Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species

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

Citation


Published Version
doi:10.1101/gr.100594.109

Citable link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:14068405

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
Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species

Dan Xie, Chieh-Chun Chen, Leon M. Ptaszek, et al.

Genome Res. 2010 20: 804-815 originally published online March 10, 2010
Access the most recent version at doi:10.1101/gr.100594.109

Supplemental Material

http://genome.cshlp.org/content/suppl/2010/03/11/gr.100594.109.DC1.html

References

This article cites 43 articles, 14 of which can be accessed free at:
http://genome.cshlp.org/content/20/6/804.full.html#ref-list-1

Creative Commons License

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here

To subscribe to Genome Research go to:
http://genome.cshlp.org/subscriptions

© 2010, Published by Cold Spring Harbor Laboratory Press
Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species

Dan Xie,1,9 Chieh-Chun Chen,1,9 Leon M. Ptaszek,2,3,4,9 Shu Xiao,1 Xiaoyi Cao,5 Fang Fang,6 Huck H. Ng,6 Harris A. Lewin,7 Chad Cowan,3,4 and Sheng Zhong1,7,8,10

1Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA; 2Cardiology Division, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA; 3Harvard Stem Cell Institute, Harvard University, Cambridge, Massachusetts 02138, USA; 4Cardiovascular Research Center, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA; 5Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA; 6Gene Regulation Laboratory, Genome Institute of Singapore, and Harvard Medical School, Boston, Massachusetts 02114, USA; 7Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA; 8Department of Statistics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

Mammalian preimplantation embryonic development (PED) is thought to be governed by highly conserved processes. While it had been suggested that some plasticity of conserved signaling networks exists among different mammalian species, it was not known to what extent modulation of the genomes and the regulatory proteins could “rewire” the gene regulatory networks (GRN) that control PED. We therefore generated global transcriptional profiles from three mammalian species (human, mouse, and bovine) at representative stages of PED, including: zygote, two-cell, four-cell, eight-cell, 16-cell, morula and blastocyst. Coexpression network analysis suggested that 40.2% orthologous gene triplets exhibited different expression patterns among these species. Combining the expression data with genomic sequences and the ChIP-seq data of 16 transcription regulators, we observed two classes of genomic changes that contributed to interspecies expression difference, including single nucleotide mutations leading to turnover of transcription factor binding sites, and insertion of cis-regulatory modules (CRMs) by transposons. About 10% of transposons are estimated to carry CRMs, which may drive species-specific gene expression. The two classes of genomic changes act in concert to drive mouse-specific expression of MTF2, which links POU5F1/NANOG to NOTCH signaling. We reconstructed the transition of the GRN structures as a function of time during PED. A comparison of the GRN transition processes among the three species suggested that in the bovine system, POU5F1’s interacting partner SOX2 may be replaced by HMGB1 (a TF sharing the same DNA binding domain with SOX2), resulting in rewiring of GRN by a trans change.

[Supplemental material is available online at http://www.genome.org. The microarray data from this study have been submitted to the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession nos. GSE18290 and GSE18319.]

Much of the information on evolution of gene regulatory networks (GRN) was generated from studies in invertebrates (Erwin and Davidson 2009). Comparison of fungi species revealed evolution of GRNs at both cis and trans levels (Scannell and Wolfe 2004). Invertebrate GRNs can be “rewired” at the “cis” level through changes in genomic regions that regulate gene transcription, or cis-regulatory modules (CRMs). CRM can gain or lose transcription factor (TF) binding sites (TFBSs) by single nucleotide mutations, insertions and deletions of neutral sequence or TFBSs (Erwin and Davidson 2009). It has been argued that at the cis level, the transcription regulatory apparatus is a “programmable” computing machine whose function can be modulated through placements of TFBSs in the regulatory region of any gene (Buchler et al. 2003; Gertz et al. 2009). A notable advantage of encoding combinatorial control in the regulatory region, as opposed to in the regulatory proteins, is evolvability (Ptashne and Gann 2002): Unlike regulatory proteins, each cis-regulatory region controls the expression of a given gene and hence can be programmed with minimal pleiotropic effects. Contrary to this argument, changes to the TF that are produced by the cell ("trans" changes) appear to be widely tolerated by Escherichia coli and may even confer evolutionary advantages (Isalan et al. 2008). Moreover, several yeast species implement compensating cis and trans changes to achieve a conserved regulatory logic (Tsong et al. 2006). This study attempted to add vertebrate data to address these arguments by analyzing a relatively conserved developmental process in three mammalian species.

Mammalian preimplantation embryonic development (PED) encompasses the period from fertilization to implantation of the embryo in the endometrial lining of the uterus. Just after fertilization, the transcriptome of the embryo is comprised of maternally deposited transcripts. After several cell divisions, maternal transcripts are specifically degraded and are replaced by zygotic transcripts produced by the new diploid nucleus containing both maternal and paternal genes. This transition is termed zygote genome activation (ZGA). The timing of ZGA and maternal transcript degradation varies among species, but in humans and bovines it reportedly occurs between the four- and eight-cell stages of...
development, while in mouse, this transition occurs between the one- and two-cell stages of development, consistent with an overall accelerated pace of murine embryonic development (Braude et al. 1988; Kanka et al. 2003; Adjaye et al. 2007; Kues et al. 2008).

