Zygotic genome activation by the totipotency pioneer factor Nr5a2

Johanna Gassler[1,2]†, Wataru Kobayashi[1]†, Imre Gáspár[1]†, Siwat Ruangroengkulrith[1]†, Adarsh Mohanan[1], Laura Gomez Hernandez[1], Pavel Kravchenko[1], Maximilian Kümmecke[2], Aleksandar Lalic[1], Nikita Rifel[1], Robert John Ashburn[1], Maciej Zaczek[2], Antoine Vallot[2], Laura Cuenca Rico[2], Sabrina Ladstätter[2], Kikuë Tachibana[1,2#]

[1] Department of Totipotency, Max Planck Institute of Biochemistry, Martinsried/Munich, Germany.

[2] Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Vienna Biocenter, Vienna, Austria.

†These authors contributed equally to this work

[#]Corresponding author. Email: tachibana@biochem.mpg.de

Abstract: Life begins with a switch in genetic control from the maternal to the embryonic genome during zygotic genome activation (ZGA). Despite its importance, the essential regulators of ZGA remain largely unknown in mammals. Based on *de novo* motif searches, we identified the orphan nuclear receptor Nr5a2 as a key activator of major ZGA in mouse 2-cell embryos. Nr5a2 is required for progression beyond the 2-cell stage. It binds to its motif within *SINE B1/Alu* retrotransposable elements found in *cis*-regulatory regions of ZGA genes. Chemical inhibition suggests that 72% of ZGA genes are regulated by Nr5a2 and potentially other orphan nuclear receptors. Nr5a2 promotes chromatin accessibility during ZGA and binds nucleosomal DNA *in vitro*. We conclude that Nr5a2 is an essential pioneer factor that regulates ZGA.

One-Sentence Summary: Nr5a2 is an essential pioneer factor that activates zygotic gene expression in mouse embryos.

The discovery that oocyte cytoplasm can reprogram somatic nuclei to totipotency has given rise to the idea that maternally provided factors reprogram and trigger ZGA in the embryo (*1*). The

transcriptional "awakening" of the embryonic genome in mice occurs in at least two waves, minor ZGA in the zygote and major ZGA (hereafter referred to as ZGA) in the 2-cell embryo (Fig. S1A) (*2, 3*). Zygotic transcription is required for progression beyond the 2-cell stage (*4*).

The transcription factors that initiate ZGA appear to be poorly conserved between species. During *Drosophila* embryogenesis, transcription factors including the diptera-specific Zelda, GAGA factor (GAF) and chromatin-linked adaptor for male-specific lethal protein (CLAMP) are essential for ZGA (*5-7*). In zebrafish and frogs, pluripotency factors belonging to the POU and Sox families as well as Nanog are required for ZGA (*8-10*). In contrast, the pluripotency factor Oct4 is not essential for murine ZGA (*11*). Nfya, Yap1, Dux, Rarg, Dppa2 and Dppa4 have been linked to mammalian ZGA (*12-17*) but are not required for progression beyond the 2-cell embryo in knockdown or genetic knockout models (*12, 18-22*). The essential transcription factors that activate mammalian zygotic genomes therefore remain largely unknown.

Transcription factors bind to regulatory elements in the genome to control gene expression and cell fate. Genomic DNA is wrapped around a histone octamer in the nucleosome, which is a barrier to transcription factor occupancy. Pioneer transcription factors (hereafter referred to as pioneer factors) are a class of transcription factors with the capacity to bind their (partial) motif on nucleosomal DNA *in vitro* and to be recruited to closed chromatin *in vivo*, eliciting local chromatin opening by diverse mechanisms (*23*). Pioneer factors cooperatively open the fruit fly and zebrafish genomes during ZGA (*6-9, 24, 25*). Whether pioneer activities initiate mammalian ZGA is not

known. Here, we identified the orphan nuclear receptor Nr5a2 as a pioneer factor that activates genome-wide gene expression in mouse embryos.

Nr5a2 is required for early embryonic development

We hypothesized that motifs of transcription factors regulating ZGA are enriched in the *cis*-regulatory regions of ZGA genes. To define ZGA genes, we compared transcription profiles of two mouse strains during the oocyte-to-embryo transition. A total of 2508 "extended" ZGA genes were upregulated >4-fold in 2-cell embryos of the two strains. The 985 genes common to both were classified as "core" ZGA genes (Fig. S1). Using *de novo* motif searches, we found a consensus sequence, comprising six motifs, that is enriched within 8 kb upstream of ~70% of core and extended ZGA genes (compared to 31% of non-ZGA genes) (Fig.1A and S2A-B). The sequence has 90% similarity to the *Short Interspersed Nuclear Element* (*SINE*) *B1* family of retrotransposons, which are related to the human *Alu* family (Fig 1A). A CA version of a variable pyrimidine/purine stretch (YR5) within the sequence is correlated with higher chromatin accessibility and histone acetylation during major ZGA (Fig. S2C and S2E), implying a functional relevance. YR5 is located in motif 1, which shows a high frequency and mean occurrence upstream of ZGA genes (Fig. S2D). The CA version of motif 1 contains the consensus sequences for orphan nuclear receptors Nr5a2 (T**CA**AGGCCA, hereafter Nr5a2 motif) and Esrrb (T**CA**AGGTCA, hereafter Esrrb motif) and nuclear hormone receptor retinoic acid receptor gamma (Rarg) (AGGT**CA**AGGTCA) (*26, 27*). Given that Rarg is not essential for ZGA since *Rarg*$^{-/-}$ females are fertile (*18*), we focused on the orphan nuclear receptors. *Nr5a2*$^{-/-}$ and *Esrrb*$^{-/-}$ embryos produced from heterozygote intercrosses die around implantation (*28, 29*). Since at least transcripts for both transcription factors are maternally provided (see below), it is conceivable that maternal proteins

enable progression beyond ZGA in homozygous knockout embryos. It is therefore unknown whether these transcription factors have functions during ZGA.

To assess whether Nr5a2 and Esrrb are ZGA regulators, we examined their expression during the oocyte-to-embryo transition. Both are maternally deposited as transcripts in metaphase II eggs and zygotes, and Nr5a2 is strongly expressed in 2-cell embryos (Fig. 1B and 1C). Nr5a2 protein is present in oocytes and enriched in the nuclei of zygotes and 2-cell embryos (Fig.1D). Esrrb protein becomes detectable in 2-cell embryos (Fig. 1E). Since Nr5a2 protein is present prior to ZGA, we focused on this transcription factor as a candidate regulator of zygotic gene expression.

To determine if Nr5a2 has an early embryonic function that would be consistent with regulating ZGA, we tested whether Nr5a2 perturbation affects development (Fig. 1F). Nr5a2 can be inhibited by the chemical compound SR1848 (*30, 31*). To evaluate its specificity, we expressed orphan nuclear receptor family members and performed a luciferase reporter assay in HEK293T cells. SR1848 inhibited Nr5a2, as expected, and also Nr5a1 and Nr2c2, but the latter two were undetectable as transcripts prior to ZGA (Fig. S3); Nr5a1 remained largely undetectable throughout pre-implantation development and Nr2c2 is transcribed during ZGA. SR1848 did not

inhibit Esrrb and Esrrg, which were both present prior to ZGA (Fig. S3). These results suggest that the main target of SR1848 in early embryos is Nr5a2.

