The role of Cdx2 as a lineage specific transcriptional repressor for pluripotent network during the first developmental cell lineage segregation

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation
Huang, D., G. Guo, P. Yuan, A. Ralston, L. Sun, M. Huss, T. Mistri, et al. 2017. "The role of Cdx2 as a lineage specific transcriptional repressor for pluripotent network during the first developmental cell lineage segregation." Scientific Reports 7 (1): 17156. doi:10.1038/s41598-017-16009-w. http://dx.doi.org/10.1038/s41598-017-16009-w.

Published Version
doi:10.1038/s41598-017-16009-w

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:34652015

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
The role of Cdx2 as a lineage specific transcriptional repressor for pluripotent network during the first developmental cell lineage segregation

Daosheng Huang1,10, Guoji Guo1,10, Ping Yuan2, Amy Ralston3, Lingang Sun1,10, Mikael Huss4, Tapan Mistri5, Luca Pinello6, Huck Hui Ng7, Guocheng Yuan6, Junfeng Ji1,10, Janet Rossant8, Paul Robson9 & Xiaoping Han1,10

The first cellular differentiation event in mouse development leads to the formation of the blastocyst consisting of the inner cell mass (ICM) and trophectoderm (TE). The transcription factor CDX2 is required for proper TE specification, where it promotes expression of TE genes, and represses expression of Pou5f1 (OCT4). However its downstream network in the developing embryo is not fully characterized. Here, we performed high-throughput single embryo qPCR analysis in Cdx2 null embryos to identify CDX2-regulated targets in vivo. To identify genes likely to be regulated by CDX2 directly, we performed CDX2 ChIP-Seq on trophoblast stem (TS) cells. In addition, we examined the dynamics of gene expression changes using inducible CDX2 embryonic stem (ES) cells, so that we could predict which CDX2-bound genes are activated or repressed by CDX2 binding. By integrating these data with observations of chromatin modifications, we identify putative novel regulatory elements that repress gene expression in a lineage-specific manner. Interestingly, we found CDX2 binding sites within regulatory elements of key pluripotent genes such as Pou5f1 and Nanog, pointing to the existence of a novel mechanism by which CDX2 maintains repression of OCT4 in trophoblast. Our study proposes a general mechanism in regulating lineage segregation during mammalian development.

Totipotency, or the ability to form both embryonic and extra-embryonic tissues, is lost upon the first lineage segregation during mouse early embryonic development. As the blastocyst forms, cells that will become trophoblast are separated from the pluripotent cells of the embryo1,2. At this crossroad, cells decide whether to establish or repress pluripotency, in establishing the inner cell mass (ICM) and trophectoderm (TE), respectively. Embryonic stem (ES) cells originate from the inner cell mass3,4, while trophoblast stem (TS) cells are derived from trophectoderm5. These stem cell lines provide expandable yet pure cell populations for genome-wide analyses of gene

1Center for Stem Cell and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou, 310058, China. 2The Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Disease, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, 510655, China. 3Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, 48824, USA. 4Science for Life Laboratory, School of Biotechnology, Royal Institute of Technology (KTH), SE-106 91, Stockholm, Sweden. 5Lovely Professional University, Jalandhar - Delhi G.T. Road, Phagwara, Punjab, 144411, India. 6Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, MA, 02215, USA. 7Stem Cell and Regenerative Biology, Genome Institute of Singapore, 60 Biopolis Street, Singapore, 138672, Singapore. 8Program in Developmental and Stem Cell Biology, Hospital for Sick Children Research Institute, Toronto, ON M5G 0A4, Canada. 9The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, 06032, USA. 10Stem Cell Institute, Zhejiang University, Hangzhou, 310058, China. Daosheng Huang, Guoji Guo, Ping Yuan and Amy Ralston contributed equally to this work. Correspondence and requests for materials should be addressed to J.R. (email: janet.rossant@sickkids.ca) or P.R. (email: paul.robson@jax.org) or X.H. (email: xhan@zju.edu.cn)

Received: 11 May 2017
Accepted: 6 November 2017
Published: xx xx xxxx
regulatory mechanisms. Studies with these stem cell lines have illuminated our understanding the genetic networks that regulate the segregation of the first two lineages in the mouse pre-implantation embryo.

