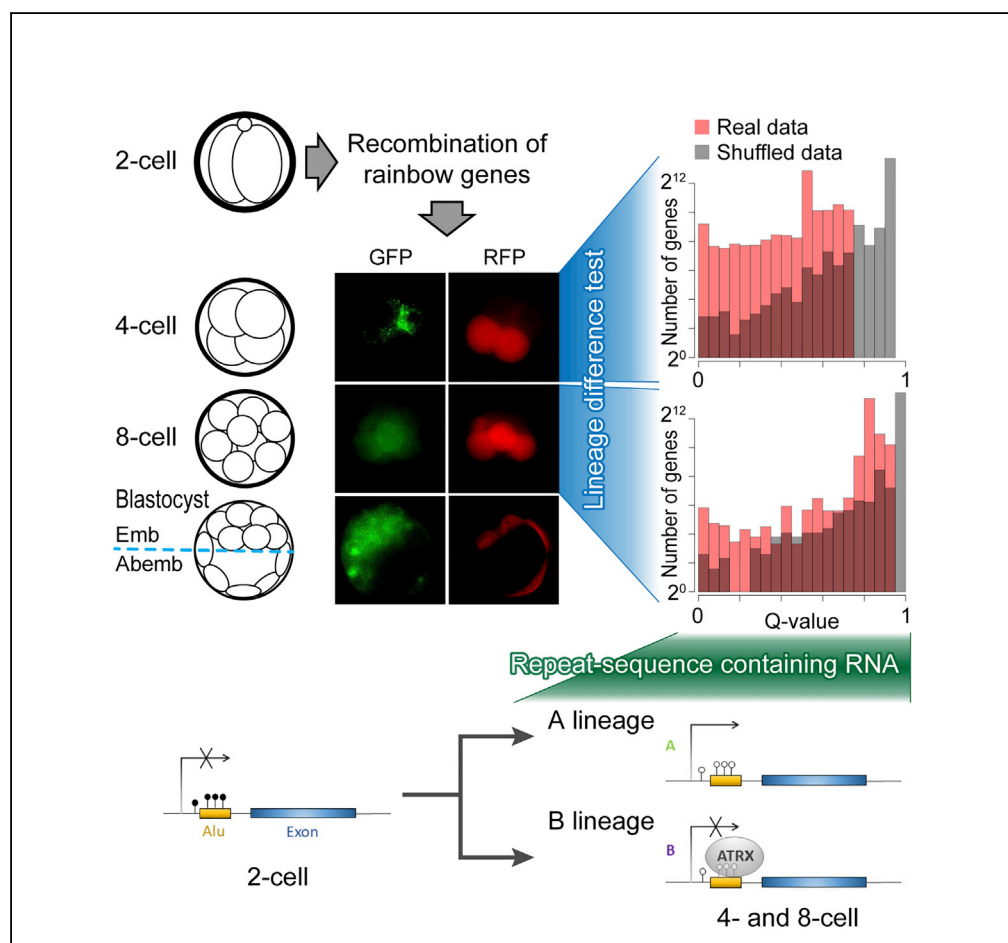


Article

Rainbow-Seq: Combining Cell Lineage Tracing with Single-Cell RNA Sequencing in Preimplantation Embryos



Fernando H. Biase, Qiuyang Wu, Riccardo Calandrelli, Marcelo Rivas-Astroza, Shuigeng Zhou, Zhen Chen, Sheng Zhong

szhong@ucsd.edu

HIGHLIGHTS

Rainbow-seq traces cell division history and reveals single-cell transcriptomes

A single gene does not exhibit unequivocal lineage difference at 4- and 8-cell stages

Lineage difference becomes clear when the transcriptome is analyzed as a whole

Novel RNA isoforms with embedded repeat sequences exhibit lineage difference

DATA AND SOFTWARE AVAILABILITY
GSE106287

Biase et al., iScience 7, 16–29
September 28, 2018 © 2018
The Authors.
<https://doi.org/10.1016/j.isci.2018.08.009>

Article

Rainbow-Seq: Combining Cell Lineage Tracing with Single-Cell RNA Sequencing in Preimplantation Embryos

Fernando H. Biase,^{1,5,6} Qiuyang Wu,^{1,2,6} Riccardo Calandrelli,¹ Marcelo Rivas-Astroza,¹ Shuigeng Zhou,³ Zhen Chen,⁴ and Sheng Zhong^{1,7,*}

SUMMARY

We developed the Rainbow-seq technology to trace cell division history and reveal single-cell transcriptomes. With distinct fluorescent protein genes as lineage markers, Rainbow-seq enables each single-cell RNA sequencing (RNA-seq) experiment to simultaneously decode the lineage marker genes and read single-cell transcriptomes. We triggered lineage tracking in each blastomere at the 2-cell stage, observed microscopically inequivalent contributions of the progeny to the two embryonic poles at the blastocyst stage, and analyzed every single cell at either 4- or 8-cell stage with deep paired-end sequencing of full-length transcripts. Although lineage difference was not marked unequivocally at a single-gene level, it became clear when the transcriptome was analyzed as a whole. Moreover, several groups of novel transcript isoforms with embedded repeat sequences exhibited lineage difference, suggesting a possible link between DNA demethylation and cell fate decision. Rainbow-seq bridged a critical gap between division history and single-cell RNA-seq assays.

INTRODUCTION

A central question to developmental biology is how cells break molecular and functional symmetry during mitotic divisions. Two models have been proposed (Altschuler and Wu, 2010). In one model, a progenitor cell first divides into a set of homogeneous cells, and then these homogeneous cells exhibit different tendencies toward different routes of differentiation (homogeneous model). In the other model, every division creates a pair of cells with slight differences in their molecular profile, and accumulation of these differences leads to cell fate commitment (non-homogeneous model).

Mammalian preimplantation development from a zygote to a blastocyst consisting of the functionally differentiated inner cell mass (ICM) and trophectoderm (TE) (Wolpert, 2015) is a perfect example of the symmetry-breaking paradigm (Gardner, 2001; Zernicka-Goetz et al., 2009). The first cell fate decision in mammals was believed to be an example of the homogeneous model (Hiiragi et al., 2006). After fertilization, a mouse zygote was thought to undergo two or more rounds of symmetric cell divisions, which create four or more homogeneous embryonic cells (blastomeres) (Leung and Zernicka-Goetz, 2015; Motosugi et al., 2005). These homogeneous cells further divide and through a stochastic process (Wennekamp and Hiiragi, 2012) eventually self-organize into ICM and TE (Wennekamp et al., 2013) (Figure 1A). However, recent work on single-cell transcriptome profiling reported reproducible inter-blastomere gene expression differences in 2- and 4-cell mouse embryos (Goolam et al., 2016; Biase et al., 2014; Burton et al., 2013; Piotrowska-Nitsche et al., 2005; Shi et al., 2015), offering support to the non-homogeneous model (Vogel, 2005; Hiiragi and Solter, 2006).

The critical missing information for using single-cell transcriptome profiling to test the competing models is the history of cell divisions (Goolam et al., 2016; Plachta et al., 2011; Torres-Padilla et al., 2007; Rodriguez-Terrones et al., 2018). We approached the challenge of connecting cell-to-cell transcriptome differences at later embryonic stages to progenitor cells by developing Rainbow-seq, which combines cell division lineage tracing (Tabansky et al., 2013) with single-cell RNA sequencing.

Advances in cell-specific genomic barcodes are revolutionizing cell lineage tracing. Leveraging genome editing tools (McKenna et al., 2016; Kalhor et al., 2017; Perli et al., 2016) (Raj et al., 2018) or combinatorial DNA recombination loci (Polylox) (Pei et al., 2017), researchers were able to insert cell-specific exogenous

¹Department of Bioengineering, University of California San Diego, San Diego, CA 92130, USA

²Department of Computer Science and Technology, Tongji University, Shanghai 201804, China

³School of Computer Science, Fudan University, Shanghai 200433, China

⁴Department of Diabetes Complications and Metabolism, City of Hope, Duarte, CA 91010, USA

⁵Present address: Department of Animal Sciences, Auburn University, Auburn, AL, USA

⁶These authors contributed equally

⁷Lead Contact

*Correspondence: szhong@ucsd.edu
<https://doi.org/10.1016/j.isci.2018.08.009>



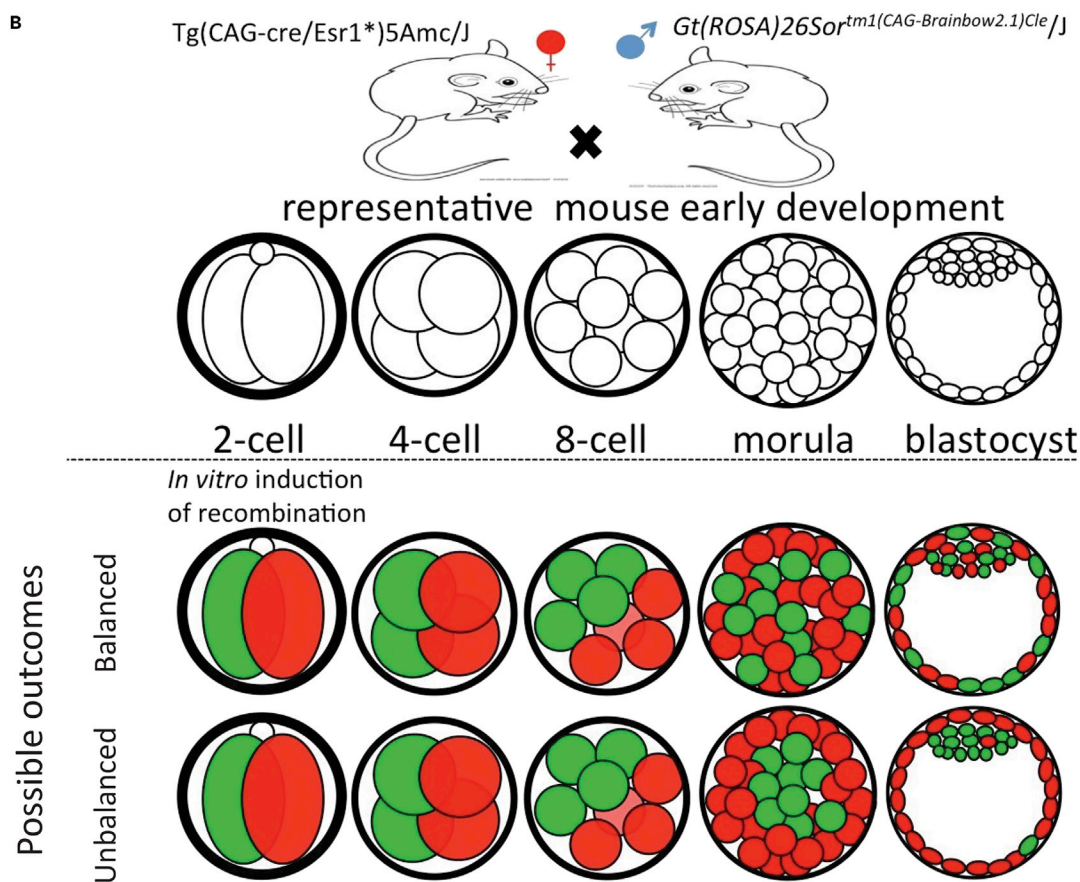
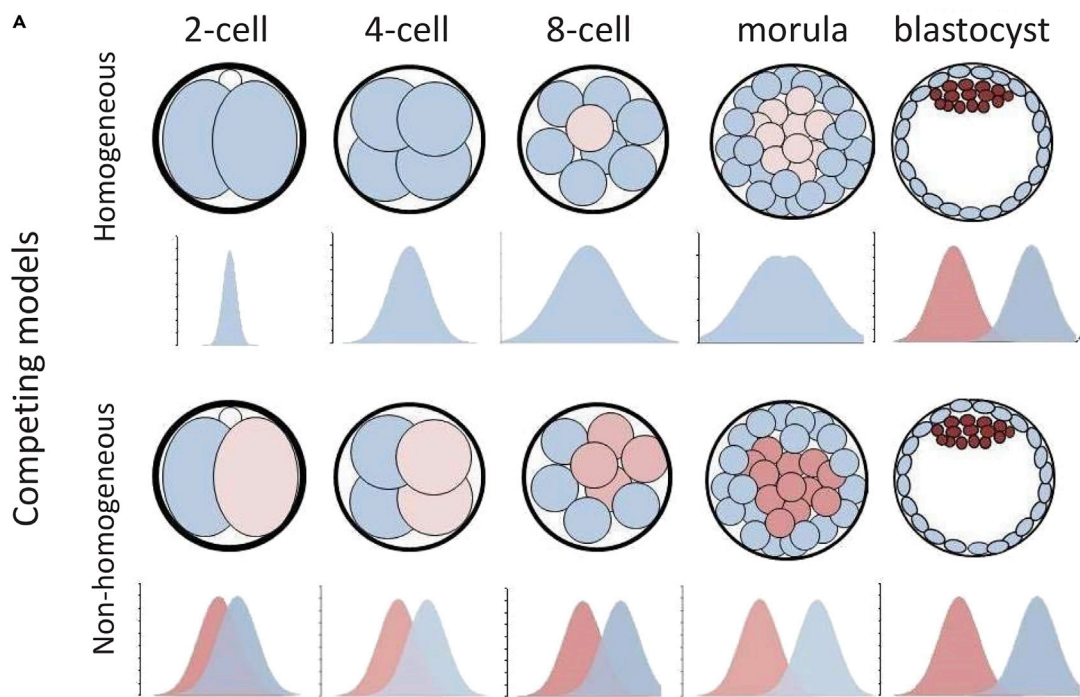


Figure 1. Competing Models and Lineage Tracking Strategy

(A) Homogeneous and non-homogeneous hypotheses differ in whether the 2 blastomeres at 2-cell stage have identical tendencies to contribute their descendant cells to inner cell mass (brown) and trophectoderm (blue-gray).