Embryos that were constitutively cultured with SR1848 failed to form blastocysts and fragmented/died at 108 hours post-fertilization (hpf) (Fig. 1G and 1H, scheme (i)). This is consistent with a function of Nr5a2 in maintaining naïve pluripotency in mouse embryonic stem (ES) cells (*26*) and suggests that it is also required for pluripotency establishment *in vivo*. Embryos treated with SR1848 from the 2-cell stage (36 hpf) onwards had a less severe phenotype than those treated from 6 hpf (Fig. 1G and 1H, scheme (ii)), suggesting that Nr5a2 activity is required not only after but also before the 2-cell stage. Indeed, transient inhibition of Nr5a2 from 6-36 hpf caused either an immediate 2-cell arrest or a >24 h developmental delay (Fig. 1G and 1H, scheme (iii)). Overall, these findings suggest that Nr5a2 plays multiple roles in development, including a hitherto unknown function between fertilization and the 2-cell stage, when ZGA occurs.

Nr5a2 and Esrrb contribute to ZGA

To monitor ZGA directly, we visualized nascent ZGA transcripts by single-molecule fluorescence *in situ* hybridization (ZGA-FISH) (Fig. S4A). Two sets of FISH probes were designed for robustness and each detected ~80 nascent ZGA transcripts that were quantified by single-molecule FISH analysis algorithms (*32-34*). The lower threshold was determined by triptolide treatment to degrade RNA polymerase II in embryos. The total copy number of nuclear transcripts was reduced

by ~62% in triptolide-treated 2-cell embryos compared to DMSO-treated controls (Fig. S4B and S4C).

Treatment of zygotes with SR1848 resulted in a dose-dependent reduction of nascent ZGA transcripts (~20% and ~40% for 5 µM and 10 µM SR1848, respectively) (Fig. S5A and S5B), suggesting that Nr5a2 regulates ZGA. To determine whether this effect could be rescued by overexpression of transcription factors, SR1848-treated embryos were microinjected with mRNA encoding orphan nuclear receptors (Fig. S5C). Esrrb-GFP was ineffective in rescuing ZGA (Fig. S5D). In contrast, Nr5a2-GFP, Nr5a1-GFP and Nr2c2-GFP rescued ZGA to nearly control levels (Fig 2A, S5E and S5F). This shows that the inhibitor is not a generic suppressor of transcription and supports the notion that SR1848 inhibits ZGA by inactivating Nr5a2 (because the other two are not detectable in early embryos). Moreover, Nr5a2-GFP but not Esrrb-GFP overexpression in otherwise unperturbed zygotes was sufficient to significantly increase transcription during ZGA in 2-cell embryos (Fig. 2B). These results suggest that multiple orphan nuclear receptor family members have the ability to activate ZGA genes, presumably by recognizing the same motif. However, since only Nr5a2 can be detected in early embryos, these results suggest that Nr5a2 is required for ZGA.

To test directly whether Nr5a2 is involved in ZGA, we used Trim-Away (*35, 36*) to target endogenous Nr5a2 for degradation. *Trim21-mCherry* mRNA and non-specific IgG or two different Nr5a2 antibodies were microinjected into zygotes and resulting 2-cell embryos were analyzed by ZGA FISH. Nr5a2 targeting resulted in ~43% reduction of ZGA transcripts, suggesting that Nr5a2

is required for ZGA (Fig. 2C; S5G, H). In an orthogonal approach, we performed siRNA-mediated knockdown of Nr5a2 in oocytes and analyzed 2-cell embryos by ZGA FISH (Fig. 2D). Nr5a2 knockdown had a negligible effect on ZGA in late 2-cell embryos (34 hpf) (Fig. S6), which could be due to zygotic transcription of *Nr5a2* overcoming the knockdown effect (Fig. 1B). To overcome this potential rescue, we examined early 2-cell embryos (26 hpf) and found that Nr5a2 knockdown reduced nascent ZGA transcripts by ~27% compared to controls (Fig. 2D). Expression of *Nr5a2* mRNA under these conditions rescued ZGA, demonstrating knockdown specificity (Fig. S5I, J). Using a similar approach in early 2-cell embryos, Esrrb knockdown repressed ZGA by ~18% (Fig. 2D). These data suggest that Nr5a2 and to some extent Esrrb contribute to efficient ZGA.

To examine ZGA genome-wide, we performed single-embryo RNA-seq of 2-cell embryos (Fig. 2E-G). Nr5a2 perturbation resulted in up- and downregulation of transcripts, including a strong decrease in *Nr5a2* abundance in SR1848-treated embryos and a moderate decrease in siRNA-treated embryos at this time-point (Fig. 2E and 2F). Since the knockdown efficiency varied between cells (Fig. S7A and S7B), the analysis focused on embryos with the fewest *Nr5a2* transcripts as strong knockdown embryos (Fig. S7B). Nr5a2 knockdown resulted in repression of *Nr2c2* transcription at ZGA, and Nr5a2 inhibition by SR1848 also resulted in downregulation of several orphan nuclear receptors including its own gene, suggesting that Nr5a2 is required for their expression during ZGA (Fig. S8).

Overall, Nr5a2 inhibition resulted in downregulation of 5,891 genes including 1,809 of the strictly defined ZGA genes, corresponding to 72% of ZGA genes (Fig. 2G and Fig. S7F). In contrast,

Nr5a2 knockdown resulted in downregulation of only 448 genes including 197 ZGA genes (Fig. 2G and Fig. S7F). There are at least two non-mutually exclusive explanations for the greater genome-wide effects seen by chemical inhibition. One possibility is that Nr5a2 depletion by siRNA is incomplete due to maternal protein contributions and/or *Nr5a2* transcription during ZGA, whereas SR1848 directly inhibits the Nr5a2 protein. Another possibility is that the inhibitor targets multiple transcription factors. Although this is difficult to exclude, the luciferase assay combined with RNA-seq of embryos suggests that the main target for SR1848 present in early embryos is Nr5a2 (Fig. S3). Consistent with this, Nr5a2 expression in SR1848-treated embryos largely rescued transcription of downregulated genes (Fig. S9). Hence, we conclude that Nr5a2 is required for expression of major ZGA genes.

Nr5a2 binds near TSS of Nr5a2-regulated ZGA genes

To determine whether Nr5a2 binds in the vicinity of ZGA genes in 2-cell embryos, we adapted CUT&Tag for ultra-low input samples and compared the results with published data on histone modifications (*37*-*39*) (Fig. 3A, S10A-C). We optimized conditions for Nr5a2 and Esrrb and obtained transcription factor binding profiles using 300 2-cell embryos. These showed an enrichment for specific and overlapping genomic regions that were not observed with IgG controls (Fig. 3B, S10D and S10E). We identified 4,035 peaks enriched for both Nr5a2 and Esrrb, 4,524 peaks unique to Nr5a2 and 13,141 peaks unique to Esrrb (Fig. 3C).

