not detected in oocytes and most of the zygotes, but were abundant (FPKM ≥ 1000, upper quartile normalized) in at least one blastomere in every 2-cell embryo (Fig. 5A; Supplemental Fig. S5). Thus, the *Fzd2* and *Zfp688* mRNAs were EGA transcripts (*P*-value = 3.5 × 10^{-4}, hypergeometric test). On the other hand, the mRNAs of *Ap2s1* and *Rsph3a* (Fig. 5B; Supplemental Fig. S5) were abundant in every zygote, whereas they had zero or near zero FPKM in precisely one blastomere in eight out of the 10 2-cell embryos, suggesting these mRNAs were differentially depleted between the sister blastomeres (maximum *P*-value = 3.5 × 10^{-4}, hypergeometric test).

Coexpression networks as further evidence for molecular asymmetry

We checked whether within-embryo differences could produce coexpression networks. After removing between-embryo variation, we clustered all the detected genes. Large coexpression modules emerged from both the 2-cell and the 4-cell stages (red blocks,
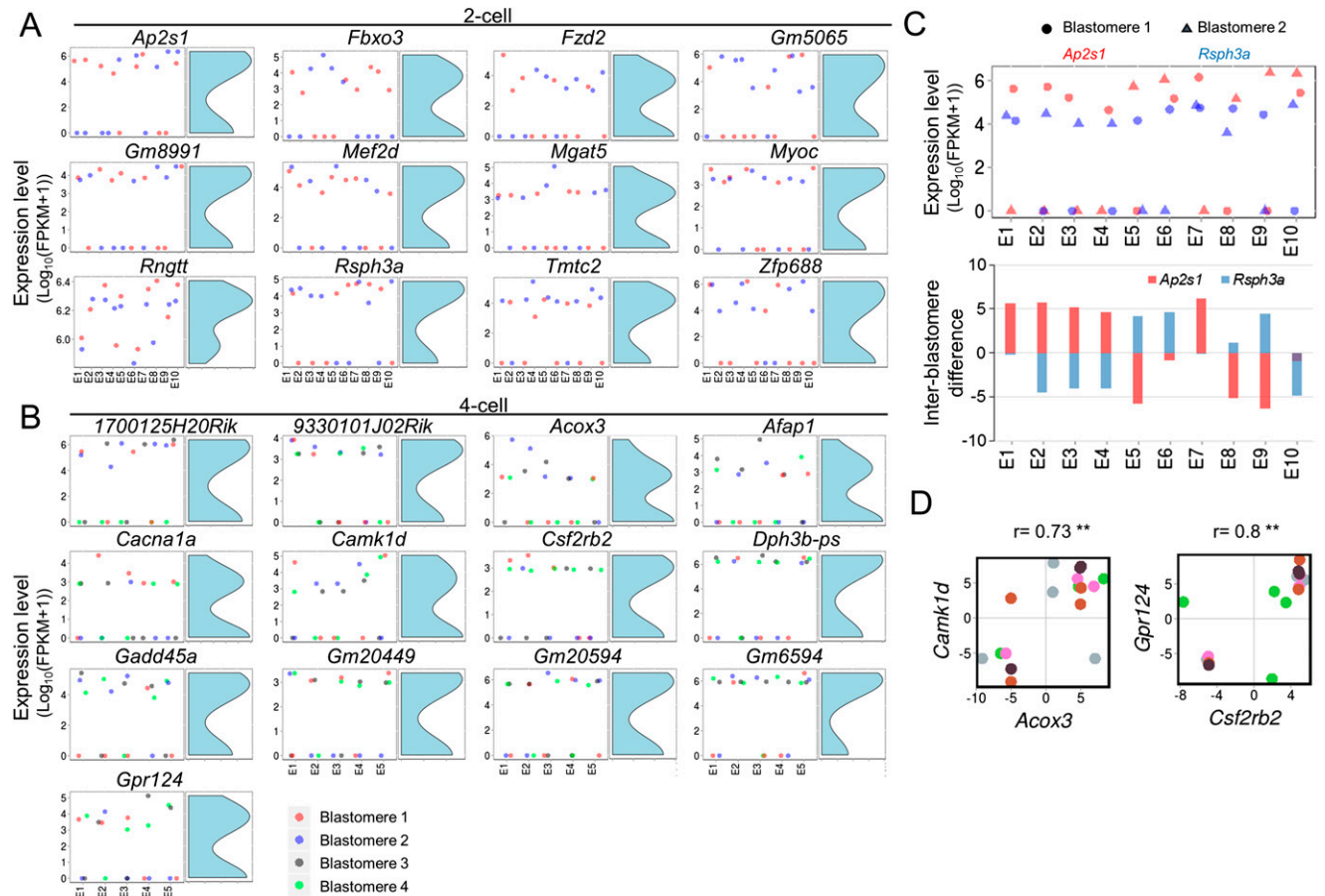
Figure 3. Bimodality of gene expression in sister blastomeres. (*A,B*) Expression levels and their histograms. (E1–E10) Embryos 1 to 10. The two (four) blastomeres of each embryo are randomly assigned to two (four) colors. (FPKM) Upper quartile normalized FPKM. (*C*) Anticorrelation of 2-cell bimodality genes *Ap2s1* and *Rsph3a*. Juxtaposition (*upper*) and inter-blastomere difference (*lower*) of expression levels in the same embryos. (*D*) Correlated genes at the 4-cell stage. Cells from the same embryo have the same color. (r) Pearson correlation. (**) FDR < 0.01.