(B) Cell lineage tracing was enabled by creation of preimplantation embryos heterozygous for Brainbow sequence and Cre and induction of Cre at 2-cell stage. At blastocyst stage, activated Brainbow genes can exhibit either relatively equal (balanced) or biased (unbalanced) distributions in inner cell mass and trophectoderm. The equal and biased distributions would lend support to the homogeneous and the non-homogeneous hypotheses respectively.

DNA barcodes to the genomes of progenitor cells. The lineage information is propagated by DNA duplication and can be recovered in the progeny by targeted sequencing of the engineered genomic sequences. We reasoned that combining lineage-specific genomic barcodes and single-cell RNA sequencing (RNA-seq) would reveal single-cell transcriptome together with cell division lineage. To maximize the chances of obtaining expressible genomic barcodes, we resorted to the well-tested Brainbow-2.1 construct that can express one of four fluorescent protein genes (green fluorescent protein [GFP], cyan fluorescent protein [CFP], red fluorescent protein [RFP], yellow fluorescent protein [YFP]) (Livet et al., 2007). These fluorescent protein genes interspersed with recombination sites serve as both genomic barcodes and expressible cell markers. Their expression can be examined microscopically, offering an intermediate step of validation before single-cell RNA-seq analysis.

RESULTS**Cell Lineage Tracing Starting with 2-Cell Embryos**

We generated mouse embryos by crossing a strain that is homozygous for the Brainbow sequence (Cai et al., 2013) with a mouse strain that is homozygous for a tamoxifen-inducible Cre recombinase gene (Figure 1B). Tamoxifen treatment was expected to induce CRE translocation to the nucleus, which in turn could induce recombination on *loxP* sites, and thus lead to activation of one and only one of the four fluorescent protein genes in the Brainbow sequence (Cai et al., 2013). We tested whether these anticipated effects could be achieved by a short application of tamoxifen to early 2-cell embryos (Figure S1A). Upon collection of early 2-cell embryos (42 hours post injection) and a 2-hr 4-hydroxytamoxifen treatment, CRE protein translocated to the nuclei (Figure S1B), and within 1 hr post treatment the nuclear concentration of CRE reduced to the background level (Figure S1B). Embryos that underwent this brief tamoxifen treatment expressed fluorescent proteins at later developmental stages including 4-cell, 8-cell, and blastocyst stages (Figure 2A). Hereafter, we name these early 2-cell-stage hybrid embryos that undergo a 2-hr 4-hydroxytamoxifen treatment as *rainbow embryos*, and the Cre-activated genes as *lineage marker genes*.

Daughter Cells of Blastomeres at 2-Cell Stage Tend to Contribute Inequivalently to Embryonic and Abembryonic Poles at the Blastocyst Stage

We leveraged the Brainbow embryos to revisit the question whether the descendant cells produced by each blastomere of a 2-cell embryo contribute equivalently to embryonic and abembryonic poles in blastocyst-stage embryos (Hiiragi and Solter, 2004; Plusa et al., 2005; Vogel, 2005). Our experiment was designed as follows. First, a total of *N* blastocyst-stage rainbow embryos would be imaged by a wide-field fluorescent microscope with filters for GFP, RFP, and DAPI (Figures 2A and 2B). Because we had to devote one of three channels to nuclei staining (DAPI) for cell counting, this experimental design would not be able to detect CFP- and YFP-expressing cells. Next, the imaged embryos with more than one-third of their cells in the embryonic pole expressing a detectable lineage marker (either GFP or RFP) would be used for data analysis. A contingency table would be built for each embryo documenting the numbers of GFP-expressing (GFP+) and non-GFP-expressing (GFP-) cells in the embryonic pole and in the abembryonic pole. An odds ratio (OR) would be calculated for each embryo, and if $1/3 < OR < 3$, this embryo would be regarded as having balanced GFP+ cells in the two poles, otherwise the embryo would be regarded as having unbalanced GFP+ cells in the two poles. The same analysis would be carried out for RFP.

We imaged 18 early blastocysts expressing RFP and/or GFP, of which 12 embryos had more than one-third of cells expressing fluorescent protein. Eight of these embryos (66%) exhibited unbalanced GFP+ or RFP+ cells in the two poles (first column, Figure 2C). For sensitivity analyses, we used two alternative counting methods to re-estimate the proportion of embryos with unbalanced distribution of cell lineage between the two poles. First, we required the embryos to be included for analysis to have at least 50% of its cells expressing either GFP or RFP, resulting in a total of 9 embryos, of which 6 embryos (66%) exhibited unbalanced GFP+ or RFP+ cells in the two poles (second column, Figure 2C). Second, we counted the GFP+ cells in the embryonic pole (*n*₁) and GFP+ cells in the abembryonic pole (*n*₂) and regarded any embryo with

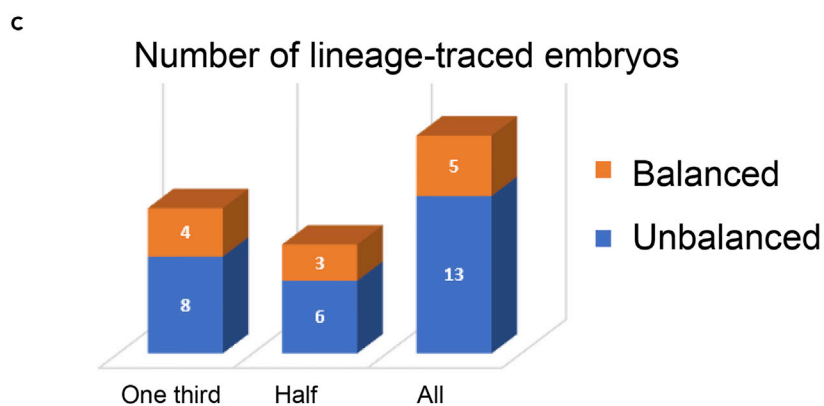
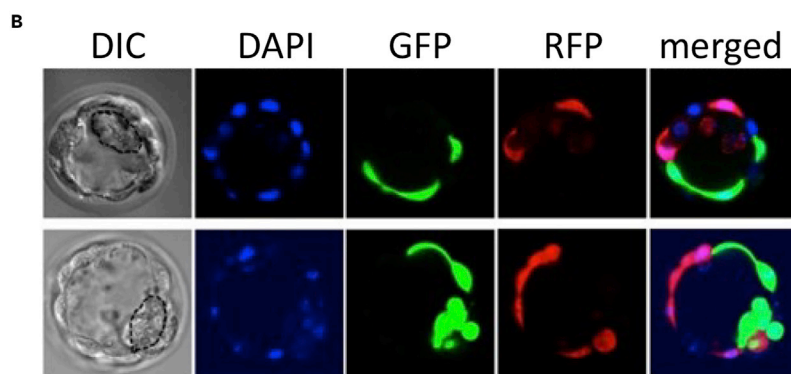
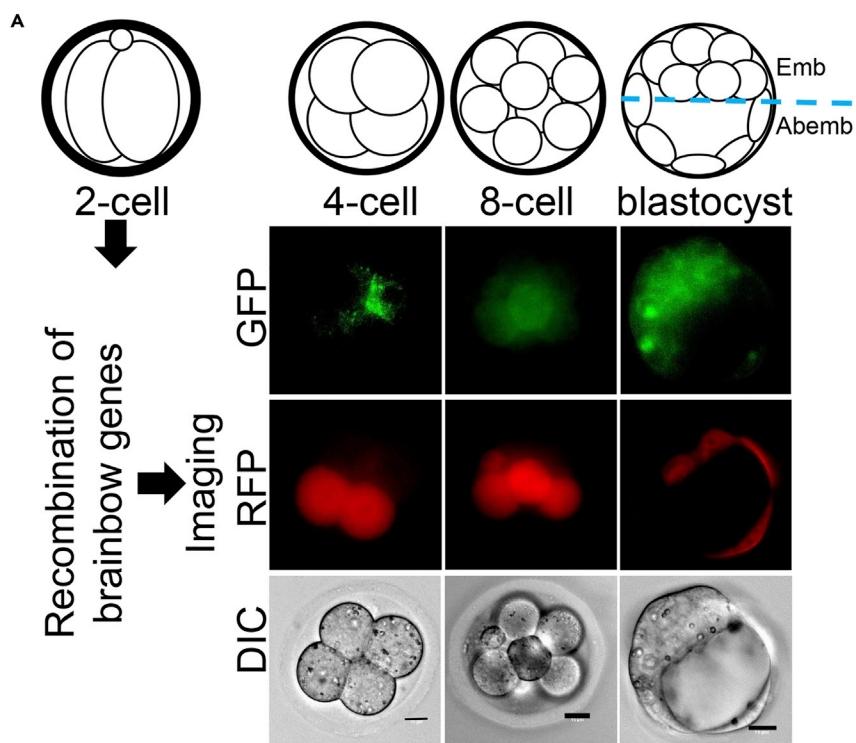


Figure 2. Cell Lineage Tracking in Mouse Preimplantation Embryos

(A) Recombination of Brainbow construct was induced in early 2-cell embryos, and images were acquired at 4-cell, 8-cell and blastocyst stages. GFP: GFP, green fluorescent protein; RFP, red fluorescent protein; DIC: differential

Figure 2. Continued

interference contrast. Scale bar: 15 μm . Images were acquired by a wide-field microscope containing filters for RFP and GFP.

(B) Two representative blastocyst-stage embryos imaged with DIC, DAPI, GFP, and RFP channels.

(C) Numbers of blastocyst-stage embryos with balanced (orange) and unbalanced (blue) lineage marker expression in the two embryonic poles, with respective counting methods (three lanes). Data from different counting methods are shown in columns, which include the method requiring at least 1/3 cells (one-third) or 1/2 cells (half) expressing either GFP or RFP, and the method of counting all the imaged embryos (all).

See also [Figure S1](#).

1/3 < n_1/n_2 < 3 as having balanced GFP+ cells, and did the same counting for RFP. All 18 embryos were included in this analysis, of which 13 embryos (72%) were unbalanced (third column, [Figure 2C](#)). Taken together, approximately two-thirds of rainbow embryos exhibited unbalanced distribution of the two cell lineages in the two embryonic poles. Previous efforts obtained discrepant estimates of such a proportion ([Piotrowska et al., 2001](#); [Hiiragi and Solter, 2004](#); [Plusa et al., 2005](#); [Vogel, 2005](#)). Our results obtained from rainbow embryos were better aligned with those suggesting more embryos exhibiting unequal contribution of the two lineages than those exhibiting equal contribution to the embryonic and abembryonic poles. These microscopic analyses confirmed that the recombination strategy introduced lineage-specific expressible marker genes and served as a sanity check for downstream Rainbow-seq analyses.

Single-Cell RNA Sequencing Analysis of Rainbow Embryos

We induced lineage tracking from 2-cell stage and carried out Rainbow-seq on nine 4-cell embryos and four 8-cell embryos. We lost one blastomere from an 8-cell embryo during the manipulation. We combined the Smart-seq2 full-length RNA-seq protocol ([Picelli et al., 2014](#)) with deep paired-end sequencing and produced paired-end 100-nt sequencing data from a total of 67 single blastomeres, including 36 from 4-cell stage and 31 from 8-cell stage ([Tables S1 and S2](#)). One 8-cell-stage blastomere did not pass our quality control as it produced fewer than 3 million uniquely mapped sequencing read pairs (Row E13, B7, [Table S2](#)). The remaining 66 cells yielded on average 21 million uniquely mapped read pairs (mm10) per cell ([Figure S2A](#)). We used single-cell RNA-seq reads mapped to the Brainbow-2.1 sequence to determine which lineage marker gene was expressed and categorized the cells from each embryo into two groups (lineage A and B) based on its expressed marker gene (Lineage column, [Tables S1 and S2](#)). Data were insufficient to resolve three 4-cell-stage embryos (12 blastomeres) and four 8-cell-stage blastomeres due to limited sequencing reads mapped onto any lineage marker gene. We deposited all sequencing data into GEO (GSE106287) for public access; however, we limited our analyses to the 50 blastomeres whose lineage can be traced to either one of the sister blastomeres at the 2-cell stage.

