Open access • Posted Content • DOI:10.1101/803890

Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos — Source link 🗗

Yang Wang, Yang Wang, Peng Yuan, Peng Yuan ...+12 more authors

Institutions: Chinese Ministry of Education, Peking University

Related papers:

- Single-Cell Multi-omics Sequencing of Human Early Embryos

- Chromatin analysis in human early development reveals epigenetic transition during ZGA

- The landscape of accessible chromatin in mammalian preimplantation embryos

- The role of DNA methylation in mammalian development

- Regulation of zygotic gene activation by chromatin structure and epigenetic factors.

Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos

Yang Wang[1,2,3,5], Peng Yuan[1,2,3,5], Zhiqiang Yan[1,2,3,4,5], Ming Yang[1,2,3], Ying Huo[1,2,3], Yanli Nie[1,2,3], Xiaohui Zhu[1,2,3], Liying Yan[1,2,3,*], Jie Qiao[1,2,3,4,*]

1 Beijing Advanced Innovation Center for Genomics, Center for Reproductive Medicine, Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing 100191, China;

2 Key Laboratory of Assisted Reproduction, Ministry of Education, Beijing 100191, China

3 Beijing Key Laboratory of Reproductive Endocrinology and Assisted Reproductive Technology, Beijing 100191, China

4 Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

5 These authors contributed equally.

Correspondence should be addressed to: L.Y. (yanliyingkind@aliyun.com) or J.Q. (jie.qiao@263.net)

Abstract

Extensive epigenetic reprogramming occurs during preimplantation embryo development and is accompanied by zygotic genome activation (ZGA) and first cell fate specification. Recent studies using single-cell epigenome sequencing techniques have provided global views of the dynamics of different epigenetic layers during this period. However, it remains largely unclear how the drastic epigenetic reprogramming contributes to transcriptional regulatory network. Here, we developed a single-cell multiomics sequencing technology (scNOMeRe-seq) that enables profiling of genome-wide chromatin accessibility, DNA methylation and RNA expression in the same individual cell with improved performance compared to that of earlier techniques. We applied this method to analyze the global dynamics of different molecular layers and their associations in mouse preimplantation embryos. We found that global DNA methylation remodeling facilitates the reconstruction of genetic lineages in early embryos and revealed that the gradual increases in heterogeneity among blastomeres are driven by asymmetric cleavage. Allele-specific DNA methylation pattern is maintained throughout preimplantation development and is accompanied by allele-specific associations between DNA methylation and gene expression in the gene body that are inherited from oocytes and sperm. Through integrated analyses of the collective dynamics between gene expression and chromatin accessibility, we constructed a ZGA-associated regulatory network and revealed coordination among multiple epigenetic layers, transcription factors (TFs) and repeat elements that instruct the proper ZGA process. Moreover, we found that inner cell mass (ICM)/trophectoderm (TE) lineage-associated cis-regulatory elements are stepwise activated in blastomeres during post-ZGA embryo stages. TE lineage-specific TFs play dual roles in promoting the TE program while repressing the ICM program, thereby separating the TE lineage from the ICM lineage. Taken together, our findings not only depict the first single-cell triple-omics map of chromatin accessibility, DNA methylation and RNA expression during mouse preimplantation development but also enhance the fundamental understanding of epigenetic regulation in early embryos.

Introduction

In mammals, embryo development starts from a unified zygote. Coincident with the first several zygotic cleavages, early embryos activate the zygotic genome, restore totipotency and further generate the inner cell mass (ICM) and trophectoderm (TE) during preimplantation development[1-3]. Failures in zygotic genome activation (ZGA) or ICM/TE lineage specification can cause early embryo developmental arrest and implantation failure in both mice and humans. With advances in low-input and single-cell epigenome sequencing, recent studies have revealed that extensive global epigenetic reprogramming, for instance, reprogramming of DNA methylation (Met), chromatin accessibility (Acc) and histone modifications, occurs in early embryos during this period[4-14]. However, it remains to be explored how these epigenome reconfigurations contribute to the establishment of proper regulatory networks of early embryos.

Acc is a hallmark of cis-regulatory elements (CREs), such as promoters and enhancers, that act coordinately with transcription factors (TFs) and epigenetic modifications to finely regulate the transcriptional activity of downstream genes and establish cell-type specific regulatory networks[15,16]. The currently optimized low-input open-chromatin sequencing techniques, such as the assay for transposase-accessible chromatin using sequencing (ATAC-seq) and low-input DNase I sequencing (liDNase-seq), are able to detect genome-wide dynamics of Acc in early embryos. However, the signal of the open regions reflects the average signal of the mixed sample, which may be confounded by highly heterogeneous and asynchronized blastomeres and even abnormal embryos[6,10,17,18]. Single-cell ATAC-seq detects only thousands of informative reads per cell on average, which might limit its application with the scarce resources of early embryos[19-21]. Moreover, because of the high heterogeneity of early blastomeres, a functional understanding of epigenomic changes requires knowledge of the transcriptional output from one individual cell. Recently, different single-cell epigenome sequencing methods and transcriptome sequencing methods have been combined to profile different combinations of molecular layers from the same individual cell, providing opportunities to explore the associations between different molecular layers[22-26]. However, factors compromising the quality of data from current single cell multi-omics technologies, such as poor genome coverage or low gene number detection, might constrain the precise interpretation of the

72   associations between different molecular layers.

73   Here, we describe a technique called **s**ingle-**c**ell **n**ucleosome **o**ccupancy, **me**thylome and **R**NA

74   **e**xpression **seq**uencing (scNOMeRe-seq) that effectively combines single-cell nucleosome

75   occupancy and methylome sequencing (scNOMe-seq) with Multiple Annealing and dC-Tailing-

76   based Quantitative single-cell RNA sequencing (MATQ-seq), showing improved performance for

77   profiling of multiple molecular layers from the same individual cell[4,27,28]. We applied scNOMeRe-

78   seq to analyze genome-wide Acc, Met and RNA expression (Expr) in mouse preimplantation

79   embryos at single-cell resolution and to provide a comprehensive overview of the functional

80   regulatory landscape in early embryos.

81

82   **RESULTS**

83   **scNOMeRe-seq profiles in mouse preimplantation embryos.**

84   To simultaneously detect genome-wide Acc, Met and Expr in the same individual cell, we

85   developed a single-cell multiomic sequencing method, scNOMeRe-seq, by combining scNOMe-

86   seq and MATQ-seq (Fig. 1a). We employed this method to profile 233 single cells isolated from

87   mouse preimplantation embryos at different stages with RNA data from 221 (94.8%) single cells

88   and DNA data from 218 (93.4%) single cells passed our stringent criteria, showing a high success

89   rate (Extended Data Fig. 1a, f and g; Supplementary Table 1). The DNA data showed better genome

90   coverages than those in a previous study using single-cell chromatin overall omic-scale landscape

91   sequencing (scCOOL-seq) (this study: WCG 3.49 million, 15.8%, GCH 31.0 million, 15.5% on

92   average per cell; scCOOL-seq: WCG 2.24 million, 10.1%, GCH 19.7 million, 9.8% on average per

93   cell)[4]. The RNA dataset showed high accuracy, high reproducibility, even coverage through genic

94   regions, and high detection sensitivity for genes expressed at low levels (Extended Data Fig. 1b-

95   e)[29]. More importantly, our RNA dataset could faithfully distinguish between ICM and TE cells in

96   embryonic day (E) 3.5 blastocysts (Extended Data Fig. 1h-i), further confirming the high quality of

97   our RNA data obtained from scNOMeRe-seq.