Fig. 6A,B, left panels). To assess the statistical significance of these coexpression modules, we compared the distribution of pairwise correlations of genes with permutation-derived distributions. All the large coexpression modules disappeared from the permutation data (Fig. 6A,B, middle panels), and the numbers of strong

Figure 4. GADD45A protein expression levels in two 4-cell mouse embryos. For each embryo (*upper* or *lower* panels), the confocal images were captured on the same z stack from an immunocytochemistry assay of GADD45A (Alexa-548, red) and alpha tubulin (Alexa-488, green). Arrows point to cells with greater expression of GADD45A. (DNA) DAPI fluorescence. (DIC) Differential interference contrast image. Scale bar, 50 µm.

(anti)correlations were greatly reduced (P-value $< 2 \times 10^{-22}$, Kolmogorov test) (Fig. 6A,B, right panels). The coexpression modules were enriched with gene ontology (GO) terms including "cell fate commitment" (Fig. 6C), and corresponded to known protein–protein or regulatory interactions (P-value $\leq 10^{-10}$, test by resampling) (Supplemental Fig. S6). The formation of coexpression modules provides further evidence favoring the asymmetric hypothesis.

Association of 4-cell stage asymmetry and ICM/TE specification

We tested whether the inter-blastomere differences at the 4-cell stage were related to the earliest cell fate decision, namely, the specification of ICM and TE. None of the 4-cell stage blastomeres could be assumed to be prone to a lineage, ruling out single gene analyses. Using the idea of second-order correlation (Dettling et al. 2005), we tested whether the gene pairs with consistent directions of inter-blastomere differences at the 4-cell stage would exhibit consistent directions of differences between ICM and TE (Fig. 6D). Because we required greater 4-cell stage positive correlations (4-cell r), more gene pairs exhibited consistent directions of expression differences between ICM and TE (P-value $< 10^{-3}$, χ^2 test). The 128 pairs of most correlated genes in 4-cell stage (r > 0.95) exhibited the same direction of differential expression
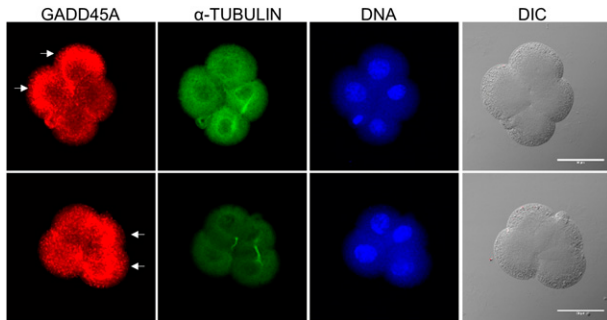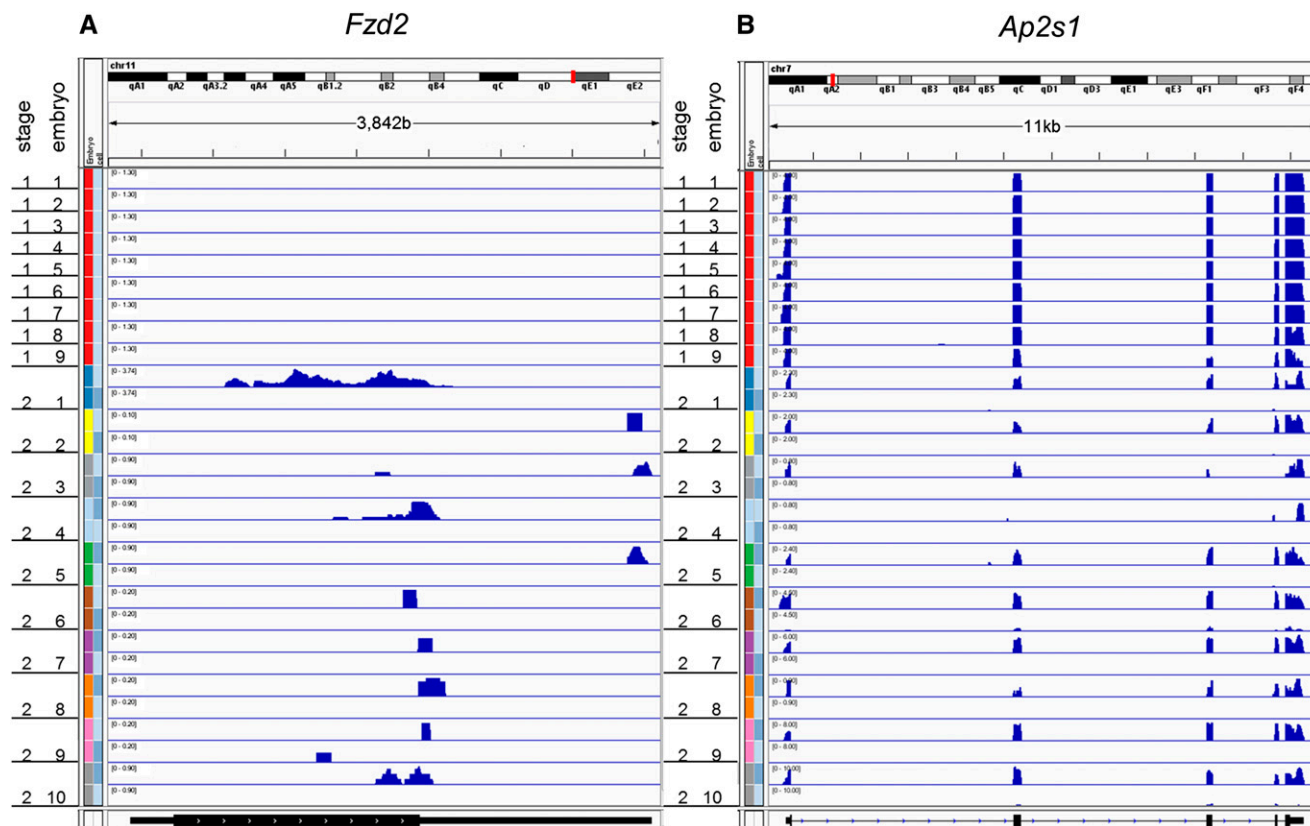
Figure 5. Coverage plots of single-cell RNA-seq reads. Each row represents a cell. The gene models are at the *bottom* of the plots. (*A*) *Fzd2*, a candidate bimodality gene due to EGA. (*B*) *Ap2s1*, a candidate bimodality gene due to blastomere-specific RNA depletion.

