Phase separation during mouse early embryonic development and underlying genetic and epigenetic correlations

Hui Quan[1,†], Sirui Liu[1,†], Yu Zhang[2,3], Wei Xie[2,3*], Yi Qin Gao[1,4*]

[1] Beijing National Laboratory for Molecular Sciences, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China.

[2] Center for Stem Cell Biology and Regenerative Medicine, MOE Key Laboratory of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing, China.

[3] THU-PKU Center for Life Sciences, Tsinghua University, Beijing, China.

[4] Biomedical Pioneering Innovation Center (BIOPIC), Peking University, Beijing 100871, China.

*Corresponding author. E-mail: xiewei121@tsinghua.edu.cn; gaoyq@pku.edu.cn

†These authors contributed equally to this work

ABSTRACT

Chromatin undergoes drastic organization and epigenetic reprogramming during embryonic development in mammals. However, the relationship among global structural change, epigenetic reprogramming, and functional implementation is largely unknown. Based on the analysis of latest published Hi-C data of post-implantation stages, we present a consistent view of the chromatin structural change and the corresponding sequence dependence. Two types of sequentially, genetically and transcriptionally distinct domains, forests and prairies, show systematic and overall increase of spatial segregation during embryonic development, but with notable mixing occurring at two stages, ZGA and implantation. The segregation level change largely coincides with the change of genetic and epigenetic properties. Detailed gene functions in specific phase-changing domains during implantation were analyzed, based on which a possible mechanism of functional realization during implantation was proposed. Interestingly, body temperature changes coincide with the change in chromatin segregation, implying that temperature is a possible factor influencing global chromatin structure.

KEYWORDS

Forests and Prairies, Hi-C, Segregation, ZGA, Implantation, Temperature.

INTRODUCTION

In mammals, chromatin undergoes drastic organization and epigenetic reprogramming after fertilization. These processes are essential for gene regulation either globally or specifically by generating chromatin environment that is permissive or repressive to gene expression[1]. Chromatin of zygotes and two-cell embryos has obscure high-order structures, existing in markedly relaxed states, which undergoes consolidation of TADs and segregation of chromatin compartments through development[2,3]. Along with remodeling of 3D chromatin architecture, genome-wide epigenetic reprogramming also takes place during embryonic development[4–6]. DNA methylation levels, as well as histone modifications and small RNAs, are globally reset[7–9], leading to the specification of the germ layers and cell differentiation.

The progression of the mammalian embryo from fertilization to germ layer formation, concurrent with transcriptional changes and cell fate transitions[1], involves an ordered series of hierarchical lineage determination that ensures the establishment of a blueprint for the whole animal body. One of the most notable transcriptional changes is the zygotic genome activation(ZGA), when the embryo changes from a state where there is little transcription to a state in which thousands of genes are transcribed[10]. ZGA is mechanistically coordinated with changes in chromatin state and cell cycle[11]. Mammalian ZGA may primarily prepare for the differentiation of the inner cell mass (ICM) and the trophectoderm (TE), which begins at the morula stage, so that the TE can develop to the extraembryonic tissue necessary for embryos to implant and to receive nutrients. After implantation, the ICM then gives rise to three germ layers (ectoderm, mesoderm and endoderm) through gastrulation, generating founder tissues for subsequent somatic development[12,13].

Thanks to recent developments in low-input chromatin analysis technologies, chromatin structural, epigenetic and transcriptional properties have been roundly explored in the early embryonic development process[14]. Taken that chromatin structural properties at the pre-implantation stage have been investigated, relevant analysis on the latest Hi-C data describing chromatin structural properties at post-implantation stages may provide a fairly complete view of the structural change during development. On the other hand, the relationship between global structural changes, epigenetic reprogramming

and functional implementation remains largely unknown. Furthermore, the DNA sequence dependence in these processes has been rarely discussed.

In our previous study, we analyzed the DNA sequence dependence in the formation of 3D chromatin structures for different cell types[15]. Based on CpG island (CGI) densities, the genome was divided into alternative forest (high CGI density) and prairie (low CGI density) domains with average lengths of 1 to 2 million bases. CGI forests and prairies largely improve the effectiveness in separating the linear DNA sequence into domains with distinctly different genetic and epigenetic properties, and the correspondence of sequential differences to 3D chromatin structure differences (including TADs and compartments). More interestingly, the two domains spatially segregate to different extents according to cell types. Along with the progressive maturation of the higher-order chromatin architecture during embryonic preimplantation development, the two domains show an increased level of segregation from each other[2,3,15]. We pointed out that the continuous mitosis resembles a repeated annealing process for the chromatin structure to evolve towards states which probably adapt to the local cellular environment. F-P segregation also increases during cell differentiation. However, the segregation degree of some somatic cells is less than that of ICM, implying deviation from a monotonic increase of segregation in the change from zygote to differentiated cell types. How the chromatin conformation changes gradually to promote cell identity establishment during development is thus an interesting open question.

In the present study, we conducted an analysis on global chromosome structural changes during embryonic post-implantation development, and present a consistent view for the chromatin structural change during early embryonic development from zygote to the post-implantation stage. Two specific stages, ZGA and implantation during which gene activation and lineage specification occur, respectively, were both found to involve mixing of the two types of genomic domains, with the latter showing a more significant change than the former. The segregation level change coincides with the change of the overall size of compartment B as well as the fraction of prairie in it, the larger value of which means a more segregated chromatin structure. The epigenetic properties in early embryonic development are also correlated with the trend of domain separation in this

process, which together indicate that chromatins of endoderm and mesoderm are more segregated than that of ectoderm. Moreover, the phase separation levels of the earliest fate-committed cell types correlate with those of the differentiated cells. Detailed functions of genes residing in more mixed or segregated domains during implantation are then analyzed, based on which we proposed that pluripotent ICM cells may first differentiate to ectoderm-like epiblast via the enhanced interactions of prairies with forests. Finally, possible factors influencing global chromatin structure, such as temperature, germ layer, and cell division, are discussed.

