Maternal *Plag1* deficiency delays zygotic genome activation and two-cell stage embryo development

Running title *Plag1* controls genome activation

Elo Madissoon[1§], Anastasios Damdimopoulos[2§], Shintaro Katayama[1§], Kaarel Krjutskov[3,4], Elisabet Einarsdottir[1,4], Katariina Mamia[5], Bert De Groef[6], Outi Hovatta[5], Juha Kere[1,7,8*], Pauliina Damdimopoulou[1,5,9*]

[1] Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden

[2] Bioinformatics and Expression Analysis core facility, Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden.

[3] Competence Centre on Health Technologies, Tartu, Estonia.

[4] Molecular Neurology Research Program, University of Helsinki and Folkhälsan Institute of Genetics, Helsinki, Finland

[5] Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden

[6] Department of Physiology, Anatomy and Microbiology, La Trobe University, Bundoora, Victoria, Australia

[7] Research Programs Unit, Molecular Neurology, University of Helsinki, and Folkhälsan Institute of Genetics, Helsinki, Finland.

[8] School of Basic and Medical Biosciences, King's College London, Guy's Hospital, London, UK

[9] Swetox, Karolinska Institutet, Unit of Toxicological Sciences, Södertälje, Sweden

[§] Equal contribution

*Corresponding authors: pauliina.damdimopoulou@ki.se and juha.kere@ki.se

Keywords: pleomorphic adenoma gene 1; preimplantation embryo; cleavage-stage embryo; zygotic genome activation

Abbreviations

2c, 4c, 8c	2-cell, 4-cell, 8-cell
BMA	Best match average
DEG	Differentially expressed gene
dpc	Days post coitus
GO	Gene ontology
HET	Heterozygous
hpf	Hours post fertilization
KO	Knockout
matPlag1KO	Maternal *Plag1* knockout
PLAG1	Pleomorphic adenoma gene 1
RNA-seq	RNA sequencing
SINE	Short interspersed nuclear element
WT	Wildtype
ZGA	Zygotic genome activation

Abstract

Pleomorphic adenoma gene 1 (*PLAG1*) is a transcription factor that is involved in cancer and growth. We discovered a *de novo* DNA motif containing a PLAG1 binding site, located within the promoters of genes activated during zygotic genome activation (ZGA) in human embryos. This site frequently overlapped with Alu elements and was conserved in mouse B1 elements. We show that *Plag1* is essential for timely preimplantation embryo development. Mouse oocytes lacking maternally loaded *Plag1* gave rise to embryos that spent significantly longer time in 2-cell stage, displayed delayed regulation of 1,089 genes and in contrast to wildtype, expressed *Plag1* from the paternal allele. The PLAG1 motif was enriched in promoters of mouse delayed-activation genes that also showed a significant overlap with human ZGA genes. By ontology, these mouse and human genes were connected to ribosome biogenesis and protein synthesis. Our data suggests that *PLAG1/Plag1* affects ZGA through a conserved DNA motif within retrotransposons influencing ribosomes and protein synthesis, a mechanism that might also explain its roles in cancer and growth.

Introduction

Early preimplantation embryo development is dependent on a number of processes, including chromatin remodeling, maternal RNA degradation and zygotic genome activation (ZGA) (Jukam et al., 2017; Niakan et al., 2012). Transcription from the newly formed zygotic genome starts gradually and a major increase in transcriptional output, also known as the major ZGA, takes place during the zygote to 2-cell (2c) transition in mice and 4c to 8c transition in humans (Jukam et al., 2017; Niakan et al., 2012). Successful ZGA is a prerequisite for the formation of a totipotent embryo with the capacity to grow, develop to all embryonic and extra-embryonic tissues, and implant into the receptive uterine endometrium. Knowledge about gene expression programs during the first steps of embryonic development is essential for understanding totipotency, lineage differentiation, and infertility, as well as for regenerative medicine. This has prompted many laboratories, including ours, to map transcriptional programs during preimplantation development (De Iaco et al., 2017; Petropoulos et al., 2016; Tohonen et al., 2015; Vassena et al., 2011; Xue et al., 2013; Yan et al., 2013).

In a previous study, we used an RNA sequencing (RNA-seq) method based on the detection of the 5' ends of transcripts to study gene expression during the first three days of human preimplantation development. Our data revealed 129 upregulated transcription start sites (or transcript far 5'-ends, TFEs) during human major ZGA and *de novo* motif analyses of the associated promoters led to our discovery of PRD-like homeodomain transcription factors as regulators of ZGA in humans (Jouhilahti et al., 2016; Madissoon et al., 2016; Tohonen et al., 2015). Several additional significant *de novo* motifs were discovered, and some of them harbored known transcription factor binding sites, including that of pleomorphic adenoma gene 1 (PLAG1).