between ICM and TE (P-value $< 10^{-10}$, Fisher's exact test). Forty-eight of these genes were annotated to biological functions potentially relevant to cell fate decisions (Supplemental Table S4). For example, the genes *Arnt*, *Cyfip1*, *Dazap1*, *Fancc*, *Napa*, *Ptcd2*, *Smad2*, and *Tacc1* were associated with the GO term "cell differentiation." In light of all the data above, we suggest that the first cell fate decision in mammals begins earlier than the 8–16 cell stage.

Discussion

Two technical considerations were important for our analysis. First, every blastomere of every embryo, especially the 4-cell embryos, has to be preserved in the analysis. The carefully matched sister blastomeres were indispensable to the observation of reproducible patterns, and provided a sufficient sample size for statistical assessments. Second, the genome-wide RNA measurement technology has to be accurate enough. Since its inception (Tang et al. 2009), single-cell RNA-seq has quickly evolved into a method with constrained technical noise, suitable for analyzing cell-to-cell variation (Ramskold et al. 2012; Brennecke et al. 2013; Shalek et al. 2013; Xue et al. 2013; Yan et al. 2013; Deng et al. 2014; Wu et al. 2014). The SMART-seq technology (Ramskold et al. 2012) and sufficient sequencing depths (Brennecke et al. 2013; Wu et al. 2014) were important for this evolution. Our single-cell real-time PCR (qPCR) experiment quantifying the expression of 96 genes in 88 blastomeres (Supplemental Fig. S8A) suggested certain degrees of differences between SMART-seq and single-cell qPCR, in quantifying

the relative expression levels across genes (Supplemental Fig. S8B) and the cell-to-cell variation of each gene (Supplemental Fig. S8C). It requires carefully designed future experiments to clarify the expected degree of reproducibility between SMART-seq and qPCR in quantifying the expression difference of a given gene among single cells.