MATERIALS AND METHODS

Definition of CGI forest/prairie and segregation ratio calculation

CGI forests are DNA domains with densely distributed CGIs, identified based on neighboring CGI distance distributions along the genome. A CGI forest was then defined as a continuous DNA region longer, and inside which all neighboring CGI distances are shorter, than a critical distance[15]. CGI prairie domains are the complements to forests excluding the longest chromosomal gap where the centromeres locate.

Based on the domain contact matrix, the inter-domain contact ratio between the same domain types was calculated as

$$R_f^s = \frac{\sum_{i,j\in F, i\neq j} D_{ij}}{\sum_{i\in F, j\in A} D_{ij}} \text{ and } R_p^s = \frac{\sum_{i,j\in P, i\neq j} D_{ij}}{\sum_{i\in P, j\in A} D_{ij}}$$

for forests and prairies, respectively. In the above equations, D_{ij} represents the sum of contacts between the two domains i and j, F is the collection for all forest domains, P is the collection for all prairie domains, and A is the union of sets F and P.

The inter-domain contact between different types was calculated similarly as

$$R_f^d = \frac{\sum_{i\in F, j\in P} D_{ij}}{\sum_{i\in F, j\in A} D_{ij}} \text{ and } R_p^d = \frac{\sum_{i\in P, j\in F} D_{ij}}{\sum_{i\in P, j\in A} D_{ij}}$$

The segregation ratio R_s was defined as the ratio of inter-domain contacts between the same types and different types

$$R_s = \frac{R_f^s}{R_f^d} \text{ or } R_s = \frac{R_p^s}{R_p^d}$$

for forests and prairies, respectively.

F-P methylation level difference

The methylation level of the open sea (defined as the genomic regions excluding CGIs, CGI shores and CGI shelves[16]) is expected to better reflect the environmental chromatin state than other regions, as they lack specificity and gene specific regulation. Thus, we quantify the methylation difference (see Ref. 15) between adjacent forests and prairies by

$$d_i = \left(q_i - \frac{q_{i-1} + q_{i+1}}{2}\right) / \left(\frac{q_{i-1} + q_i + q_{i+1}}{3}\right)$$

where q_i, $q_{i-1}$and q_{i+1} are the average open sea methylation levels for the *i*th domain and its two flanking domains.

Gene function analysis

We analyzed the functional enrichment for genes located in various selected regions by ClusterProfiler and DAVID, both yielding similar results. The results were demonstrated via GO terms with p-values.

Identification of lineage specific genes

To identify lineage specific genes, we used a Shannon-entropy-based method, as previously described[17]. Genes with entropy score less than 1.7 were selected as candidates for stage-specific genes. Among them, we selected E6.5 epiblast specific genes satisfying the following conditions: the gene is highly expressed in E6.5 epiblast (FPKM>=1), its relative expression is higher than 1/7, and its mean expression of epiblast is higher than that of ICM. These genes were then reported in the final lineage specific gene lists.

Significantly more strongly segregated or more mixed domains

To identify forest or prairie domains that are significantly more or less segregated in one tissue when compared to another, we analyze the difference between the R_s values of the two samples. To identify domains of significantly different segregation levels in two samples, we first calculated the ratio between the two R_s values obtained from the Hi-C data of two ICM replicates from the same laboratory (denoted as R_s(ICM)). If for a domain, the R_s ratio between one examined sample and another is larger than 97.5% (smaller than 2.5%) of R_s(ICM), the domain in the former is considered to be more segregated(mixed) than that in the latter. The 2.5% and 97.5% of R_s(ICM) are denoted as R_s^L and R_s^U, respectively.

For each pair of states considered, we calculated the ratio of R_s between the two states for each forest and prairie domain. Significantly more strongly segregated domains were identified as those with the ratio of R_s higher than R_s^U (97.5% percentile), and

significantly more mixed domains were defined as those with the ratio of R_s lower than R_s^L (2.5% percentile). For convenience, we denoted more strongly segregated and more mixed domains in forest and prairie as F_{seg}, P_{seg}, F_{mix}, and P_{mix}, respectively.

Overall and stringent relative segregation ratios

Overall and stringent relative segregation ratios (R_{or} and R_{sr}) of one tissue over another are identified:

$$R_{or} = ln\left(\frac{N_{ij}^o}{N_{ji}^o}\right) \text{ and } R_{sr} = ln\left(\frac{N_{ij}^s}{N_{ji}^s}\right)$$

respectively. Here N_{ij}^o is the number of domains with a higher R_s ratio for tissue *i* than tissue *j*, and N_{ij}^s is the number of significantly more segregated domains for tissue *i* over tissue *j*. A positive value of R_{or} or R_{sr} indicates that the chromatin of the former tissue is more segregated than that of the latter. The parameters R_{or} and R_{sr} can generally and strictly reflect domain segregation, respectively.

RESULTS

Phase separation trend during mouse early embryonic development

In our previous study[15], based on the division of the mouse genome into forest and prairie domains, we defined the segregation ratio R_s as the inter-domain contact ratio between domains of the same types and different types (see methods). A higher R_s for a sample indicates a higher degree of segregation of same type-genomic domains. Based on the small-scale in situ Hi-C (sisHi-C)[2,14] data, the calculated R_s for different stages during the early embryonic development of mouse is given in Fig. 1A. As seen from this figure, the segregation ratio R_s generally increases when the early embryonic cells develop from zygotes to the germ layer stage, but two dips are observed at the early to late 2-cell stage and the ICM to E6.5 epiblast stage, respectively. The decrease is more significant during the second stage than the first one for prairie domains, while for forest domains the second one is less pronounced. Since the change of R_s provides a description for the variation of global domain segregation in early embryonic development, these results show that the degree of domain segregation increases during the embryonic development from zygote to E7.5 ectoderm, except for the early to late 2-cell and ICM to E6.5 epiblast transitions. Interestingly, the former process corresponds to ZGA[18], and the latter coincides with the embryo implantation into the uterus[19]. These results suggest that the deviation from a monotonic increase of segregation during the development from zygote to differentiated cells is mainly caused by chromatin domain mixing during ZGA and implantation. Except for these two stages, the chromatin generally becomes more sequentially segregated during the embryonic development. Correspondingly, the average gene expression levels increase during both of these two stages (Supplementary Fig. S1). The domain-level F-P chromatin mixing suggested by decrease of R_s is possibly related to gene activation as well as an environmental change in implantation. We try to obtain more details on these two phase-mixing events in the following analysis.