The first two principle components explained 8% of sample variations, where the developmental stage was a major source of variation ([Figure S2B](#)). With Fragments Per Kilobase of transcript per Million mapped reads (FPKM) > 1 as the threshold for calling expressed genes, the 4-cell-stage blastomeres on average expressed 6,852 coding genes per cell. A total of 15,218 coding genes were expressed in at least one 4-cell-stage blastomere. The 8-cell blastomeres on average expressed 6,474 coding genes per cell, with a total of 15,342 coding genes expressed in at least one 8-cell blastomere.

Transcriptome-wide Expression Difference between Cell Division Lineages

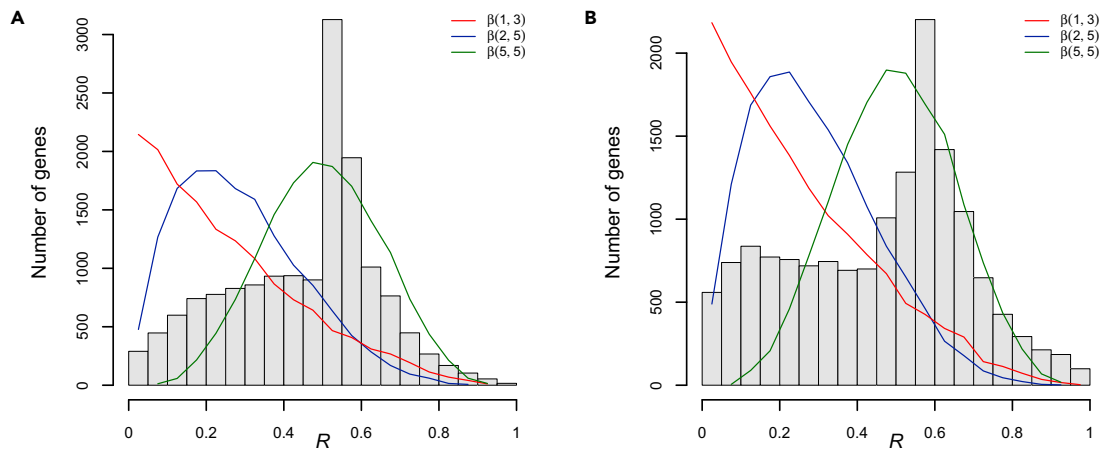
Recognizing that data of individual genes may not provide conclusive evidence to either the homogeneous or the non-homologous model, we analyzed data of the entire transcriptome. We carried out two transcriptome-wide tests. First, we tested whether there is non-zero probability for any subset of genes to exhibit greater between-lineage variation than embryo-to-embryo variation. To this end, we fit a generalized linear model to every gene (g) and calculated the sum of squares to account for the between-lineage variation ($SS_{lineage}^g$) and embryo-to-embryo variation (SS_{embryo}^g). For every gene, we calculated the ratio (r^g) of lineage sum of squares to combined sum of squares of embryo and lineage. The empirical distribution of this ratio (denoted as R) exhibited greater tail probabilities than a series of three β distributions, $\beta(2,5)$, $\beta(1,3)$, and $\beta(5,5)$, each of which has non-zero probability in any non-singular interval near 1 ($(1-\delta, 1)$, ($\delta > 0$)) ([Figures 3A–3D](#)). The non-zero probability for a subset of genes to exhibit greater lineage variation than embryo variation suggests a transcriptome-wide difference between the two division lineages.

Second, after a gene-by-gene test of the null hypothesis that this gene does not exhibit between-lineage expression differences (ANOVA, [Transparent Methods](#)), we obtained the empirical distribution of the

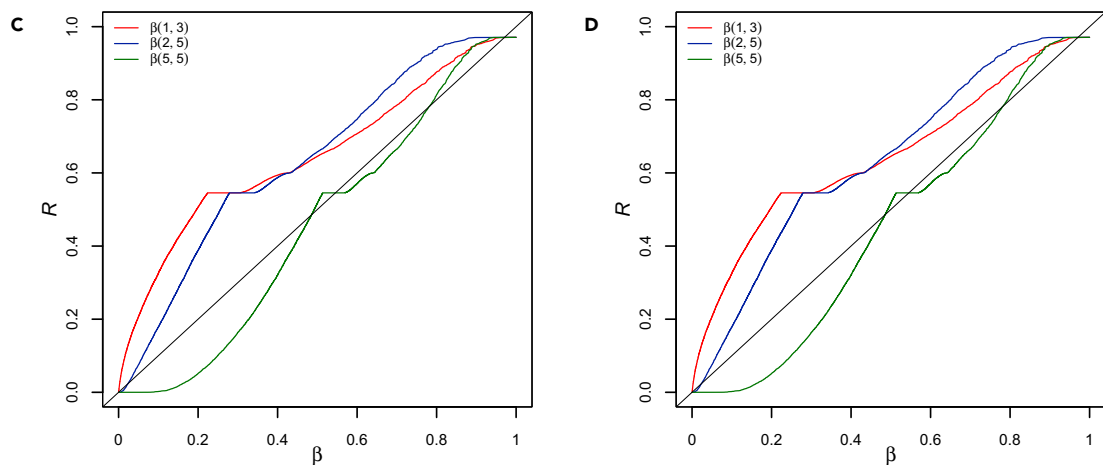
4-cell stage, n=15,218

8-cell stage, n=15,342

Distribution of the ratio of between-lineage variation to combined variation (R)



Q-Q plots comparing between-lineage ratio (R) distribution to β distributions



Q-values from lineage equivalence test

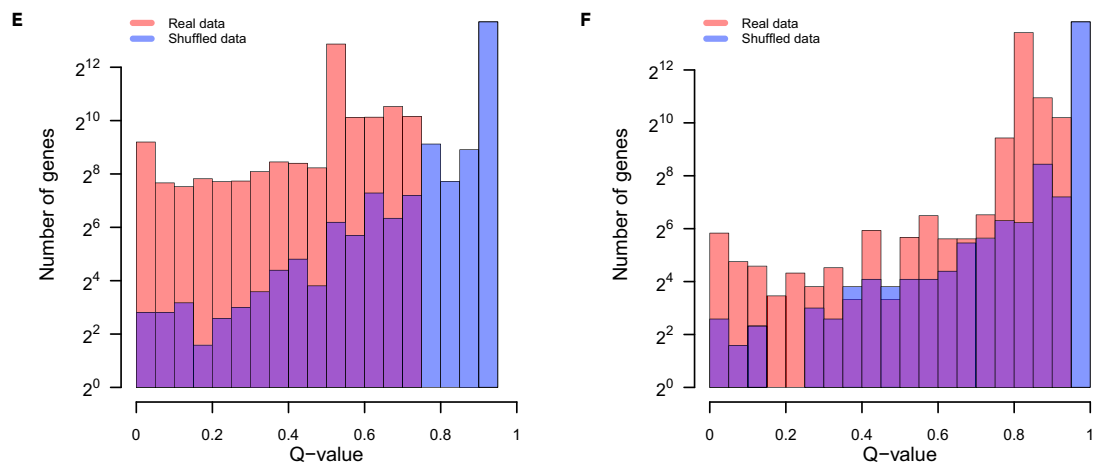


Figure 3. Transcriptome-wide Differences between Two Lineages

(A and B) Empirical distributions of random variable R (ratio of between-lineage variation to combined variation) in 4-cell (A) and 8-cell-stage (B) embryos, superimposed with a series of β distributions [red: $\beta(1,3)$, blue: $\beta(2,2)$, green: $\beta(5,5)$].

(C and D) Q-Q plots of R versus a series of β distributions in 4-cell (C) and 8-cell (D) stage embryos, where red, blue, and green represent $\beta(1,3)$, $\beta(2,2)$, and $\beta(5,5)$, respectively.

(E and F) Distributions of q-values from lineage equivalence tests based from real data (red) and shuffled data (blue) in 4-cell (E) and 8-cell-stage (F) blastomeres.

See also Figure S2 and Tables S1 and S2.

q-values for all the genes from such a test (red bars, Figures 3E and 3F). To derive a background distribution, we shuffled the lineage labels on every blastomere and carried out the same test (blue bars, Figures 2E and 2F). The distribution of q-values from real data was skewed toward lower values when compared with the q-values from shuffled data (p value $< 2.2 \times 10^{-16}$, Kolmogorov test), reflecting a transcriptome-wide difference between the two cell division lineages.

Specific Genes with Expression Differences between Cell Division Lineages

The genes with the largest between-lineage differences in transcript abundance at 4-cell stage were involved in the basic transcription machinery including *Gtf2f1* (General Transcription Factor IIF Subunit 1, TFIIIF) and *Eloa* (RNA Polymerase II Transcription Factor SIII Subunit A1, SIII), negative regulations of P53 activity (*Aurkb*) and WNT signaling (*Ctnnbip1*, Inhibitor of Beta-Catenin and *Tcf4*), nuclear pore complex (*Nup35*), protein synthesis in cytoplasm (*Rbm3*) and mitochondrion (*Mrpl10*), endoplasmic reticulum (ER) biogenesis (*Atl2*), and glycogen synthesis (*Fbp2*) (Figure 4A). The genes with the largest between-lineage variations at 8-cell stage were transcriptional suppressors *Bclaf1* and *Smad7*, the chromatin remodeler *Atrx* involved in deposition of H3K9me3, histone demethylase *Kdm2b* related to removal of active histone marks including H3K4me3, vesicle trafficking and secretory pathways including Golgi membrane proteins *Glg1* and *B4galnt2*, trans-Golgi network, *Snx15* and Golgi-to-ER transport *Stx8* (vesicle fusion, Golgi-to-ER transport), cell shape modulator *Fez2*, and the lincRNA genes *Gm12514* and *GM14617* (Figure 4B). Interestingly, suppression of H3K4 demethylase *Kdm2b* in mouse oocytes resulted in impaired blastocyst formation (see extended Figure 6E in Liu et al. (2016)). Even though the above-mentioned genes appeared to support the non-homologous model, the absolute expression difference is small (Figure 4) and statistical significance is obscure due to limited sample size and multiple comparisons. At this point, it became clear that the statistical tests based on the entire transcriptome (Figure 3) served to better evaluate the two competing models than any possible test based on any individual gene.

Transposon-Related Novel Transcript Isoforms

We tested whether any novel transcript isoforms were specifically expressed in early-stage preimplantation embryos. To this end, we combined all the Rainbow-seq datasets to obtain over 2.1 billion paired-end reads, of which over 1.6 billion read pairs could be uniquely mapped to the genome (mm10). Using Cufflinks (v2.2.1 with g parameter), we recorded a total of 20,371 novel transcript isoforms, corresponding to 7,594 annotated genes. Among these novel isoforms, 3,632 (17.8% of all novel isoforms) contained transposon sequences according to Repeatmasker database (February 2015 release) (Figure 5A), which will be referred to as Transposon Related Novel Isoforms (TRENIs). These 3,632 TRENIs corresponded to 1,701 Refseq genes, in which “poly(A) RNA binding,” “RNA binding,” “nucleotide binding,” and “nucleic acid binding” were enriched (Benjamini adjusted p value $< 1.1 \times 10^{-7}$; Gene Ontology functions [DAVID v6.8]) (Figure 5B), suggesting a recurring theme in RNA binding and processing.

We asked whether the incorporation of these transposons in TRENIs were supported by junction reads spanning across the transposon and the nearest known Refseq exons. To this end, we computed the ratio of junction reads and the total number of reads aligned to each transposon, hereafter denoted as junction ratio (JR, Transparent Methods). We identified 3,902 transposons contained in TRENIs, of which 2,817 (72.19%) exhibited a JR of 0.9 or above (Figure 5C).