The inter-blastomere differences of mRNA abundance within a 2-cell or a 4-cell embryo were primarily attributed to random noise (Zernicka-Goetz et al. 2009). This attribution was based on the small and non-reproducible between-blastomere fluctuations of candidate genes before the 8–16 cell stage (Dietrich and Hiiragi 2007; Jedrusik et al. 2008; Zernicka-Goetz et al. 2009; Guo et al. 2010; Morris et al. 2013). Our genome-wide analyses revealed nontrivial and reproducible inter-blastomere differences in 2-cell and 4-cell embryos. These inter-blastomere differences were often larger than between-embryo differences (Figs. 2B,C, 3A,B). Moreover, the genes with the largest cell-to-cell variation were enriched with those exhibiting bimodal expression between sister blastomeres (Fig. 3A,B; Supplemental Fig. S7). In nearly every embryo, these bimodal genes consistently expressed at a high level in some blastomere(s) and at a minimal level in the other blastomere(s). Although protein synthesis is not necessarily contemporary with transcript synthesis (Vigneault et al. 2009), we found further evidence at the protein expression level for one (*Gadd45a*) out of four bimodal transcripts tested. This reproducible pattern cannot be explained by random noise, and it is reminiscent of noninvasive lineage tracing results (Plusa et al. 2005b; Bischoff et al. 2008; Tabansky et al. 2013).
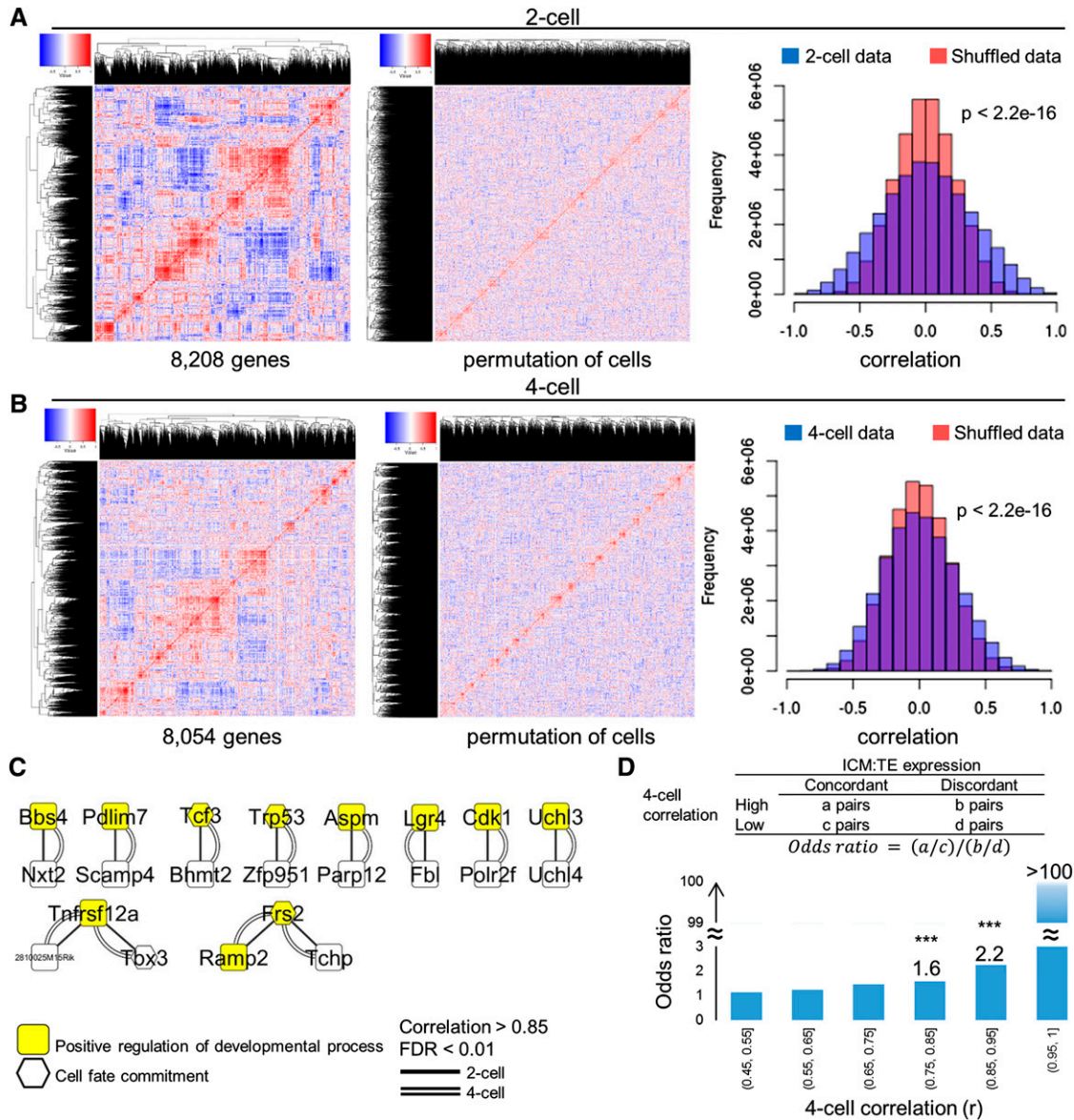
Figure 6. Coexpression networks resulted from inter-blastomere differences. (*A,B*) Correlation heatmaps of all genes after removing between-embryo variation (*left*) indicates the existence of coexpression networks (red blocks). The networks disappeared after permuting the cell labels (*middle*), which essentially produces data from the equivalence hypothesis. There were considerably more correlated and anticorrelated gene pairs in the real data than in the permuted data (*right*). (*C*) A subset of gene pairs with strong correlations at both 2-cell and 4-cell stages. This subset contained the genes with GO annotations "positive regulation of developmental processes" or "cell fate commitment." (*D*) Consistency of gene expression differences at 4-cell and blastocyst stages. Gene pairs are categorized by their correlation (r) at the 4-cell stage (*x*-axis). A large r suggests that a gene pair shares this same direction of inter-blastomere differences. Odds ratio (OR) was used to represent the chances of these gene pairs exhibiting the same direction of expression differences between ICM and TE. OR had a monotone increase with respect to 4-cell r. (***) *P*-value $< 10^{-3}$. (0.45, 0.55] is equivalent to $0.45 < r \leq 0.55$.

Probable mechanisms

Several probable mechanisms are compatible with the asymmetric hypothesis, including polarized cell division and differential gene regulation by signaling (Morris et al. 2012), TFs, epigenetics, post-transcriptional processing and transport, and degradation (for reviews, see Zernicka-Goetz 2004; Vogel 2005; Zernicka-Goetz et al. 2009; Takaoka and Hamada 2012; Saiz and Plusa 2013). Supportive data for these mechanisms have become evident for the 8–16 cell stage and onward, but not as clear for the earlier stages (Torres-Padilla et al. 2007; Plachta et al. 2011). The majority of genes or gene pairs implicated to be associated with 2-cell or 4-cell transcriptome asymmetry in this analysis are signaling molecules, TFs, epigenetic modifiers, and RNA processors (Supplemental Tables S5–S7).

WNT signaling

WNT signaling is overwhelmingly represented by 2-cell bi-modality genes (Supplemental Table S4), starting with its ligand biogenesis enzyme *Mgat5* (Matthews et al. 2009), receptor *frizzled homolog 2* (*Drosophila*) (*Fzd2*), frizzled-binding modulator *myocilin* (*Myoc*) (Kwon et al. 2009), to its ligand endocytosis

factor *Ap2s1* (Blitzer and Nusse 2006), TF-binding *Rsph3a* (Grimsby et al. 2004), and to direct transcriptional targets *Mgat5* (Sengupta et al. 2010) (potentially forming a feed forward loop) and TF gene *Zeb1* (Sanchez-Tillo et al. 2011). Besides, both the *Wnt* inducing TF gene *Tbx3* (also known as *Brachury*) and the canonical TF gene of the WNT pathway *Tcf3* showed up in correlated gene pairs (Fig. 6C; Supplemental Table S6). WNT antagonist *Apc* strongly correlated with DNA methyltransferase *Dnmt1* in both 2-cell and 4-cell stages (Supplemental Fig. S5). More signaling pathways were implicated in the 4-cell bimodality genes and within-embryo correlated gene pairs (Fig. 6C; Supplemental Tables S6, S7), including FGF receptor substrate *Frs2*, G protein-coupled receptor *Gpr124*, and WNT agonist receptor *Lgr4*.