It is well known that the entire genome can be partitioned into two spatial compartments (A and B)[20]. It was also found that forest domains reside mainly in compartment A and prairie domains mainly in compartment B for differentiated cells[15]. To understand how forests and prairies' positioning in compartments affects segregation, we divided the genome into four components, Af, Bf, Ap and Bp, according to the positioning

of forests and prairies within compartments A and B, and investigated how the four components evolve in early embryonic development. During development, the proportion of Af is nearly constant, whereas that of Bp varies greatly with developmental stages. Interestingly, the latter changes in the same trend as the level by which chromatin segregates, as seen in Fig. 1B. For all the stages analyzed, zygote at the PN3 stage has the smallest segregation ratio and also the lowest proportion of Bp. The Bp ratio increases during the transition from PN3 to ICM stages, except for the early to late 2-cell stage. ICM to E6.5 epiblast also exhibits a decrease in the size of Bp. When E6.5 epiblast develops to E7.5 ectoderm, the size of Bp appears to slightly increase. The higher degree the global domain segregation of chromatin, the greater size of Bp is observed. Consistently, we found that the size of compartment B also exhibits the same trend of variation (Supplementary Fig. S2). On the other hand, the forest contribution to compartment B remains to be small and largely invariant during embryonic development. The small proportion of Bf domain signifies that forests are mostly at active or poised states.

To gain more information on the sequential properties of these four components (especially to compare Af with Ap, and Bf with Bp), we calculated their GC contents and CpG densities during development (Fig. 1C, Supplementary Fig. S3). Af and Bp have consistently highest and lowest GC contents among the four components, respectively, signifying the role of forest and prairie domains in chromatin compartmentalization. Notably, the GC content of the Ap component changes in the same trend as the segregation degree. Since the proportion of Bp and the segregation level are highly correlated, when the degree of segregation increases, a higher portion of prairie domains partition into compartment B, leaving the remaining prairie sequences in compartment A more “bivalent” (low GC content regions of forests or high GC content regions of prairies[15]). Moreover, both forest components (Af and Bf) have higher GC contents than the two prairie components (Ap and Bp) before the 2-cell stage, whereas the two compartment A components (Af and Ap) have higher GC contents than the two components of compartment B (Bf and Bp) after the 2-cell stage. The analysis on CpG densities yielded similar results. These results indicate that compartmentalization become more prominent after the 2-cell stage, which implies the increased establishment of three-dimensional structures. We also note here that the dominant contribution of forests and prairies to

compartments A and B, respectively, manifests the trend of DNA segregation into spatially separated domains of similar sequence properties.

Relationship between epigenetic marks and structures

In the earlier study we showed that differences in the methylation levels of forests and prairies correlate with chromatin spatial packing[15]. In the following, we show that the open sea methylation difference between forest and prairie domains in different cell types corresponds to their difference in segregation degrees. During embryonic development, the absolute value of the F-P open sea methylation difference increases from 2-cell to ICM stage, as seen in Fig. 2A, consistent with the trend of chromatin structural segregation. During the further development to the E7.5 stage, this methylation difference also increases. Specifically, for the three germ layers, the methylation difference for endoderm is greater than that of mesoderm and ectoderm, which would suggest the forest-prairie domains of endoderm to be more segregated than those of mesoderm and ectoderm. The correspondence between methylation difference and segregation degree during embryonic development further suggests a connection between this epigenetic mark and the chromatin structure. Since forests are the main contributor to the more accessible chromatin regions, they are presumably more prone to both DNA demethylation and methylation than prairies.

In addition to the chromatin segregation level discussed above, it is interesting to examine whether F-P open sea methylation differences of differentiated tissues show a germ layer dependence. We therefore analyzed average methylation difference levels of 68 mouse differentiated cells and 45 human differentiated cells (19, 16, 33 cells originating from endoderm, mesoderm and ectoderm respectively for mouse and 13, 19, 13 for human)[21–23]. As shown in Fig. 2B and Supplementary Fig. S4, the absolute values of methylation difference for ectoderm-derived tissues are generally smaller than those for mesoderm/endoderm-derived tissues, in the same order as the segregation degree for the three germ layers. These results all suggest that the epigenetic marks do reflect to a large degree the phase segregation level for the various early development stages, and the

orders of segregation degree for the three germ layers are in agreement with those of differentiated tissues.

ZGA and implantation related domain mixing

Based on the Hi-C data of two ICM replicates[2], we identified threshold values of R_s ratio variation ($R_s^U = 1.167$ and $R_s^L = 0.878$, respectively, see methods, Supplementary Fig. S5). As ZGA mainly occurs at the late 2-cell stage[24], we calculated R_s ratio between early and late 2-cell embryos for each forest and prairie domain (Fig. 3A). We also calculated the difference of R_s value between ICM and E6.5 epiblast stages, during which implantation takes place (Fig. 3B). Both processes are accompanied by a general lessening of the domain segregation. By comparing to R_s^L or R_s^U defined earlier, we identified forest or prairie domains that become either significantly more strongly segregated or mixed (for convenience, denoted as F_{seg}, P_{seg}, F_{mix}, P_{mix} respectively, see methods).

Next, we examined the relation between the domain segregation and compartmentalization. During implantation, for the four types of components defined above, we calculated their corresponding change of the compartment vector index during the ICM to E6.5 epiblast transition. Along with this transition, the majority of F_{seg} P_{mix} domains (83.1% and 86.9%, respectively) observe an increase of the compartment vector indices (becoming more compartment A-like), whereas for 50.6% of F_{mix} and 42.4% of P_{seg} domains, the indices decrease (Fig. 3C). These results show that, along with implantation, the interactions increase between forests and prairies and compartment B significantly downsizes. During ZGA, 71.4% of F_{seg} and 62.3% of P_{mix} domains show an increase of the compartment vector indices, whereas for 57.2% of F_{mix} and 72.6% of P_{seg} domains, their indices decrease (Supplementary Fig. S6). For both processes, when chromosomal domains tend to have more contact with forests (F_{seg} and P_{mix}), they tend to have an increased compartment vector index (to become more likely resided in compartment A). In contrast, a domain with an increase contact with prairie (F_{mix} and P_{seg}), is more likely to have their compartment vector index reduced.