98   To detect the abnormal blastomeres in our early embryos, we analyzed the copy number variations

99   (CNVs) with the RNA data and DNA data for each individual cell. Consistent results were obtained

100   from both datasets regarding the inferred CNVs, even for the partial chromosome CNVs (Fig. 1b

101   and Extended Data Fig. 2a). Unexpectedly, we also found several parthenogenetic (PG) embryos

102   among our detected early embryos using single-nucleotide polymorphism (SNP)-separated allelic

103   reads (Fig. 1c and Extended Data Fig. 2a). Then, we sought to determine whether the embryo

104   abnormalities would cause aberrant embryo development. Notably, along the developmental

105   trajectory inferred from the RNA dataset, most PG blastomeres showed delayed development after

106   ZGA compared to that of normal and aneuploid blastomeres (Extended Data Fig. 2b-c). Although

107   only 58% of ZGA genes were activated in PG blastomeres at the 2-cell stage, the PG blastomeres

108   became more similar to the rest of the blastomeres beginning at the same stage after ZGA, indicating

109   that the PG embryos were able to go through at least partial ZGA and develop to further stages

110 (Extended Data Fig. 2b and d). The Met in PG blastomeres were clearly distinct from that in normal

111 and aneuploid blastomeres, while Acc did not differ among PG, normal and aneuploid blastomeres

112 (Extended Data Fig. 2f-g and 3a). Together, these results revealed that the aneuploid cells were able

113 to undergo preimplantation development as well as proper epigenetic reprogramming; however, the

114 PG cells showed delayed development and an aberrant Met pattern.

115 Furthermore, k-means clustering with the Acc of the transcription start site (TSS) could not

116 distinguish the abnormal cells from the normal cells; however, the cells from the two clusters

117 showed significant differences in global Acc levels and correlations between Acc and Expr at the

118 TSS regions for each stage (Extended Data Fig. 3b). Notably, the cells from the cluster with the

119 relatively higher Acc level (cluster_2) consistently showed lower correlations between Acc and

120 Expr at the TSS regions, without differences in global Met levels, and correlations between Met and

121 Expr were observed between the two clusters (Extended Data Fig. 3b). A previous study revealed

122 globally increased Acc at S phase during mitosis; such changes in Acc are not associated with

123 transcriptional regulation but are potentially linked to DNA duplication[30]. Therefore, these two

124 clusters in each stage reflected the highly asynchronous cell cycles among the blastomeres of the

125 early embryos. Furthermore, we detected the nucleosome-depleted regions (NDRs) using an

126 aggregated Acc dataset from single cells in each cluster at each stage. Regardless of the genome

127 coverage, cluster_1 (low Acc level and high correlation between Acc and Expr) exhibited more

128 NDRs than cluster_2 for each stage (Extended Data Fig. 3c-d). The NDRs in cluster_1 at each stage

129 showed greater fractions overlapping with previously defined open chromatin in early embryos than

130 those in cluster_2 (Extended Data Fig. 3e-f)[4,6,10]. Together, these results reveal that the

131 heterogeneity of Acc among blastomeres of the same stage might mainly derive from highly

132 asynchronous blastomere cell cycles and suggest that the mixed asynchronous populations may

133 increase the background noise and compromise the discovery of transcription-related Acc signals.

134 To focus on transcriptional regulation-related epigenome characteristics during preimplantation

135 development, the Acc datasets of cells from cluster_2 and the Met datasets of PG cells were removed

136 for downstream analysis. Then, we explored the dynamics and associations of different molecular

137 layers in each single cell during preimplantation development. Both unsupervised clustering and

138 principal component analysis (PCA) revealed that cells of the same stage clustered more closely

139 within each molecular layer, consistent with the findings of previous studie (Fig. 1d)[4]. The global

140 Met levels were relatively stable in earlier stages but sharply decreased at the blastocyst stage (Fig.

141 1e, Extended Data Fig. 2e). The correlations between Met and Expr in the TSS and gene body

142 regions showed the highest associations in zygotes and gradually decreased in the following stages

143 (Fig. 1e). In contrast, global Acc was drastically decreased at the 2-cell stage and restored at the 4-

144 cell stage before gradually increasing at later stages (Fig. 1f, Extended Data Fig. 2e). Notably, the

145 correlations between Acc and Expr at the TSS regions were the most positive at the 2-cell stage

146 among all preimplantation stages, coinciding with ZGA; this finding suggests that the drastic Acc

147 reprogramming at the 2-cell stage might contribute to proper ZGA (Fig. 1f).

148 **Reconstruction of genetic lineages reveals the source of heterogeneity in early embryos**

149 Given the insufficient maintenance of Met levels during mitosis in early embryos, a previous study

150 sought to reconstruct the genetic lineages of 4-cell embryos using single-cell genome-wide CpG

151 Met datasets in both humans and mice and successfully elucidated the lineages[5]. To test whether our

152 single-cell Met (WCG) datasets could be used to infer the genetic lineages of early embryos, we

153 first computed the pairwise correlations among blastomeres in each individual 4-cell or late 4-cell

154 embryo (see Methods). We repeatedly observed two pairs of cells with highly negatively correlated

155 Met levels in each individual embryo, consistent with previous findings (Fig. 2a, b and d)[5]. Then,

156 we validated that the two cells in each pair originated from the same mother 2-cell blastomere (Fig.

157 2e). Interestingly, we also observed a conserved pairwise correlation of Met levels among

158 blastomeres for each analyzed 8-cell embryo, implying that it might be possible to reconstruct

159 genetic lineages for 8-cell embryos using single-cell Met datasets (Fig. 2c and d). To verify the Met

160 correlation patterns among blastomeres at the 8-cell stage derived from the same blastomeres in 2-

161 cell and 4-cell embryos, we microinjected FITC to label one blastomere each in 2-cell and 4-cell

162 embryos and performed single-cell bisulfite sequencing (scBS-seq) for each individual cell when

163 these embryos developed to the 8-cell stage. Blastomeres in 8-cell embryos derived from the same

164 blastomeres in 4-cell embryos exhibited highly positively correlated Met levels; in contrast, 2 pairs

165 of blastomeres in 8-cell embryos derived from the same blastomeres in 2-cell embryos exhibited

166   highly positively correlated Met levels within pairs, but cells from different pairs exhibited highly

167   negatively correlated Met levels (Fig. 2f-g). Therefore, these results demonstrate that we can

168   accurately construct the full lineages from the zygote stage to the 8-cell embryo stage using single-

169   cell Met datasets.

170   Furthermore, we investigated when unified zygotes generate heterogeneity in different molecular

171   layers among blastomeres. We first computed the correlations between blastomeres within each

172   embryo (intraembryonic correlations) versus those between blastomeres from different embryos

173   (interembryonic correlations) at the same stage for each molecular layer. We found that the

174   intraembryonic correlations were consistently higher than the interembryonic correlations for each

175   molecular layer throughout the preimplantation development stages, suggesting highly

176   asynchronous development among different embryos at the same stage (Fig. 2h). Moreover, the

177   correlations in Expr levels were highest at the zygote stage and gradually decreased at later stages,

178   suggesting that the heterogeneity among blastomeres in the same embryo was generated during

179   ZGA and gradually increased with preimplantation development (Fig. 2h). We also noticed that the

180   correlations in both the Met and Acc levels were highest at the 2-cell stage, indicating that the

181   epigenome was robustly reprogrammed for each individual cell during ZGA (Fig. 2h). Leveraging

182   this lineage tracing information, we further explored the dynamics of heterogeneity between

183   daughter cells during the first three cleavages. In the transcriptome, the correlations between

184   blastomeres from the same mother cells gradually decreased during the first three cleavages,

185   whereas the correlations between blastomeres from the same mother cells at the late 4-cell stage

186   were comparable to those at the 4-cell stage (Fig. 2h). Moreover, the correlations between

187   blastomeres from the same grandmother cells were higher than those of blastomeres from different

188   grandmother cells in 8-cell embryos (Fig. 2h). These results demonstrated that asymmetric cleavage

189   might have been the major source of the transcriptome heterogeneity. Although the heterogeneity in

190   the epigenome seemed not to be associated with asymmetric cleavage, we notably observed that the

191   correlations in Met levels between blastomeres from the same mother cells were higher in 8-cell

192   embryos than in 4-cell and late 4-cell embryos, indicating that increased Met maintenance occurs

193   to some extent during DNA duplication at the 4-cell stage (Fig. 2h).