EGA is implicated to the creation of asymmetry

Fifteen percent of the mRNAs with large SSwe (SSwe > 4 × SSbe) were not detectable in mature oocytes (Ramskold et al. 2012; Xue et al. 2013), and thus belonged to EGA transcripts. A subset of 2-cell bimodal genes was transcribed by the embryo genome in one and only one blastomere in 2-cell embryos (Fig. 5A; Supplemental Fig. S5A). This reproducible pattern cannot be explained by transcription noise alone. It is tempting to hypothesize that WNT-related TFs and epigenetic modifiers (Supplemental Tables S4–S7) regulate blastomere-specific EGA.

Blastomere-specific RNA depletion could account for another subset of bimodal mRNAs, especially those abundant in zygotes (Fig. 5B; Supplemental Fig. S5B). This depletion pattern could be explained by polarized cell division (Johnson and Ziomek 1981; Fleming and Johnson 1988) or cell-specific RNA degradation (Giraldez et al. 2006; Lykke-Andersen et al. 2008), even though no evidence of either mechanism has been found prior to the 8-cell stage. Surprisingly, at least four out of the 12 correlated gene pairs (Supplemental Table S7) are involved in regulation of cell polarity (Supplemental Table S8). These data suggest that, as early as the 2-4 cell stages, the blastomeres could have prepared differently for future asymmetric divisions. Consistent with this idea, bimodal *Afap1* encodes a protein that interacts with atypical protein kinase C (aPKC) (Qian et al. 2002), which adopt polarized localization from the 8-cell stage onward and regulate cell fate decisions (Plusa et al. 2005a). In summary, these single-cell data from matched sister blastomeres argue against the equivalence hypothesis, and lead to several mechanistic insights into the proposed early asymmetry.

Capacity versus preference

Every 2-cell or 4-cell stage blastomere is capable of contributing progenies to ICM or TE (Tarkowski et al. 2001; Katayama et al. 2010). This totipotent capacity does not guarantee an identical tendency of lineage contribution among the blastomeres. This study adds molecular data to the notion of nonidentical lineage preferences of 2-cell and 4-cell blastomeres in the endogenous developmental process. The dynamic nature of gene transcription during the first two cleavages (Wang et al. 2004; Zernicka-Goetz et al. 2009; Xie et al. 2010) leaves little room for the bimodal gene expression to be consecutive at both 2-cell and 4-cell stages. Nonetheless, the overlap of strongly correlated gene pairs between 2-cell and 4-cell stages may indicate a continuity of the inter-blastomeric difference. The data cannot be explained by equivalent cells transiently visiting distinct states due to the shared directions of within-embryo differential expression between the 4-cell and the blastocyst stages (Fig. 6D).

Methods

Embryo collection

Adult female mice (C57BL/6) were super-ovulated by intraperitoneal hormone injections with a 48-h interval between injections. The hormones administered were either PMSG followed by hCG (5 IU/ 0.1 mL, Sigma-Aldrich), or 0.1 mL of PG600 (Intervet) administered twice, equivalent to 2.5 IU of hCG and 5I U of PMSG. Following the second hormonal treatment, the females were mated with males (1:1, C57BL/6). Zygotes, 2-cell, 4-cell, and blastocysts were collected 26, 48, 52, and 94 h post-hCG or the second injection of PG600. These procedures were approved by the Institutional Animal Care and Use Committees of the University of Illinois at Urbana Champaign and the University of California San Diego.

Prior to blastomere collection, the zona pelucida was removed by Tyrode's acid solution. The 2- and 4-cell embryos were then immersed in Trypsin solution (Invitrogen) with RNase inhibitor (1 IU/μL, Clontech) and BSA (50 μg/μL) for the separation of the blastomeres. The separated blastomeres were snap frozen and maintained at −80°C until processed.

ICM was isolated by immunosurgery, using anti-mouse IgG and complement (Sigma-Aldrich), following previously described procedures (Handyside and Barton 1977; Harlow and Quinn 1979), and snap frozen in TRIzol reagent (Invitrogen). TE was isolated by exposure of the blastocyst to concanavalin A conjugated to magnetic beads, followed by trypsin treatment for cell dismemberment. The cells were immersed in PBS (RNase inhibitor 1 IU/μL, BSA 50 μg/ μL). Following magnetic separation of the trophoblast cells, they were snap frozen in TRIzol reagent. The total RNA was extracted from ICM or TE cells with TRIzol (Life Technologies) and the supplement of 1 μg of Glycoblue (Invitrogen).

RNA library preparation and sequencing

The zygotes, blastomeres, and RNA extracted from ICM and TE were used for cDNA amplification using the SMARTer Universal Low Input RNA kit (Clontech). All amplifications were carried out in parallel with positive and no-template controls for quality assurance. For all samples, the total amplified cDNA was used for library preparation using Nextera XT DNA Sample Prep Kit (Illumina). In order to avoid batch effects, all the libraries for cells of the same embryo were prepared at the same time. The measurements of cDNA amplification library quality control were taken with Qubit (Invitrogen) and Bioanalyzer (Agilent). The libraries of cells composing one embryo were subsequently sequenced in the same lane of a flow cell to 100-base read length on HiSeq 2000 or HiSeq 2500 sequencers (Illumina).

Alignment of sequencing reads to the mouse genome

The 100-base long single-end reads were aligned to the mouse genome (mm10, Ensembl; Flicek et al. 2013), using the STAR aligner (Dobin et al. 2013), allowing for up to four mismatches. Only the reads aligned uniquely to one genomic location, and with alignment quality above 10 were retained for subsequent analyses. The duplicate reads were eliminated with Picard tools (http://picard. sourceforge.net).