The transcription factor CDX2 acts early during the blastocyst formation, playing an instructive role in the formation of trophoblast. Loss of Cdx2 in the embryos leads to ectopic expression of pluripotency markers in the TE, and over-expression of Cdx2 in ES cells is sufficient to direct the formation of TS cells. How CDX2 achieves its role via transcriptional regulation is therefore a central question. Nishiya et al. characterized the genome-wide early responsive CDX2 targets when Cdx2 was overexpressed in ES cells, and could not demonstrate direct binding of CDX2 to the regulatory regions of pluripotency genes. Rather, CDX2 interfered with a pro-pluripotency transcriptional complex during the early stages of Cdx2 over-expression. However, the long-term activities of CDX2 in maintaining cell fate, in stem cell lines and in vivo, have not been extensively characterized.

Given the importance of understanding CDX2 targets in a biologically relevant setting, direct examination of CDX2 function in the embryonic TE tissues is needed. We first analyzed global gene expression in the TE of wild type embryos from single cell RNA-seq. We then developed micro-genomic methodologies to profile gene expression in individual Cdx2 knockout blastocysts. We performed CDX2 ChIP-seq in TS cells, which identified CDX2 targets relevant to TE biology. Finally, we defined putative lineage-specific silencer regulatory regions that possess unique chromatin features, on a genome-wide level. Ultimately, we have integrated these data to present a holistic model of how CDX2 regulates the ICM/TE lineage segregation during mouse embryo development.

Results

Comparison of in vitro trophoblast stem cell lines and in vivo trophobectoderm progenitors. TS cells derived from blastocysts or Cdx2-overexpressing ES cells provide a useful platform to investigate gene regulatory networks of early cell commitment in vitro. However, the properties of the two cell line systems are not exactly the same and both are likely to be different from the embryonic trophobectoderm. To test this hypothesis, we analyzed global gene expression pattern from the different cell sources.

We utilized the inducible Cdx2 over-expression ES cell system as previous reports to measure transcriptional changes upon single gene perturbation. Time-course microarray analysis was performed on three different inducible clones at day 0, day 0.25, day 1, day 2 as well as day 6. Changes in individual gene expression during the time-course are shown in Fig. 1a. CDX2-induced gene activation or repression may start as early as 6 hours after over-expression. On day 6, the TE transcriptional program (including Cdx2, Tgif2c and Id2) is fully activated, while the ES transcriptional program (including Pou5f1, Sox2 and Nanog) is completely repressed. Notably, a list of genes including Hoxa9, Hoxa10, Hoxb6, Foxh1, Phy19, Nkx1-2 and Sox7 is transiently induced during the early time points, but eventually repressed on day 6. As the chromatin state of ES cells is relatively open, forced expression of Cdx2 may activate targets that are irrelevant to trophoblast development.

In order to understand the whole-genome gene expression profiles of in vivo TE, we analyzed recently published mouse embryo single cell RNA-Seq data. We analyzed 61 single cells from 64-cell stage mouse embryo, and defined 32 ICM cells and 29 TE cells, as shown in t-SNE plot (Fig. 1b). A comparison of individual gene FPKM value between the two cell type reveals the TE/ICM differential expressions (Fig. 1b, and Supplementary Table S1). We sorted genes by their expression fold difference between whole blastocysts and ICMs; and then defined TE enriched genes based on methods exploited in Seurat.

Identification of CDX2 functional targets in Cdx2 knockout embryos. We next used Cdx2 knockout embryos to identify genes whose expression level depends on Cdx2. Cdx2 is first activated at E2.5 at 8-cell stage. Cdx2 null embryos die at around E4.5: they still form a blastocoeI but fail to maintain blastocyst integrity. In order to characterize Cdx2 functional targets in vivo, we applied high throughput micro-fluidic qPCR gene expression profiling on E3.75 blastocysts from Cdx2+/− crosses. In total, 27 blastocysts were assayed against 176 genes selected by the availability of Taqman probes, as well as lineage specific markers defined in our previous work. Out of 27 embryos do not have any detectable Cdx2 mRNA (Fig. 2a, and Supplementary Table S2). These embryos also have significantly higher Neo expression, which was used to replace the Cdx2 alleles. These embryos were designated as Cdx2−/− embryos. In addition, 4 embryos had negligible Neo levels: these were designated as wild-type embryos. As shown in Fig. 2a, expression of Pou5f1, Nanog and Sox2 in the Cdx2 null embryos...
were up-regulated compared to the heterozygous and wild-type, suggesting that CDX2 is required to repress ES pluripotent markers during early embryogenesis. The expression of the TE marker Gata3 is unperturbed, consistent with a previous study demonstrating that Cdx2 regulates trophoblast development in parallel to Gata3.[13]. In total, more than 50 percent of all assayed genes show reproducibly altered gene expression levels in the Cdx2 null blastocysts. The hierarchical clustering of gene expression profiles of the 27 blastocysts clearly demonstrated that the 6 Cdx2 knockout blastocysts exhibited distinct global gene expression patterns (Fig. 2b).

