Nr5a2 is dispensable for zygotic genome activation but essential for morula development

Nicola Festuccia, Sandrine Vandormael-Pournin, Almira Chervova, Anna Geiselman, Francina Langa, Rémi-Xavier Coux, Inma Gonzalez, Guillaume Giraud Collet, Michel Cohen-Tannoudji, Pablo Navarro

Title: *Nr5a2* is dispensable for zygotic genome activation but essential for morula development

Authors: Nicola Festuccia[1*], Sandrine Vandormael-Pournin[1], Almira Chervova[1], Anna Geiselmann[1,2], Francina Langa-Vives[3], Rémi-Xavier Coux[1], Inma Gonzalez[1], Guillaume Giraud Collet[1,4], Michel Cohen-Tannoudji[1†], and Pablo Navarro[1†]

Affiliations:

1. Department of Developmental and Stem Cell Biology, Institut Pasteur, Université Paris Cité, CNRS UMR3738, Epigenomics, Proliferation, and the Identity of Cells Unit; Paris, France.
2. Sorbonne Université, Complexité du Vivant, F-75005, Paris, France.
3. Mouse Genetics Engineering Center, C2RA; Institut Pasteur, Paris, France.
4. Université Paris Cité, BioSPC, F-75013, Paris, France.

† These authors contributed equally to this work
* Corresponding author: nicola.festuccia@pasteur.fr

Abstract:

Early embryogenesis is driven by transcription factors (TFs) that first activate the zygotic genome and then specify the lineages constituting the blastocyst. While the TFs specifying the blastocyst's lineages are well characterized, those playing earlier roles remain poorly defined. Using mouse models of the TF *Nr5a2*, we show that *Nr5a2*$^{-/-}$ embryos arrest at the early morula stage and exhibit altered lineage specification, frequent mitotic failure and substantial chromosome segregation defects. Whilst NR5A2 plays a minor but measurable role during zygotic genome activation, it predominantly acts as a master regulator at the 8-cell stage, controlling expression of lineage-specifying TFs and genes involved in mitosis, telomere maintenance and DNA repair. We conclude that NR5A2 coordinates proliferation, genome stability and lineage specification to ensure correct morula development.

One-Sentence Summary:

Pre-implantation development is strictly dependent on NR5A2, a transcription factor dispensable for zygotic genome activation, but required for the formation of a viable morula.

In mammals, pre-implantation development corresponds to the transition from a single cell, the zygote, to a complex structure, the blastocyst, constituted of three lineages that give rise to all extra-embryonic and embryonic tissues. The specification of the lineages of the blastocyst starts in the morula, which consists of 16 to 32 cells resulting from the successive cleavages of the zygote and compaction, and is mediated by the hierarchical action of several transcription factors (TFs). TEAD4, GATA3 and CDX2 specify the extra-embryonic trophectoderm (TE), while SOX2 and OCT4, among others, support the formation of the inner cell mass (ICM), giving rise to the early blastocyst. NANOG and GATA6 subsequently segregate the embryonic pluripotent epiblast (EPI) from the extra-embryonic primitive endoderm (PrE), to form the late blastocyst competent for implantation and further development (*1*). While the importance of these and other TFs in the formation of the blastocyst has been clearly established, whether and which TFs ensure the formation of a viable morula remains unclear.

TFs have also been associated with the first regulatory step at the onset of mouse development, the zygotic genome activation (ZGA), which initiates transcription at the 2-cell (2C) stage (*2*). In contrast to lineage determinants, many of which perform essential roles in the formation of the blastocyst, the disruption of TFs involved in ZGA, including DUX, RARG, DPPA2/4, NFYA and YAP1, typically does not compromise ZGA or progression beyond the 2C stage (*3-10*). Recently, however, two classes of TFs have been proposed to play a preponderant role during ZGA, OBOX TFs (*11*) and the orphan nuclear receptor NR5A2 (*12*). However, the role of NR5A2 remains debated. Upon chemical inhibition of NR5A2, development is arrested in cultured embryos as early as the 2C stage, and this phenotype is linked to a largely compromised ZGA (*12*). However, other studies, using base editing and knockdown techniques, suggested that NR5A2 plays a major role later, controlling the transcription of lineage determinants (*13-15*). These partially conflicting reports pointed to the importance of NR5A2 as a regulator of early embryogenesis but did not base their conclusions on stable knockout (KO) models of *Nr5a2*. Thus, uncertainty remains as to whether embryogenesis can proceed *in utero* in the complete absence of NR5A2, and if the observed developmental phenotypes are mediated by the role of NR5A2 as an initiator of ZGA – implying an essential function of its maternal inheritance – or through gene regulation at later stages, after its zygotic expression. Unambiguously establishing whether NR5A2 performs essential functions is also complicated by the possible redundancy with other nuclear receptors, such as ESRRB, given their combined role in maintaining the activity of the pluripotency gene regulatory network in mouse embryonic stem cells (ESCs) (*16, 17*).

We thus established stable and inducible knockout mouse models to determine the role of *Nr5a2* before implantation and studied any potential functional interaction with *Esrrb*. We show that while maternal and zygotic *Nr5a2* expression transiently contribute to activating a small fraction of 2C stage-specific genes, neither NR5A2 nor ESRRB is required, alone or in combination, to properly execute ZGA and progress beyond the 2C stage *in utero*. However, after ZGA, NR5A2 alone becomes essential for the development of a mature morula capable of initiating lineage specification and developing into a blastocyst. Therefore, our work establishes NR5A2 as a key TF during a crucial period of pre-implantation development: that which links the initiation of gene expression to the first cell fate determination events leading to the emergence of the three early embryonic lineages.

Results

Maternally expressed NR5A2 and ESRRB are dispensable for embryogenesis.

To study the roles of NR5A2 and ESRRB during pre-implantation development, we generated female mice carrying previously described conditional knockout alleles for *Esrrb* (*18*) and/or *Nr5a2* (*19*), combined with *Zp3:Cre* (*20*). *Zp3:Cre* drives allelic recombination during oocyte growth, and has been widely used to study the role of maternal developmental factors. This strategy allowed us to test whether development proceeds in the absence of maternal (mKO), zygotic (zKO) or maternal and zygotic (mzKO) expression of *Nr5a2* and *Esrrb* (NrKO and EKO, respectively) or both genes (DKO; see methods and **Table S1**). Recently, NR5A2 has been reported to function as a major regulator of ZGA (*12*), implying that ablating its maternal contribution should trigger severe developmental defects. We first examined the effects of the loss of maternal *Nr5a2*/*Esrrb*. Contrary to expectation, females with oocytes devoid of NR5A2, ESRRB, or both TFs (oKO), gave birth to litters of equal size to matched controls, efficiently transmitting the deleted alleles (**Fig 1A**, **Table S1**). This rules out the possibility that the maternal inheritance of NR5A2 or ESRRB performs a determinant function. In accordance with this, NR5A2 expression above background levels, defined by analyzing oKO oocytes or mzKO embryos, was not observed by immunofluorescence before or around the time of ZGA (**Fig 1B**, **S1A**), in contrast to later stages. Low *Nr5a2* expression was confirmed by Ribo-seq (**Fig S1B**) (*21*), and RNA-seq (**Fig S1C**) (*22*). It is notable, however, that NR5A2 translation began early, reaching substantial levels already at the mid 2C stage. Similarly, microscopic analysis of oocytes and embryos from *Esrrb-TdTomato* reporter mice revealed that ESRRB was not

maternally inherited at substantial levels (**Fig 1C**, **S1D**), and became detectable in late 2C embryos. Taken together, these results suggest that NR5A2 and ESRRB are not decisive maternal regulators of ZGA. They may however exert non-essential functions at the onset of ZGA as soon as the respective proteins increase in abundance.

A minor role for NR5A2 during ZGA.

The normal development of *Nr5a2* mKOs past the 2C stage contradicts the recent proposal that NR5A2 functions as a key ZGA regulator (*12*). To further assess a potential role of NR5A2 during ZGA, we performed full-length single-cell RNA sequencing (FLASH-seq) (*23*) based gene expression analysis of 102 control or *Nr5a2* oKO/mKO, zKO and mzKO oocytes and early 2C (E2C) embryos. We also analysed 122 control or mzKO embryos lacking either *Nr5a2*, *Esrrb* or both genes, immediately after ZGA in late 2C (L2C) embryos, in early 4C, and in 8C embryos (**Table S2, S6**). The deletion of *Nr5a2* in growing oocytes resulted in minimal expression differences before fertilization (**Fig 2A-B**), indicating that mutants and control embryos began development from an equivalent starting point. A small number of genes were affected around the time of ZGA: in mzNrKO embryos the number of differentially expressed genes (DEGs) was small at the early 2C stage (E2C: 272 up, 360 down). Similarly, limited changes were observed in L2C embryos in mzDKOs (152 up, 255 down; **Fig 2A-B**), or mzNrKOs (**Fig S2A**). Subsequently, an increase in gene dysregulation was evident: first to a lesser extent, with 358 up- and 426 down-regulated genes at the 4C stage in mzDKOs, or 198 and 280 in mzNrKOs, and then more substantially, with more than 2000 genes affected at the 8C stage (1703/1398 up, 1247/982 down; **Fig 2A-B, S2A**). In addition, the genes dependent on NR5A2 in early 2C embryos were not affected in mutants from the late 2C stage onwards, when a distinct set of genes started to be differentially expressed in mzNrKOs or mzDKOs (**Fig 2B-C, S2A**). This indicates a sharp shift between two non-overlapping functions for NR5A2/ESRRB, likely dependent on the different transcriptional landscapes that characterize the embryo before and after ZGA (*22*, *24*, *25*).

Next, we directly assessed the effect of the loss of NR5A2 on the expression of a validated set of 2107 genes (**Fig S2B**, see methods) that are activated during ZGA (*12*). We did not observe global changes in their profile of expression from oocytes to 8C embryos suggesting that ZGA was properly executed in all genotypes and that NR5A2/ESRRB do not play an essential role in this process (**Fig 2D**). The exclusive loss of maternal *Nr5a2*, which we would have expected

to be determinant had this TF played a key role in initiating ZGA, disrupted the activation of only 11 genes in early 2C embryos (**Fig 2E left, S2C-D**). Expression of 5% of ZGA genes was dysregulated only after additionally ablating zygotic *Nr5a2* expression (99/2107; **Fig 2E middle, S2C-D**). At the late 2C stage, when ZGA is normally largely completed, the combined mzKO of *Nr5a2*/*Esrrb* affected only 71 genes (**Fig 2E right**). Similar changes were observed after *Nr5a2* mzKO and zKO (**Fig S2D**), suggesting that zygotically expressed NR5A2 was responsible for most of these effects. In addition, responsive genes showed weak enrichment in the proximity of regions bound by NR5A2 in 2C stage embryos (*15*), indicating indirect regulation (**Fig 2F**). Thus, we conclude that the combination of maternal and zygotic *Nr5a2* expression contributes to the regulation of a small fraction of ZGA genes. In addition to single copy genes, ZGA is also characterized by transcription of several repetitive elements, with specific groups (*MERVLs*, *MaLRs* and *B1/Alu SINE* elements) promoting expression of nearby genes (*26, 27*). Despite the reported binding of NR5A2 to *SINE* elements (**Fig S2E**) and more sporadically to *ERVLs*, only a slight tendency of *MERVLs* to be downregulated was observed, especially in late 2C embryos (**Fig 2G**), corroborating the notion that *Nr5a2* function is largely dispensable during ZGA.

Notably, we found specific regulators of E2C embryos, including *Zscan4* and *Tcstv* genes (*28, 29*), among the few differentially expressed genes at this stage. We thus analyzed the expression of 358 genes peaking in early 2C embryos (**Fig S2B**). Among these, 10 and 66 genes were downregulated in *Nr5a2* mKOs and mzKOs, respectively (**Fig 2H**), although also in this case these targets were not enriched in proximity of regions bound by NR5A2 (**Fig 2I**). This suggests that NR5A2 mediates the activation of a small subset of early 2C-specific genes. We next checked whether known regulators of ZGA, *Obox (11)* and *Dux* genes (*10*), were affected by NR5A2, but these did not show increased expression, ruling out the possibility that they functionally compensate for the loss of NR5A2 (**Table S2**). Moreover, we did not observe an overlap between the genes affected by the mzKO of *Nr5a2*/*Esrrb* and the mzKO of *Dux (10)* at the 2C stage (**Fig S2F**). However, and in accordance with the reported activation of *Nr5a2* by OBOX (*11*), a partial overlap in targets was observed with OBOX factors (**Fig S2G**), suggesting that NR5A2 might function downstream of these genes. We also did not find that the nuclear receptors expressed at these stages or in MII oocytes were upregulated (**Fig S2G**). Thus, the lack of substantial expression deficits in our genetic experiments cannot be ascribed to functional compensation or transcriptional adaptation processes, as described in other systems (*30, 31*).

Altogether, we conclude that zygotically and maternally expressed NR5A2 begins contributing to the regulation of gene expression immediately after the initiation of zygotic transcription, as previously reported (*12*). However, NR5A2 is not an essential regulator of ZGA. While NR5A2 supports the activation of a small fraction of ZGA and E2C-specific genes, these effects are transient, possibly indirect, and depend on the combined zygotic and maternal expression of NR5A2. By the late 2C stage, ZGA has been correctly executed in all mutant embryos, NR5A2 expression increases, and this TF, together with ESRRB, begins regulating a distinct and much larger set of genes.

Zygotically expressed NR5A2 is essential for morula development.

Given the extensive transcriptional dysregulation observed at the 8C stage, with more than 2000 differentially expressed genes at the 8C stage, we next sought to assess the effects of ablating maternal and zygotic expression of *Nr5a2*/*Esrrb* on the progression of development beyond ZGA. The genetic disruption of *Esrrb* is known to trigger developmental defects in extra-embryonic lineages only after implantation (*32*). In keeping with this, *Esrrb* mzKO did not affect pre-implantation development: morphologically normal mutant blastocysts which had completed lineage segregation could be recovered at E4.25 (**Fig 3A,C**, **Table S1**). Defects in the epiblast have been reported in *Nr5a2* zKO embryos (*33-35*), but only around the time of implantation, which contrasts with the early effect of depleting NR5A2 in cultured embryos (*12, 14, 15*). Supporting an early role for NR5A2, all E3.5-E4.25 *Nr5a2* mzKO embryos showed profound morphological abnormalities and a marked developmental delay (**Fig 3B,C**; **Table S1**), accompanied by reduced cell numbers, degenerated nuclear morphology, and disordered progression through cell division (**Fig 3D**). To exclude potential biases due to the genetic background in our (prevalently C57BL/6) and previous genetic KOs, and to exclude specific effects due to the combined loss of maternal and zygotic NR5A2, we analyzed *Nr5a2* zKOs generated in a mixed CD-1/C57BL/6 background, which recapitulated the phenotype of mzKOs. (**Fig S3A,B**). Therefore, zygotic expression of *Nr5a2* appears to perform dominant functions. To confirm the prevalence of zygotic NR5A2 after the 4C stage, and rule out the impact of any earlier role, we used two complementary strategies: inducible deletion of *Nr5a2* just before the appearance of morphological defects in mzKOs and rescue of NR5A2 deficiency from the 4C stage. For the first approach, we used a tamoxifen inducible $ROSA^{Cre\text{-}ERt2}$ driver (*36*). $Nr5a2^{flox/flox}$ females were crossed with $ROSA^{Cre\text{-}ERt2\ KI/+}$; $Nr5a2^{\ del/+}$ males and embryos

were collected at the 4C stage and treated with tamoxifen for 24h. Analysis after one additional day of culture, when untreated or wild-type/heterozygous embryos had reached the early blastocyst stage, confirmed the absence of NR5A2 and revealed that the phenotypic abnormalities observed in mzKO mutants were reproduced by triggering the loss of *Nr5a2* after ZGA is completed (**Fig S3C,D and Table S1**). Conversely, lethality and developmental delay could be rescued by injecting mz*Nr5a2* KO embryos with *in vitro* transcribed *Nr5a2* mRNA at the 4C stage, after ZGA (**Fig 3E and Table S1**). These results exclude any essential contribution of maternal *Nr5a2* expression, or any determinant function during ZGA that could explain the defects observed from the 8C stage onwards and establish that the developmental failure in mutant embryos is due to the loss of zygotic NR5A2.

To define when the first major defects appear in the mutants, we cultured *Nr5a2* mzKO embryos isolated at the 4C stage. Live imaging revealed that developmental delay/defects occurred progressively after the 8C stage, and by the time control embryos formed blastocysts, mutant morulae fully degenerated (**Fig S3E and Table S1**). *In utero* at E2.75 (8-16C), the expression of NANOG and GATA3 in some cells of the embryos was compromised in both *Nr5a2* and double *Nr5a2*/*Esrrb* mzKO embryos, but most cells continued to express GATA6. These TFs are key regulators of the EPI, TE and PrE, respectively (**Fig 3F**, **S3F**). Other pluripotency TFs were either less affected (SOX2) or not perturbed (OCT4; **Fig S3G**). At E2.5, mutants already showed a developmental delay (**Fig 3C**), and incipient NANOG expression was abolished (**Fig S3H**). Therefore, zygotic NR5A2 is essential for establishing the expression of key determinants of the first lineage segregation, and for morula formation.

***NR5A2 is dominant over ESRRB, and controls genes involved in lineage specification, cell division and genomic stability*.**

Given the severe phenotype observed at the 8C stage, we further explored the transcriptional effects of ablating maternal and zygotic *Nr5a2/Esrrb* in embryos developing past ZGA. As noted before, and in keeping with the appearance of evident defects in mutant embryos, gene dysregulation increased from the late 2C to the 8C stage (**Fig 4A, S4A, 2A**), as validated by absolute transcript counts (**Fig S4B**). Gene expression trends at the 2C and 4C stage were largely concordant, with a stronger overlap between DEGs that fail to be expressed in mutant embryos (DOWN, **Fig S4A, 2C**). Genes affected at the 4C stage also showed concordant changes at the 8C stage, but the global similarity in gene regulation was less obvious, due to

the large number of responsive genes that started being affected only at this time. Assessing the individual contribution of NR5A2 and ESRRB gave results compatible with the phenotype observed in single mutant embryos: *Esrrb* mzKO had minor effects at the 8C stage whereas *Nr5a2* mzKO recapitulated the effect of the double *Esrrb*/*Nr5a2* mzKO (**Fig 4B,C**). Nevertheless, a modest but clear effect of ESRRB was observed at the late 2C stage, where *Esrrb* mzKO and mzDKO DEGs appeared to be concordantly regulated. Visualizing the combination of these effects, principal component analysis of single embryos delineated a developmental trajectory that showed how *Nr5a2* mzKO and mzDKO embryos start diverging from wild-type and *Esrrb* mzKO embryos only at the 4C stage, and segregate from controls at the 8C stage (**Fig 4D**).

To characterize the genes regulated by NR5A2 from the late 2C to the 8C stage, we performed gene ontology analysis after combining all DEGs under the experimental conditions (L2/4/8C-stages, **Table S3**). Clustering the first 200 most enriched terms revealed several groups of related ontologies that correlated with the observed phenotypes (**Fig 4E**). Developmental terms such as blastocyst and placenta development (Groups 1 and 6) were among the most enriched, and included genes coding for key effectors of lineage specification for the EPI (*Nanog*, *Klf4*, and *Sox15*) or the TE (*Tead4*, *Tfap2c*, *Gata3*, *Klf5* and *Cdx2*). These genes were downregulated in *Nr5a2* mzKO and mzDKO embryos mainly at the 8C stage (**Fig 4F**). Some exceptions exist, such as genes with an early developmental function (*Sox15*, *Tead4) (37, 38),* which were downregulated from the L2C stage, and PrE genes, that were upregulated (*Gata4, Sox17, Pdgfra*; **Fig 4F**). Thus, NR5A2 appears to regulate expression of many key TFs controlling blastocyst formation, in line with the results of our immunofluorescence analyses displaying perturbed NANOG and GATA3 expression, and previous knockdown studies (*15*). Another cluster of gene ontology terms (Group 2) was related to the mitotic and cell division defects observed in *Nr5a2* mzKO and mzDKO embryos (**Fig 4E**). This cluster included genes important for the mitotic regulation of the centromeres (*Cenpa*, *Haspin*, *Borealin* and *Shugosin*) or more general regulators of the cell cycle (members of the APC/C complex – *Cul3* and *Ube2c* –, the mitotic *Cdk1* kinase and its regulators *Cdc25* and *Mcph1*, or the G1/S regulator *Cyclin E*). Many of these genes were downregulated from the 2C or 4C stage (**Fig 4F**), likely underlying the impaired ability of mutant blastomeres to proliferate and properly partition the genetic material during mitosis. Finally, a cluster of genes associated with DNA repair, homologous recombination and telomere maintenance (Group 7) might also be related to the low cell numbers observed in *Nr5a2* mzKO and mzDKO embryos (**Fig 4E**). These include

telomeric proteins and genes involved in DNA replication and repair that have a central function in telomere maintenance (**Fig S4C**), preferentially by the alternative lengthening of telomeres (ALT) pathway (*39*), which is active in early embryos (*40*). Notably, nuclear receptors participate in this pathway by binding to variant telomeric repeats in cancer cells (*41*). Thus, the dysregulation of this set of genes might belong to a broader response to telomere damage in the absence of NR5A2/ESRRB, which bind and affect chromatin accessibility at specific subtelomeric regions in embryos and ESCs (**Fig S4D, E**).