PLAG1 encodes a C2H2 zinc finger transcription factor that was first characterized in 1997 through positional cloning in pleomorphic adenomas of the salivary gland (Kas et al., 1997).

It belongs to the same protein family as the imprinted tumor suppressor PLAG-like 1 (*PLAGL1*) and the oncogene PLAG-like 2 (*PLAGL2*). *PLAG1* and *PLAGL2* have overlapping functions; they can bind to the same DNA elements, induce insulin-like growth factor (*IGF2*) expression, and stimulate mitogenic signaling and tumor growth in mice (Hensen et al., 2002). Ectopic expression of *PLAG1* and *PLAGL2* resulting from chromosomal translocation events can be found in malignant tumors (Juma et al., 2016). Associations with cancer have been the focus of most studies on *PLAG1* family transcription factors and consequently, less is known about their role in normal physiology. There are no reports on *PLAG* family genes in preimplantation embryo development or pluripotency.

Plag1 knockout (KO) mice are growth restricted and subfertile; *Plag1KO* males have impaired spermatogenesis (Juma et al., 2017) and *Plag1KO* females give birth to small litters (Hensen et al., 2004). The poor breeding success of the *Plag1KO* mice combined with our discovery of PLAG1 binding sites in the promoters of human ZGA genes prompted us to study the reproductive phenotype of *Plag1KO* animals. Here, we describe the role of *Plag1* in the murine female reproductive tract and preimplantation embryos. Our results show that *Plag1* is dispensable in normal ovarian and uterine function but essential for ZGA.

Results

A de novo *motif containing a PLAG1 binding site is enriched in the promoters of human ZGA genes*

Analysis of the 129 promoters upregulated during major ZGA in human embryos revealed 13 significant *de novo* DNA motifs (Tohonen et al., 2015), including a 31 bp element harboring a known PLAG1 binding site (MA0163.1 in JASPAR, p=0.0076 by TomTom) (Fig. 1a). This PLAG1 *de novo* motif was found in 74 of the 129 promoters upregulated during ZGA ($E=2.6\times10^{-646}$ by MEME; File S1) and similar sites were found in 93 altogether ($p<0.0001$ by MAST; Fig. 1b & File S1). On average, each promoter contained 4.3 motifs and they were preferably located approximately 1,250 bp upstream of the transcription start site (Fig. 1b). In addition, the PLAG1 *de novo* motif frequently overlapped with Alu elements in a manner similar to the PRD-like homeodomain transcription factor motif we discovered in our previous study (Tohonen et al., 2015) (Fig. 1b, c & File S1). In fact, these two motifs are both located within a conserved region of Alu elements in close vicinity to an RNA polymerase III promoter element, A-box (Ludwig et al., 2005) (Fig. 1c). Alu elements are primate-specific retrotransposable short interspersed nuclear elements (SINE) that emerged from a duplication of the 7SL RNA gene (Ullu and Tschudi, 1984) and their evolutionary counterparts in rodents are the B1 elements (Labuda et al., 1991). Interestingly, the segment of Alu elements containing the PRD and PLAG1 *de novo* motifs was well conserved in rodent B1 elements and the putative PLAG1/Plag1 binding within the elements was nearly identical to the consensus PLAG1 binding motif (Voz et al., 2000) (Fig 1c). Human PLAG1 and mouse Plag1 proteins are 94% identical by sequence and their DNA binding domains are identical (Text S1). Collectively, these data suggested that PLAG1/Plag1 could be involved in mouse and human ZGA through binding to Alu and B1 elements. If true, disruption of PLAG1/Plag1 could lead to reproductive failure.

Plag1 *knockout affects growth and reproduction in mice*

126 To study the role of *Plag1* in fertility, we obtained *Plag1* knockout (KO) mice (Hensen et al.,
127 2004). Pups from heterozygous intercrosses (HET×HET) were genotyped to study the effect
128 of *Plag1* deficiency on embryonic lethality. Genotype ratios from 16 litters encompassing 164
129 pups did not significantly deviate from the expected Mendelian distribution, although 43%
130 fewer KO pups were born (n=27) than expected (n=47) (Fig. 2a). Weights recorded at
131 weaning (3 weeks of age) confirmed growth retardation in KO animals: *Plag1KO* females
132 were 40% smaller and *Plag1KO* males 41% smaller than WT pups (Fig 2b) (Hensen et al.,
133 2004). Compared to HET intercrosses, breeding pairs with a *Plag1KO* female and a
134 *Plag1HET* male produced significantly fewer pups per litter (Fig. 2c). When we reversed
135 parental genotypes (*Plag1KO* males with *Plag1HET* females), litter size was not affected
136 (Fig. 2c). Homozygous KO×KO intercrosses produced the smallest litters (Fig. 2c). The
137 significant reduction in litter size of *Plag1KO* mothers was seen as early as 7.5–8.5 days post
138 coitum (dpc), when fewer implantation marks were observed in the uterus compared to the
139 *Plag1HET* and WT mothers (Fig. 2d, e).

