Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages

Yusuke Ohnishi¹, Wolfgang Huber², Akiko Tsumura³, Minjung Kang⁴, Panagiotis Xenopoulos⁴, Kazuki Kurimoto⁵,⁶, Andrzej K. Oleś², Marcos J. Araúzo-Bravo⁵, Mitinori Saitou³,⁵,⁶,⁸, Anna-Katerina Hadjantonakis⁴ and Takashi Hiiragi¹,⁹

It is now recognized that extensive expression heterogeneities among cells precede the emergence of lineages in the early mammalian embryo. To establish a map of pluripotent epiblast (EPI) versus primitive endoderm (PrE) lineage segregation within the inner cell mass (ICM) of the mouse blastocyst, we characterized the gene expression profiles of individual ICM cells. Clustering analysis of the transcriptomes of 66 cells demonstrated that initially they are non-distinguishable. Early in the segregation, lineage-specific marker expression exhibited no apparent correlation, and a hierarchical relationship was established only in the late blastocyst. Fgf4 exhibited a bimodal expression at the earliest stage analysed, and in its absence, the differentiation of PrE and EPI was halted, indicating that Fgf4 drives, and is required for, ICM lineage segregation. These data lead us to propose a model where stochastic cell-to-cell expression heterogeneity followed by signal reinforcement underlies ICM lineage segregation by antagonistically separating equivalent cells.

Mammalian preimplantation development gives rise to three lineages in the blastocyst: the EPI and two extraembryonic tissues, the PrE and trophotroctoderm. Lineage segregation between EPI and PrE occurs within the ICM of the blastocyst and involves two successive phases. First, at the morula stage (embryonic day (E)2.5; 8–16 cells), the EPI-specific transcription factor Nanog and PrE-specific Gata6 (refs 2,3) become evident and are expressed by all ICM cells. This overlapping expression persists until E3.5 (64–90 cells) when two distinct cell populations emerge as PrE precursors activate a sequence of transcription factors (Gata6, Sox17, Gata4 and Sox7; ref. 4), and EPI precursors co-express pluripotency-associated factors (for example, Nanog and Sox2). As EPI and PrE markers establish mutually exclusive expression, they become arranged in a salt-and-pepper distribution. Even though biased to a specific lineage, ICM cells exhibit a plasticity preceding their sorting to respective positions when the PrE begins to epithelialize at E4.5 (>150 cells).

In the mouse this segregation of EPI and PrE lineages is regulated by FGF/MAPK signalling. Modulation of FGF/MAPK signalling shifts the balance of EPI and PrE cells: excess of Fgf4 converts all ICM cells to adopt a PrE identity, whereas when FGF signalling is blocked, all ICM cells become Nanog-positive. How the heterogeneity in FGF signalling is established remains an open question. Two, apparently disparate, models have been proposed: a random or cleavage-history-dependent mechanism. Two-to-three asymmetric cell divisions (8-to-16-cell, 16-to-32-cell and 32-to-64-cell) generate the ICM cells. Consequently, it has been proposed that cells internalized during the first wave exhibit a greater bias towards EPI, whereas cells internalized later are biased to PrE (ref. 15). This notion was challenged by another study that showed an apparently random generation of EPI and PrE precursors, irrespectively of internalization timing. Importantly, an absolute correlation between lineage and cleavage pattern has not been evident from any study.

As the emergence of lineage precursors within the ICM is preceded by stochastic gene expression variability, we reasoned that single-cell gene expression profiling would be requisite for

¹Developmental Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. ²Genome Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. ³Institute for Integrated Cell-Material Sciences, Kyoto University, Yoshida-Ushinomiya-cho, Sakyo-ku, Kyoto 606-8501, Japan. ⁴Developmental Biology Program, Sloan-Kettering Institute, 1275 York Avenue, Box 371, New York, New York 10065, USA. ⁵Department of Anatomy and Cell Biology, Graduate School of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. ⁶JST, ERATO, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. ⁷Computational Biology and Bioinformatics Group, Max Planck Institute for Molecular Biomedicine, Röntgenstrasse 20, 48149 Münster, Germany. ⁸Center for IPS Cell Research and Application, Kyoto University, 53 Kawahara-cho, Shogoin Yoshida, Sakyo-ku, Kyoto 606-8507, Japan.

Correspondence should be addressed to T.H. (e-mail: hiiragi@embl.de)

Received 10 May 2013; accepted 18 October 2013; published online 1 December 2013; DOI: 10.1038/ncb2881
understanding the mechanisms driving lineage segregation. Recent technical advances enable quantitative gene expression profiling at the single-cell level using quantitative PCR (qPCR; ref. 16), microarrays17,18, or RNA-seq (refs 19,20). It is now widely recognized that cell-to-cell expression variation and multi-lineage gene activation exist early on during lineage commitment21–23. Recent single-cell expression studies demonstrated that the expression of key factors is independently regulated in the transition from self-renewal to lineage-committed states in haematopoiesis24,25, and that early stochastic gene expression is followed by the establishment of a hierarchy during cellular reprogramming26. Although the changes in expression during blastocyst lineage specification began to be characterized at the single-cell level using defined cohorts of genes16, a comprehensive and unbiased view is still missing. Prompted by the availability of characterized lineage-specific markers, and recent studies proposing underlying mechanisms2,3,7,13,15,27, we focused our single-cell transcriptomic analysis on the ICM cells of E3.25 (32–50 cells) to E4.5 blastocysts.

RESULTS

Single-cell analysis establishes a lineage map

To assess the inherent heterogeneities and population dynamics associated with the emergence of EPI versus PrE cells at the single-cell level, we sought to build on previous methods17,18 and extend our studies by expression profiling individual cells within the ICM of developing blastocyst-stage mouse embryos. Having formulated a method for collecting live single cells from ICMs recovered by immunosurgery28 (Fig. 1a), we established a robust protocol for the amplification of messenger RNAs from them. Embryos were staged according to the average total number of cells in littermates. Reflecting the quality of the sample preparations, the mRNA isolation protocol produced a representative amplification output for the detection of a control ‘spike’ RNA (Supplementary Fig. 1), a uniform level of expression for housekeeping genes (for example, Gapdh) and bimodal distribution of EPI and PrE lineage-specific gene expression at E4.5 (Fig. 1b). For the ensuing microarray analyses, we selected 66 single-cell samples that provided a linear output for the detection of ‘spike-in’ RNAs with as little as 20 copies (Supplementary Fig. 1 and Methods), implying that mRNAs expressed with more than 20 copies could be analysed quantitatively.