In order to reveal the hierarchy of CDX2-regulated gene expression, we generated single cell expression profiling data from the wild-type E3.75 blastocyst using our previous methods[16] and plotted genes according to their correlation with Cdx2 in both single cell data and the knockout blastocyst data (Fig. 2c). Based on the 2D plot, there are mainly three groups of genes that are of interest. The first group of genes are highly correlated with Cdx2 expression and thus can be confirmed as being positively regulated by CDX2. The second group of genes are independent of Cdx2 activation, however they are specific to the CDX2-positive TE, which indicates they function in parallel with Cdx2. Finally, the third cluster genes include most of ICM pluripotent markers that are negatively regulated by Cdx2. Our results clearly demonstrate that Cdx2 activates the TE transcriptional program and represses the pluripotent network during blastocyst formation.

**Whole-genome ChIP-Seq analysis reveals diverse targets of CDX2 in TS cells.** To characterize genome-wide direct targets of CDX2, we performed chromatin immuno-precipitation sequencing (ChIP-Seq) experiment within TS cells using a highly specific CDX2 antibody (Supplementary Fig. S1a). Enriched CDX2 binding DNA fragments were analyzed by high-throughput sequencing. Using model-based ChIP-seq analysis[21], we found a total of 16736 confident peaks (Supplementary Table S3). We performed de novo motif discovery with...
CisFinder. The main consensus-binding motif cluster turns out to be a known CDX motif (Fig. 3a). This motif is overrepresented in ChIP-enriched regions, as we looked at motif counts across 4 kb windows centered on the CDX2 binding sites (Supplementary Fig. S1b).

Traditional ChIP-Seq analysis usually associates a TF binding peak with a gene based on the distance between the peak and the transcriptional start site of the gene. Ouyang et al. made a significant improvement by integrating the surrounding peak intensity and the proximity to genes to define the association strength between TF and individual genes. We used this method to calculate CDX2 ChIP-seq association score for each gene, ranked the genome according to the association score and defined top 3000 associated genes as CDX2 targets (Supplementary Table S1). To associate CDX2 bindings with its function in TS cells, we reanalyzed published...
gene expression data for TS cells as well as differentiated TS cells. We identified genes that are significantly up- or down-regulated during 6 days of TS cell differentiation based on fold change ranking (cut-off = 2, p-value < 0.05) (Supplementary Table S1). Cdx2 was dramatically down-regulated upon TS differentiation, together with other TS cell markers such as Eomes and Elf5, as previously shown. We overlapped our defined CDX2 targets with the significantly up or down-regulated genes among TS cell differentiation. Here, we define significantly up-regulated genes upon TS cell differentiation as TS Differentiation markers, and down-regulated genes as TS stem cell markers. Among the significantly down-regulated genes, top CDX2 binding associated genes include TS stem cell markers Elf5, Lin28, Cdx2, Id2 and Eomes (Fig. 3b). TS Differentiation markers such as Elf3, Dlx3 and Sox15 also have high CDX2 binding association. We sorted the genome according to expression fold difference between differentiated and undifferentiated TS cells. We then looked for changes in the average association score of CDX2.

Figure 3. Cdx2 ChIP-Seq analysis in TS cells reveals direct targets of Cdx2 (See also Fig. S1). (a) Cdx2 main binding motif clusters identified with CisFinder via 200 bp sequences centered at ChIP sites. (b) Venn diagram showing the overlap between Cdx2 target list and the significantly up/down regulated genes after 6 days of TS cell differentiation. Representative Cdx2 targets are listed. (c) Blue line: relationship between gene expression difference and TF ChIP-Seq association score. X-axis shows the gene rank after sorting the genome according to expression fold change between differentiated and undifferentiated TS cells (Kidder and Palmer, 2010). Y-axis shows the average Cdx2 binding association score from a sliding window of 500 genes. (d) Oct4 ChIP-Seq peaks (from ES cells) and Cdx2 ChIP-Seq peaks (from TS cells) in the Pou5f1 gene region viewed with USCS mouse mm8 browser. (e) Oct4 ChIP-Seq peaks (from ES cells) and Cdx2 ChIP-Seq peaks (from TS cells) in the Nanog gene region viewed with USCS mouse mm8 browser. (f) Analysis of Cdx2 ChIP-seq results from our TS cell system and the ES cell Cdx2 overexpression system (Nishiyama et al., 2012) reveals strikingly similar core Cdx2 binding motifs. (g) Although Cdx2 does not bind to Pou5f1, Sox2 and Nanog in the ES cell TE differentiation system (Nishiyama et al., 2012), we have observed significant repressive bindings of Cdx2 on pluripotent genes in the established TS cell system.
in a sliding 500 gene window. Interestingly, TS stem cell markers or potential CDX2 positively regulated targets have the highest association score (Fig. 3c). Genes whose expression did not change during differentiation have extremely low association scores. However, TS differentiation markers or potential CDX2 negatively regulated targets tend to have moderate CDX2 binding. The CDX2 binding association curve suggests that CDX2 is actively involved in both gene activation and repression within the TS cells.

