

Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst

Guoji Guo,^{1,2,4} Mikael Huss,^{3,4} Guo Qing Tong,^{2,5} Chaoyang Wang,² Li Li Sun,² Neil D. Clarke,³ and Paul Robson^{1,2,*}

¹Department of Biological Sciences, National University of Singapore, Singapore 117543

²Stem Cell & Developmental Biology

³Computational and Systems Biology Group

Genome Institute of Singapore, Singapore 138672

⁴These authors contributed equally to this work

⁵Present address: Department of Obstetrics and Gynecology, Nanjing Maternal and Child Health Care Hospital, Nanjing Medical University, Nanjing 210004, China

*Correspondence: robsonp@gis.a-star.edu.sg

DOI 10.1016/j.devcel.2010.02.012

SUMMARY

Three distinct cell types are present within the 64-cell stage mouse blastocyst. We have investigated cellular development up to this stage using single-cell expression analysis of more than 500 cells. The 48 genes analyzed were selected in part based on a whole-embryo analysis of more than 800 transcription factors. We show that in the morula, blastomeres co-express transcription factors specific to different lineages, but by the 64-cell stage three cell types can be clearly distinguished according to their quantitative expression profiles. We identify *Id2* and *Sox2* as the earliest markers of outer and inner cells, respectively. This is followed by an inverse correlation in expression for the receptor-ligand pair *Fgfr2/Fgf4* in the early inner cell mass. Position and signaling events appear to precede the maturation of the transcriptional program. These results illustrate the power of single-cell expression analysis to provide insight into developmental mechanisms. The technique should be widely applicable to other biological systems.

INTRODUCTION

Mouse preimplantation development provides an attractive model to study regulatory networks in the control of cell fate decisions. This developmental period begins at fertilization and proceeds from the 1-cell zygote to the blastocyst. Just prior to implantation, the embryo consists of three different cell types, the trophectoderm (TE), the primitive endoderm (PE), and the epiblast (EPI) (Rossant and Tam, 2009). The TE is a functional epithelium that is responsible for mediating attachment with and implantation into the uterine wall and subsequently contributing to the placenta, the PE is an extra-embryonic cell type whose descendants provide patterning cues and nutrient supplies to the developing embryo, and the pluripotent EPI gives rise to all cell types of the embryo proper and is the source of the

embryonic stem (ES) cell. Though much insight has been gained from classical developmental studies, there still remains only a rudimentary understanding of the developmental genetic regulatory architecture controlling these first cell fate decisions.

Though zygotic activation of transcription is initiated early at the 1- to 2-cell stage (Schultz, 2002), cellular differentiation occurs much later after embryonic compaction. While early cleavage patterns appear to bias blastomeres to a particular fate (Jedrusik et al., 2008), the formation of the blastocyst is primarily regulative in nature, with all 16-cell blastomeres retaining the ability to contribute to any of the three blastocyst cell lineages (Rossant and Lis, 1979; Rossant and Vijn, 1980; Suwinska et al., 2008; Ziomek et al., 1982), position within the morula being the most significant contributor to eventual cell fate decisions (Rossant and Tam, 2009).

The TE is first to form at approximately the 32-cell stage (E3.25–E3.5). It is derived from the maturation of cells positioned on the outer surface of the morula (Johnson and McConnell, 2004). Tead4 is known to be the earliest required transcription factor (TF) for TE formation; though ubiquitously expressed in the preimplantation embryo, it is functionally activated via the Hippo pathway (Nishioka et al., 2008, 2009; Yagi et al., 2007). Lying downstream of Tead4, Cdx2 and Gata3 are two TE-specific TFs coexpressed from the 8-cell stage through to the blastocyst (Home et al., 2009; Ralston et al., 2010; Ralston and Rossant, 2008; Strumpf et al., 2005). Their expression is not initially restricted to outside cells but subsequently becomes restricted to the nascent TE. Though neither is essential for TE specification, both drive trophoblast-specific gene expression.

The EPI and PE are formed from the inner cell mass (ICM). Though the mechanism of formation of these two lineages remains unclear (Rossant and Tam, 2009), it has been shown that these ICM cell fates are largely determined prior to positioning of the PE on the blastocoel-facing surface of the ICM (Chazaud et al., 2006; Gerbe et al., 2008; Plusa et al., 2008), with the most comprehensive gene expression analysis of this done by single-cell microarray analysis (Kurimoto et al., 2006).

The molecular control of these ICM fate decisions is most understood with respect to the EPI. Oct4, Sox2, and Nanog are TFs all known to be essential for the formation and/or maintenance of the EPI, and null embryos for each of these are

incapable of giving rise to ES cells (Avilion et al., 2003; Mitsui et al., 2003; Nichols et al., 1998). Oct4 and Nanog null blastocysts have neither EPI nor PE characteristics but give rise only to trophoblast derivatives (Nichols et al., 1998; Silva et al., 2009), whereas in Sox2 null embryos the ICM is initially thought to be established but the EPI cells are not maintained and at least partially become fated to the PE (Avilion et al., 2003).

Both Gata6 and Gata4 are early markers of the PE (Chazaud et al., 2006; Kaji et al., 2007; Plusa et al., 2008) and are capable of driving the expression of a PE program when overexpressed in ES cells (Fujikura et al., 2002), but neither of the respective knockouts lead to PE defects. PE-defective phenotypes are evident in knockouts involving FGF signaling. *Fgf4* (ligand), *Fgfr2* (receptor), and *Grb2* (a downstream effector) null embryos lack PE and/or its postimplantation derivatives (Arman et al., 1998; Chazaud et al., 2006; Feldman et al., 1995). This indicates an essential role for FGF signaling in the differentiation of ICM cells into PE, but it has not been clear how this is established as there is no evidence for differential expression of these components within the early ICM.

Recent reports describing coexpression of lineage-restricted TFs in individual blastomeres within the morula (Dietrich and Hiiragi, 2007; Plusa et al., 2008; Ralston and Rossant, 2008) emphasize the need to more comprehensively define expression patterns that accompany the generation of the three distinct blastocyst cell types. Therefore, we sought to analyze at the single-cell level an extensive set of regulators that, in combination, may better define the events involved in the development of the blastocyst. Our focus was on capturing the first transcriptional differences, as these would precede commitment to particular cell fates. We analyzed mRNA levels of 48 genes in parallel in over 500 individual cells from the 1-celled zygote through to the 64-celled blastocyst, thus providing unprecedented insight into the earliest cell fate decisions of the developing mouse embryo.

RESULTS

Identification of Candidate Transcription Factors for Single-Cell Analysis

As TFs are essential for defining cell-type specific phenotypes, we first established a high-quality data set containing the preimplantation expression dynamics of a comprehensive set of these regulatory molecules. We first identified the TFs represented within the 131,845 expressed sequence tags (ESTs) derived from mouse preimplantation embryos in the UniGene database (Wheeler et al., 2003) at NCBI. Of the 2348 predicted TFs encoded in the mouse genome (Panther data base; www.pantherdb.org), 885 (38%) had at least one preimplantation EST, indicating the transcriptional complexity of this brief period in development. From these and additional TFs identified from ESTs generated from ES and trophoblast stem (TS) cell libraries, two cell lines derived from the blastocyst (Evans and Kaufman, 1981; Tanaka et al., 1998), we selected 802 TFs to profile by TaqMan real-time PCR (see Supplemental Information available online).

We analyzed mRNA levels at seven developmental time points (oocyte, 2-cell, 4-cell, 8-cell, morula, E3.5 blastocyst, and E4.25 blastocyst) in addition to E4.25 ICMs isolated by immunosurgery. The data (Table S1 and deposited at www.informatics.jax.org under accession number J:140465) provide comprehensive

gene expression dynamics through preimplantation development in addition to providing cell-type specific expression (TE versus EPI/PE) within the blastocyst. Whole-blastocyst versus ICM comparisons provide specificity scores for TE and EPI/PE expression, as validated by the fact that known TE markers are found to have high TE-specificity scores (Figure S1A). Although this analysis cannot differentiate between the PE and EPI cell types within the ICM, we were able to assign some ICM-restricted genes as probable EPI or PE markers based on data from single-ICM cell microarray experiments (Kurimoto et al., 2006).