De novo motif analyses revealed that Nr5a2 motifs were enriched in 77% of all Nr5a2 CUT&Tag peaks, whereas Esrrb motifs were enriched in 52% of all Esrrb CUT&Tag peaks (Fig. 3D). The

motifs that emerged from the detected peaks contained (A/G)(A/G)T upstream of the consensus sequences (Fig. 3D), which was also detected in motif 1 in the *SINE B1/Alu 5YR* upstream of ZGA genes (Fig. 1A and S2D). Indeed, 70% of Nr5a2 peaks and 54% of Esrrb peaks overlapped with *SINE B1/Alu* (Fig. 3E and 3F), suggesting that *SINE B1/Alu* retrotransposons are major targets for Nr5a2 and Esrrb in 2-cell embryos. We cannot exclude that Nr5a2 recruitment outside canonical *SINE B1/Alu* (30% of the peaks) also contributes to ZGA regulation. However, these "isolated" Nr5a2 motifs also show signatures of degenerate *SINE B1/Alu* elements, implying that most Nr5a2 motifs are derived from retrotransposon propagation (Fig. S2F).

We examined the distance from the transcription start site (TSSs) of ZGA genes to the nearest Nr5a2 peak. Nr5a2 peaks were substantially closer to ZGA genes (median: 20.52 kb) than to non-ZGA genes (median: 73.55 kb) (Fig. 3G). Nr5a2 peaks were also much closer to the TSSs of genes that were downregulated vs. upregulated in Nr5a2 knockdown 2-cell embryos. Focusing on downregulated genes, we found that the distance for knockdown-specific (median: 18.73 kb) and SR1848-specific genes (25.57 kb) was much closer than for non-differentially regulated genes (median: 72.3 kb) (Fig. S7G). These findings suggest that the inhibitor-specific genes are similar to knockdown-specific genes with respect to the distance of Nr5a2 binding to TSS and imply that the former are unlikely to be off-target effects (Fig. S7G).

To analyze the correlation between gene expression changes during ZGA and Nr5a2 occupancy, we conservatively defined occupancy as the sum of Nr5a2 CUT&Tag signals over Nr5a2 motifs in the 8 kb upstream region of genes (Fig. S10F). The gene expression changes between G2-phase

zygotes and 2-cell stages were significantly higher for genes with higher Nr5a2 occupancy and stronger gene expression changes correlated with increased Nr5a2 occupancy (Fig. 3H and S10G). We further examined the correlation between gene expression changes in Nr5a2 knockdown and SR1848-treated embryos and Nr5a2 occupancy. Genes with Nr5a2 occupancy were significantly more downregulated by Nr5a2 perturbation than unoccupied genes. Among genes occupied by Nr5a2, downregulation of gene expression was also inversely correlated with Nr5a2 occupancy (Fig. 3I, 3J, S10H and S10I). Further analysis showed that >50% (921/1809) of ZGA genes downregulated by SR1848 treatment showed some Nr5a2 CUT&Tag signal in their extended promoter regions, whereas both the number and proportion of Nr5a2 occupied regions were lower for unchanged or upregulated ZGA genes (Fig 3K). Based on these data, we propose that Nr5a2 binding close to ZGA genes promotes their transcriptional activation at ZGA, although it cannot be excluded that distant binding of Nr5a2 also contributes to ZGA regulation. These data also indicate that Nr5a2 controls expression of many ZGA genes directly.

Nr5a2 and Esrrb target cell-type specific enhancers

To analyze whether the regions bound by Nr5a2 and Esrrb in 2-cell embryos could be *cis*-regulatory elements (cREs), we compared our CUT&Tag binding profiles with published ATAC-seq and histone modification ChIP-seq data from 2-cell embryos (*40, 41*) and found that these transcription factors bind to open chromatin (Fig. 4A). We classified enhancer-like signatures (ELS) as regions with high H3K27ac and ATAC-seq and low H3K4me3 signals, vs. promoter-like signatures (PLS) with high H3K4me3 and ATAC-seq and low H3K27ac signals (*42*) (Fig. S11A and S11B). Nr5a2-binding was enriched at distal enhancer-like signatures (dELS) (Fig. 4B), whereas Esrrb-binding was detected at ELS, PLS and other regions (Fig. 4B). These findings

suggest that Nr5a2 and Esrrb target common and distinct loci, and Nr5a2 preferentially binds enhancers in 2-cell embryos.

Esrrb, and to a lesser extent Nr5a2, are also expressed in ES cells (*26*). We examined published data to determine if these transcription factors are recruited to distinct cREs in totipotent 2-cell embryos vs. pluripotent ES cells (*43*, *44*). This comparison suggests that Nr5a2 and Esrrb bind to prominent ATAC-seq and H3K27ac peaks (ELS) that are present (i) in 2-cell embryos only, (ii) in 2i medium-cultured ES cells only, or (iii) in both 2-cell embryos and ES cells (Fig. 4C). We identified 9,099 2-cell-embryo-specific ELS (2C-ELS), 6,460 2i-mESC-specific ELS (2i-ELS) but only 110 common ELS sites (Fig. S11C). Aggregation plot analysis showed that both Nr5a2 and Esrrb are specifically bound to each cell-type specific ELS with H3K27ac enrichment and chromatin accessibility (Fig. 4D, S11D and S11E). These data suggest that Nr5a2 and Esrrb are involved in setting up largely distinct gene-regulatory networks during different stages of embryogenesis by defining development (cell-type) specific enhancers.

Nr5a2 directly promotes chromatin accessibility

To test whether Nr5a2 and Esrrb are required for chromatin accessibility in 2-cell embryos, we developed a microscopy-based approach to quantify open chromatin in single cells, which we termed ChARM (Chromatin Accessibility Revealed by Microscopy). Similar to ATAC-see (*45*), ChARM uses Tn5-mediated insertion of adaptor DNA into accessible chromatin but uses hybridization chain reaction (HCR) to amplify signals of the inserted adaptor DNA. This approach generated quantifiable spot-like patterns rather than the diffusive signals observed by ATAC-see

(Fig. S12A and S12B). As a proof of concept, we examined Bromodomain 4 (Brd4)-dependent open chromatin. Two-cell embryos treated with the Brd4 inhibitor JQ-1 showed reduced H3K27 acetylation, as expected (Fig. S12C and S12D), and also significantly reduced ChARM signal (Fig. S12C and S12E).

To test whether Nr5a2 and Esrrb promote chromatin accessibility, we performed knockdown in oocytes and analyzed 2-cell embryos by ChARM at 26 hpf. Chromatin accessibility was reduced in Nr5a2- and Esrrb-siRNA embryos compared to controls, suggesting that both transcription factors contribute to chromatin accessibility (Fig. 5A and 5B). Similar results were obtained for 2-cell embryos treated with SR1848 (Fig. S12F). Together, these findings suggest that these transcription factors promote chromatin accessibility.

To determine whether Nr5a2 functions as a pioneer factor, we tested whether Nr5a2 is required for chromatin opening at sites where it is bound. SR1848 was used to inhibit Nr5a2 and obtain sufficient cell numbers to perform Omni Assay for Transposase-Accessible Chromatin using sequencing (Omni ATAC-seq) (Fig. S13A-C) (*46*). Changes in accessibility at TSSs were significantly correlated with gene expression changes (Fig. S13D and S13E), suggesting that changes in chromatin accessibility reflect gene expression changes that depend on Nr5a2.

By comparing ATAC-seq peaks from SR1848 vs DMSO treated embryos, we identified 492 differentially accessible regions (DAR) that all showed loss of accessibility (10% false discovery

rate (FDR), Fig. 5C). Nearly half of the DARs overlapped with Nr5a2 CUT&Tag peaks in 2-cell embryos (Fig 5D). To determine whether the DARs are dependent on Nr5a2, we performed motif enrichment analysis and examined Nr5a2 occupancy. Although Nr5a2 and similar motifs occurred with similar frequencies in non-DAR and DAR regions (42% and 56%, Fig. S13F), DARs showed higher occupancy of Nr5a2 in 2-cell embryos (Fig. 5E). These regions are largely inaccessible in zygotes (Fig. S13G), suggesting that their opening occurs in 2-cell embryos.