To test whether TRENIs exhibit between-lineage expression difference, we derived a q-value for each of the 3,632 TRENIs by ANOVA (Transparent Methods). For comparison, we shuffled the lineage labels on all blastomeres and derived q-values again (background q-values). The distribution of real data q-values differed with that of background q-values (p value $< 2.2 \times 10^{-16}$, Kolmogorov test) (Figures 5D, 5E, and S3). More importantly, a total of 37 TRENIs exhibited q-values less than 0.01, whereas none of the background q-values were less than the same threshold (Figure S3). Finally, all the repeats expressed in

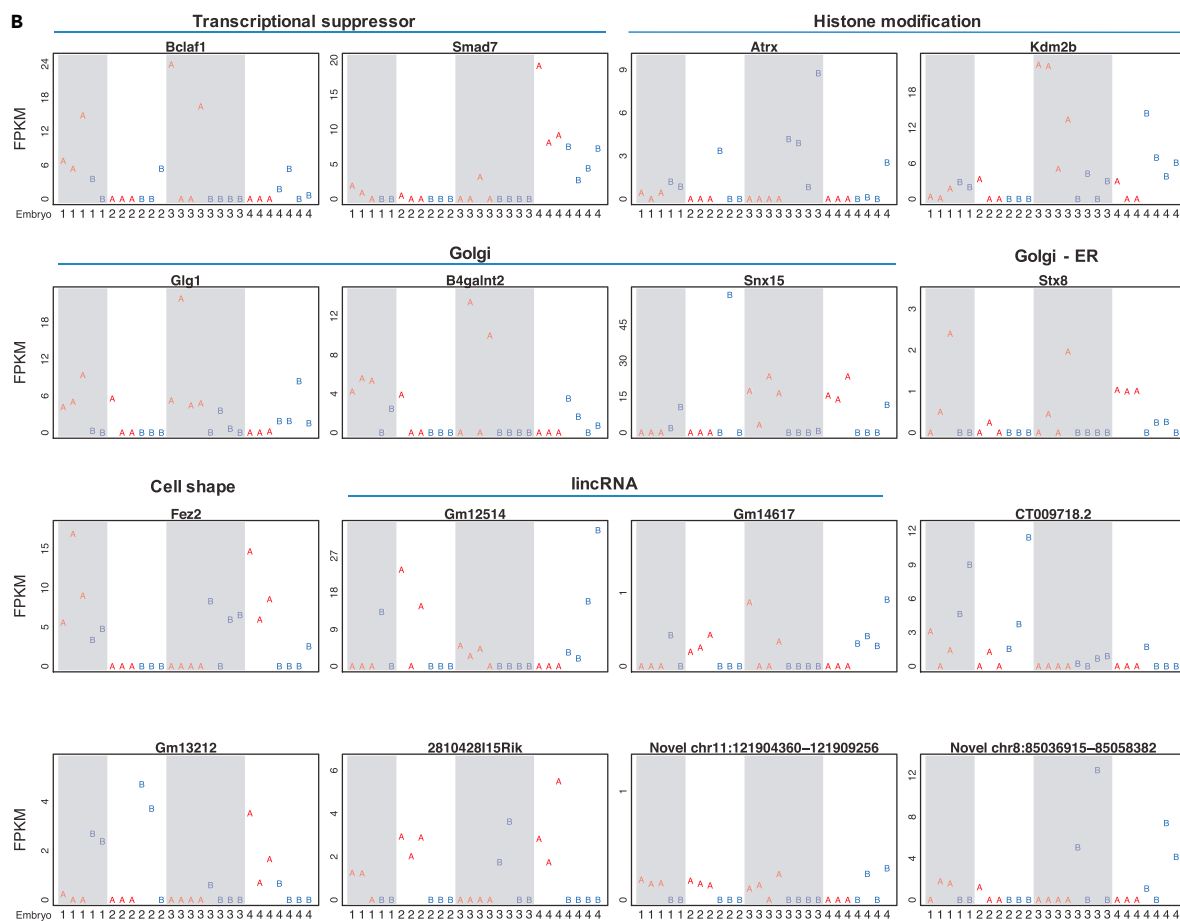
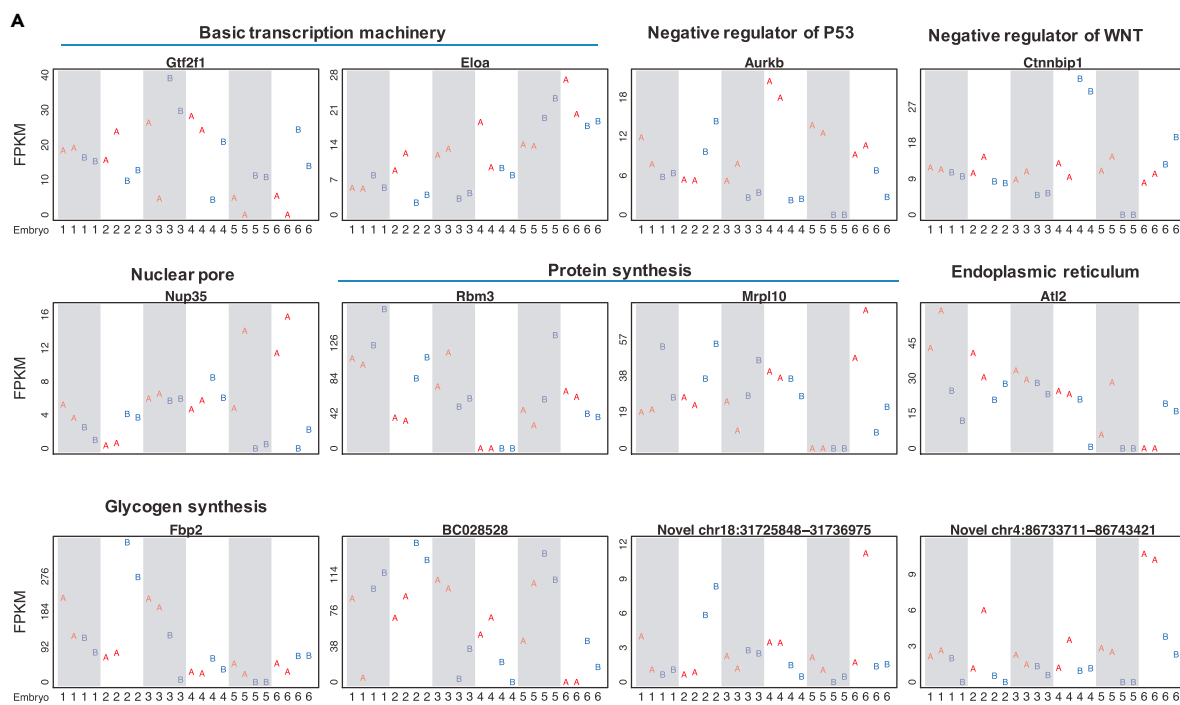


Figure 4. Genes with differential expression between two lineages.

Genes with the between-lineage differences at 4- (A) and 8-cell (B) stages. FPKM (y axis) of every blastomere (column) in each embryo (marked by embryo number in columns). Shaded columns delineate different embryos. The blastomeres of the two lineages are marked with A (red) and B (blue).

4- and 8-cell embryos, regardless of whether they were contained within TRINEs, did not exhibit statistically discernable between-lineage differences (the distribution of q-values derived from real data was not different from that of shuffled data) (Figure S4).

Murine Alu Makes the Largest Contribution to Novel Transcript Isoforms

We asked whether different repeat families contributed equally to TRENIs. In the total of 58 repeat families (RepeatMasker) 18 families had 20 or more genomic copies included in TRENIs (Figure 5F). To account for genome-wide copy numbers of each repeat family, we calculated the OR for each repeat family to be contained in TRENIs. When repeat families were ranked by this OR, murine Alu, B4, ID, and B2 were most enriched in TRENIs (Figure 5G). We asked whether the association between any repeat family and TRENIs would become more pronounced when only high-confidence transposon-containing TRENIs were included in the analysis. To this end, we repeated the above analysis using only the 2,817 transposons with $JR \geq 0.9$, which resulted in similar ORs with Alu being the most enriched repeat family (Figure S5).

Next, we asked whether different repeat families contributed equally to TRENIs' 5' UTR. There were a total of 673 (18.53%) TRENIs in which the transposons were inserted in the 5' upstream to previously annotated start codons (5' UTR-TRENIs) (Figure 5H), among which 418 contained murine Alu, simple repeat, low complexity, and B4 and B2 repeats in 5' UTRs, where murine Alu alone accounted for more than three hundred 5' UTRs (Figure 4H). We calculated the OR to quantify the association of transposons in each transposon family to 5'UTR-TRENIs. Murine Alu exhibited the largest enrichment to 5' UTRs in these novel transcript isoforms (OR = 3.03, p value = 0, chi-square test) (Figure 4H). In contrast, the other repeat families exhibited either moderate enrichment ($1 < OR < 1.5$) or depletion ($OR < 1$) in 5' UTR-TRENIs (Figure 5I). Finally, 9 and 26 5' UTR-TRENIs exhibited expression differences between the two division lineages at 4- and 8-cell stages, respectively, including novel transcript isoforms of *Gata4* (Figure S6A) and *Khdca1a* (Figure S6B).

DISCUSSION**Lack-of-Continuity Dilemma**

Not a single gene has been reported to exhibit continuous between-lineage expression difference starting from the two blastomeres at 2-cell stage until the formation of ICM and TE at blastocyst stage. This is not completely unexpected considering that comprehensive transcriptome analyses revealed highly dynamic gene transcription in early-stage mouse preimplantation embryos (Kidder and Pedersen, 1982; Xie et al., 2010), likely attributable to combined effects of cleavage-related split of RNA pool, RNA degradation, and zygotic genome activation (Biase et al., 2014). The lack of identified single gene(s) with continuous expression divergence posed a dilemma to the non-homologous hypothesis, in that if the two blastomeres at 2-cell stage are poised for cell fate decisions, what could be the molecular means to record and later exhibit such a difference? Hereafter we call this dilemma the lack-of-continuity dilemma.

Reconciliations to Lack-of-Continuity Dilemma

The dilemma was rooted in the non-homogenous model and may be resolved in several ways. First, the cell fate decision is initiated after the first cleavage. This idea essentially refutes the non-homogeneous model for the first cleavage. Second, the lineage difference is passed from one gene activated at an earlier stage, for example, 2-cell stage, onto another gene activated at a later stage. Given that the transcription rate and mRNA half-life are both an order of magnitude smaller than the time of a cleavage, this model is within the boundaries of biophysics principles. Third, the lineage difference is recorded in the order of expression levels of a pair or list of genes, and this order is preserved throughout preimplantation development. For example, at 2-cell stage the expression level order is $A > B$ in one blastomere and $B > A$ in the other blastomere, and such an order is preserved in the ICM and TE. The third explanation was supported by single-cell RNA-seq data (Biase et al., 2014). Taken together, there are plausible means to explain the dilemma while withholding the homologous model.

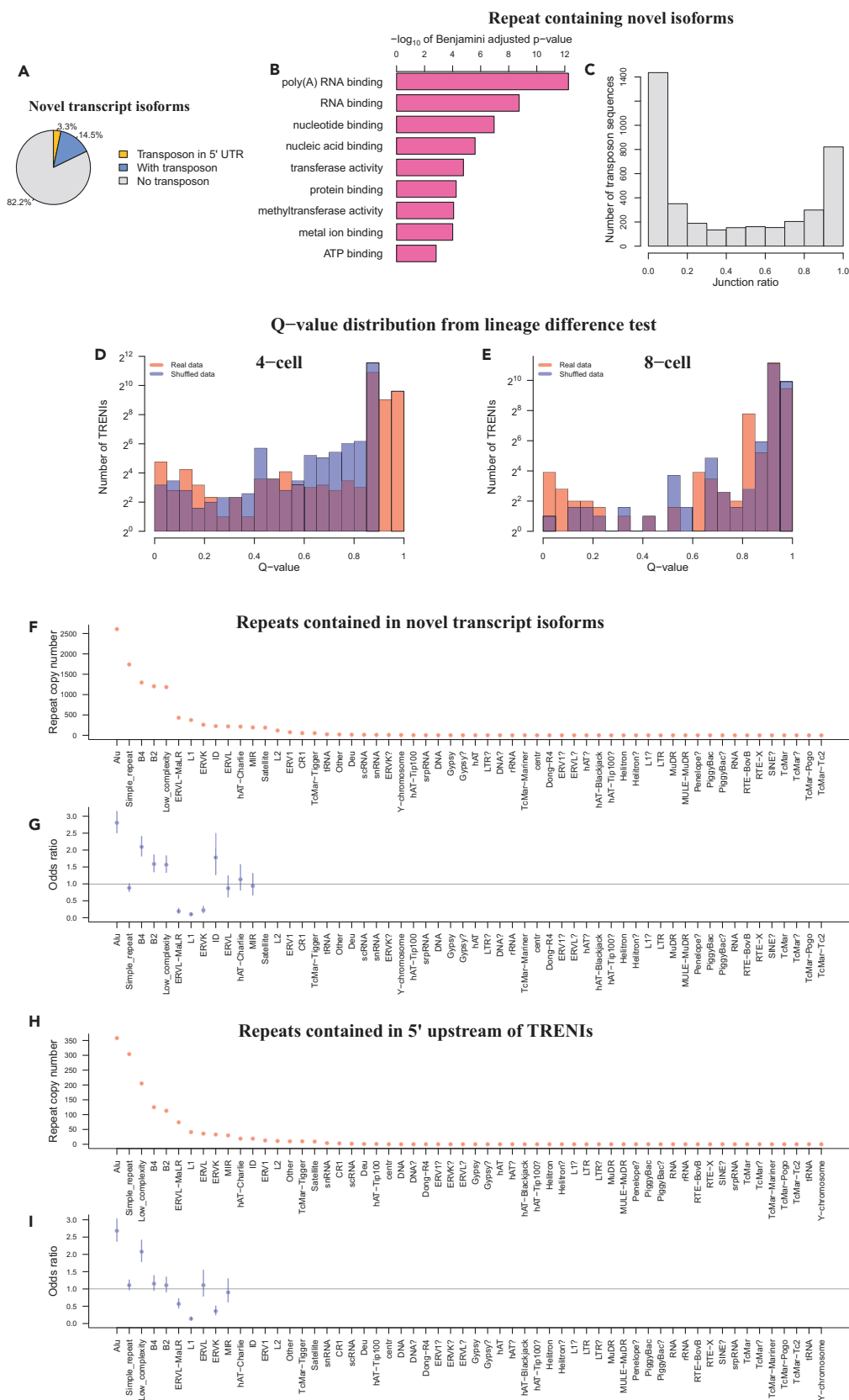


Figure 5. Transposon Related Novel Isoforms (TRENI)

(A) Pie chart of novel transcript isoforms, including 16,739 that do not contain any repeat sequence (gray) and 3,064 and 673 isoforms with transposons located downstream to start codons (blue) and in 5' UTRs (yellow).