With this temporal and spatial expression data, we could begin to identify the molecular dynamics of cell-type formation as the zygote transitions into the three distinct cell types within the blastocyst. The challenge in defining cell types as they form through development was apparent from the complexity of these TF expression profiles. By tracking multiple markers, with specific expression within the E4.25 blastocyst, for any one of the three distinct cell-types, it was evident that there were at least three distinct developmental expression patterns emerging (Figures S1C and S1D). Using TE-specific TFs as an example, *Tcfap2c* levels remained unchanged throughout preimplantation development, *Cdx2* and *Gata3* were upregulated at the 8-cell-to-morula transition, and *Gata2* and *Tcfap2a* were upregulated at the morula-to-blastocyst transition.

Single-Cell Expression Profiling

Our initial data expanded the known repertoire of cell-type-specific TFs in the blastocyst and hinted at the complex transcriptional dynamics that led to this expression. These data, however, were generated from pools of cells, and, as cell fate decisions are made by individual cells, this averaged expression may mask interesting single-cell dynamics. Thus we next analyzed expression at the single-cell level. Considering the complexity of preimplantation developmental expression profiles, we felt it was essential to profile multiple genes in parallel to capture cell identity based on the expression of an oligarchy of regulatory genes rather than just one or two. We used 48.48 Dynamic Array chips (Fluidigm), which allow quantitative analysis of 48 genes in parallel at the single-cell level. The analysis involved the manual separation of single blastomeres/cells, followed by lysis, cDNA synthesis, and sequence-specific preamplification in a single tube, and quantitation of gene expression using TaqMan real-time PCR on the BioMark system (Fluidigm).

Crucial to the success of this was the selection of genes to profile. We primarily focused on TFs contained within our initial screen under the assumption that these are likely drivers of cellular fate. The remaining genes were selected based on known differential expression within the blastocyst, or because of known function in early development, or both. The final 48 genes (Table S2), including 27 TFs, were selected from 129 genes that we initially tested for their utility in blastocyst single-cell gene expression detection.

Using this 48 gene set, we initially analyzed a total of 442 single cells obtained at different stages of development from zygote through blastocyst (Table S3). The biological reproducibility of the data is clear among cells obtained from ~64-celled blastocysts, leading to unambiguous assignments of most cells to a particular cell type. Hierarchical clustering of the expression profiles from the 159 cells obtained from ~64-cell blastocysts

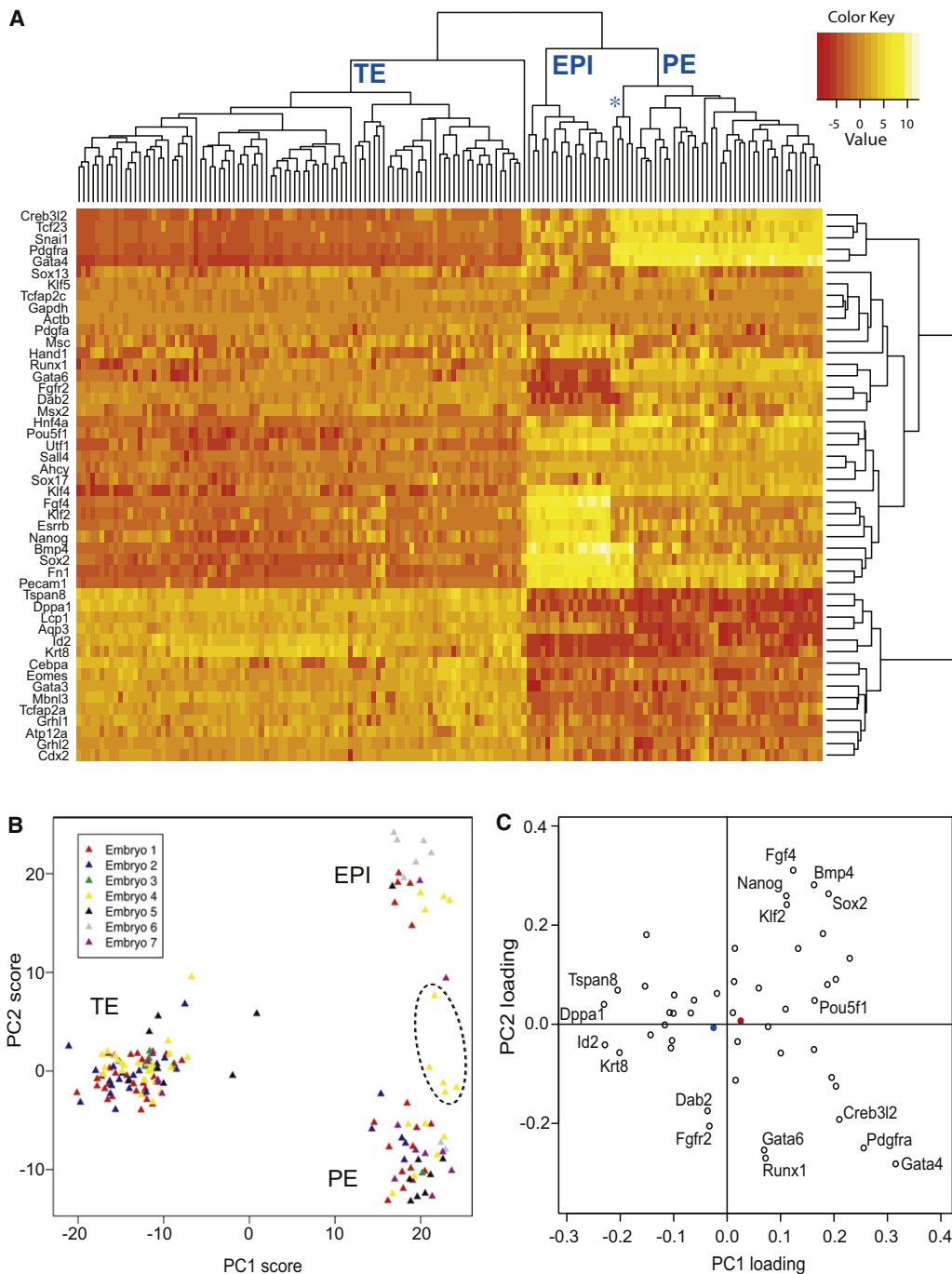


Figure 1. Three Molecularly Defined Populations at the ~64-Cell Stage

(A) A heat map of expression levels for 48 genes (see also Figure S1 and Table S2 for how these were selected) from 159 individual cells collected from ~64-cell stage blastocyst. Cells are defined as trophoblast (TE), epiblast (EPI), and primitive endoderm (PE) based on their expression of known markers *Cdx2*, *Nanog*, and *Gata4*, respectively. The asterisk (*) marks five transitional cells with PE and EPI expression characteristics.

(B) Principal component (PC) projections of the 159 ~64-cell stage cells colored according to their embryo of origin. Encircled by a dashed line are the same five cells marked by an asterisk in (A).

(C) PC projections of the 48 genes, showing the contribution of each gene to the first two PCs. The first PC can be interpreted as discriminating between TE and ICM; the second between PE and EPI. The position of endogenous control genes *Actb* (blue) and *Gapdh* (red) are shown.

unequivocally reveals the three cell types known to exist at this stage (Figure 1A). Ninety-five cells (60%) were highly enriched in TE-specific markers such as *Cdx2* and *Krt8*. Forty cells

(25%) were specifically enriched in the PE markers *Gata4* and *Pdgra*, and eighteen cells (11%) were specifically enriched in EPI-restricted genes including *Nanog* and *Sox2*. Interestingly,