We also observed that both single-copy genes and *MERVL/MaLR* transcripts characteristic of ZGA displayed elevated levels in 8C mutant embryos (**Fig S4F, G**), suggesting that NR5A2 may repress 2C-specific transcripts as development proceeds. Accordingly, elevated expression of these transcripts was also observed in KO ESCs (**Fig S4H, I**) (*16*), which are known to sporadically transit into a state that resembles that observed in 2C embryos (2C-like) (*42*). Thus, we turned our attention to ESCs, to assess whether the upregulation of 2C-specific transcripts in mzKO 8C embryos could be due to a direct repressive role of NR5A2 (*43*). Using previously described ESCs (*17*), we found that the fast and concomitant depletion of both ESRRB and NR5A2 led to an increase in the fraction of 2C-like ESCs (**Fig S4J**), as judged by the expression of reporters driven by *MERVL LTRs* (*42*). SINE elements, which normally remain expressed in 8C embryos, were downregulated at this stage (**Fig S4F**), suggesting that the positive regulation of this class of repeats by NR5A2 serves roles that go beyond 2C-specific, and ZGA-related functions (*12*, *15*, *44*).

These results suggest that, in addition to regulating lineage drivers and cell cycle genes, NR5A2 may play a role in telomere homeostasis and contribute to silence the 2C-state when development proceeds. We conclude that NR5A2 controls the progression of the embryo from ZGA to the morula stage, coordinating the activation of lineage determinants with that of genes ensuring the execution of essential housekeeping functions.

NR5A2 supports chromatin accessibility at a subset of its targets.

We then asked if the observed gene expression changes can be ascribed to the direct transcriptional control by NR5A2 and attempted to clarify the underlying mechanisms. Previous studies have proposed that NR5A2 is a pioneer TF, binding nucleosomal DNA to promote chromatin accessibility in early embryos (*45*). We thus set out to investigate the effects

of the complete loss of expression of this TF and link changes in chromatin status to gene expression taking two complementary approaches. First, we measured changes in chromatin accessibility performing ATAC-seq (assay for transposase-accessible chromatin using sequencing) (*46*) in embryos developing in the complete absence of NR5A2 until the 8C stage, processing single mzKO or control embryos, and genotyping after tagmentation. Alternatively, we analyzed pools of embryos in which NR5A2 loss-of-function was induced from the early 2C stage, when zygotic *Nr5a2* is first expressed, driving allelic recombination using *ROSA*$^{Cre\text{-}ERt2}$ in culture. Accessibility was measured on pools of embryos less than one day after completing tamoxifen treatment to mitigate the inherent limitations of single-embryo analyses and capture relatively short-term, and likely direct, responses. To concentrate on direct regulation, we also focused our analysis on a set of 16,958 genomic regions bound by NR5A2 in 8C embryos (*15*).

We observed that NR5A2 bound in similar proportions accessible and low/non-accessible regions (8228 and 8730, respectively; **Fig 5A, S5A and Table S4**). Hence, while targeting non-accessible regions is in line with a pioneer TF activity (*45*), NR5A2 binding does not necessarily lead to accessibility. To identify potential differences between the two classes of sites, we analyzed the DNA sequence of these loci. NR5A2 motifs were enriched at the vast majority of binding regions, suggesting that recruitment is driven by direct binding to DNA. Homeobox motifs, which are targeted by OBOX factors (*11*), were also present throughout NR5A2 occupied regions. However, the presence of high quality NR5A2 or homeobox motifs did not correlate with accessibility. In keeping with this, repetitive elements associated with NR5A2 and OBOX binding (*11, 12*) showed limited differential enrichments at NR5A2 binding regions: *SINE* elements are enriched in non-accessible chromatin and *ERVLs* concentrate at accessible regions (**Fig 5A**), but this distinction is not absolute and the presence of a particular repeat family does not explain accessibility at NR5A2 sites. In contrast, motifs for other TFs that play key roles in early embryos, especially KLF factors (*47*) and TFAP2c (*13*), concentrated in the proximity of NR5A2 sites at the most accessible regions (**Fig 5A, S5A**). We conclude that NR5A2 can bind DNA alone but cooperates with specific classes of developmental TFs to make chromatin accessible.

Accordingly, the mzKO of *Nr5a2* resulted in a moderate reduction in ATAC signal at approximately half of accessible targets, while only 8% of these sites showed increased accessibility (**Fig S5B-C**, ATAC Up and Down, respectively). Moreover, while loss of

accessibility could also be observed after inducible *Nr5a2* inactivation, sites displaying increased ATAC signal in mzKOs showed little changes, suggestive of indirect effects (**Fig S5B-C**). Therefore, NR5A2 appears to contribute to increasing chromatin accessibility at its targets.

Gene regulation by NR5A2 employs distinct mechanisms.

We finally sought to relate changes in chromatin accessibility to gene regulation. Regions proximal (+/-50 Kb) to genes activated by NR5A2 more frequently lost accessibility in KO embryos, compared to those near repressed targets (**Fig 5B, Table S5**). Similarly, NR5A2 regulated genes were found to be enriched near NR5A2 bound and accessible regions. In particular, genes activated by NR5A2 showed strong enrichment near regions that lost accessibility in KO embryos (**Fig 5C**). A reciprocal relation, between repressed genes and regions gaining accessibility, was also observed, although it appeared weaker. Indications of such direct control are evident both globally, and when focusing on the regulation of genes previously identified as linked to the determination of ICM or TE fate (*48*), or involved in housekeeping functions such as mitosis (**Fig 5D**, see methods). Regions near positive NR5A2 target genes belonging to these classes showed a more pronounced reduction in accessibility than those near non-responsive or repressed genes, or near distal regulatory elements or promoter regions not bound by NR5A2, which remain largely unaffected (**Fig 5D**). However, NR5A2 does not require establishing accessibility at binding sites to regulate gene expression, since non-accessible but bound regions are also enriched near regulated genes (**Fig S5D**). Therefore, NR5A2 appears to directly activate expression of genes important for embryonic development and lineage specification, as exemplified by *Tead4* (**Fig 5E**), often, but not necessarily, using mechanisms that involve the enhancement of chromatin accessibility at regulatory regions bound by other TFs. Intriguingly, we observed that, at the 8C stage, genes regulated by NR5A2 generate chimeric transcripts originating in *ERVL* elements (**Fig 5F**), as more commonly reported during ZGA (*26, 27*). These *ERVL* elements, linked to positive targets, are often immediately adjacent to *SINE* repeats bound by NR5A2, which is rarely observed for *ERVLs* (**Fig S5E**). Coupled to our previous analyses, this indicates that binding to *SINE* elements enables gene regulation by NR5A2 well beyond the 2C stage, and reveals an original, if not widespread, partnership between *SINE* and *ERVL* repeats in the control of gene expression.

Altogether our analysis identifies NR5A2 as an essential developmental TF that directly regulates gene expression during the period that links the first wave of embryonic transcription to the beginning of lineage specification.

Discussion

In this study, we have analysed embryos developing *in utero* in the complete absence of NR5A2. We show that this orphan nuclear receptor performs key stage-specific functions: NR5A2 from maternal and zygotic sources starts acting during the period of ZGA to modestly influence transcription in 2C embryos, playing a minor but measurable role in the activation of the embryonic genome; however, it is only after ZGA that zygotic NR5A2 progressively manifests its essential role, coordinating basic functions such as proliferation and the maintenance of genomic stability with the control of the transcriptional events that prepare the onset of lineage specification.

We show that *Nr5a2* mzKO and zKO embryos arrest at the early morula stage. Previous knockout studies (*33-35*) reported a later phenotype for zKO mutants. The basis of this discrepancy is unclear but might be linked to different genetic backgrounds, since previous reports used 129Sv/MF1 or mice of unspecified background, and our animals are mostly C57BL/6 (see methods). Irrespectively, independent work has also shown profound detrimental effects on pre-implantation development upon *Nr5a2* knockdown, the use of chemical inhibitors, and gene editing in zygotes (*12, 14, 15*). Therefore, together with our findings, these reports establish NR5A2 as a key TF for early mouse embryogenesis.

However, major contradictions between previous studies made it difficult to appreciate the precise function and importance of NR5A2. First suggested to affect expression of a vast number of genes during ZGA, leading to an extremely early embryonic arrest (*12*), NR5A2 was subsequently shown to have a later role (*14, 15*). All of the above studies analyzed embryos *ex vivo*, after interfering with *Nr5a2* by a variety of technical means, which might explain the diverging results. In contrast, we ablated either maternal, zygotic or the combined maternal and zygotic expression of *Nr5a2* using gene knockouts, which is crucial to test an involvement during the first stages of mouse development. Analyzing transcription before, at the onset, and after ZGA, we exclude an essential role for NR5A2 in initiating embryonic transcription. However, we show that disrupting both maternal and zygotic *Nr5a2* affects a small set of genes

characteristically expressed in early 2C embryos, including known regulators of this stage, which could, in turn, contribute to the minor transcriptional changes we report. Of note, while challenging previous conclusions, our results are not incompatible with a study in which *Nr5a2* was knocked-down by small interfering RNA in oocytes, that reported the downregulation of only a few hundred genes at the E2C stage, and negligible effects at the L2C stage (*12*). In the light of these findings, the results obtained through the use of chemical inhibitors, which have a strong impact on ZGA, can be put in context. While directed against NR5A2, SR1848 strongly inhibits NR2C2 (*12*) and possibly more nuclear receptors expressed at the 2C stage. Given that these TFs often share the same DNA binding motif, that we have shown a high redundancy in their chromatin association (*16, 17*), and that their expression is not affected in *Nr5a2*/*Esrrb* knockouts, it is possible that not NR5A2 alone, but this class of TFs as a whole, may be essential to initiate zygotic transcription in mammals. Two non-mutually exclusive scenarios can be envisioned. First, NR5A2 may promote expression of key regulators of the 2C stage in concert with other nuclear receptors such as RARG, which is active in this context (*6*). Indeed, the motif identified at genes responding to NR5A2 in 2C stage embryos, which matches *SINE B1/Alu* repeats (*12*), is constituted of a direct tandem of two basic nuclear receptor response elements (DR0), a configuration that promotes heterodimerization between nuclear receptors (*49*), as shown for RXR-RAR (*50*). Second, the ability of NR5A2, ESRRB and other nuclear receptors such as NR2C2 (*41*) to bind subtelomeric regions, where the genes for some 2C regulators such as *Zscan4* and *Tcstv1/3* are located, might also promote expression of these genes. More generally, and as illustrated by DPPA2 (*9*), KLF5 (*47*) or DUX (*51*), our results on NR5A2 highlight the notion that many sequence specific TFs, while not maternally inherited and themselves upregulated during ZGA, engage in a reinforcing feedback loop that boosts gene activation. None of these factors is individually essential for ZGA, suggesting that these activities might operate redundantly downstream of one or more essential maternal triggers of the process. The recent description of the importance of maternally inherited OBOX factors during ZGA (*11, 25*), and the observation that *Nr5a2* is activated by these proteins and controls an overlapping set of genes, supports this hierarchical model.

Furthermore, our work not only clarifies the role of NR5A2 during ZGA, but also unmasks its potential involvement in terminating this stage. Indeed, we observe ectopic expression of 2C-specific genes and *ERVL* repeats in mutants at the 8C stage. Together with the finding that a similar upregulation is observed in ESCs lacking NR5A2/ESRRB, these results suggest that the

two orphan receptors, possibly through a change in cofactors, might later switch to the repression of 2C genes, conferring robustness to developmental progression.

Our study further establishes that *Nr5a2* function becomes determinant at the 8C stage. This is precisely the period when a second major wave of transcription, named mid-preimplantation gene activation (MGA), has been proposed to take place (*52*). Indeed, gene expression steadily increases after ZGA, with expression of many genes peaking only at the 4-8C stage, when conventional features of the transcriptional cycle are also gradually established (*53*). It will therefore be important to test whether NR5A2 and other nuclear receptors, that can interact with Mediator and the general transcriptional machinery (*17, 54*), participate in establishing these mature traits of gene regulation. Gene regulation during this transition period seems to be executed through *SINE* elements, which remain robustly bound by NR5A2 at the 8C stage. We show that NR5A2 binds to these repeats both on closed and open chromatin, confirming its role as a pioneer TF (*45*), but we also report that NR5A2 may require cooperative binding by other TFs to establish accessibility. Opening chromatin does not seem to be a prerequisite for regulating transcription, as an enrichment of bound and inaccessible regions is observed near responsive genes. Studying the mechanisms through which nuclear receptors activate and repress gene expression during early development will shed light on the basic principles through which TFs determine cell fate.

The regulation of this wave of gene expression by NR5A2 is required for morula formation, as already shown by studies analysing embryo development in culture (*15*). This is a role of crucial importance: while the players controlling lineage determination have been extensively characterized, which factors prepare the embryo for the onset of these cell fate choices is only beginning to be understood. NR5A2 activates both TE and EPI genes and should therefore be considered as a TF promoting developmental bi-potentiality, as previously proposed (*13*), and as also shown for KLF5 and TFAP2c (*13, 47, 55*). Notably, inactivation of genes coding for TFs controlling TE fate, such as *Tead4* (*38, 56*) and *Cdx2* (*57, 58*), leads to the closest phenotypes to that observed in *Nr5a2* mutants. Given the fact that expression of *Tead4* and *Cdx2* is severely altered in *Nr5a2* KO embryos, as is the expression of *Klf5* and *Tfap2c*, it is possible that these genes might be important mediators of the lineage specification phenotype of *Nr5a2* KO embryos. In contrast, interfering with ICM/epiblast regulators has consequences that usually become evident later (*59-63*). This is possibly because that of the ICM is a transitory state that resolves only later, during the formation of the PrE and Epibalst. In this

regard, it has been shown that NR5A2 prepares ICM/epiblast enhancers for future activation (*15, 64*). Since NR5A2 is required, together with ESRRB, for the maintenance of ESCs, it now remains to be understood in detail how the two nuclear receptors take part in the establishment of pluripotency *in vivo*. Finally, NR5A2 appears to be unique in its early role in supporting proliferation and genomic stability during cleavage stages. This activity might be exerted either indirectly, by controlling genes important for assembling the mitotic machinery or repairing DNA, or by a yet to be determined direct involvement in the execution of these processes, including the maintenance of telomere homeostasis.

Given the multiple functions observed for NR5A2, exerted at different stages, the reported roles of other nuclear receptors (*6*), and the potential for a collective activity (*16, 17*), we speculate that a changing set of nuclear receptors might control the ordered progression of pre-implantation development. NR5A2, with partners such as RARG or NR2C2, which are involved in controlling the 2C-stage and telomere homeostasis (*6, 41*), might guide progression through the first divisions and ZGA (*12*). Next, NR5A2 alone, or with the support of yet to be identified nuclear receptors, would confer blastomere bi-potentiality and prepare for the segregation of ICM and TE (*13-15*). Finally, ESRRB and NR5A2might regulate together the segregation of the PrE (*65-68*) from the epiblast (*16, 69, 70*), allowing the establishment of pluripotency.

Materials and Methods

Animals and embryos production

Mice

All experiments were conducted according to the French and European regulations on care and protection of laboratory animals (EC Directive 86/609, French Law 2001-486 issued on June 6, 2001) and were approved by the Institute Pasteur ethics committee (dap160016, MESR 23784). Animals were kept on an inverted day/night cycle with 12 hours light cycle from 1 pm to 1 am.

Originally, the *Nr5a2*$^{flox/flox}$ (B6.129S-*Nr5a2*^{tm1Sakl}/DjmJ, RRID:IMSR_JAX:024054) is a congenic strain on C57BL/6 background (*19*), the *Esrrb*$^{flox/flox}$ (*Esrrb*$^{tm1.1Nat}$/J, RRID:IMSR_JAX:007674) strain is on mixed but predominantly C57BL/6 background (*18*), and *Zp3Cre*$^{tg/tg}$ (Tg(Zp3-cre)93Knw/J) is maintained on a mixed 129Sv/C57BL/6 background. The mating schemes used to generate *Zp3Cre*$^{tg/0}$*; Nr5a2*$^{flox/del}$ and/or *Esrrb*$^{flox/del}$ females, and *Nr5a2*$^{del/+}$ and/or *Esrrb*$^{del/+}$ males, progressively brought the genetic background to C57BL/6. *CAG::H2b-Egfp* (*71*) mice are essentially on CD-1 background (*20*), and *Rosa:CreER*^{t2} (Gt(ROSA)26Sortm2(cre/ERT2)Brn) was on a predominantly C57BL/6 background, and has

been brought to C57BL/6. To generate the litters quantified in Fig. 1a, *Zp3Cre*$^{tg/0}$; *Esrrb*$^{del/flox}$, *Zp3Cre*$^{tg/0}$; *Nr5a2*$^{del/flox}$, or *Zp3Cre*$^{tg/0}$; Nr5a2$^{del/flox}$; *Esrrb*$^{del/flox}$ females were crossed with wild-type CD1 males. Littermate females (*Zp3Cre*$^{tg/0}$*; Esrrb*$^{floxl+}$ and/or *Nr5a2*$^{flox/+}$) were used as controls. The number of females used and litters obtained for each genotype is listed in Table S1. P-values for the differences in litter sizes between KOs and the respective controls were obtained performing a Mann-Whitney test.

Generation of Esrrb-TdTomato mice

Esrrb$^{TdTomato/+}$ ES cells were generated by targeting an IRES-TdTomato-LoxP-BlasticidinR-LoxP cassette at the stop codon of *Esrrb*. This was achieved by electroporating 10 µg of a linearised plasmid including homology arms to the locus in CK35 ES cells (*72*). Single resistant colonies were expanded and cells carrying correctly targeted alleles identified by PCR on genomic DNA and sequencing. ESCs were injected in mouse (C57BL/6N) blastocysts which were then transferred to pseudopregnant (C57BL/6 x DBA/2)F1 females. Chimeric males in the progeny were crossed with C57BL/6J albinos (B6(C)/Rj-*Tyr*$^{c/c}$) females to establish germ line transmission. The blasticidin resistance floxed cassette was removed by crossing with *PgkCre*$^{tg/0}$ mice (*73*).The reporter *Esrrb*^{TdTomato} line was backcrossed to CD-1 mice 3 times and maintained through intercrosses.

Embryo and oocyte production

All embryos were staged according to the date of the vaginal plug (E0.5). Except for E4.25 (late blastocyst) embryos, that were recovered from natural mating, embryos were collected from females hormonally stimulated by intraperitoneal injection of pregnant mare serum gonadotropin (PMSG) (2.5 IU/mouse, Chrono-gest ref A214A01 Cosmo Bio), followed 42-48h after by an additional injection of human chorionic gonadotropin (hCG) (5IU/mouse, Chorulon ref A242A01 Cosmo Bio). Immediately after the second injection females were mated with appropriate males. Embryos were collected from females 20, 30, 52-4, 65-75 or 95 hours after hCG in EmbryoMax FHM HEPES Buffered Medium w/o Phenol Red (Sigma Aldrich, Cat# MR-025D) for zygote, early 1.5 (early 2C), late E1.5 (late 2C/early 4C), E2.5 (8C / morula) or E3.5 (blastocyst) stages respectively.