Previous studies have revealed that CDX2 is also involved in ICM/TE lineage segregation by repressing ES core pluripotent gene Oct4 (Pou5f1)\(^{20,21}\). Early studies have employed the Cdx2-overexpressing ES cell system to demonstrate that CDX2 directly interacts with OCT4 to repress its transcriptional activity in ES cells. ChiP-qPCR assay showed that CDX2 prevented OCT4 protein from binding to an auto-regulatory element (ARE) of Pou5f1\(^{21}\), but CDX2 did not directly bind to Pou5f1 and Nanog regulatory elements during the initial phases of CDX2 overexpression in ES cells\(^22\). In our CDX2-ChiP-Seq experiments on TS cells, however, we found clear and significant CDX2 binding to cis-acting sequences associated with the core ES transcription factors: Pou5f1 and Nanog (Fig. 3d,e). CDX2 binds to the first intron of Pou5f1 with more than 40-fold enrichment, validated by ChiP-qPCR across the binding region (Supplementary Fig. S1c). In addition, it is noteworthy that CDX2 binds to Pou5f1 at the first intron instead of promoter or other cis-regulatory DNA elements of Pou5f1. The exact role of this intronic region in regulating Pou5f1 expression remains to be explored.

To identify more genes whose expression is regulated by CDX2 binding, we compared CDX2 ChiP-seq data from TS cells and the CDX2 overexpression ES system from published data\(^23\). Interestingly, although the core CDX2 binding motifs are remarkably the same (Fig. 3f), the actual targets vary in the two systems (Fig. 3g). Contrary to our CDX2-ChiP-Seq data in TS cells, CDX2 does not bind to the core pluripotent genes, Sox2, Pou5f1 and Nanog, during the initial step of ES cell differentiation upon CDX2 over-expression. However, binding was observed in association with several Klf genes. As our data show in Fig. 1d, the gene expression patterns and regulatory network in the CDX2 overexpressing ES cell system are not exactly identical with those in the TS cell system. Because of this, we used ES and TS cell comparisons to define CDX2 regulatory activity in trophoblast development in the rest of the studies.

**Cdx2 as a lineage-specific transcriptional repressor during ICM/TE segregation.** To compare CDX2 and OCT4 binding sites, we integrated OCT4-ChiP-seq data in ES cells\(^24\) with our TS data. By evaluating the genome association score, we defined the top 3000 OCT4 binding targets in ES cells (Supplementary Table S1). Remarkably, OCT4 and CDX2 share 711 downstream targets (Supplementary Fig. S2a). Binding site overlap studies also revealed 449 co-occupied loci. Many important pluripotent markers such as Nanog, Sox2, Klf4, Esrrb and Utf1 are within the overlapping targets list (Supplementary Fig. S2b). To associate CDX2 and OCT4 binding with function in TS cells and ES cells respectively, we overlapped the top 3000 CDX2 binding targets with TS-specific genes and ES-specific genes, and did the same with OCT4 binding targets. Venn diagrams reveal that both CDX2 and OCT4 not only bind to TS-expressed genes, but also to ES-expressed genes (Supplementary Fig. S2c,d). Notably, ES cell-specific genes Jarid2 and Sox2 are among the top 5 genes bound by both CDX2 and OCT4, suggesting that CDX2 and OCT4 play active, but opposite, roles in regulating Jarid2 and Sox2 expression. Furthermore, we sorted the genome according to the expression fold difference between TS and ES cells and looked for change in the average association score of CDX2 in a sliding 1000 gene window. Interestingly, gene activation or potential positive targets have the strongest binding association (Fig. 4a), whereas gene repression or potential negative targets tend to have moderate binding association. Similar patterns were also found for OCT4 (Fig. 4b). The significance of expression fold change between TS and ES cells is positively correlated with binding association score regardless of whether it is gene activation or repression.