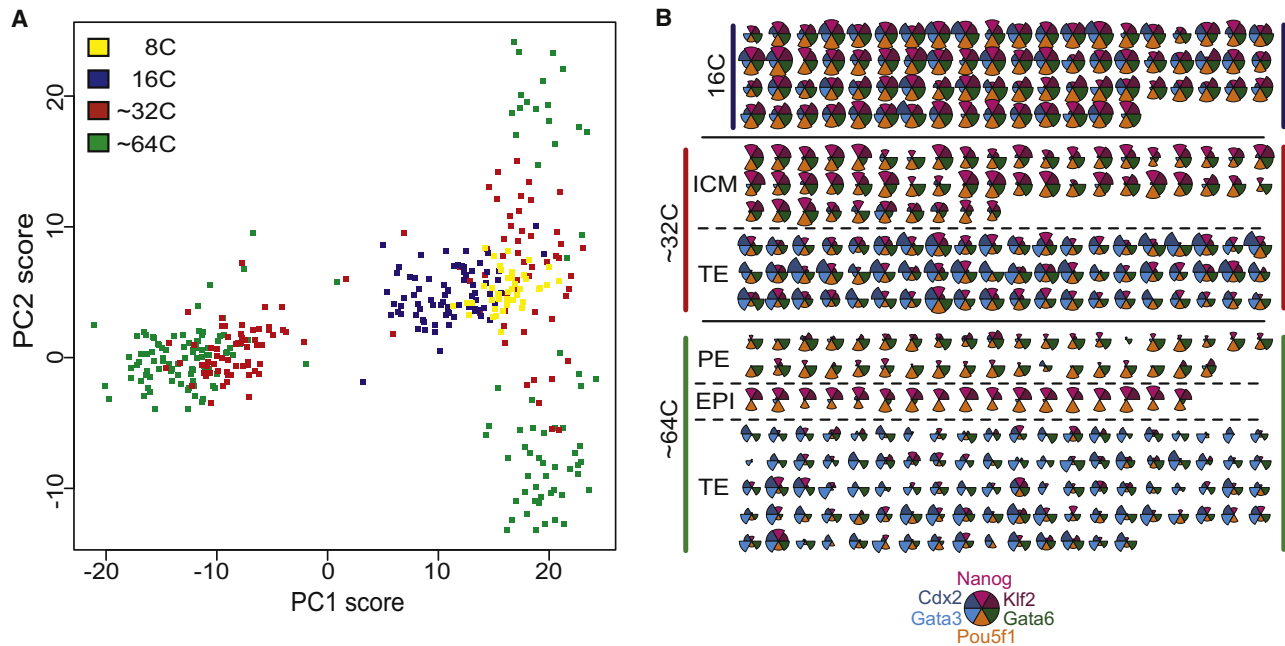


Figure 2. Developmental Progression to Three Distinct Blastocyst Cell Types

(A) Position, based on the expression of 48 genes, of individual cells from 8-, 16-, 32-, and 64-cell stage embryos projected onto the first two PCs of the 64-cell stage data (see Figure 1B). See also Figure S2 for earlier stages.

(B) Expression levels from 8-cell through to the ~64-cell stage for six lineage specific TFs—*Klf2/Nanog* (EPI), *Cdx2/Gata3* (TE), *Gata6* (PE), and *Pou5f1* (EPI/PE)—are plotted as “slices” of “pies” representing individual cells. The radius of a slice reflects the expression level of the TF. The cells from the ~32-cell stage and the ~64-cell stage are subdivided into TE and ICM or TE, EPI, and PE based on their projected positions in (A). Gene expression levels are background-subtracted based on a titration series (see Supplemental Experimental Procedures) and not normalized to endogenous controls.

a small set of five cells, all from the same embryo, expressed some markers of both EPI and PE, suggesting cells in transition. Only one of the 159 cells obtained from ~64-cell embryos appeared truly anomalous, not closely aligning to any of the three cell types.

To better visualize the data, we used principal component analysis (PCA). PCA takes data points in a high dimensional space and defines new axes (components) that cut across that space such that the first component captures as much of the variance in the data as possible, the second component (orthogonal to the first) captures as much of the remaining variance, and so on. Here, the data points are cells, and the space is 48-dimensional (48 genes), with the coordinate in each dimension being the normalized gene expression value for a given gene in that cell. Importantly, because the components cut across this 48D space, each component has contributions from all of the 48 genes. Applied to the expression data derived from the 159 cells of the ~64-celled embryos, we find that the first principal component (PC1) explains 58.6% of the observed variance while the second principal component (PC2) explains 13.5%. A projection of the cells’ expression patterns onto PCs 1 and 2 clearly separates individual cells into three distinct clusters (Figure 1B). The TE cluster (characterized, for example, by *Cdx2* expression) can be distinguished from the ICM cells along the PC1 axis, whereas the two different types of ICM cells (EPI, characterized by *Nanog*, and PE, characterized by *Gata4*) can be distinguished from one another along the PC2 axis. Note that cells were not identified based on their position in the

dissected embryo but instead are defined based on key features of their expression pattern.

Because PCA components consist of contributions from all 48 genes, it is possible to identify the most information-rich genes in classifying the three cell types at the 64-cell stage (Figure 1C). Though technical artifacts may mask the extent of *Cdx2* and *Gata3* specificity (see Supplemental Information), by this criterion *Id2*, *Dppa1*, *Tspan8*, and *Krt8* are the most specific markers of the TE. For the EPI it is *Fgf4*, *Bmp4*, *Sox2*, *Nanog*, and *Klf2*, and for PE, *Gata4*, *Pdgfra*, and *Creb3l2*. Other genes are characteristic of two of the cell types but absent or poorly expressed in the third. For example, the archetypal pluripotent TF *Pou5f1/Oct4*, though enriched in the ICM, does not differentiate between the EPI and PE lineages at this stage, as it is equally expressed in these cell types. Similarly, genes such as *Gata6*, *Fgfr2*, and *Dab2* are not expressed in the EPI lineage, but are highly expressed in both TE and PE lineages. These expression patterns are reflected in the intermediate positions that these genes assume in the PCA projection (Figure 1C).

Mixed Lineage Expression in Individual 16-Cell Blastomeres

Cellular differentiation is thought to initiate after compaction and polarization events occurring in the late 8-cell stage (Johnson and McConnell, 2004). Consistent with this idea, we found no distinguishing characteristics among individual cells at the 2-, 4-, or 8-cell stages (Figure S2). To visualize the changes that occur subsequently, we projected the expression patterns of cells

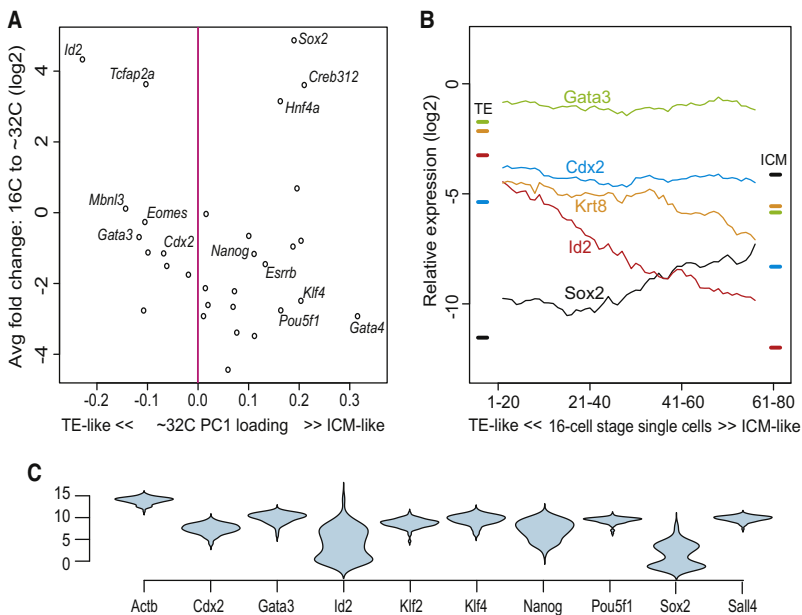


Figure 3. Heterogeneous Expression of Sox2 and Id2 in the Morula

(A) The average fold-change in expression between single cells from the 8-cell stage and either the ICM or TE cells of the 32-cell stage plotted against ICM/TE specificity. The average fold change was computed using either the TE or ICM ~32-cell average depending on which was higher. The ICM/TE specificity is measured by each gene's contribution to PC1 (Figure 1C).

(B) *Id2* and *Sox2* expression levels vary across the 16-cell stage population and are negatively correlated. The 16-cell stage cells on the x axis are sorted according to their projection score for PC1 (based on the expression of 48 genes) so that TE-like cells are on the far left and ICM-like cells are on the far right. The traces represent moving averages of the given gene's expression level in overlapping windows of 20 cells. Expression levels are given relative to endogenous controls. Colored bars marked "TE" and "ICM" show the average expression levels in ~32-cell stage TE and ICM, respectively.