Allele-specific regulation of gene expression in early embryos

194

195   Drastic epigenetic reprogramming occurs in parental genomes after fertilization. The Acc levels of

196   the parental genomes were comparable in most individual cells throughout preimplantation

197   development (Fig. 3a). The Met level in the paternal allele was consistently higher than that in the

198   maternal allele for each individual zygote (Fig. 3b). From the 2-cell to the 8-cell stage, the global

199   differences in Met levels between parental alleles varied in the different individual cells; however,

200   after the morula stage, the Met level in the maternal allele was consistently higher than that in the

201   paternal allele for each individual cell (Fig. 3b). In addition, we found that 16.7%-29.7% of regions

202   showed significant allelic differences (FDR < 0.01) in Acc levels, and 7.9%-40.2% of regions

203   showed significant allelic differences (FDR < 0.01) in Met levels across preimplantation stages

204   (Extended Data Fig. 4a-b). The differences in allelic Acc levels were widely distributed in the whole

205   genome and showed no preference for a particular parental allele (Fig. 3c, Extended Data Fig. 4c).

206   Notably, the maternal hypermethylated regions were highly enriched in genic regions, whereas the

207   paternal hypermethylated regions were highly enriched in distal intergenic regions throughout the

208   preimplantation stages, consistent with previous findings (Fig. 3d, Extended Data Fig. 4d)[4]. Given

209   that the oocyte genome is highly methylated at actively transcribed genic regions and

210   hypomethylated at intergenic regions, while the sperm genome is highly methylated at intergenic

211   regions, our results indicate that global differences in Met levels between parental alleles in gametes

212   could be largely maintained throughout preimplantation development[4,12].

213   We next sought to determine whether the allelic epigenome differences were associated with allelic

214   transcriptional regulation. First, we overlapped the differential allelic epigenetic regions with known

215   imprinting control regions (ICRs)[31]. Four germline ICRs overlapped with our differential allelic Acc

216   regions, and all showed corresponding differential allelic Acc patterns in at least one

217   preimplantation stage (Extended Data Fig. 4e). In the other hand, six known germline ICRs

218   overlapping with differential allelic Met regions showed the expected differential allelic Met

219   patterns throughout preimplantation development, validating the accuracy of our analysis (Extended

220   Data Fig. 4f). Furthermore, we assessed the correlations between allelic epigenetic modification

221   levels and Expr in each individual cell. Both parental alleles showed similar correlation patterns

222   between allelic Acc and Expr, mimicking the overall Acc vs Expr associations (Fig. 3e). Notably,

223   we observed clearly different correlation patterns between allelic Met and Expr in parental alleles:

224   the Met levels of the paternal genome at the gene body regions showed no correlations with Expr,

225   unlike those in the maternal genome (Fig. 3f). To determine whether the allele-specific correlations

226   between Met and Expr at gene bodies were caused by inherent correlations from maternal factors,

227   we further compared the correlations between allelic Met and the Expr of maternal genes (transcript

228   per million mapped reads (TPM) $\geq$ 1 in the zygote) with nonmaternal genes. The Met levels at gene

229   body regions were clearly higher for maternal genes than for nonmaternal genes in maternal alleles

230   throughout preimplantation stages (Fig. 3g). As expected, major Met differences between parental

231   alleles were observed in maternal genes but not in nonmaternal genes (Fig. 3g). Furthermore, we

232   found that the correlations between maternal Met and Expr at gene body regions were clearly weaker

233   in nonmaternal genes than in maternal genes, indicating that the observed positive correlations

234   between maternal Met and Expr at gene body regions were mainly inherited from oocytes (Fig. 3h).

scNOMeRe-seq reveals a ZGA-associated regulome

236   To reveal ZGA-associated CREs, we measured the correlations between the Acc of each

237   promoter/distal NDR and the Expr of its corresponding ZGA gene (2-cell vs zygote, fold change $\geq$

238   4, FDR < 0.01; Supplementary Table 2) across single cells during the transition from the zygote to

239   the 2-cell stage. We found that 338 promoter NDRs and 7822 distal NDRs were positively linked to

240   301 and 2239 ZGA genes, respectively, while 356 promoter NDRs and 2728 distal NDRs were

241   negatively linked to 317 and 1226 ZGA genes, respectively (Fig. 4a-b; Supplementary Table 3).

242   Although the Met in those CREs was not significantly associated with ZGA gene expression (data

243   not shown), the overall Met levels of these positively correlated CREs were lower in 2-cell embryos

244   than in zygotes (Extended Data Fig. 5a and c). Notably, the Acc of these positively correlated CREs

245   was specifically increased in each individual cell in 2-cell embryos, but this increase was

246   accompanied by a drastic global Acc decrease during this period (Extended Data Fig. 5b and d).

247   These results suggest that robust chromatin reprogramming occurs during ZGA to remove

248   regulatory memory from gametes and rebuild the zygotic regulatory network.

249   To explore how ZGA is regulated in early embryos, we further comprehensively analyzed the

250   enrichment of repeat elements and histone modifications in ZGA-associated CREs. The positively

251   correlated CREs, but not the negatively correlated CREs, were preferentially enriched with Alu, B2,

252   B4 and ERVL repeat classes as well as active histone modifications (H3K4me3 and H3K27ac) in

253   both promoter and distal regions (Fig. 4c-d). H3K4me3 was gradually established at the majority of

254   positively correlated CRE loci from the MII oocyte stage to the 2-cell stage and was colocalized

255   with H3K27ac in both promoter and distal regions, while repressive histone modifications

256   (H3K27me3 and H3K9me3) were gradually removed from these regions (Extended Data Fig. 6a-

257   c). Moreover, the positively correlated CREs were clustered in regions enriched with active histone

258   modifications and deficient in repressive histone modifications, implying a high-dimensional

259   regulatory structure of ZGA CREs (Extended Data Fig. 6d-e). Furthermore, we investigated which

260   TFs might be responsible for the establishment of ZGA-associated CREs. Notably, both positively

261   correlated promoter CREs and distal CREs were highly enriched with *Arnt*, *Bcl6*, *Klf5*, *Nkx3-2*,

262   *Nr5a2*, *Rara*, *Rarg*, *Pitx1*, and *Thrb* motifs; however, the negatively correlated CREs showed no

263   enrichment with TFs in either promoter or distal regions (Fig. 4e). Considering the global decreases

264   in Acc during the ZGA process and the characteristics of the negatively correlated CREs described

265   above, these negative correlations seemed to simply reflect global changes in Acc rather than

266   representing repressive regulation during ZGA. We next calculated TF activity (see Methods) in

267   each individual cell (Fig. 4f). Notably, we found that *Klf4*, *Nkx3-2*, *Nr5a2* and *Rarg* showed high

268   TF activity and high expression levels in 2-cell embryos compared to zygotes (Fig. 4g). More

269   importantly, the TF activity of *Rarg*, *Nr5a2* and *Klf4* was strongly positively correlated with the

270   expression levels of these genes, further supporting their potential roles in regulating ZGA-

271   associated CREs (Fig. 4h-i). It is worth noting that among these three TFs, *Klf4* already showed

272   high expression levels and high TF activity at the zygote stage, while both *Rarg* and *Nr5a2* showed

273   almost no TF activity at the zygote stage, implying that *Klf4*, as a maternal factor, might contribute

274   to initiating the ZGA process as early as the zygote stage (Fig. 4f and i).

275   **Mutually exclusive regulome confers ICM/TE lineage segregation**

276   Along with gradual increases in heterogeneity among blastomeres in preimplantation embryos,

277   establishment of cell lineage-specific transcription regulatory networks occurred beginning in

278   unified totipotent zygotes that generated ICM and TE cells to enable further embryo development.

279   To reveal the potential active CREs during this process, we determined the correlations between the

280   Acc of each promoter/distal NDR and the Expr of its corresponding ICM/TE-specific expressed

281   genes (specifically expressed in ICM: 766 genes, TE: 930 genes; Supplementary Table 4) across

282   single cells during preimplantation development (Fig. 5a-b). The NDRs significantly correlated with

283   ICM- or TE-specific expressed genes were termed ICM.CREs (positive: 497 in promoters, 4086 in

284   distal regions; negative: 210 in promoters, 1559 in distal regions) or TE.CREs (positive: 774 in

285   promoters, 5109 in distal regions; negative: 424 in promoters, 3445 in distal regions), respectively

286   (Fig. 5a-b; Supplementary Table 5). Consistent with the ZGA-associated CREs, the positively

287   correlated ICM/TE CREs also showed strong enrichment for active histone markers and depletion

288   of repressive histone markers (Extended Data Fig. 7a-b). Notably, all of the known enhancers for

289   three key ICM/TE TFs (*Pou5f1*, *Nanog*, and *Cdx2*) that we analyzed were revealed to be present in

290   preimplantation embryos or in embryonic stem cells, confirming that the CREs identified by our

291   correlation analysis could cover known active enhancers (Fig. 5c-d, Extended Data Fig. 7c-f)[32-34].