Standard number of embryos (12, 15) were used as biological replicates for RNA-seq (4 to 29 embryos per condition) and ATAC-seq (8 controls and 13 mzNrKOs). For ATAC-seq analysis of tamoxifen-induced *Nr5a2* deletion, 9 controls and 8 induced KO embryos were pooled to increase the depth of the analysis. For other inducible KO experiments, for rescue experiments, or for the analysis of genotype frequencies observed in crosses generating mzEKOs, mzNrKOs, mzDKOs, zNrKOs, or control embryos, the number of embryos per genotype, and number of females used are listed in Table S1.

For the immunofluorescence presented in Figure 3A-B: embryos were collected from either *Zp3Cre*$^{tg/0}$*; Nr5a2*$^{flox/del}$ females mated with *Nr5a2*$^{del/+}$ males or from *Zp3Cre*$^{tg/0}$*; Esrrb*$^{flox/del}$ females mated with *Esrrb*$^{del/+}$males. For immunostaining and the analyses presented in Figure S3A,G on zNrKO embryos, *Nr5a2*$^{del/+}$ mice were backcrossed once to CD-1 before being intercrossed. For all transcriptional analyses on mzNrKOs, mzEKOs, mzDKOs or control DHets performed at or after the late 2C stage and for the immunofluorescence presented in Figure 3F and S3H: embryos were collected from *Zp3Cre*$^{tg/0}$*; Nr5a2*$^{flox/del}$; *Esrrb*$^{flox/del}$ females, mated with *Nr5a2*$^{del/+}$; *Esrrb*$^{del/+}$ males, and inherited either WT or deleted *Nr5a2*/*Esrrb* alleles from the male. For the transcriptional analyses on zNrKOs and controls at the late 2C stage:

embryos were collected from *Zp3Cre*$^{tg/0}$*; Nr5a2*$^{flox/+}$; *Esrrb*$^{flox/+}$ females, mated with *Nr5a2*$^{del/+}$; *Esrrb*$^{del/+}$ males, and inherited two deleted *Nr5a2* alleles and at least one intact Esrrb allele. For the transcriptional analyses on mNrKO, zNrKO and mzNrKO mutants, or Het and quasi-WT controls at the early 2C stage presented in Figure 2, or for embryos presented in Figure 1B: quasi-WT embryos were generated from *Nr5a2*$^{flox/+}$ females not expressing *Zp3:Cre* and inherited an intact *Nr5a2* allele from a *Nr5a2*$^{del/+}$ male; Het embryos were generated from *Nr5a2*$^{flox/+}$ females expressing *Zp3:Cre* mated with a *Nr5a2*$^{del/+}$ male, and inherited one intact and one deleted allele from either parent; zNrKO embryos were generated from *Nr5a2*$^{flox/+}$ females expressing *Zp3:Cre which were mated with* a *Nr5a2*$^{del/+}$ male, and inherited two deleted *Nr5a2* alleles; mNrKO embryos were generated from *Nr5a2*$^{flox/del}$ females expressing *Zp3:Cre, and* inherited an intact *Nr5a2* allele from a *Nr5a2*$^{del/+}$ male; mzNrKO embryos were generated from *Nr5a2*$^{flox/del}$ females expressing *Zp3:Cre, and* inherited a deleted *Nr5a2* allele from a *Nr5a2*$^{del/+}$ male. For the transcriptional analyses on oNrKO and controls MII oocytes presented in Figure 2, or for oocytes presented in Figure 1B: quasi-WT oocytes were generated from *Nr5a2*$^{flox/+}$ females not expressing *Zp3:Cre*, Het oocytes were generated from *Nr5a2*$^{flox/+}$ females expressing *Zp3:Cre*; oNrKO oocytes were generated from *Nr5a2*$^{flox/del}$ females expressing *Zp3:Cre*. For Figure S3C,D, inducible NrKO embryos were recovered at the 4C stage from *Nr5a2*$^{flox/flox}$ females, crossed with *Rosa*$^{CreERt2\ KI/+}$; *Nr5a2* $^{del/+}$ males. Embryos were then cultured for an additional 43h until controls reached the blastocyst stage, in the presence or absence of 750 nM 4-hydroxytamoxifen (Sigma Cat #H7904) for the first 24h. For ATAC-seq on inducible NrKO embryos, *Nr5a2*$^{flox/flox}$ females were crossed with *Nr5a2*$^{flox/flox}$ males, with either parent also homozygous for *Cre* KI (*Rosa*$^{CreERt2\ KI/KI}$), and embryos were dissected at the early 2C stage and kept in culture for additional 45h until they reached the 8C stage, treating or not with 750nM 4-hydroxytamoxifen (Sigma, Cat # H7904) for the first 24h. For the microinjection experiments presented in Figure 3E, mzNrKO and control Het embryos were collected at early 4-cells stage from *Nr5a2*$^{flox/del}$ females expressing *Zp3:Cre, and* inherited either a deleted or intact *Nr5a2* allele from a *Nr5a2*$^{del/+}$ male. To visualize nuclei/chromatin during the live imaging experiments presented in Figure 1C and S1D, female *Esrrb*$^{TdT/+}$ mice were crossed with heterozygous *CAG::H2b-Egfp* mice.

Embryo genotyping

Embryos were lysed in 5-15 µl of lysis buffer (10 mM Tris pH8 ; 50 mM KCl , 0.01 % gelatine + 300 µg/ml PK) at 56 °C for 1h followed by 10min at 95 °C in a thermocycler. 5 µl of lysate were used for 21 rounds of amplification using Phusion High-Fidelity DNA Polymerase (NEB, M0530L) in a total volume of 25 µl (13 µl H2O, 5 µl buffer 5X, 1.25 µl P1+P2+P3 (10 µM each), 0.5 µl dNTP (10 mM), 0.25 µl Phusion). 1 µl of the pre-amplified material was then used for 21-24 cycles of a similar nested PCR reaction, also in 25 µl. The amplification program for both steps was as follows. Tubes were held 30 sec at 98°C, followed by 21-24 cycles of: 15 sec 98°C, 15 sec 64°C, 1 min 72°C. Then tubes were held 3 min at 72°C. Primers for 1st step *Esrrb*: *Esrrb* del fw CACGGTCAGCTTCCACTTTT, *Esrrb* flox fw ACCATTCAAGGTGCGTGGTG, *Esrrb* rv TGAGTCTTAGAGTTGAAATCCTTGT. Primers for 2nd step *Esrrb*: *Esrrb* del fw GGGTCTCTGATTTGAAGTTTACG, *Esrrb* flox fw TATTAATTCCAAGTCTCGTTTCCTG, *Esrrb* rv TATTAATTCCAAGTCTCGTTTCCTG. Primers for 1st step *Nr5a2*: *Nr5a2* del fw CATAAGGGCTCAGTGGCAC, *Nr5a2* flox fw TAATAACTAAGAAGCAGAAAGCATGC, *Nr5a2* rv CTTCACTGGCTGCCAAGCTG. Primers for 2nd step *Nr5a2*: *Nr5a2* del fw: TACTGGTGATTTCCTGAGAGTACA, *Nr5a2* flox fw TAAGAAGCAGAAAGCATGCCAAG, *Nr5a2* rv CATTCTTCGGCAGTTGAGAGTGA.

RNA in vitro synthesis and embryos microinjection

The coding sequence for NR5A2 (ENSEMBL Nr5a2-205) was cloned into a pRN3 plasmid (*74*) and PCR products derived from this construct using the primers T7-PRN3 fw (CTGTCTGTAATA CGACTCACTATAGGGAGGGAACAAAAGCTGGAGCTAAGCTT) and PRN3 rv (GGGTACCGGCCGACT TGGCCCTGCA). H2B-mCherry was amplified from the vector pCAG-H2B-mCherry with primers H2B-mCherry RNA T7-F (GTAATACGACTCACTATAGGGATGCCAGAGCCAGCGAAGTCTGCT) and mCherry-R (TTACTTGTACAGCTCGTCCATGCC). PCR products were used as template for *in-vitro* transcription using an mMesseage mMachine T7 ultra kit (Thermo Scientific, AM1345). mRNAs were purified using LiCl/ethanol precipitation and resuspended in Brinster's buffer (10 mM Tris pH7.5; 0.25mM EDTA). For microinjections, embryos were transferred to a depression slide in EmbryoMax FHM HEPES Buffered Medium w/o Phenol Red and injected using a Leica inverted microscope equipped with a Leitz micromanipulator. For *Nr5a2* and *H2B-mCherry* expression, mRNA was diluted at 50 ng/µl and 300 ng/µl respectively and injected into the cytoplasm of each blastomere of 4-cell stage embryos. Embryos were cultured for 48h in Embryo Love medium (Embryotech, ETECH-EL-50) before fixation, imaging and genotyping. Four and two injections experiments were performed for *Nr5a2* and *H2B-mCherry* mRNA, respectively.

ES cell culture conditions

ES cell culture conditions

ES cells were cultured on 0.1% gelatine (SIGMA, G1890-100G) in DMEM + GlutaMax-I (Gibco, 31966-021), 10% FCS (Gibco 10270-098), 100 µM 2-mercaptoethanol (Gibco, 31350-010), 1× MEM non-essential amino acids (Gibco, 1140-035) and 10 ng ml−1 recombinant LIF (MILTENYI BIOTEC, 130-099-895). Cells were passaged 1:10/1:20 every 2–3 days.

ESRRB/NR5A2–AID 2C:TdT ESC derivation and flow cytometry

ESRRB/NR5A2–AID cells from (*17*), in which NR5A2/ESRRB can be depleted by Auxin treatment, were stably transfected with a reporter of the 2C-like state(*42*), driving expression of *TdTomato* under the control of an *ERVL* LTR. A clonal population in which TdTOMATO and ZSCAN4 expression correlates was selected and cells treated with or without auxin for 3 days before analysing fluorescence levels with an LSR II flow cytometer system (BD Biosciences). Cells showing a fluorescence signal at least one logarithmic unit over background were considered as positive.

Imaging

Embryo immunofluorescence

Embryos were fixed for 20 min at room temperature in 4% paraformaldehyde (Euromedex ref 15714) or on ice in 8% paraformaldehyde, 1X PhosSTOP for NR5A2 IF detection. Embryos were incubated in permeabilized/blocking solution in PBS with 0.1% Triton-X100 and 10% donkey serum (or 1% BSA for NR5A2 staining) for 30 minutes at room temperature. Embryos were treated with primary antibodies diluted in PBS with 0.1% Tween20 and 10% Donkey serum solution at 4°C overnight. After three washes with PBS with 0.1 % Tween20, embryos were incubated with secondary antibodies diluted 1/300 in PBS with 0.1% Tween20 for at least 1-2 hours. Hoechst 33342 (1.6 µM, Thermo Fisher) was included during incubation with

secondary antibodies. The following antibodies were used: anti-NANOG (1:100; CosmoBio ref REC-RCAB0002PF), OCT4 (1:100; Santa Cruz ref sc5279), SOX2 (1:200; eBioscience, ref 14-9811-80), NR5A2 (1:100; R&D/Perseus ref PP-H2325-00), GATA6 (1:100; Cell Signaling Technology, ref 5851), GATA3 (1:100; R&D, ref AF2605), AlexaFluor 488 Donkey anti-mouse (1:300; Invitrogen, ref A21202) or anti-rat (1:300; Invitrogen, ref A21208), AlexaFluor 546 Donkey anti-rabbit (1:300; Thermo Fisher, ref A10040) and AlexaFluor 647 Donkey anti-goat (1:300; Thermo Fisher ref A21447). The images presented are representative of the results obtained in multiple embryos from independent females. In specific, where not specified in figure panels, embryo numbers are as follows. For figure 1B, Oocytes: Het n=11, oNrko n=16; 2C: Het n=12, mzNrKO=9; 4C: WT/Het n=9, zNrKO=3; 8/16C: WT/Het n=9, zNrKO=6. For figure S1A, L2C n= 6, 4C n=11, 8C n=14, EB n=1, MB n=6. For figure 3A, Het n=12, mzEKO n=17. For figure 3B, Het=16, NrKO n=14. For figure 3F, from mzDKO crosses, DHet n=9, mzEKO n=7, mzNrKO n=4, mzDKO n=5; from equivalent experiments on embryos from mzNrKO crosses, Het n=15, mzNrKO n=9; from equivalent experiments on embryos from zNrKO crosses, WT/Het n=33, mzNrKO n=7. For figure S3A, WT/Het n=63, KO n=15. For figure S3G WT/Het n=17, NrKO n=4. For Figure S3H, DHet n=6, EKO n=5, NrKO n=3, DKO n=5. For plots in figure S3F, the mean immunofluorescence signal was quantified after 3D segmentation of the nuclei using the Imaris software (Oxford Instruments), based on Hoechst staining. For each channel, the mean signal of each nucleus in heterozygous control embryos was plotted against the respective Z position (center of segmented volume), and an exponential curve fitted to the data and used to normalize the decay in fluorescence over the Z imaging axis, in all nuclei plotted. The minimal normalized fluorescence value for each channel was further subtracted from each data point.

Imaging of embryos

All images were obtained by using a Zeiss LSM 900 Airyscan 2 Multiplex 4Y confocal microscope. Embryos were placed individually in wells of a customized micro fabricated device (*75*) in PBS1X, 0.1 % Tween20. Images were acquired using a long working distance Plan-Apochromat objective without immersion (20×/0.8), bi-directional scanning, 2x zoom and 1 airy unit pinhole. Stacks were created with a 2µm step, generating 16-bit images at confocal resolution (1293x1293 pixels).

For live imaging, embryos were cultured for 4 days in a 24-hour pre-equilibrated EmbryoMax KSOM medium (Sigma Aldrich, Cat# MR-121D) covered with FertiCult mineral oil (FertiPro, MINOIL050) and maintained in a humidified incubation chamber at 37°C in 8% CO_2. To minimize phototoxicity for H2B-EGFP and ESRRB-TdT signal acquisition, only 11 z stacks of low resolution (512 x 512 pixels size) were generated per embryo, with a pinhole aperture of 100 µm and 15 min time interval. For figure 1C, 5 embryos were imaged for the “mat only” group and 3 for the “mat + zyg” group.

Single-embryo RNA-seq

Library preparation

The following protocol is derived from FLASH-seq (*23*), with two main modifications: 1/ Genotyping primers are added during the cDNA amplification step; 2/ A phosphate group is added to the 3’ of the template switching primer to exclude the possibility it might prime DNA synthesis during the cDNA amplification step, which would result in tagging single transcripts with multiple UMI barcodes. Only polyadenylated transcripts are detected by this protocol. Embryos were isolated and directly transferred by mouth pipetting to PCR tubes in 5 µl of the

a lysis solution containing: 0.1 µl Triton X-100 (10% v/v, Sigma-Aldrich, T8787), 2 µl dNTP mix (25 mM each, ThermoFisher Scientific, R1121), 0.46 µl 5'Bio-AAGCAGTGGTATCAACGCAGAGTACT$_{30}$VN-3', 100 µM (supplied by IDT), 0.46 µl 5' Biotin-AGTGGTATCAACGCAGAGTACGATCNNNNNNNNrGrGrG-3′Phospate, 100 µM (supplied by IDT), 0.15 µl recombinant RNase Inhibitor (40 U/µL, ThermoFisher Scientific, AM2684), 0.06 µl Dithiothreitol (DTT, 100 mM; part of the Superscript IV kit, ThermoFisher Scientific, 18091050), 1 µl betaine (5M, Sigma-Aldrich, 61962-50G), 0.45 µl dCTP (100 mM, Sigma-Aldrich, DNTP100), water to 5 µL final. Note that the first of the two primers anneals to the poly-A tail of transcripts and the second is a template switching primer that is added to the 5' of transcript during reverse transcription. AGTGGTATCAACGCAGAGTAC is a suppression PCR sequence. The following 4 bases "G(ATC)" form a 3bp barcode that can be varied if multiple samples are to be mixed before tagmentation, and only UMI reads are to be considered. Random Ns form the UMI tag, and rGrGrG are required for template switching. As mentioned, the 3' phosphate inhibits priming in following steps. After transfer in the lysis solution, tubes were incubated 3 min at 72°C in a thermocycler, and immediately transferred on ice. RT-PCR was performed by adding 20 µl of the following mix to each sample: 1.19 µl Dithiothreitol (DTT, 100 mM; part of the Superscript IV kit, ThermoFisher Scientific, 18091050), 4 µl betaine (5M, Sigma-Aldrich, 61962-50G), 0.23 µl magnesium chloride (1 M, ThermoFisher Scientific, AM9530G), 0.48 µl recombinant RNase Inhibitor (40 U/µl, ThermoFisher Scientific, AM2684), 0.25 µl Superscript IV Reverse Transcriptase (200 U/µl, ThermoFisher Scientific, 18091050), 12.5 µl KAPA HiFi Hot-Start ReadyMix (2x, Roche; cat. KK2502), water to a final volume of 20 µl. Samples were then incubated in a thermocycler set to execute the following program: hold at 50°C, and after placing tubes proceed to 60 min at 50°C, 3 min at 90°C, hold at 4°C. Tubes were transferred on ice and 0.5µl of a 100 µM solution of amplification primer (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGCAGTGGTATCAACGCAGAGTACG) were added to each sample. Note that AGATGTGTATAAGAGACAG is the sequence Tn5 uses for transposition. 1µl of a solution of genotyping primers (see section 1d/ Embryo genotyping) at 10µM each was also added to each tube. Tubes were then transferred to a thermocycler, set to the following program: hold at 4°C and after placing tubes proceed to 20 sec 98°C, 20 sec 64°C, 6 min 72°C, followed by 20 cycles of 20 sec 98°C, 20 sec 67°C, 6 min 72°C, adding 12 seconds per cycle. Amplification products were purified using 1x volume of SPRI magnetic beads (homemade, see the SPRI bead preparation and use section), eluting in 15 µl of water. 1 µl was set aside to perform the second step of the nested PCR genotyping protocol (see Embryo genotyping section above). Fragment size distribution was checked on a Genomic DNA Screentpape (Agilent Technologies, 5067-5365/6), with typical lengths ranging from less than 1 kb to more than 10 kb. For tagmentation, each sample was diluted to 150 pg/µl. 2 µl of diluted cDNA was mixed with 2µl of Amplicon Tagment Mix (ATM) and 4 µl of Tagment DNA Buffer (TD) from a Nextera XT DNA Library Preparation Kit (Illumina, FC-131-1024), and incubated for 8 min at 55°C, in a thermocycler. Tn5 was released from DNA by adding 2 µL of Neutralization Buffer (NT), followed by a 5 min RT. Enrichment of UMI containing fragments in the tagmented DNA was performed by adding 5 µL of Nextera PCR Master Mix (NPM) and 1 µl of 5 µM amplification primer (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGCAGTGGTATCAACGCAGAGTACG, as before). KAPA HiFi Hot-Start ReadyMix (2x, Roche; cat. KK2502) can substitute the Nextera PCR Master Mix, adapting volumes as required. Tubes were incubated as follows: 72°C for 3 min, 95°C for 30 sec. This was followed by 5-10 cycles of linear enrichment PCR (cycle numbers can be varied depending on how many UMI reads are required): 95°C for 10 sec, 55°C for 30 sec, 72°C for 30 sec. Holding at 4°C, 4 µl of a 2.5 µM solution of each of the following amplification primers were added on ice: Universal primer,

AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTGTATAAGAGACAG, Indexed primer CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG. The 8 Ns were replaced with the Illumina index sequence of choice. Tubes were returned to the thermocycler held at 4°C, and then subjected to 14 iterations of the cycle: 95°C for 10 sec, 55°C for 30 sec, 72°C for 30 sec. This was followed by 72°C for 5 min. Amplification products were purified using 1-1.4x volume of SPRI magnetic beads, eluting in 30 µl water. Fragment size distribution was checked on a DNA Hi Sensitivity D1000 HS Screentape (Agilent Technologies, 5067-5584/5).