Another major event during PED is the differentiation and specialization of cells. Immediately after ZGA, embryos undergo compaction, a process during which dividing embryonic cells become a tightly organized cell mass with an indistinguishable membrane. In most mammalian embryos, compaction is usually observed between the eight- and 16-cell stages. After compaction, the embryo is named the morula, which becomes the blastocyst through formation of an internal cavity and further cellular differentiation. At least two cell types can be distinguished in the blastocyst: an inner cell mass (ICM) and an outer layer of cells termed the trophectoderm. The ICM contributes to all three germ layers present in the mature adult (known as pluripotency). Embryonic stem (ES) cells were derived from ICM, and ES cells have been used as surrogates to analyze the molecular mechanisms in ICM and preimplantation embryos (Ogawa et al. 2008). The progression of PED is highly conserved among mammals: All progress through the same morphologic stages. Perhaps the most marked difference is the amount of time spent at each stage. The other notable interspecies differences appear after the blastocyst stage, including that bovine blastocysts initiate the gastrulation process before implantation, and the formation and functions of the placenta and yolk sac (Gossler 1992). In this study, we chose the highly conserved PED stages, i.e., from zygote to blastocyst stages for cross-species GRN comparison.

**Results**

**Dynamic gene expression landscapes of preimplantation development in three species**

Human, mouse, and bovine embryos were collected at representative stages of PED, including oocyte (bovine only), one-, two-, four-, eight-, 16-cell (bovine only), morula, and blastocyst. Developmental stage was determined by direct microscopic visualization of each collected embryo (Fig. 1A). RNA from each stage was collected and analyzed with species-specific Affymetrix GeneChips. Biological replicates of the same developmental stage exhibited strong correlations, which were comparable to the sample correlations in previous bovine and mouse studies (Fig. 1B; Hamatani et al. 2004; Wang et al. 2004; Kues et al. 2008). Hierarchical clustering analyses revealed global transcriptional similarities among the biological replicates of the same developmental stages together, with a few exceptions of the samples at neighboring stages (Fig. 1C). The hierarchical order of all the samples almost perfectly reflected the relative time differences of all the developmental stages. The largest changes of expression were observed between four- and eight-cell stages in human and mouse and between eight- and 16-cell stages in the bovine system, in keeping with the previously reported timing of ZGA in these species.

In order to explore the conservations and changes of gene expression patterns across species, we developed a Bayesian clustering method to simultaneously cluster the timecourse expression data and automatically determine the cluster number (Methods). Nine clusters were generated for 5941 orthologous gene triplets (including 1036 TF triplets) (Supplemental Fig. S1). Of these orthologous triplets, 2122 (35.7%, 383 TF triplets) fell into clusters with consistent expression patterns in all the three species, and the remaining 3290 orthologous triplets (55.4%, 444 TF triplets) showed at least two different expression patterns among these species. The large differences in the TF expression patterns suggested
that there might be a substantial interspecies difference of GRN structures. To test if this result is sensitive to the input genes, we repeated this analysis with the 927 orthologous gene triplets (including 371 TF triplets) that exhibited significantly variable expression (ANOVA $P$-value $< 0.05$) during PED in at least two of the three species. Among them, 383 (40.2%) orthologous triplets (149 TF triplets) showed different expression patterns among these species, consistent to the estimated large-scale GRN changes.

The three mammals share more maternally deposited than ZGA-activated RNAs

The transition from predominance of maternal transcripts to zygotic transcripts is a critical aspect of PED. We identified maternally deposited (MT) and ZGA-activated transcripts (ZGAT) in each species based on trends in transcript levels (Fig. 2B). Comparing the conservation levels of the human MTs vs. human ZGAT,

![Figure 2](https://example.com/figure2.png)
we found 69.4% human MT and 40.0% human ZGAT were shared in
mouse (ratio = 69.4/40.0 = 1.74, P-value < 10^{-10}); 63.8% HMt
and 18.5% human ZGAT were shared in bovine (ratio = 3.45,
P-value < 10^{-20}); 49.0% human MT and 8.5% human ZGAT were
shared in both mouse and bovine (ratio = 5.76, P-value < 10^{-20}).
These data showed that human MT is more conserved than human
ZGAT in mammals (Fig. 2A,B). Comparisons benchmarked against
mouse or bovine MT and ZGAT generated similar results. Changing
the algorithmic thresholds for detection of MT and ZGAT did not
affect this conclusion (Supplemental Fig. S2A).

Rewiring of ZGA genes by TFBS usage

We hypothesized that the large-scale changes of ZGAT across spe-
cies is associated with changes of TFBSs of TFs that activate gene
expression in morulas and ICM. To test this hypothesis, we inter-
sected human and mouse ZGAT genes with the genes that are bound
by POU5F1, SOX2, and NANOG in human ES cells (Lister et al.
2009) and mouse ES cells (Chen et al. 2008) (Methods). POU5F1,
SOX2, and NANOG are hereafter referred to as ZGA TFs. Among
the ZGA genes that are conserved, human-specific and mouse-specific,
58, 64, and 335 of them are bound by ZGA TFs in a conserved,
human-specific and mouse-specific manner (Supplemental Fig.
S2B). Surprisingly, the TFBSs of ZGA TFs are not enriched in the
conserved ZGA genes (χ^2 test with continuity correction P-value >
0.31). Instead, mouse-specific ZGA genes are enriched with mouse-
specific binding of ZGA TFs (P-value < 2 × 10^{-4}). The large changes
of ZGATs and the TFBSs of ZGA TFs across species may suggest
the extent of rewiring associated with ZGA.