(C) Violin plot representation of expression (fold change above background) in individual cells for a number of genes at the 16-cell stage; note the bimodal distribution of *Id2* and *Sox2*.

from the uncompact 8-cell stage through to the ~32-cell stage onto the first two PCs computed for the ~64-cell stage data (Figure 2A). The transitions through each of the three cleavage divisions captured in this developmental time frame contribute to the progression toward the PC coordinates of the three distinct cell types discernable at the ~64-cell stage. In the 8- to 16-cell transition the cells, as a group, move closer to the TE lineage. By the ~32-cell stage, most of the cells are either TE-like or resemble one of the ICM cell types, but cell identity in the ICM is not fully resolved. By the ~64-cell stage, essentially all cells are well resolved into one of the three cell types.

As already noted, the three cell types found in ~64-cell embryos are characterized by high expression of distinct sets of TFs. Interestingly, many of the TFs that subsequently become cell-type restricted in the blastocyst are coexpressed at high levels in the majority of individual 16-cell blastomeres, and the expression levels are comparable to what is later found in the lineage-restricted individual cells at the ~64-cell stage. Shown are representative markers of TE (*Cdx2*, *Gata3*), EPI (*Nanog*, *Klf2*), EPI and PE (*Pou5f1*), and of PE (*Gata6*) (Figure 2B), but this is also seen for other TFs that are subsequently lineage restricted, including *Esrrb*, *Klf4*, *Runx1*, *Snai1*, and *Grhl2* (Figure S4). Recent reports describing colocalization of *Nanog* with *Cdx2* and *Gata6* are consistent with our data (Dietrich and Hiiragi, 2007; Plusa et al., 2008). Maternal transcripts cannot account for the high mRNA transcript levels in the 16-cell blastomeres, as substantial zygotic activation of transcription before the 16-cell stage is apparent (Figure S4). Thus, for many of these TFs, it is the removal of their mRNAs in rival lineages as opposed to an increase within their own respective lineage that occurs as cell types form in the transition from the morula to the blastocyst.

Earliest Differences in Inner and Outer Cells

We next aimed to identify the earliest significant expression differences in the morula to blastocyst transition to perhaps gain some mechanistic insight into the corresponding cell fate

decisions. We began by further analyzing the data generated from single cells from 16-cell and 32-cell embryos. Cells from the latter were classified as ICM or TE based on their PCA projections, and the gene expression changes from 16-cell embryos to each of these cell types was determined. Among TFs, *Id2* and *Sox2* stood out for their exceptionally strong induction from the 16-cell to 32-cell stage, with *Id2* induction presumably occurring in nascent TE cells and *Sox2* in nascent ICM cells (Figure 3A). Again, this is in contrast to many of the other TFs such as *Nanog* and *Pou5f1*, levels for which remain virtually unchanged (though high) at the single-cell level from the 16-cell blastomeres to the 32-cell ICM.

We next looked for evidence of even earlier upregulation of *Id2* and *Sox2* in different cells within the 16-cell morula, the first stage at which there are inside and outside cells. Although we do not resolve TE and ICM cell types at this stage, there is some dispersion of cells along the PC1 axis, suggesting some differential expression (Figure 2A). Ordering the 16-cell stage cells according to their PC1 score revealed an *Id2/Sox2* inverse correlation, with *Sox2* highest in cells at the ICM-end and *Id2* highest in cells at the TE-end of the PC1 axis (Figure 3B). In comparison, the variation of the TE-specific markers *Cdx2* and *Gata3* appears to be less significant than *Id2* across the morula cell populations.

To address whether these gene expression differences were a result of differences between distinct cell populations or just stochastic noise, we analyzed our data with violin plots (Figure 3C). Population noise and gene expression noise should exhibit unimodal distribution around a reference level in these density plots, whereas a multimodal distribution is indicative of distinct gene expression differences between cell populations. As expected, the distribution in expression levels of the endogenous control *Actb* is unimodal, with a very narrow peak indicative of low variations between individual cells. Although most genes maintain such a unimodal distribution at the 16-cell stage, both *Sox2* and *Id2* clearly show

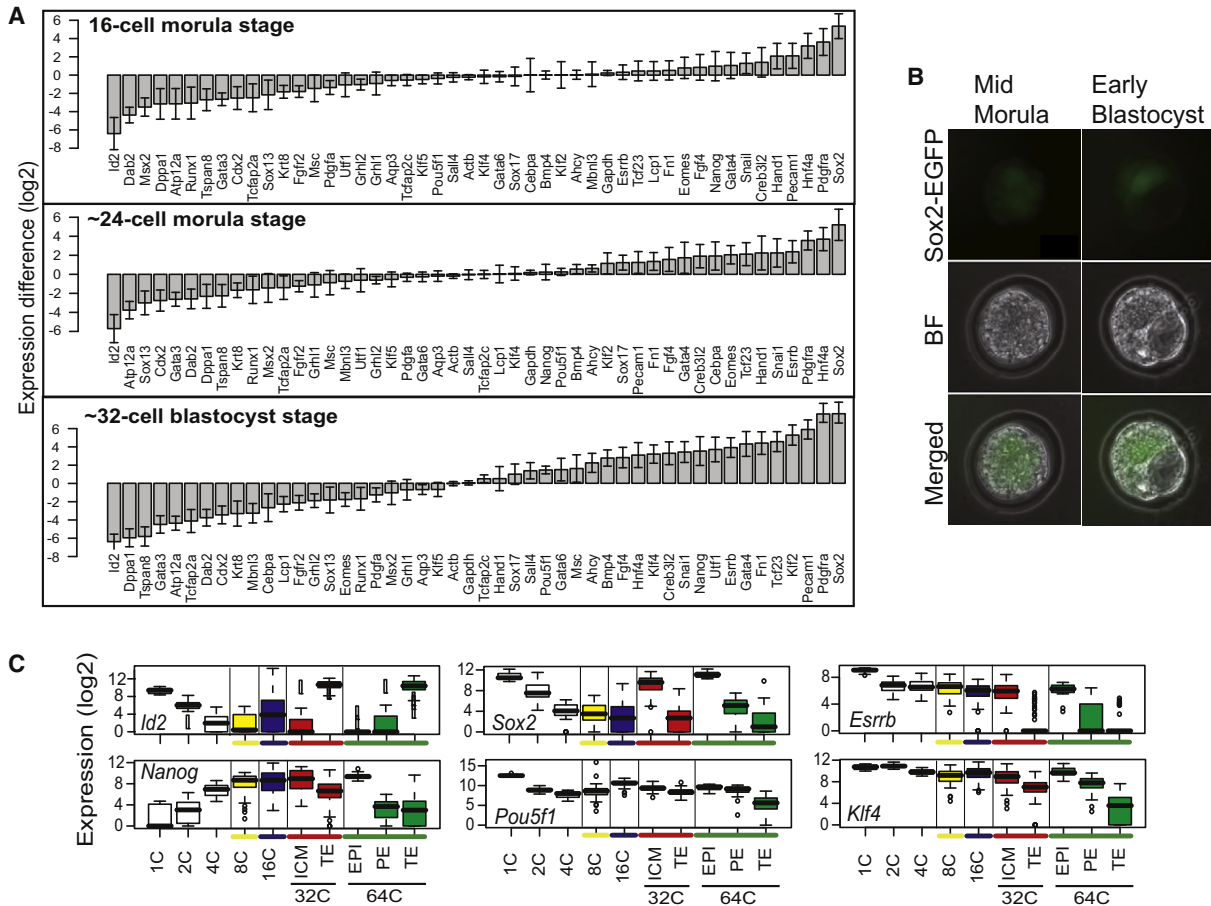


Figure 4. Earliest Detectable Gene Expression Differences between Inner and Outer Cells

(A) Expression fold-difference (log₂ value) between outer single cells and inner single cells from 16-cell morula to 32-cell blastocysts. Outer and inner cells were identified based on high and low fluorescence, respectively, by membrane labeling with a fluorescent dye applied to the embryos prior to dissociation. See also Figure S3 for principal component projections of identified cells. The error bars represent 95% confidence intervals for the difference between inner and outer cell expression.

(B) Images of morula and early blastocyst stage embryos capturing the fluorescence resulting from a paternally derived *EGFP* allele knocked in to the *Sox2* locus.