The data obtained from qPCR analysis of a total of 137 single cells ranging from E3.25 to E4.5 revealed distinct behaviours in gene expression dynamics as the two ICM lineages arise (Fig. 1b). At least two distinct mechanisms can give rise to bimodal lineage-specific gene expression. In the first, bimodal gene expression is achieved from an initial state whereby all ICM cells express certain genes, followed by resolution into mutually exclusive lineage-specific patterns, presumably through lineage-specific gene repression. This was the case for Sox2 and Gata6, in agreement with previous findings19. Alternatively, we noted cases where an initially negligible level of gene expression evolves into lineage-specific gene activation and mutually exclusive expression. This was the case for Gata4. Notably, the expression of the Fgf4 gene was detected only in some cells at E3.25, therefore presaging the segregation of EPI or PrE progenitors at E3.5.

Among the 154 single-cell samples (see Methods for details), complementary RNAs derived from the highest quality 66 individual ICM cells (as assessed by expression of spike RNA) were hybridized to the GeneChip Mouse Genome 430 2.0 arrays. Overall, 10,958 distinct mRNAs were detected above background in these samples. The single-cell data established a transcriptome map of lineage segregation between EPI and PrE in the mouse blastocyst. To visualize the main features of this map, we used principal component (PC) projections of individual cells based on the expression of the 100 most variable genes in all cells (Fig. 1c). In this map, PC1 approximately corresponded to the stage of development (time), whereas PC2 aligned with the lineage difference (EPI or PrE). These data reveal that the EPI and PrE lineages become progressively segregated within a cohort of initially equivalent ICM cells during E3.25–E4.5 blastocyst stages.

Unsupervised clustering of the data obtained from single ICM cells at E3.5 and E4.5 (22 and 8 cells, respectively) using the expression of the 100 most variable genes identified two stable clusters, which we conclude corresponded to EPI and PrE lineages on the basis of the expression of markers for each lineage. Thus, these data collectively provide the most comprehensive unbiased list of markers for EPI or PrE lineage at E3.5 and E4.5 (Supplementary Table 1). An unsupervised clustering stability analysis (Fig. 1d) demonstrated that ICM cells in E3.5 embryos showed strong evidence for falling into two clusters, whereas those at E3.25 did not reproducibly segregate into clusters (Fig. 1e). These data therefore reveal that at E3.25 ICM cells are not readily distinguishable in terms of their gene expression profile. Consequently, the transcriptome data do not favour what would be predicted from a model of predetermined19, in which distinct ‘waves’ of cell divisions generate distinctly identifiable types of inner cell; however, the data also do not exclude the possibility that more subtle differences—for example in single messages, or in other molecules—between ICM cells could underlie their eventual cell fate specification (see Discussion).

Progressive establishment of correlation

To begin to unravel the general principles of lineage emergence and segregation within the early mouse embryo, we validated several lineage markers newly identified in the microarray analysis of 66 cells (Supplementary Table 1) using qPCR for a total of 137 single cells (Fig. 2a). Genes analysed included: Cldn4 and Enox1 for EPI, and Aldh1b1a1, Amn, Col4a1, Col4a2, Cabin, Foxq1, Lamb1, P4ha2, Serpinh1 and Tom111 for PrE. Among them, the PrE-specific expression of Amn, Cabin and Col4 is in agreement with immunofluorescence staining in ref. 29, and that of Lamb1 with ref. 30. Immunostaining of Serpinh1 and P4ha2 also confirmed their specific expression in PrE at E4.5 (Supplementary Fig. 2). Differentially expressed lineage-specific markers exhibited stochastic expression that seemed uncorrelated between genes, early in the lineage segregation process (Fig. 2a).

We identified several lineage markers that allow characterization of the stage of PřE differentiation, because these genes were progressively activated during lineage specification (Fig. 2b). These marker genes were defined in two steps (see Methods for details); after screening the microarray data for lineage-specific genes that were progressively upregulated from E3.25 to E3.5, and to E4.5, the identified candidate genes were verified by qPCR of additional single-cell complementary DNA samples. This allowed identification of 7 PrE differentiation stage markers (Fig. 2b) whose gene expression is progressively upregulated during the PrE lineage differentiation. It should be noted that the
Figure 1  Single-cell expression analysis of the lineage segregation within the ICM of the mouse blastocyst. (a) Schematic of the experimental method of single-cell isolation and gene expression profiling. cDNA was processed, stored and used for qPCR and microarray analyses. (b) Gene expression profiles of 137 cells isolated from the ICM at E3.25 (33 cells from 4 embryos), E3.5 (43 cells from 3 embryos) and E4.5 (61 cells from 3 embryos) analysed by qPCR. Each bar represents the expression of indicated genes in individual cells, with the same horizontal positions representing the same cells. The red line indicates the minimal level of gene expression detectable quantitatively (20 copies). (c) PCA plot of the microarray expression profiles characterizing the relative position of individual cells from blastocysts (66 cells including 36 cells from 6 embryos at E3.25, 22 cells from 3 embryos at E3.5, and 8 cells from one embryo at E4.5) in a map of lineage segregation. Note that the PCA was performed in an unsupervised manner, that is, without information on cell stage or lineage. (d) Schematic of the cluster stability analysis to identify subpopulations among cells. If distinguishable subgroups exist (marked in green and blue on the right), repeated bootstrap-sampled unsupervised clustering segregates them reproducibly (right panel). If repeated clustering produces incongruent results, no stably identifiable subgroups exist (left, grey). (e) Results of the cluster stability analysis (using a version of k-means clustering, partitioning around medoids, with \( k = 2 \) for E3.25 and E3.5 cells. Left: membership probabilities of each cell in the consensus clustering. Each dot represents the relative frequency at which a cell was assigned to one of the two consensus clusters in 250 random samplings. For E3.5, these frequencies had a bimodal distribution at 0 and 1, whereas for E3.25, they were diffuse. Right: box plot of cluster agreement score of 250 random samplings with the consensus. The central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to 1.5 times the interquartile range (25th to 75th percentile), and outliers are plotted individually. Consistently high agreement was seen for E3.25, whereas the score was close to random expectation for E3.25. The agreement score distributions between E3.25 and E3.5 were significantly different (\( P = 2 \times 10^{-16} \), Wilcoxon test).
Figure 2 Correlation and hierarchy of gene expression is progressively established during lineage segregation within the ICM of the mouse blastocyst. (a) Expression of lineage-specific markers analysed by single-cell qPCR (137 cells in total, including 33 cells from 4 embryos at E3.25, 43 cells from 3 embryos at E3.5, and 61 cells from 3 embryos at E4.5). Genes marked in red represent newly identified lineage markers. Each column represents the expression profile of an individual cell, with the colour code at the bottom right representing the estimated copy number for each gene. (b) Progressive upregulation of newly identified PrE differentiation marker genes. Box plots showing the expression level for each gene, collected for each stage from single-cell qPCR analysis (137 cells in total, including 33 cells from 4 embryos for E3.25, 21 and 22 cells from 3 embryos for E3.5 EPI and PrE, and 30 and 31 cells from 3 embryos for E4.5 EPI and PrE, respectively). The central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to 1.5 times the interquartile range (25th to 75th percentile), and outliers are plotted individually. (c) Hierarchical relationships of the activation of PrE differentiation marker genes. Each column represents one cell; dark blue indicates upregulation of genes during the transition from E3.25 to E3.5 (left) or from E3.5 to E4.5 (right). Upregulation during a transition was operationally defined as a gene expression value more than the midpoint of the average expression levels for E3.25 and E3.5 cells, or for E3.5 and E4.5 cells, respectively (Methods and Supplementary Figs 3 and 4d for detailed method). Hierarchy in gene activation was significantly stronger at the E3.5 to E4.5 transition than at the E3.25 to E3.5 transition (P = 2 × 10^{-16}, t-test).