140 The number of litters per month over a three-month continuous breeding period did not differ
141 between HET×HET crosses and pairs consisting of a *Plag1KO* female and a *Plag1HET* male
142 (Fig. 2f). When we reversed the parental genotypes and crossed *Plag1KO* males with
143 *Plag1HET* females, two of the three pairs did not manage to maintain the approximate one-
144 litter-per-month rate (Fig. 2f). When we intercrossed homozygous *Plag1KO* mice, litter
145 frequency was significantly reduced compared to *Plag1HET* intercrosses: the three KO×KO
146 breeding pairs produced only two litters during the entire three-month test period (Fig. 2f).

147 In summary, *Plag1KO* females produced small litters with normal frequency; *Plag1KO* males
148 had normal litter size but irregular frequency, and KOxKO pairs produced very small and very
149 infrequent litters.

150 **Plag1 *deficiency does not affect the ovaries or uterus***

To study whether the reduced litter size of *Plag1KO* females could depend on impaired ovarian function, we carried out superovulation experiments and dissected the ovaries for histological evaluation. Oocyte yield and ovarian weights did not differ between WT and *Plag1KO* females (Fig. 3a,b). The general histology of *Plag1KO* ovaries was similar to WT: the number of follicles in different stages and *corpora lutea* did not differ between the genotypes when normalized by cross section surface area (Fig. 3c,d). The *Plag1KO* uteri were significantly smaller than WT uteri, but all normal uterine structures were present (Fig. 3e,f). When the weights were corrected for body weight the significance between genotypes was lost (p=0.07 by Student's t-test). To investigate possible differences at the transcriptomic level, we performed RNA-seq from a piece of uterine horn, plus mucosa scraped from the endometrial surface, and from outer muscle layer myometrium (Fig. S1). Principal component analysis separated the samples by tissue type (uterus, mucosa, myometrium) but not by genotype (Fig. 3g). Taken together, these data show that *Plag1KO* and WT ovaries and uteri were not significantly different.

Maternal* Plag1 *deficiency delays 2-cell stage embryo development

We used embryos derived from *Plag1KO* females crossed with WT males to study the effect of maternal *Plag1* deficiency on embryonic development. These breeding pairs produced *Plag1HET* embryos that lack the maternal *Plag1* allele, and will hereafter be referred to as *matPlag1KO* embryos. We observed preimplantation development of WT and *matPlag1KO* embryos by time-lapse microscopy from zygote up to late morula or early blastocyst stages (Movie 1 and 2). We recorded the time each embryo spent in different developmental stages, and discovered that *matPlag1KO* embryos spent significantly longer time at the 2c stage compared to WT embryos (Fig. 4). The time spent at the zygote, 2c and 4c stages was 34.1±0.5, 26.1±0.4 and 14.8±0.3 h (mean±SEM) in the *matPlag1KO* embryos compared to 33.1±0.3, 23.1±0.6 and 14.3±0.4 h in the WT, respectively. When 50% of WT embryos had processed to the 4c stage, all *matPlag1KO* embryos were still arrested at stage 2c (Fig 4, middle panel). When the *matPlag1KO* embryos proceeded beyond the 2c stage, normal

178 developmental pace was regained and the time spent at the 4c stage was not significantly
179 different from WT (Fig. 4). The survival of the embryos (zygote to morula) did not differ
180 between genotypes (73% in *matPlag1KO*, 60% in WT). In conclusion, maternal *Plag1*
181 deficiency caused a delay in 2c stage embryo development without compromising embryo
182 viability or further preimplantation development.