Transposons contribute to species-specific gene expression

There is a significant overlap between transposon-associated ele-
ments and TFBS-rich regions in mammalian genomes. One nota-
ble example includes association between the TP53 binding sites
and transposons in the human genome (Wang et al. 2007; Bourque
et al. 2008), and one of the TP53 binding transposons was stabi-
lized in humans (Wang et al. 2007); however, it is yet to be
determined to what extent the transcriptome is modulated by
transposons. Therefore, we attempted to assess the overall impact
of murine-specific transposons on the PED transcriptome.

Previously reported chromatin immunoprecipitation with
massively parallel sequencing (ChIP-seq) mapping of mouse ES
cells identified binding sites of 16 transcriptional regulators to the
mouse genome (Chen et al. 2008). Based on this information, we
identified the binding regions that resided in murine-specific
transposable regions (Fig. 3A). CTCF-binding regions exhibited
the largest number of overlap with known transposons (10,313 over-
lapping regions, representing 26.0% of total CTCF-binding regions),
among which 9074 (22.9%) overlapped with the B2 family of
transposons. In addition to the CTCF-binding regions, the binding
regions for the following factors exhibited nontrivial overlaps
(>10% of the binding regions of a factor) with transposons: ESRRB,
SOX2, NANOG, SMAD1, TFCFP21, STAT3, EP300, and POU5F1.
Transposon families B2, ERVK, and B4 were most strongly associ-
ated with the binding regions of these transcription regulators.
This suggests that these families of transposable elements contain
the DNA binding motifs of the listed TFs. On the contrary, 10
(0.2%) and 1 (0%) of Suz12 and RNA polymerase II (Pol II) binding
regions overlapped with transposons, far below the expected
amount of overlaps (16.3%), suggesting that the insertion of
transposable elements to the binding regions may negatively affect
the function of these factors.

A number of the TFs included in the original ChIP-seq analysis
(Chen et al. 2008), especially NANOG and POU5F1, exhibited
increase in expression after ZGA. In addition, these TFs are known
to specify ICM expression of their target genes. Therefore, if the
binding of these TFs to the murine-specific transposons manifests
functional consequences to the transcriptome, we would expect to
see mouse-specific expression of the genes near these transposons.
Among the 7450 orthologous gene triplets in the PED data set,
3369 (45.2%) exhibited mouse-specific expression patterns. Within
the 3066 gene triplets whose mouse gene had a nearby transcription-
factor-bound transposon, 1403 (45.8%) exhibited mouse-
specific expression patterns. The genes near TF-bound transposons
did not exhibit a significantly larger percentage of mouse-specific
expressed genes, suggesting that the majority of binding
events between the transcription regulators to the transposons do
not manifest functional consequences. Therefore, the majority of
mutants generated by the insertion of TFBS-carrying transposons
are likely to be selectively neutral (Kimura 1991).

We hypothesized that the transposons containing binding
regions of multiple TFs (CRM-containing transposons) were more
likely to affect the expression of neighboring genes. We tested this
hypothesis by first classifying the transposons by the number of
transcription regulators bound to them and then analyzing the
effect on gene expression of each class of transposons (Fig. 3B).
A total of 20,784 murine transcription-regulator bound transposons
were identified, among which 2212 (10.2%) transposons were
bound with at least two transcription regulators. For each class
of transposons, we identified the subset whose nearest genes ex-
hibited mouse-specific expression patterns (Methods). First, we
compared mouse PED to human PED, ignoring the bovine ex-
pression data. For the transposons bound by one transcription
regulator, 65% of the nearest mouse genes exhibited mouse-
specific expression changes, which is comparable to the total
percentage of genes that exhibit mouse-specific expression changes
(64%). For the transposons bound by at least two (and up to six)
factors, the percentages of the nearest genes with mouse-specific
expression changes increased consistently from 67.8% to 85.2%
(Fig. 3C). Second, we compared mouse PED to both human PED
and bovine PED. From the single-factor-bound transposons to the
six-factor-bound transposons, the percentages of the nearby genes
that exhibited mouse-specific expression increased consistently from
45.1% to 70%. These data suggest that species-specific transposons
bound by a single TF generally do not change gene
expression, whereas CRM-carrying transposons bound with mul-
tiple TFs are statistically associated with species-specific expres-
sion. About 10% of the transposons in our analyses carry CRMs.