Since the interaction and partition of the forest and prairie have been shown to relate to gene activation/repression, we analyzed next the functional enrichment for genes located

in each of the four regions using ClusterProfiler[25]. In both ZGA and implantation, a higher fraction of domains and genes change to more mixed than to more strongly segregated states. In ZGA, the majority of the significantly more mixed domains are forests. The functional assignments of the associated F_{mix} genes during ZGA by Gene Ontology (GO) terms are characterized by cellular immunity represented by GO terms such as "STAT protein", " immune response", "B cell proliferation" and "T cell activation" (Fig. 3D). Genes located on the other types of domains, F_{seg}, P_{mix}, P_{seg} regions, are enriched in terms of metabolic process, cellular response to amino acid and cellular differentiation, respectively (Supplementary Fig. S7).

In contrast, in implantation the majority of the DNA domains become pronouncedly less segregated are prairies. The associated P_{mix} genes are characterized by "sensory perception of taste", "regulation of lactation", "keratinocyte differentiation", "epidermal cell differentiation", "epidermis development", "skin development", "mammary gland development" and "long-term synaptic potentiation" (Fig. 3E). Interestingly, these terms are related to mammary glands, epidermis and nervous system that are all differentiated from ectoderm[26], indicating that the mixing of prairie domains with forests during implantation is highly pertinent to ectoderm differentiation. Moreover, genes lying in F_{mix}, F_{seg} and P_{seg} domains are related to cell cycle, cellular immunity and metabolic process, respectively (Supplementary Fig. S8).

To confirm the roles of segregation change on functional implementation, we further analyzed the segregation change of lineage-specific genes. We first found 552 lineage-specific genes for E6.5 epiblast stage (see methods), of which 474 genes located at forest domains and 78 genes prairie domains. When mapping these genes to four segregation-changing domains, we found that 97% of the lineage-specific genes in prairie domains belong to P_{mix}, whereas only 16.4%, 16.7% of forest lineage-specific genes are located at F_{seg} and F_{mix} regions, respectively. On the other hand, the logarithm of segregation ratio changes for each gene $ln\frac{R_{s_ICM}}{R_{s_E6.5}}$ shows that segregation factors of most prairie genes decrease for the ICM to E6.5 epiblast transition (Fig. 3F). These results suggest that the mixing of prairies into forests to play an important role during implantation for E6.5 epiblast and further developmental stages.

Phase segregation of germ layers

The chromosome domain segregation level of E7.5 ectoderm is similar to that of E6.5 epiblast (Fig. 4A), suggesting that chromatin takes at an ectoderm-like state shortly after implantation. The above analysis showed that during implantation the prairie domains containing many ectoderm differentiation relevant genes establish interactions with forests. The average expression level of more strongly segregated prairie genes decreases when ICM develops to E6.5 epiblast, whereas that of more mixed ones increases (difference values of the logarithms of gene expression are -0.4285 and 0.0063, respectively). It implies that many ectoderm related genes become less segregated during the ICM to E6.5 epiblast transition, and increase expression. The similarity between E6.5 epiblast and E7.5 ectoderm found here is consistent with earlier studies which showed the post-implantation epiblast cells to subsequently develop towards the ectoderm[27].

It was shown earlier that forests and prairies segregate to different extents in different lineages. For example, they are more spatially segregated in liver and spleen than in brain cells[15]. Consistently, liver, spleen and brain are originated from endoderm, mesoderm and ectoderm, respectively. As ectoderm chromatin has a weaker segregation than that of the endoderm (as suggested by epigenetic properties mentioned above), we wonder whether the degree of segregation for differentiated cells could also be dependent on the germ layer, that is, cells originated from endoderm is likely to have a more highly segregated chromosome than those originated from ectoderm. We then analyzed Hi-C data for 14 human tissues[28] and compared the segregation ratio of these tissues derived from different germ layers. To quantify the difference between two tissues, we defined and calculated both overall and stringent relative segregated ratios (see methods) of one tissue over another, a larger value of which represents that chromatin in the former tissue is more segregated than that in the latter. For tissues belong to ectoderm (cortex and hippocampus), their overall relative segregated ratios over those derived from endoderm and mesoderm are indeed generally negative, indicating the chromatin of the latter to be segregated than that of the former. Similarly, the overall relative segregation ratio of tissues derived from mesoderm over those derived from endoderm is also mainly less than 0 (Fig. 4B and Supplementary Fig. S9). Stringent relative segregation ratio gives

similar results, which are mainly negative (Fig. 4B and Supplementary Fig. S9). These results are all consistent with the notion that the segregation level of a differentiated cell shows a rough germ layer signature.

DISCUSSION

Developmental mechanism from a phase separation/chromatin segregation perspective

We try to provide in this study a description on the chromosome structural change during the development from zygote to E7.5 ectoderm. We found that during this process the extent of phase separation generally increases (The increasing trend was found earlier to continue in cell differentiation and senescence[15].), but phase mixing does occur at two important stages, ZGA, in which zygotic genes activate to express, and implantation, in which differentiation starts to form germ layers. The functional analyses of genes showed that genes locating at the more mixed prairie domains are related to ectoderm differentiation, which account for the majority of phase-changing genes. The trend of these prairie genes mixing into forest domains results in an ectoderm-like E6.5 epiblast. This analysis suggests that pluripotent ICM cells first differentiate to ectoderm-like epiblast state via prairie mixing. On the other hand, the chromatin domains are implied to be more segregated in mesoderm and endoderm formation (Fig. 4C).

Possible factors influencing phase separation

As we have suggested before[15], temperature may affect chromatin domain segregation, as it can sensitively affect phase separation. An increase of temperature was argued to favor the 'precipitation' of the prairies both dynamically and/or thermodynamically, leading to a stronger F-P segregation, and a lower temperature may lead to a more forest-prairie mixed chromatin structure. Such a change was indeed observed for rice genome[29].