Our ChiP-Seq data indicates that CDX2 extensively binds and represses other ES-specific genes in addition to Pou5f1 in TS cells. Therefore, how CDX2 binding leads to transcriptional repression instead of activation becomes a key question. Mapping DNase I hypersensitive sites (DHSs) by DNase-Seq has been a valuable tool for identifying different types of regulatory DNA elements, including promoters, enhancers, and silencers\(^25\). Here, we integrated published DNase-Seq data in TS and ES cells, respectively\(^26\). Results show that DHSs and CDX2 represesive binding sites co-occupy the first intron of Pou5f1 in TS cells (Fig. 4c). Similarly, the CDX2 binding sites at other ES pluripotent markers such as Nanog, Sox2 and Klf4, are also DNase I hypersensitive sites (Supplementary Fig. S3a). Previous study has shown that core transcription factors OCT4, SOX2, and NANOG share substantial target genes\(^22\) and therefore we further analyzed OCT4, SOX2, and NANOG ChiP-Seq, plus DNase-Seq in ES cells\(^24\). Interestingly, the three transcription factors co-occupy a Cdx2 intron with high DNase I hypersensitive signal (Fig. 4c). Similar binding patterns also exist in other TS-specific genes, including Id2, Tcfap2a, and Msx2 (Supplementary Fig. S3b). Notably, there are high level of H3K27me3 signals surrounding these binding sites, indicating that the regions are transcriptionally inactive\(^28\). Intriguingly, we found that DNase I hypersensitive sites do not exactly overlap with H3K27me3 peaks. As shown in Fig. S3b, DNase I hypersensitive site actually overlap with the H3k27me3 "valley" in Fgfr2 first intron. This is an interesting phenomenon, because transcription factor needs to bind to the DNase I hypersensitive site, and then recruit silencing complex to add H3K27me3 modification to the nearby DNA regions. Together, this type of binding site can be considered as lineage-specific silencers, based on their unique chromatin state and lineage-specific transcriptional repression features.

To characterize lineage-specific silencers on a genome-wide scale, we defined silencer candidates in TS cells based on possession of three properties: CDX2 binding, DNase I hypersensitivity, and enriched H3K27me3. This identified 1500 putative CDX2-regulated TS-specific silencers in TS cells (Fig. 4d, and Supplementary Table S4). Similarly, we defined silencer candidates in ES cells as sharing Oct4-sox2-Nanog (OSN) co-binding peak, DNase I hypersensitive sites, and enriched H3K27me3. This produced a list of 1223 putative OSN-regulated silencer regions in ES cells (Fig. 4d, and Supplementary Table S4). Candidate silencer elements were annotated to the nearest TSS and its associated gene. GO analysis revealed that top associated molecular function of silencer-related genes in both TS cells and ES cells includes "stem cell population maintenance" and "cell fate commitment", which
Figure 4. Cdx2 directly competes with Oct4 on genome-wide regulation of lineage segregation (See also Figs S2 and S3). (a) Relationship between gene expression difference of TS/ES and Cdx2-ChIP-Seq association score. X-axis shows the gene rank after sorting the genome according to expression fold change between TS and ES cells. Y-axis shows the average Cdx2 binding association score from a sliding window of 500 genes. (b) Relationship between gene expression difference of TS/ES and Oct4-ChIP-Seq association score. X-axis shows the gene rank after sorting the genome according to expression fold change between TS and ES cells (Kidder and Palmer, 2010). Y-axis shows the average Oct4 binding association score from a sliding window of 500 genes. (c) Cdx2 ChIP-Seq peaks, H3K27me3 Peaks, DNase Peaks from TS cells in the Pou5f1 gene region viewed with IGV; OSN ChIP-Seq peaks, H3K27me3 Peaks, DNase Peaks from ES cells in the Cdx2 gene region viewed with IGV. OSN: Oct4-Sox2-Nanog. (d) Venn diagram show silencer candidates in TS cell (left); Venn diagram show silencer candidates in ES cell (right). (e) GO analysis of silencer-related genes in TS cells; GO analysis of silencer-related genes in ES cells.
suggests silencer-associated genes function in transcriptional regulation (Fig. 4e, and Supplementary Table S4). Besides, the top associated biological process of silencer-related genes in both ES cells includes “embryo development” term and transcription binding-related terms (Fig. 4e, and Supplementary Table S4).