Gain of functional TFBSs and CRM by species-specific
transposons

To confirm the functional importance of the transposons to gene
regulation, we further characterized two murine-specific trans-
posons that carry functional TFBS and CRM and, in turn, regulate
mouse-specific gene expression in early development. One such
transposon is Lx8, a member of non-LTR retrotransposons, which
we estimated to be inserted into the murine genome after the di-
vergence of murine and primate species, ~60 million years ago
(Supplemental Fig. S3). A copy of Lx8 is present upstream of the
cold shock domain–containing protein A (Csda) gene in the mouse genome. Csda is a transcription repressor, highly conserved among eukaryotes from yeast to human. In the mouse, Csda is known to be critical for several processes in reproduction and development, including spermatogenesis and in utero fetal development. The function of Csda in other mammals has not been tested. Transcription of Csda in mouse PED starts at the time of ZGA (two-cell stage), with an apparent pause during the four-cell stage, followed by continued increase from the eight-cell stage to the morula stage (Fig. 4). The increase in Csda transcription during mouse PED was correlated with Pou5f1 expression (R = 0.41, permutation P-value = 0.046). Like Pou5f1, Csda transcripts were restricted to the ICM at the blastocyst stage (Yoshikawa et al. 2006); however, ZGA of CSDA was not conserved in human or bovine, suggesting that the mouse Csda is associated with species-specific regulatory sequences. The insertion of Lx8 upstream of the mouse Csda gene produces a consensus POU5F1 binding site ([T/A]ATG[C/T]AAAT). This POU5F1 TFBS was functional in mouse ES cells: Analysis of published ChIP-seq data (Chen et al. 2008) revealed that the POU5F1–SOX2-bound regions together with the Pol II-bound Csda promoter. ChIP-chip analyses of human ES cells did not identify any POU5F1, SOX2, or NANOG binding sites upstream of the human CSDA gene. Furthermore, motif scans based on the POU5F1 and NANOG motifs did not detect any high-affinity TFBS within 20 kilobases (kb) upstream or downstream of the human CSDA transcription start site (TSS). The data produced by gene expression, DNA motif, and TF binding analyses collectively suggest that the unique, transposon-mediated change in the regulatory sequence of the mouse Csda gene is associated with a gain of function.

A second transposon in the mouse genome is associated with species-specific gene expression during PED. This transposon, Orr1b1-int, a member of the ERV3 family of retroviruses, is inserted upstream of the murine Mtf2 gene (Fig. 5A,D). Orr1b1-int is a “younger” transposon than Lx8 (Supplemental Fig. S3), but may...
have more potent gene regulatory effects than Lx8. Orr1b1-int contains a CRM comprised of two TFBSs (POU5F1 and a NANOG), in contrast with the single POU5F1 site in Lx8. ChIP-seq data revealed strong binding of POU5F1 and NANOG to their corresponding TFBSs within this Orr1b1-int, as reflected by large sequence counts: (the NANOG site ranked 434 among 10,345 detected NANOG sites, and the POU5F1 site ranked 66 among 3736 detected POU5F1 sites) (Chen et al. 2008). The POU5F1 binding on this Orr1b1-int region was five times stronger than on the Lx8 Csda enhancer. For both NANOG and POU5F1, the minimal overlaps of the ChIP-seq reads pinpointed to sites that matched their known DNA binding motifs (Fig. 5D), further confirming that the TF–DNA binding was a result of the presence of TFBSs within the Orr1b1-int sequence. These data, together with the gain of another NANOG TFBS by mutation (see below), offer an explanation for the dramatic difference in MTF2 expression among the three mammals analyzed in this study (Fig. 2A). Although it is not known why MTF2 transcripts are present in only human and bovine oocytes, the dramatic, mouse-specific induction of the Mtf2 gene during PED is consistent with the gain of murine-specific regulatory sequences.

**Turnover of TFBSs caused by single nucleotide mutations**

Besides transposons, we postulated that single nucleotide changes may cause TFBS turnover and thus be responsible for interspecies differences in mammalian PED. Associating TFBSs with single nucleotide mutation with the genes with species-specific expression, we found three single nucleotide mutations that appear to influence gene expression by modulating the binding affinities of the TFBSs. Two of the three mutations locate within the same TFBS upstream to the FRAT2 gene, and the third mutation locates upstream to the MTF2 gene. **FRAT2** belongs to the GSK-3-binding protein family and is a positive regulator of the WNT signaling pathway. Expression of the human and bovine **FRAT2** genes is strongly induced in late PED, whereas the mouse **Frat2** gene appears to be transcriptionally...
silent throughout PED. (Supplemental Fig. S4A). The absence of expression in mouse is attributable to an A-to-C mutation at the center position of the POU5F1 binding site (Supplemental Fig. S4C,D), which eliminated the POU5F1 binding, as indicated by a chromatin immunoprecipitation with a microarray hybridization (ChIP-chip) peak on the human POU5F1 site (Boyer et al. 2005), but no ChIP-seq signal in the genomic neighborhood of the mouse Frat2 gene (Chen et al. 2008) (Supplemental Fig. S4B). This mutation at the center of the DNA binding motif caused a loss of the TFBS. Furthermore, the human Frat2 exhibits larger induction of expression in response to the increased concentration of POU5F1 during PED, as compared with the bovine system. This delicate difference is attributable to a primate-specific T-to-A mutation at the second position (a peripheral position) of the same POU5F1 TFBS (Matys et al. 2003). In summary, in the case of an orthologous group of POU5F1 TFBSs, two mutations are present in DNA motif. The central mutation was consistent with a dramatic change on gene expression, and the peripheral mutation was reflected by a delicate change of expression.

A NANOG TFBS resides upstream of the Mtf2 gene in the mouse, but not in human or cow. ChIP-seq data accumulated over 62 sequence reads, overlapping on a single nucleotide in the mouse genome. (Supplemental Fig. S4F). This gain is associated with mouse-specific regulation of Mtf2. Sixty-two NANOG ChIP-seq reads overlapped a single nucleotide upstream of mouse Mtf2,
at a location corresponding to a NANOG binding site (Fig. 5A). The
position in question is occupied by an A in the mouse and a T in all
other mammals. This suggests that the murine lineage contains
a distinct T-to-A mutation (Fig. 5B). We sought to assess whether
the putative “ancestral” sequence (not containing the T-to-A mu-
tation) might harbor a NANOG TFBS. To this end, we mutated the
functional binding site from A (mouse) back to T (ancestral). This
mutation abrogated NANOG binding to the site in vitro (Fig. 5C).
An A-to-C mutation had the same effect. These data suggest that
the mouse TFBS was “born” from this mutation. Flanking CTCF
insulator TFBSs grouped this newborn mouse NANOG TFBS to-
gether with the Osr1-int transposon reported above and the Pol
II bound Mtf2 promoter, creating an autonomous cis-regulatory
region within (Fig. 5A). This autonomous cis-regulatory region
contained two murine-specific enhancers, which were generated by
different evolutionary mechanisms, but synergistically contributed
to the dramatic change of Mtf2 expression (Fig. 2A).