Comparable EPI markers were more difficult to identify, because E3.25 ICM cells more closely resemble the E3.5 EPI than the PrE cohort, and upregulation of the expression of EPI markers is generally limited during differentiation (Fig. 1c).

Using these 7 PrE differentiation stage markers, we examined potentially hierarchical relationships of the activation in the lineage markers by investigating whether the genes could be ordered so that within each individual cell, expression of a gene is seen only if the
Heterogeneity in protein expression level of the EPI and PrE lineages. (a) A single-section immunofluorescence image of the E3.5 blastocyst simultaneously stained for Serpinh1 (a newly identified PrE marker), Gata6 and Nanog. In the merged image, Serpinh1, Gata6 and Nanog are labelled in blue, red and green, respectively. Scale bar, 10 μm. (b) Quantitative plots showing the normalized mean fluorescent intensity of Gata6 relative to Nanog, Serpinh1 relative to Nanog, and Serpinh1 relative to Gata6. Each dot represents one blastomere with different colours representing different embryos (56 cells from 4 embryos at E3.5). The expression intensity value of the respective gene is normalized against the level of DAPI signal. The average background fluorescence level is 0.032, 0.001 and 0.027 for Gata6, Nanog and Serpinh1, respectively. Correlation of protein expression levels is evident between Nanog and Gata6, Nanog and Serpinh1, and Gata6 and Serpinh1 ($r = -0.62$ and $P = 3 \times 10^{-7}$, $r = -0.46$ and $P = 3 \times 10^{-4}$, respectively; Pearson’s correlation coefficient).

**Figure 3**

Cell position influences gene expression

Positional information has been proposed to play a prominent role in the patterning of early embryos\(^2,3,32\). However, there are limited data\(^3,29\) to suggest that a cell’s position within the ICM influences its overall gene expression. To address this question and determine whether gene expression differences within the ICM reflect the position of individual cells, we established a method to identify, selectively isolate and expression profile cells located on the surface of the ICM adjacent to the blastocyst cavity versus those located deeper within the ICM (Fig. 4a). Expression profiling and comparison of these two populations revealed that cells facing the blastocyst cavity more closely resembled the PrE lineage from E3.5 onwards (Fig. 4b,c). These data therefore suggest that positional information may play an instructive role influencing the differential gene expression observed within the ICM at E3.5.

**Fgf4 is required for EPI versus PrE segregation**

Next, we wished to identify the symmetry-breaking signals driving lineage segregation within ICM cells. To do this we sought to characterize the genes that segregate into two distinct ICM

preceding gene is activated (Fig. 2c; Methods and Supplementary Figs 3 and 4 for detailed Methods). Remarkably, an approximate hierarchy in gene activation was observed at the E3.5 to E4.5 transition, whereas evidence for hierarchy was much weaker at E3.25 to E3.5 ($P = 2 \times 10^{-16}$, $t$-test), suggesting that the activation of lineage-specific marker gene expression establishes a hierarchical relationship only at the late blastocyst stage.

We also wished to evaluate variability in the expression of the lineage markers at the protein level. To do so, we performed a quantitative analysis of protein expression of a newly identified PrE marker, Serpinh1 (also known as heat shock protein 47, Hsp47; ref. 31), in relation to the lineage markers, Gata6 and Nanog (Fig. 3a). Serpinh1 is localized exclusively in the cytoplasm of PrE cells in E4.5 blastocysts (Supplementary Figs 2a), in agreement with its reported function as a chaperone for collagen synthesis. To evaluate any potential variability in protein expression during EPI versus PrE segregation, E3.5 blastocysts (having a total of 70–90 cells) were immunostained simultaneously for Serpinh1, Gata6 and Nanog, as well as DNA and cell membrane for $z$ axis normalization and cell/nucleus segmentation. This allowed us to perform quantitative measurements of the levels of protein expression for 56 individual ICM cells derived from 4 embryos (Fig. 3b, and Supplementary Video 1; see Methods for details). Although positive or negative correlation of protein expression levels is evident between Nanog and Gata6, Nanog and Serpinh1, and Gata6 and Serpinh1, high variability in their expression levels at E3.5 does not allow separation of the two cell populations, in contrast to E4.5 ICM cells (see Supplementary Fig. 2). This is consistent with our findings made at the RNA level (Fig. 2a), and favours a model in which EPI and PrE lineages stochastically emerge within a cohort of initially equivalent ICM cells, rather than being predetermined by two distinct division histories\(^15\).
Figure 4 Cell position influences gene expression. (a) Schematic of the method to label the cells on the surface of the ICM facing the blastocyst cavity. Immunosurgery was combined with manual bisection and isolation of the embryonic half of the blastocyst, followed by fluorescent labelling of the exposed surface cell layer (see Methods for details). (b) Multi-dimensional scaling plot of the labelled and non-labelled E3.5 and E4.5 inner cells, based on the expression of 10 highly variable genes, as identified from the E3.5 and 4.5 microarray data (Cotl1, Cth, Cubn, Fgf4, Lama1, Morc1, Pdgfra, Sepinh1, Sox17, Srgn), and quantified by additional single-cell qPCR measurements (43 cells in total including 23 cells from 6 embryos at E3.5, and 20 cells from 2 embryos at E4.5). (c) Number of label-positive and -negative cells in PrE and EPI groups, in which the lineage identity is assigned by marker gene expressions. Clear segregation of the PrE and EPI cells at E4.5 indicates that this labelling method can clearly distinguish the PrE cells from the EPI cells in the E4.5 blastocyst. In E3.5, label-positive cells are strongly enriched in the PrE group (odds ratio 12, P = 0.01, Fisher’s exact test).