292   Specifically, we found three positively correlated CREs (#2, #3, and #4) corresponding to three

293   known enhancers of *Pou5f1*; importantly, CRE #4 showed the highest positive correlation

294   coefficient in our analysis, consistent with previous findings that the known enhancer corresponding

295   to CRE #4 is the dominant enhancer regulating *Pou5f1* expression during preimplantation

296   development (Fig. 5c-d)[32]. Thus, these results validate the accuracy of our analysis.

297   To explore how ICM/TE-associated regulatory networks are regulated in early embryos, we

298   comprehensively analyzed different epigenetic molecular layers in these CREs. First, we calculated

299   the correlations between Met and Expr for each CRE-gene pair. Interestingly, we observed mainly

300   negatively correlated Met vs Expr CRE-gene pairs (27 positive vs 324 negative pairs in promoters;

301   1119 positive vs 4194 negative pairs in distal regions) for the positively correlated Acc/Expr CREs,

302   while we observed mainly positively correlated Met vs Expr CRE-gene pairs (105 positive vs 15

303   negative pairs in promoters; 1504 positive vs 388 negative pairs in distal regions) for the negatively

304   correlated Acc/Expr CREs (Fig. 5e). These results not only reveal the complex interplay between

305   Met and Acc in regulating ICM/TE lineage-associated regulatory networks during preimplantation

306   development but also confirm our speculation that the CREs with Acc/Expr positive and negative

307   correlations might be bound by activators and repressors, respectively. Regardless, the Met levels

308   of both ICM.CREs and TE.CREs were lower in TE cells than in ICM cells, reflecting the more

309   extensive erasure of genome-wide Met in TE cells (Fig. 5f). Next, we measured the dynamics of

310   different molecular layers in these CREs during preimplantation development. Clearly, Acc and

311   active histone modifications (H3K4me3 and H3K27ac) gradually increased in both positively and

312   negatively correlated Acc/Expr CREs beginning at the 2-cell stage, while repressive epigenetic

313   modifications (Met, H3K9me3 and H3K27me3) were already depleted in zygotes and remained

314   depleted throughout preimplantation development in these regions, suggesting an overall priming

315   of the epigenetic environment during ICM/TE lineage differentiation (Extended Data Fig. 8a-f).

316   Interestingly, we also noticed that ICM.CREs were activated earlier than TE.CREs, as we observed

317   clearly higher levels of active epigenetic modifications in ICM.CREs than in TE.CREs at the 2-cell

318   stage (Extended Data Fig. 8a, b, e and 9a-b). Subsequently, TE.CREs were quickly activated and

319   showed higher levels of active epigenetic modifications than ICM.CREs at the 8-cell stage,

320   suggesting a stepwise activation of ICM.CREs and TE.CREs during preimplantation development

321   (Extended Data Fig. 8a, b, e and 9a-b).

322   Finally, we investigated which TFs might be responsible for the establishment of differential

323   regulatory networks in ICM and TE lineages. Notably, we found that the commonly enriched TFs

324   in positively correlated ICM/TE CREs were ZGA drivers that showed high TF activity as early as

325   the 2-cell embryo stage, such as *Nr5a2*, *Rarg*, *Rara*, *Bcl6*, etc., indicating that the earliest initiation

326   of both ICM and TE programs occurs during the ZGA process (Fig. 5g and Extended Data Fig. 9c).

327   In addition, three TFs, *Crx*, *Arnt* and *Pitx1,* were more enriched in ICM.CREs, while *Ctcf*,

328   *Klf3/4/6/9/10*, *Gata1/2/4/6*, *Tead1/2/3/4*, *Sfpi1*, *Tcfap2a* and *Sp1/2* were more enriched in TE.CREs

329   (Fig. 5g). Interestingly, we observed that TE lineage-specific TFs, such as *Tcfap2a* and *Gata* family

330   TFs, were enriched in the negatively correlated ICM.CREs, suggesting their repressive roles in

331   regulating the ICM program (Fig. 5g). Moreover, most TE.CRE-specifically enriched TFs showed

332   higher activity and expression levels in TE cells than in ICM cells, while ICM.CRE-associated TFs

333   showed higher expression levels in ICM cells than in TE cells (Fig. 5h-i and Extended Data Fig.

334   9c). Together, these results suggest that a mutually exclusive regulatory network is adopted to

335   gradually establish and stabilize the different ICM and TE lineage fates, especially for the TE

336   lineage; specific drivers of this lineage establish a TE program while repressing the ICM program,

337   forcing the TE lineage to separate from the ICM lineage.

Discussion

339   In conclusion, we have developed a single-cell multiomics sequencing technology, scNOMeRe-seq,

340   that can be used to profile transcriptomes, DNA methylomes and Acc in parallel in the same

341   individual cell with high accuracy, sensitivity and genome coverage. Taking advantage of this

342   powerful tool, we have also characterized multiple molecular layers of mouse preimplantation

343   embryos at single-cell resolution and have explored the associations between different epigenome

344   layers and transcriptional output, providing new insights to enhance functional understanding of

345   epigenetic reprogramming during mouse preimplantation development. Specifically, our results

346   reveal that PG blastomeres show delayed development and abnormal DNA methylomes, in contrast

347   to aneuploid blastomeres. The changes in Acc not only reflect the dynamic regulatory landscape but

348   also may be substantially derived from asynchronous cell cycles of blastomeres that are irrelevant

349   with transcriptional regulation, highlighting the importance of functional interpretations of

350   epigenetic reprogramming at single-cell resolution. Using the DNA methylomes of all the individual

351   cells within individual embryos, we reconstructed genetic lineages from zygotes to 8-cell embryos

352   and revealed that asymmetric cleavage might be the major driver of the gradual increases in

353   transcriptome heterogeneity among blastomeres that occur during the first three cleavages. Despite

354   global demethylation in early embryos, allele-specific Met patterns inherited from oocytes and

355   sperm are maintained throughout preimplantation development. The associations between Acc/Met

356   and Expr at promoter regions in single cells are consistent with the findings from bulk samples;

357   however, the positive correlations between Met and Expr at gene body regions are largely inherited

358   from maternal genomes and are absent in the paternal genomes of early embryos. The Acc of

359   parental memory-related regions appears to be substantially erased during the ZGA process and

360   reconfigured in concert with the influences of histone modifications, Met, repeats, TFs, and possible

361   high-dimensional chromatin structures to ensure proper activation of the zygotic genome. The

362   overall-primed ICM/TE lineage-associated CREs are partially activated as early as the 2-cell

363   embryo stage and are asynchronously activated in the following preimplantation stages. Intriguingly,

364   TE lineage-specific TFs seem to play dual roles in activating the TE program and repressing the

365   ICM program, thereby segregating the TE fate from the ICM fate. Taken together, our findings not

366   only provide new insights into the functional regulatory landscape in preimplantation development

367   but also elucidate the fundamental mechanisms of epigenetic regulation.

368

Methods

Methods and any associated references are available in the supplemental sections.

Acknowledgements

Author contributions

Y.W. conceived the method. L.Y. and J.Q. conceived the project. Y.W. and P.Y. performed experiments with the help of M.Y., Y.H., Y.N., X.Z. Y.W., Z.Y., and Y.P. performed computational analysis. Y.W., P.Y. and Z.Y. wrote the manuscript with feedback from all authors.

Competing financial interests

The authors declare no competing financial 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**, 723-734, doi:10.1038/nrm3885 (2014).

2   Zernicka-Goetz, M., Morris, S. A. & Bruce, A. W. Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat Rev Genet* **10**, 467-477, doi:10.1038/nrg2564 (2009).

3   Wang, Y., Liu, Q., Tang, F., Yan, L. & Qiao, J. Epigenetic Regulation and Risk Factors During the Development of Human Gametes and Early Embryos. *Annu Rev Genomics Hum Genet* **20**, 21-40, doi:10.1146/annurev-genom-083118-015143 (2019).