(C) Box plots of expression level distributions across developmental stages and (for the last two stages) predicted cell types for six genes. Each developmental stage is shown in a separate color. The boxed region represents the middle 50% of expression values, the black bar the median, and the “whiskers” the extreme values. Values considered outliers are represented by a circle. A background of Ct = 28 was used to obtain an absolute expression level.

bimodal distributions. Combined with the *Id2/Sox2* inverse correlation, there thus appear to be distinct populations of *Id2*^{high}/*Sox2*^{low}- and *Id2*^{low}/*Sox2*^{high}-expressing cells within the morula.

To determine if the *Id2*^{high}/*Sox2*^{low} and *Id2*^{low}/*Sox2*^{high} cell populations correlated with position within the morula, we next labeled intact embryos with a general cell membrane labeling fluorescent dye (PKH26) prior to dissociation and single-cell expression analysis. This allows for the specific marking of outside cells based on an increased fluorescence. In addition, we aimed to capture embryos in transition from the 16-cell to the 32-cell stage and thus harvested embryos within such a time frame. A total of 134 individual cells harvested from 16-cell, ~24-cell, and ~32-cell embryos were analyzed with the 48-gene set and classified as high fluorescence (outer cells) or low fluorescence (inner cells), with intermediate fluorescing cells being excluded. Not only did this analysis confirm that indeed

inverse *Id2/Sox2* expression was the earliest recognizable expression difference, but it indicated that the *Id2*^{low}/*Sox2*^{high} cells represented inner cells of the morula (Figure 4A; see Figure S3 for principal component projections). Notably, ICM markers such as *Nanog*, *Esrrb*, and *Klf2* are initially expressed equivalently in the 16-cell stage single cells but then gradually become correlated with the inner cell position and eventually show high specificity to the 32-cell ICM (Figure 4A). This result indicates that within the morula positional allocation preceded differential mRNA levels for most genes.

To provide independent confirmation of inner cell-specific upregulation of *Sox2* expression, we next analyzed mice that contained an *EGFP* reporter targeted to the *Sox2* locus (Ellis et al., 2004). To avoid confounding the analysis with oocyte-derived *EGFP*, as *Sox2* transcripts and protein are detected in the oocyte, we tracked expression of the paternally derived *EGFP* allele by mating heterozygous males with wild-type

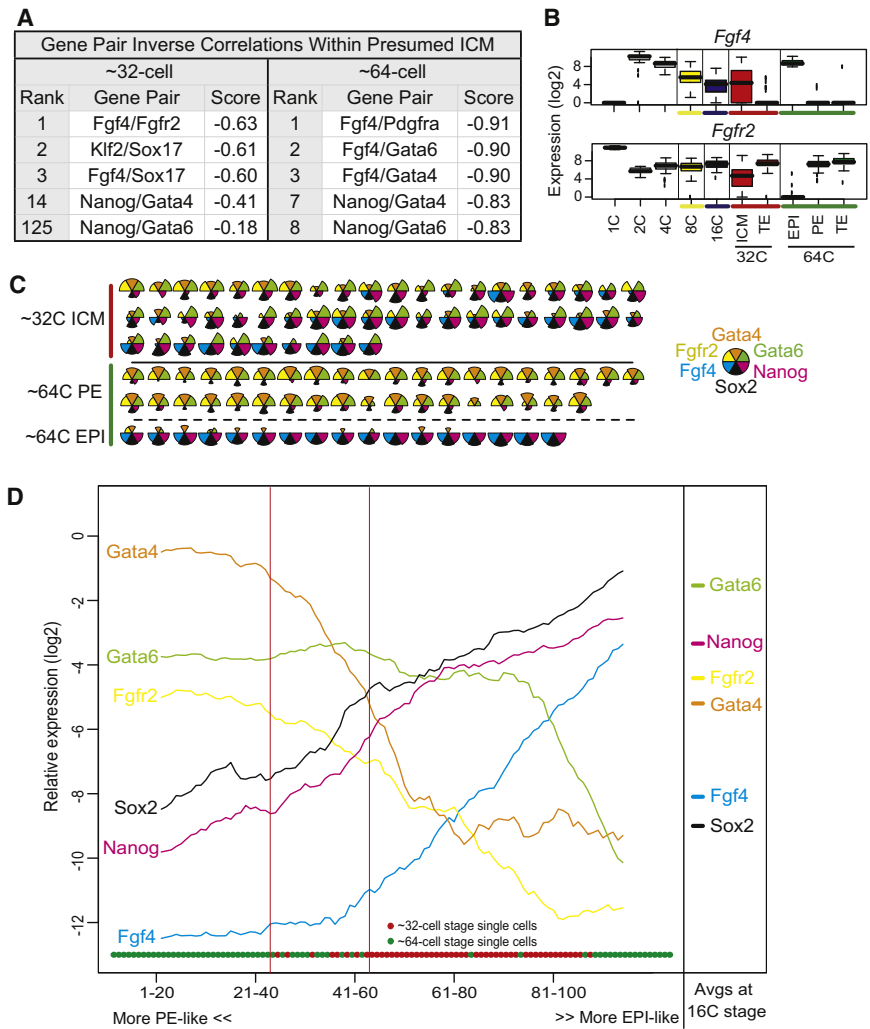


Figure 5. *Fgf4/Fgfr2* Inverse Correlation within the Early ICM

(A) The top three anticorrelated gene pairs within the ~32-cell and ~64-cell ICM in addition to gene pairs (*Nanog/Gata4*) previously described to be anticorrelated within the ICM. See also Table S6 for top 20 inverse and positive correlations.

(B) Box plots of expression level distributions across developmental stages and (for the last two stages) predicted cell types for *Fgf4* and *Fgfr2*. Box plots are generated as described in Figure 4C.

(C) Expression levels of EPI- and PE-associated signaling molecules and TFs shown as pie plots according to the same logic as Figure 2B.

(D) Variation of EPI- and PE-associated genes across ICM cells at the ~32-cell and ~64-cell stages. On the x axis, cells from both stages (~32C, red; ~64C, green) are sorted according to their projection score to PC2 (Figure 1B); based on the expression of 48 genes). The traces represent moving averages of the given gene's expression level in overlapping windows of twenty cells. Expression levels are given relative to endogenous controls. The two vertical red lines capture transitional cells including the five cells marked by an asterisk in Figure 1A. The colored side bars mark the average expression levels in 16-cell stage blastomeres.

regulation in transcript levels within the other two blastocyst lineages rather than an inner cell/EPI-specific transcriptional increase. In contrast, *Sox2* mRNA is at its lowest level at the 8-cell stage (having decreased from a high level of maternal transcripts) and is only then increased in inner cells in 16-cell embryos. Thus, in

females. Thus resulting embryos would be either wild-type or heterozygous for *Sox2*^{EGFP} in which EGFP expression is derived from the paternal allele. At the 8-cell stage, none of the embryos expressed EGFP. EGFP expression was first detectable in approximately 50% of the embryos (the expected fraction) in inner cells of the late morula, and this expression subsequently remained restricted to the ICMs of the early blastocyst (Figure 4B), thus independently confirming, at least from the paternal allele, the inner cell-specific increase in *Sox2* mRNA levels detected in our single-cell analysis.

As *Sox2* is one of the key TFs in the pluripotency regulatory network, evidenced by the lack of maintenance of the EPI in *Sox2* null embryos (Avilion et al., 2003) and its utility in inducing reprogramming of differentiated cells to a pluripotent state (i.e., iPS cells) (Takahashi and Yamanaka, 2006), our finding of its inner cell-specific upregulation is highly significant with respect to the establishment of the endogenous pluripotency regulatory network. This expression is in stark contrast to other pluripotency and reprogramming factors. *Oct4* (*Pou5f1*), *Nanog*, *Esrrb*, and *Klf4* are all abundantly expressed in virtually all blastomeres of the 8-cell and 16-cell embryos (Figure 4C). Restriction to EPI cells for these TFs occurs as a result of down-

regulation in transcript levels within the other two blastocyst lineages rather than an inner cell/EPI-specific transcriptional increase. In contrast, *Sox2* mRNA is at its lowest level at the 8-cell stage (having decreased from a high level of maternal transcripts) and is only then increased in inner cells in 16-cell embryos. Thus, in the path from the fertilized egg to the pluripotent ground state (Nichols and Smith, 2009), *Sox2* is the last of the reprogramming factors to be transcriptionally activated, but the first to mark cell population differences in the developing embryo.