(B) Enriched GO molecular functions in TRENI, ranked by multiple hypothesis testing adjusted p values.

(C) Histogram of junction ratios for all TRENI-containing transposons.

(D and E) Histogram of TRENI's q-values derived from tests of equivalent expression between lineages at 4- (D) and 8-cell (E) stages, based from real data (red) and shuffled data (blue).

(F) Copy number of repeats contained in TRENI, categorized by transposon family (columns).

(G) Odds ratios between TRENI and each repeat family (columns). Vertical bars: 95% confidence intervals. Odds ratio was not calculated for those repeat families when the TRENI associated with a repeat family were fewer than 20.

(H) Copy number of repeats contained in TRENI's 5 UTRs, categorized by transposon family (columns).

(I) Odds ratios between 5'UTR-TRENI and each repeat family (columns). Vertical bars: 95% confidence intervals. Odds ratio was not calculated for those repeat families when the 5'UTR-TRENI associated with a repeat family were fewer than 20.

See also [Figures S3–S6](#).

The Rainbow-seq Approach

Rainbow-seq addresses the lack-of-continuity dilemma from another perspective. The main idea is to combine cell division tracing with single-cell RNA-seq. A genetic cell-lineage tracer was activated at 2-cell stage and analyzed at later stages. This method has two advantages. First, activation of distinct tracer genes in each blastomere's genome was achieved by a relatively simple procedure. Second, through a genetic tracer embedded in the genome, DNA duplication ensured inheritance of the tracer during cell divisions. We used Rainbow-seq to test the consequential events of two cell division lineages created by the first cleavage.

The Lineage Difference Lies in the Collective Behavior of Many Genes

When each gene is individually analyzed, even the most significant genes only exhibited small expression differences between the lineages ([Figure 4](#)). However, the difference of the two lineages became clear when the entire transcriptome was analyzed as a whole. In other words, the collective distribution of transcriptome exhibited a significant difference in the two cell division lineages ([Figure 3](#)). This observation has important implications. On the one hand, it repeated once again the negative results of previous studies that not a single gene exhibited clear difference between the two division lineages created by the first cleavage. On the other hand, the two division lineages in fact exhibit pronounced difference, which lies in the integral expression of many genes. To translate this difficult concept, we would like to make an analogy that "the individual seed beads all look similar, but two bead-woven crafts can exhibit very different patterns."

Recent studies have reported inter-blastomere expression difference of several specific genes at the 4-cell stage or later stages ([Biase et al., 2014](#); [Burton et al., 2013](#); [Piotrowska-Nitsche et al., 2005](#)). This study pushed the initiation time point of between-blastomere difference to the 2-cell stage. The top-ranked genes with between-lineage differences appeared to emphasize a theme of negative regulation, including inhibitors of the WNT pathway *Cttnbip1* (*Inhibitor of Beta-Catenin and Tcf4*) ([Tago et al., 2000](#)) and *Smad7* ([Han et al., 2006](#)), *aurora kinase B* that negatively regulates P53's transcriptional activity by promoting its degradation ([Gully et al., 2012](#)), chromatin remodeler *Atrx* that promotes repressive histone mark H3K9me3 and formation of heterochromatin ([Sadik et al., 2015](#)), histone demethylase *Kdm2b* that removes active histone mark H3K4me3 ([Tsukada et al., 2006](#)), and glycosyltransferase *B4galnt2* involved in negative regulation of cell-cell adhesion ([Kawamura et al., 2005](#)) ([Figure 4](#)). Considering that negative feedback as exemplified in Delta-Notch signaling is perhaps the best known molecular mechanism for neighboring cells to take on divergent functions ([Artavanis-Tsakonas et al., 1999](#)), the emergence of negative regulators from Rainbow-seq analysis may allude to cell-cell communication between the first two cell lineages. The other emerged theme of vesicle trafficking and secretion as exemplified by *Atf2*, *Glg1*, *B4galnt2*, *Snx15*, and *Stx8* is consistent with the idea of cell-cell communication at very early embryonic stages ([Calarco-Gilam, 1986](#)) ([Sozen et al., 2014](#)). Finally, *Fbp2*'s divergent presence in the two lineages is reminiscent of a sperm-related embryo patterning hypothesis ([Zernicka-Goetz, 2002](#); [Takaoka and Hamada, 2014](#)), considering the presence of *Fbp2* mRNA in mouse sperm ([Schuster et al., 2016](#)) and interactions of FBP2 and ALDOA, a sperm-head protein essential for sperm-egg interaction ([Petit et al., 2013](#)).

Contribution of Transposon Related Novel RNA Isoforms (TRENI) and DNA Demethylation to Lineage Difference

A group of novel transcript isoforms of known genes that contained repeat sequences, called TRENI, exhibited lineage difference, especially at the 8-cell stage ([Figure 5D](#)). The lineage difference is more

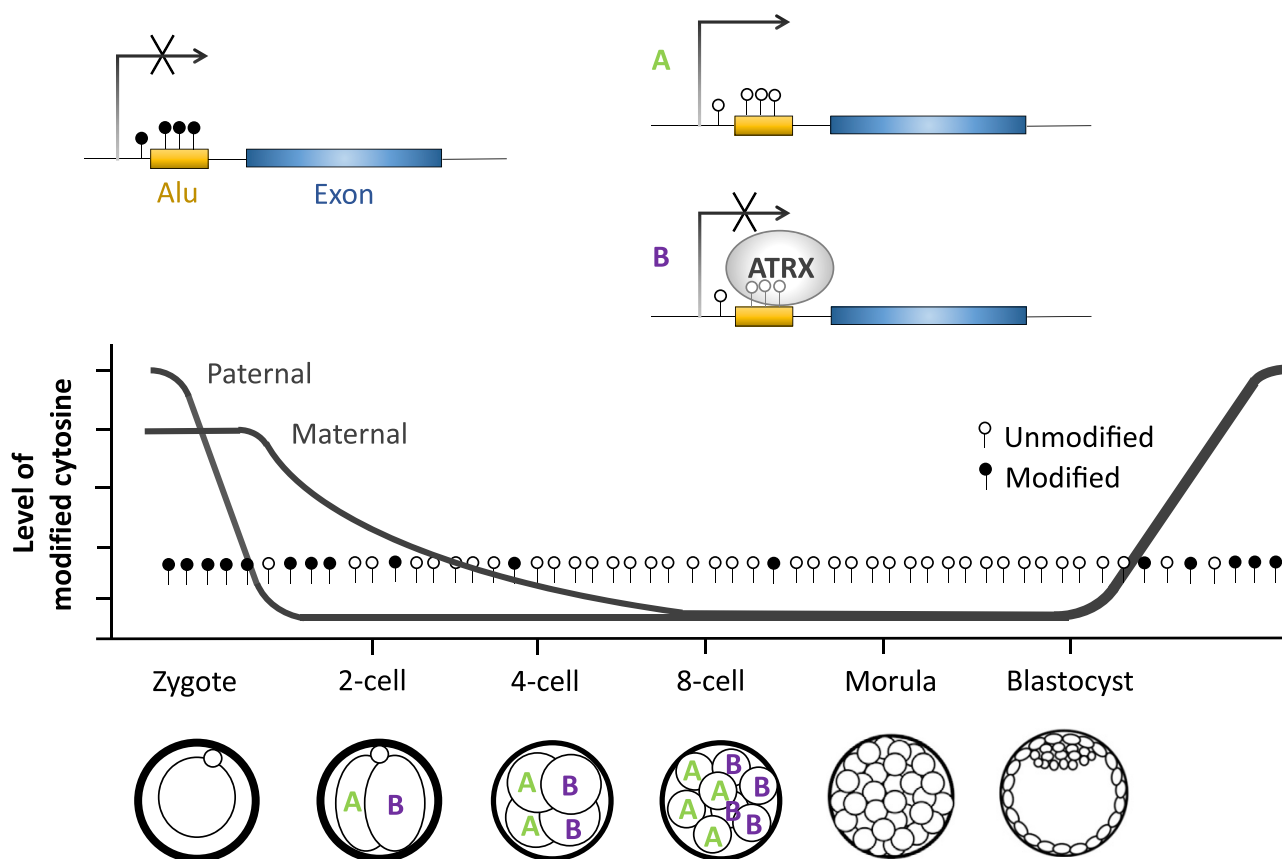


Figure 6. A Model of Repeat Sequence-Related Lineage Difference

Decrease of modified cytosine (filled circles) allows repeat sequence (yellow) to be expressed (A). *Atrx* is expressed in one lineage where ATRX preferentially attaches to hypomethylated (empty circles) repeats and suppresses their expression (B).

pronounced in each group of TRENIs (Figures 5D and 5E) than an individual TRENi (Figure S6), reinforcing the idea that lineage difference at such early stages of development is marked by the cumulative behavior of many genes. At 8-cell stage, chromatin remodeler *Atrx* that responds to DNA demethylation and suppresses the expression of hypomethylated repeat sequence in preimplantation embryos (He et al., 2015) exhibited between-lineage expression difference (Figure 4B). These data suggest a model where DNA demethylation on repeat sequence is asymmetrically compensated by ATRX in the two cell lineages, leading to differential expression of transposon-containing transcripts (Figure 6). Murine Alu repeats were most enriched in TRENIs, and the murine Alu family was the only repeat family with significant enrichment at the 5'UTR of TRENIs. These expression data may offer an explanation to the selective retention of Alu repeats to upstream and intronic regions of known genes (Tsirigos and Rigoutsos, 2009), where the selectively retained Alu repeats are part of rare transcript isoforms, which are specifically expressed in preimplantation embryos. Furthermore, asymmetric suppression of these repeats leads to asymmetric expression of genes required for the first cell fate decision, which is the functional force for selective retention. Consistent with this idea, the same group of genes involved in "RNA binding" were enriched in TRENIs and were reported by an independent analysis as the neighboring genes to selectively retained Alu repeats (Tsirigos and Rigoutsos, 2009). However, the other transcripts of repeat sequences that were expressed in preimplantation embryos (Rodríguez-Terrones and Torres-Padilla, 2018) but not contained within TRENIs did not exhibit statistically discernable lineage differences (Figure S4).

Limitations of the Study

Tradeoff between the Number of Embryos and Sequencing Depth

Our experimental design was determined by our analysis goals. The goals of identifying potentially very small inter-blastomere differences and novel transcript isoforms required deep paired-end

sequencing of full-length transcriptome in single cells. Most single cells in this study produced 20 to 30 million uniquely mapped 100-nt read pairs, which are 40–60 million uniquely mapped 100 nt reads per cell. Sequencing cost remains a limitation at this sequencing depth. If future applications do not care about transposon-produced transcripts or novel RNA isoforms, further experiments can significantly reduce the sequencing depth, use single-end sequencing, and even replace with other protocols that do not sequence full-length transcripts. As a tradeoff, more single cells can be sequenced.

METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

DATA AND SOFTWARE AVAILABILITY

The GEO accession number for the data reported in this paper is GEO: GSE106287.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, six figures, and two tables and can be found with this article online at <https://doi.org/10.1016/j.isci.2018.08.009>.

ACKNOWLEDGMENTS

The authors thank Drs. Bo Huang and Xiaoyi Cao for useful discussions and help on graphical presentation. This work is supported by NIH R00HL122368 (Z.C.) and DP1HD087990 (S.Zhong.), and the National Key Research and Development Program of China 2016YFC0901704 (S.Zhou).

AUTHOR CONTRIBUTIONS

Conceptualization, S.Zhong; Methodology, F.H.B., Q.W., R.C., M.R.A., and S.Zhong; Investigation, F.H.B., Q.W., R.C., M.R.A., S.Zhou., Z.C., and S.Zhong.; Writing – Original Draft, F.H.B., Q.W., R.C., and S.Zhong.; Writing – Review & Editing, Z.C. and S.Zhong.; Funding Acquisition, S.Zhou, Z.C., and S.Zhong.; Resources, F.H.B. and S.Zhong.; Supervision, S.Zhong.

DECLARATION OF INTERESTS

S.Zhong. is a co-founder and a board member of Genemo Inc.