183 ***Two-cell stage* matPlag1KO *embryos exhibit delayed transcriptional program and***
184 ***express Plag1 from paternal allele***

185 In order to better understand the effect of maternal *Plag1* deficiency on embryo development,
186 we collected MII oocytes (21–23 h post chorionic gonadotropin treatment), 2c embryos
187 (45–47 h post fertilization, hpf) and 8c embryos (71–73 hpf) for single embryo RNA-seq.
188 Altogether, we sequenced and analyzed 45 WT and 45 *matPlag1KO* oocytes and embryos
189 collected at three separate time points (Table S1). We used spike-in RNA for data
190 normalization to correct for the large general changes in cellular RNA content during these
191 developmental stages (Tohonen et al., 2015). The size of the spike-in reference RNA library
192 recovered by RNA-seq did not vary between different genotypes or developmental stages,
193 showing the robustness of the normalization strategy (Fig. S2). In contrast, the spike-in
194 normalized embryo RNA-seq libraries showed a dramatic drop in embryonic poly(A) mRNA
195 content from oocyte to 2c stage and a subsequent increase during 2c–8c transition (Fig. 5a).
196 There were no significant differences in general mRNA amount between *matPlag1KO* and
197 WT embryos at any stage (Fig. 5a). Analysis of differentially expressed genes (DEGs)
198 showed that the majority of DEGs were downregulated in both genotypes from oocyte to 2c
199 stage when maternal RNA degradation is known to take place. In that transition, there were
200 6,832 and 6,044 DEGs downregulated in WT and *matPlag1KO* compared to 2,910 and 2,851
201 being upregulated, respectively (Fig. 5b). More DEGs were upregulated in the 2c–8c
202 transition, 3,636 in WT and 3,586 in *matPlag1KO*, when the zygotic genome becomes fully
203 active (Fig. 5b).

We then compared WT and *matPlag1KO* embryos, and found that they differed most at the 2c stage. The 2c *matPlag1KO* embryos had 530 downregulated and 559 upregulated DEGs compared to 2c WT embryos (Fig. 5c). A heat map clustered the embryos into three primary groups by developmental stage (MII oocyte, 2c, 8c) but at the 2c stage, the embryos also clustered by genotype (Fig. 5d). Principal component analysis likewise separated the embryos primarily by developmental stage, but also by genotype at the 2c stage (Fig. 5e). Finally, cell trajectory (pseudotime) analysis yielded a similar pattern, and further showed that the development of the *matPlag1KO* embryos transcriptionally lagged behind WT at the 2c stage (Fig. 5f).

Next, we studied the expression pattern of the DEGs in *matPlag1KO* 2c stage embryos by plotting their mean expression levels from oocyte to 8c stage. We found that the DEGs upregulated in *matPlag1KO* embryos as compared to WT (n=559) were actually maternally loaded transcripts whose degradation was delayed in the *matPlag1KO* embryos as compared to WT embryos (Fig. 6a). Similarly, the DEGs downregulated in *matPlag1KO* embryos at the 2c stage (n=530) were in fact genes whose upregulation was delayed in the *matPlag1KO* embryos as compared to WT (Fig. 6a). We hereafter refer to these two sets of genes as “delayed-degradation” and “delayed-activation” genes. Interestingly, the expression level of the delayed-activation and delayed-degradation genes in *matPlag1KO* reached WT levels by the 8c stage, suggesting that the transcriptional dysregulation caused by the lack of maternal *Plag1* allele was temporary (Fig. 6a).

We next considered the expression profile of *Plag1* itself in the embryos. *Plag1* transcripts were present in WT oocytes, showing that *Plag1* is a maternally loaded factor (Fig. 6b). We also stained ovarian tissue sections and observed positive staining in secondary follicles within the oocyte nucleoli, confirming expression in growing oocytes (Fig. S3). In WT embryos, the maternally loaded *Plag1* transcripts were degraded by the 2c stage, and there was no further expression at the 8c stage, suggesting that *Plag1* is not normally expressed from the zygotic genome this early (Fig. 6c). A similar expression pattern was confirmed in

two independent embryo RNA-seq datasets in human, where *PLAG1* transcripts were present from oocyte to 4c stage, when ZGA takes place, and then downregulated (Fig. S4).

As expected, *Plag1* transcripts were not detected in *matPlag1KO* oocytes. Surprisingly, they were present in the 2c stage *matPlag1KO* embryos (Fig. 6b). As there is no maternal *Plag1* allele in these embryos, the expression must stem from the paternal allele. Expression of paternal *Plag1* returned to baseline by the 8c stage (Fig. 6b). The expression levels of the other *Plag1* family members *Plagl1* and *Plagl2* were unaffected in *matPlag1KO* embryos as compared with WT (Fig. S5).