SPRI bead preparation

Preparation: 1 ml Sera-Mag Magnetic SpeedBeads, carboxylated, 1 µm, 3 EDAC/PA5 (GE Healthcare Life Sciences, Cat #65152105050250) were washed 3 times with a TE-Tween solution (10 mM Tris HCl pH 8, 1 mM EDTA, 0.05% Tween 20, pH 8.0) and resuspended in TE-Tween-20% PEG 8000 solution (10 mM Tris HCl pH 8, 1 mM EDTA, 0.05% Tween 20, pH 8.0). DNA purification: 1 or 1.2x sample volumes of SPRI beads were added to each tube and samples transferred to a 96 well plate. After incubating for 5 min, the plate was put on a 96S Super Ring Magnet (Alpaqua, A001322), beads were allowed to separate completely, and the supernatant removed without disrupting the bead pellet. Beads were washed twice with approximately 100 µl of 70% Ethanol and the supernatant completely removed. DNA was eluted in water after separating the beads one last time.

Single-embryo ATAC-seq

ATAC-seq was performed on either i) single freshly dissected 8C mz *Nr5a2* KO or control embryos, or ii) on pools of approximately 20 *Nr5a2* inducible KO (iNrKO) embryos (*Rosa*$^{CreERt2\ KI/+}$*; Nr5a2*$^{flox/flox}$), dissected at the early 2C stage, and kept in culture for an additional 45h until they reached the 8C stage, with or without 750nM 4-hydroxytamoxifen (Sigma, Cat # H7904) treatment for the first 24h. The method was adapted from (*46*) and performed with the following variations. First, the zona pellucida was removed by treatment with acid Tyrode's. After blocking in EmbryoMax FHM HEPES Buffered Medium w/o Phenol Red (Sigma Aldrich, Cat# MR-025D), embryos were transferred in acetylated BSA(1 mg/ml in 1X PBS), and lysed for 3 min in resuspension buffer (RSB; 10 mM Tris-HCl pH 7.4 10 mM, 5M NaCl, 3mM MgCl2) supplemented with 0.1% NP40, 0.1% Tween-20 and 0.01% digitonin in Terasaki wells. Embryos were then washed for 3 min in RSB with 0.1%Tween, and tagmentation performed for 30 min at 37°C in 25 µl total volume of the following reaction buffer: 12.5 ul TD buffer (20mMTris-HCl pH 7.6,10 mM MgCl2, 20 % dimethyl Formamide), 8.25 ul PBS, 5ul water, supplemented with 0.01% digitonin, 0.1% Tween-20, and 1ul of Tagment DNA Enzyme (Illumina, Cat# 20034210). 125 ul of buffer PB were added directly to stop the reaction, and the material loaded on MinElute PCR Purification columns (Quiagen, Cat# 28006), eluting in 17 µl of water. The eluted 14.5ul of transposed DNA, was then mixed with 1.5 µl of 25 µM Ad1.noMX primer, 2.5 µl of 25 µM Ad2 primer from (*76*), 1.5 µl of 10 µM genotyping primers, 19 µl of KAPA HiFi HotStart 2x master Mix (Roche; cat. KK2502) and 1ul of pre-diluted (1:10) of Quant-iT Picogreen dye (Invitrogen, P11496). The amplification was performed on a Roche LightCycler 480 II instrument (Cat# 05015243001) using the following program: 72°C 5 min, 98°C 30 sec. This was followed by repetitions of: 98°C 10 sec, 63°C 30 sec, 72°C 1 min, until reaching mid-exponential phase. Amplified material was then purified on SPRI beads (1.4 sample/beads volume). 1ul was used to measure the DNA concentration with a Qubit 3 and the provided reagents (Invitrogen, Q33218). 1 -5 ng

of DNA were used to check fragment size with a D1000 High Sensitivity Screentape and appropriate reagents (Agilent Technologies, 5067-5584/5) on an Agilent 2200 Tapestation. Genotyping was performed on pre-amplified material as for FLASH-seq samples.

Computational Methods

Single-embryo RNA-seq analysis

Single-end sequences (50-85bp, using 50bp sequencing kits) were generated using Illumina Next-seq 500 or 2000 instruments. After trimming Illumina adapter sequences with cutadapt (*77*) using options -b CTGTCTCTTATA --minimum-length=1, reads were split in UMI-containing or not with Julia (*78*), trimming 11bp after the sequence "AAGCAGTGGTATCAACGCAGAGTACGATC" (see the FLASH-seq section above). For UMI, reads were further processed in Julia to remove the invariant adapter sequence, extract the identifier, and the UMI/ transcript sequence pairs used to remove duplicates. Typically UMI libraries displayed 30% of unique reads. Note that a much lower duplication rate is to be expected while analysing conventional reads, as in our protocol UMIs are anchored to the 5' of transcripts, and thus reads are not randomly distributed, and less diverse. Conventional or UMI reads were then aligned to the mm10 genome using STAR (*79*) and quantified by RSEM (*80*) through the RSEM-STAR pipeline, using gencode.v25 annotations and additional options "--seed 1618 --calc-pme --calc-ci --estimate-rspd --forward-prob 0.5" for conventional reads and "--seed 1618 --calc-pme --calc-ci --estimate-rspd --forward-prob 1" for UMI reads. RSEM estimated read counts per sample were rounded for use with DESeq2 (*81*). For all differential expression tests DESeq2 was run with standard parameters, using a model design accounting for both batch effects in different replicates (or preparation days) and experimental conditions tested. Genes were considered as differentially expressed if their mean expression exceeded 10 TPM in at least one genotype for a given stage, and using a <> 1.5 fold change and <0.1 FDR cutoff. The significance of the overlap between targets identified at successive stages was calculated performing standard one-tailed Fisher's exact tests. Tables detailing gene expression changes in Dux and Obox KO embryos were obtained from (*10*, *11*). For visualizing the enrichment of overlap between DEGs, we displayed the -log10 one-tailed Fisher's P-value of observing the detected number of genes differentially expressed at one stage among the genes responding at the following developmental stage analysed, given the number of genes expressed at both stages (mean expression exceeding 10 TPM).

ATAC-seq and Cut&Run read alignment, peak calling and region definitions

First, fastq files generated in this work, or obtained from the GEO series GSE66581 (*82*), and GSE229740 (*15*) were trimmed using the cutadapt software, and aligned with Bowtie2 (*83*) to the mm10 genome, with options --no-unal, and --no-mixed --no-discordant when relevant. Reads were then filtered based on the alignment score to allow a maximum of 2 mismatches in high quality positions running awk '(int(substr($12,6))>-15)', or awk '((int(substr($(NF-1),6))>-15) && (int(substr($12,6))>-15))' on single or paired-end bam files respectively, and retaining only reads with mates. Reads with MAPQ lower than 30 were then excluded, and duplicates removed. Experimental replicates for each stage were then merged before peak calling. Specifically: for ATAC data from (*82*) or Cut&Run data from (*15*) all experimental replicates were merged separately for the 8C or 2C stage; for our data obtained using single mz *Nr5a2* KO or control 8C embryos, all samples were merged; for our data obtained using pools of inducible *Nr5a2* KO embryos, treated and untreated samples were merged. Peaks were then

called using MACS2 (*84*) with options -q 0.2 -g 1.87e9. This defined NR5A2 bound regions. Accessible regions were instead retained if a peak was called in at least 2 out of 3 datasets: our mzKO or inducible *Nr5a2* KO analyses, and data from (*82*). Accessible and bound regions were defined as the union of overlapping accessible and bound peaks. Blacklisted regions (*85*) were excluded from all datasets. ATAC-seq regions were further classified as losing, or increasing in accessibility based on the mean of two ratios in ATAC signal: that between *Nr5a2* mzKO / *Nr5a2* mzHet and that between untreated / tamoxifen-treated samples (FC <>1.5).

ATAC-seq, Nr5a2 binding and gene expression correlations

To associate genomic regions with genes we followed the following approach. First, the coordinates of the TTS for all genes expressed at the 8C or 2C stage were determined. For this, coordinates of all transcripts annotated in the mm10 genome were obtained from ENSEMBL using the biomaRt package (*86*) (dataset = "mmusculus_gene_ensembl", host="apr2020.archive.ensembl.org"). For each gene, and separately at each embryonic stage, the transcript showing maximal expression was selected by quantifying the count of 5' anchored reads in our UMI datasets (see section describing RNA-seq analyses) in a window of -50/+150 bp from the annotated TSS, using the bamsignals R package (*87*). Only genes showing counts greater than zero were considered. Accessible and/or NR5A2 bound regions falling in a window of +/- 50Kb from the TSS of each gene in a particular class (NR5A2 DEGs, ZGA genes, 2C-specific genes, ICM genes, …) were then selected for further analyses. To assess differences in the distributions of accessibility changes, standard non-parametric one-tailed Kolmogorov-Smirnov tests were used after removing regions with 0 counts in at least one condition. Genes upregulated during ZGA were obtained from (*12*). "2C-specific" genes were selected using tables reporting gene expression during pre-implantation murine development, obtained from GSE45719 (*22*). Genes were selected with expression greater than 10 RPKM at the early, mid or late 2C-stage, and so that their relative expression, setting max to 1 and min to 0, was <0.5 at 4C and <0.3 at all other stages. ICM and TE specific gene lists were obtained from (*48*). To each list were further added all genes associated with the term "inner cell mass" or "trophectoderm" and all child terms, respectively, in the GO consortium database (https:// geneontology.org). These were retrieved using the Bioconductor org.Mm.eg.db. R package. Genes with a role in cell division and partitioning the genetic material were also obtained from the GO consortium, as above, using terms: spindle, chromosome segregation, chromosome separation, cytokinesis, chromosome localization, metaphase, centromeric, chromatid, mitosis. For enrichment plots, we displayed the -log10 one-tailed Fisher's P-value of the enrichment of genes from the group of interest (upregulated/downregulated), at increasing distances from specific genomic regions of interest. For this, we used the test to compare the number of regions falling in a given distance window from the TSS of the selected genes, to the number of regions found at the same distance to all other expressed genes. The test was repeated for discrete increments of the distance window (X axis), and P-values plotted on the Y axis.

Heatmaps of ATAC-seq, Nr5a2 binding, repeats distributions, and motif analyses

To identify occurrence of TF motifs the relative pfm or pwm matrixes were obtained from Jaspar, using the Jaspar2014 (*88*) R package. (KLF5: MA05991; homeobox/OTX2: MA07122, NFYA: MA00603;TFAP2c MA08141; NR5A2: derived from NR5A2 bound and accessible regions using the RSAT software, matches closely MA0505.1), and occurrence and score of TF motifs identified using the TFBT R package (*89*), setting quality thresholds of: 85% for NR5A2, homeobox/OTX2, NFYA and TFAP2c; 95% for KLF5. To plot histograms of motif qualities, for each region, scores returned by TFBT for motifs falling into a +/-150 bp window from the

center of the region analyzed were summed up. Before plotting, cumulative scores were min/max normalized assigning relative scores ranging from 0 to 1. To plot heatmaps of motif distributions in the regions, each motif was extended by +- 40 bp to help visualization, and the score of the motif (corrected for the quality threshold applied to call motifs) added to the bases covered by the motif in the region. Cumulative score matrixes were then min-max normalized. To plot heatmaps of repetitive elements localisation an analogous process was used, but using the RepeatMasker coordinates and score (*90*) for each element derived from the genomic annotation files available for download at repeatmasker.org. All heatmaps were produced using ComplexHeatmap (*91*). NR5A2 bound regions were classified as overlapping a *SINE* or *ERVL* element if one such element was present in RepeatMasker annotations in a window of +-100 bp from the position of the NR5A2 motif closest to the C&R summit. For plotting ATAC-seq signal heatmaps, NR5A2 bound and accessible regions were centered on ATAC-seq summits, and NR5A2 bound but non-accessible regions on NR5A2 Cut&Run summits. NR5A2 Cut&Run heatmaps were always centered on NR5A2 Cut&Run summits. For heatmaps of motifs, those of TF enriched in accessible regions were centered on ATAC-seq summits, and those of NR5A2 and homeodomain TFs on NR5A2 Cut&Run summits, as were the heatmaps of repetitive element distributions.

Repetitive element expression

Non-UMI reads were mapped to the mm9 genome using bowtie2, and split by the RepEnrich2_subset command of RepEnrich2 (*92*), using a mapq value of 30 to subset uniquely mapping from multi-mapping reads. Repeat expression was then quantified by RepEnrich2 using the default repeatmasker file mm9.fa.out.gz downloaded from repeatmasker.org. Differential enrichment analysis was performed using a generalized linear model in EdgeR (*93*), following a standard procedure (https://github.com/nerettilab/RepEnrich2). Repeats were called as differentially expressed using a <> 1.5 fold change and <0.1 FDR cutoff.

Other computational analyses

Gene ontology analysis was performed pooling all differentially express genes, and using the enrichGO function of the ClusterProfiler R package (*94*), with options “ont ="BP", pAdjustMethod = "BH" (Bonferroni-Holm correction for multiple comparisons), minGSSize = 0, maxGSSize = 180 (exclude terms associated with more than 180 genes), pvalueCutoff = 0.01”. Results were plotted with the command emapplot, after calculating the pairwise term similarities and displaying the first 206 enriched terms. Violin and box plots were generated in R using ggplot2 (*95*). Venn diagrams were made with eulerr (*96*). Bigwig files were created from the alignments by the bamCoverage command of deeptools (*97*) with --binSize 20, and coverage and splicing graphs generated from Bigwigs using IGV integrative genomics viewer, and the sashimi plot option. PCA plots were generated in R considering genes differentially expressed after the loss of *Nr5a2*/*Esrrb* at any stage, using the prcomp command, with options center = T.

!

References and Notes

1. J. Rossant, Genetic Control of Early Cell Lineages in the Mammalian Embryo. *Annu Rev Genet* **52**, 185-201 (2018).
2. K. N. Schulz, M. M. Harrison, Mechanisms regulating zygotic genome activation. *Nat Rev Genet* **20**, 221-234 (2019).
3. C. Yu *et al.*, Oocyte-expressed yes-associated protein is a key activator of the early zygotic genome in mouse. *Cell Res* **26**, 275-287 (2016).
4. A. Bhattacharya *et al.*, The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation. *Cancer Res* **63**, 8167-8172 (2003).
5. F. Lu *et al.*, Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell* **165**, 1375-1388 (2016).
6. A. Iturbide *et al.*, Retinoic acid signaling is critical during the totipotency window in early mammalian development. *Nat Struct Mol Biol* **28**, 521-532 (2021).
7. O. Wendling, N. B. Ghyselinck, P. Chambon, M. Mark, Roles of retinoic acid receptors in early embryonic morphogenesis and hindbrain patterning. *Development* **128**, 2031-2038 (2001).
8. O. Kubinyecz, F. Santos, D. Drage, W. Reik, M. A. Eckersley-Maslin, Maternal Dppa2 and Dppa4 are dispensable for zygotic genome activation but important for offspring survival. *Development* **148**, (2021).
9. Z. Chen, Z. Xie, Y. Zhang, DPPA2 and DPPA4 are dispensable for mouse zygotic genome activation and pre-implantation development. *Development* **148**, (2021).
10. Z. Chen, Y. Zhang, Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat Genet* **51**, 947-951 (2019).
11. S. Ji *et al.*, OBOX regulates mouse zygotic genome activation and early development. *Nature* **620**, 1047-1053 (2023).
12. J. Gassler *et al.*, Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science* **378**, 1305-1315 (2022).
13. L. Li *et al.*, Lineage regulators TFAP2C and NR5A2 function as bipotency activators in totipotent embryos. *Nat Struct Mol Biol*, (2024).
14. Y. Zhao *et al.*, Nr5a2 ensures inner cell mass formation in mouse blastocyst. *Cell Rep* **43**, 113840 (2024).
15. F. Lai *et al.*, NR5A2 connects zygotic genome activation to the first lineage segregation in totipotent embryos. *Cell Res* **33**, 952-966 (2023).
16. N. Festuccia, N. Owens, A. Chervova, A. Dubois, P. Navarro, The combined action of Esrrb and Nr5a2 is essential for murine naive pluripotency. *Development* **148**, (2021).
17. A. Chervova *et al.*, Mitotic bookmarking redundancy by nuclear receptors in pluripotent cells. *Nat Struct Mol Biol*, (2024).
18. J. Chen, J. Nathans, Estrogen-related receptor beta/NR3B2 controls epithelial cell fate and endolymph production by the stria vascularis. *Dev Cell* **13**, 325-337 (2007).
19. Y. K. Lee *et al.*, Liver receptor homolog-1 regulates bile acid homeostasis but is not essential for feedback regulation of bile acid synthesis. *Mol Endocrinol* **22**, 1345-1356 (2008).
20. W. N. de Vries *et al.*, Expression of Cre recombinase in mouse oocytes: a means to study maternal effect genes. *Genesis* **26**, 110-112 (2000).

21. Z. Xiong *et al.*, Ultrasensitive Ribo-seq reveals translational landscapes during mammalian oocyte-to-embryo transition and pre-implantation development. *Nat Cell Biol* **24**, 968-980 (2022).
22. Q. Deng, D. Ramskold, B. Reinius, R. Sandberg, Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193-196 (2014).
23. V. Hahaut *et al.*, Fast and highly sensitive full-length single-cell RNA sequencing using FLASH-seq. *Nat Biotechnol* **40**, 1447-1451 (2022).
24. B. Liu *et al.*, The landscape of RNA Pol II binding reveals a stepwise transition during ZGA. *Nature* **587**, 139-144 (2020).
25. M. Sakamoto *et al.*, Detection of newly synthesized RNA reveals transcriptional reprogramming during ZGA and a role of Obox3 in totipotency acquisition. *Cell Rep* **43**, 114118 (2024).
26. A. E. Peaston *et al.*, Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* **7**, 597-606 (2004).
27. V. Franke *et al.*, Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. *Genome Res* **27**, 1384-1394 (2017).
28. G. Falco *et al.*, Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Dev Biol* **307**, 539-550 (2007).
29. Q. Zhang *et al.*, Tcstv1 and Tcstv3 elongate telomeres of mouse ES cells. *Sci Rep* **6**, 19852 (2016).
30. M. A. El-Brolosy *et al.*, Genetic compensation triggered by mutant mRNA degradation. *Nature* **568**, 193-197 (2019).
31. A. Rossi *et al.*, Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524**, 230-233 (2015).
32. J. Luo *et al.*, Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. *Nature* **388**, 778-782 (1997).
33. C. Labelle-Dumais, M. Jacob-Wagner, J. F. Pare, L. Belanger, D. Dufort, Nuclear receptor NR5A2 is required for proper primitive streak morphogenesis. *Dev Dyn* **235**, 3359-3369 (2006).
34. J. F. Pare *et al.*, The fetoprotein transcription factor (FTF) gene is essential to embryogenesis and cholesterol homeostasis and is regulated by a DR4 element. *J Biol Chem* **279**, 21206-21216 (2004).
35. P. Gu *et al.*, Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. *Mol Cell Biol* **25**, 3492-3505 (2005).
36. D. Hameyer *et al.*, Toxicity of ligand-dependent Cre recombinases and generation of a conditional Cre deleter mouse allowing mosaic recombination in peripheral tissues. *Physiol Genomics* **31**, 32-41 (2007).
37. M. Maruyama, T. Ichisaka, M. Nakagawa, S. Yamanaka, Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. *J Biol Chem* **280**, 24371-24379 (2005).
38. N. Nishioka *et al.*, Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech Dev* **125**, 270-283 (2008).
39. J. M. Zhang, L. Zou, Alternative lengthening of telomeres: from molecular mechanisms to therapeutic outlooks. *Cell Biosci* **10**, 30 (2020).
40. L. Liu *et al.*, Telomere lengthening early in development. *Nat Cell Biol* **9**, 1436-1441 (2007).