We also investigated whether accessibility changes in CREs are related to Nr5a2 binding. Chromatin accessibility of PLS and ELS occupied by Nr5a2 is more reduced by SR1848 treatment than unbound regions (Fig. 5F). These results suggest that Nr5a2 binding promotes opening of chromatin, which is a hallmark of pioneer factors.

Nr5a2 and Esrrb bind nucleosomal DNA *in vitro*

Another hallmark of pioneer factors is the ability to target their (partial) motif on nucleosomal DNA (*23*). To test whether Nr5a2 and Esrrb possess this ability, we purified full-length and DNA-binding domains (DBDs) of mouse Nr5a2 and Esrrb, and mouse histones as recombinant proteins (Fig. S14A-D). Mass photometry showed that Nr5a2 forms a monomer, whereas Esrrb forms a dimer or multimer in solution (Fig. S14E).

To examine the binding specificity to naked DNA, we performed fluorescence polarization (FP) measurements and electrophoretic mobility shift analysis (EMSA). Since Esrrb showed

promiscuous binding to DNA (Fig. S14F), we performed experiments in the presence of low concentrations of competitor DNA. Nr5a2 bound to its own motif and the Esrrb motif with comparable affinity (K_d of 5.62 ± 1 nM and 6.49 ± 0.35 nM, respectively). In contrast, Esrrb bound to its own motif with a higher affinity than the Nr5a2 motif (K_d of 8.56 ± 0.69 nM and 625.53 ± 133 nM, respectively) (Fig. 6A and S14G). The same motif specificity was observed by EMSA (Fig. S6H and S6I).

To test whether Nr5a2 and Esrrb bind nucleosomal DNA, we performed SeEN-seq (Selected Engagement on Nucleosome sequencing), in which motifs are tiled throughout the Widom 601 nucleosome positioning sequence (Fig. 6B) (*47*). We prepared nucleosome libraries with 5 bp shifts in the position of each motif (Fig. S15A, S15C, Table S13). EMSA showed a shift of Nr5a2 with the nucleosome library but not with the 601 "template" nucleosome lacking a motif, suggesting that the Nr5a2-nucleosome complex forms in a motif-dependent manner (Fig. S15B). Similarly, Esrrb showed specific band shifts with the nucleosome library (Fig. S15D). Transcription factor-bound and -unbound fractions were purified and sequenced. SeEN-seq revealed that Nr5a2 and Esrrb preferentially bound at the entry-exit sites on the nucleosome (Fig. 6C), reminiscent of how the pioneer factors Oct4-Sox2 and GATA3 binding to nucleosomal DNA (*47*, *48*).

To investigate Nr5a2 and Esrrb binding at specific motif positions on nucleosomes, we selected high enrichment sites and reconstituted nucleosomes with motifs at superhelical locations (SHL) -6, +5.5 and +6 for Nr5a2; and -6 and +5.5 for Esrrb (Fig. 6D and S15E). Consistent with the

SeEN-seq results, a bandshift was detected for Nr5a2 binding to nucleosomes containing the motif at SHL-6, SHL+5.5 and SHL+6 (Fig. 6E and S15F). Motif-specific binding was also detected for the Nr5a2 DBD (Fig. S15H). Esrrb full-length and DBD showed similar nucleosome binding efficiencies (Fig. 6F, S15G, and S15I). A stronger bandshift was detected for Esrrb binding to nucleosomes with motifs at SHL-6 and SHL+5.5 than for 601 template (Fig. 6F and S15G). Since it is conceivable that the binding to entry/exit sites of nucleosomal DNA is due to the tight binding of 601 DNA to histones, we also tested whether Nr5a2 could bind to an endogenous sequence in which the motif was closer to the dyad axis. We found that Nr5a2 could also bind nucleosomes containing this endogenous sequence (Fig. S15J and S15K), suggesting that Nr5a2 can recognize its motif in different superhelical locations.

To test the specificity of the transcription factor-nucleosome interactions, we performed competition assays with naked DNA. Specific but not non-specific DNA outcompeted binding of Nr5a2 and Esrrb to nucleosomes (Fig. 6G and 6H). Thus, both Nr5a2 and Esrrb directly engage with their own motifs on nucleosomal DNA. Overall, our data show that Nr5a2 and Esrrb have properties consistent with pioneer factor activity *in vivo* and *in vitro*.

Discussion

We provide evidence that the orphan nuclear receptor Nr5a2 is a pivotal pioneer factor that activates up to 72% of major ZGA genes in mouse embryos. Nr5a2 binds to *SINE B1*/*Alu* retrotransposable elements in the *cis*-regulatory regions of nearly half of all ZGA genes and this binding correlates with transcriptional changes. The genome-wide regulation of ZGA by Nr5a2

and potentially other orphan nuclear receptors exceeds that of Nfya, which affects ~15% of ZGA genes (*13*) and is comparable to Zelda in *Drosophila* (~75% of ZGA genes) (*5*) and the collective activity of three pluripotency factors in zebrafish (>75% of ZGA genes) (*8, 9*). Based on genome-wide binding profiles, chromatin accessibility and *in vitro* nucleosome binding assays, we propose that Nr5a2 acts locally to promote chromatin opening and functions as a pioneer factor to initiate ZGA in mouse embryos.

The classification of Nr5a2 and Esrrb as pioneer factors suggests that the mechanism of multiple pioneer factors triggering ZGA is evolutionarily conserved from fly to mouse and possibly human, despite differences in transcription factor identities. The species-specific regulation of some ZGA genes is supported by the recent finding that human-specific TPRXs contribute to ZGA (*49*). However, ZGA is a fundamental process that initiates control of the zygotic genome for all multicellular organisms. It is therefore important to note that Nr5a2 is conserved in all taxa of metazoa that we examined and maternally provided in embryos of model organisms and human, implying an ancestral function in early development (Fig. S16) (*50, 51*). Although *SINE B1* are murine-specific, the human genome harbors the related *Alu* retrotransposable elements that also originated from 7SL RNA. *Alu* elements contain two full and one degenerated Nr5a2 motifs (Fig. S16B). We speculate that the regulation of ZGA by orphan nuclear receptors such as Nr5a2 is a conserved mechanism, at least amongst mammals.

Our work provides a conceptual "*ex uno plura*" (many from one) framework for ZGA activation based on three findings: 1) Nr5a2 directly activates transcription of many ZGA genes, 2) Nr5a2 is

required for transcription of its own gene, *Nr2c2* and other orphan nuclear receptors during ZGA (Fig. S8), and 3) orphan nuclear receptors that are normally expressed during ZGA such as Nr2c2 can principally activate ZGA genes when overexpressed at an earlier stage (Fig. S5F). We therefore propose that Nr5a2 activates transcription of ZGA genes, which includes orphan nuclear receptors that can recognize motif 1 and potentiate the activation of ZGA genes (Fig. 6I). Current data do not allow us to distinguish whether Nr5a2 co-operates with newly synthesized transcription factors or whether each functions redundantly after ZGA initiation by Nr5a2. Whether transcription factors that bind to the other motifs in *SINE B1/Alu* contribute to ZGA also remains to be elucidated.