populations at the earliest stage, corresponding to E3.5 in our analysis. Fgf4 was identified as one of such genes exhibiting the greatest differential expression between EPI and PrE cells (Figs 1b and 2a and Supplementary Table 1). To comprehensively characterize the involvement of Fgf signalling in the EPI versus PrE lineage segregation, the expression levels of all Fgf ligands, receptors and downstream cytoplasmic signalling components in the developing blastocyst were analysed using the 66 single-cell ICM transcriptome data (Fig. 5a and Supplementary Fig. 5). Several Fgf ligands (Fgf3, 4 and 13) and all Fgf receptors (Fgfr1–4) were found to be differentially expressed within the ICM, thus possibly contributing to the EPI versus PrE lineage segregation. In contrast, cytoplasmic signalling components exhibited no differential expression, suggesting that any differential regulation would predominantly be at the post-transcriptional level. The overlapping expression of ligands and receptors suggests the presence of redundant functions within Fgf signalling pathway components. A statistically significant correlation (positive or negative) in gene expression levels is discernible at the single-cell level for Fgf4 against Fgfr2 (in agreement with ref. 16), Fgf4 against Fgfr3, Fgf3 with Fgfr3, and Fgfr2 at E3.5 and E4.5 ICMs (Fig. 5b). Among those genes expressed in the blastocyst, Fgf4 and Fgfr2 exhibit differential expression the earliest (E3.25), followed by Fgfr1. The higher variability, and bimodality (Fig. 5b), in the expression of Fgf4 than of Fgfr2 at E3.25 suggests that Fgf4 may be the driver for the observed differential gene expression and EPI versus PrE lineage segregation.

We recently demonstrated that Fgf4 is required for the establishment of a salt-and-pepper distribution of EPI/PrE lineage precursors at E3.5, as well as the specification of PrE within the ICM (ref. 13). To comprehensively characterize the impact of loss of Fgf4 on EPI versus PrE lineage segregation we performed single-cell gene expression analyses on the ICMs of Fgf4−/− mutant embryos. The expression profiles of individual ICM cells derived from Fgf4−/− blastocysts at E3.25–E4.5 were overlaid on the lineage map established using the wild-type single-cell expression profiles (shown in Fig. 1c). The samples’ coordinates allowed us to characterize the differentiation status of Fgf4−/− ICM cells. Surprisingly we noted that the differentiation...
Comprehensive characterization of expression of Fgf signalling components in the early mouse embryo. (a) Box plots showing the mRNA expression levels of Fgf ligands and receptors detectable in the early mouse embryo, collected for each stage from single-cell microarray analysis (66 wild-type (WT) cells including 36 cells from 6 embryos for E3.25, 11 and 11 cells from 3 embryos for E3.5 EPI and PrE, and 4 and 4 cells from one embryo for E4.5 EPI and PrE cells, respectively; and 35 Fgf4−/− (Fgf4-KO) cells including 17 cells from 3 embryos for E3.25, 8 cells from one embryo for E3.5 and 10 cells from one embryo for E4.5). The central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to 1.5 times the interquartile range (25th to 75th percentile), and outliers are plotted individually. Those Fgf ligands whose expression level is negligible are shown in Supplementary Fig. 5. (b) Scatter plots with each dot representing the mRNA expression levels of specific Fgf ligand and receptor pairs in one blastomere. The colour code of the dot is the same throughout this study, shown in the inset of Fig. 1c, with pink representing E3.25 cells, light blue and green E3.5 EPI and PrE cells, and blue and green E4.5 EPI and PrE cells, respectively. Those with statistically significant correlation (positive or negative) are shown (r = −0.77, P = 7 × 10−7 (Fgf4 versus Fgfr2); r = −0.42, P = 2 × 10−2 (Fgf4 versus Fgfr3); r = 0.82, P = 4 × 10−8 (Fgf3 versus Fgfr3); r = 0.76, P = 1 × 10−6 (Fgf3 versus Fgf4); Pearson’s correlation coefficient).

not only of PrE but also of EPI cell lineage was arrested in Fgf4−/− mutants (Fig. 6a), indicating that Fgf4 is required for segregating these two lineages. Moreover, in Fgf4−/− mutants, expression of PrE lineage-specific markers was significantly suppressed and maintained at the level of wild-type E3.25, whereas loss of Fgf4 had a more variable effect on EPI markers (Fig. 6b). It should be noted that although E4.5 Fgf4−/− ICM cells are positioned relatively close to EPI cells in the two-dimensional principal component analysis (PCA) projection (Fig. 6a), their expression profiles are significantly distinct from E3.5 and E4.5 wild-type EPI cells (Fig. 6c), indicating that Fgf4−/− ICM cells are not simply differentiating into the EPI lineage. Similarly, although E3.5 Fgf4−/− cells seem to overlap with E3.25 wild-type cells, a more detailed analysis of their expression profiles indicates that they represent a distinct population (Fig. 6d). Moreover, E3.25 Fgf4−/− ICM cells seem to be distributed differently from wild-type cells, suggesting that there might be a distinct role for Fgf signalling at an early stage. Additional qPCR analysis of Fgfr2, Nanog and Gata6 expression in E3.25 and E3.5 ICM cells (Supplementary Fig. 6) revealed that whereas in wild-type cells their gene expression levels show positive or negative correlation at the single-cell level, Fgf4−/− cells tend to lose such correlations. These data suggest the requirement of Fgf signalling in establishing the gene regulatory network for EPI versus PrE lineage segregation. Loss of Fgf4 alone does not induce compensatory expression of other Fgf ligands, and the expression patterns of other Fgf signalling components are generally unaltered at the E3.25/3.5 stage (Fig. 5a and Supplementary Fig. 5). Genes downregulated in E3.5 Fgf4−/− cells (Supplementary Table 2) would include putative targets of Fgf signalling in the early mouse embryo. Collectively, our data suggest that heterogeneity in the expression, and thus availability, of Fgf4 is critical for lineage segregation and couples it to the salt-and-pepper distribution of EPI/PE cells within the E3.5 ICM (ref. 13).

**DISCUSSION**

In this study we have developed a framework for the isolation of single cells from the ICMs of developing mouse blastocysts, expression profiling and data analysis. These data represent the first comprehensive