Discussion

In this study, we use single cell qPCR and RNA-seq on normal and Cdx2 knockout embryos to provide new insights into the likely downstream targets of Cdx2 function in the developing blastocyst. In Cdx2 knockout embryos, a large portion of ICM markers are up-regulated while TE markers are down-regulated, demonstrating that Cdx2 is a key regulator, playing a dual function in TE formation through both repressing pluripotency genes and activating TE genes.

To investigate the direct targets of CDX2 in the whole genome, we performed CDX2 ChIP-Seq in TS cells. It turns out that CDX2 binds to a wide range of targets, including both markers for TS self-renewal and differentiated state. Furthermore, we integrated CDX2 function with other TS specific transcription factors and compared our CDX2 target list with the published EOMES and TCFAP2C target lists in TS cell (Kidder and Palmer, 2010) (Supplementary Table S1). We showed that CDX2, EOMES and TCFAP2C co-occupy 346 genes (Supplementary Fig. S4a). On the top of this overlapping target list, there are many characterized TE lineage markers such as id2, Elf5 and Hand1. Interestingly, the top overlapping targets of CDX2, EOMES and TCFAP2C are TE-specific activation targets. In the ChIP-seq data from the established TS cultures, we also observed CDX2 binding to many important pluripotent genes such as Pou5f1, Nanog, Sox2 and Esr2, genes that are also upregulated in Cdx2 mutant blastocysts. This might suggest that direct repressive occupancy by CDX2 is involved in maintaining repression of pluripotent gene expression in the trophoblast lineage. By contrast, in the Cdx2-inducible overexpression ES system, it had previously been shown that Cdx2 does not directly bind to ES cell core regulatory genes. However, this system may be more representative of the early transient phases of lineage activation, rather than the stable regulatory system needed for TS cell lineage maintenance.

A subset of CDX2 binding sites in TS cells possess a unique chromatin state, with DNase I hypersensitivity, indicative of open chromatin, and enrichment of the H3K27me3 modification. These sites are enriched in associations with known pluripotency regulatory genes and are proposed to function as lineage specific silencers mediating transcriptional repression of pluripotent markers (Pou5f1, Sox2 and Nanog) to maintain TS cell identity. This is consistent with in vivo studies showing that Cdx2 knockout embryos fail to down-regulate Pou5f1 and Nanog in the trophectoderm. Strikingly, re-analyzing the published OSN ChIP-Seq data in ES cells reveals that the ES cell core transcription factors, OSN, co-occupy regulatory regions of TS-specific genes with DNase I hypersensitive feature and enriched H3K27me3 modification, similar to CDX2 targets in TS cells. Most work on ES cell gene regulatory networks has focused on defining enhancers that function in activating the ES cell pluripotent program. However, potential silencing in ES cells defined by our study offers a new perspective to broaden the understanding of the pluripotent state as one in which both activation of epiblast gene expression and inhibition of extra-embryonic gene expression is needed for stability of the stem cell state. Future studies such as identification of co-activators or partners binding to silencers are needed to reveal the molecular mechanisms underlying how candidate silencers function in the process of transcriptional repression. Importantly, our study provides valuable resources to study the function of silencers, one of the cis-regulatory DNA elements, in regulating gene transcription in diverse biological processes.

In conclusion, our findings provide evidence that lineage-specific silencers exist in both TS and ES cells on a genome-wide scale. We identify Cdx2 as lineage-specific repressor that transcriptionally represses ICM pluripotency genes via binding to specific silencer elements. Finally, we propose that lineage-specific transcriptional repression through silencers may serve as a novel mechanism to establish TS/ES unique gene expression patterns, and promote ICM and TE lineage segregation during first cell fate determination.

Methods

Culture of ES cells and TS cells. Mouse ES cells were maintained in Dulbecco’s modified Eagle’s medium (DMEM, Gibco-BRL), with 20% heat-inactivated ES-qualified fetal bovine serum (FBS, Gibco-BRL), 0.055 mM β-mercaptoethanol (Gibco-BRL), 2 mM L-glutamine, 0.1 mM MEM nonessential amino acid, 5000 U/ml penicillin/streptomycin and 1000 U/ml leukemia inhibitory factor (LIF, Chemicon) without MEFs. TS cells are from Rossant lab, and maintained in DMEM, 20% of FBS, 0.1 mM sodium pyruvate, 0.1 mM nonessential amino acids, 0.055 mM mercaptoethanol, 2 mg/ml of sodium heparin (Sigma), and 20 ng/ml of recombinant FGF4 (Sigma) in the presence of 70% of the MEF-conditioned medium.