Interestingly, during early embryonic development, the variation of BBT (basal body temperature) of human corresponds nicely to the change in domain segregation (Supplementary Fig. S10). Firstly, BBT rises by at least 0.2℃ immediately after

ovulation and remains high even after implantation. Correspondingly, the segregation ratio R_s increases along with the development from zygote to ICM, except for the early to late 2-cell stage. This stage is characterized by ZGA, and interestingly corresponds closely to a small temperature dip nearly two days after ovulation. This temperature dip at ZGA may be beneficial for the chromatin to become less condensed and thus more mixed. Researchers did find that an ultimately-successful pregnancy is characterized by significantly more min/day of relative low CBT (core body temperature) in the second day after mouse pairing which rightly correspond to ZGA, than for pregnancies which subsequently terminated prematurely[30].

Moreover, a one-day implantation dip of BBT occurs nearly one week after ovulation. Interestingly, a domain mixing is observed at this stage of development, from ICM to E6.5 epiblast. As is widely known, implantation dip is effective in detecting pregnancy in humans[31]. It has been found that continuous BBT measurement reliably allows for the earliest and most accurate detection of pregnancy[30]. These observations give rise the possibility that temperature shifts are closely coupled with large scale change of chromatin structural segregation during embryonic development.

For other species, temperature is also found to be important for embryonic development. One interesting example is from examination of the BBT of hibernating bears. The body temperature of the pregnant female bears decreases at the implantation stage, but remains 3-4 degrees higher and significantly more stable than non-pregnant females or males[32], indicating the importance of a high and steady temperature in embryo development. In addition, researchers also reported chromosome de-condensation in rice seedlings underwent cold stress[29]. All these results are consistent with the postulation that an increase of temperatures is favorable for the chromatin structure segregation, and imply that temperature could be a non-negligible factor influencing global chromatin structure, thus a possible parameter for influencing chromatin reprogramming or modification.

Besides temperature, the phase separation levels of the earliest fate-committed cell types are also correlated with those of the differentiated cells. Tissues differentiated from endoderm generally have more segregated chromatin structures than those from ectoderm. Such a difference between ectoderm and endoderm can be seen as early as from an E7.5

embryo. Moreover, researchers recorded the tissue temperature of interscapular brown adipose tissue (IBAT), liver, and the cortical (T_{IBAT}, T_L, T_{CO}, respectively), and found that T_{CO} was significantly lower than T_{IBAT} and T_L, and T_{IBAT} was slightly lower than T_L[33]. The order of temperature is consistent with that of segregation level. For these samples, chromatins at high temperatures are more likely to have a segregated structure. The degree of phase separation appears to increase in the order of cortical, IBAT and liver, which are originated from ectoderm, mesoderm and endoderm respectively, also consistent with degree of phase separation of the different germ layers. Although much more data is needed and many more factors are likely involved in affecting the chromatin structure, the cause of such correlations is worth of more detailed investigations.

Cell division is another factor that might influence chromatin domain segregation. The establishment of TAD structure was reported to require DNA replication but not zygotic transcription[2,3], implying that cell division could play a role in 3D structural formation. On the other hand, the current and earlier analyses[15] indicate that the DNA methylation can reflect the chromatin conformation and PMD, the age related hypomethylation, appears to track the accumulation of cell divisions[34]. We speculate that along with PMD hypomethylation, prairie domains tend to accumulate more strongly in the less active spatial domains (heterochromatin)[35]. But as discussed above chromatin from different germ layers segregates to different levels (endoderm is likely to be more strongly segregated than ectoderm), different tissues could show very different age dependence of domain segregation and thus different epigenetic changes. In fact, the forest-prairie methylation level difference for brain decreases with aging, while the trend is the opposite for endodermic and mesodermal cells (Supplementary Fig. S11).

CONCLUSION

In this study, we cognized chromatin structural and epigenetic reprogramming in mouse early embryonic development based on the sequence-based phase separation model, and correlated the overall chromatin structural change with functional implementation during implantation. Two distinct domains gradually separate from each other along embryonic development, but they show a tendency of mixing when

transcription and implantation happen. Increased interactions between the two types of domains are accompanied by body temperature changes, which indicates the start of transcription or lays the foundation for further cell differentiation from lineage commitment. Genes changing segregation states show lineage relevant functions and important roles in corresponding process, thus more detailed analyses on gene expression and regulation of transcription factors are in need for understanding the related molecular mechanisms, and more importantly, to predict biological functions at the molecular level from the perspective of phase separation.

Supplemental Information is available in the online version of the paper.

Acknowledgement

This work was supported by National Natural Science Foundation of China [21573006, 21233002 and 91427304 to Yi Qin Gao] and the National Key Research and Development Program of China [2017YFA0204702 to Yi Qin Gao].

Author Contributions

Wei Xie provided the sisHi-C data of post-implantation stages. Yi Qin Gao designed the research. Hui Quan and Sirui Liu performed the research. Hui Quan, Sirui Liu and Yi Qin Gao wrote the manuscript. Yu Zhang contributed to the manuscript writing.

Competing Interests

The authors declare that they have no competing interests.

References

1. Burton, A. & Torres-Padilla, M.-E. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 723–735 (2014).
2. Du, Z. *et al.* Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 (2017).
3. Ke, Y. *et al.* 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell* **170**, 367-381.e20 (2017).