What is the consequence of the MTF2-associated gain
of regulatory sequences on GRN?

We suspect that the altered regulatory sequence is responsible for
the murine-specific expression of Mtf2. We sought to further ex-
ploring the impact of this gain-of-function on the GRN of mouse
blastocyst embryos. To this end, mouse ES cells and trophoblast
stem (TS) cells, derived from the ICM and the trophoderm of
blastocysts, respectively, were used to analyze tissue-specific gene
regulation. Mtf2 exhibited a 4.22-fold increase of expression level
in ES cells than in TS cells (Fig. 5E), consistent with ICM-specific
expression of POU5F1 and Nanog in the mouse (Yoshikawa et al.
2006). The up-regulation of Mtf2 in ES cells is consistent with a
previous computational result, which predicted Mtf2 should bear
functional importance in mouse ES cells (Zhou et al. 2007). To test
this, we knocked down Mtf2 in mouse ES cells with shRNA. Upon
two days of transfection, two shRNA constructs both reduced Mtf2
expression to about 33% of the expression levels in the ES cells
transfected with control plasmids. These cells were subjected to
expression microarray analysis. Mtf2 knockdown resulted in the
expression of 14 genes that are involved in signaling and transcrip-
tion (false discovery rate = 0.06), notably Notch3 (ranked ninth among
28,064 genes, Fig. 5F). Notch3 is one of the receptors in the Notch
signaling pathway, which modulates specification of germ layers and
differentiation of ES cells. These data are consistent with the notion that the gain of regulatory
sequences detailed here may implement a Pou5f1/Nanog — Mtf2 —
Notch3 pathway in the ICM of mouse blastocyst embryos. This
model predicted that the expression of Notch3 should be restricted
to the ICM. This prediction was confirmed by whole-mount, in
situ hybridization (Yoshikawa et al. 2006).

Dynamic changes in GRNs during PED

The static GRNs active in human ES cells were recently described
(Muller et al. 2008), although the dynamic processes that are capa-
ble of generating pluripotent GRNs remain elusive (Mikkelsen
et al. 2008). Since many varieties of ES cells are procured from the
inner cell mass of blastocyst-stage embryos, a more comprehensive
understanding of the molecular events that drive PED could illu-
minate the mechanisms by which pluripotency-associated GRNs
active in ES cells are constructed. Species-specific differences in the
dynamic development of these GRNs may also help us better un-
derstand the different behaviors of ES cells from different species.

We therefore reconstructed the dynamic transitions of GRNs
during PED in the human, murine, and bovine systems. For each
stage of PED, we identified the interacting modules of genes (in
each species) by jointly modeling the expression data and a refer-
ence protein interaction network with the MATISSE program as
described in Methods (Fig. 6A; Ulitsky and Shamir 2007). The ex-
pression data from a given developmental stage were used in conjunction with expression data from adjoining stages: e.g., one-
two-, and four-cell expression data were used for reconstructing the
interacting modules at the two-cell stage. The reference protein
interaction network was generated by integrating the data from the
Human Protein Reference Database (Mishra et al. 2006) with
high throughput yeast two-hybrid screens (Ulitsky and Shamir
2007) and mass spectrometry analyses (Wang et al. 2006; Ewing
et al. 2007). After constructing the state transition of the GRNs
(http://biocomp.bioen.uiuc.edu/ped), we focused our analyses on the
modules that contained the key regulatory genes that were
induced in ES cells, as opposed to differentiated cells (Fig. 6B).
Although expression data from the blastocyst-stage embryos were
generated from a mixture of both ICM and trophectodermal cells,
the reconstructed GRN in human blastocysts resembles the GRN of
human ES cells (Muller et al. 2008). This similarity is highlighted
by the interaction of Pou5f1, Sox2, Nanog, Sall4, and C1orf103.
The resemblance of the two GRNs lends credibility to the GRN
reconstruction procedure and also suggests the genes involved in the
Pou5f1 interaction module exhibited a strong coexpression
pattern during the morula-to-blastocyst transition.

We found a few shared protein interactions among the three
species. For example, HDAC2 is predicted to interact with Pou5f1
(mouse, bovine) or Nanog (human) in early PED stages (four-cell
human and bovine, one-cell mouse). HDAC2 is one of the histone
deacetylase (HDAC) family proteins and is responsible for tran-
scriptional repression. The fact that HDAC2 exhibited conserved
interacting partners may indicate that its role of regulating cleav-
age of mouse preimplantation embryos (Ito et al. 2000) is con-
served across species as well. An interaction module involving
Pou5f1, Sox2, Nanog, Sall4, and Ilf2 is shared between hu-
man and mouse blastocyst embryos. In particular, the predicted
Pou5f1–Ilf2 interaction at the blastocyst stage is shared in all
three species. Ilf2, also known as Nf45, is a TFI required for T-cell
expression of the interleukin 2 gene. Although its role in non-T cells
is largely unexplored, Ilf2 does physically interact with Nanog and
Pou5f1 in mouse ES cells (Wang et al. 2006). The conserved Pou5f1–
Ilf2 interaction might suggest a novel transcriptional regulatory mod-
ule responsible for regulating blastocyst embryos and ES cells.