Received: February 12, 2018

Revised: August 1, 2018

Accepted: August 10, 2018

Published: September 28, 2018

REFERENCES

- Altschuler, S.J., and Wu, L.F. (2010). Cellular heterogeneity: do differences make a difference? *Cell* **141**, 559–563.
- Artavanis-Tsakonas, S., Rand, M.D., and Lake, R.J. (1999). Notch signaling: cell fate control and signal integration in development. *Science* **284**, 770–776.
- Biase, F.H., Cao, X., and Zhong, S. (2014). Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing. *Genome Res.* **24**, 1787–1796.
- Burton, A., Muller, J., Tu, S., Padilla-Longoria, P., Guccione, E., and Torres-Padilla, M.E. (2013). Single-cell profiling of epigenetic modifiers identifies PRDM14 as an inducer of cell fate in the mammalian embryo. *Cell Rep.* **5**, 687–701.
- Cai, D., Cohen, K.B., Luo, T., Lichtman, J.W., and Sanes, J.R. (2013). Improved tools for the Brainbow toolbox. *Nat. Methods* **10**, 540–547.
- Calarco-Gillam, P. (1986). Cell-cell interactions in mammalian preimplantation development. *Dev. Biol. (N. Y.)* **108**, 329–371.
- Gardner, R.L. (2001). Specification of embryonic axes begins before cleavage in normal mouse development. *Development* **128**, 839–847.
- Goolam, M., Scialdone, A., Graham, S.J.L., Macaulay, I.C., Jedrusik, A., Hupalowska, A., Voet, T., Marioni, J.C., and Zernicka-Goetz, M. (2016). Heterogeneity in Oct4 and Sox2 targets biases cell fate in 4-cell mouse embryos. *Cell* **165**, 61–74.
- Gully, C.P., Velazquez-Torres, G., Shin, J.H., Fuentes-Mattei, E., Wang, E., Carlock, C., Chen, J., Rothenberg, D., Adams, H.P., Choi, H.H., et al. (2012). Aurora B kinase phosphorylates and instigates degradation of p53. *Proc. Natl. Acad. Sci. USA* **109**, E1513–E1522.
- Han, G., Li, A.G., Liang, Y.Y., Owens, P., He, W., Lu, S., Yoshimatsu, Y., Wang, D., Ten Dijke, P., Lin, X., and Wang, X.J. (2006). Smad7-induced beta-catenin degradation alters epidermal appendage development. *Dev. Cell* **11**, 301–312.
- He, Q., Kim, H., Huang, R., Lu, W., Tang, M., Shi, F., Yang, D., Zhang, X., Huang, J., Liu, D., and Songyang, Z. (2015). The Daxx/atrx complex protects tandem repetitive elements during DNA hypomethylation by promoting H3K9trimethylation. *Cell Stem Cell* **17**, 273–286.
- Hiragi, T., Louvet-Vallee, S., Solter, D., and Maro, B. (2006). Embryology: does prepatterning occur in the mouse egg? *Nature* **442**, E3–E4, discussion E4.

- Hiiragi, T., and Solter, D. (2004). First cleavage plane of the mouse egg is not predetermined but defined by the topology of the two apposing pronuclei. *Nature* 430, 360–364.
- Hiiragi, T., and Solter, D. (2006). Fatal flaws in the case for pre patterning in the mouse egg. *Reprod. Biomed. Online* 12, 150–152.
- Kalhor, R., Mali, P., and Church, G.M. (2017). Rapidly evolving homing CRISPR barcodes. *Nat. Methods* 14, 195–200.
- Kawamura, Y.I., Kawashima, R., Fukunaga, R., Hirai, K., Toyama-Sorimachi, N., Tokuhara, M., Shimizu, T., and Dohi, T. (2005). Introduction of Sd(a) carbohydrate antigen in gastrointestinal cancer cells eliminates selectin ligands and inhibits metastasis. *Cancer Res.* 65, 6220–6227.
- Kidder, G.M., and Pedersen, R.A. (1982). Turnover of embryonic messenger RNA in preimplantation mouse embryos. *J. Embryol. Exp. Morphol.* 67, 37–49.
- Leung, C.Y., and Zernicka-Goetz, M. (2015). Mapping the journey from totipotency to lineage specification in the mouse embryo. *Curr. Opin. Genet. Dev.* 34, 71–76.
- Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H., et al. (2016). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* 537, 558–562.
- Livet, J., Weissman, T.A., Kang, H., Draft, R.W., Lu, J., Bennis, R.A., Sanes, J.R., and Lichtman, J.W. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* 450, 56–62.
- McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., and Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* 353, aaf7907.
- Motosugi, N., Bauer, T., Polanski, Z., Solter, D., and Hiiragi, T. (2005). Polarity of the mouse embryo is established at blastocyst and is not prepatterned. *Genes Dev.* 19, 1081–1092.
- Pei, W., Feyerabend, T.B., Rossler, J., Wang, X., Postrach, D., Busch, K., Rode, I., Klapproth, K., Dietlein, N., Quedenau, C., et al. (2017). Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* 548, 456–460.
- Perli, S.D., Cui, C.H., and Lu, T.K. (2016). Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* 353, <https://doi.org/10.1126/science.aag0511>.
- Petit, F.M., Serres, C., Bourgeon, F., Pineau, C., and Auer, J. (2013). Identification of sperm head proteins involved in zona pellucida binding. *Hum. Reprod.* 28, 852–865.
- Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using smart-seq2. *Nat. Protoc.* 9, 171–181.
- Piotrowska-Nitsche, K., Perea-Gomez, A., Haraguchi, S., and Zernicka-Goetz, M. (2005). Four-cell stage mouse blastomeres have different developmental properties. *Development* 132, 479–490.
- Piotrowska, K., Wianny, F., Pedersen, R.A., and Zernicka-Goetz, M. (2001). Blastomeres arising from the first cleavage division have distinguishable fates in normal mouse development. *Development* 128, 3739–3748.
- Plachta, N., Bollenbach, T., Pease, S., Fraser, S.E., and Pantazis, P. (2011). Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. *Nat. Cell Biol.* 13, 117–123.
- Plusa, B., Hadjantonakis, A.K., Gray, D., Piotrowska-Nitsche, K., Jedrusik, A., Papaioannou, V.E., Glover, D.M., and Zernicka-Goetz, M. (2005). The first cleavage of the mouse zygote predicts the blastocyst axis. *Nature* 434, 391–395.
- Raj, B., Wagner, D.E., McKenna, A., Pandey, S., Klein, A.M., Shendure, J., Gagnon, J.A., and Schier, A.F. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* 36, 442–450.
- Rodriguez-Terrones, D., Gaume, X., Ishiuchi, T., Weiss, A., Kopp, A., Kruse, K., Penning, A., Vaquerizas, J.M., Brino, L., and Torres-Padilla, M.E. (2018). A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat. Genet.* 50, 106–119.
- Rodriguez-Terrones, D., and Torres-Padilla, M.E. (2018). Nimble and ready to mingle: transposon outbursts of early development. *Trends Genet.*
- Sadic, D., Schmidt, K., Groh, S., Kondofersky, I., Ellwart, J., Fuchs, C., Theis, F.J., and Schotta, G. (2015). Atrx promotes heterochromatin formation at retrotransposons. *EMBO Rep.* 16, 836–850.
- Schuster, A., Tang, C., Xie, Y., Ortogero, N., Yuan, S., and Yan, W. (2016). SpermBase: a database for sperm-borne RNA contents. *Biol. Reprod.* 95, 99.
- Shi, J., Chen, Q., Li, X., Zheng, X., Zhang, Y., Qiao, J., Tang, F., Tao, Y., Zhou, Q., and Duan, E. (2015). Dynamic transcriptional symmetry-breaking in pre-implantation mammalian embryo development revealed by single-cell RNA-seq. *Development* 142, 3468–3477.
- Sozen, B., Can, A., and Demir, N. (2014). Cell fate regulation during preimplantation development: a view of adhesion-linked molecular interactions. *Dev. Biol.* 395, 73–83.
- Tabansky, I., Lenarcic, A., Draft, R.W., Loulier, K., Keskin, D.B., Rosains, J., Rivera-Feliciano, J., Lichtman, J.W., Livet, J., Stern, J.N., et al. (2013). Developmental bias in cleavage-stage mouse blastomeres. *Curr. Biol.* 23, 21–31.
- Tago, K., Nakamura, T., Nishita, M., Hyodo, J., Nagai, S., Murata, Y., Adachi, S., Ohwada, S., Morishita, Y., Shibuya, H., and Akiyama, T. (2000). Inhibition of Wnt signaling by ICAT, a novel beta-catenin-interacting protein. *Genes. Dev.* 14, 1741–1749.
- Takaoka, K., and Hamada, H. (2014). Origin of cellular asymmetries in the pre-implantation mouse embryo: a hypothesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369.
- Torres-Padilla, M.E., Parfitt, D.E., Kouzarides, T., and Zernicka-Goetz, M. (2007). Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* 445, 214–218.
- Tsirigos, A., and Rigoutsos, I. (2009). Alu and b1 repeats have been selectively retained in the upstream and intronic regions of genes of specific functional classes. *PLoS Comput. Biol.* 5, e1000610.
- Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439, 811–816.
- Vogel, G. (2005). Embryology. Embryologists polarized over early cell fate determination. *Science* 308, 782–783.
- Wennekamp, S., and Hiiragi, T. (2012). Stochastic processes in the development of pluripotency in vivo. *Biotechnol. J.* 7, 737–744.
- Wennekamp, S., Mesecke, S., Nedelec, F., and Hiiragi, T. (2013). A self-organization framework for symmetry breaking in the mammalian embryo. *Nat. Rev. Mol. Cell Biol.* 14, 452–459.
- Wolpert, L. (2015). *Principles of Development*, Oxford, United Kingdom (Oxford University Press).
- Xie, D., Chen, C.C., Ptaszek, L.M., Xiao, S., Cao, X., Fang, F., Ng, H.H., Lewin, H.A., Cowan, C., and Zhong, S. (2010). Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res.* 20, 804–815.
- Zernicka-Goetz, M. (2002). Patterning of the embryo: the first spatial decisions in the life of a mouse. *Development* 129, 815–829.
- Zernicka-Goetz, M., Morris, S.A., and Bruce, A.W. (2009). Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat. Rev. Genet.* 10, 467–477.

ISCI, Volume 7

Supplemental Information

Rainbow-Seq: Combining Cell Lineage

Tracing with Single-Cell RNA Sequencing

in Preimplantation Embryos

Fernando H. Biase, Qiuyang Wu, Riccardo Calandrelli, Marcelo Rivas-Astroza, Shuigeng Zhou, Zhen Chen, and Sheng Zhong

Transparent Methods

Generation of rainbow mouse embryos

All the procedures for animal handling and breeding were approved by Institutional Animal Care and Use Committees of University of California San Diego. For the generation of rainbow embryos, we crossed two transgenic strains obtained from Jackson Laboratory. The strain STOCK Gt(ROSA)26Sortm1(CAG-Brainbow2.1)Cle/J (stock number 013731, RRID:SCR_004894) is homozygous for one copy of the brainbow construct v2.1, and the strain STOCK Tg(CAG-cre/Esr1*)5Amc/J (IMSR Cat# JAX:004453, RRID:IMSR_JAX:004453) is homozygous with one copy of the gene coding CRE protein. We crossed mice from each of the strains (1:1) to generate hybrid embryos containing one copy of the brainbow construct v2.1 and one copy of the Cre gene.

Adult female mice were estrous-synchronized and super-ovulated by two consecutive intra-peritoneal injections of hormones (to 2.5 IU of hCG and 5IU of PMSG, PG600, Intervet). The injections were administered 48 hours apart, and following the second injection, females and males were mated (1:1). Two-cell embryos were collected 46 hours post second hormonal injection (hpi) and incubated (37°C, 5%CO₂, 100% humidity) for two hours with 4-Hydroxytamoxifen (100µm, Sigma) in KSOM media (Millipore).

Embryo imaging

Tamoxifen-treated embryos were cultured *in vitro* (37°C, 5%CO₂, 100% humidity) until reaching 4-cell, 8-cell or blastocyst stages (~ 54, 70 and 90hpi, respectively), briefly treated with NucBlue™ Live Cell Stain (Invitrogen) in KSOM media and taken to wild-field inverted fluorescent microscope (Olympus). Embryos were imaged in KSOM media, with filter cubes for DAPI (Ex: 350/50x, Em: 460/50m), GFP (Ex: 480/40x, Em: 535/50m), RFP (Ex: 545/25x, Em: 605/70m) and differential interference contrast, and a 40x dry objective (NA=0.95, WD=0.18mm). Images were captured with an ORCA R2 monochromatic CCD camera (Hamamatsu) at 2µm spacing intervals in the z-axis, and rendered in MetaMorph software (Molecular Devices). Images were further evaluated in ImageJ software (Schneider et al., 2012).