41. P. Marzec *et al.*, Nuclear-receptor-mediated telomere insertion leads to genome instability in ALT cancers. *Cell* **160**, 913-927 (2015).
42. T. S. Macfarlan *et al.*, Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57-63 (2012).
43. S. Fujii *et al.*, Nr0b1 is a negative regulator of Zscan4c in mouse embryonic stem cells. *Sci Rep* **5**, 9146 (2015).
44. S. X. Ge, Exploratory bioinformatics investigation reveals importance of "junk" DNA in early embryo development. *BMC Genomics* **18**, 200 (2017).
45. W. Kobayashi *et al.*, Nucleosome-bound NR5A2 structure reveals pioneer factor mechanism by DNA minor groove anchor competition. *Nat Struct Mol Biol*, (2024).
46. M. R. Corces *et al.*, An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962 (2017).
47. M. Kinisu *et al.*, Klf5 establishes bi-potential cell fate by dual regulation of ICM and TE specification genes. *Cell Rep* **37**, 109982 (2021).
48. Y. Zhang *et al.*, Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nat Genet* **50**, 96-105 (2018).
49. E. R. Weikum, X. Liu, E. A. Ortlund, The nuclear receptor superfamily: A structural perspective. *Protein Sci* **27**, 1876-1892 (2018).
50. A. Penvose, J. L. Keenan, D. Bray, V. Ramlall, T. Siggers, Comprehensive study of nuclear receptor DNA binding provides a revised framework for understanding receptor specificity. *Nat Commun* **10**, 2514 (2019).
51. P. G. Hendrickson *et al.*, Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet* **49**, 925-934 (2017).
52. T. Hamatani *et al.*, Global gene expression profiling of preimplantation embryos. *Hum Cell* **19**, 98-117 (2006).
53. K. Abe, T. Schauer, M. E. Torres-Padilla, Distinct patterns of RNA polymerase II and transcriptional elongation characterize mammalian genome activation. *Cell Rep* **41**, 111865 (2022).
54. D. L. van den Berg *et al.*, An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6**, 369-381 (2010).
55. S. C. Lin, M. A. Wani, J. A. Whitsett, J. M. Wells, Klf5 regulates lineage formation in the pre-implantation mouse embryo. *Development* **137**, 3953-3963 (2010).
56. S. Menchero *et al.*, Transitions in cell potency during early mouse development are driven by Notch. *Elife* **8**, (2019).
57. A. Jedrusik, A. Cox, K. B. Wicher, D. M. Glover, M. Zernicka-Goetz, Maternal-zygotic knockout reveals a critical role of Cdx2 in the morula to blastocyst transition. *Dev Biol* **398**, 147-152 (2015).
58. S. Blij, T. Frum, A. Akyol, E. Fearon, A. Ralston, Maternal Cdx2 is dispensable for mouse development. *Development* **139**, 3969-3972 (2012).
59. J. Silva *et al.*, Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722-737 (2009).
60. I. Chambers *et al.*, Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230-1234 (2007).
61. K. Mitsui *et al.*, The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-642 (2003).

62. A. A. Avilion *et al.*, Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**, 126-140 (2003).
63. J. Nichols *et al.*, Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-391 (1998).
64. L. Li *et al.*, Multifaceted SOX2-chromatin interaction underpins pluripotency progression in early embryos. *Science* **382**, eadi5516 (2023).
65. D. Olivieri *et al.*, Cooperation between HDAC3 and DAX1 mediates lineage restriction of embryonic stem cells. *EMBO J* **40**, e106818 (2021).
66. K. Uranishi, T. Akagi, H. Koide, T. Yokota, Esrrb directly binds to Gata6 promoter and regulates its expression with Dax1 and Ncoa3. *Biochem Biophys Res Commun* **478**, 1720-1725 (2016).
67. S. Herchcovici Levy *et al.*, Esrrb is a cell-cycle-dependent associated factor balancing pluripotency and XEN differentiation. *Stem Cell Reports* **17**, 1334-1350 (2022).
68. W. B. Hamilton *et al.*, Dynamic lineage priming is driven via direct enhancer regulation by ERK. *Nature* **575**, 355-360 (2019).
69. N. Festuccia *et al.*, Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell* **11**, 477-490 (2012).
70. G. Martello *et al.*, Esrrb is a pivotal target of the gsk3/tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* **11**, 491-504 (2012).
71. A. K. Hadjantonakis, V. E. Papaioannou, Dynamic in vivo imaging and cell tracking using a histone fluorescent protein fusion in mice. *BMC Biotechnol* **4**, 33 (2004).
72. C. Kress, S. Vandormael-Pournin, P. Baldacci, M. Cohen-Tannoudji, C. Babinet, Nonpermissiveness for mouse embryonic stem (ES) cell derivation circumvented by a single backcross to 129/Sv strain: establishment of ES cell lines bearing the Omd conditional lethal mutation. *Mamm Genome* **9**, 998-1001 (1998).
73. Y. Lallemand, V. Luria, R. Haffner-Krausz, P. Lonai, Maternally expressed PGK-Cre transgene as a tool for early and uniform activation of the Cre site-specific recombinase. *Transgenic Res* **7**, 105-112 (1998).
74. P. Lemaire, N. Garrett, J. B. Gurdon, Expression cloning of Siamois, a Xenopus homeobox gene expressed in dorsal-vegetal cells of blastulae and able to induce a complete secondary axis. *Cell* **81**, 85-94 (1995).
75. S. Vandormael-Pournin, E. Frachon, S. Gobaa, M. Cohen-Tannoudji, Microfabricated Device for High-Resolution Imaging of Preimplantation Embryos. *Methods Mol Biol* **2214**, 11-30 (2021).
76. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).
77. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).
78. J. Bezanson, A. Edelman, S. Karpinski, V. Shah, Julia: A fresh approach to numerical computing. *https://arxiv.org/abs/1411.1607*, (2015).
79. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
80. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
81. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

82. J. Wu *et al.*, The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652-657 (2016).
83. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
84. J. Feng, T. Liu, B. Qin, Y. Zhang, X. S. Liu, Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**, 1728-1740 (2012).
85. H. M. Amemiya, A. Kundaje, A. P. Boyle, The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354 (2019).
86. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184-1191 (2009).
87. A. Mammana, bamsignals: Extract read count signals from bam files. R package version 1.20.0. (2020).
88. A. Mathelier *et al.*, JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* **42**, D142-147 (2014).
89. G. Tan, B. Lenhard, TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* **32**, 1555-1556 (2016).
90. S. Tempel, Using and understanding RepeatMasker. *Methods Mol Biol* **859**, 29-51 (2012).
91. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).
92. S. W. Criscione, Y. Zhang, W. Thompson, J. M. Sedivy, N. Neretti, Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics* **15**, 583 (2014).
93. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
94. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287 (2012).
95. H. Wickham, in *Use R!,*. (Springer International Publishing : Imprint: Springer,, Cham, 2016), pp. 1 online resource (XVI, 260 pages 232 illustrations, 140 illustrations in color.
96. J. Larsson, eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. R package version 6.1.0. (2020).
97. F. Ramirez, F. Dundar, S. Diehl, B. A. Gruning, T. Manke, deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191 (2014).

Acknowledgments: The authors acknowledge the Mouse Genetics Engineering Center, Biomics, Photonic BioImaging and Image Analysis Hub platforms of Institut Pasteur. We thank Diana Voicu and Michael Imbeault for the critical reading of the manuscript and for discussions on repetitive elements and their analyses. We also thank Margaret Buckingham for proofreading and commenting on the manuscript.

Funding: This study was supported by recurrent funding from the Institut Pasteur (PN), the CNRS (NF, MCT, PN), Revive Investissement d'Avenir ANR-10-LABX-73 (MCT and PN), by an European Research Council grant (ERC-CoG-2017 BIND) (PN), and an ANR JCJC grant (Dev-NucleR; ANR-22-CE13-0008) (NF).

Author contributions: NF performed gene expression, ChIP-seq, and most computational analyses. Embryo collection, culture, and processing was carried out by SVP. AC aligned and statistically analyzed FLASH-seq data. AG contributed preparatory embryo immunostainings. IG, RXC and FLV derived the ESRRB-TDT reporter mice. GCG quantified immunofluorescence data, and prepared ATAC-seq samples. NF conceived the project, with the contribution of MCT. NF, MCT and PN supervised the work and analyzed the data. NF and PN wrote the manuscript.**Competing interests:** The authors declare that they have no competing interests.

Data and materials availability: FLASH-seq samples are summarized in **Table S6**, which also indicates preparation day, embryonic stage, and genotype. Fastq files from the listed samples, generated by trimming and splitting UMI and non-UMI reads, are available in GEO (GSE223673). The ATAC-seq data generated in this study are also available in GEO (GSE267536). Published ATAC-seq data were obtained from the GEO series GSE66581 (*82*), and Cut&Run data from series GSE229740 (*15*). Processed RNA-seq and Ribo-seq data are publicly available (*21*, *22*), as are tables detailing gene expression changes in *Dux* and *Obox* KO embryos (*10*, *11*). All data generated for this study are available in the main text or the supplementary materials.

Supplementary Materials

Figs. S1 to S5

Tables S1 to S6

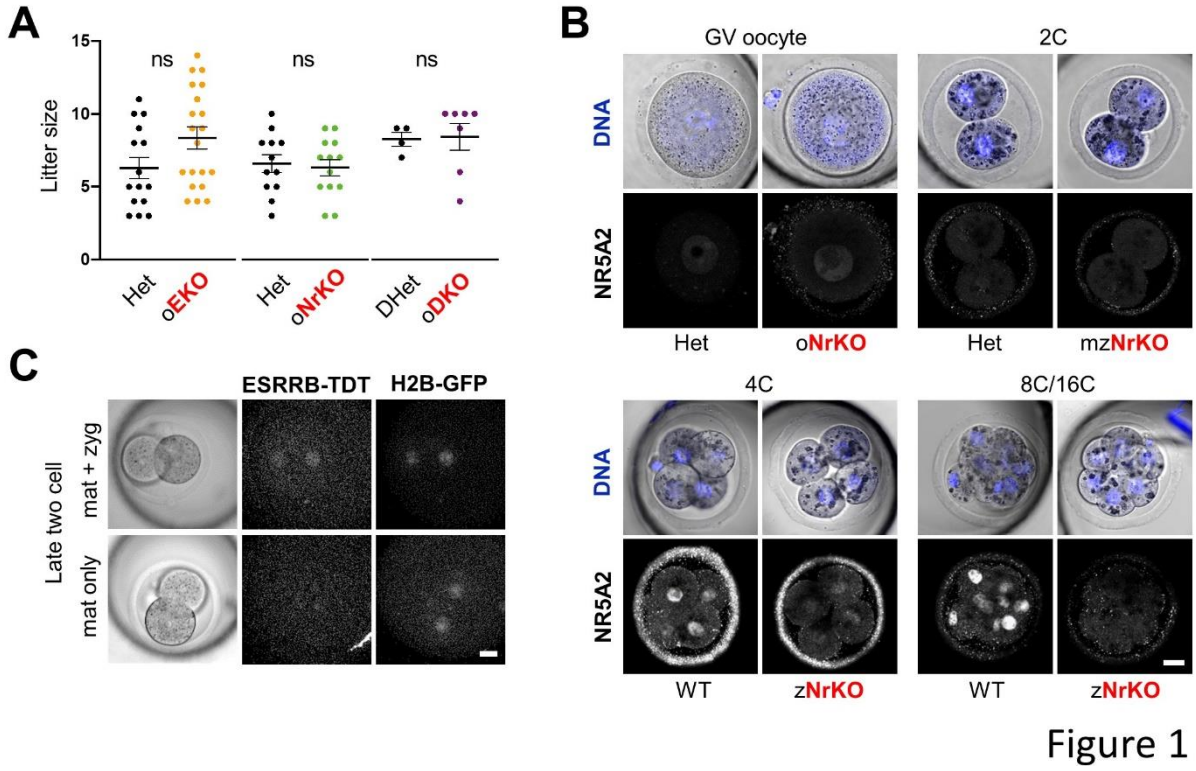


Fig. 1. Pre-implantation development does not require maternal NR5A2 and ESRRB.

A: Size of litters from crosses between females with oocytes devoid of NR5A2 and/or ESRRB (Essrb: oEKO, *Nr5a2:* oNrKO, both genes: oDKO) and wild-type males. Littermate females heterozygous for either or both genes were used as controls (see SM, materials and methods, and Table S1). Ns: Non-significant (Mann-Whitney P-value 0.06, 0.78, 0.38 for oEKO, oNrKO and oDKO, respectively). **B**: Immunofluorescence analysis for NR5A2 in control and oNrKO germinal vesicle (GV) oocytes, maternal and zygotic *Nr5a2* KO (mzNrKO) embryos at the 2C stage, or zygotic *Nr5a2* KO (zNrKO) embryos at the 4C, and 8C stages shows clear expression over background only after the 2C stage. Scale bars equal 20 μm in all panels. **C**: Expression of ESRRB-TDT fusion protein in late 2C embryos obtained from a cross between an $Esrrb^{TdT/+}$ knock-in female with an *H2b-Egfp* transgenic male. Upper panels show an embryo having inherited both the $Esrrb^{TdT}$ knock-in allele and any maternal mRNAs and protein store (mat +zyg), while the embryo shown in lower panels did not inherit the knock-in allele (mat only).

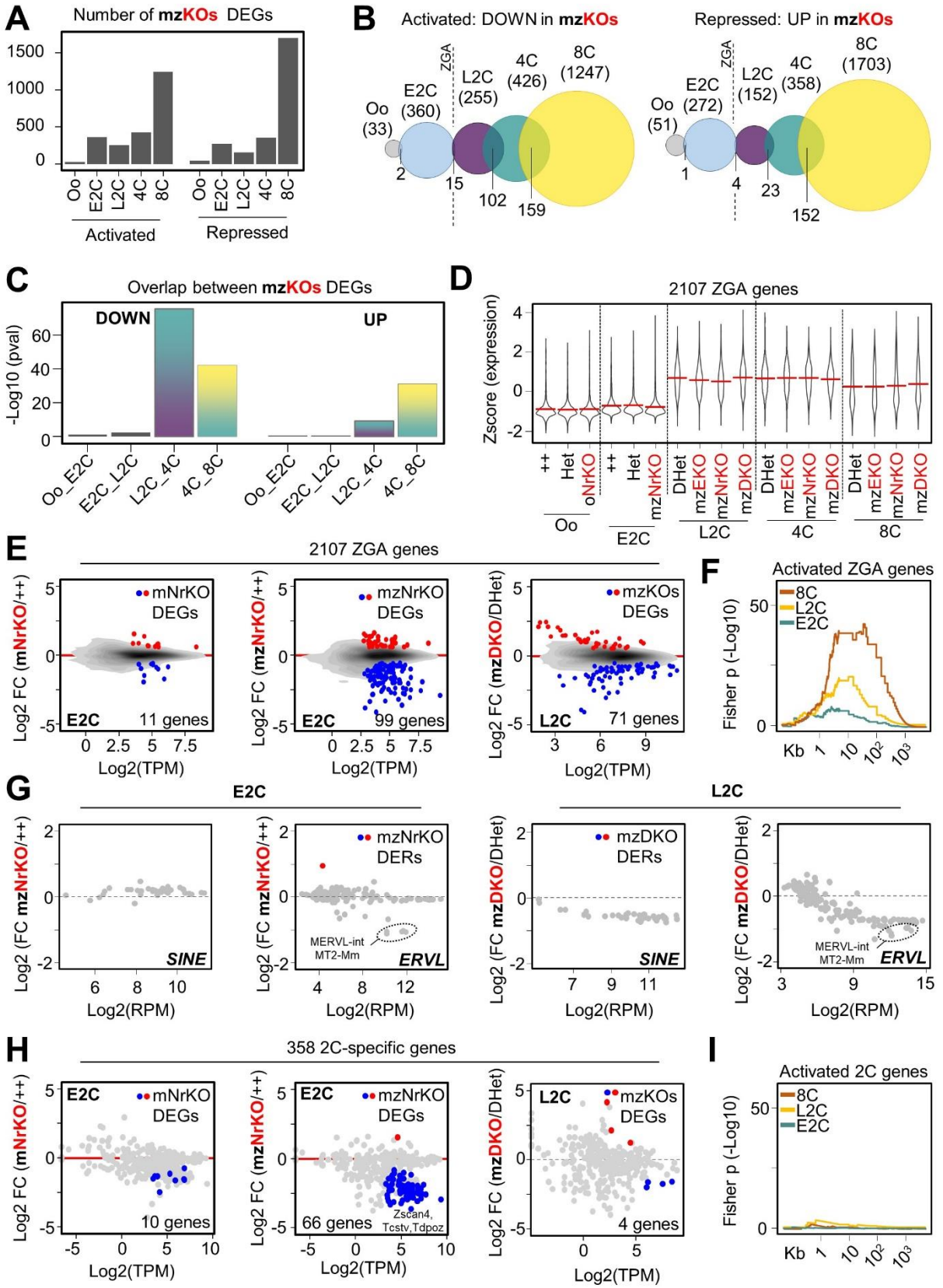

A
Number of mzKOs DEGs
1500
1000
500
0
Oo E2C L2C 4C 8C
Activated
Repressed
B
Activated: DOWN in mzKOs
Repressed: UP in mzKOs
ZGA
Oo (33)
E2C (360)
L2C (255)
4C (426)
8C (1247)
2
15
102
159
Oo (51)
E2C (272)
L2C (152)
4C (358)
8C (1703)
1
4
23
152
C
Overlap between mzKOs DEGs
DOWN
UP
-Log10 (pval)
Oo_E2C
E2C_L2C
L2C_4C
4C_8C
D
2107 ZGA genes
Zscore (expression)
++ Het oNrKO ++ Het mzNrKO DHet mzEKO mzNrKO mzDKO DHet mzEKO mzNrKO mzDKO DHet mzEKO mzNrKO mzDKO
Oo
E2C
L2C
4C
8C
E
2107 ZGA genes
mNrKO DEGs
E2C 11 genes
mzNrKO DEGs
E2C 99 genes
mzKOs DEGs
L2C 71 genes
Log2 FC (mNrKO/++)
Log2 FC (mzNrKO/++)
Log2 FC (mzDKO/DHet)
Log2(TPM)
F
Activated ZGA genes
8C
L2C
E2C
Fisher p (-Log10)
Kb
G
E2C
L2C
mzNrKO DERs
mzDKO DERs
SINE
ERVL
MERVL-int MT2-Mm
Log2 (FC mzNrKO/++)
Log2 (FC mzDKO/DHet)
Log2(RPM)
H
358 2C-specific genes
mNrKO DEGs
10 genes
mzNrKO DEGs
66 genes
Zscan4, Tcstv,Tdpoz
mzKOs DEGs
4 genes
I
Activated 2C genes

Figure 2

Fig. 2. Zygotic genome activation proceeds with minor defects in the complete absence of NR5A2.

A-B: Number (A) and overlap (B) between differentially expressed genes (DEGs) - Fold change <>1.5; False discovery rate (FDR) < 0.1; cutoff 10 transcripts per million (TPM) - in *Nr5a2* KO MII oocytes (Oo), after the maternal and zygotic KO of *Nr5a2* at the early 2C stage (E2C), or after the KO of either *Nr5a2*, *Esrrb*, or both genes at the late 2C (L2C), 4C and 8C stages (mzKOs). **C**: Significance of overlap (−log10 one-tailed Fisher's exact test P-value) between the genes differentially expressed (defined as in A-B) at consecutive developmental stages. **D**: Z-score of expression of 2107 genes activated during ZGA (as defined by (*12*) and expressed in our datasets), in *Nr5a2* KO oocytes (oNrKO) and embryos devoid of maternal and zygotic expression of *Nr5a2* (mzNrKO), *Esrrb* (mzEKO), or both genes (mzDKO), at the indicated developmental stages. Heterozygous or quasi-WT embryos are shown as controls (see methods). **E**: Fold change in expression of ZGA genes (see methods) in response to the maternal or the maternal and zygotic KO (mNrKO, mzNrKO) of *Nr5a2* at the E2C stage, or to the maternal and zygotic KO of *Esrrb* and *Nr5a2* (mzDKO) at the L2C stage. Red and blue identify significantly upregulated or downregulated genes in mzNrKO embryos for the E2C stage, or in either mzNrKO, mzEKO or mzDKO embryos (mzKOs) for the L2C stage (defined as in A-B; the numbers reported in the panels refer only to downregulated genes). The gray shading represents the density distribution of all genes, divided in 10 bins. **F**: Statistical enrichment (−log10 one-tailed Fisher's exact test P-values) of ZGA genes changing in response to the mzKO of *Nr5a2* at the early 2C stage, or of *Nr5a2*, *Esrrb* or both genes at the late 2C and 8C stages, at a given distance from regions bound by NR5A2 (Binding data from (*15*)). The test is repeated for discrete increments of the distance window (X axis), and P-values plotted on the Y axis. **G**: Fold change in expression of *SINE* or *ERVL* repeats in response to the mzKO of *Nr5a2* at the E2C stage, or of *Nr5a2* and *Esrrb* at the late 2C stage (mzDKO). Red and blue identify significantly upregulated or downregulated repeats (DERs) in mzNrKO (E2C) or mzDKO (L2C) embryos (FC<>1.5; False discovery rate (FDR) < 0.1). **H**: Fold change in expression of 2C-genes (see methods) plotted as in E, and under the same conditions. **I**: Statistical enrichment of 2C-genes at increasing distances from regions bound by NR5A2, represented as in F.