4   Guo, F. *et al.* Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Res* **27**, 967-988, doi:10.1038/cr.2017.82 (2017).

5   Zhu, P. *et al.* Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat Genet* **50**, 12-19, doi:10.1038/s41588-017-0007-6 (2018).

6   Wu, J. *et al.* The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652-657, doi:10.1038/nature18606 (2016).

7   Zhang, B. *et al.* Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553-557, doi:10.1038/nature19361 (2016).

8   Liu, X. *et al.* Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558-562, doi:10.1038/nature19362 (2016).

9   Dahl, J. A. *et al.* Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548-552, doi:10.1038/nature19360 (2016).

10   Lu, F. *et al.* Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell* **165**, 1375-1388, doi:10.1016/j.cell.2016.05.050 (2016).

11   Xia, W. *et al.* Resetting histone modifications during human parental-to-zygotic transition. *Science* **365**, 353-360, doi:10.1126/science.aaw5118 (2019).

12   Guo, H. *et al.* The DNA methylation landscape of human early embryos. *Nature* **511**, 606-610, doi:10.1038/nature13544 (2014).

13   Wang, C. *et al.* Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nature cell biology* **20**, 620-631, doi:10.1038/s41556-018-0093-4 (2018).

14   Li, L. *et al.* Single-cell multi-omics sequencing of human early embryos. *Nature cell biology* **20**, 847-858, doi:10.1038/s41556-018-0123-2 (2018).

15   Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82, doi:10.1038/nature11232 (2012).

16   Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**, 207-220, doi:10.1038/s41576-018-0089-8 (2019).

17   Wu, J. *et al.* Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256-260, doi:10.1038/s41586-018-0080-8 (2018).

18   Gao, L. *et al.* Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. *Cell* **173**, 248-259 e215, doi:10.1016/j.cell.2018.02.028 (2018).

19   Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-490, doi:10.1038/nature14590 (2015).

428   20   Chen, X., Miragaia, R. J., Natarajan, K. N. & Teichmann, S. A. A rapid and robust method
429   for single cell chromatin accessibility profiling. *Nat Commun* **9**, 5345, doi:10.1038/s41467-
430   018-07771-0 (2018).
431   21   Buenrostro, J. D. *et al.* Integrated Single-Cell Analysis Maps the Continuous Regulatory
432   Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535-1548 e1516,
433   doi:10.1016/j.cell.2018.03.074 (2018).
434   22   Hou, Y. *et al.* Single-cell triple omics sequencing reveals genetic, epigenetic, and
435   transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res* **26**, 304-319,
436   doi:10.1038/cr.2016.23 (2016).
437   23   Clark, S. J. *et al.* scNMT-seq enables joint profiling of chromatin accessibility DNA
438   methylation and transcription in single cells. *Nat Commun* **9**, 781, doi:10.1038/s41467-
439   018-03149-4 (2018).
440   24   Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of
441   single cells. *Science* **361**, 1380-1385, doi:10.1126/science.aau0730 (2018).
442   25   Bian, S. *et al.* Single-cell multiomics sequencing and analyses of human colorectal cancer.
443   *Science* **362**, 1060-1063, doi:10.1126/science.aao3791 (2018).
444   26   Liu, L. *et al.* Deconvolution of single-cell multi-omics layers reveals regulatory
445   heterogeneity. *Nat Commun* **10**, 470, doi:10.1038/s41467-018-08205-7 (2019).
446   27   Pott, S. Simultaneous measurement of chromatin accessibility, DNA methylation, and
447   nucleosome phasing in single cells. *Elife* **6**, doi:10.7554/eLife.23203 (2017).
448   28   Sheng, K., Cao, W., Niu, Y., Deng, Q. & Zong, C. Effective detection of variation in single-
449   cell transcriptomes using MATQ-seq. *Nature methods* **14**, 267-270,
450   doi:10.1038/nmeth.4145 (2017).
451   29   Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic,
452   random monoallelic gene expression in mammalian cells. *Science* **343**, 193-196,
453   doi:10.1126/science.1245316 (2014).
454   30   Luo, H. *et al.* Cell identity bookmarking through heterogeneous chromatin landscape
455   maintenance during the cell cycle. *Hum Mol Genet* **26**, 4231-4243,
456   doi:10.1093/hmg/ddx312 (2017).
457   31   Xie, W. *et al.* Base-resolution analyses of sequence and parent-of-origin dependent DNA
458   methylation in the mouse genome. *Cell* **148**, 816-831, doi:10.1016/j.cell.2011.12.035
459   (2012).
460   32   Yeom, Y. I. *et al.* Germline regulatory element of Oct-4 specific for the totipotent cycle of
461   embryonal cells. *Development* **122**, 881-894 (1996).
462   33   Blinka, S., Reimer, M. H., Jr., Pulakanti, K. & Rao, S. Super-Enhancers at the Nanog Locus
463   Differentially Regulate Neighboring Pluripotency-Associated Genes. *Cell Rep* **17**, 19-28,
464   doi:10.1016/j.celrep.2016.09.002 (2016).
465   34   Rayon, T. *et al.* Notch and hippo converge on Cdx2 to specify the trophectoderm lineage
466   in the mouse blastocyst. *Developmental cell* **30**, 410-422,
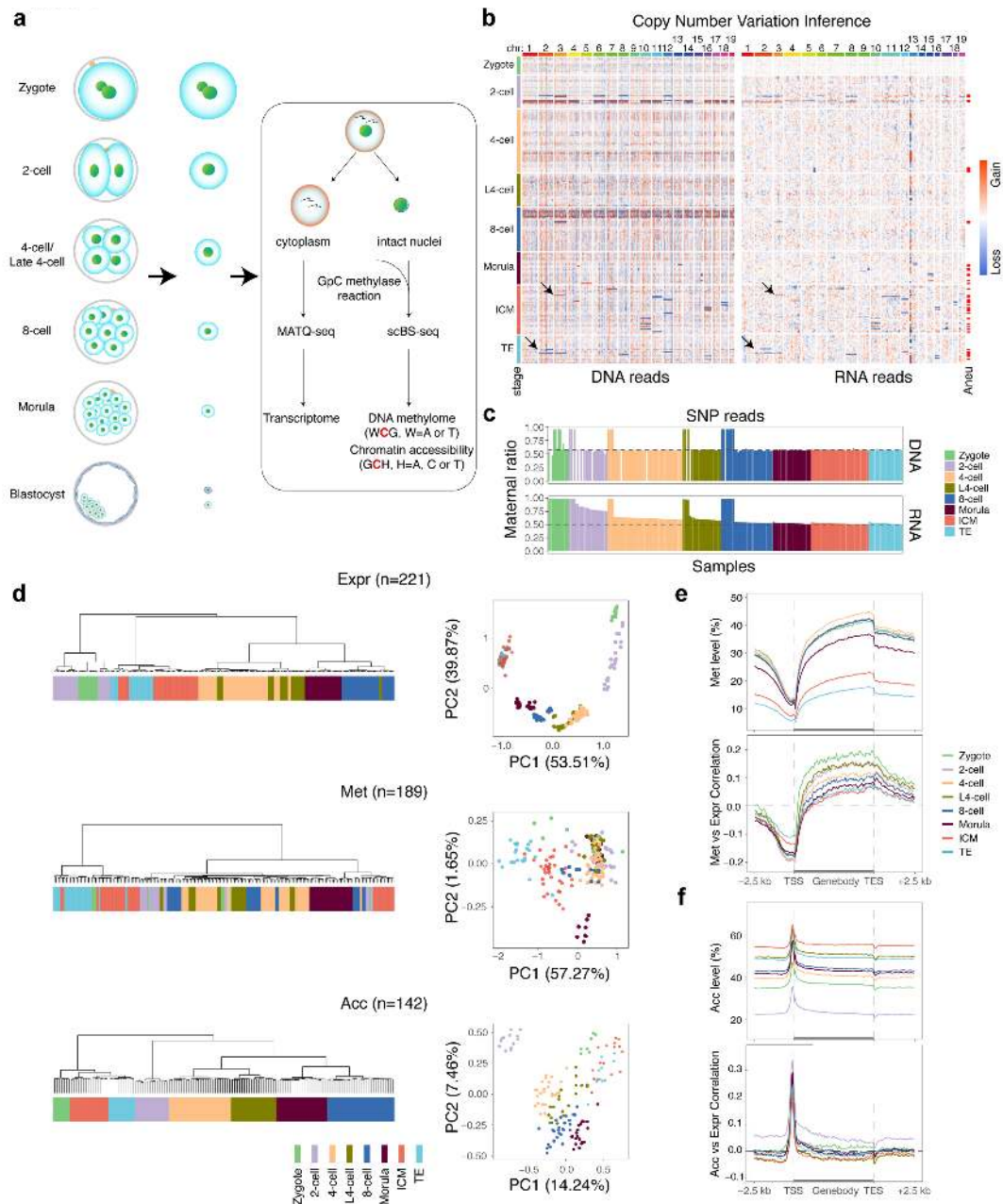467   doi:10.1016/j.devcel.2014.06.019 (2014).
468

Figure 1 scNOMeRe-seq profiles in mouse preimplantation embryos.