A notable interspecies difference is that Sox2 is recruited to
the Pou5f1 module in human and mouse morulae and blastocysts,
but not in bovine morula or blastocysts. Instead, Pou5f1 is pre-
dicted to interact with Hmgb1 in bovine morula (Fig. 6, dashed
circles). Analyses of the probe sequences for the Hmgb1 and Sox2
genes excluded cross-hybridization as a confounding factor to this
prediction. Hmgb1 (high-mobility group box 1) is a DNA-binding
protein containing the HMG-box DNA binding domain. Sox2 is
characterized by the same HMG-box protein domain, which is the
only protein domain identified thus far in Sox2 (Supplemental
Fig. S5A). A reanalysis of gene expression levels using real-time PCR
confirmed the increased expression level of Hmgb1 in bovine
morulae (Supplemental Fig. S5A). The expression and protein
domain data suggest that the Pou5f1–Sox2 interaction may have been
replaced by the Pou5f1–Hmgb1 interaction without chang-
ing target CRMs in bovine preimplantation embryos. Further
experiments will be required to test this hypothesis.
Discussion

An evolutionary model for TFBS insertion by transposons

It has been recently observed that species-specific transposable elements carried a substantial fraction of TFBSs that are associated with TFs in vivo (Wang et al. 2007; Bourque et al. 2008). The influence of these transposable elements on the transcriptome has not been extensively explored. Our analyses suggest that although tens of thousands of transposons bind with TFs in vivo, the majority of such TF-bound transposons do not affect the expression of the surrounding genes. We estimate that 10% of murine-specific transposons, generally those that carry CRMs, may affect gene expression. These data suggest a model for the evolution of cis-regulatory sequences by insertion of transposons. The TFBSs associated with transposons are inserted into the host genome and bind with corresponding TFs with similar stoichiometric coefficients as the parent TFBSs. The majority of these inserted TFBSs do not affect the expression of the surrounding genes and are evolutionarily neutral. A small fraction of the transposons, especially those that carry CRMs and attract a module of TFs, can change the expression of surrounding genes, and may be subject to positive or negative selections.

ES cells better represent ICM than EpiSCs

Pluripotent stem cells have been derived from several embryonic source tissues in mouse: ES cells are derived from the inner cell mass of blastocyst stage embryos, and epiblast stem cells are derived from the late epiblast layer of post-implantation mouse embryos (mEpiSCs) (Brons et al. 2007). mEpiSCs and mES cells exhibit nonidentical TF binding patterns and gene expression patterns. In terms of both enhancer usage and gene expression patterns, mES cells resemble ICM, whereas mEpiSCs are different. Based on these data, it was suggested that mES cells and mEpiSCs are two distinct pluripotent states representing cells of the preimplantation embryo and later epiblast cells (Brons et al. 2007), which is in line with our approach of using TF-binding data in ES cells to study TF-binding in PED.

Pluripotency is a conserved phenotype that can be sustained by evolvable GRNs

Human and mouse ES (hES and mES) cells are similar in the sense that they are both pluripotent cell populations that are capable of self-renewal that are derived from the ICM of blastocyst embryos. Even so, the growth factor requirements for hES and mES maintenance are different: mES cells require the presence of Lif, whereas hES cells do not. The differences in media conditions are thought to reflect differences in critical signaling networks responsible for maintenance of pluripotency and self-renewal capability between species. For example, TF Foxd3 is required for mES cell self-renewal, but its expression appears to be nonessential for hES (Ginis et al. 2004). Similarly, STAT3, a TF downstream to LIF signaling, is also required for self-renewal and the maintenance of pluripotency of mES cells, but it seems to be dispensable in hES cells (Schuringa et al. 2002). Compared to mES cells, mEpiSCs are functionally more similar to hES cells in several ways, including their responses to BMB4 and Activin/Nodal signaling (Brons et al. 2007).
observations led to the hypothesis that pluripotent cell identity can be established and maintained through more than one GRN. These GRNs may share core components universally indispensable for pluripotency. Peripheral components, though critical for cell fate, can be implemented using alternative designs. Because ES cells are derived from the ICM of blastocyst embryos, the natural processes of building pluripotent GRN are reflected by the transitions of the maternal cellular environment to ZGA-activated GRN. The comparison of GRN across multiple stages of PED suggested novel factors, such as MTF2 and HMGB1 for building a pluripotent GRN. These factors serve as candidates of dispensable reprogramming factors, which may be tested for potential enhancement of reprogramming efficiency. Since these factors may already be expressed in some tissues, they may contribute to customizing the combination of reprogramming factors for each tissue.

Methods

IVF embryos

The IVF conditions were optimized for each species. To minimize the influence of IVF conditions to cross-species comparison, we subsequently based the comparisons on either relative expression changes or coexpression patterns (Stuart et al. 2003; Gilad et al. 2006; Tirosi et al. 2007).