**Figure 5** Comprehensive characterization of expression of Fgf signalling components in the early mouse embryo. (a) Box plots showing the mRNA expression levels of Fgf ligands and receptors detectable in the early mouse embryo, collected for each stage from single-cell microarray analysis (66 wild-type (WT) cells including 36 cells from 6 embryos for E3.25, 11 and 11 cells from 3 embryos for E3.5 EPI and PrE, and 4 and 4 cells from one embryo for E4.5 EPI and PrE cells, respectively; and 35 Fgf4−/− (Fgf4-KO) cells including 17 cells from 3 embryos for E3.25, 8 cells from one embryo for E3.5 and 10 cells from one embryo for E4.5). The central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to 1.5 times the interquartile range (25th to 75th percentile), and outliers are plotted individually. Those Fgf ligands whose expression level is negligible are shown in Supplementary Fig. 5. (b) Scatter plots with each dot representing the mRNA expression levels of specific Fgf ligand and receptor pairs in one blastomere. The colour code of the dot is the same throughout this study, shown in the inset of Fig. 1c, with pink representing E3.25 cells, light blue and green E3.5 EPI and PrE cells, and blue and green E4.5 EPI and PrE cells, respectively. Those with statistically significant correlation (positive or negative) are shown (r = −0.77, P = 7 × 10−7 (Fgf4 versus Fgfr2); r = −0.42, P = 2 × 10−2 (Fgf4 versus Fgfr3); r = 0.82, P = 4 × 10−8 (Fgf3 versus Fgfr3); r = 0.76, P = 1 × 10−6 (Fgf3 versus Fgf4); Pearson’s correlation coefficient).
Figure 6 Fgf4 is required for driving lineage segregation between EPI and PrE in the early mouse embryo. (a) PCA plot of the microarray expression profiles of Fgf4−/− (Fgf4-KO) cells (35 Fgf4−/− cells including 17 cells from 3 embryos for E3.25, 8 cells from one embryo for E3.5 and 10 cells from one embryo for E4.5) overlaid on the EPI versus PrE lineage map established using the WT cell profile (66 wild-type cells including 36 cells from 6 embryos for E3.25, 11 and 11 cells from 3 embryos for E3.5 EPI and PrE, and 4 and 4 cells from one embryo for E4.5 EPI and PrE cells, respectively). Note that the position of wild-type cells is identical to that shown in Fig. 1c and is used here as a reference map. (b) Impact of the loss of Fgf4 on the expression of lineage markers analysed by microarray. Box plots show the expression of PrE and EPI markers (including differentiation markers), collected for each stage from single-cell microarray analysis (similarly to Fig. 5a). The central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to 1.5 times the interquartile range (25th to 75th percentile), and outliers are plotted individually. (c) Cluster stability analysis (250 random samplings) for Fgf4−/− E4.5 cells together with wild-type E3.5 EPI and PrE cells (upper row), or with E4.5 EPI and PrE cells (lower row). Shown are the membership probabilities of the consensus clustering, analogous to the analysis in Fig. 1e. Unsupervised clustering faithfully recovers the grouping into wild-type E3.5 EPI cells, wild-type E3.5 PrE cells, wild-type E4.5 EPI and PrE cells, and Fgf4−/− E4.5 cells. (d) Cluster stability analysis (250 random samplings) for Fgf4−/− E3.5 cells together with wild-type E3.25 cells. Shown are the membership probabilities of the consensus clustering. The analysis demonstrates that Fgf4−/− E3.5 cells form a single, tight cluster.
Figure 7 Schematic model for EPI versus PrE lineage segregation in the early mouse embryo, contrasting with mechanisms for embryo patterning in non-mammalian species. (a) In many non-mammalian species, localized determinants play a key role in embryonic patterning. (b) In the ICM of the mouse blastocyst, EPI and PrE lineages are progressively segregated within a cohort of initially equivalent cells. Cell-to-cell variability generated by stochastic onset of gene expression (genes A, B, C represent the lineage marker for blue cells, and D, E, F for green cells) is progressively enhanced by signalling activities and feedbacks as well as cell–cell interactions, and forms a salt-and-pepper pattern, with two emerging populations. This process eventually leads to establishing two distinct cell lineages (blue or green cells) with specific gene regulatory networks (GRNs) in the context of positional information. In the absence of Fgf4, reinforcement by the signalling cascade may fail and lineage segregation is halted without differentiation into either of the two lineages.

and unbiased single-cell resolution lineage map of the ICM of mammalian blastocyst. The finding that inner cells at E3.25 show no apparent distinction favours a model of stochastic and progressive segregation of EPI and PrE lineages7. However, these data do not exclude the possibility that some difference may exist among cells within the ICM at E3.25, as was postulated previously26 on the basis of the inverse correlation between Fgf4 and Fgfr2, which we also noted in our samples (see Fig. 5b). The statistical cell subpopulation analysis used in this study provides evidence against a consistent, widespread gene expression pattern reflecting predetermination or lineage commitment at E3.25, although our analysis would not detect a difference that is restricted to a small number of genes or non-mRNA molecules. E3.25 ICM cells, however, do not exhibit a ‘uniform’ gene expression status, perhaps reminiscent of the ground state of embryonic stem cells33, but instead are a mixture of cells with stochastic gene expression variability. Stochastic fluctuations of gene expression may offer a greater repertoire of combinatorial gene expression23, which may underlie the developmental plasticity and highly regulative capacity of the preimplantation mouse embryo before E4.5 (ref. 34).

Our single-cell data allowed us to comprehensively identify EPI and PrE lineage markers. Newly identified genes that are specifically expressed early in the PrE differentiation include extracellular matrix components and factors involved in their synthesis. Presumptive PrE cells may need to produce a large amount of structural proteins that need to be incorporated into the basement membrane at the interface between the newly forming PrE epithelial layer and adjacent inner EPI cells.

We also determined the impact of loss of a key signalling molecule, Fgf4, through the analysis of ICM cells in a mutant13. Embryos lacking Fgf4, or the effector Grb2 (ref. 2), exhibit a profound defect within the ICM characterized by an absence of PrE cells, a phenotype that can be recapitulated using Fgf signalling inhibitors8. If Fgf4 and Grb2 are critical non-redundant points in the pathway, several questions remain
concerning the identity of the receptors and downstream intracellular effectors required for transducing the Fgf4/Grb2 signal, the cells in which the pathway is normally active, and the mechanism by which Fgf4-positive and -negative cells are generated in E3.25 ICMs.

It is now well established that molecular heterogeneities presage marker restriction and lineage segregation\(^3\). Live imaging of a fluorescent reporter for the PrE-lineage reporter embryos demonstrated that cell sorting and position-dependent regulation of gene expression may help resolve the molecular heterogeneities into the pattern\(^7\). A combination of live imaging embryos expressing lineage-specific fluorescent reporters and single-cell gene expression profiling should eventually allow dissection of the underlying mechanisms. A recent study investigating neural tube patterning in zebrafish revealed that cell sorting rearranges an initial mixture of different neural progenitors formed by heterogeneous signalling activity into sharply bordered domains\(^8\). Thus, the generation and resolution of molecular heterogeneities might represent a conserved mechanism for driving pattern formation in various contexts during embryonic development across species\(^9\).