An intriguing finding is that the same transcription factors are important for totipotency and later for pluripotency during mammalian development. Nr5a2 and Esrrb target distinct enhancer-like sequences in 2-cell embryos and 2i-ES cells. How these transcription factors selectively establish cell-type specific enhancers remains to be determined. Since *SINE B1/Alu* elements contain several motifs, it is conceivable that cooperative binding of multiple transcription factors including Nr5a2 and Esrrb establishes 2-cell embryo-specific active enhancers (Fig. 6J). Our findings imply that maternally provided Nr5a2 initiates a cascade of transcription factor bindings that lead to the transcriptional waves at the start of life.

Methods summary

A detailed materials and methods section is provided in the supplementary materials. Briefly, the care and use of the mice at IMBA were carried out in agreement with the authorizing committee according to the Austrian Animal Welfare law and the guidelines of the International Guiding

Principles for Biomedical Research Involving Animals (CIOMS, the Council for International Organizations of Medical Sciences). All animals housed at MPIB were sacrificed prior to the removal of organs in accordance with the European Commission Recommendations for the euthanasia of experimental animals (Part 1 and Part 2). Breeding and housing as well as the euthanasia of the animals are fully compliant with all German (e.g. German Animal Welfare Act) and EU (e.g. Directive 2010/63/EU) applicable laws and regulations concerning care and use of laboratory animals.

In vitro maturation and fertilization were performed as described (52) with some changes. For siRNA knockdown, isolated GV oocytes were microinjected with sets of two siRNAs against targets or control. Trim-Away was performed according to the published method (*35*,*36*). For single embryo RNA-seq, sequencing libraries were prepared using the SmartSeq2 protocol described (*53*).

For the Fluorescent In Situ Hybridization (ZGA-FISH), FISH probes were designed as three ($ZGA^{\#1}$, $ZGA^{\#2}$ and $ZGA^{\#3}$) categories based on their transcript abundance in 2-cell embryos. Embryos at the appropriate developmental time were fixed with 4% PFA. Prehybridization was done in 1x pre-HYBEC supplemented with 50 µg/ml heparin by placing the embryos into a humid chamber in a water bath. Preheated hybridization mixture was applied to the embryos and hybridization was carried out. For Chromatin Accessibility Revealed by Microscopy (ChARM), split-initiator oligonucleotides were applied to the prehybridized embryos. Excess initiator molecules were washed away, and prepared hairpins were applied to the embryos to initiate a hybridization chain reaction.

CUT&Tag was performed as described previously (*37*) with a few modifications. Briefly, cells with intact zona pellucida were incubated with ice-cold extraction buffer. Pre-extracted cells were further lightly fixed by DPBS with 0.1% formaldehyde for 2 min at room temperature.

Cells were incubated with appropriate antibodies, and prepared DNA libraries were sequenced on a NextSeq 500. Omni ATAC-seq was performed as described previously (*46*).

The recombinant Nr5a2 and Esrrb were bacterially expressed and were purified. Mouse histones H2A, H2B, H3.3 and H4 were expressed and purified according to published protocols (*54*). The nucleosomes were reconstituted by salt dialysis method and were further purified by polyacrylamide gel electrophoresis using a Prep Cell apparatus. SeEN-seq assay was performed as described previously with a few modifications (*47*). Nr5a2 motif (TCAAGGCCA) or Esrrb motif (TCAAGGTCA) was tiled 5 bp interval across the entire Widom 601 DNA sequence (*55*). DNA libraries were prepared as described previously (*47*).

References and Notes

1. J. B. Gurdon, The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J Embryol Exp Morphol* **10**, 622-640 (1962).

2. F. Aoki, D. M. Worrad, R. M. Schultz, Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev Biol* **181**, 296-307 (1997).10.1006/dbio.1996.8466.

3. G. Flach, M. H. Johnson, P. R. Braude, R. A. Taylor, V. N. Bolton, The transition from maternal to embryonic control in the 2-cell mouse embryo. *EMBO J* **1**, 681-686 (1982).

4. K. I. Abe *et al.*, Minor zygotic gene activation is essential for mouse preimplantation development. *Proc Natl Acad Sci U S A* **115**, E6780-E6788 (2018).10.1073/pnas.1804309115.

5. H. L. Liang *et al.*, The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* **456**, 400-403 (2008).10.1038/nature07388.

6. M. M. Gaskill, T. J. Gibson, E. D. Larson, M. M. Harrison, GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. *Elife* **10**, (2021).10.7554/eLife.66668.

7. J. Duan *et al.*, CLAMP and Zelda function together to promote Drosophila zygotic genome activation. *Elife* **10**, (2021).10.7554/eLife.69937.

8. M. T. Lee *et al.*, Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360-364 (2013).10.1038/nature12632.

9. M. Leichsenring, J. Maes, R. Mossner, W. Driever, D. Onichtchouk, Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **341**, 1005-1009 (2013).10.1126/science.1242527.

10. G. E. Gentsch, T. Spruce, N. D. L. Owens, J. C. Smith, Maternal pluripotency factors initiate extensive chromatin remodelling to predefine first response to inductive signals. *Nat Commun* **10**, 4269 (2019).10.1038/s41467-019-12263-w.

11. G. Wu *et al.*, Establishment of totipotency does not depend on Oct4A. *Nat Cell Biol* **15**, 1089-1097 (2013).10.1038/ncb2816.

12. 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).10.1038/cr.2016.20.

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

14. A. De Iaco *et al.*, DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* **49**, 941-945 (2017).10.1038/ng.3858.

15. 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).10.1038/ng.3844.

16. J. L. Whiddon, A. T. Langford, C. J. Wong, J. W. Zhong, S. J. Tapscott, Conservation and innovation in the DUX4-family gene network. *Nat Genet* **49**, 935-940 (2017).10.1038/ng.3846.

17. 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).10.1038/s41594-021-00590-w.

18. D. Lohnes *et al.*, Function of retinoic acid receptor gamma in the mouse. *Cell* **73**, 643-658 (1993).10.1016/0092-8674(93)90246-m.

19. A. De Iaco, S. Verp, S. Offner, D. Grun, D. Trono, DUX is a non-essential synchronizer of zygotic genome activation. *Development* **147**, (2020).10.1242/dev.177725.

20. Z. Chen, Z. Xie, Y. Zhang, DPPA2 and DPPA4 are dispensable for mouse zygotic genome activation and preimplantation development. *Development **148**, (2021).*10.1242/dev.200178

21. 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).10.1242/dev.200191*

22. 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).10.1038/s41588-019-0418-7.

23. M. Iwafuchi-Doi, K. S. Zaret, Pioneer transcription factors in cell reprogramming. *Genes Dev* **28**, 2679-2692 (2014).10.1101/gad.253443.114.

24. K. N. Schulz *et al.*, Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res* **25**, 1715-1726 (2015).10.1101/gr.192682.115.

25. Y. Sun *et al.*, Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. *Genome Res* **25**, 1703-1714 (2015).10.1101/gr.192542.115.

26. 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).10.1242/dev.199604.

27. Y. Wang *et al.*, Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. *Nat Commun* **12**, 1247 (2021).10.1038/s41467-021-21409-8.

28. J. Luo *et al.*, Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. *Nature* **388**, 778-782 (1997).10.1038/42022.