(**a**) Schematic illustration of scNOMeRe-seq, including key steps, methods of library preparation and mouse preimplantation stages analyzed in this work outlined in the text.

(**b**) Heat map showing copy number variations (CNV) inferred by DNA reads (left) and gene expression level (right, normalized each gene by average closest 100 genes) in mouse preimplantation blastomeres. Arrows indicate the examples of matched CNV inferences from DNA and RNA reads.

(**c**) Bar plot showing the ratios of SNP tracked maternal DNA (top) or RNA (bottom) reads in total SNP tracked parental reads in each individual cell across preimplantation development.

(**d**) Unsupervised clustering (left) and principle component analysis (right) of preimplantation blastomeres using gene expression level (top), DNA methylation level of 5 kilobases (kb) tiles (middle) and chromatin accessibility of all stage merged NDRs (bottom). n, the cell numbers of each dataset.

(**e**) Profiles showing DNA methylation level (top), the weighted Pearson correlation coefficients of DNA methylation level vs gene expression level (bottom) along the gene bodies and 2.5 kb upstream of the transcription start sites (TSS) and 2.5 kb downstream of the transcription end sites (TES) of all genes for each stage. Met, WCG methylation; Expr, RNA expression.

(**f**) Profiles showing chromatin accessibility (top), the weighted Pearson correlation coefficients of chromatin accessibility vs gene expression level (bottom), along the gene bodies and 2.5 kb upstream of the TSS and 2.5 kb downstream of the TES of all genes for each stage. Acc, GCH methylation.
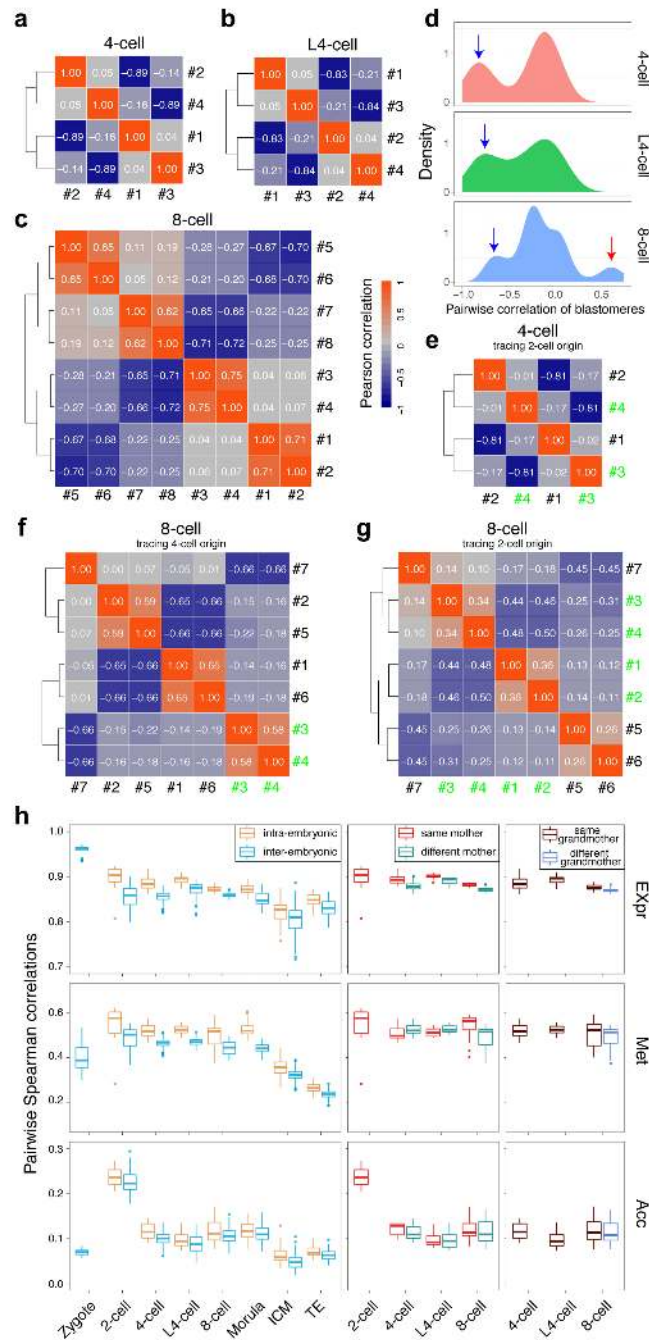
Figure 2 Reconstruction of genetic lineages reveals the source of heterogeneity in early embryos.

(**a-c**) Heat map showing the Pearson correlation coefficients in representative 4-cell embryo (**a**), late 4-cell embryo (**b**), and 8-cell embryo (**c**), using z-score normalized DNA methylation level of 1 millionbases (Mb) bins in individual blastomere from the same embryo. The numbers in white color showing the Pearson correlation coefficients.

(**d**) Distribution of the pairwise Pearson correlation coefficients for DNA methylation level of 1 Mb bins in individual blastomere from the same 4-cell embryo (top), late 4-cell embryo (middle) and 8-cell embryo (bottom). Blue arrows indicate the pairs of blastomeres from the same 2-cell blastomere, and the red arrow indicates the pairs of blastomeres from the same 4-cell blastomere.

(**e**) Heat map showing the Pearson correlation coefficients in a 4-cell embryo. #3-4 (green labeled) cells divided from the same 2-cell blastomere validated by injection with FITC-dye. The numbers in white color showing the Pearson correlation coefficients.

(**f**) Heat map showing the Pearson correlation coefficients in an 8-cell embryo. #3-4 (green labeled) cells divided from the same 4-cell blastomere validated by injection with FITC-dye. The numbers in white color showing the Pearson correlation coefficients.

(**g**) Heat map showing the Pearson correlation coefficients in an 8-cell embryo. #1-4 (green labeled) cells divided from the same 2-cell blastomere validated by injection with FITC-dye. The numbers in white color showing the Pearson correlation coefficients.