Our single-cell data showed that ICM cells maintain the same level of gene expression variability despite the lack of Fgf4 (standard deviation of log2 expression measurements of the 100 most variable genes in E3.25: 1.7 ± 1.2 (wild type) versus 1.5 ± 1.2 (Fgf4\(^{-/-}\))), suggesting that Fgf4 is not required for generating the initial molecular heterogeneity. Consequently, it would be conceivable to separate early blastocysts with cell-to-cell gene expression variability into two phases. In the first phase, expression of individual genes exhibits stochastic variability, possibly independent from one another. In contrast, in the second phase, a correlation of gene expression levels gradually emerges, probably owing to the activation of lineage-specific signalling cascades (for example, Fgf; Fig. 7). The second phase may correspond to the blastocyst stage in which a salt-and-pepper pattern of expression\(^2\) can be defined by the onset of Gata4 expression, and restriction of Gata4 to Gata4-positive PrE lineage ‘precursors’, or cells with a propensity to contribute to the emergent PrE. However, as demonstrated in this study (Figs 2 and 3), when evaluated with a number of genes/proteins simultaneously, ICM cells at this stage still exhibit a high degree of expression variability, and future studies would require a comprehensive and quantitative description of molecular heterogeneities. Taken together, we propose that an initial phase of stochastic gene expression followed by signal reinforcement may drive lineage segregation by antagonistically separating a cohort of initially equivalent cells (Fig. 7). Thus, the inherent molecular heterogeneity, and subsequent salt-and-pepper pattern of lineage precursors, within the ICM may form the foundation for segregating distinct EPI and PrE lineages within an initially equivalent population of cells.

Methods and any associated references are available in the online version of the paper.

Note: Supplementary Information is available in the online version of the paper

Acknowledgements

We thank A. Courtin for help with image analysis, R. Niwayama for quantitative protein expression analysis, S. Salvenmoser and R. Bloech for technical assistance, and EMBL Genomics Core Facility for technical support. We also thank the members of the Hiiragi, Hadjantonakis, Huber and Saitou laboratories for helpful and stimulating discussions. Work in the laboratory of T.H. is supported by the Max Planck Society, European Research Council under the European Commission FP7, Stem Cell Network North Rhine Westphalia, German Research Foundation (Deutsche Forschungsgemeinschaft), and World Premier International Research Initiative, Ministry of Education, Culture, Sports, and Technology, Japan. Work in the laboratory of A.-K.H. is supported by the National Institutes of Health (NIH) NIH RO1-HD052115 and RO1-DK084391 (AKH) and NYSTEM. W.H. acknowledges financial support from the European Commission FP7-Health through the RADIANT project. Y.O. is supported by Naito and Uehara Memorial Foundation fellowships, and by Marie Curie FP7 IIF fellowship (no. 273193).

Author Contributions

Y.O., A.T. and T.H. designed the study, Y.O. performed most of the experiments, A.T., K.K. and M.S. contributed to establishing the method of single-cell gene expression analysis in the mouse preimplantation embryo. Y.O., A.T., M.K. and P.X. collected the single-cell samples. W.H. and A.K.O. performed statistical analysis, and M.J.A.-B. contributed to initial analyses of the data. Y.O., A.-K.H. and T.H. wrote the manuscript.

Competing Financial Interests

The authors declare no competing financial interests.

Published online at www.nature.com/doifinder/10.1038/ncb2881

Reprints and permissions information is available online at www.nature.com/reprints