Assessment of transcript coverage

The relative coverage of the reads along the entire length of the transcripts was used to assess the quality of single-cell RNA-seq. The RefSeq mRNA sequences, corresponding to mm10, were downloaded from the UCSC Genome Browser (Kuhn et al. 2012),

indexed with Bowtie 2 v2.1.0 (Langmead and Salzberg 2012), and then aligned with the RNA-seq reads. The reads that aligned once to a sequence were retained for further filtering of duplicates with Picard tools. The coordinate of a read mapping (transcript and nucleotide) was used to calculate the position of the read relative to the total length of the transcript (0%–100%). The relative positions of all the reads were summarized into a cumulated density map.

Estimation of gene expression levels in single cells

The non-duplicated reads uniquely aligned to the mouse genome were used to estimate the gene expression levels (FPKM) by Cufflinks v2.0.2 (Trapnell et al. 2012) with the option of upper quartile normalization, which multiplied a scaling factor by the unnormalized FPKM (Dillies et al. 2013). Gene models were obtained from the Ensembl annotation (GRCm38.70). Requiring FPKM ≥ 1 in at least one blastomere of every 2-cell embryo, or at least two blastomeres of every 4-cell embryo, we obtained 8208 and 8054 mRNAs from the 2-cell and the 4-cell stages, respectively. Only the detected mRNAs were used in this analysis.

Building the web browser

The single-cell genome browser was built based on the portable codes of UCSC Genome Browser (Kent et al. 2002). We removed duplicated reads and converted the remaining aligned reads of each blastomere into a bigWig track. We normalized the read intensity in every track by the number of non-duplicated uniquely aligned reads of that library. We grouped the tracks by each embryo and ordered them by developmental stages. Genome annotation and sample information were provided alongside data tracks. This browser currently hosts 49 single cell tracks, four ICM tracks, three TE tracks, as well as two previously published sperm (Gan et al. 2013) and five previously published mature oocyte data sets (Ramskold et al. 2012; Xue et al. 2013).

Removing between-embryo variation

Between-embryo variation was removed in a subset of the analyses (Figs. 3, 5). Let y_{ij} be the $\log_{10}(\text{FPKM} + 1)$ of the gene in embryo i blastomere j. The following transformation is applied to every gene: $z_{ij} = y_{ij} - \overline{y_i}$, where $\overline{y_i}$ is the mean of y_{ij} in embryo i, and z_{ij} is the relative expression level after removing between-embryo variation.

Testing the deviations from unimodal distributions

The Dip Test (Hartigan and Hartigan 1985) was used to assess the deviation of data from unimodal distributions. The null hypothesis was the expression levels followed a unimodal distribution and the alternative was a multimodal distribution. FDR was calculated subsequently (Pollard et al. 2005).

Permutation test of correlation

One hundred permutations were performed after removing the between-embryo variation. The cell labels of each gene were independently swapped in each permutation. The distribution of correlations was derived from sampling 336,815 and 324,133 gene pairs in each permutation from the 2-cell and the 4-cell stages, respectively. These sampled gene pairs added up to the same number of gene pairs in real data from the 100 permutations.

Association of coexpression in 4-cell embryos and in ICM/TE

Pairwise correlation was calculated for 4-cell data, after removing between-embryo variation. An association test was devised to as-

sess whether the highly correlated gene pairs at the 4-cell stage tend to share the same direction of differential expression between ICM and TE. Gene pairs were binned by their 4-cell correlation, and then the numbers of gene pairs in each bin that shared the same direction of differential expression between ICM and TE were counted. These counts were used to compute odds ratio (OR), for which significance was assessed by either χ^2 or Fisher's exact test.

Immunocytochemistry assay (IC)

Four genes were selected for further validation of RNA-seq results. Antibodies for the proteins are as follows: AP2S1 (Abcam, ab128950), ZFP688 (Abcam, ab72605), ACOX3 (Sigma, HPA035840), and GADD45A (Sigma, G3548). Embryos were collected at 2- or 4-cell stages and the zona pelucida was removed as described above. Embryos were fixed in methanol at $-20°C$ overnight, washed in PBS (supplemented with 0.1% BSA), and incubated for 5 min in permeabilization solution (PBS, 0.1% BSA, 0.1% Triton X-100). Embryos were then incubated for 1 h in blocking solution (PBS, 0.1% BSA, 10% bovine fetal serum), following incubation with primary antibody (1:100 dilution in blocking solution) for 2 h at room temperature. After PBS (0.1% BSA) washes, indirect detection was possible by immersing the embryos in blocking solution with anti-rabbit antibody conjugated with Alexa Fluor-546 Goat Anti-Rabbit IgG (1:500 dilution, A11010, Invitrogen). Following washes with PBS (0.1% BSA), embryos were immersed in blocking solution containing Mouse anti-Tubulin-Alexa-488 (1:25 dilution, 32-2588, Invitrogen) for 1 h. Embryos were washed in PBS (0.1% BSA) supplemented with NucBlue Fixed Cell Stain ReadyProbes (Invitrogen), and preserved with Prolong Gold (Invitrogen). Embryos were scanned with a Fluoview FV1000 confocal microscope (Olympus), with the aid of a 40× oil objective, excitation wavelength 405/473/559, and emission wavelength 461/520/572 (DAPI, Alexa-488 and Alexa-545, respectively). Scanning was performed at consecutive z-planes separated by 1 μm.