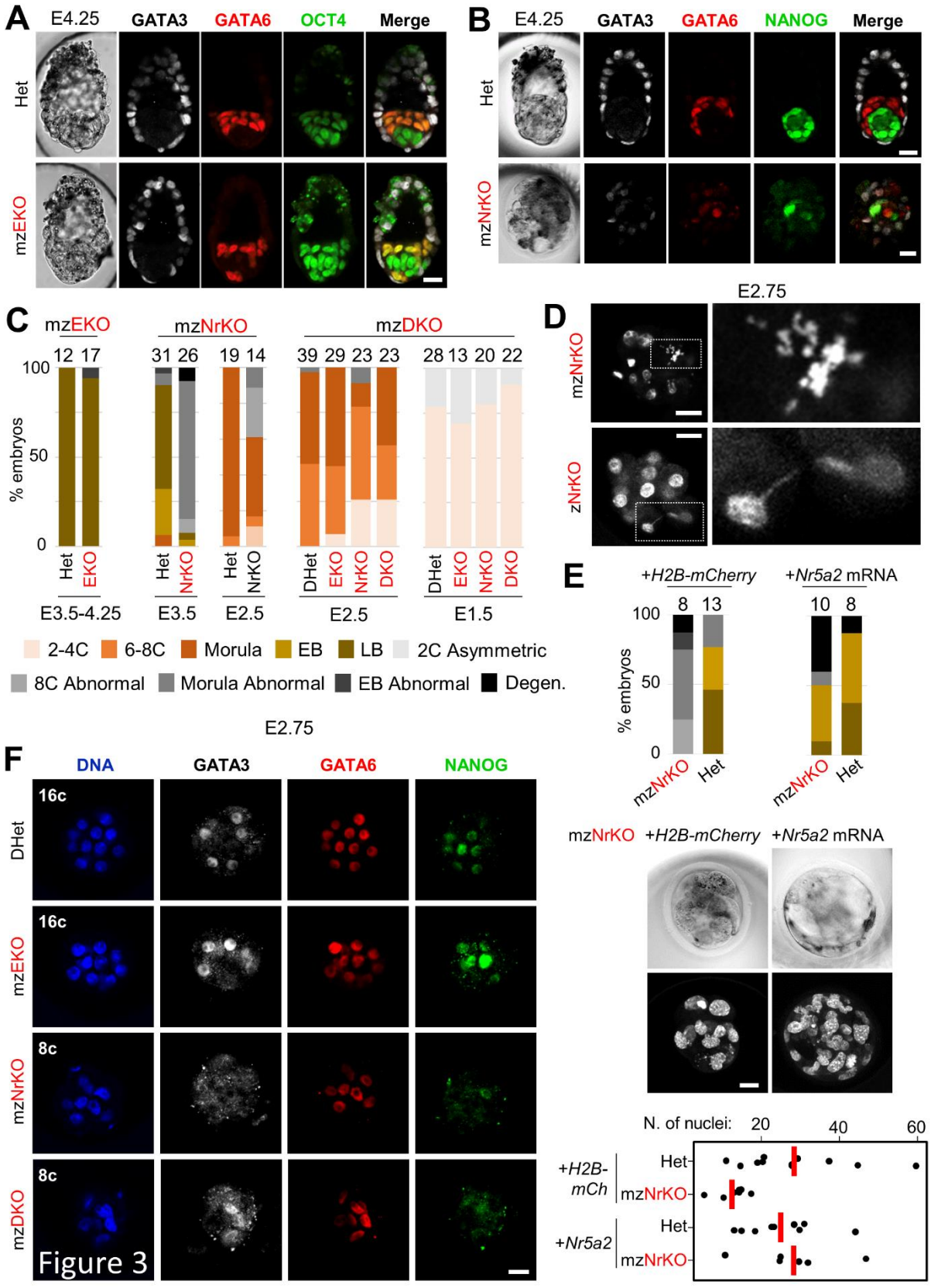

A
E4.25
GATA3
GATA6
OCT4
Merge
Het
mzEKO
B
E4.25
GATA3
GATA6
NANOG
Merge
Het
mzNrKO
C
mzEKO
mzNrKO
mzDKO
12 17
31 26
19 14
39 29 23 23
28 13 20 22
100
50
0
% embryos
Het
EKO
Het
NrKO
Het
NrKO
DHet
EKO
NrKO
DKO
DHet
EKO
NrKO
DKO
E3.5-4.25
E3.5
E2.5
E2.5
E1.5
2-4C
6-8C
Morula
EB
LB
2C Asymmetric
8C Abnormal
Morula Abnormal
EB Abnormal
Degen.
D
E2.75
mzNrKO
zNrKO
E
+H2B-mCherry
+Nr5a2 mRNA
8 13
10 8
100
50
0
% embryos
mzNrKO
Het
mzNrKO
Het
mzNrKO
+H2B-mCherry
+Nr5a2 mRNA
N. of nuclei:
20
40
60
+H2B-mCh
Het
mzNrKO
+Nr5a2
Het
mzNrKO
E2.75
F
DNA
GATA3
GATA6
NANOG
DHet
16c
mzEKO
16c
mzNrKO
8c
mzDKO
8c
Figure 3

Figure 3

Fig. 3. NR5A2 is required for the formation of a viable morula.

A: GATA3, GATA6 and OCT4 immunostaining of heterozygous, or maternal and zygotic *Esrrb* KO (mzEKO) embryos at E4.25. Scale bars equal 20 μm in all panels. **B**: GATA3, GATA6 and NANOG immunostaining of heterozygous, or maternal and zygotic *Nr5a2* KO (mzNrKO) embryos at E4.25. **C**: Rate of development observed for each genotype in E1.5 to E4.25 litters from crosses generating mzEKO, mzNrKO, or *Esrrb* and *Nr5a2* double mzKO (mzDKO) embryos. Degen: Degenerated. EB and LB: Early and late blastocyst. **D**: Hoechst DNA staining highlights mitotic defects - misaligned chromosomes at metaphase (up) and chromatin bridge (down) - in *Nr5a2* zygotic KO (zNrKO) or mzNrKO embryos collected at E2.75. **E**: Rate of development observed in *Nr5a2* heterozygous or mzKO embryos recovered at the 4C stage (54h post-hCG), injected with in-vitro transcribed mRNA for *Nr5a2* and *H2b-mCherry*, or *H2b-mCherry* alone, and cultured for an additional 48 hours in vitro. The legend for the barplots is common to panel C. Representative images for control and rescued mzNrKO embryos, and the number of nuclei per embryo observed in each genotype (excluding degenerate embryos where counting was impossible) are shown in the middle and bottom panels, respectively. **F**: GATA3, GATA6 and NANOG immunostaining of mzEKO, mzNrKO, mzDKO and heterozygous littermate embryos at E2.75.

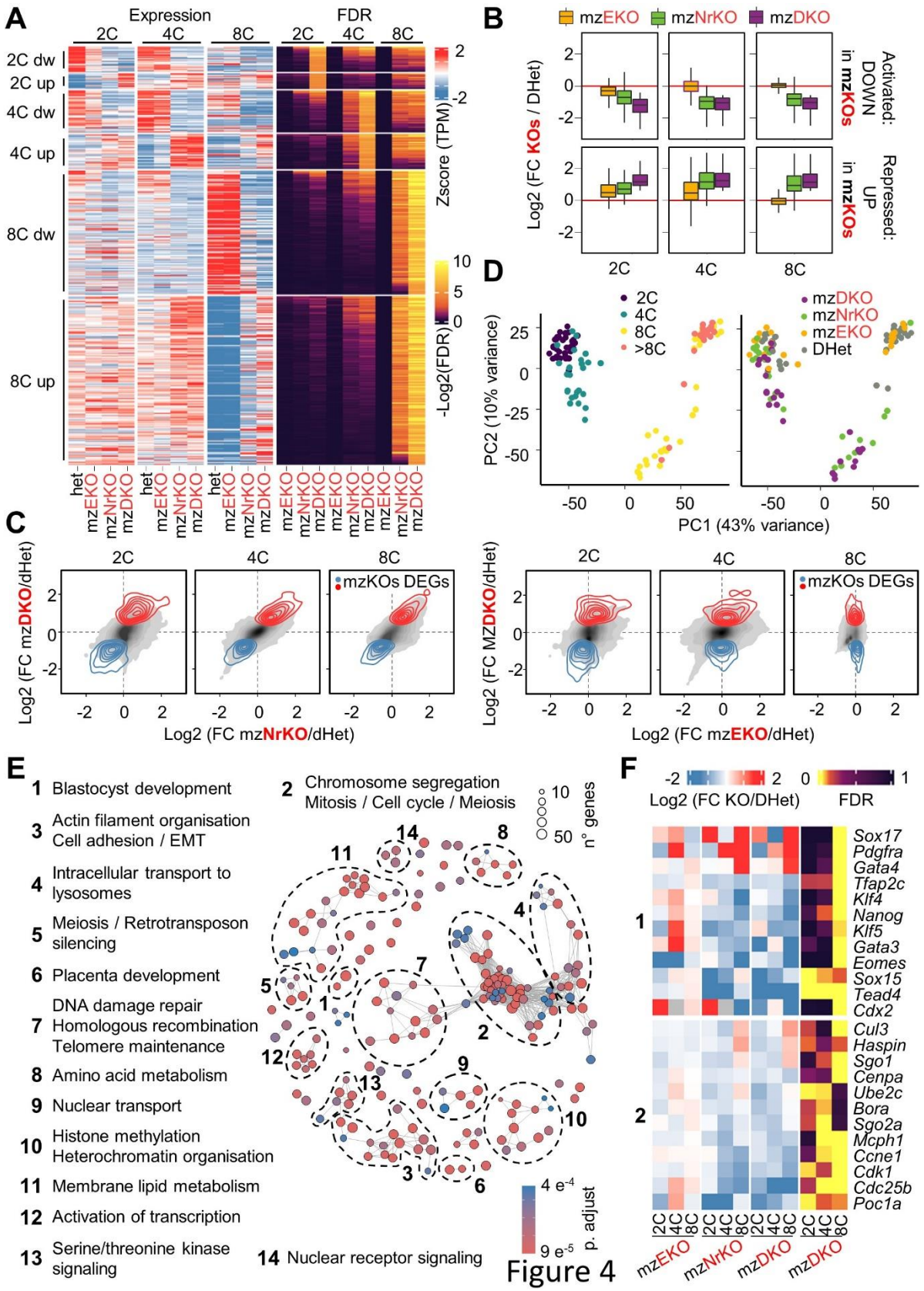

A
Expression
FDR
2C
4C
8C
2C dw
2C up
4C dw
4C up
8C dw
8C up
Zscore (TPM)
-Log2(FDR)
het
mzEKO
mzNrKO
mzDKO
B
mzEKO
mzNrKO
mzDKO
Log2 (FC KOs / DHet)
Activated: DOWN in mzKOs
Repressed: UP in mzKOs
C
Log2 (FC mzDKO/dHet)
Log2 (FC mzNrKO/dHet)
Log2 (FC MZDKO/dHet)
Log2 (FC mzEKO/dHet)
mzKOs DEGs
D
PC2 (10% variance)
PC1 (43% variance)
DHet
E
1 Blastocyst development
2 Chromosome segregation
Mitosis / Cell cycle / Meiosis
3 Actin filament organisation
Cell adhesion / EMT
4 Intracellular transport to lysosomes
5 Meiosis / Retrotransposon silencing
6 Placenta development
7 DNA damage repair
Homologous recombination
Telomere maintenance
8 Amino acid metabolism
9 Nuclear transport
10 Histone methylation
Heterochromatin organisation
11 Membrane lipid metabolism
12 Activation of transcription
13 Serine/threonine kinase signaling
14 Nuclear receptor signaling
n° genes
p. adjust
F
Log2 (FC KO/DHet)
FDR
Sox17
Pdgfra
Gata4
Tfap2c
Klf4
Nanog
Klf5
Gata3
Eomes
Sox15
Tead4
Cdx2
Cul3
Haspin
Sgo1
Cenpa
Ube2c
Bora
Sgo2a
Mcph1
Ccne1
Cdk1
Cdc25b
Poc1a

Figure 4

Fig. 4. Transcription is progressively dysregulated in *Nr5a2* KO embryos, including that of master lineage TFs, genes involved in DNA repair, telomere maintenance and cell division.

A: Heatmap showing relative expression (Z-scores) of genes significantly affected - Fold change >1.5; False discovery rate (FDR) < 0.1; cutoff 10 transcripts per million (TPM) - after the loss of maternal and zygotic expression of *Esrrb* (mzEKO), *Nr5a2* (mzNrKO) or both genes (mzDKO) at the 2, 4 or 8-cell stages. The FDR for differential expression is shown alongside. **B**: Fold change in expression of genes significantly affected (as in A) in response to the mzKO of either *Nr5a2*, *Esrrb* or both genes (mzKOs). **C**: Fold change in gene expression in response to the mzKO of *Nr5a2* or *Esrrb* versus the inactivation of both genes (mzDKO) at different stages of development. Red and blue distributions identify significantly upregulated or downregulated genes (as in A) in either mzNrKO, mzEKO or mzDKO (mzKOs) at each stage. **D**: Principal component analysis of the transcriptome from individual embryos color coded to highlight differences in genotypes (right) or developmental stage (left). **E**: Map showing the relation between gene ontology terms enriched among the gene that respond to the loss of *Nr5a2* and/or *Esrrb* in 2, 4 and 8C embryos. Terms sharing associated genes are linked by gray lines and cluster together. The number of responsive genes associated with each term is indicated by the circle size, the adjusted p-value for the enrichment of each term by the color scale. Numbering reflects ranking based on the p-value of the most enriched term in each cluster. **F**: Heatmap showing the fold change in expression of genes significantly affected (as in A) after the mzKO of *Esrrb*, *Nr5a2* or both genes at the 2, 4 or 8-cell stage. The FDR for the differential expression of each gene in mzDKO embryos is plotted alongside.

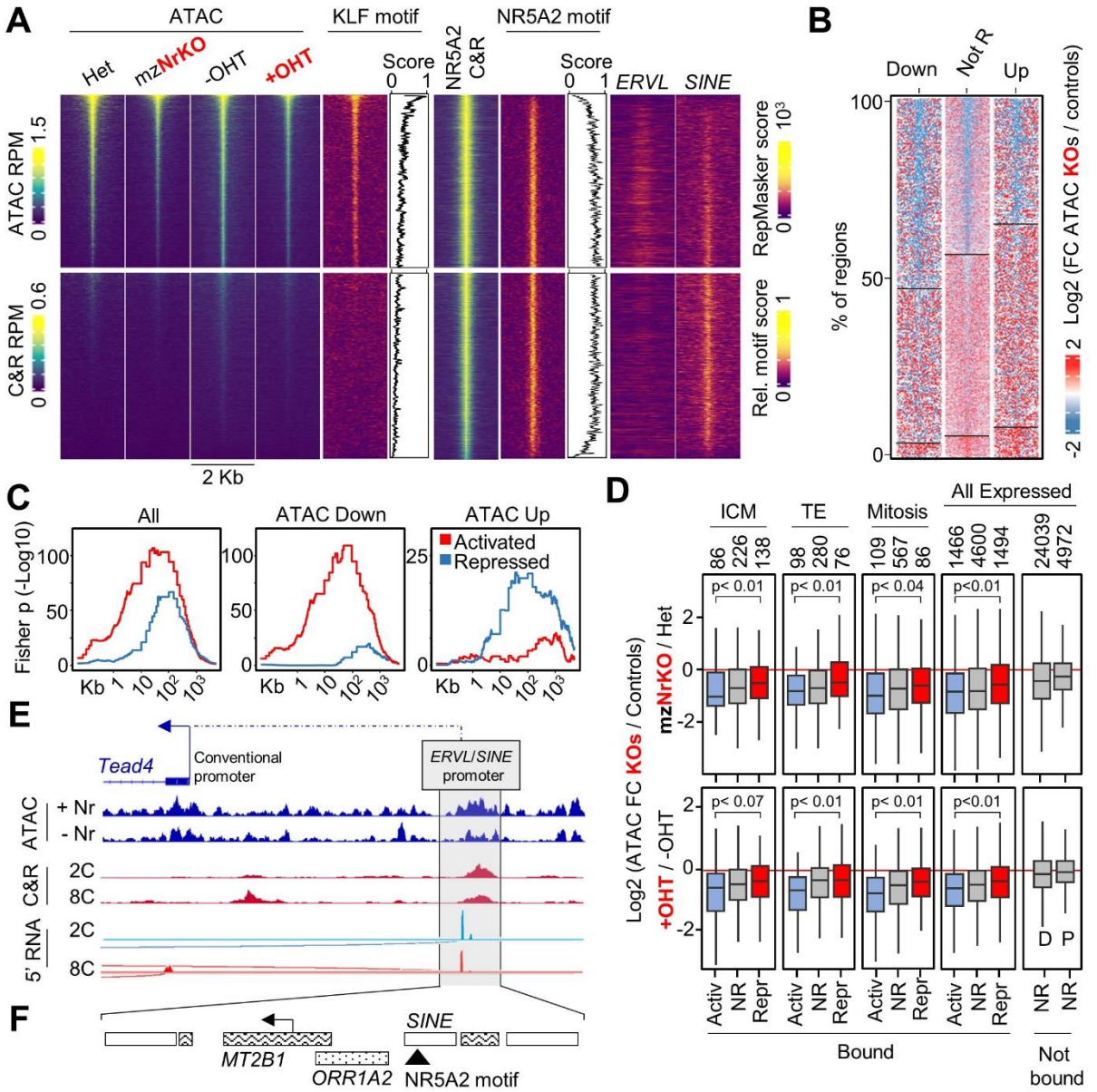


Figure 5

Fig. 5. To directly control expression of master lineage TFs, and genes involved in cell division and DNA repair, NR5A2 promotes chromatin accessibility at nearby regulatory regions.

A: Heatmaps showing chromatin accessibility in maternal and zygotic *Nr5a2* KO (mzNrKO) and control 8C embryos, or in *Rosa*$^{Cre\text{-}ERT2}$ embryos, with or without 4-hydroxytamoxifen (+/-OHT) treatment to induce recombination of the floxed *Nr5a2* alleles. The maps show regions bound by NR5A2 at the 8C stage, together with the distribution and relative cumulative scores of transcription factor motifs, and the location of repetitive elements. Regions are divided into accessible, or presenting no/low accessibility (see methods; NR5A2 binding data from (*15*)). For ATAC-seq and KLF motif profiles, accessible and bound regions are centered on ATAC-seq peak summits. All low accessibility regions, as well as those represented in the maps of NR5A2 binding, NR5A2 motifs, and SINE/ERVL profiles, are centered on CUT&RUN (C&R) peak summits. All maps extend for 1 kb on either side. **B**: Heatmaps showing the changes in chromatin accessibility in *Nr5a2* mutants 8C embryos (average between mzKOs and inducible

KOs) at NR5A2 bound and accessible regions in the proximity (+/- 50kb) of genes activated by, repressed by, or not responding (Not R) to NR5A2 and/or ESRRB at the 8C stage (see methods). **C**: Statistical enrichment (−log10 one-tailed Fisher's exact test P-values) of genes activated or repressed by NR5A2 and/or ESRRB at the 8C stage, over other expressed genes, at a given distance from accessible regions, and specifically from those showing increased or decreased accessibility in response to the loss of *Nr5a2*. The test is repeated for discrete increments of the distance window (X axis), and P-values plotted on the Y axis. **D**: Boxplots showing the change in accessibility in *Nr5a2* mutants at regions proximal (+/- 50kb) to genes regulated by, or irresponsive (Not R) to, NR5A2 and/or ESRRB at the 8C stage. The analysis is shown for all expressed genes, for TE or ICM specific genes (*48*), or for genes involved in the control of cell division (see methods). Changes at promoters (P) or distal regions (D) not bound by NR5A2 are shown as a control. The number of regions in each group is shown on top. One-tailed Kolmogorov-Smirnov test P-values (outliers excluded; see methods) were obtained comparing NR5A2 bound and accessible regions near activated versus repressed genes (shown over brackets), or versus distal regions not bound by NR5A2 located near non responsive genes ($p<0.01$ for all). **E**: Example of chromatin accessibility profiles (ATAC) in control and *Nr5a2* mutant 8C embryos, merging mz and inducible KOs, and NR5A2 binding profiles (C&R - data from (15)) near *Tead4*. The 5' anchored RNA-seq coverage (from UMI reads, see methods) and the inferred splice junctions observed in 2C or 8C control embryos (5' RNA) are also shown. **F**: Zoomed-in inset presenting the arrangement of repetitive elements at a non-canonical promoter active in early embryos, with ERVLs filled with a black wave pattern, ERVL-MaLRs with dots, *SINEs* in solid white, and a NR5A2 consensus binding motif shown as a triangle. Note that one transcript originates from an MT2B1 LTR juxtaposed to a NR5A2-bound *SINE* element (black arrow). Only this transcript is detected in 2C embryos, while at the 8C stage the conventional *Tead4* promoter is also active.