From Inner Cells to EPI and PE

The ICM gives rise to the EPI and PE. To gain insight into the mechanism of EPI/PE segregation, we returned to our single-cell data to identify the earliest expression differences within the ICM. These differences should be evident in cells classified as ICM along PC1 (Figure 2A) in the transition from the ~32-cell to the ~64-cell blastocyst. Therefore, we calculated the correlation in gene expression for each of the 1035 pairs of genes (excluding *ActB* and *Gapdh*, which are used for normalization) using the 50 ICM cells in the ~32-cell data set and, separately, the 55 ICM cells in the ~64-cell data set.

Consistent with salt and pepper staining for *Nanog* and *Gata4/6* within the E3.5 ICM (Chazaud et al., 2006; Plusa et al., 2008), we find that the *Nanog/Gata4* gene pair and the *Nanog/Gata6* gene pair expression levels are inversely correlated in 64-cell ICM cells but show insignificant or weak correlations in cells of the 32-cell ICM (Figure 5A). Indeed, virtually all 32-cell

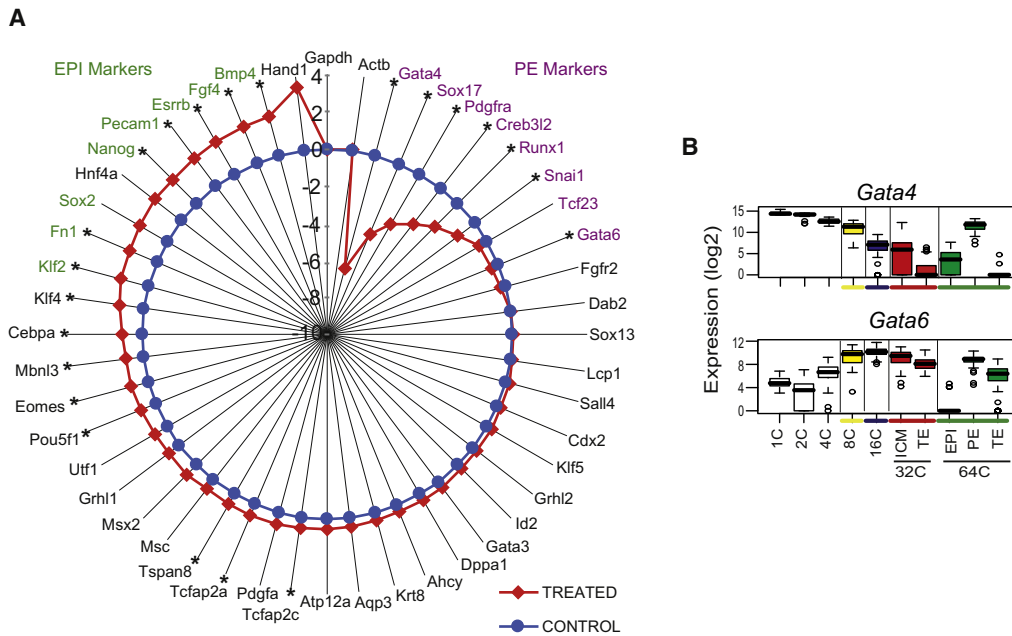


Figure 6. Inhibition of FGF Signaling Blocks the Upregulation of the PE Transcriptional Program

(A) TaqMan real-time PCR gene expression analysis generated from individual embryos treated with 10 μ M SU5402 from the 16-cell stage (E3.0) for 24 hr. Gene expression levels are expressed relative to vehicle controls. Each data point is the average of three biological replicates. Genes are ordered clockwise based on increasing Log2 fold change. Genes labeled with an asterisk passed a two-sided t test adjusted for multiple comparisons (Bonferroni corrected $p < 0.01$). Thus, *Gata4* represents the most significantly downregulated and *Bmp4* the most significantly upregulated gene in the treated embryos. EPI markers ($PC1 > 0.05$, $PC2 > 0.1$) are marked green and PE markers ($PC1 > 0.05$, $PC2 < -0.05$) are marked purple based on PC projections from Figure 1C.

(B) Box plots of expression level distributions across developmental stages and (for the last two stages) predicted cell types for *Gata4* and *Gata6* in the original single-cell analysis. Box plots are generated as described in Figure 4C. See also Figure S4 for similar profiles for all other genes.

ICM cells contain high mRNA levels of both *Nanog* and *Gata6*, while *Gata4* is initially low. The later PE specificity of *Gata4* is achieved by strong activation of expression in a subpopulation of ICM cells in transition from the 32-cell to the 64-cell stage (Figures 5C and 5D). EPI and PE specificity for *Nanog* and *Gata6*, respectively, is the result of downregulation in the opposing cell type in transition to the 64-cell stage (Figures 5C and 5D).

Intriguingly, the gene pair with the strongest inverse correlation among the early ICM (32-cell) cells does not involve TFs at all, but instead encodes the ligand/receptor pair *Fgf4/Fgfr2* (Figure 5A), which is highly significant considering their essential role in PE formation (Arman et al., 1998; Feldman et al., 1995). To our knowledge, this is the first evidence of differential expression of these genes within the early ICM. This inverse correlation is apparently established by a decrease in *Fgfr2* (highly expressed at the 16-cell stage) and an increase of *Fgf4* (low at the 16-cell stage) in a subpopulation of 32-cell ICM cells (Figures 5B–5D). Within the 64-cell ICM, *Fgf4* is restricted to the EPI and *Fgfr2* to the PE (Figures 5B–5D).

To position FGF signaling within the context of the ICM developmental regulatory network, we next treated embryos during the morula (E3.0) to blastocyst (E4.0) transition with the specific FGFR inhibitor SU5402, a treatment known to improve ES cell derivation efficiency (Ying et al., 2008). We analyzed individual blastocysts with our repertoire of lineage markers and, though not providing as comprehensive insight as single-cell analysis,

we can infer lineage-specific transcriptome changes based on our wild-type single-cell data. Remarkably, all the inhibitor-treated embryos indicated significantly lower expression of the majority of the PE-specific developmental regulators (e.g. *Gata4*, *Sox17*) compared to the vehicle controls. This was in contrast to the EPI-specific markers (e.g., *Nanog* and *Esrrb*), many of which showed a 2- to 4-fold upregulation (Figure 6A). TE-specific genes such as *Cdx2*, *Gata3*, and *Krt8* were unaffected by FGF signal inhibition. Interestingly, levels of the PE marker *Gata6* remained virtually unaltered, suggesting this PE-specific marker's expression is independent of FGF signaling. An analysis of the developmental expression of *Gata6* clearly indicates that its elevated expression precedes the *Fgf4/Fgfr2* inverse correlation (Figure 6B). This is consistent with the findings of Plusa et al. (2008). This *Gata6* expression is in contrast to that for *Gata4* (Figure 6B) and other PE markers (Figure S4) in which the PE-specific upregulation occurs after the *Fgf4/Fgfr2* inverse correlation is established. This block in PE formation by FGF inhibition is supported by another recent study (Nichols et al., 2009). Our data position FGF signaling upstream of the transcriptional upregulation of the PE-specific TFs *Gata4*, *Sox17*, and *Creb3l2*.

DISCUSSION

Here we apply single-cell analysis to capture the expression dynamics of 48 genes from hundreds of cells harvested over

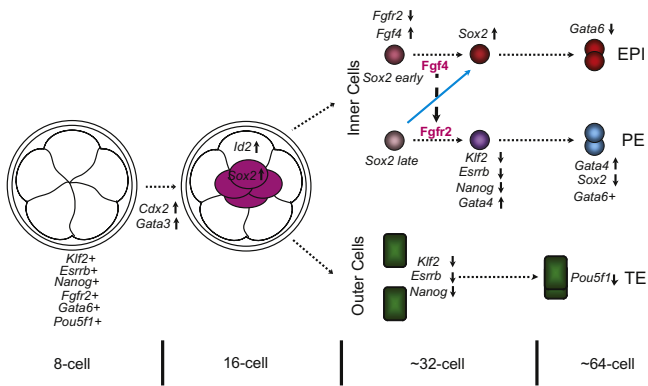


Figure 7. Schematic Model of Gene Expression Changes in the Development of the Three Blastocyst Lineages

The blue arrow indicates fate of PE progenitors upon inhibition of FGF signaling.

the first 4 days of mouse development. The single-cell analysis of these developmental regulators across multiple developmental time points provides a rich data set that allows us to visualize the formation of the first three cell types, salient features of which are highlighted in our model (Figure 7).