Human embryos

Supernumerary human embryos were obtained from patients at the Boston IVF Clinic through informed consent with the approval of the Internal Review Board of Harvard University. Embryos thus obtained were cultured in a two-stage culture system (Sage/Biopharma) in microdrops at 37°C and 5% CO2 under oil (Sage) as described previously (Kaferlein and Powers 2007). Embryos from each stage were segregated into three groups. Each group was considered to be a biologic replicate: Embryos within each replicate were pooled for further analysis. The blastocyst embryos were graded according to the Gardner scale (Brook and Gardner 1997; Van Royen et al. 1999).

Mouse embryos

Mouse embryos were obtained as described previously (Wang et al. 2004). Briefly, female mice were impregnated, and embryos were flushed from the oviducts. Individual embryos were morphologically staged by light microscopy.

Bovine embryos

Ovaries were washed, and 2–6 mm follicles were aspirated. Only oocytes surrounded by several cell layers of dense cumulus were selected for culture. Oocytes were washed three times in TL-HEPES (0.22 mM sodium pyruvate, 3 mg/mL BSA at pH 7.4), and we placed maturation media that had been equilibrated in 5% carbon dioxide and air (39°C, high humidity). The maturation medium consisted of TC199 supplemented with 3 mL of bovine LH and FSH (Siuoh Biochemical), 0.25 mM sodium pyruvate, and 10% fetal calf serum. After 20 h of incubation in maturation media, oocytes were removed, washed in TL-HEPES and placed in groups of 10 oocytes each into 44-μL microdrops of IVF-Talp (Biowhittaker) supplemented with 0.22 mM sodium pyruvate, and 6 mg/mL essentially fatty acid-free BSA. Live sperm were separated by centrifugation in a percoll gradient (Sigma, P1985). Eggs were fertilized by adding one million/mL of sperm, 2.0 μg of heparin (Sigma), pencillamine, hypotaurine, and epinephrine (Sigma) to each drop of fertilization media. Oocytes were removed from the fertilization media after 18–22 h, washed in TL-HEPES, placed in a 1.5-mL tube, and vortexed for 2 min to remove cumulus cells. Twenty-five cumulus-free oocytes were placed in each 50-μL drop of SOF culture medium, supplemented with 8 mg/mL of fatty acid–free BSA, nonessential and essential amino acids, and pyruvate.

Transcription profiling

RNA was isolated from embryos using the PicoPure Isolation system (Arcturus) according to the manufacturer’s instructions. Two rounds of linear amplification were performed for each sample to generate cDNA probes from total RNA as previously described (Baugh et al. 2001) using the PicoAmp system (Arcturus). The cDNA produced by this system was converted to aRNA and labeled using a T7-based Invitro-Transcription Labeling system (Enzo Lifesciences). Labeled aRNA was hybridized to the Affymetrix U133 2.0 plus, MOE430A 2.0 and Bovine Genome Genescips array (Affymetrix) according to the manufacturer’s instructions. These expression data were submitted to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) under accession no. GSE18290. The raw data for each species were normalized and modeled with dChip software.

Clustering analysis

We developed a Bayesian clustering method to cluster the time-course expression data (W Huang, X Cao, G Lin, D Xie, and S Zhong, in prep.; see Supplemental material). The cluster generating process is modeled with a Dirichlet Process (Rasmussen 2000) and the time course profile of each cluster is modeled with a mixed-effects model with splines (Luan and Li 2003). The advantages of this method are: (1) it explicitly utilizes the timecourse data structure to minimize the adverse effects of noise in the data; (2) it automatically identifies the proper cluster number; (3) it is based on a fully described probabilistic model and therefore it minimizes the use of heuristic algorithmic steps. The difference in developmental time had little contribution to the detected difference of expression patterns. For example, maternal deposited RNA for TCF7, the canonical TF for the WNT signaling pathway, started degradation at four-, two-, and eight-cell stages in human, mouse, and bovine, respectively (Fig. 2A), but TCF7 was grouped into clusters of consistent expression patterns.

Identifying genes with mouse-specific expression changes

For the mouse vs. human comparison, we defined mouse-specific expression change as a gene exhibiting a non-flat expression trajectory in mouse PED, but its orthologous human gene exhibiting a flat expression trajectory in human PED, ignoring the bovine expression data. For the mouse vs. human and bovine comparison, we defined mouse-specific expression change as a gene exhibiting a non-flat expression trajectory in mouse PED, but its orthologous genes exhibiting flat expression trajectories in both human PED and bovine PED. To detect if a gene has a flat expression trajectory, we compared the expression levels between the first time point and every later time point, using t-tests. If any of the t-tests showed a significant difference (P-value < 0.05), then this gene was considered to have a non-flat trajectory. If none of the t-tests showed a significant difference, this gene was considered to not have any expression changes.

Detection of maternal and ZGA transcripts

The MT genes were required to have presence call in all the replicates at one-cell stage (oocyte and one-cell stages for bovine). The ZGA genes were identified with the following criteria: There is a detectable increase in expression levels during PED. To detect an
increase of expression, we compared every time point to its previous time point. A detectable increase was defined as: fold change > 2, difference of dChip gene expression index > 100 and two-sample comparison P-value < 0.05.

ChIP-seq data

ChIP-seq data in human and mouse ES cells were obtained from GSE17917, GSE18292 (Lister et al. 2009), and GSE11431 (Chen et al. 2008). ChIP-seq identified TFBSs were associated with a gene if it is within a 50,000-flanking region (25,000 upstream and downstream) of the transcription start site of the gene. A ZGA gene is called bound by a ZGA TF, if it is bound by any of the three TFs: POU5F1, SOX2, and NANOG.