Single cell RNA-Seq analysis methods of mouse embryo data. We firstly filtered 61 64-cell stage cells from recently published data. Downstream analysis was performed by single cell RNA-seq data analysis tools Seurat in R language. Nearly 2 thousands variable genes were selected for dimension reduction, ten principle components was selected for t-SNE reduction, than 2 distinct clusters was defined. Differential expression genes were determined using default function.

Gene expression profiling of individual blastocysts by high throughput microfluidic qPCR. Total RNA was extracted from individual mouse embryos using the PicoPure RNA isolation kit (Arcturus Bioscience). The entire RNA preparation was used for cDNA synthesis at 37 °C for 2 hrs. using the high capacity cDNA archive kit (Applied Biosystems). One eighth of each cDNA preparation was pre-amplified using the TaqMan primers for genes of interest by 16 cycles of amplification (each cycle: 95 °C for 15 Sec and 60 °C for 4 min) using the TaqMan PreAmp Master Mix Kit (Applied Biosystems). These preamplified products were diluted 5-fold before analysis. Real-time reactions were performed in technical triplicate with master mix (Applied Biosystems) in 48.48 Dynamic Arrays on a BioMark System (Fluidigm). Threshold cycle (Ct) values
High throughput single cell qPCR. Equal volumes of each inventoried TaqMan Gene Expression Assay (20X, Applied Biosystem) were pooled and then diluted using TE buffer so that each assay was at a final concentration of 0.2X. These pooled assays were for use in the pre-amplification step. Individual cells were harvested directly into the 10 µL RT-PreAmp Master Mix (5.0 µL CellsDirect 2X Reaction Mix (CellsDirect qRT-PCR kit, Invitrogen); 2.5 µL 0.2X Assay Pool; 0.2 µL RT/Tag Enzyme (CellsDirect qRT-PCR kit, Invitrogen); 2.3 µL RNase-free water. The harvested single cell samples were immediately frozen and stored at −80 °C. Cell lysis and sequence-specific reverse transcription were performed at 50 °C for 20 min. The reverse transcriptase was inactivated by heating to 95 °C for 2 min. Subsequently, in the same tube, cDNA went through sequence-specific amplification by denaturing at 95 °C for 15 s, and annealing at 60 °C for 4 min for 18 cycles. The pre-amplified products were diluted 5-fold and then analyzed by TaqMan PCR. Real-time reactions were performed with Universal PCR Master Mix and inventoried TaqMan gene expression assays (Applied Biosystems) in 48.48 Dynamic Arrays on a BioMark System (Fluidigm). Threshold cycle (Ct) values were calculated from the system's software (BioMark Real-time PCR Analysis, Fluidigm).

Cdx2 ChIP-Seq with TS cells. TS cells were cross-linked with 1% (w/v) formaldehyde for 10 min at room temperature, and formaldehyde was then inactivated by the addition of 125 mM glycine. Chromatin extracts containing DNA fragments with an average size of 500 bp were immunoprecipitated, using anti-Cdx2 (CDS-88, Biogenex). The ChIP enriched DNA was then decross-linked and sequenced by Genome Analyzer II (Illumina) according to Illumina's manuals.

ChIP-seq data analysis. Peak calling based on the Cdx2 ChIP-seq data was performed using MACS; tag size was set to be 35, mfold was set to be 8, mm9 was used as the reference genome. De novo motif discovery was performed with CisFinder. The association strength between TF and individual genes were calculated with the published method29. For the TF association score, we calculate association strength of TFj on gene i as a weighted sum of intensities of all of the peaks of TFj according to the published method29. D0 is set to be 10000 bp so as to give more weights for enhancer and expression regions. Only peaks within 1Mbp distance of a gene are considered in the calculation. The association score pattern along the gene expression change window was generated using EXCEL. Briefly, genes were ordered according to their expression fold change between the compared samples, then the average association scores in the sliding 500 gene window were calculated across the sorted genome. ChIP-seq and DNase-seq peaks were visualized with IGV software. Overlapping motif analysis was performed using CDX2 ChIP-seq data and previously published OCT4 ChIP-seq data34. We allowed up to 200 bp between the borders of two peaks. To explore silencer candidates in genome-wide scale, we integrated peak calling data from MACS. We exploited BEDtools31 to find silencer candidates with properties that overlapped with binding peak, DNase I hypersensitive sites and enriched H3K27me3 signal. We than assigned silencer candidates to genes defined in the RefSeq (NCBI37/MM9) gene annotations by calculating the distance from the center of the silencer candidates to the TSS of each gene. The silencer candidates was then assigned to the closest gene. We did gene ontology analysis using limma package in R language platfrom to determine whether the set of silencer-associated genes in TS cell was statistically enriched for genes that were important for the maintenance of ES cell identity, and silencer-associated genes in ES cell was statistically enriched for genes that were important for the maintenance of TS cell state.