4. Xu, Q. & Xie, W. Epigenome in Early Mammalian Development: Inheritance, Reprogramming and Establishment. *Trends Cell Biol.* **28**, 237–253 (2018).
5. Morgan, H. D., Santos, F., Green, K., Dean, W. & Reik, W. Epigenetic reprogramming in mammals. *Hum. Mol. Genet.* **14 Spec No 1**, R47-58 (2005).
6. Burton, A. & Torres-Padilla, M.-E. Epigenetic reprogramming and development: a unique heterochromatin organization in the preimplantation mouse embryo. *Brief. Funct. Genomics* **9**, 444–454 (2010).
7. Zheng, H. *et al.* Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Mol. Cell* **63**, 1066–1079 (2016).
8. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
9. Lee, H. J., Hore, T. A. & Reik, W. Reprogramming the Methylome: Erasing Memory and Creating Diversity. *Cell Stem Cell* **14**, 710–719 (2014).
10. Lu, F. *et al.* Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell* **165**, 1375–1388 (2016).
11. Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic Genome Activation in Vertebrates. *Dev. Cell* **42**, 316–332 (2017).
12. Arnold, S. J. & Robertson, E. J. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **10**, 91–103 (2009).
13. Lawson, K. A., Meneses, J. J. & Pedersen, R. A. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Dev. Camb. Engl.* **113**, 891–911 (1991).
14. Zhang, Y. *et al.* Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nat. Genet.* **50**, 96–105 (2018).
15. Liu, S. *et al.* From 1D sequence to 3D chromatin dynamics and cellular functions: a phase separation perspective. *Nucleic Acids Res.* gky633–gky633 (2018). doi:10.1093/nar/gky633
16. Sandoval, J. *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6**, 692–702 (2011).
17. Xie, W. *et al.* Epigenomic Analysis of Multilineage Differentiation of Human Embryonic Stem Cells. *Cell* **153**, 1134–1148 (2013).
18. Svoboda, P., Franke, V. & Schultz, R. M. Chapter Nine - Sculpting the Transcriptome During the Oocyte-to-Embryo Transition in Mouse. in *Current Topics in Developmental Biology* (ed. Lipshitz, H. D.) **113**, 305–349 (Academic Press, 2015).
19. Bedzhov, I. & Zernicka-Goetz, M. Self-Organizing Properties of Mouse Pluripotent Cells Initiate Morphogenesis upon Implantation. *Cell* **156**, 1032–1044 (2014).
20. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
21. Schultz, M. D. *et al.* Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015).
22. Lister, R. *et al.* Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* **341**, 1237905 (2013).
23. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
24. Hamatani, T., Carter, M. G., Sharov, A. A. & Ko, M. S. H. Dynamics of Global Gene Expression Changes during Mouse Preimplantation Development. *Dev. Cell* **6**,

117–131 (2004).
25. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).
26. Li, L. *et al.* Location of transient ectodermal progenitor potential in mouse development. *Development* **140**, 4533–4543 (2013).
27. Ji, X. *et al.* 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell* **18**, 262–275 (2016).
28. Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* **17**, 2042–2059 (2016).
29. Liu, C., Cheng, Y.-J., Wang, J.-W. & Weigel, D. Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis. *Nat. Plants* **3**, 742–748 (2017).
30. Smarr, B. L., Zucker, I. & Kriegsfeld, L. J. Detection of Successful and Unsuccessful Pregnancies in Mice within Hours of Pairing through Frequency Analysis of High Temporal Resolution Core Body Temperature Data. *PLOS ONE* **11**, e0160127 (2016).
31. Refinetti, R. & Menaker, M. The circadian rhythm of body temperature. *Physiol. Behav.* **51**, 613–637 (1992).
32. Friebe, A. *et al.* Factors Affecting Date of Implantation, Parturition, and Den Entry Estimated from Activity and Body Temperature in Free-Ranging Brown Bears. *PLoS ONE* **9**, e101410 (2014).
33. Dewasmes, G., Loos, N., Delanaud, S., Ramadan, W. & Dewasmes, D. Liver Temperature During Sleep. *Sleep* **26**, 948–950 (2003).
34. Zhou, W. *et al.* DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* **50**, 591–602 (2018).
35. DEEP Consortium *et al.* A comprehensive analysis of 195 DNA methylomes reveals shared and cell-specific features of partially methylated domains. *Genome Biol.* **19**, (2018).

Figures and Legends

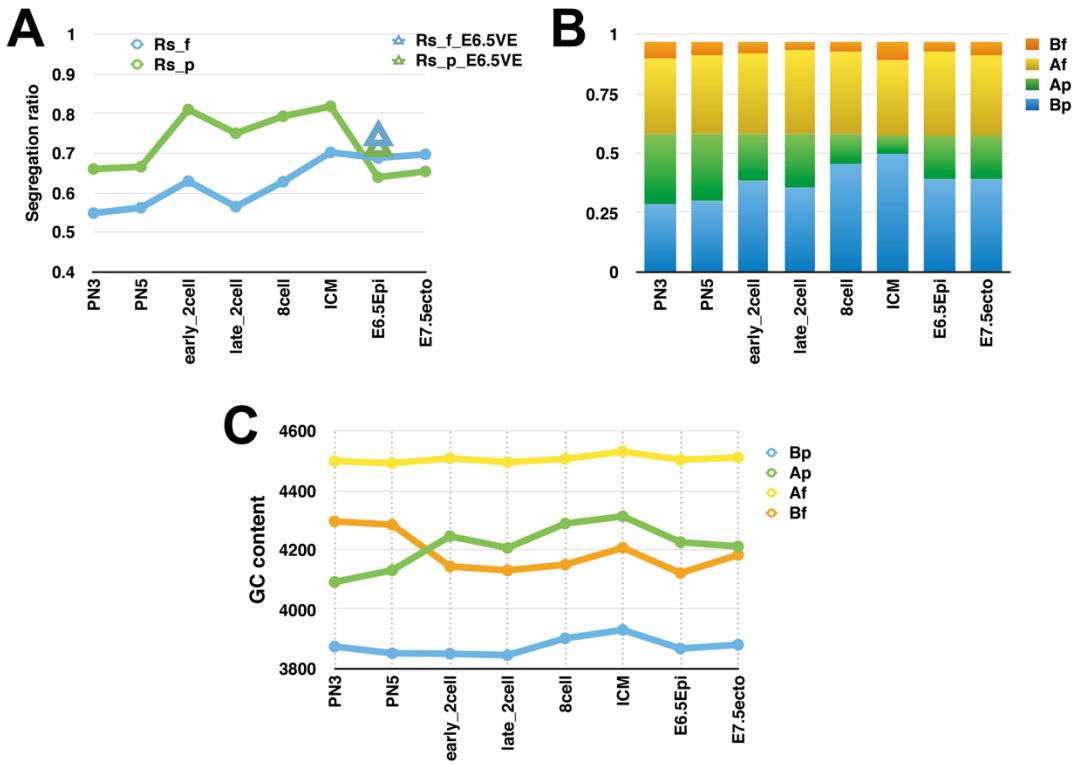

Figure 1. Phase separation trend during mouse early embryonic development.
(A) The chromatin segregation ratio for forest and prairie at each developmental stage of mouse.
(B) The proportions of four components, Af, Bf, Ap and Bp, according to the positioning of forests and prairies in compartments A and B, for each stage in early embryonic development.
(C) The average GC contents of the four components in development.