Estimation of transposon ages

All the Lx8 and Orr1b1-int transposons were downloaded from the RepeatMasker track on the UCSC Genome Browser (http://genome.ucsc.edu). The ages of the repeats were estimated by age = divergence/substitution rate (Bourque et al. 2008). The substitution rate of the mouse genome used in this analysis was 4.5 × 10⁻⁹ (Mouse Genome Sequencing Consortium 2002).

Detection of TFBS-carrying transposons

Coordinates of ChIP-seq detected binding regions of 16 transcription regulators in mouse ES cells were downloaded from GEO (Chen et al. 2008). Coordinates of transposable elements were downloaded from the UCSC Genome Browser, and the murine specific transposons were identified by the annotation of RepBase (Jurka et al. 2005). We took the intersection of these two sets of sequences and obtained a list of TFBS-carrying transposons.

Detection of species-specific expression

The map of orthologous human, mouse, and bovine probe sets was downloaded from the Affymetrix website (http://www.affymetrix.com/Auth/analysis/downloads/na28/ivt/Bovine.na28.ortholog.csv). To resolve ambiguous mapping, for each gene we chose the probe set with the largest mean signal as the representative probe set for that gene. A gene was identified as exhibiting species-specific expression when either of the following two criteria was satisfied: (1) It was called “absent” in all time points in one species, but it had "presence" calls in other species. (2) When it was present in all species, its fold change of early (one-to-eight-cell) and late (morula and blastocyst) embryos were computed. The fold change was more than twice as large in one species as compared to another.

shRNA-mediated RNA knockdown

Feeder-free E14 mouse ES cells were cultured at 37°C with 5% CO₂. All cells were maintained on gelatin-coated dishes in DMEM (GIBCO), supplemented with 15% heat-inactivated FBS (GIBCO), 0.055 mM β-mercaptoethanol (GIBCO), 2 mM l-glutamine, 0.1 mM MEM nonessential amino acid, 5000 units/mL penicillin–streptomycin, and 1000 units/mL LIF (Chemicon), as described previously. Transfection of shRNA constructs was performed using Lipofectamine 2000 (Invitrogen) according to manufacturer’s instructions. Briefly, 1.5 μg of plasmid DNA was transfected into ES cells on 60-mm plates for RNA extraction. Puromycin (Sigma) selection was introduced 1 d after transfection at 1.0 μg/mL and maintained for 2 and 4 d before harvesting.

shRNA targeting specific genes was designed as previously described (Reynolds et al. 2004; Ui-Tei et al. 2004). The 19-nucleotide (nt) hairpin-type shRNAs with a 9-nt loop were cloned into pSUPER.puro (BgIII and HindIII sites, Oligoengine). Two shRNA, targeting different regions of respective transcripts, were designed for MTF2 to ensure specificity. These short nucleotides are: GGA GACTGTTATTTAAAAGA, GCCACCATGCGTTTACA, pSuperpuro constructs expressing shRNA against luciferase (Firefly) were used as controls.

EMSA

Electrophoretic mobility shift assay (EMSA) was used to analyze the binding affinities of putative NANOG TFBSs. The mouse wild-type “AGTGCTCA” was compared to two site-directed mutagenesis sites “AGTGCTGt” and “AGTGCTCc”, denoted as “A>T” and “A>C,” respectively. A recombinant NANOG protein (GST-tagged) was used for EMSA. The full-length NANOG protein was cloned into the pET42b (Novagen) vector, expressed and purified with GST-sepharose beads (Amersham). DNA oligonucleotides (Proligo) labelled with biotin at the 5’ end of the sense strands were annealed with the antisense strands and were purified with agarose gel DNA extraction kit (Qiagen). A LightShift Chemiluminescent EMSA kit (Pierce Biotechnologies) was used. One-hundred nanograms of protein were added to a 5-μL reaction mixture and incubated for 20 min at room temperature. The binding mixture was resolved on 6% native polyacrylamide gels. The gels were transferred to Bio-dyne B nylon membranes (Pierce Biotechnologies) using Western blot techniques and detected using chemiluminescence.

Microarray data for Mtf2 knockdown ES cells and trophoblast stem (TS) cells

Two Mtf2 knockdown ES cell samples and two samples with control vectors were analyzed with Agilent Whole Mouse Genome Oligo Microarray Kit, representing over 41,000 mouse genes and transcripts. These expression data were submitted to GEO under accession no. GSE18319. Affymetrix MU74Av2 microarray data for ES and trophoblast stem (TS) cells were obtained from the GEO database (http://www.ncbi.nlm.nih.gov/geo/, accession no. GSE3766).

Acknowledgments

We thank the Douglas Melton lab for initiating this collaboration by supplying the human IVF embryos, all duly obtained by full review of an ESCRO committee and IRB with certified consent forms. We thank Carole Wilson and Mark Band for microarray experiments. We thank Tetsuya Tanaka, Yanyon Li, and Wei Huang for useful discussions. This work is partially supported by NSF DBI 08-45823 (S.Z.) and USDA Agricultural Research Service AG 58-1265-7-317 (H.A.L.).

References


Xie et al.
Rewirable gene regulation in preimplantation


Received September 11, 2009; accepted in revised form March 1, 2010.