Supplementary Materials for

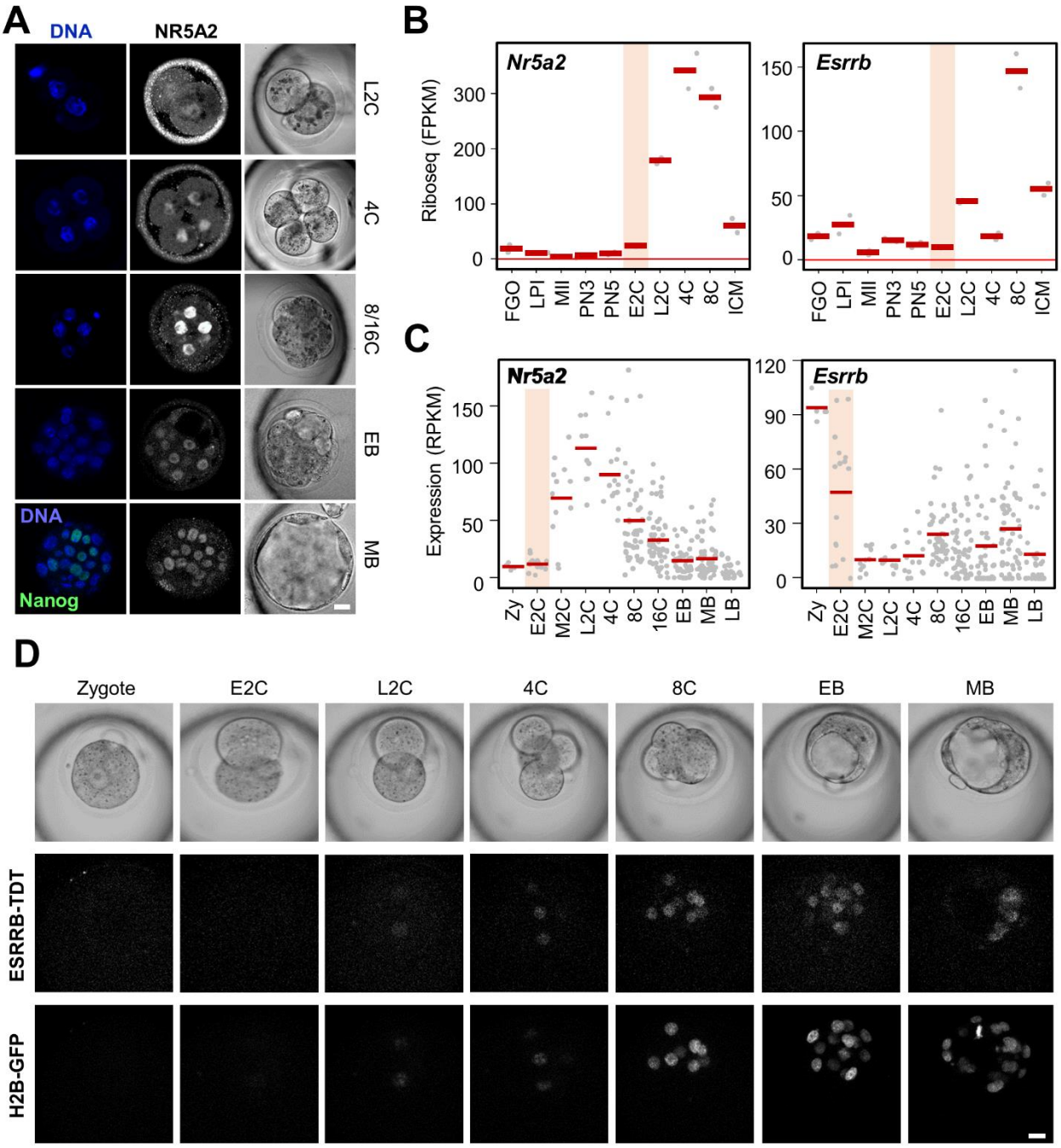
Nr5a2 is dispensable for ZGA but essential for morula development

Nicola Festuccia, Sandrine Vandormael-Pournin, Almira Chervova, Anna Geiselmann, Francina Langa-Vives, Rémi-Xavier Coux, Inma Gonzalez, Guillaume Giraud Collet, Michel Cohen-Tannoudji1, and Pablo Navarro
Corresponding author: nicola.festuccia@pasteur.fr

The PDF file includes:

Figs. S1 to S5
Table S1

Fig. S1.



No or marginal levels of NR5A2 and ESRRB maternal stores.

A: Immunofluoresce for NR5A2 in wild-type embryos at different stages of development (abbreviations defined below). **B-C**: Levels of mRNA undergoing translation (B), or total mRNA (C), in oocytes or embryos/single blastomeres at different developmental stages. The brown bars highlight the onset of major ZGA. Scale bars equal 20 μm in all panels. Data from (*21*). **D**: Snapshots of a time-lapse movie of an embryo with a maternal *Esrrb:TdTomato* knock-in allele and a paternal *H2b-Egfp* transgenic allele. Zy: zygote; FGO: fully grown

oocyte; LPI: late prometaphase I oocyte; MII: MII oocyte; PN3/5: zygote at pronuclear stage 3/5. E/M/L2C: early, mid, late 2-cell stage; E/M/LB: early, mid, late blastocyst. ICM: inner cell mass.

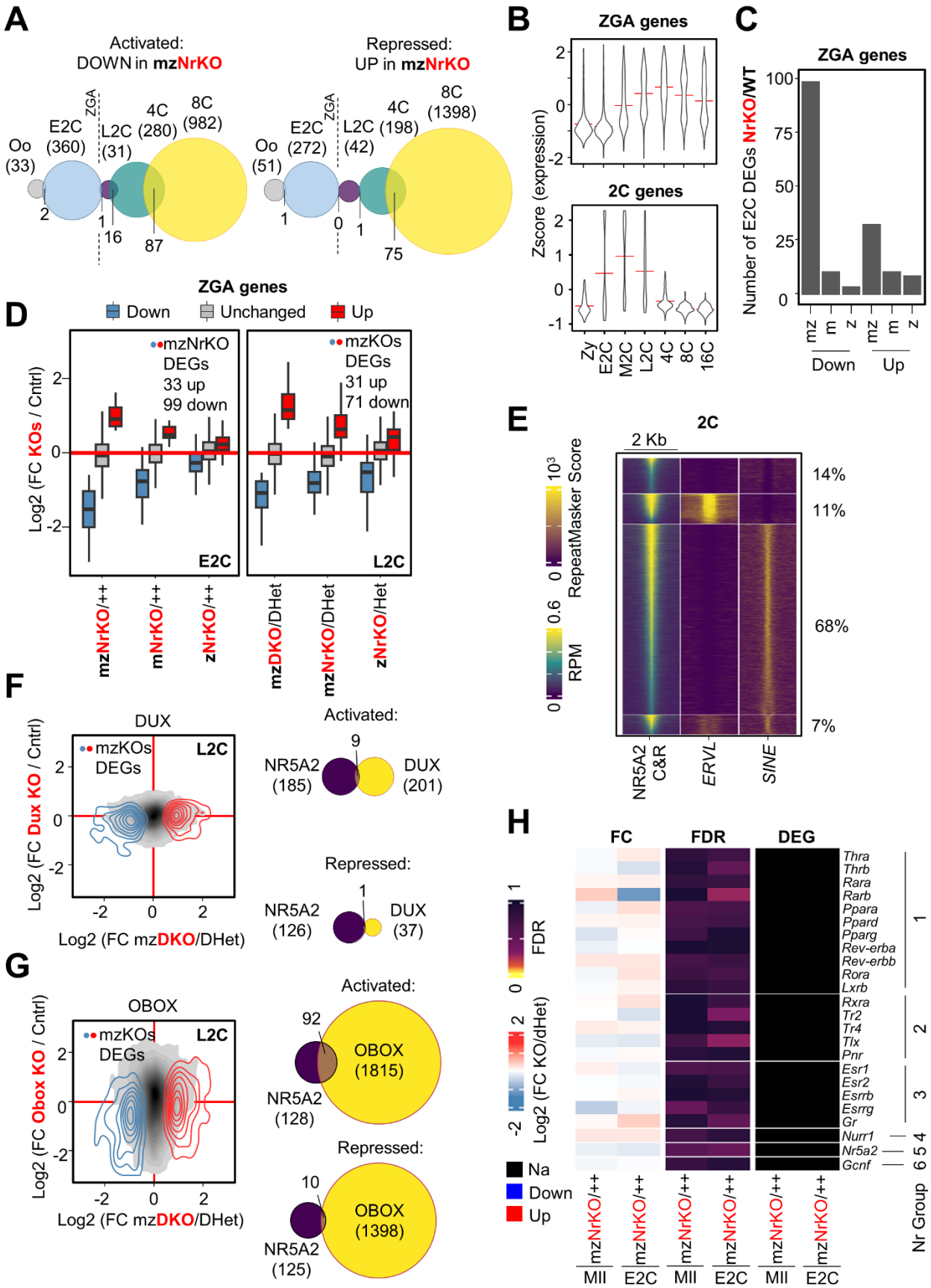
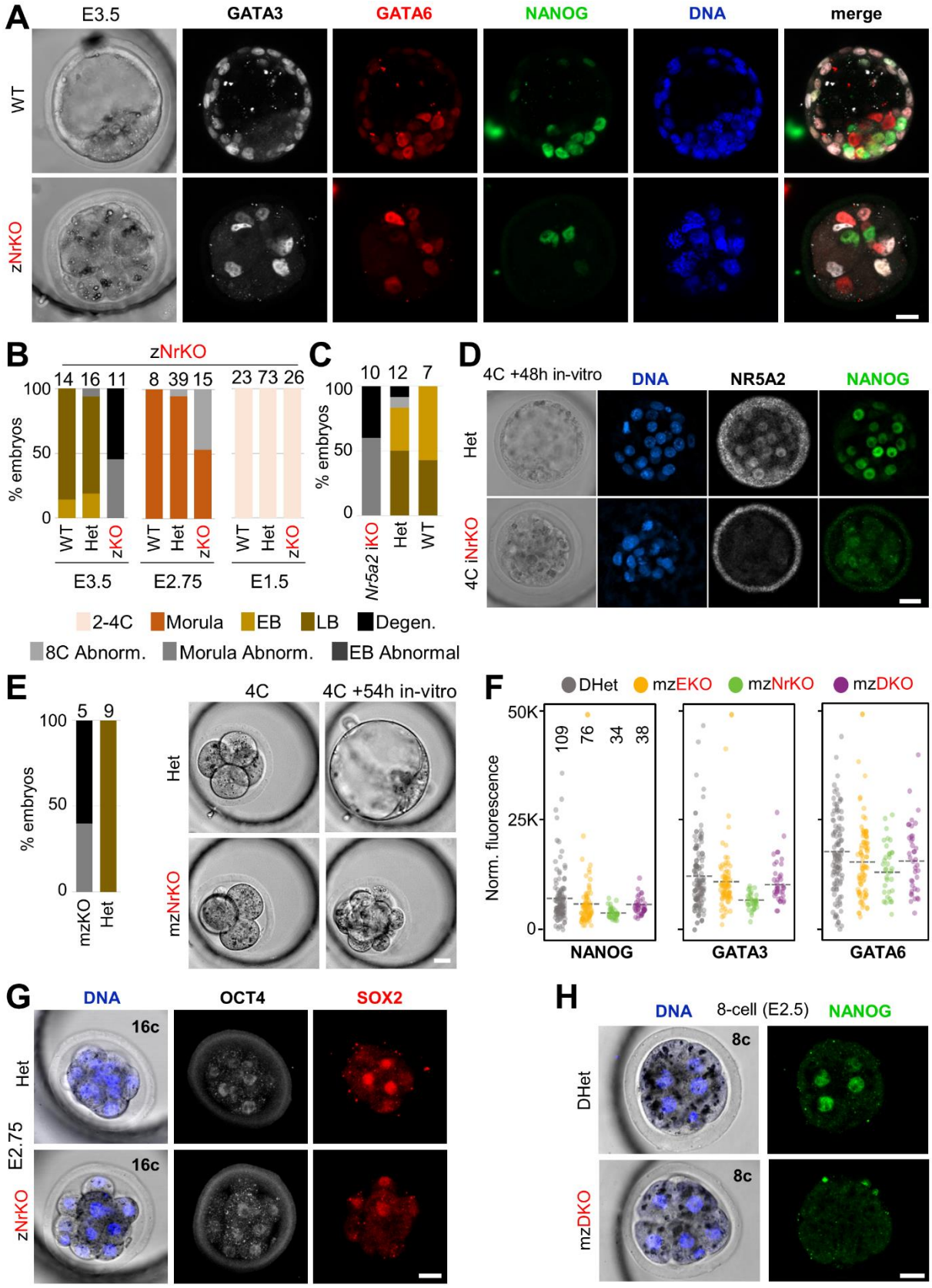

A
Activated:
DOWN in mzNrKO
Repressed:
UP in mzNrKO
ZGA
Oo (33)
E2C (360)
L2C (31)
4C (280)
8C (982)
2
1
16
87
Oo (51)
E2C (272)
L2C (42)
4C (198)
8C (1398)
1
0
1
75
B
ZGA genes
Zscore (expression)
2C genes
Zy
E2C
M2C
L2C
4C
8C
16C
C
ZGA genes
Number of E2C DEGs NrKO/WT
mz
m
z
Down
Up
D
ZGA genes
Down
Unchanged
Up
Log2 (FC KOs / Cntrl)
mzNrKO DEGs
33 up
99 down
E2C
mzKOs DEGs
31 up
71 down
L2C
mzNrKO/++
mNrKO/++
zNrKO/++
mzDKO/DHet
mzNrKO/DHet
zNrKO/Het
E
2C
2 Kb
RepeatMasker Score
RPM
14%
11%
68%
7%
NR5A2 C&R
ERVL
SINE
F
DUX
mzKOs DEGs
L2C
Log2 (FC Dux KO / Cntrl)
Log2 (FC mzDKO/DHet)
Activated:
NR5A2 (185)
9
DUX (201)
Repressed:
NR5A2 (126)
1
DUX (37)
G
OBOX
mzKOs DEGs
L2C
Log2 (FC Obox KO / Cntrl)
Log2 (FC mzDKO/DHet)
Activated:
92
NR5A2 (128)
OBOX (1815)
Repressed:
10
NR5A2 (125)
OBOX (1398)
H
FC
FDR
DEG
FDR
Log2 (FC KO/dHet)
Na
Down
Up
Thra
Thrb
Rara
Rarb
Ppara
Ppard
Pparg
Rev-erba
Rev-erbb
Rora
Lxrb
Rxra
Tr2
Tr4
Tlx
Pnr
Esr1
Esr2
Esrrb
Esrrg
Gr
Nurr1
Nr5a2
Gcnf
Nr Group
mzNrKO/++
MII
E2C

Fig. S2.

Zygotic genome activation proceeds with minor defects in the complete absence of NR5A2.

A: Overlap between differentially expressed genes (DEGs) [fold change (FC) <1.5 or >1.5; false discovery rate (FDR) < 0.1; cutoff 10 transcripts per million (TPM)] in *Nr5a2* KO MII oocytes (Oo), or after the maternal and zygotic KO of *Nr5a2* at the early 2C (E2C), late 2C (L2C), 4C and 8C stages (mzNrKO). **B:** Z-score of expression at successive developmental stages of 2107 genes activated during ZGA (top) [as defined in (*12*) and expressed in our datasets] or of 358 genes peaking at the 2C stage [bottom, see SM, materials and methods]. Expression data is from (*22*). Zy: zygote; E/M/L2C: early, mid, late 2C stages. **C**: Number of genes differentially expressed [defined as in A] in maternal (m), zygotic (z), or maternal and zygotic (mz) *Nr5a2* KO embryos at the early 2C stage. **D**: Boxplot showing the fold change in expression of ZGA genes in response to the zKO, mKO, or mzKO of *Nr5a2* at the E2C stage (left box), or the mzKO of *Nr5a2* and *Esrrb* (mzDKO), the mzKO of *Nr5a2*, or the zKO of *Nr5a2* at the L2C stage (right box). Red, blue and grey identify significantly upregulated, downregulated, or unresponsive genes (as in C) in *Nr5a2* mzKOs at the E2C stage, or either mzNrKOs, mzEKOs or mzDKOs (mzKOs) at the L2C stage. **E**: Heatmap showing NR5A2 occupancy and the distribution of *ERVL* and *SINE* elements, colored by their RepeatMasker alignment score, at NR5A2 bound regions in 2C embryos [binding data from (*15*)]. Regions are centered on NR5A2 CUT&RUN peak summits, and extend for 1 Kb on either side. The proportion of NR5A2 bound regions overlapping these classes of repeats is shown [SM, materials and methods]. **F-G**: Left: Comparison between the changes in gene expression elicited in response to the mzDKO of *Nr5a2* and *Esrrb* at the late 2C stage, or in response to the inactivation of *Dux* (*10*) (E), or *Obox* genes (*11*) (F). Red and blue distributions identify significantly upregulated or downregulated genes in either mzNrKOs, mzEKOs or mzDKOs (mzKOs), the gray shading represents the density distribution of all genes. Right: Venn diagrams comparing significantly upregulated or downregulated genes in the same conditions. **H**: Heatmap showing the fold change in expression of nuclear receptors in *Nr5a2* KO oocytes, or after the mzKO of *Nr5a2*, at the E2C stage, compared to quasi-WT controls (left). The FDRs for the differential expression of each gene, as well as the classification into activated, repressed and unresponsive genes [defined as in A] are shown in the middle and at right. Gene names and the respective nuclear receptor family group are also indicated.