(**h**) Box plot showing the pairwise Spearman correlation coefficients of RNA expression level (top), DNA methylation level of 5 kb bins (middle) and chromatin accessibility of NDRs (bottom) at indicated stages. intra-embryonic, within the same embryo; inter-embryonic, in different embryos of the same stage; same mother, the blastomeres origin from a same mother blastomere; same grandmother, the blastomeres origin from a same grandmother blastomere.
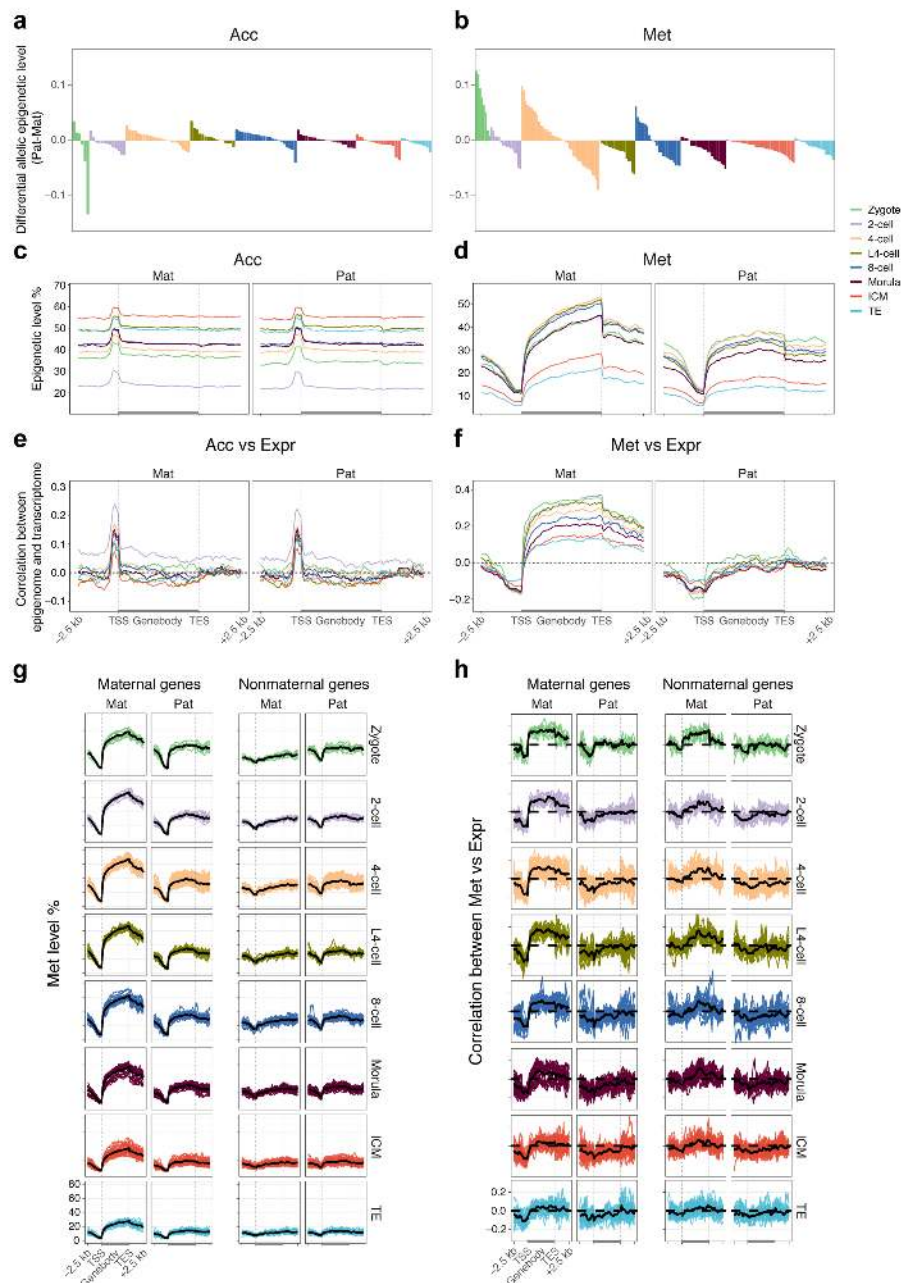
Figure 3 Allele-specific regulation of gene expression in early embryos.

(**a-b**) Bar plot showing differences of global chromatin accessibility (**a**) and DNA methylation level (**b**) between paternal and maternal genomes in each individual cell of preimplantation embryos.

(**c-d**) Allelic chromatin accessibility (**c**) and DNA methylation level (**d**) around genic regions of each stage. Mat, maternal allele; Pat, paternal allele.

(**e**) The Pearson correlations between chromatin accessibility around genic regions of maternal allele or paternal allele vs gene expression level of each stage.

(**f**) The Pearson correlations between DNA methylation level around genic regions of maternal allele or paternal allele vs gene expression level of each stage.

(**g**) Allelic DNA methylation level around genic regions of maternal (TPM ≥ 1 in zygote stage) and nonmaternal genes in each stage.

(**h**) The Pearson correlations between DNA methylation level around genic regions of maternal allele or paternal allele vs expression level of maternal and nonmaternal genes.
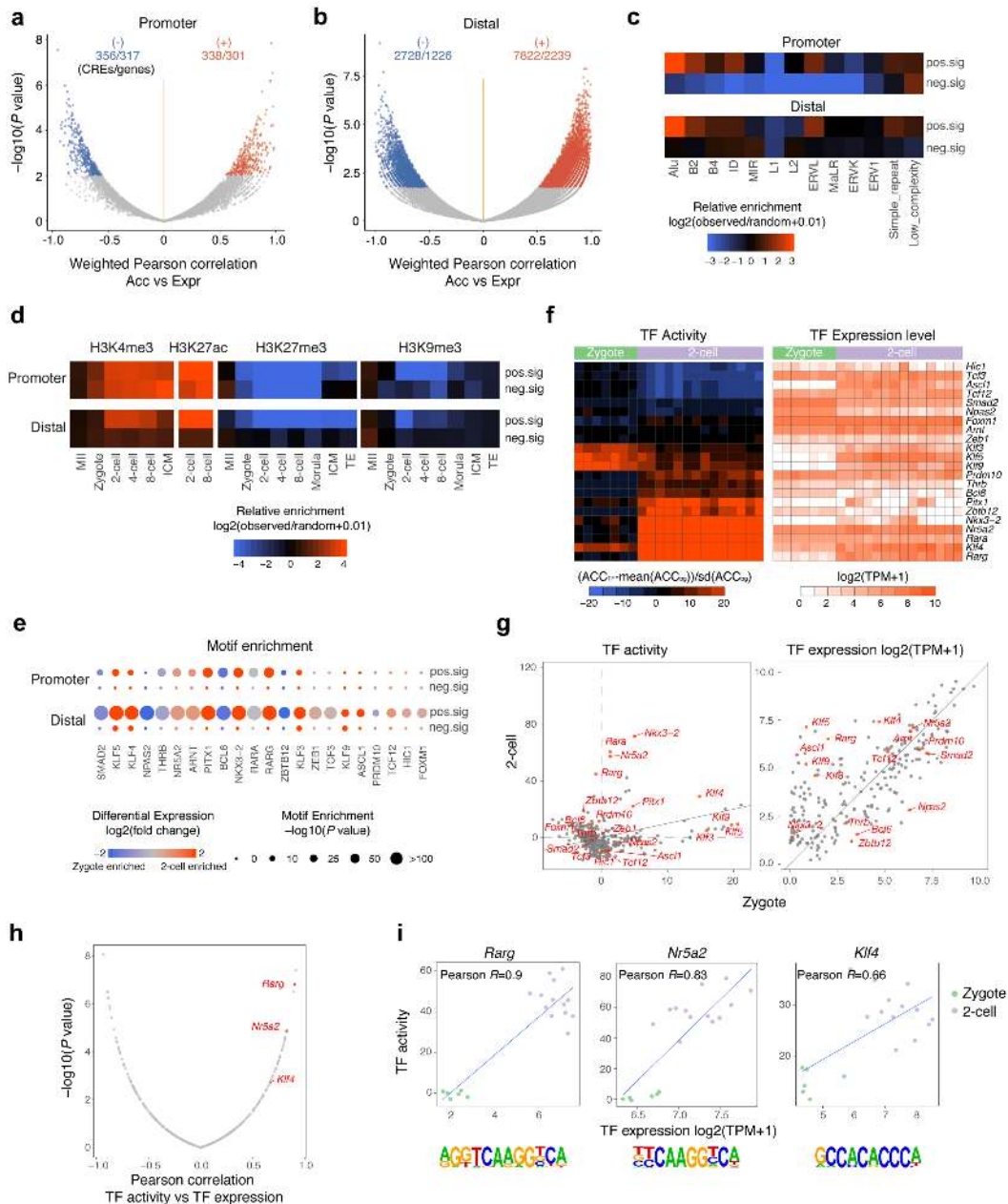
Figure 4 scNOMeRe-seq reveals a ZGA associated regulome.

(**a-b**) Volcano plot showing the weighted Pearson correlations between chromatin accessibility of promoter-NDRs (**a**) / distal-NDRs (**b**) and the expression level of corresponding ZGA genes across cells from zygote to 2-cell stage. Significant associations (FDR < 0.1) are in red (positive) and blue (negative). The number of CREs (CRE, the NDR that was significantly correlated with ZGA gene) and unduplicated genes are labeled.