We next carried out gene ontology (GO) analysis on the delayed-activation and delayed-degradation gene sets to understand their functions. The delayed-activation genes clustered into categories relevant for ribosome biogenesis, RNA processing, and translation (Fig. 6c, File S2). The delayed-degradation genes showed less clear clustering. The largest clusters represented diverse GO categories, such as neurogenesis, cell communication, developmental and morphogenesis events (Fig 6c, File S2). We then compared these GO categories to those associated with normal mouse embryo development by GO-annotating the genes significantly up- and downregulated during 2c–8c transition in WT (File S3). The delayed-activation gene GOs showed higher semantic similarity with genes typically upregulated during 2c–8c transition in the WT [best match average (BMA) 0.792] than with those downregulated (BMA 0.531), while the delayed-degradation gene GOs showed higher similarity to genes normally downregulated between 2c–8c stages (BMA 0.714) than to those upregulated (BMA 0.445). These data confirmed that embryos that lack the maternal *Plag1* allele were characterized by delayed regulation of both maternal mRNA degradation and ZGA compared to WT embryos.

Mouse delayed-activation genes overlap with human ZGA genes and show enrichment for the* PLAG1 de novo *motif

In order to compare our earlier human embryo RNA-seq dataset (Tohonen et al., 2015) to our current mouse data, we converted the mouse and human genes to orthologous gene as well as to proteins, and used both gene and protein family-based approaches in subsequent comparisons. First, we compared gene expression changes during major ZGA between humans (4c–8c transition) and mice (2c–8c transition) in general using gene set analysis, and found highly significant overall similarities (Fig. 7a). We then compared the human ZGA genes with the genes expressed in mouse embryos at the 2c stage, and discovered that they were significantly enriched among the delayed-activation genes (Fig. 7b). We obtained the same result when using the more conservative χ^2 test that is based on significant DEGs only (whereas gene set analyses are based on ranked expression of all detected genes) (Fig. 7c). DEGs upregulated during human major ZGA showed a significant overlap with the *matPlag1KO* delayed-activation genes but not with delayed-degradation genes (Fig. 7c). Comparison of protein families yielded similar results (Fig. 7d).

We next studied the promoters of the delayed-activation genes. When enrichment of the database PLAG1 binding site (MA0163.1 in JASPAR) was studied, no significant differences were found between delayed-activation, delayed-degradation and not regulated genes (Fig. 7e). Interestingly, when we used our PLAG1 *de novo* motif instead, significant enrichment was discovered in promoters of delayed-activation genes as compared to genes whose regulation was not affected by maternal lack of *Plag1* (Fig. 7e). Altogether, 132 delayed-activation genes had at least one PLAG1 *de novo* motif within –2,000 to +500 bp of their transcription start site (File S4). We then compared these mouse genes with those 94 human ZGA genes that contained a PLAG1 *de novo* motif (Fig 1b, File S1). Gene set analysis revealed that there was a significant overlap between the mouse and human genes, suggesting that ZGA genes containing a PLAG1 *de novo* motif are conserved between mice and humans (Fig 7f).

We finally wanted to know the functions of these conserved mouse and human genes. We GO-annotated them and compared the semantic similarity between the top 100 GO

categories associated to each gene set. Hierarchical clustering revealed that human ZGA and mouse delayed-activation genes clustered together to categories representing ribosomes, protein transport, translation, RNA processing, and protein metabolism (Fig 7g, File S5). Conversely, the GOs associated to mouse delayed-degradation genes clustered to groups representing morphogenesis, p53 signalling, and immune responses and showed little overlap with the human ZGA clusters (Fig 7g, File S5).

Collectively, our findings show that there is a conserved sets of genes harboring *PLAG1* binding motifs in their promoters that are activated during ZGA in mouse and human. Our data further suggests that these genes play roles in central cellular processes that relate to ribosomes, RNA and protein metabolism.

Discussion

In the present study, we have discovered a new role for oncogene *PLAG1* as a regulator of ZGA. We identified a motif containing a PLAG1 binding site among the promoters upregulated during ZGA in human embryos, and showed that the lack of *Plag1* leads to an incomplete ZGA in mice on a transcriptional level with consequences for the timing of cleavage stage development.