We see little consistent difference in the expression of the 48 genes we analyzed among individual blastomeres up to and including the uncompact 8-cell stage. At the 8-cell stage virtually all blastomeres are expressing a number of lineage-specific TFs (e.g., *Nanog*, *Pou5f1*, and *Gata6*) at a level equivalent to that found later in the lineage-restricted blastocyst cell-types. The transition to the morula sees the upregulation of two key TE TFs, *Gata3* and *Cdx2* (Home et al., 2009; Ralston et al., 2010; Strumpf et al., 2005). Though the direct molecular wiring has yet to be identified, *Cdx2* and *Gata3* are dependent on *Tead4*, thus implicating the Hippo signaling pathway in their activation (Nishioka et al., 2009; Ralston et al., 2010; Yagi et al., 2007).

At the 16-cell stage, we find that not all blastomeres express equivalent levels of mRNA for these TE TFs, perhaps reflecting their previous cleavage history (Jedrusik et al., 2008). Nonetheless, many 16-cell blastomeres do coexpress high levels of these and many other lineage-restricted TFs, representative of all three blastocyst lineages. This finding potentially provides the molecular foundation for the known developmental plasticity of 16-cell blastomeres (Rossant and Lis, 1979; Rossant and Vihj, 1980; Suwinska et al., 2008). Importantly, our findings indicate a decrease of transcript levels in cells of opposing lineages for many of these TFs, rather than a lineage-restricted increase, as blastocyst cell types develop from the uncommitted blastomeres of the morula.

Two key observations in our work, achieved by single-cell analysis of many cells, are the inner cell-specific upregulation of *Sox2* and the inversely correlated expression of *Fgf4*/*Fgfr2* within the early ICM. All three molecules are essential to fate decisions within the ICM (Arman et al., 1998; Avilion et al., 2003; Feldman et al., 1995), and Fgf signaling is essential to the establishment of the PE (Chazaud et al., 2006; Nichols et al., 2009). Direct molecular wiring linking these events is plausible, as *Sox2* has been identified to bind directly to both *Fgf4* and *Fgfr2* in pluripotent cells (Chen et al., 2008).

Inner cells do not all form at the same time (Fleming, 1987; Pedersen et al., 1986), thus temporal differences in inner cell formation potentially create temporal differences in *Sox2* upregulation. We would propose that such differences create the heterogeneity within the developing ICM that subsequently leads to the epiblast and primitive endoderm fate choices (Figure 7). In this model, elevated *Sox2* mRNA levels in the earliest formed inner cells would subsequently lead to the direct transcriptional activation of *Fgf4* (Yuan et al., 1995) and downregulation of *Fgfr2* (Masui et al., 2007). Inner cells generated later, and already expressing *Fgfr2*, are exposed to the nascent *Fgf4* signal, the immediate effect of which is the downregulation of a module of pluripotency regulators (i.e., *Nanog*, *Esrrb*, and *Klf2*) and transcriptional activation of PE TFs such as *Gata4*. In addition to the downregulation of *Fgfr2*, cells expressing *Sox2* earlier would be in a position to replenish depleting levels of maternal *Sox2*, thus maintaining the expression of key pluripotency *Sox2* target genes such as *Nanog* (Rodda et al., 2005; Silva et al., 2009).

In summary, the insights we have obtained in this study are a consequence of analyzing many genes in parallel at the single-cell level over developmental time. Multiple genes provide a more accurate view of cellular phenotypes, a critical point as developmental decisions are made at the cellular level and are likely affected by the relative expression levels of many genes. Our single-cell analysis offers intriguing new insights into the formation of the mammalian blastocyst. Position and signaling events appear to precede the segregation of lineage-specific transcriptional programs in these earliest of fate decisions. In general, the application of these methods to other biological systems will no doubt add clarity to the underlying molecular mechanisms controlling cellular behavior.

EXPERIMENTAL PROCEDURES

Embryo Collection and Culture

The BRC IACUC (Biopolis) approved all mouse work. Embryos were derived from superovulated F1 (B6CBAF1) females crossed to CD-1 males and collected in M2 medium. Extended embryo culture was in KSOM medium under paraffin oil in 5% CO₂ at 37°C. For the 802 TF screen, pools of 150 freshly harvested embryos for each developmental stage and isolated ICMs (via immunosurgery) were used. For the single-cell study, embryos were collected at the 1-cell (E0.5), 2-cell (E1.5), and 8-cell (E2.5) stages and either directly used or cultured in vitro to the appropriate stages (i.e., 2-cell to 4-cell or 8-cell to each subsequent stage). One-cell embryos were treated with hyaluronidase to remove cumulus cells. Cell numbers were confirmed by single-cell counting after dissociation. Not all cells from all embryos were successfully harvested. The total cell numbers and embryo numbers analyzed for each stage are summarized in Table S3. For FGF inhibition experiments, embryos were treated with 10 μM SU5402 (Calbiochem) or DMSO from the morula (E3.0) to expanded blastocyst stage (E4.0).

Single-Cell Isolation

After removal of zona pellucida (acid tyrode's solution), embryos were incubated in trypsin-EDTA for 10 min at 37°C and transferred into PBS-FBS buffer. Single cells were manually collected by mouth pipette aided by a finely pulled glass tip. Labeling of outer cells was by incubation in PKH26 (Red Fluorescent Cell Linker Mini Kit, Sigma) for 3 min prior to segregation. Labeled single cells were inspected under a fluorescent microscope (Axio Observer D1, Zeiss) and classified as "outer (bright)," "inner (dark)," or "undetermined (weak signal)" based on fluorescent intensity.

TaqMan Gene Expression Analysis

For the 802 TF study, total RNA was purified using a PicoPure RNA isolation kit (Arcturus Bioscience), and cDNA was synthesized with a high-capacity cDNA archive kit (Applied Biosystems; ABI). Total RNA from the FGFR inhibitor experiments was similarly prepared. Equal volumes of cDNA and TaqMan Universal PCR Master Mix (ABI) were combined and loaded into the ports of TaqMan custom low-density arrays following the manufacturer's instructions. TaqMan assays used are listed in Table S1. Real-time PCR was performed on Prism 7900HT Sequence Detection System 2.2 (ABI), and Ct values were calculated by the system software. High expression is defined by Ct values of 31 or lower, and undetectable as a Ct value of 35 or greater.

High-Throughput Single-Cell qPCR

Inventoried TaqMan assays (20 \times , Applied Biosystem) were pooled to a final concentration of 0.2 \times for each of the 48 assays. Individual cells were harvested directly into 10 μ l RT-PreAmp Master Mix (5.0 μ l CellsDirect 2 \times Reaction Mix (Invitrogen); 2.5 μ l 0.2 \times assay pool; 0.5 μ l RT/Taq enzyme [CellsDirect qRT-PCR kit, Invitrogen]; 2.0 μ l TE buffer). The harvested single-cell samples were immediately frozen and stored at -80°C . Cell lysis and sequence-specific reverse transcription were performed at 50°C for 20 min. The reverse transcriptase was inactivated by heating to 95°C for 2 min. Subsequently, in the same tube, cDNA went through sequence-specific amplification by denaturing at 95°C for 15 s, and annealing and amplification at 60°C for 4 min for 18 cycles. These preamplified products were diluted 5-fold prior to analysis with Universal PCR Master Mix and inventoried TaqMan gene expression assays (ABI) in 48.48 Dynamic Arrays on a BioMark System (Fluidigm). Ct values were calculated from the system's software (BioMark Real-time PCR Analysis; Fluidigm). cDNA from the FGFR inhibitor experiment was identically analyzed but using 1/8 of the total cDNA and running 16 cycles of preamplification.