Immunocytochemistry assay

Two-cell embryos were fixed in methanol at -20°C overnight, washed in PBS (supplemented with 0.1% BSA), and incubated for 5 min in permeabilization solution (PBS, 0.1% BSA, 0.1%

Triton X-100). Embryos were then incubated for 1 h in blocking solution (PBS, 0.1% BSA, 10% bovine fetal serum), following incubation with Monoclonal Anti-Cre antibody (Sigma-Aldrich Cat# C7988, RRID:AB_439697, 1:100 dilution in blocking solution) for 2 h at room temperature. After three PBS (0.1% BSA) washes, indirect detection was possible by immersing the embryos in blocking solution with Goat Anti-Mouse IgG1 (y1) Antibody, Alexa Fluor 594 Conjugated (Innovative Research Cat# 322588, RRID:AB_1502303, 1:500 dilution). Following three washes with PBS (0.1% BSA), embryos were immersed in blocking solution containing Mouse anti-Tubulin-Alexa-488 (Innovative Research Cat# 322588, RRID:AB_1502303, 1:25 dilution) for 1 h. Embryos were washed in PBS (0.1% BSA) supplemented with NucBlue Fixed Cell Stain ReadyProbes (Invitrogen), and preserved with Prolong Gold (Invitrogen). Embryos were scanned with a Fluoview FV1000 confocal microscope (Olympus), with the aid of a 40x oil objective.

Assessment of color-coded distribution of cells in blastocysts

Blastocyst images were visually inspected by one investigator. Cells were identified with the by their nucleus staining with DAPI. The cells were classified according to two criteria: First, cells were assigned to the embryonic (em) or abembryonic (abe) pole of the embryo. Second, the cells were color coded as red, green or no-color.

Single-cell RNA-sequencing

Four- (n=9) and 8-cell (n=4) embryos imaged and having two or more cells expressing at least one of the fluorescent proteins (RFP or GFP) were subjected to single-cell RNA-sequencing. Embryos were briefly treated with Tyrode's acid solution (Sigma) for removal of zona pelucida. The embryos were washed in PBS and incubated in Trypsin solution (Invitrogen) with RNase inhibitor (1IU/ μ l, Clontech) and BSA (50 μ g/ μ l) for the separation of the blastomeres. The blastomeres were separated by gentle pipetting, washed in PBS supplemented with RNase inhibitor and BSA, and snap frozen in individual micro-centrifuge tubes.

We amplified the cDNA of individual blastomeres according to the SMART-seq2 protocol (Picelli et al., 2014), followed by library preparation using Nextera XT DNA Sample Prep Kit (Illumina). The yields of cDNA and library amplification were assessed by Qubit (Invitrogen), and the quality of the distribution was assessed by Bioanalyzer (Agilent). Libraries were sequenced to produce pair-end reads of 100 nucleotides each in a HiSeq 2500 sequencer (Illumina).

Alignment of sequencing reads to fluorescent protein genes in brainbow construct v2.1

Sequencing reads from each cell were aligned to DNA sequences of the four fluorescent protein genes present on brainbow construct v2.1 (Addgene plasmid repository 18723) using STAR aligner. Reads that aligned exclusively to one of the sequences were used for counting. Cells were assigned a color coding (RFP, YFP, CFP or GFP) according to the greatest number of reads matching to one of the fluorescent proteins, or NA when no read aligned to any of the DNA sequences.

Alignment to the reference genome

STAR (STAR_2.5.1b, default parameters) was used to align single-cell sequencing reads to the reference genome (mm10, UCSC). Only uniquely aligned reads were used for the following analyses. Annotation files for Refseq genes and RepeatMasker defined repeats were downloaded from UCSC (February 2015), which contained a total of 24,225 genes and 5,138,231 repeats, respectively. These annotation files and the uniquely mapped reads in each cell were given as input to HTSeq-count (version 0.9.1) for counting of reads associated with each gene or repeat.

Principle component analysis

Principle component analysis was carried out based on all the genes that exhibited FPKM>1 in at least one blastomere.

Analysis of variance (ANOVA) analysis

A generalized linear model was fit to every gene as follows:

$$Y_{ijk}^g = u^g + \alpha_i^g + \beta_{ij}^g + \varepsilon_{ijk}^g$$

where Y_{ijk}^g is the observed expression level ($\log_2(\text{FPKM}+1)$) of gene g in embryo i , lineage j ($j = 0,1$), cell k ; u^g is the average expression level of gene g in all cells; α_i^g is the embryo-effect to average expression; β_{ij}^g is the difference of expression levels between the two cell division lineages; ε_{ijk}^g is the error term. R package {stats} was used to implement this linear model and test for every gene the null hypothesis that no reproducible expression differences between the two division, namely $\beta_{ij}^g = 0$. To account for multiple hypothesis testing, R package {qvalue} was used to compute q-values from ANOVA derived p-values.

ANOVA analysis for TRENIs was carried out in the same way as for genes, except that index g in the above generalized linear model was used as the index for TRENIs.

Comparison of lineage variation and embryo variation

For every gene (g) we calculated the ratio (r^g) of lineage sum of squares to combined sum of squares of embryo and lineage as

$$r^g = SS_{lineage}^g / (SS_{embryo}^g + SS_{lineage}^g).$$

The random variable (R) of ratio of lineage sum of squares and combined sum of squares of embryo and lineage was defined as

$$R = SS_{lineage} / (SS_{embryo} + SS_{lineage}).$$

Identification of novel transcript isoforms

Novel transcript isoforms were identified by Cufflinks (v2.2.1 with -g parameter) (Trapnell et al., 2012) with all Rainbow-seq data. Repeat sequences that overlapped with any Refseq annotated exons were removed from further analysis. TRENIs were identified by intersecting the remaining repeats with novel transcript isoforms using bedtools (2.27.0) (Quinlan and Hall, 2010).

Calculation of junction ratio

For each repeat sequence (indexed by i), HTSeq-count (version 0.9.1) was used to count the number of reads mapped to this repeat (T_i) and the number of reads spanning the junction of this repeat and nearby exons (S_i). Junction ratio (JR) for the i^{th} repeat was calculated as $JR_i = T_i/S_i$.

Analysis of association between repeat families and TRENIs

Repeat families and the correspondence of individual repeat sequences to repeat families were retrieved from RepeatMasker (February 2015). A contingency table was built for each repeat family and TRENIs as follows.

	Contained within TRENIs	Not contained in TRENIs
Is a member of this family	a	b
Not a member	c	d

In the above table a, b, c, d are counts of repeat sequences and they sum to the total number repeat sequences (5,138,231). Odds ratio representing the degree of association between this repeat family and TRENI was calculated as

$$OR = (a \times d)/(b \times c).$$

Ninety five percent confidence interval for odds ratio was calculated as

$$e^{\log(OR) \pm [1.96 \times SE(\log(OR))]},$$

where $SE(\log(OR)) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$.

Data and software availability

We deposited all sequencing data into GEO (GSE106287) for public access.

Reference

- PICELLI, S., FARIDANI, O. R., BJORKLUND, A. K., WINBERG, G., SAGASSER, S. & SANDBERG, R. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*, 9, 171-81.
- QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841-2.
- SCHNEIDER, C. A., RASBAND, W. S. & ELICEIRI, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*, 9, 671-5.
- TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 7, 562-78.

SUPPLEMENTARY FIGURES

Figure S1. Tamoxifen induced nuclear localization of Cre recombinase, related to Figure 2. (A) Overview of time and events. Tamoxifen is applied in a 2-hour window (yellow bar) at the early 2-cell stage. Microscopic analyses are carried out prior to (*), immediately after (**), and 1 hour after tamoxifen treatment (***). (B) Representative embryos (rows). Nuclear concentration of CRE was induced at the end of tamoxifen treatment (2nd, 3rd rows, two different embryos) and decreased to background level within 1 hour after removing tamoxifen (4th row).

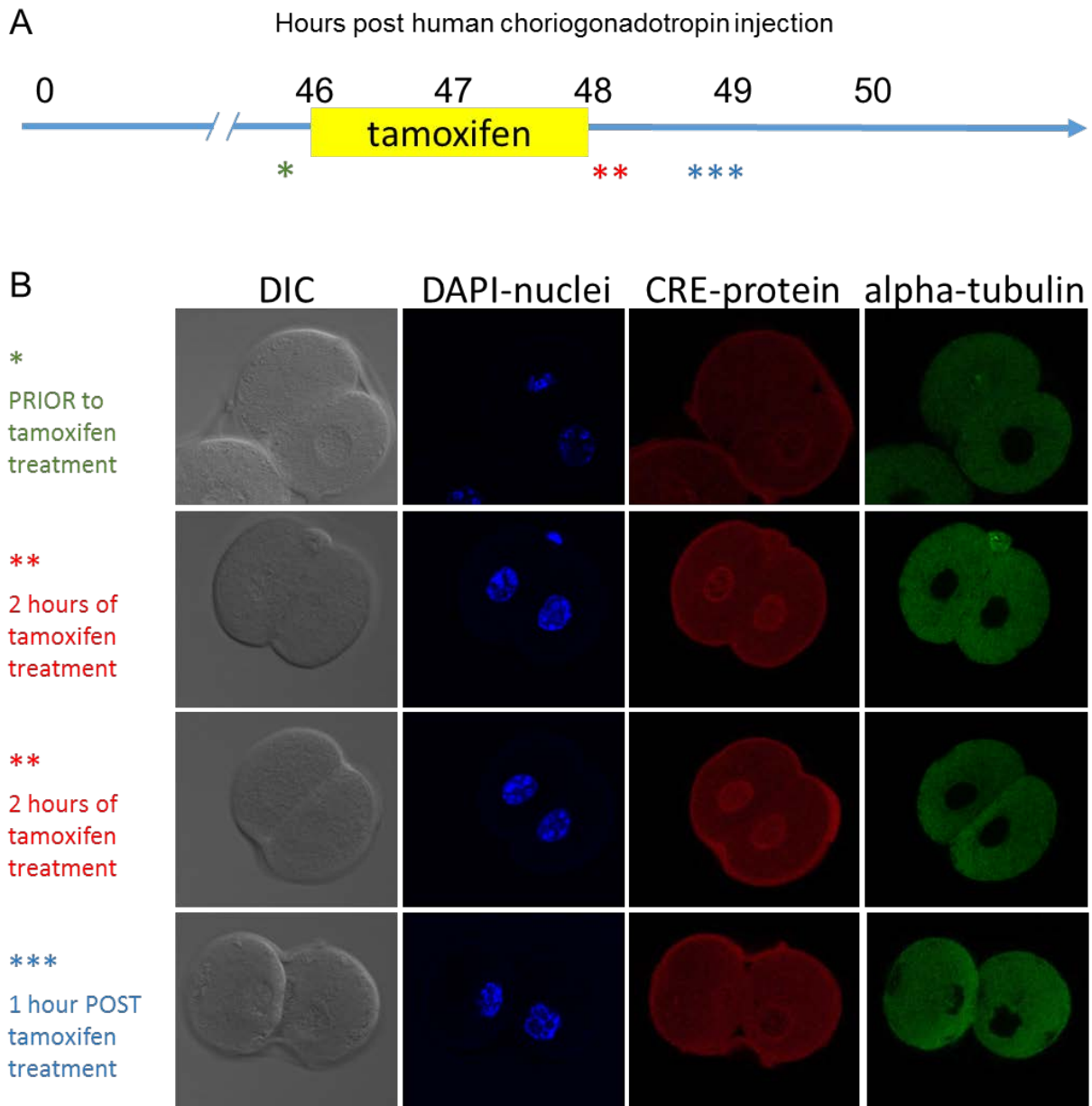


Figure S2. Overview of the single-cell sequencing data, related to Figure 3 and Figure 4. (A) Number of uniquely mapped read pairs (y axis) from each single cell (column). Each single cell is marked with an embryo number (Embryo), a cell number (Cell) and a binary lineage indicator (A or B). N: unresolved. (B) Principle component analysis of lineage resolved blastomeres at 4- (red) and 8-cell stages (green), using all genes with FPKM>1 in a least one blastomere.