Our studies on *Plag1KO* mice confirmed many of the earlier observations, including growth restriction and subfertility (Hensen et al., 2004). *Plag1KO* females produced small litters regardless of the paternal genotype, implying that this phenotype was caused by the lack of *Plag1* allele in the mother. Still, we could not find any significant differences in histology or function of reproductive organs between *Plag1KO* and WT females. Their ovaries contained normal numbers of follicles at all developmental stages, an equivalent number of oocytes were ovulated, and their litter frequency was not affected. *Plag1KO* uteri had normal histology and did not transcriptionally differ from WT. We note that the *Plag1KO* uteri remained slightly smaller than WT even after body weight adjustment, which could perhaps

impose spatial constrains to implantation. However, “uterine crowding” has not been found to affect fetal viability in rabbits or mice (Argente et al., 2017; Bruce and Wellstead, 1992), suggesting that smaller uterine size does not necessarily reduce implantation rate or compromise post-implantation survival. We conclude that *Plag1* is not required for normal ovarian or uterine function. This suggests that embryos derived from *Plag1KO* females, regardless of paternal genotype, do not have the same developmental potential as WT embryos do, and are even different from HET embryos that have inherited the mutant allele from the father. In order to focus on the maternal effect of *Plag1*, we compared *Plag1KO* oocytes fertilized by WT studs (*matPlag1KO* embryos) with WT embryos.

We discovered that *matPlag1KO* embryos spent significantly more time in the 2c stage compared to WT embryos. Interestingly, major ZGA takes place in the 2c stage of mouse development, and failure to activate the zygotic genome leads to a developmental arrest. For example, knockdown of the pluripotency factor *Lin28* in mouse embryos leads to arrest at the 2c–4c stage (Vogt et al., 2012), and KO of the maternal-effect gene *Mater* to arrested embryonic development at the 2c stage (Tong et al., 2000). ZGA is accomplished with the help of maternally loaded factors, *i.e.,* transcripts and proteins that are deposited into the oocyte. We showed that *Plag1* was a maternally loaded factor that could be detected in nucleoli of oocytes within growing follicles by staining, and in mature mouse and human oocytes by RNA-seq. *MatPlag1KO* embryos lacked maternally loaded *Plag1* and were delayed at 2c stage development suggesting that *Plag1* is required for normal cleavage-stage embryo development.

Despite lacking maternal *Plag1*, *matPlag1KO* embryos developed to the blastocyst stage with equal efficiency as WT embryos. Interestingly, *matPlag1KO* embryos started expressing *Plag1* from the paternal allele, which WT embryos did not do, and this may have rescued their development. Although these data argue that *Plag1* is essential for embryo development, we noted that even homozygous *Plag1KO* intercrosses occasionally produced litters. We hypothesize that maternally loaded *Plagl2*, the *Plag1* family member with

redundant activity to *Plag1,* might rescue embryo development in some cases. Testing the hypothesis would require generation of *Plag1/PlagL2* double KO mice, which is impossible due to the severe phenotype of *PlagL2* mice. *Plagl2KO* pups die shortly after birth to starvation due to their inability to absorb chylomicrons (Van Dyck et al., 2007).

Although *matPlag1KO* embryos regained normal developmental pace after exiting the 2c stage, their developmental success was not the same as that of WT embryos. Interestingly, human embryos can be scored after *in vitro* fertilization through morphokinetic measurements, and the time an embryo spends in the 2c and 4c stages is a significant determinant of developmental potential (Wong et al., 2010). The disrupted regulation of over 1,000 genes at the 2c stage in *matPlag1KO* embryos, when not only ZGA but also epigenetic changes take place, could have permanent effects on the competence of the blastocysts. Moreover, synchronous preparation of both the embryo and endometrium for implantation is a prerequisite for successful implantation during the window of receptivity. Therefore, a simple delay in preimplantation development could lead to some blastocysts missing this critical window (Wang and Dey, 2006).

The PLAG1 *de novo* motif in humans frequently localized within Alu elements in the promoters of the ZGA genes. Alu elements are transposable elements ubiquitously present in primate genomes with involvement in gene regulation by various mechanisms (Elbarbary et al., 2016). The counterparts of Alu elements in the mouse genome are B1 elements (Labuda et al., 1991). Mouse B1 and primate Alu have split from the common ancestor 7SLRNA over 80 million years ago and have evolved and retrotransposed independently ever since (Ullu and Tschudi, 1984). Despite independent evolution, the densities of these elements in promoters of orthologous genes in humans and mice are surprisingly correlated (Mouse Genome Sequencing et al., 2002; Polak and Domany, 2006), and genes containing Alu and B1 elements belong to highly correlated functional categories (Tsirigos and Rigoutsos, 2009). Both Alu and B1 elements are highly enriched in promoters that are activated during ZGA, and the number of copies in the promoter correlates with the level of

gene upregulation (Ge, 2017). It is interesting that the component of the Alu element containing the PLAG1 binding site was well-conserved in the mouse B1 element. Collectively, these findings suggested that the SINEs Alu and B1 play a role in ZGA in humans and mice by attracting transcription factors such as PLAG1 to gene promoters. Our data further showed that the database PLAG1 binding site alone did not show any enrichment in the promoters of delayed-activation genes. This suggested that PLAG1 works in concert with other transcription factors and co-regulators and therefore the short binding site does not accurately reflect the DNA sequence required for PLAG1-stimulated gene expression. This is a largely unexplored area that should be a focus of follow-up studies.