Single-Cell Data Processing

All Ct values (Tables S4 and S5) obtained from the BioMark System were converted into relative expression levels by subtracting the values from the assumed baseline value of 28. Cells with low or absent endogenous control gene expression levels were removed from analysis ($\sim 10\%$). The resulting values were at times normalized to the endogenous control by subtracting, for each cell, the average of its *Actb* and *Gapdh* expression levels. As the Ct scale is logarithmic (a difference of one Ct corresponds to a doubling of measured transcript), a subtraction of the average of two genes on this scale corresponds to taking the geometric mean on a linear scale. Data shown in Figures 1A–1C, 2A, 3A, 3B, 4A, 5A, and 5D have been normalized against endogenous controls.

Single-Cell Data Visualization

Principal component analysis was performed using the *svd* command in R on $\sim 64\text{C}$ expression data, normalized against endogenous controls as above, from which the mean expression levels for each gene had been subtracted. For projections of data from other developmental stages onto the $\sim 64\text{C}$ principal components, the $\sim 64\text{C}$ means were subtracted from the expression data for consistency. The *star* command in R was used for simultaneously visualizing the expression levels of multiple TFs. Here, to facilitate comparisons, expression levels were converted to a gene-specific scale where the lowest expression level of the gene at $\sim 64\text{C}$ corresponds to 0 and the highest expression level at $\sim 64\text{C}$ corresponds to 1. Thus, the radii of the "pies" in Figures 2B and 5C can be compared across cells (and time points). For constructing Figures 2B, 3B, 4C, 5A–5D, and 6B, we used principal component analysis to classify cells into hypothetical cell types. At the $\sim 64\text{C}$ stage, the cells are classified as TE, PE, EPI, or uncertain, while at the $\sim 32\text{C}$ stage, they are classified as TE, ICM, or uncertain. For the $\sim 64\text{C}$ stage, cells with a PC1 score below 0 were classified as TE, cells with PC1 score ≥ 10 and a PC2 score ≥ 10 were classified as EPI, and cells with PC1 score ≥ 10 and PC2 score < -5 were classified as PE; cells fulfilling none of these criteria were classified as uncertain. For the $\sim 32\text{C}$ stage, cells with a PC1 score < 0 were classified as TE, cells with PC1 score ≥ 10 classified as ICM, and the remainder were considered to be uncertain.

ACCESSION NUMBERS

The TaqMan low-density array transcription factor data were deposited in the Mouse Genome Informatics database under accession number J:140465.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and seven tables and can be found at [doi:10.1016/j.devcel.2010.02.012](https://doi.org/10.1016/j.devcel.2010.02.012).

ACKNOWLEDGMENTS

We would like to thank Martin Pieprzyk (Fluidigm) for technical discussion and Andrew Hutchins for critical comments on the manuscript. This work was supported by the Agency for Science, Technology and Research (A*Star) and a Singapore Stem Cell Consortium grant (SSCC-06-007).

Received: August 10, 2009

Revised: November 9, 2009

Accepted: February 8, 2010

Published: April 19, 2010

REFERENCES

- Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J.K., and Lonai, P. (1998). Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. *Proc. Natl. Acad. Sci. USA* 95, 5082–5087.
- Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* 17, 126–140.
- Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev. Cell* 10, 615–624.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106–1117.
- Dietrich, J.E., and Hiiragi, T. (2007). Stochastic patterning in the mouse pre-implantation embryo. *Development* 134, 4219–4231.
- Ellis, P., Fagan, B.M., Magness, S.T., Hutton, S., Taranova, O., Hayashi, S., McMahon, A., Rao, M., and Pevny, L. (2004). SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. *Dev. Neurosci.* 26, 148–165.
- Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotent cells from mouse embryos. *Nature* 292, 154–156.
- Feldman, B., Poueymirou, W., Papaioannou, V.E., DeChiara, T.M., and Goldfarb, M. (1995). Requirement of FGF-4 for postimplantation mouse development. *Science* 267, 246–249.
- Fleming, T.P. (1987). A quantitative analysis of cell allocation to trophectoderm and inner cell mass in the mouse blastocyst. *Dev. Biol.* 119, 520–531.
- Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., Miyazaki Ji, J., and Niwa, H. (2002). Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev.* 16, 784–789.
- Gerbe, F., Cox, B., Rossant, J., and Chazaud, C. (2008). Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst. *Dev. Biol.* 313, 594–602.
- Home, P., Ray, S., Dutta, D., Bronshteyn, I., Larson, M., and Paul, S. (2009). GATA3 is selectively expressed in the trophectoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. *J. Biol. Chem.* 284, 28729–28737.
- Jedrusik, A., Parfitt, D.E., Guo, G., Skamagki, M., Grabarek, J.B., Johnson, M.H., Robson, P., and 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, M.H., and McConnell, J.M. (2004). Lineage allocation and cell polarity during mouse embryogenesis. *Semin. Cell Dev. Biol.* 15, 583–597.
- Kaji, K., Nichols, J., and Hendrich, B. (2007). Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. *Development* 134, 1123–1132.
- Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K.D., Yamada, R.G., Ueda, H.R., and Saitou, M. (2006). An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res.* 34, e42.
- Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* 9, 625–635.
- Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 113, 631–642.
- Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. *Cell Stem Cell* 4, 487–492.
- Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 95, 379–391.
- Nichols, J., Silva, J., Roode, M., and Smith, A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* 136, 3215–3222.
- Nishioka, N., Yamamoto, S., Kiyonari, H., Sato, H., Sawada, A., Ota, M., Nakao, K., and Sasaki, H. (2008). Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech. Dev.* 125, 270–283.
- Nishioka, N., Inoue, K., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R.O., Ogonuki, N., et al. (2009). The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Dev. Cell* 16, 398–410.
- Pedersen, R.A., Wu, K., and Balakier, H. (1986). Origin of the inner cell mass in mouse embryos: cell lineage analysis by microinjection. *Dev. Biol.* 117, 581–595.
- Plusa, B., Piliszek, A., Frankenberger, S., Artus, J., and Hadjantonakis, A.K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* 135, 3081–3091.
- Ralston, A., and Rossant, J. (2008). Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Dev. Biol.* 313, 614–629.
- Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J.S., and Rossant, J. (2010). Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* 137, 395–403.
- Rodda, D.J., Chew, J.L., Lim, L.H., Loh, Y.H., Wang, B., Ng, H.H., and Robson, P. (2005). Transcriptional regulation of Nanog by Oct4 and Sox2. *J. Biol. Chem.* 280, 24731–24737.
- Rossant, J., and Lis, W.T. (1979). Potential of isolated mouse inner cell masses to form trophectoderm derivatives in vivo. *Dev. Biol.* 70, 255–261.
- Rossant, J., and Vihj, K.M. (1980). Ability of outside cells from preimplantation mouse embryos to form inner cell mass derivatives. *Dev. Biol.* 76, 475–482.
- Rossant, J., and Tam, P.P. (2009). Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* 136, 701–713.
- Schultz, R.M. (2002). The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. *Hum. Reprod. Update* 8, 323–331.
- Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. *Cell* 138, 722–737.
- Strumpf, D., Mao, C.A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* 132, 2093–2102.
- Suwinska, A., Czolowska, R., Ozdzinski, W., and Tarkowski, A.K. (2008). Blastomeres of the mouse embryo lose totipotency after the fifth cleavage division: expression of Cdx2 and Oct4 and developmental potential of inner and outer blastomeres of 16- and 32-cell embryos. *Dev. Biol.* 322, 133–144.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676.
- Tanaka, S., Kunath, T., Hadjantonakis, A.K., Nagy, A., and Rossant, J. (1998). Promotion of trophoblast stem cell proliferation by FGF4. *Science* 282, 2072–2075.
- Wheeler, D.L., Church, D.M., Federhen, S., Lash, A.E., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Sequeira, E., Tatusova, T.A., et al. (2003). Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* 31, 28–33.
- Yagi, R., Kohn, M.J., Karavanova, I., Kaneko, K.J., Vullhorst, D., DePamphilis, M.L., and Buonanno, A. (2007). Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development* 134, 3827–3836.
- Ying, Q.L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519–523.
- Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev.* 9, 2635–2645.
- Ziomek, C.A., Johnson, M.H., and Handyside, A.H. (1982). The developmental potential of mouse 16-cell blastomeres. *J. Exp. Zool.* 221, 345–355.