29. 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).10.1002/dvdy.20996.

30. S. Busby *et al.*, in *Probe Reports from the NIH Molecular Libraries Program*. (Bethesda (MD), 2010).

31. C. A. Corzo *et al.*, Antiproliferation activity of a small molecule repressor of liver receptor homolog 1. *Mol Pharmacol* **87**, 296-304 (2015).10.1124/mol.114.095554.

32. A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods*. **5**, 877–879 (2008).

33. S. C. Little, K. S. Sinsimer, J. J. Lee, E. F. Wieschaus, E. R. Gavis, Independent and coordinate trafficking of single Drosophila germ plasm mRNAs. *Nat. Cell Biol.* **17**, 558–568 (2015).

34. I. Gaspar, F. Wippich, A. Ephrussi, Enzymatic production of single-molecule FISH and RNA capture probes. *RNA.* **23**, 1582–1591 (2017).

35. D. Clift *et al*. A Method for the Acute and Rapid Degradation of Endogenous Proteins. *Cell* **171**, 1692-1706 (2017).10.1016/j.cell.2017.10.033.

36. D. Clift, C. So, W.A. McEwan, L.C. James, M. Schuh. Acute and rapid degradation of endogenous proteins by Trim-Away. *Nat Protoc.* **13**, 2149-2175 (2018). doi:10.1038/s41596-018-0028-3

37. H. S. Kaya-Okur *et al.*, CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930 (2019).10.1038/s41467-019-09982-5.

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

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

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

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

42. ENCODE Project Consortium et al. Expanded encyclopaedias of DNA elements in the human and mouse. *Nature* **583**, 699-710(2020).10.1038/s41586-020-2493-4

43. Y. Atlasi *et al.*, Epigenetic modulation of a hardwired 3D chromatin landscape in two naive states of pluripotency. *Nat Cell Biol* **21**, 568-578 (2019).10.1038/s41556-019-0310-9.

44. C. Galonska, M. J. Ziller, R. Karnik, A. Meissner, Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming. *Cell Stem Cell* **17**, 462-470 (2015).10.1016/j.stem.2015.07.005.

45. X. Chen *et al.*, ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat Methods* **13**, 1013-1020 (2016).10.1038/nmeth.4031.

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).10.1038/nmeth.4396.

47. A. K. Michael *et al.*, Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science* **368**, 1460-1465 (2020).10.1126/science.abb0074.

48. H. Tanaka *et al.*, Interaction of the pioneer transcription factor GATA3 with nucleosomes. *Nat Commun* **11**, 4136 (2020).10.1038/s41467-020-17959-y.

49. Z. Zou, *et al*. Translatome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome activation. *Science* abo7923 (2022).10.1126/science.abo7923.

50. Z.H. Zhao, T.G. Meng, A. Li, H. Schatten, Z.B. Wang, Q.Y. Sun. RNA-Seq transcriptome reveals different molecular responses during human and mouse oocyte maturation and fertilization. *BMC Genomics* **21**, 475 (2020).10.1186/s12864-020-06885-4

51. W. Hu *et al*. Single-cell transcriptome and translatome dual-omics reveals potential mechanisms of human oocyte maturation. *Nat Commun*. **13**, 5114 (2022).10.1038/s41467-022-32791-2

52. J. Gassler *et al.,* A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J* **36,** 3600-3618 (2017).10.15252/embj.201798083.

53. S. Picelli *et al*., Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181 (2014).10.1038/nprot.2014.006.

54. T. Kujirai *et al*., Methods for Preparing Nucleosomes Containing Histone Variants. *Methods Mol Biol* **1832**, 3-20 (2018).10.1007/978-1-4939-8663-7_1.

55. P. T. Lowary, J. Widom, New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol 276, 19-42 (1998).10.1006/jmbi.1997.1494.

56. I. Gaspar, F. Wippich, A. Ephrussi, Enzymatic production of single-molecule FISH and RNA capture probes. RNA 23, 1582-1591 (2017).10.1261/rna.061184.117.

57. I. Gaspar, V. Sysoev, A. Komissarov, A. Ephrussi, An RNA-binding atypical tropomyosin recruits kinesin-1 dynamically to oskar mRNPs. EMBO J 36, 319-333 (2017).10.15252/embj.201696038.

58. I. Gaspar et al., Klar ensures thermal robustness of oskar localization by restraining RNP motility. J Cell Biol 206, 199-215 (2014).10.1083/jcb.201310010.

59. B. Liu et al., The landscape of RNA Pol II binding reveals a stepwise transition during ZGA. Nature 587, 139-144 (2020).10.1038/s41586-020-2847-y.

60. A. Sonn-Segev et al., Quantifying the heterogeneity of macromolecular machines by mass photometry. Nat Commun 11, 1772 (2020).10.1038/s41467-020-15642-w.

61. D. Wu, G. Piszczek, Standard protocol for mass photometry experiments. Eur Biophys J 50, 403-409 (2021).10.1007/s00249-021-01513-9.

62. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34, 525-527 (2016).10.1038/nbt.3519.

63. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res 4, 1521 (2015).10.12688/f1000research.7563.2.

64. 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).10.1186/s13059-014-0550-8.

65. K. K. Dey, C. J. Hsiao, M. Stephens, Visualizing the structure of RNA-seq expression data using grade of membership models. PLoS Genet 13, e1006599 (2017).10.1371/journal.pgen.1006599.

66. L. Li. GADEM: a genetic algorithm guided formation of spaced dyads coupled with an EM algorithm for motif discovery. J Comput Biol 16, 317-329 (2009).10.1089/cmb.2008.16TT.

67. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357-359 (2012).10.1038/nmeth.1923.

68. F. Ramirez et al., deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44, W160-165 (2016).10.1093/nar/gkw257.

69. R. Stark, G. Brown. DiffBind: differential binding analysis of ChIP-Seq peak data. (2011). Bioconductor. Available online at: http://bioconductor.org/packages/release/bioc/html/DiffBind.html.

70. T.E. Keenan et al., Molecular correlates of response to eribulin and pembrolizumab in hormone receptor-positive metastatic breast cancer. Nat Commun. 12, 5563 (2021).10.1038/s41467-021-25769-z.

71. S. Heinz et al., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576-58 (2010).10.1016/j.molcel.2010.05.004.

72. Y. Zhang et al., Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137 (2008).10.1186/gb-2008-9-9-r137.

73. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: scanning for occurrences of a given motif. Bioinformatics 27, 1017-1018 (2011).10.1093/bioinformatics/btr064.

74. G. Yu, L. G. Wang, Q. Y. He, ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382-2383 (2015).10.1093/bioinformatics/btv145

75. Q. Wang et al., Exploring Epigenomic Datasets by ChIPseeker. Curr Protoc 2, e585 (2022).10.1002/cpz1.585

76. M. S. Lawrence et al., Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499, 214-218 (2013).10.1038/nature12213.

77. H. Li et al., The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079 (2009).10.1093/bioinformatics/btp352.

78. J. Ruan et al., TreeFam: 2008 Update. Nucleic Acids Res. 36, D735-D740 (2008).10.1093/nar/gkm1005.

79. S. Guindon S et al., New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59, 307-321 (2010). 10.1093/sysbio/syq010

80. M. Hasegawa, H. Kishino, T. Yano, Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22, 160-174 (1985).10.1007/BF02101694

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

82. R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research, 32, 1792–1797 (2004).10.1093/nar/gkh340

83. L.N. Voong et al., Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. Cell. 167, 1555-1570 (2016).10.1016/j.cell.2016.10.049.