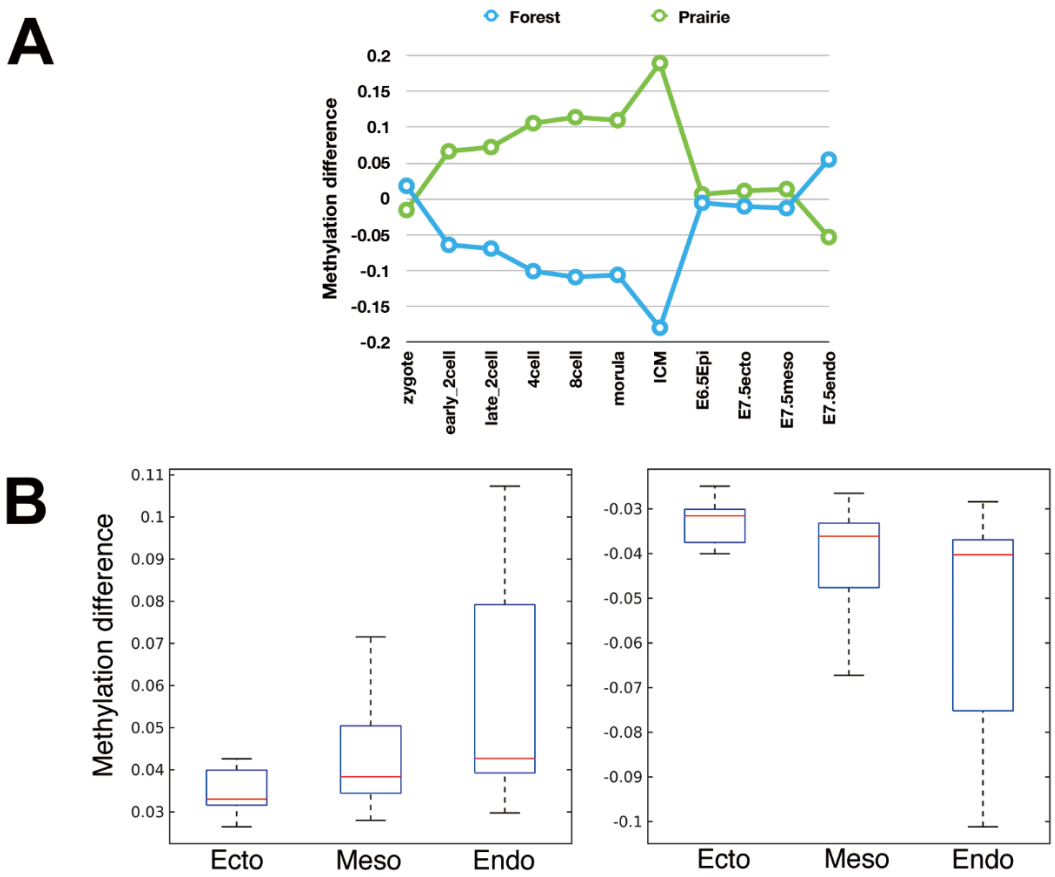

Figure 2. Relationship between epigenetic marks and structures and a schematic picture of development mechanism from ICM to E7.5 germ layers.

(A) The average forest/prairie open sea methylation level difference at different stages during early embryonic development.

(B) The box plots for average open sea methylation level difference of mouse differentiated cells originating from endoderm, mesoderm and ectoderm respectively, for forest (left) and prairie (right) domains.