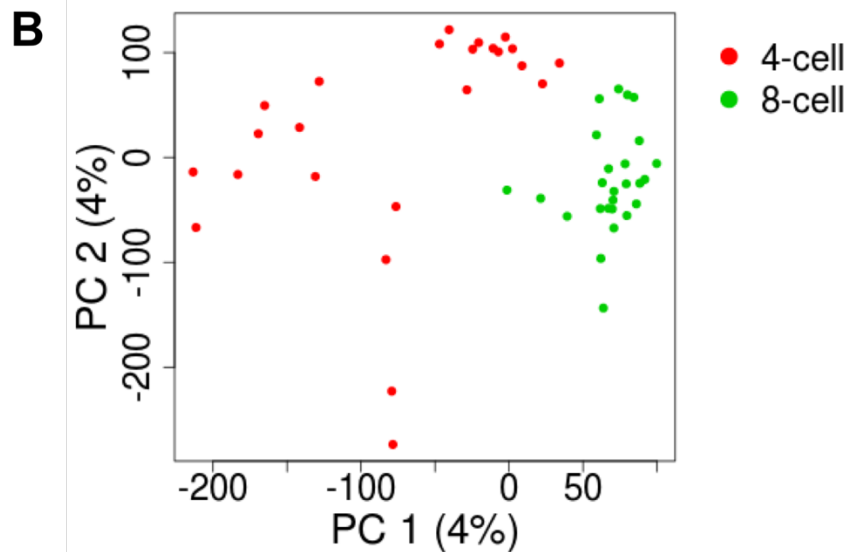
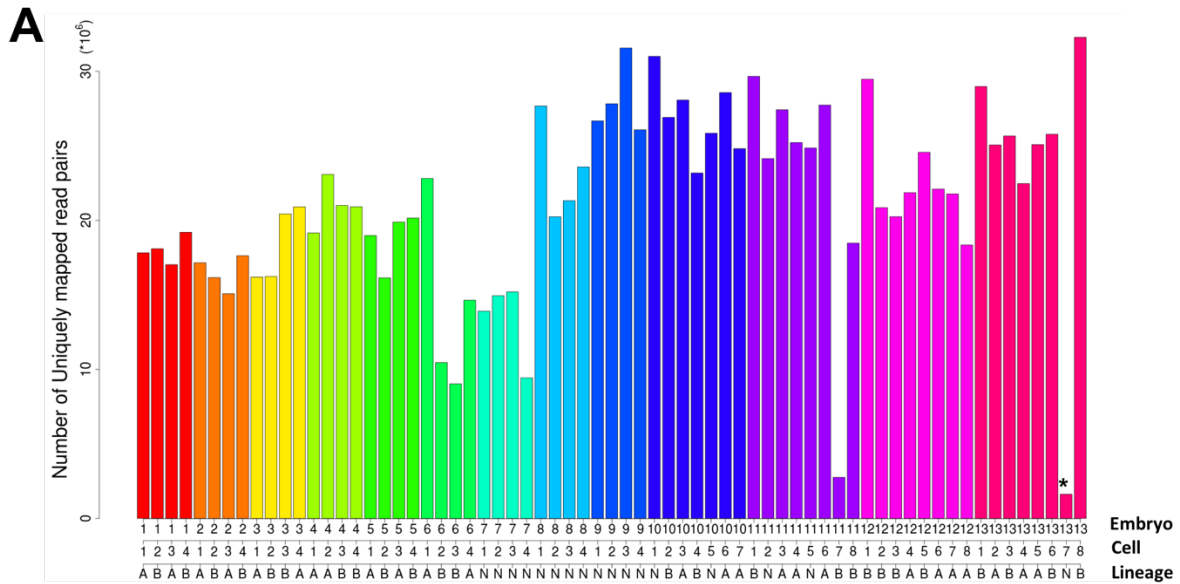


Figure S4. Tests of equivalent expression between lineages for all expressed repeat sequences, related to Figure 5. Repeats' q-values derived from real (red) and shuffled data (blue) exhibited similar histograms, in both 4- (A) and 8-cell (B) stages.

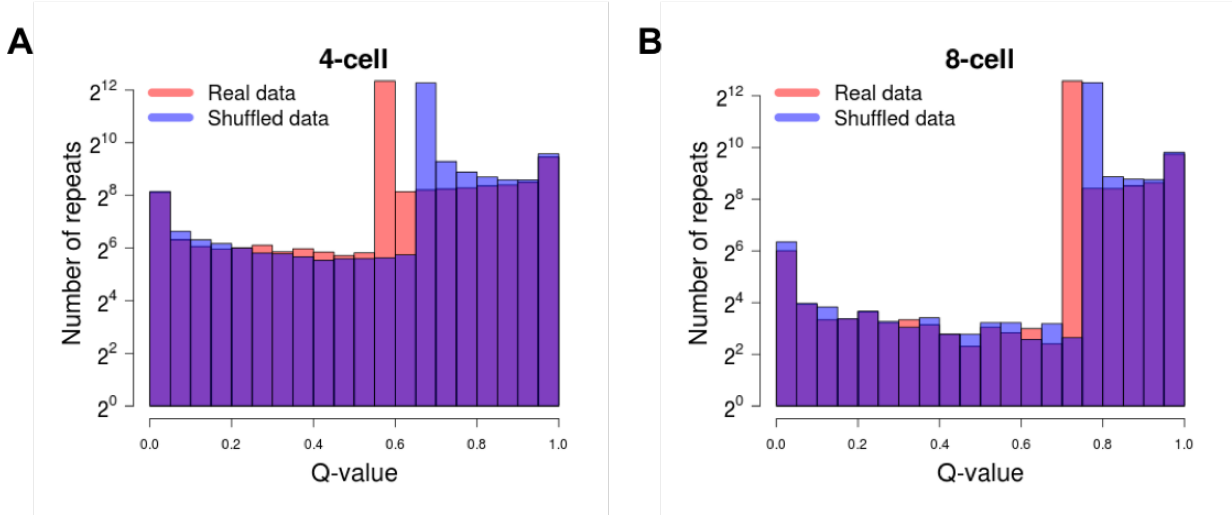
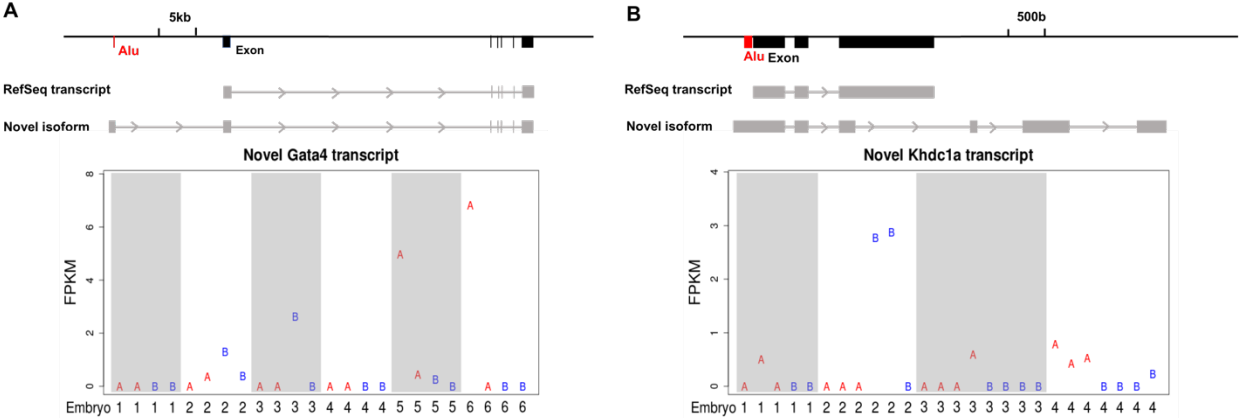


Figure S5. Sensitivity analysis of transposon related novel isoforms (TRENIs), related to Figure 5. Odds ratios between TRENIs and each transposon family (columns), calculated from only the repeat sequences with junction ratio (JR) > 0.9 (blue) were not very different from those calculated from all TRENIs embedded repeats (red). Odds ratio was not calculated when the TRENIs of a repeat family were fewer than 20. Vertical bars: 95% confidence intervals.



Figure S6. Novel transcript isoforms of *Gata4* and *Khdc1a* containing murine Alu in 5' UTRs, related to Figure 5 and Figure 6. (A) Expression levels of a novel transcript isoform of *Gata4* in 4-cell blastomeres. (B) A novel transcript isoform of *Khdc1a* in 8-cell blastomeres. FPKM (y axis) of every blastomere (column) in each embryo (marked by embryo number in columns). Shaded columns delineate different embryos. The blastomeres of the two lineages are marked with A (red) and B (blue), respectively.



SUPPLEMENTARY TABLES

Table S1. Summary of Rainbow-seq data from 4-cell stage blastomeres, related to Figure 1, Figure 3, and Figure 4. Each blastomere (row) is indexed by embryonic stage (Stage), embryo number (Embryo), cell number (Cell), cell division lineage (Lineage), the total number of Rainbow-seq reads (# of reads), number (# uniquely mapped) and the percentage (Genome mapping rate) of reads uniquely mapped to the mm10 genome, and the percentage of reads mapped to the mitochondria genome (Mitochondria mapping rate). NA: undetermined.

Stage	Embryo	Cell	Lineage	# of reads	# uniquely mapped	Genome mapping rate	Mitochondria mapping rate
4-cell	E1	B1	A	21314126	17819219	83.60%	3.66%
		B2	B	21985294	18088361	82.27%	4.98%
		B3	A	20235470	17028354	84.15%	1.47%
		B4	B	23513067	19197409	81.65%	6.23%
E2	E2	B1	A	20704467	17146739	82.82%	5.54%
		B2	B	20058467	16160003	80.56%	7.39%
		B3	A	18278090	15080241	82.50%	5.29%
		B4	B	21622583	17630799	81.54%	7.12%
E3	E3	B1	A	18775617	16185096	86.20%	2.57%
		B2	B	18793826	16228376	86.35%	3.96%
		B3	B	23865180	20424658	85.58%	3.30%
		B4	A	24696310	20903704	84.64%	4.72%
E4	E4	B1	A	25954803	19148778	73.78%	5.02%
		B2	A	33404660	23079578	69.09%	6.49%
		B3	B	29530032	21001228	71.12%	7.07%
		B4	B	31340190	20908747	66.72%	7.94%
E5	E5	B1	A	26738657	18974243	70.96%	5.28%
		B2	B	28170918	16139013	57.29%	30.15%
		B3	A	28801088	19877650	69.02%	4.79%
		B4	B	28915622	20154724	69.70%	21.79%
E6	E6	B1	A	29629536	22814400	77.00%	3.00%
		B2	B	12599757	10457443	83.00%	6.21%
		B3	B	10829517	9033260	83.41%	2.13%
		B4	A	17688808	14640436	82.77%	1.81%
E7	E7	B1	NA	16610220	13908974	83.74%	12.83%
		B2	NA	17463191	14944165	85.58%	5.43%
		B3	NA	18099348	15200717	83.98%	7.91%
		B4	NA	11109883	9425697	84.84%	6.59%
E8	E8	B1	NA	32556574	27666944	84.98%	3.47%
		B2	NA	23739102	20243893	85.28%	2.85%
		B3	NA	25140202	21316100	84.79%	3.13%
		B4	NA	28565034	23582751	82.56%	2.94%
E9	E9	B1	NA	32027811	26666074	83.26%	8.36%
		B2	NA	36935316	27818485	75.32%	7.19%
		B3	NA	44737967	31564011	70.55%	8.46%
		B4	NA	35022576	26073115	74.45%	13.18%

Table S2. Summary of Rainbow-seq data from 4-cell stage blastomeres, related to Figure 1, Figure 3, and Figure 4. Each blastomere (row) is indexed by embryonic stage (Stage), embryo number (Embryo), cell number (Cell), cell division lineage (Lineage), the total number of Rainbow-seq reads (# of reads), number (# uniquely mapped) and the percentage (Genome mapping rate) of reads uniquely mapped to the mm10 genome, and the percentage of reads mapped to the mitochondria genome (Mitochondria mapping rate). NA: undetermined.

Stage	Embryo	Cell	Lineage	# of reads	# uniquely mapped	Genome mapping rate	Mitochondria mapping rate
8-cell	E10	B1	NA	35510447	30994895	87.28%	2.55%
		B2	B	31310655	26908056	85.94%	4.20%
		B3	A	32364998	28069936	86.73%	3.32%
		B4	B	26879062	23168806	86.20%	3.03%
		B5	NA	30026238	25835885	86.04%	2.17%
		B6	A	33728624	28568517	84.70%	2.24%
		B7	A	29108703	24806968	85.22%	8.06%
		B8	B	38092133	29663146	77.87%	5.68%
	E11	B1	NA	30719208	24136791	78.57%	4.52%
		B2	A	35052929	27408245	78.19%	7.15%
		B3	A	34218985	25218809	73.70%	7.74%
		B4	NA	32137812	24846522	77.31%	7.91%
		B5	A	33420557	27727130	82.96%	7.07%
		B6	B	3650871	2758113	75.55%	11.98%
		B7	B	22651287	18476451	81.57%	10.44%
		B8	B	34015604	29465623	86.62%	3.54%
	E12	B1	B	24573413	20847753	84.84%	4.69%
		B2	B	23764288	20252119	85.22%	4.72%
		B3	A	26254436	21857322	83.25%	4.68%
		B4	B	29526449	24568457	83.21%	4.23%
		B5	A	26966760	22102352	81.96%	3.96%
		B6	A	28616287	21771577	76.08%	5.21%
		B7	A	22578208	18345514	81.25%	4.97%
		B8	A	34150397	28982368	84.87%	3.90%
	E13	B1	A	29319315	25052456	85.45%	3.33%
		B2	B	31016258	25656830	82.72%	4.22%
		B3	A	26313847	22461091	85.36%	3.46%
		B4	A	29808796	25077571	84.13%	2.66%
		B5	B	31502252	25769136	81.80%	3.55%
		B6	NA	2967324	1618930	54.56%	0.01%
		B7	B	41717993	32283499	77.39%	5.65%
		B8	B				