Functional annotation categories associated to delayed-activation genes in the *matPlag1KO* embryos were mainly related to ribosome biogenesis, maturation and function. Even when we restricted the analysis to only those genes containing a PLAG1 *de novo* motifs, GO-analyses suggested roles in ribosome biogenesis, RNA and protein metabolism. Importantly, human ZGA genes with PLAG1 *de novo* motif had a very similar GO profile. Alu elements have been suggested to function as enhancer elements that regulate gene transcription (Su et al., 2014), and Alu elements are enriched in promoters associated to ribosome biogenesis, protein biosynthesis and RNA metabolism (Polak and Domany, 2006) as well as to ZGA (Ge, 2017). It is a plausible hypothesis that Alu elements in the promoters of ribosomal function genes attract PLAG1 that then regulates the associated gene. Many oncogenes have effects on ribosome biogenesis, which enables cancerous cells to increase protein synthesis and grow rapidly (Pelletier et al., 2018). Although *PLAG1* is an oncogene, its potential role in ribosome biogenesis and protein synthesis has not been evaluated. Interestingly, one of the most striking phenotypes of the *Plag1KO* mice is their small size (Hensen et al., 2004). In addition, *PLAG1* polymorphisms associate with body size and growth in farm animals and humans (Fink et al., 2017; Rubin et al., 2012; Utsunomiya et al., 2013; Zhang et al., 2016). Based on our data, we hypothesize that *PLAG1/Plag1* associated growth effects could result from modulation of protein synthesis that affects cell size and division rate.

We conclude that the lack of maternally loaded *Plag1* leads to delayed ZGA and 2c stage development, and reduced embryo competence for implantation. The effect on ZGA genes arose through retrotransposition of Alu and B1 elements that harbour conserved PLAG1 binding sites. The genes affected by the lack of PLAG1 have roles in ribosome biogenesis, RNA and protein metabolism. Follow-up studies should focus on Alu and B1 elements in preimplantation development and a deeper analysis of PLAG1 involvement in protein synthesis, as this is a mechanism that would explain many of the reported biological activities of PLAG1, including tumorigenicity, cell proliferation, and growth.

Materials and Methods

Mouse colony and handling

All experiments were approved by the Swedish Board of Agriculture (#S5-14) and performed in accordance to the ethical licence. *Plag1KO* mice (Hensen et al., 2004), backcrossed to the CD1 strain, were a kind gift from Prof. Wim Van de Ven (University of Leuven, Belgium) and Dr. Carol Schuurmans (University of Calgary, Canada). The local colony at Karolinska Institutet was established through embryo transfers. HET animals were kept in continuous breeding, and the litters were earmarked, weaned and weighed at the age of 3–4 weeks. Ear punches were used for genotyping with primers Plag1MT_F (5'-CAGTTCCCAGGTGTCCAACAAG-3'), Plag1MT_R (5'-AATGTGAGCGAGTAACAACCCG-3'), Plag1WT_F (5'- CGGAAAGACCATCTGAAGAATCAC -3'), and Plag1WT_R (5'-CGTTCGCAGTGCTCACATTG -3'). The animals were housed in standard conditions (19–21°C, 55% humidity, lights 6:00 am–6:00 pm) with free access to feed (irradiated Global 18% diet 2918; Envigo, Huntingdon, UK) and tap water. In all experiments, animals were sacrificed by cervical dislocation.

Superovulation and embryo imaging

Sexually mature 1-3 months old *Plag1KO* and WT females were superovulated by i.p. injections of 5 IU pregnant mare serum (Folligon; Intervet, Dublin, Ireland), followed two days later by 5 IU human chorionic gonadotropin (hCG) (Chorulon; Intervet, Dublin, Ireland), and mated with trained WT studs. MII oocytes (no mating) and zygotes were collected from oviducts the following morning and their numbers recorded. Cumulus cells were removed with hyaluronidase (0.3 mg/ml, Sigma-Aldrich, St.Louis, MO, USA) treatment. For imaging, zygotes were placed 2–4 per well into Primo Vision embryo culture dishes (Vitrolife, Göteborg, Sweden) under a Nikon Ti-E spinning disk wide-field microscope with a live-cell imaging incubator. The microscope was programmed to take bright-field images every 30 min with an Andor EM-CCD camera using the 20× objective for a total of 90 h 30 min. The imaging was repeated three times with embryos from a total of 5 *Plag1KO* (providing 103

426 embryos) and 6 WT (providing 89 embryos) females with both genotypes present at every
427 session. Cleavage events until the 8c stage were scored manually from the images by a
428 researcher blind to the genotypes.