1. Rossant, J. & Tam, P. P. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development 136, 701–713 (2009).
2. Chazaud, C., Yamanaka, Y., Pawson, T. & Rossant, J. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Dev. Cell 10, 615–624 (2006).
3. Plusa, B., Piliszek, A., Frankenbeg, S., Artus, J. & Hadjantonakis, A. K. Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. Development 135, 3081–3091 (2008).
4. Schrode, K. et al. Anatomy of a blastocyst: cell behaviors driving cell fate choice and morphogenesis in the early mouse embryo. Genesis 51, 219–233 (2013).
5. Grabarek, J. B. et al. Differential plasticity of epiblast and primitive endoderm precursors within the ICM of the early mouse embryo. Development 139, 129–139 (2012).
6. Lanner, F. & Rossant, J. The role of FGF/Erk signaling in pluripotent cells. Development 137, 3351–3360 (2010).
7. Yamanaka, Y., Lanner, F. & Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development 137, 2673–2674 (2010).
8. Arman, H., Haffner-Krausz, R., Chen, Y., Heath, J. K. & Lonai, P. Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in gastrulation mammalian development. Proc. Natl Acad. Sci. USA 95, 5082–5087 (1998).
9. Nichols, J., Silva, J., Rode, M. & Smith, A. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development 136, 3215–3222 (2009).
10. Cheng, A. M. et al. Mammalian Grb2 regulates multiple steps in embryonic development and malignant transformation. Cell 95, 793–803 (1998).
11. Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M. & Goldfarb, M. Requirement of FGF-4 for postimplantation mouse development. Science 267, 246–249 (1995).
12. Wilder, P. J. et al. Inactivation of the Fgf4 gene in embryonic stem cells alters the growth and/or the survival of their early differentiated progeny. Dev. Biol. 192, 614–629 (1997).
13. Kang, M., Piliszek, A., Artus, J. & Hadjantonakis, A. K. FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. Development 140, 267–279 (2013).
14. Chisholm, J. C. & Houliston, E. Cytokeratin filament assembly in the preimplantation mouse embryo. Development 101, 565–582 (1987).
15. Morris, S. A. et al. Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. Proc. Natl Acad. Sci. USA 107, 6364–6369 (2010).
16. Guo, G. et al. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Dev. Cell 18, 675–685 (2010).
17. Kurimoto, K. et al. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. Nucleic Acids Res. 34, e42 (2006).
18. Kurimoto, K., Yabuta, Y., Ohinata, Y. & Saitou, M. Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. Nat. Protocols 2, 739–752 (2007).
19. Tang, F. et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods 6, 377–382 (2009).
20. Tang, F. et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat. Protocols 5, 516–535 (2010).
21. Pelkmans, L. Cell Biology. Using cell-to-cell variability—a new era in molecular biology. Science 336, 425–426 (2012).
22. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. Nature 467, 167–173 (2010).
23. Hu, M. et al. Multilineage gene expression precedes commitment in the hemopoietic system. Genes Dev. 11, 774–785 (1997).
24. Pina, C. et al. Inferring rules of lineage commitment in haematopoiesis. Nat. Cell Biol. 14, 287–294 (2012).
25. Moignard, V. et al. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. Nat. Cell Biol. 15, 544 (2013).
26. Buganim, Y. et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell 150, 1209–1222 (2012).
27. Frankenberg, S. et al. Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. Dev. Cell 21, 1005–1013 (2011).
28. Solter, D. & Knowles, B.B. Immonosurgery of mouse blastocyst. Proc. Natl Acad. Sci. USA 72, 5099–5102 (1975).
29. Gerbe, F., Cox, B., Rossant, J. & Chazaud, C. Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst. Dev. Biol. 313, 594–602 (2008).
30. Artus, J., Piliszcz, A. & Hadjantonakis, A. K. The primitive endoderm lineage of the mouse blastocyst: sequential transcription factor activation and regulation of differentiation by Sox17. Dev. Biol. 350, 393–404 (2011).
31. Widmer, G. et al. Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINC1 and its structure-specific client recognition. Proc. Natl Acad. Sci. USA 109, 13243–13247 (2012).
32. Tarkowski, A. K. & Wroblewska, J. Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage. J. Embryol. Exp. Morphol. 18, 155–180 (1967).
33. Silva, J. & Smith, A. Capturing pluripotency. Cell 132, 532–536 (2008).
34. Wennekamp, S., Mesecke, S., Nedelec, F. & Hiiragi, T. A self-organization framework for symmetry breaking in the mammalian embryo. Nat. Rev. Mol. Cell Biol. 14, 454–461 (2013).
35. Dietrich, J. E. & Hiiragi, T. Stochastic patterning in the mouse pre-implantation embryo. Development 134, 4219–4231 (2007).
36. Xiong, F. et al. Specified neural progenitors sort to form sharp domains after noisy shh signaling. Cell 153, 550–561 (2013).
37. Kay, R. R. & Thompson, C. R. Forming patterns in development without morphogen gradients: scattered differentiation and sorting out. Cold Spring Harb. Perspect. Biol. 1, a001503 (2009).
Single-cell cDNA amplification. Single-cell cDNA amplification was performed as previously described [13, 14]. Briefly, single blastomeres were lysed in individual tubes without purification, and first-strand cDNAs were synthesized using a modified protocol [13, 14]. First, the lysed trophectoderm (TE) and IC were digested by repeated pipetting. ICM was further dissociated into single blastomeres by pipetting in PBS (25 mM HEPES, 137 mM NaCl, 5 mM KCl, 0.7 mM Na2HPO4, 6 mM dextrose, 0.9 mM CaCl2, and 0.5 mM MgCl2, at pH 7.05) supplemented with 1 mM EDTA (H-EDTA) after ~5–7 min incubation at 37 °C in H-EDTA supplemented with 1% trypsin (Sigma-Aldrich, T-4549). It usually took ~80–90 min to isolate and lyse ICM cells and recover their RNAs after euthanization of the mouse, and great care was taken to minimize the time. The developmental stage of embryos subjected to the single-cell gene expression analysis was defined as follows. On recovery, an average-size embryo was selected for subsequent experiments, and the remaining littersmates were fixed in PBS supplemented with 4% paraformaldehyde (Electron Microscopy Sciences, 19208) and stained in PBS with DAPI (Molecular Probes, D3571, 1:1,000) as well as with either Alexa Fluor 633 or 564 phalloidin (Molecular Probes, A22284 or A22283, respectively). Between 10–17 blastomeres were identified under confocal microscopy, and gene expression was analysed using single-cell cDNA amplification and qPCR as described above. Samples (43 cells from one embryo, 49 cells from 3 embryos at E3.5 and 61 cells from 3 embryos at E4.5) in 5 independent experiments, were not randomized, and the investigators were not blinded to allocation during the study. The background intensity was defined as the average of mean fluorescence intensities of 15 randomly chosen spots located outside the embryo, divided by the average of all mean DAPI intensities. A Quicktime video (Supplementary Video 1) showing the entire z-scans of immunofluorescence staining was generated using Photoshop (CSS, Adobe). Labeling of the cells located on the surface of the inner cell mass facing to the blastocyst cavity. After removal of the zona pellucida with a brief treatment of pronase, blastocysts were incubated for 25 min at 37 °C in KSO supplemented with anti-mouse antibody (Cedarlane, CL2301, 1:8), and manually bisected using a 27-G needle in KSO containing HEPES (ref. 38). The surface of the resultant embryos containing polar trophectoderm and ICM was stained by 2 x 1 s incubation in KSO supplemented with Cell Mask (Invitrogen, C10045, 1:100), followed by 15–20 min incubation at 37 °C in KSO supplemented with guinea pig complement (Cedarlane, CL5000F, 1:2). After removal of the lysed trophectoderm by pipetting, the ICM was dissociated into single blastomeres by 5 min incubation in H-EDTA followed by 7 min incubation at 37 °C in H-EDTA supplemented with 0.05% trypsin, and further pipetting in PBS. Fluorescently labelled outer or non-labelled inner cells were identified by cupolocal confocal microscopy, and gene expression was analysed using single-cell cDNA amplification and qPCR as described above. Samples (43 cells) were collected from a total of 8 embryos (23 cells from 6 embryos at E3.5 and 20 cells from 2 embryos at E4.5) in 5 independent experiments, each time collecting the sample cells from one embryo. Occasionally, outer cells seemed to be not entirely removed, owing possibly to the modified immunosurgery protocol, and those single-cell samples in which qPCR detected the expression of trophectoderm markers (for example Cdx2 and Id2) were eliminated from further analysis.

Statistical analyses. All statistical procedures were developed by a statistician (W.H.), carefully checked for robustness both to choice of method and natural variability in the data, and the analyses were performed using R/Bioconductor software. An R package named Hiiragi2015 including the complete data and software scripts is available as an executable document ('vignette') at www.bioco conductor.org. No statistical method was used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment. Statistical tests were chosen to meet the properties of the data. t-tests were performed with the Welch approximation to the degrees of freedom to allow for unequal variances. Extensive data exploration and visualization provided no indication of heteroscedasticity-induced problems. Each sample was represented by the average cell number of the littermates, but excluding samples with the maximum and minimum cell numbers, was determined and used to represent the developmental stage of the experimental sample for single-cell analysis.
resampled data sets (each containing a random subset of 67% of cells, sampled without replacement), constructing a consensus clustering, and comparing the B individual clustering results with the consensus. Specifically, for each sample, its cluster assignment probabilities were computed, and for each of the B clusterings, their agreement with the consensus was measured by the Euclidean dissimilarity D of the membership matrices, that is, the square root of the minimal sum of the squared differences of U and all column permutations of V, where U and V are the cluster membership matrices. The cluster agreement scores shown in Figs 1e and 6c are 1 −D/ M, where M is an upper bound for the maximal Euclidean dissimilarity. Computations were performed using the R package clue41.