Acknowledgements: We would like to thank B. Kunkel, R. Hornberger, K. Straßer, H. Jung and E. Krstevska-Vulic for technical assistance. We thank all members of the K.T. laboratory including B. J. Dequeker, E. E. Chatzidaki, S. Feng and H. Marvanova for assistance with oocyte isolation. We thank L. Zeitler (Murray lab, MPIB) for technical assistance with the luciferase assay. HEK293T cells were a kind gift from Prof. Dr. Reinhard Fässler (MPIB). Expression vectors used in the luciferase reporter assay were a kind gift from Patrick R. Griffin (The Scripps Research Institute). JQ-1 was a kind gift from J. Zuber (IMP). We thank R. S. Grand (Schübeler lab, FMI) for advice on SeEN-seq analysis. We thank R. H. Kim for sequencing single embryo samples at the NGS facility in the Department of Totipotency, MPIB. Illumina sequencing of the developmental transcriptome analysis was performed by the NGS facility at Vienna BioCenter

Core Facilities (VBCF). We thank M. Novatchkova (IMP) for RNA-seq analysis of MZT data. We would like to thank J.-M. Peters and Life Science Editors for critical reading of the manuscript.

Funding:

L'Oréal Austria Fellowship for Women in Science (JG)

Austrian Science Fund (FWF) DK Chromosome Dynamics grant W1238-B20 (KT, JG)

JSPS Overseas Research Fellowship (WK)

European Research Council grant ERC-CoG-818556 TotipotentZygotChrom (KT)

Human Frontier Science program RGP0057-2018 (KT)

Austrian Academy of Sciences (KT)

Max Planck Society (KT)

Author contributions: JG collected MZT samples for developmental RNA-Seq. JG, AM, AV, LCR and LGH performed oocyte microinjections and collected samples for FISH and RNA sequencing analyses. JG performed Esrrb and Nr5a2 immunofluorescence assays. JG carried out embryo developmental competence assays. IG designed and IG and LGH carried out the ZGA-FISH experiments. IG and NR analysed the ZGA-FISH experiments. IG, PK and SR analysed single embryo RNA-seq data. WK and SL established the method of CUT&Tag. WK performed CUT&Tag and Omni ATAC-seq using mouse oocytes and embryos. SR, IG and PK analyzed

CUT&Tag data. IG and JG performed ChARM and analysed data. SR, IG and PK analysed Omni ATAC-seq data. WK, MK and MZ performed protein purification, SeEN-seq and biochemical analyses. SR analyzed SeEN-seq data. AL, RJA and NR carried out the luciferase reporter assay. KT conceived the project and supervised the work. WK, IG, JG, SR, and KT planned the project, designed the experiments, and wrote the manuscript. All authors discussed the results and commented on the manuscript.

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

Data and materials availability: Requests for plasmids generated in this study should be directed to the corresponding author. All RNA sequencing data have been deposited to the Gene Expression Omnibus (GEO) under the accession number GSE178661. All CUT&Tag and Omni ATAC-seq data have been deposited to the Gene Expression Omnibus (GEO) under the accession number GSE178234. ATAC-seq and histone modification ChIP-seq in 2-cell embryo datasets were downloaded from GEO accession GSE66390 and GSE72784, respectively. For the dataset in mESCs, ChIP-seq data of Nr5a2, Esrrb, and H3K27ac were obtained from GSE92412. H3K4me3

ChIP-seq was obtained from GSE56312. MNase-seq data in mESCs was obtained from GSE82127.

Supplementary Materials

Materials and Methods

Figs. S1 to S16

Tables S1 to S14

References (56–83)

Figure 1. Nr5a2 is required for early embryonic development

(**A**) Sequence logo of the identified super motif with individual motifs highlighted. Motif #1 resembles the cognate binding sequence of Nr5a2 or Esrrb. (**B and C**) Transcript abundance of main protein coding isoforms of Nr5a2 (B) and Esrrb (C) during the oocyte to 2-cell (2C) embryo transition from pure B6 (B6xB6) and B6CASTF1 (B6xCAST) mice (see methods). (**D and E**) Representative immunofluorescence images depicting Nr5a2 (D) and Esrrb (E) in embryonic stages. Maternal and paternal zygotic nuclei are indicated by symbols. H3K27ac immunofluorescence is shown to indicate proper antibody penetration and to outline the nuclei. Depicted are single z-slices. Scale bar represents 20 µm. (**F**) Schematic of early embryo stages including timing and length of inhibitor treatment. (**G**) Stereomicroscopic example images of the embryonic stages observed at the indicated time points in different conditions. Mock control embryos were treated with DMSO. Scale bars are 150 µm. (**H**) Quantification of four replicate experiments of embryonic development. Sample sizes are: mock: n= 19, 23, 16, 33; SR1848 (6-108 hpf): n= 18, 20, 21; SR1848 (6-36 hpf): n=18, 20, 20, 30 cells; SR1848 (36-108 hpf): n=20, 21, 17, 29; each experiment comprising of 6-12 females.

Figure 2. Nr5a2 and Esrrb are required for efficient ZGA

(**A**) Representative nascent ZGA-FISH images of DMSO control, SR1848-treated and SR1848-treated and Nr5a2-GFP mRNA microinjected 2-cell embryos. Right panel shows a quantification of total nascent ZGA-FISH signal within nuclei of 2-cell embryos. Biological replicates are shown with different bullet styles. Black dots and bars show the mean and 95% confidence interval per replicate, red dots and bars indicate the mean and 95% confidence interval of the experimental condition with all replicates merged (also for B-D). Sample sizes: control: n = 16, 13, 13, 9, 19; SR1848: n = 21, 18, 12, 16, 13; SR1848+Nr5a2-GFP: n = 28, 20, 12, 16, 20 nuclei. Scale bars are 5 µm. (**B**) Representative nascent ZGA-FISH images of 2-cell embryos microinjected with Nr5a2-GFP or Esrrb-GFP mRNA. Right panel shows a quantification of total nascent ZGA-FISH signal within nuclei of 2-cell embryos. Sample sizes: control: n = 20, 22, 19 and 34, 24; Nr5a2-GFP: n = 21, 30, 23 and 29, 24; Esrrb-GFP: n = 23, 18, 23 nuclei. Scale bars are 5 µm. (**C**) Schematic illustration showing Trim-Away-mediated Nr5a2 knockdown in 2-cell embryos. Representative nascent ZGA-FISH images of 2-cell embryos microinjected with control IgG or Nr5a2 antibodies. Right panel shows a quantification of total nascent ZGA-FISH signal within nuclei of 2-cell

embryos in the respective conditions. Sample sizes: control IgG: n = 35, 31, 18; anti Nr5a2 n = 28, 26, 22 nuclei. Scale bars are 5 µm. (**D**) Representative nascent ZGA-FISH images of *Esrrb*, *Nr5a2* knockdown embryos and their controls during early ZGA (26 hpf). Right panel shows a quantification of total nascent ZGA-FISH signal within nuclei of *Nr5a2*, *Esrrb* knockdown and control 2-cell embryos in early ZGA in two replicates. Sample sizes: control: n = 15, 15; Esrrb knockdown: n = 19, 16; *Nr5a2* knockdown: n = 16, 18. Scale bars are 5 µm. (**E**) Bland–Altman (MA) plot comparing strong Nr5a2 knockdown embryos to control by DESeq2. (**F**) MA plot comparing chemically inhibited 2-cell embryos to control by DESeq2. (**G**) Euler diagrams showing the overlap of major ZGA genes, down-regulated genes in strong Nr5a2 knockdown and in SR1848-treated 2-cell embryos.