Single-cell quantitative PCR

Independent of the SMART-seq analyzed embryos, another 12 zygotes, 20 2-cell, and nine 4-cell embryos were collected for single-cell qPCR. The single cells were separated and frozen as described above. In addition, we prepared a pool of 10 zygotes and a pool of 50 zygotes as positive controls. The cells and pools were subjected to reverse transcription with SuperScriptVILO cDNA Synthesis kit (Life Technologies) as recommended by the manufacturer. DELTAgene Assays were custom designed for 94 genes associated with embryonic development or regulation of gene expression, as well as two housekeeping genes (Supplemental Table S9). The cDNA from cells and pools were subjected to target specific amplification for 20 cycles. Pre-amplified template was diluted and subjected to primer specific amplification with SsoFast EvaGreen Supermix (Bio-Rad) on a 96.96 array chip and BiomarkHD System (Fluidigm). Relative quantification of gene expression ($\log_2$ space) was obtained by subtracting the Ct values from the baseline value of 22.

Data access

The sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE57249. Processed single cell data can be visualized at http://singlecell.cepbrowser.org. Single-cell qPCR data have been submitted to GEO under accession number GSE59892. Additional files associated with this study are available in the Supplemental Material and at http://systemsbio.ucsd.edu/singlecellped.

Acknowledgments

References

Alarcon VB, Marikawa Y. 2005. Unbiased contribution of the first two blastomeres to mouse blastocyst development. *Mol Reprod Dev* **72:** 354–361.

Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17:** 126–140.

Bischoff M, Parfitt DE, Zernicka-Goetz M. 2008. Formation of the embryonic-abembryonic axis of the mouse blastocyst: relationships between orientation of early cleavage divisions and pattern of symmetric/asymmetric divisions. *Development* **135:** 953–962.

Blitzer JT, Nusse R. 2006. A critical role for endocytosis in Wnt signaling. *BMC Cell Biol* **7:** 28.

Brennecke P, Anders S, Kim JK, Kolodziejczyk AA, Zhang X, Proserpio V, Baying B, Benes V, Teichmann SA, Marioni JC, et al. 2013. Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods* **10:** 1093–1095.

Deng Q, Ramskold D, Reinius B, Sandberg R. 2014. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343:** 193–196.

Dettling M, Gabrielson E, Giovanni P. 2005. Searching for differentially expressed gene combinations. *Genome Biol* **6:** R88.

Dietrich JE, Hiiragi T. 2007. Stochastic patterning in the mouse pre-implantation embryo. *Development* **134:** 4219–4231.

Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, et al. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform* **14:** 671–683.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21.

Fleming TP, Johnson MH. 1988. From egg to epithelium. *Annu Rev Cell Biol* **4:** 459–485.

Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. 2013. Ensembl 2014. *Nucleic Acids Res* **42:** D749–D755.

Gan H, Wen L, Liao S, Lin X, Ma T, Liu J, Song CX, Wang M, He C, Han C, et al. 2013. Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis. *Nat Commun* **4:** 1995.

Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312:** 75–79.

Grimsby S, Jaensson H, Dubrovska A, Lomnytska M, Hellman U, Souchelnytskyi S. 2004. Proteomics-based identification of proteins interacting with Smad3: SREBP-2 forms a complex with Smad3 and inhibits its transcriptional activity. *FEBS Lett* **577:** 93–100.

Guo G, Huss M, Tong GQ, Wang C, Li Sun L, Clarke ND, Robson P. 2010. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell* **18:** 675–685.

Handyside AH, Barton SC. 1977. Evaluation of the technique of immunosurgery for the isolation of inner cell masses from mouse blastocysts. *J Embryol Exp Morphol* **37:** 217–226.

Harlow GM, Quinn P. 1979. Isolation of inner cell masses from mouse blastocysts by immunosurgery or exposure to the calcium ionophore A23187. *Aust J Biol Sci* **32:** 483–491.

Hartigan JA, Hartigan PM. 1985. The dip test of unimodality. *Ann Stat* **13:** 70–84.

Huang S. 2010. Cell lineage determination in state space: a systems view brings flexibility to dogmatic canonical rules. *PLoS Biol* **8:** e1000380.

Jedrusik A, Parfitt DE, Guo G, Skamagki M, Grabarek JB, Johnson MH, Robson P, Zernicka-Goetz M. 2008. Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo. *Genes Dev* **22:** 2692–2706.

Johnson MH, Ziomek CA. 1981. The foundation of two distinct cell lineages within the mouse morula. *Cell* **24:** 71–80.

Johnston RJ Jr, Desplan C. 2014. Interchromosomal communication coordinates intrinsically stochastic expression between alleles. *Science* **343:** 661–665.

Katayama M, Ellersieck MR, Roberts RM. 2010. Development of monozygotic twin mouse embryos from the time of blastomere separation at the two-cell stage to blastocyst. *Biol Reprod* **82:** 1237–1247.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12:** 996–1006.

Kuhn RM, Haussler D, Kent WJ. 2013. The UCSC genome browser and associated tools. *Brief Bioinform* **14:** 144–161.

Kurotaki Y, Hatta K, Nakao K, Nabeshima Y, Fujimori T. 2007. Blastocyst axis is specified independently of early cell lineage but aligns with the ZP shape. *Science* **316:** 719–723.

Kwon HS, Lee HS, Ji Y, Rubin JS, Tomarev SI. 2009. Myocilin is a modulator of Wnt signaling. *Mol Cell Biol* **29:** 2139–2154.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359.

Lykke-Andersen K, Gilchrist MJ, Grabarek JB, Das P, Miska E, Zernicka-Goetz M. 2008. Maternal Argonaute 2 is essential for early mouse development at the maternal-zygotic transition. *Mol Biol Cell* **19:** 4383–4392.

Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, et al. 2009. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res* **37:** D619–D622.

Morris SA, Guo Y, Zernicka-Goetz M. 2012. Developmental plasticity is bound by pluripotency and the Fgf and Wnt signaling pathways. *Cell Rep* **2:** 756–765.