For the analysis of hierarchical relationships among gene activations, the differentiation stage markers were first identified as follows: expressed in only one of the lineages at E4.5; and expressed an average fold-change of at least 8 from E3.25 to E4.5, as well as average fold-changes of at least 1.4 in the individual transitions from E3.25 to E3.5, and from E3.5 to E4.5. We then used qPCR of additional single-cell cDNA samples for validation, and identified 7 PrE differentiation stage markers (Fig. 2b) whose gene expression is progressively upregulated during the PrE lineage differentiation, without change in the EPI lineage. For each of the 7 genes, the average levels in the conditions E3.25, E3.5 (PrE) and E4.5 (PrE) were computed, and two thresholds were defined corresponding to the midpoint between the averages of E3.25 and E3.5, and the midpoint between the average of E3.5 and E4.5 (see Supplementary Fig. 3). Data were binned into two states, on and off, as follows. For either the E3.25 to E3.5 transition or the E3.5 to E4.5 transition, a gene was considered on in a cell if its expression value exceeded the threshold associated with the transition. For a particular ordering of the seven identified genes, a hierarchy mismatch score was defined by counting the number of instances when an on gene preceded an off gene in the ordering. The minimum score was determined over all 7! = 5,040 possible orderings, and normalized to the range from 0 to 1 by dividing it by the number of gene pair comparisons. All possible orders with the minimum score are depicted in Supplementary Fig. 4. To assess the statistical significance of the observed difference between the hierarchy mismatch score of the E3.25 to E3.5 transition and that of the E3.5 to E4.5 transition, the procedure was bootstrap-resampled.

Accession number. The microarray data have been deposited to the ArrayExpress database with the accession number E-MTAB-1681.

38. Ohnishi, Y. et al. Small RNA class transition from siRNA/piRNA to miRNA during pre-implantation mouse development. Nucleic Acids Res. 38, 5141–5151 (2010).
39. Irizarry, R. A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4, 249–264 (2003).
40. Kauffmann, A., Gentleman, R. & Huber, W. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. Bioinformatics 25, 415–416 (2009).
41. Hornik, K. A CLUE for CLUster Ensembles. J. Statist. Software 14 (2005).
Supplementary Figure 1 Performance of spike RNA amplification. Each blue line represents the outcome of spike RNA amplification for each experimental sample that is used for microarray (66 cells in total including 36 cells from 6 embryos for E3.25, 22 cells from 3 embryos for E3.5, and 8 cells from one embryo for E4.5). Boxplot shows the performance of spike RNA amplification for all samples including those used only for additional qPCR (grey, 154 cells in total including 50 cells from 6 embryos for E3.25, 43 cells from 3 embryos for E3.5 and 61 cells from 3 embryos for E4.5). Those single-cell cDNAs of highest quality with minimal deviation from the ideal value (red line) are processed for microarray analysis. Based on this performance, we defined 20 copies as the minimum amount of mRNAs that we can detect quantitatively.
Supplementary Figure 2 Immunofluorescence single-section images of the E4.5 (>150 cell stage) blastocyst stained for Serpinh1 (a) and P4ha2 (b), PrE markers newly identified in the microarray analysis, indicating the lineage-specific expression in PrE. Scale bars; 10 μm.
**Supplementary Figure 3** qPCR data for the expression of seven PrE differentiation stage markers used in Fig. 2b,c. Each dot represents the gene expression pattern of single cells derived from E3.25 ICM (purple), E3.5 PrE (light green), and E4.5 PrE (dark green) cells with Y-axis indicating the estimated copy number (86 cells in total including 33 cells from 4 embryos for E3.25, 22 cells from 3 embryos for E3.5 PrE, and 31 cells from 3 embryos for E4.5 PrE). The within-group means and the binning thresholds are shown as horizontal dotted lines (light grey) and horizontal solid lines (dark grey), respectively.
Supplementary Figure 4  All possible and equally optimal orders of the genes (Y-axis) used in Fig.2c to examine the potential hierarchy in gene activation during the E3.25 to E3.5 transition (see Methods). A total of seven equally optimal solutions are available for aligning the genes upregulated during the E3.25 to E3.5 transition, including one shown in Fig. 2c. Note that there was only one solution for the E3.5 to E4.5 transition, as shown in Fig. 2c.
**Supplementary Figure 5** Comprehensive characterisation of expression of Fgf signalling components within the early mouse embryo. Box plots showing the mRNA expression level of Fgf ligands and downstream cytoplasmic signal effectors, collected for each stage from single-cell microarray analysis (66 WT cells including 36 cells from 6 embryos for E3.25, 11 and 11 cells from 3 embryos for E3.5 EPI and PrE, and 4 and 4 cells from one embryo for E4.5 EPI and PrE cells, respectively; and 35 Fgf4⁻/⁻ cells including 17 cells from 3 embryos for E3.25, 8 cells from one embryo for E3.5 and 10 cells from one embryo for E4.5).
Supplementary Figure 6 Scatter plots showing the early lineage marker expressions in individual WT and Fgf4−/− ICM cells. Each dot represents the expression of lineage markers in single blastomere, analysed by qPCR (33 cells from 4 embryos for E3.25 WT and 9 cells from one embryo for E3.25 Fgf4−/−, and 43 cells (21 and 22 cells for EPI and PrE, respectively) from 3 embryos for E3.5 WT and 8 cells from one embryo for E3.5 Fgf4−/−). The gene expression levels are normalised to that of Gapdh (x or y = 0). The colour code is the same as shown in Fig. 6a. In WT cells, each combination of two marker genes exhibits statistically significant correlation (E3.25: r = 0.35, p = 4 x 10⁻² (Gata6 vs. Fgfr2); r = -0.46, p = 7 x 10⁻³ (Nanog vs. Fgfr2) and E3.5: r = -0.42, p = 5 x 10⁻³ (Nanog vs. Gata6); r = 0.54, p = 2 x 10⁻⁴ (Gata6 vs. Fgfr2); r = -0.66, p = 2 x 10⁻⁶ (Nanog vs. Fgfr2); Pearson’s correlation coefficient), except for Nanog vs. Gata6 at E3.25 (r = 0.07, p = 0.7). However, the correlation is lost in Fgf4−/− cells (E3.25: r = 0.34, p = 0.4 (Gata6 vs. Nanog); r = 0.01, p = 1 (Gata6 vs. Fgfr2); r = 0.30, p = 0.4 (Nanog vs. Fgfr2) and E3.5: r = 0.25, p = 0.5 (Nanog vs. Gata6); r = 0.05, p = 0.9 (Gata6 vs. Fgfr2); r = -0.04, p = 0.9 (Nanog vs. Fgfr2); Pearson’s correlation coefficient).
Supplementary Table Legends

**Supplementary Table 1** Lists of the genes expressed at the level with highest difference between EPI and PrE clusters within the ICM of the E3.5 and E4.5 blastocysts. Lineage specificity is shown as minus for EPI and plus for PrE, with the higher absolute value indicating the higher degree of differential expression.

**Supplementary Table 2** Lists of the genes upregulated or downregulated in Fgf4-/- cells at E3.5.

**Supplementary Table 3** List of the gene-specific primers used for qPCR.

Supplementary Video Legend

**Supplementary Video 1** Immunofluorescence staining of the E3.5 blastocyst. Z-scanning sections of one of the four embryos used for the quantitative protein expression analysis in Fig. 3b. Serpinh1, Gata6 and Nanog are labelled in blue, red and green, respectively.