(**c**) Heat map showing the enrichment of ZGA associated CREs in repeats. The enrichment was calculated as log2 ratio for the numbers of observed CREs that overlap with repeats divided by the numbers of random regions that overlap with repeats.

(**d**) Heat map showing the enrichment of ZGA associated CREs in histone modifications of early embryos from available published data (ref. 7-9). The enrichment was calculated as log2 ratio for the number of observed CREs that overlap with histone modification peaks divided by the number of random regions that overlap with histone modification peaks.

(**e**) TF motifs identified from ZGA associated CREs. Only TFs with the *P* value $< 1 \times 10^{-10}$ and TPM ≥ 5 at least at one stage were included. The color of point represents the differential gene expression level between 2-cell stage and zygote. The size of point represents the significance of motif enrichment.

(**f**) Heat map showing the TF activity (left) and expression level (right) of ZGA enriched TFs in each individual cell of zygote and 2-cell embryos.

(**g**) Scatter plot showing the difference of TF activity (left) and expression level (right) between zygote and 2-cell stage. The genes labeled in red indicate the TFs showing significantly differential activity (left, FDR < 0.1) or expression level (right, FDR < 0.01).

(**h**) Volcano plot showing the Pearson correlations between TF activity and expression level across cells from zygote to 2-cell stage. Significantly correlated (FDR < 0.1) TFs that enriched in ZGA associated CREs and showed higher TF activities and expression level in 2-cell embryos are labeled.

(**i**) Scatter plot showing TF activity and expression level of *Rarg*, *Nr5a2* and *Klf4* in each single cell of zygote and 2-cell embryos. The corresponding enriched DNA-binding motifs are shown at the bottom of each TF.
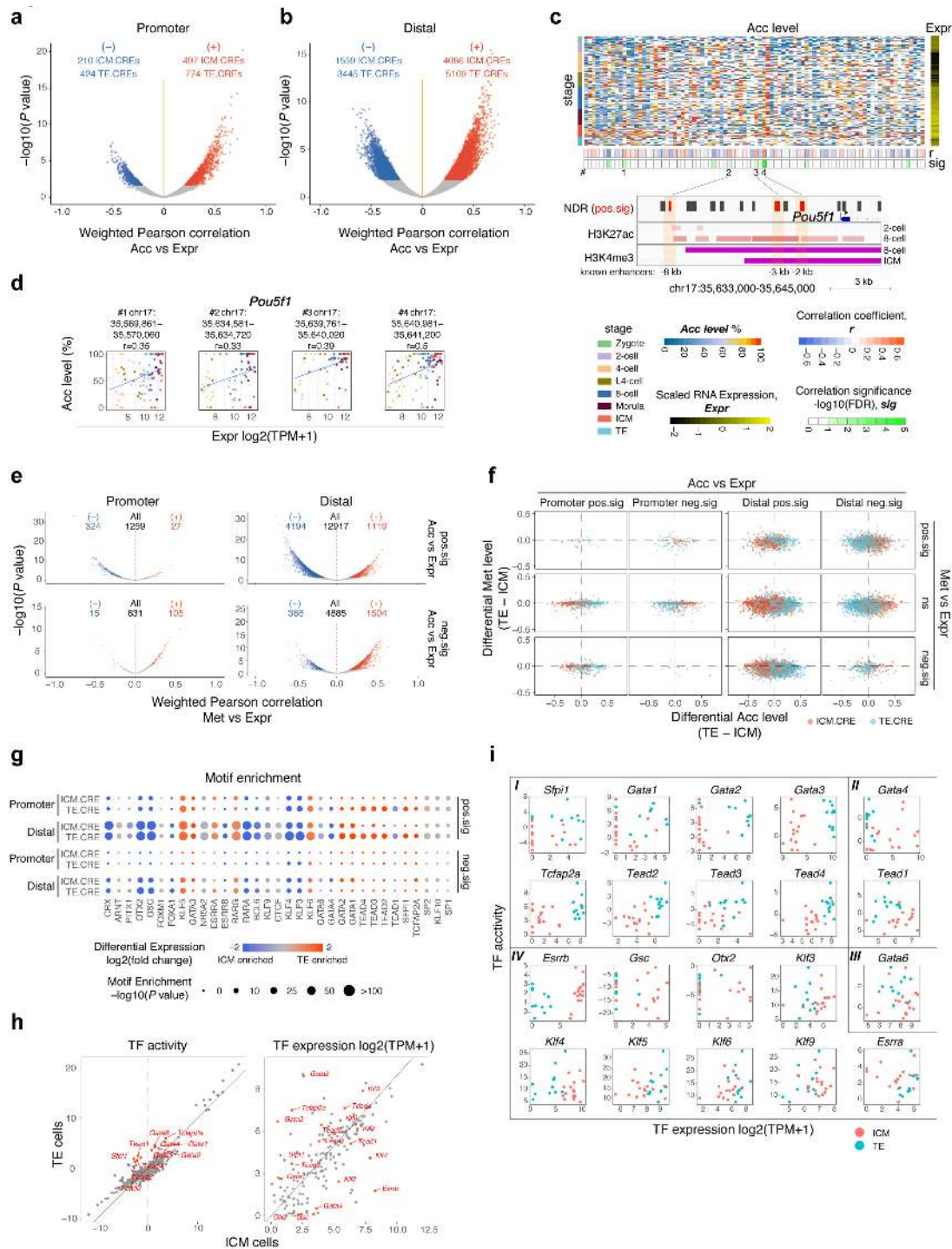
Figure 5 Mutually exclusive regulome confers ICM/TE lineage segregation.

(**a-b**) Volcano plot showing the weighted Pearson correlations between chromatin accessibility of NDRs at promoter (**a**) / distal (**b**) regions and the expression level of corresponding ICM/TE specific expressed genes across preimplantation development. Significant associations (FDR < 0.1) are in red (positive) and blue (negative). ICM.CRE, the NDR that was significantly correlated with ICM-specific-expressed gene; TE.CRE, the NDR that was significantly correlated with TE-specific-expressed gene. The number of CREs are labeled.

(**c**) (top) Heat map showing chromatin accessibility (Acc level) of *Pou5f1* gene locus surrounding NDRs (from TSS upstream 100 kb to TES downstream 100 kb), and expression level (Expr, scaled

log2(TPM+1)) of *Pou5f1* in each individual cell of early embryos; the weighted Pearson correlation coefficients (r) and significance (sig) are shown. #, the labels of positive-correlated CREs. (bottom) IGV snapshot showing the distribution of *Pou5f1* surrounding NDRs (positive-correlated CREs are in red), H3K27ac and H3K4me3 peaks. Three known enhancers of *Pou5f1* are shaded.

(**d**) Scatter plot showing the expression level of *Pou5f1* and chromatin accessibility of positive-correlated CREs which have been labeled in (**c**). The genomic coordinates of CREs and the correlation coefficients are shown.

(**e**) Volcano plot showing the weighted Pearson correlation between DNA methylation level and gene expression level of positive-correlated (top) and negative-correlated (bottom) CREs.

(**f**) Scatter plot showing the differential chromatin accessibility and DNA methylation level of ICM.CREs and TE.CREs between ICM and TE cells at promoter (left) and distal (right) regions. Positive value indicates a higher epigenetic level in TE cells, negative value indicates a higher epigenetic level in ICM cells.

(**g**) TF motifs identified from ICM.CREs and TE.CREs. Only TFs with the *P* value $< 1 \times 10^{-10}$ and TPM ≥ 5 at least at one stage were included. The color of points represents the differential gene expression level between ICM and TE. The size of point represents the significance of motif enrichment.

(**h**) Scatter plot showing the difference of TF activity (left) and expression level (right) between ICM and TE. The genes labeled in red indicate the TFs showing significantly differential activity (left, FDR < 0.1) or expression level (right, FDR < 0.01) between ICM and TE.

(**i**) Scatter plot showing the TF activity and expression level in each individual cell of ICM and TE. I, the TFs showing higher activity and expression level in TE cells; II, the TFs showing higher activity in TE cells but showing higher expression level in ICM cells; III, the TF showing higher activity in TE cells but showing no difference in gene expression level; IV, the TFs showing no difference in activity but showing higher expression level in ICM cells.