Figure 3. Genomic localization of Nr5a2 and Esrrb during ZGA

(**A**) Schematic illustration of CUT&Tag on 2-cell embryos. (**B**) Representative integrative genomics viewer (IGV) snapshot showing enrichment of IgG control (grey), Nr5a2 (red) and Esrrb (blue) within the indicated region on chromosome 3. The CUT&Tag signal for each factor was merged from two replicates. (**C**) Euler diagram showing overlap between Nr5a2 and Esrrb peaks in 2-cell embryos. (**D**) DNA sequence identified by Homer *de novo* motif analysis from Nr5a2 and Esrrb peaks in comparison to Nr5a2 (MA0505.1) and Esrrb (MA0141.1) motifs reported in the JASPAR database. The p-value of the motif comparison and percent of peaks containing *de novo* motifs are indicated. Black dot squares show three extended nucleotides that are identified from CUT&Tag in 2-cell embryos. (**E**) Pie chart showing the percentage of peaks with *SINE B1*. (**F**) Heat maps showing the enrichment of repeats (subfamily) in Nr5a2-unique, Esrrb-unique and Nr5a2/Esrrb-overlap peaks in 2-cell embryos. (**G**) Violin plots showing the distance from TSS to the nearest Nr5a2 peaks between down-regulated, up-regulated and unaffected genes in Nr5a2 KD / SR1848-treated embryos, ZGA genes and non-ZGA genes. Red dashed line shows 8 kb. (**H-J**) Box plots showing the expression change of genes with no, weak, moderate or strong Nr5a2 CUT&Tag signal at Nr5a2 motifs in their 8 kb upstream regions (H) between G2 zygotes and 2-cell embryos, (I) in Nr5a2 and (J) KD in SR1848-treated 2-cell embryos. Bonferroni corrected p-values of pairwise Mann-Whitney U tests against genes with no Nr5a2 occupancy are shown. (**K**) Bar chart representing the extended ZGA genes according to their expression changes upon

chemical inhibition of Nr5a2 (y-axis) and the total Nr5a2 CUT&Tag signal measured in their 8k bp upstream regions (black: strong, gray: weak, white: no occupancy)

5

Figure 4. Cell-type specific distribution of Nr5a2 and Esrrb

(**A**) Line plots (above) and heatmaps (below) of Nr5a2, Esrrb, ATAC-seq, H3K27ac and H3K4me3 enrichments in 2-cell embryos (*Z*-score normalized RPKM value). Each row is classified by Nr5a2 unique, Esrrb-unique and Nr5a2/Esrrb-overlap regions. (**B**) Classification of cREs bound by Nr5a2 and Esrrb according to epigenetic signatures. We defined 4 major groups: enhancer-like signatures (ELS), promoter-like signatures (PLS), ATAC only, and other. Number of cREs bound by Nr5a2 and Esrrb are shown as bar graphs. The criteria of proximity are described in the methods section. pELS: proximal ELS. dELS: distal ELS. (**C**) Representative IGV snapshots show the enrichment of Nr5a2 (red), Esrrb (blue), ATAC-seq (yellow) and H3K27ac (green) in 2-cell embryo and 2i-mESC. Nr5a2 and Esrrb peaks are highlighted as blue. (**D**) Average profiles of Nr5a2, Esrrb, ATAC-seq, H3K27ac and H3K4me3 in 2C-ELS and 2i-ELS. The signal in a ± 2 kb window flanking the peak center is shown. Blue and orange lines indicate peaks on 2C-ELS and 2i-ELS, respectively.

Figure 5. **Nr5a2 regulates chromatin accessibility during ZGA**

(**A**) Representative images of ChARM (green) in 2-cell embryos. The top panel shows merged signal between ChARM (green) and DNA labeled by DAPI (gray). The bottom panel shows only ChARM channel with the nuclear outlines (dashed line). Scale bars are 5 µm. (**B**) Scatterplot shows the relative percentage of the number of ChARM foci per nucleus and normalized signal intensity. Experimental replicates are shown with different bullet styles. Red dots and bars indicate the mean and 95% confidence interval of the experimental condition with all replicates merged. Number of nuclei analyzed in the replicates: control: n = 24, 13; *Nr5a2* knockdown: n = 9, 12; Esrrb knockdown: n = 14, 9. (**C**) MA plot showing differentially accessible regions (DAR) analysis between SR1848-treated vs. DMSO (control) 2-cell embryos (FDR<10%). Loss DAR are shown in blue. (**D**) Bar chart representing the ratios of DAR loss and non-DAR that overlap with Nr5a2 peaks from CUT&Tag. (**E**) Aggregation plots and heat map comparing accessibility (omni-ATAC, DMSO (control) and SR1848) and Nr5a2 signal at regions that lost and showed no change in accessibility upon chemical inhibitor treatment. (**F**) Aggregation plot comparing accessibility at different classes of CREs with and without Nr5a2 peaks.

Figure 6. Nr5a2 and Esrrb specifically recognize nucleosomal target DNA

(**A**) DNA binding measured by fluorescence polarization for Nr5a2 (left) and Esrrb (right) using DNA with non-specific (black lines), Nr5a2 motif (red lines) and Esrrb motif (blue lines). The average values of three independent experiments are shown with the SD values. (**B**) Schematic illustration of SeEN-seq. Nucleosome libraries were reconstituted with TF motif (green) tiled in 601 DNA. TF-bound and unbound fractions were recovered and sequenced for revealing position-specific enrichments. Star indicates a sequence of specific enrichment as an example. (**C**) SeEN-seq enrichment profiles of Nr5a2 and Esrrb. The enrichments (log_2) were plotted against each SHLs (from SHL -6.5 to SHL +6.5). The average values of two independent experiments are shown with the SD values. (**D**) Left panel shows a schematic of SHL positions on nucleosome (PDB ID: 1KX5). Right panel shows the location where Nr5a2 (red) or Esrrb (blue) motif is inserted on 601 DNA sequence. (**E and F**) Nucleosome binding assays with Nr5a2 (E) or Esrrb (F). Left panel shows representative data of EMSA with the 601 template nucleosome and SHL-6 nucleosome. Right panel shows graphical representation. The average values of three independent experiments are shown with the SD values. (**G and H**) Competition assays with Nr5a2 (G) and Esrrb (H). Nr5a2 or Esrrb (0.5 μM) was incubated with SHL-6 nucleosome containing their own motifs (50 nM) in the presence of 5-, 10-, 20- and 40-fold molar excess of specific competitor DNA ("s" lanes) or non-specific DNA ("ns" lanes). Asterisk indicates the competitor DNA-bound complex. Quantification of the results shown in the right panel. The average values of three independent experiments are shown with the SD values. (**I)** Model of Nr5a2-dependent ZGA

regulation in mouse embryos. **(J)** Model of Nr5a2's distinct regulation of totipotency and pluripotency networks during mammalian development.

5