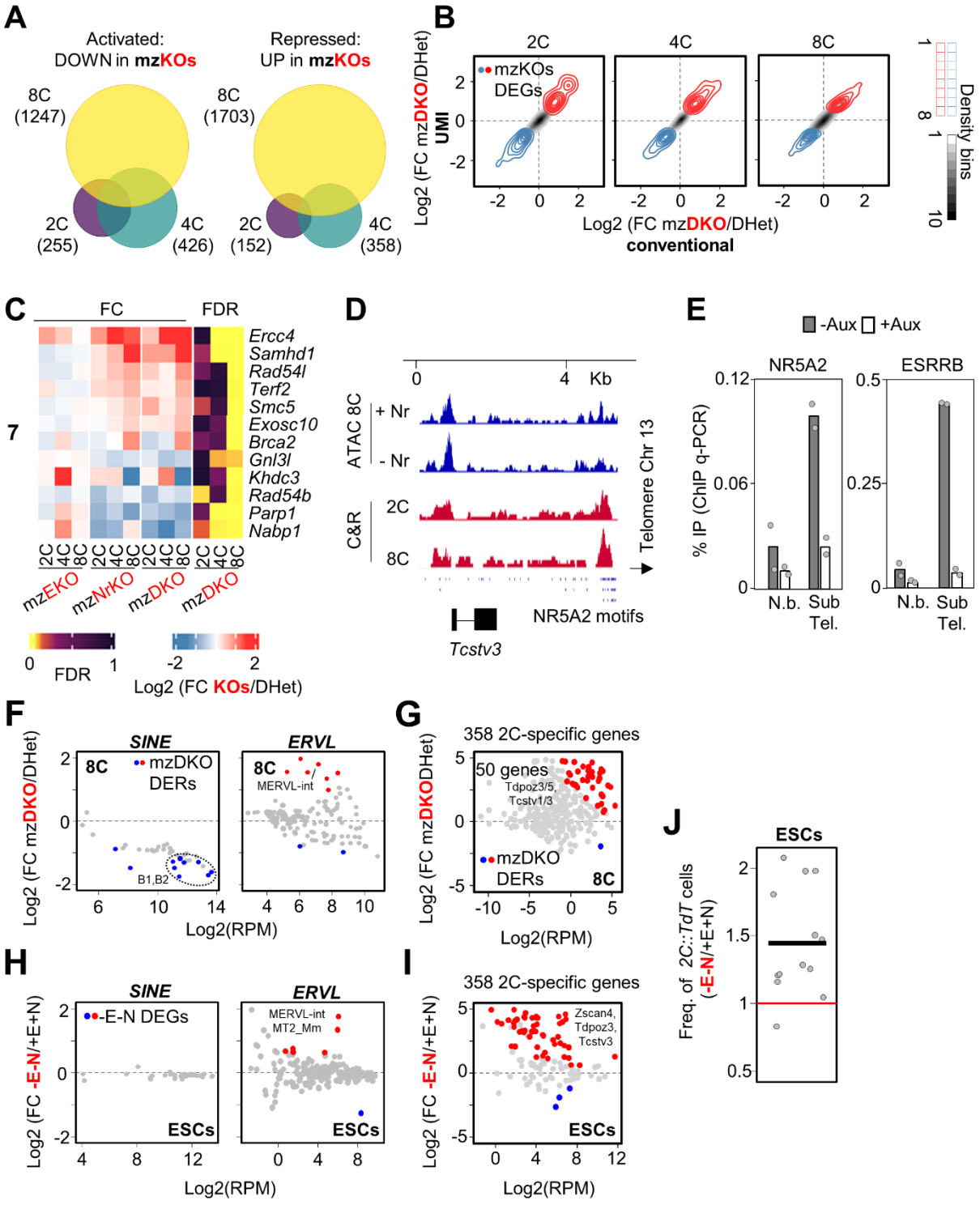
Fig. S3.


A
E3.5
GATA3
GATA6
NANOG
DNA
merge
WT
zNrKO
B
zNrKO
14 16 11
8 39 15
23 73 26
100
50
0
% embryos
WT
Het
zKO
E3.5
E2.75
E1.5
C
10 12 7
Nr5a2 iKO
Het
WT
D
4C +48h in-vitro
DNA
NR5A2
NANOG
Het
4C iNrKO
2-4C
Morula
EB
LB
Degen.
8C Abnorm.
Morula Abnorm.
EB Abnormal
E
5 9
mzKO
Het
4C
4C +54h in-vitro
mzNrKO
F
DHet
mzEKO
mzNrKO
mzDKO
50K
25K
0
Norm. fluorescence
109 76 34 38
NANOG
GATA3
GATA6
G
DNA
OCT4
SOX2
16c
E2.75
H
8-cell (E2.5)
8c
DHet
mzDKO

NR5A2 is required for the formation of a viable morula.

A: GATA3, GATA6 and NANOG immunostaining of *Nr5a2* zygotic KO (zNrKO) or control embryos at E3.5. Scale bars, 20 μm in all panels. **B**: Rate of development observed for each genotype at E1.5 to E3.5 in litters from crosses generating zNrKO embryos. Degen: Degenerated. EB and LB: Early and late blastocyst. **C**-**D**: Panel C shows the rate of development observed in wild-type, heterozygous, or inducible *Nr5a2* KO (iNrKO) embryos [see table S1 for genotypes], dissected at the 4C stage and cultured for 2 days in-vitro, in the presence of 4-hydroxytamoxifen for the first 24h. Legend is common to panel B. Panel D shows representative NR5A2, and NANOG immunostainings in iNrKO and control embryos. **E**: Rate of development observed in *Nr5a2* maternal and zygotic KO (mzNrKO) and control embryos dissected at the 4C stage and cultured for 54 hours in-vitro, along with representative images. Legend is common to panel B. **F**: Signal, normalized for the decay over the imaging Z axis [see SM, materials and methods], after immunofluorescence for NANOG, GATA3, and GATA6 expression, in the E2.75 heterozygous (mzDHet), *Esrrb* mzKO (mzEKO), *Nr5a2* mzKO (mzNrKO) or double mzKO (mzDKO) littermate embryos shown in Fig. 3F. The gray dashed lines show the mean of each distribution. The total number of nuclei analyzed for each genotype is shown in the left panel. Note that while most cells showed clear GATA6 staining, only few cells discernibly expressed NANOG and GATA3 at this stage, corresponding to the sparse bright events in the plots. **G**: OCT4 and SOX2 immunostaining of zNrKO or control littermate embryos at E2.75. **H**: NANOG immunostaining of mzDKO and control littermate embryos at the 8C stage.

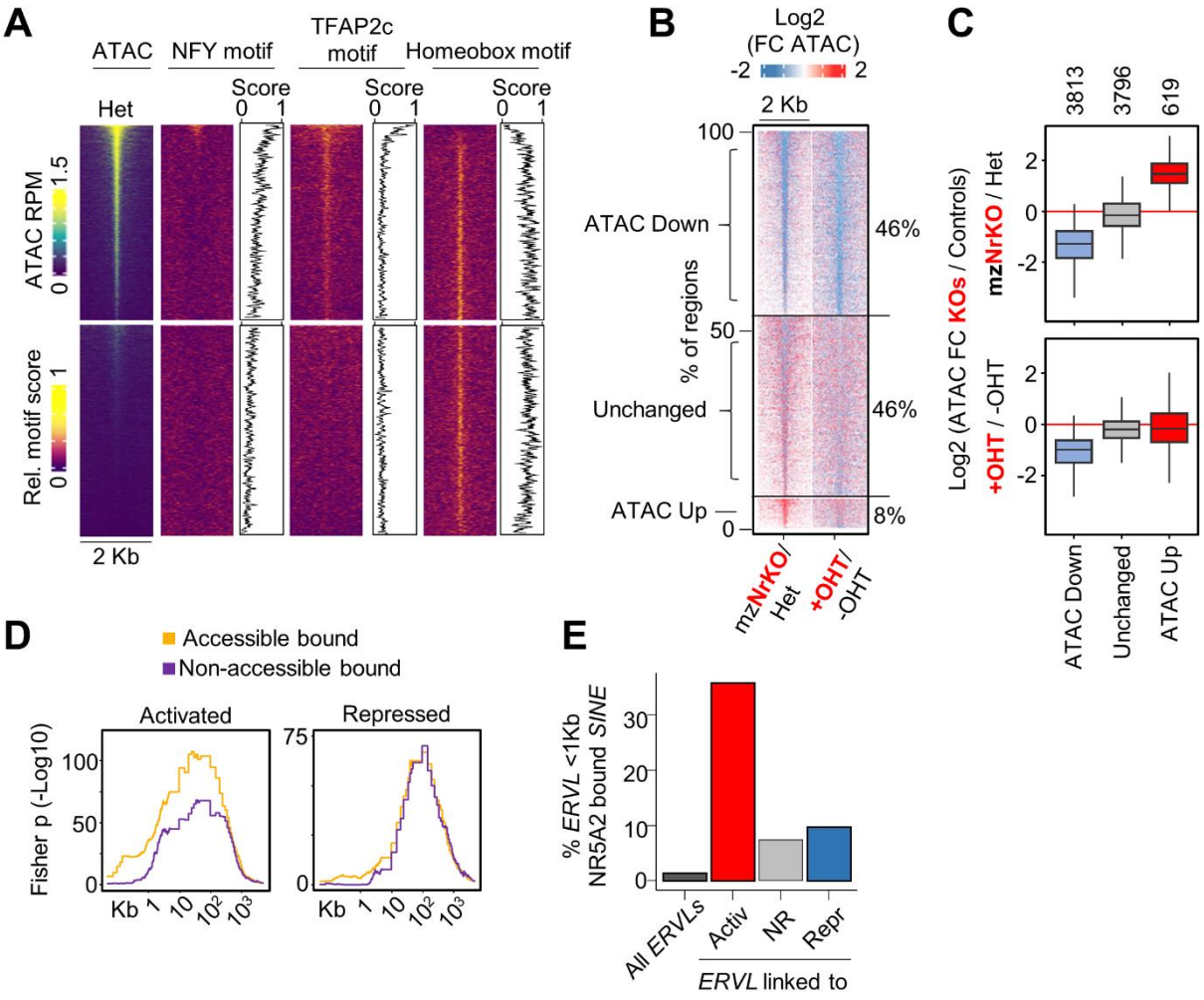
Fig. S4.


A
Activated:
DOWN in mzKOs
Repressed:
UP in mzKOs
8C
(1247)
2C
(255)
4C
(426)
8C
(1703)
2C
(152)
4C
(358)
B
2C
4C
8C
mzKOs
DEGs
Log2 (FC mzDKO/DHet)
UMI
Log2 (FC mzDKO/DHet)
conventional
Density bins
C
FC
FDR
Ercc4
Samhd1
Rad54l
Terf2
Smc5
Exosc10
Brca2
Gnl3l
Khdc3
Rad54b
Parp1
Nabp1
7
mzEKO
mzNrKO
mzDKO
mzDKO
FDR
Log2 (FC KOs/DHet)
D
0
4 Kb
ATAC 8C
+ Nr
- Nr
C&R
2C
8C
Telomere Chr 13
NR5A2 motifs
Tcstv3
E
-Aux +Aux
NR5A2
ESRRB
% IP (ChIP q-PCR)
N.b.
Sub Tel.
F
SINE
ERVL
8C
mzDKO DERs
MERVL-int
B1,B2
Log2 (FC mzDKO/DHet)
Log2(RPM)
G
358 2C-specific genes
50 genes
Tdpoz3/5,
Tcstv1/3
mzDKO DERs
8C
Log2 (FC mzDKODHet)
Log2(RPM)
H
SINE
ERVL
-E-N DEGs
MERVL-int
MT2_Mm
ESCs
Log2 (FC -E-N/+E+N)
Log2(RPM)
I
358 2C-specific genes
Zscan4,
Tdpoz3,
Tcstv3
ESCs
Log2 (FC -E-N/+E+N)
Log2(RPM)
J
ESCs
Freq. of 2C::TdT cells
(-E-N/+E+N)

Transcription is progressively dysregulated in *Nr5a2* KO embryos, including that of master lineage TFs, genes involved in DNA repair, telomere maintenance and cell division.

A: Number and overlap between the genes differentially expressed [fold change (FC) <1.5 or >1.5; false discovery rate (FDR) < 0.1; cutoff 10 transcripts per million (TPM)] in response to the maternal and zygotic loss of *Nr5a2*, *Esrrb* or both genes at the late 2C (L2C), 4C and 8C stage. **B**: Fold change in gene expression in response to the maternal and zygotic KO of *Nr5a2* and *Esrrb* (mzDKO) at the L2C, 4C and 8C stages, as judged quantifying all reads, or relying on UMI counts. Red and blue distributions identify significantly upregulated or downregulated genes [defined as in A] in response to the maternal and zygotic loss of *Nr5a2*, *Esrrb* or both genes (mzKOs) at each stage, based on conventional read analysis. The gray shading represents the density distribution of all genes, divided in 10 bins. **C**: Heatmap showing the fold change in expression of genes significantly affected (as in A) after the mzKO of *Esrrb*, *Nr5a2* or both genes at the L2C, 4C or 8C stage. The FDR for the differential expression of each gene in mzDKO embryos is plotted at right. **D**: NR5A2 binding (data from (*15*)) and chromatin accessibility in *Nr5a2* mutants (merging mz and inducible *Nr5a2* KOs, see Fig. 5) and controls, in the proximity of the subtelomeric region near *Tcstv3* on Chr 13. Note the presence of a cluster of degenerate NR5A2/ESRRB binding sites at the edge of the repetitive telomeric sequences (right). **E**: NR5A2 and ESRRB binding at the same sub-telomeric regions as in D (Sub Tel.), assessed by ChIP q-PCR in ES cells in which NR5A2 and ESRRB can be depleted by addition of Auxin (+Aux) (*17*). N.b (Not bound) is a control locus located 3' of the *Esrrb* gene. **F-H**: Fold change in expression of *SINE* or *ERVL* repeats in response to (F) the mzDKO of *Nr5a2/Esrrb* at the 8C stage, or (H) 2 days after *Nr5a2/Esrrb* loss-of-function in ESCs (-E-N) (*16*). Red and blue identify differentially expressed repeats (DERs) [FC <1.5 or >1.5; FDR < 0.1] in mzDKOs, or -E-N conditions. **G-I**: Fold change in expression of 2C-specific genes [see SM, materials and methods] in response to (G) the mzDKO of *Nr5a2* and *Esrrb* at the 8C stage, or (I) 2 days after *Nr5a2* and *Esrrb* loss-of-function in ESCs (-E-N) (*16*). Red and blue identify significantly upregulated or downregulated genes (defined as in A) in mzDKOs, or -E-N conditions. **J**: Fold change in the frequency of ESCs expressing a *TdTomato* reporter driven by a *MERVL* LTR [known as 2C-like cells (*42*)], 3 days after the depletion of ESRRB/NR5A2 using auxin inducible degrons (-E-N) (*17*).

Fig. S5.



To directly control expression of master lineage TFs, and genes involved in cell division and DNA repair, NR5A2 tends to favor chromatin accessibility at nearby regulatory regions.

A: Heatmaps showing chromatin accessibility at NR5A2 bound regions at the 8C stage, along with the distribution and relative cumulative scores of transcription factor motifs. The regions, identical to those presented in Fig 5A, are divided into accessible, or presenting no or low accessibility [see SM, materials and methods]. Regions are centered on ATAC-seq peak summits if accessible, or on Cut&Run peak summits if inaccessible, and extend for 1 Kb on either side. **B-C**: Heatmaps (B) and boxplots (C) of the fold change in chromatin accessibility at NR5A2 bound and accessible regions showing decreased, unchanged, or increased ATAC-seq signal in *Nr5a2* maternal and zygotic (mzNrKO) and control 8C embryos, or in *Rosa*$^{Cre\text{-}ERt2}$ embryos with or without 4-hydroxytamoxifen (+/- OHT) treatment to induce recombination of the floxed *Nr5a2* alleles [see SM, materials and methods]. The regions are identical to those presented in the top panels of Fig 5A. In B, regions were centered on ATAC-seq peak summits, and extended of 1 Kb on either side. **D**: Statistical enrichment [−log10 one-tailed Fisher's exact test P-values] of genes activated or repressed by NR5A2 and/or ESRRB at the 8C stage at a given distance from accessible and bound, or accessible but not bound regions (NR5A2 binding data from (*15*)). The test is repeated for discrete

increments of such distance (X axis), and P-values plotted on the Y axis. **E**: Proportion of *ERVL* linked to a gene activated (Activ), repressed (Repr), or not responding (NR) to NR5A2, that are juxstaposed (<1 Kb) to an NR5A2 bound *SINE* element. The same proportion is shown for all *ERVLs* as a control.

Table S1.

Genotypes of the live progeny and embryos recovered at different stages of development, after the maternal, zygotic, maternal and zygotic, or inducible deletion of *Nr5a2*/*Esrrb*.

A/ Transmission of deleted Nr5a2 and Esrrb alleles to the live progeny of Zp3Cre$^{tg/0}$; Esrrb$^{del/flox}$ or Nr5a2$^{del/flox}$ females occurs at the expected frequency

Genotype of live pups from *Zp3Cre$^{tg/0}$*; *Esrrb$^{del/flox}$* females crossed with wild-type CD1 males

	transmitted allele				
female genotype	*wt*	*flox*	*del*	*ND*	Total
Zp3Cre $^{tg/0}$; Esrrb$^{flox/+}$	23	0	32	3	58
Zp3Cre $^{tg/0}$; Esrrb$^{flox/del}$	0	0	72	0	72

n=4 females per genotype

Genotype of live pups from *Zp3Cre$^{tg/0}$*; *Nr5a2$^{del/flox}$* females crossed with wild-type CD1 males

	transmitted allele				
female genotype	*wt*	*flox*	*del*	*ND*	Total
Zp3Cre $^{tg/0}$; Nr5a2 $^{flox/+}$	10	0	17	0	27
Zp3Cre $^{tg/0;}$ Nr5a2 $^{flox/del}$	0	0	62	1	63

n= 3 control and n=4 mutant females respectively

B/ mzKO o zKO embryos are recovered at the expected frequencies

***Esrrb* mzKO**

Genotype of embryos from *Zp3Cre$^{tg/0}$*; *Esrrb$^{del/flox}$* females crossed with *Esrrb$^{del/+}$* males

	Number of embryos per genotype			
Stage	*Esrrb$^{del/+}$*	*Esrrb$^{del/del}$*	*ND*	Total
E3.0	*3*	*4*	*0*	7
E 3.5 + 20h	6	11	0	17
E 4.0	6	6	0	12
Total	15	21	0	36

n=1 female at E3.0; n=2 females at E3.5 + 20h and n=3 females at E4.0

***Nr5a2* mzKO**

Genotypes of embryos from *Zp3Cre$^{tg/0}$*; *Nr5a2$^{del/flox}$* females crossed with *Nr5a2$^{del/+}$* males

	Number of embryos per genotype			
Stage	*Nr5a2$^{del/+}$*	*Nr5a2$^{del/del}$*	*ND*	*Total*
E 2.5	19	14	3	36
E 3.5	15	13	1	29
E 4.0	16	14	0	30
Total	50	41	4	95

n=4 females at E2.5; n=2 females at E3.5 and n=3 females at E4.0

***Nr5a2* zKO**

Genotypes of embryos from *Nr5a2$^{del/+}$* females crossed with *Nr5a2$^{del/+}$* males

Stage	Number of embros per genotype				
	Nr5a2$^{+/+}$	*Nr5a2$^{del/+}$*	*Nr5a2$^{del/del}$*	*ND*	*Total*
E1.5	5	6	4	2	*17*
E 2,5	*8*	*39*	*15*	*8*	*70*
E 3.5	14	16	11	0	41
Total	27	61	30	10	128

n=1 female at E1.5; n=5 females at E2.5; n=3 females at E3.5

mzDKO

Genotypes of embryos from *Zp3Cre$^{tg/0}$*; *Esrrb$^{del/flox}$*; *Nr5a2$^{del/flox}$* females crossed with *Esrrb$^{del/+}$; Nr5a2$^{del/+}$* males

Stage	Number of embros per genotype					
	DKO *Esrrb$^{del/del}$ Nr5a2$^{del/del}$*	**EKO** *Esrrb$^{del/del}$ Nr5a2$^{del/+}$*	**NrKO** *Esrrb$^{del/+}$ Nr5a2$^{del/del}$*	**DHet** *Esrrb$^{del/+}$ Nr5a2$^{del/+}$*	*ND*	*Total*
E1.5	24	19	20	28	13	104
E 2,5	23	29	23	37	4	116
Total	47	48	43	65	17	220

n=7 female at E1.5; n=6 females at E2.5

C/ The inducible KO of Nr5a2 from the 4C stage recapitulates the phenotype of mzKOs.

Genotypes of inducible *Nr5a2* KO embryos recovered at E1.5 and cultured 2 days in vitro, in the presence or absence of 4-hydroxytamoxifen for the first 24h. *Nr5a2$^{flox/flox}$* (n=3) females were crossed with *Rosa$^{CreERt2\ KI/+}$; Nr5a2$^{del/+}$* males, so that not all embryos expressed *Cre-ERt2*.

Treatment	Number of embryos per genotype					
	Nr5a2$^{del/+}$	*Nr5a2$^{del/del}$*	*Nr5a2$^{flox/+}$*	*Nr5a2$^{flox/del}$*	*ND*	*Total*
wo OHT	0	0	7	7	5	19
with OHT	8	10	7	4	0	29