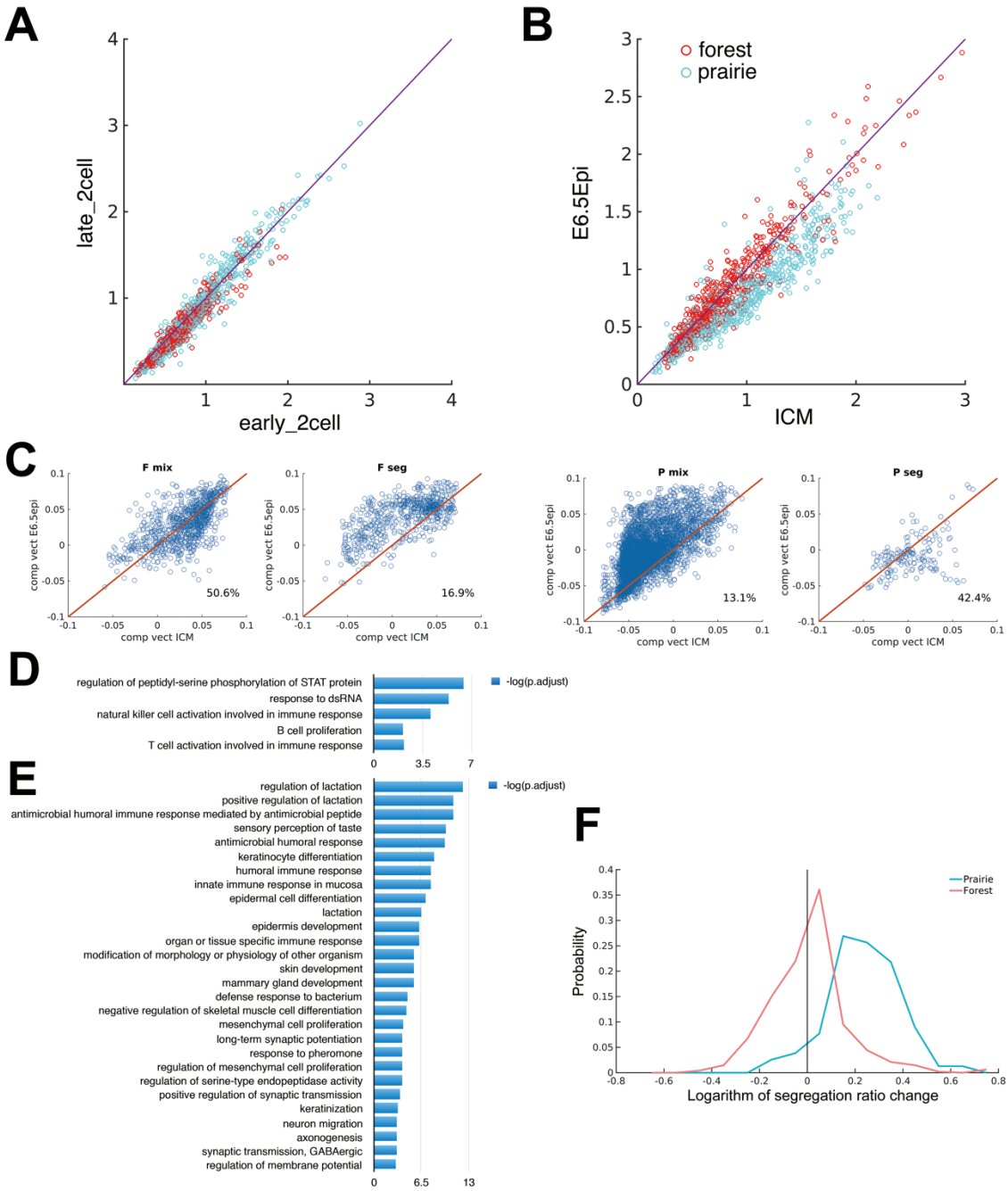

Figure 3. ZGA and implantation related domain mixing.
(A-B) The scatter plot of R_s ratio for each forest and prairie domain between early and late 2-cell embryos (A) and between ICM and E6.5 epiblast stages (B).
(C) The change of compartment vector index for the four kinds of segregated domains from ICM to E6.5 epiblast stage.
(D) Functional assignment of the F_{mix} genes during ZGA and (E) the P_{mix} genes during implantation.
(F) The probability of logarithm of segregation ratio change for forest/prairie genes from ICM to E6.5 epiblast stage.

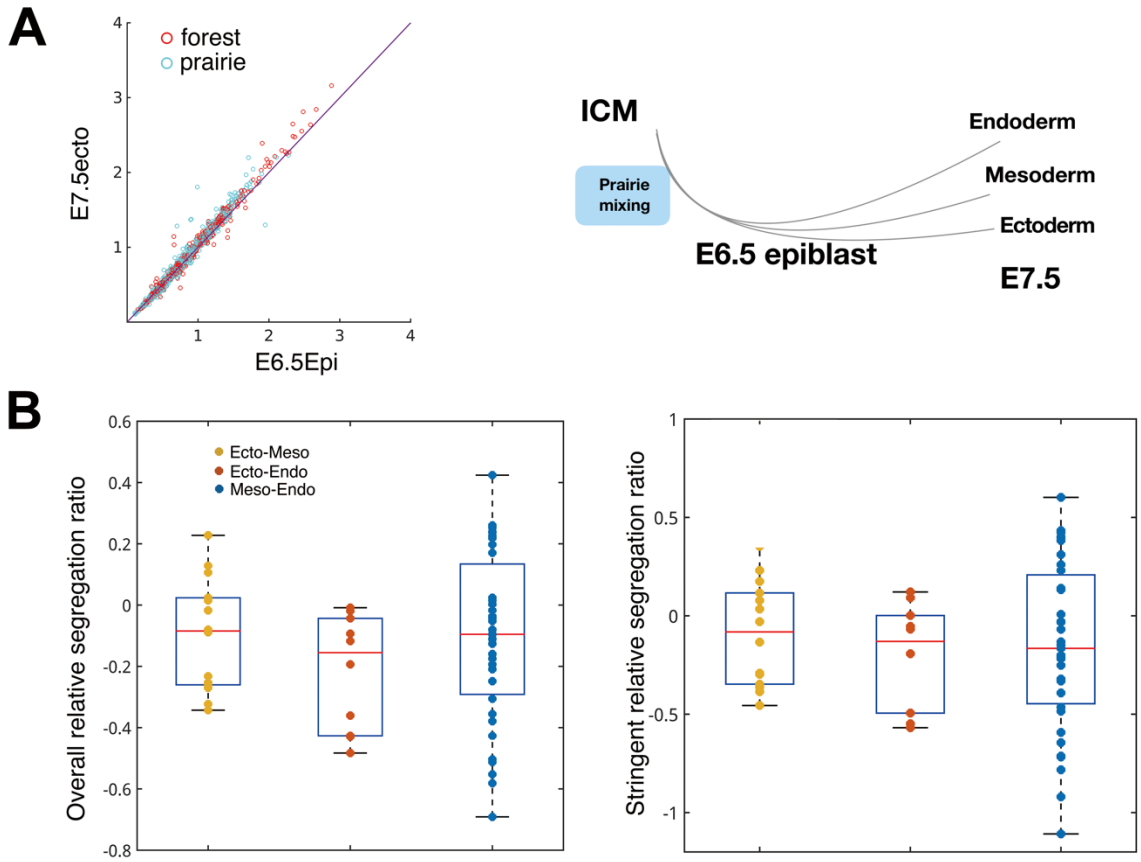

Figure 4. Germ layer formation and domain separation.

(A) The scatter diagrams of R_s ratio for each forest and prairie domain between E6.5 epiblast and three E7.5 germ layers, ectoderm (left), mesoderm (middle), and endoderm (right).

(B) Box-plots of overall (left) and stringent (right) relative segregation ratio of one tissue over another for prairies. The three categories are tissues derived from ectoderm over those derived from mesoderm, ectoderm over endoderm, and mesoderm over endoderm.

(C) A schematic picture of developmental mechanism from ICM to E7.5 germ layers.