Data availability. The data generated or analyzed during this study are included in this published article (and its Supplementary Information files). Raw sequencing data is uploaded to figshare. (Data link: https://figshare.com/s/s8119390159270f9a17).
13. Ralston, A. et al. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development 137, 395–403, https://doi.org/10.1242/dev.088828 (2010).
14. Cambuli, F. et al. Epigenetic memory of the first cell fate decision prevents complete ES cell reprogramming into trophoblast. Nat Commun 5, 5538, https://doi.org/10.1038/ncomms6538 (2014).
15. Posfai, E. et al. Position- and Hippo signaling-dependent plasticity during lineage segregation in the early mouse embryo. Elife 6, https://doi.org/10.7554/eLife.22906 (2017).
16. Guo, G. et al. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Developmental cell 18, 675–685, https://doi.org/10.1016/j.devcel.2010.02.012 (2010).
17. Rugg-Gunn, P. J., Cox, B. J., Ralston, A. & Rossant, J. Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proceedings of the National Academy of Sciences of the United States of America 107, 10783–10790, https://doi.org/10.1073/pnas.0914507107 (2010).
18. Latos, P. A. et al. Elf5-centered transcription factor hub controls trophoblast stem cell self-renewal and differentiation through stoichiometry-sensitive shifts in target gene networks. Genes & development 29, 2435–2448, https://doi.org/10.1101/gad.268821.115 (2015).
19. Lohnes, D. The Cdx1 homeodomain protein: an integrator of posterior signaling in the mouse. BioEssays: news and reviews in molecular, cellular and developmental biology 25, 971–980, https://doi.org/10.1002/bies.10340 (2003).
20. Ralston, A. & Rossant, J. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. Developmental biology 313, 614–629, https://doi.org/10.1016/j.ydbio.2010.10.054 (2008).
21. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137, https://doi.org/10.1186/gb-2008-9-9-r137 (2008).
22. Sharov, A. A. & Ko, M. S. Exhaustive search for over-represented DNA sequence motifs with CisFinder. DNA research: an international journal for rapid publication of reports on genes and genomes 16, 261–273, https://doi.org/10.1093/dnares/dsp014 (2009).
23. Ouyang, Z., Zhou, Q. & Wong, W. H. ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America 106, 21521–21526, https://doi.org/10.1073/pnas.0904863106 (2009).
24. Whyte, W. A. et al. Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell 153, 307–319, https://doi.org/10.1016/j.cell.2013.03.035 (2013).
25. Boyle, A. P. et al. High-resolution mapping and characterization of open chromatin across the genome. Cell 132, 311–322, https://doi.org/10.1016/j.cell.2007.12.014 (2008).
26. Calabrese, J. M. et al. Site-specific silencing of regulatory elements as a mechanism of X inactivation. Cell 151, 951–963, https://doi.org/10.1016/j.cell.2012.10.037 (2012).
27. Young, R. A. Control of the embryonic stem cell state. Cell 144, 940–954, https://doi.org/10.1016/j.cell.2011.01.032 (2011).
28. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. Cell 128, 707–719, https://doi.org/10.1016/j.cell.2007.01.015 (2007).
29. Nishiyama, A. et al. Uncovering Early Response of Gene Regulatory Networks in ESs by Systematic Induction of Transcription Factors. Cell stem cell 5, 420–433, https://doi.org/10.1016/j.stem.2009.07.012 (2009).
30. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol 33, 495–502, https://doi.org/10.1038/nbt.3192 (2015).
31. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics 47, 11 12 11–34, https://doi.org/10.1002/0471250953.bi111247 (2014).

Acknowledgements
We thank Y. Zhou, L. Shen, S. Ying and D. Zhou for help on experiments. We thank Y. Hong, H. NG and C. Neil for insightful suggestions for the project. This work was supported by funding from Fundamental Research Funds for the Central Universities (G.G.) and 1000 Youth Talent Plan (G.G.). G.G. is a professor from Zhejiang University.

Author Contributions
Experiments were designed by G.G., carried out by G.G., X.H., P.Y., A.R., L.S., M.H., T.M., L.P., H.N., G.Y. and J.J., analyzed by D.H., G.G. The paper was written by D.H., X.H., J.R., P.R. and G.G. This work was supported by Zhejiang University Stem Cell Institute.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-16009-w.

Competing Interests: The authors declare that they have no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017