Morris SA, Graham SJ, Jedrusik A, Zernicka-Goetz M. 2013. The differential response to Fgf signalling in cells internalized at different times influences lineage segregation in preimplantation mouse embryos. *Open Biol* **3:** 130104.

Motosugi N, Bauer T, Polanski Z, Solter D, Hiiragi T. 2005. Polarity of the mouse embryo is established at blastocyst and is not prepatterned. *Genes Dev* **19:** 1081–1092.

Nishioka N, Yamamoto S, Kiyonari H, Sato H, Sawada A, Ota M, Nakao K, Sasaki H. 2008. Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech Dev* **125:** 270–283.

Plachta N, Bollenbach T, Pease S, Fraser SE, Pantazis P. 2011. Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. *Nat Cell Biol* **13:** 117–123.

Plusa B, Frankenberg S, Chalmers A, Hadjantonakis AK, Moore CA, Papalopulu N, Papaioannou VE, Glover DM, Zernicka-Goetz M. 2005a. Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. *J Cell Sci* **118:** 505–515.

Plusa B, Hadjantonakis AK, Gray D, Piotrowska-Nitsche K, Jedrusik A, Papaioannou VE, Glover DM, Zernicka-Goetz M. 2005b. The first cleavage of the mouse zygote predicts the blastocyst axis. *Nature* **434:** 391–395.

Pollard K, Dudoit S, van der Laan M. 2005. Multiple testing procedures: the multtest package and applications to genomics. In *Bioinformatics and computational biology solutions using R and bioconductor* (ed. Gentleman R et al.), pp. 249–271. Springer, New York.

Qian Y, Baisden JM, Cherezova L, Summy JM, Guappone-Koay A, Shi X, Mast T, Pustula J, Zot HG, Mazloum N, et al. 2002. PC phosphorylation increases the ability of AFAP-110 to cross-link actin filaments. *Mol Biol Cell* **13:** 2311–2322.

Raj A, van Oudenaarden A. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135:** 216–226.

Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC, et al. 2012. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* **30:** 777–782.

Saiz N, Plusa B. 2013. Early cell fate decisions in the mouse embryo. *Reproduction* **145:** R65–R80.

Sanchez-Tillo E, de Barrios O, Siles L, Cuatrecasas M, Castells A, Postigo A. 2011. β-catenin/TCF4 complex induces the epithelial-to-mesenchymal transition (EMT)-activator ZEB1 to regulate tumor invasiveness. *Proc Natl Acad Sci* **108:** 19204–19209.

Sengupta PK, Bouchie MP, Kukuruzinska MA. 2010. *N*-glycosylation gene *DPAGT1* is a target of the Wnt/β-catenin signaling pathway. *J Biol Chem* **285:** 31164–31173.

Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D, et al. 2013. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498:** 236–240.

Tabansky I, Lenarcic A, Draft RW, Loulier K, Keskin DB, Rosains J, Rivera-Feliciano J, Lichtman JW, Livet J, Stern JN, et al. 2013. Developmental bias in cleavage-stage mouse blastomeres. *Curr Biol* **23:** 21–31.

Takaoka K, Hamada H. 2012. Cell fate decisions and axis determination in the early mouse embryo. *Development* **139:** 3–14.

Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, et al. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6:** 377–382.

Tarkowski AK, Ozdzenski W, Czolowska R. 2001. How many blastomeres of the 4-cell embryo contribute cells to the mouse body? *Int J Dev Biol* **45:** 811–816.

Torres-Padilla ME, Parfitt DE, Kouzarides T, Zernicka-Goetz M. 2007. Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* **445:** 214–218.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L, et al. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7:** 562–578.

Vigneault C, McGraw S, Sirard MA. 2009. Spatiotemporal expression of transcriptional regulators in concert with the maternal-to-embryonic transition during bovine in vitro embryogenesis. *Reproduction* **137:** 13–21.

Vogel G. 2005. Embryology. Embryologists polarized over early cell fate determination. *Science* **308:** 782–783.

Wang QT, Piotrowska K, Ciemerych MA, Milenkovic L, Scott MP, Davis RW, Zernicka-Goetz M. 2004. A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. *Dev Cell* **6:** 133–144.

Wu AR, Neff NF, Kalisky T, Dalerba P, Treutlein B, Rothenberg ME, Mburu FM, Mantalas GL, Sim S, Clarke MF, et al. 2014. Quantitative assessment of single-cell RNA-sequencing methods. *Nat Methods* **11:** 41–46.

Xie D, Chen CC, Ptaszek LM, Xiao S, Cao X, Fang F, Ng HH, Lewin HA, Cowan C, Zhong S. 2010. Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res* **20:** 804–815.

Xue Z, Huang K, Cai C, Cai L, Jiang CY, Feng Y, Liu Z, Zeng Q, Cheng L, Sun YE, et al. 2013. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500:** 593–597.

Yagi R, Kohn MJ, Karavanova I, Kaneko KJ, Vullhorst D, DePamphilis ML, Buonanno A. 2007. Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development* **134:** 3827–3836.

Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J, et al. 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* **20:** 1131–1139.

Zernicka-Goetz M. 2004. First cell fate decisions and spatial patterning in the early mouse embryo. *Semin Cell Dev Biol* **15:** 563–572.

Zernicka-Goetz M, Morris SA, Bruce AW. 2009. Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat Rev Genet* **10:** 467–477.

Zhang J, Tam WL, Tong GQ, Wu Q, Chan HY, Soh BS, Lou Y, Yang J, Ma Y, Chai L, et al. 2006. Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of *Pou5f1*. *Nat Cell Biol* **8:** 1114–1123.