429 ***Histological assessment of ovaries and uteri***

430 Ovaries and uteri were collected during superovulation experiments, their weights recorded,
431 and tissues stored in 4% (w/v) paraformaldehyde. Fixative was changed to 70% ethanol,
432 tissue embedded in paraffin and processed to 4-µm hematoxylin-eosin stained sections, and
433 the slides were digitalized with Mirax Slide Scanner (Zeiss, Göttingen, Germany). Transverse
434 sections from the middle of the uterine horn were used for histological examination. One
435 section from the middle of the ovary was used for assessment of follicles in different
436 developmental categories (preantral, antral, atretic and corpora lutea) and their number
437 adjusted for the ovary surface area using Pannoramic Viewer software (3DHistech,
438 Budapest, Hungary). Altogether, 22 *Plag1KO* and 13 WT animals were used for ovary
439 assessment, and 8 *Plag1KO* and 8 WT for uterus histology.

440 ***X-gal staining***

441 Ovaries were collected from 3 *Plag1KO* and 3 WT females, aged 5 months, fixed,
442 cryosectioned and subjected to X-gal staining and Nuclear Fast Red counterstaining as
443 described before (Juma et al., 2017).

444 ***Implantation study***

445 *Plag1KO* and WT females, aged 2-4 months, were mated with trained WT studs. The uteri
446 were collected 7.5–8.5 dpc and the number of implantation sites counted. The experiment
447 was carried out twice, with a total of 8 *Plag1KO* and 8 WT females.

448 ***Single-embryo RNA-seq***

449 Zygotes collected from superovulated females (see “superovulation and embryo imaging”
450 above) were treated with hyaluronidase and placed in KSOM medium (EMD Millipore,
451 Nottingham, UK) under ovoil-100 (Vitrolife) in IVF 4-well plates (Sigma-Aldrich) for culture at
452 37°C under 5% CO_2/95% air. Single embryos or uncultured MII oocytes were picked into 4 µl

library lysis buffer containing 5 mM Tris-HCl pH 8.0 (Sigma-Aldrich), 2 mM dNTP mixture (ThermoFisher Scientific, Waltham, MA, USA), 10 mM DTT (Sigma-Aldrich), 0.05% Triton X-100 (Sigma-Aldrich), 400 nM anchored oligo(dT) primer biotin-TTAAGCAGTGGTATCAACGCAGAGTCGAC$(T)_{29}$V where V is LNA nucleotide (Exiqon, Vedbaek, Denmark), and 4 U RiboLock RNase inhibitor (ThermoFisher). Embryos were picked 21–23 h post hCG treatment (MII oocytes, no culture), 45–47 hpf (2c stage) and 71–73 hpf (8c stage) on three different occasions, yielding a total of 16 zygotes, 14 2c stage embryos, and 15 8c stage embryos for both *matPlag1KO* and WT (Table S1). Two separate 46-plex libraries (libA and libB) were made as described previously (Krjutskov et al., 2016) with the following modifications: barcoded 10 µM template-switching oligonucleotides were added prior to reverse transcriptase, ERCC spike-in Mix A was diluted 15,300-fold with clean water, and 1 µl was taken per library reverse transcriptase master mix. Twenty cycles of PCR were used for the first round of amplification and ten additional cycles for the second round to introduce Illumina-compatible universal sequences. Both libraries contained all developmental stages and genotypes.

Uterus RNA-seq

Uterine horns from 8 *Plag1KO* and 8 WT females used in the superovulation experiments were collected in RNAlater (Ambion, Foster City, CA, USA). At dissection, uterine horns were cut in half longitudinally, and the endometrial side was gently scraped with a scalpel (Fig. S1). RNA was extracted with the RNeasy mini kit (Qiagen, Hilden, Germany) and quality measured with the Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). Ten nanograms of high-quality RNA (RIN>8) was used for RNA-seq. The RNA-seq was performed according to the modified STRT protocol (Krjutskov et al., 2016).

RNA-seq data analysis

The analysis was performed as described previously (Krjutskov et al., 2016). In short, the reads were filtered, samples de-multiplexed, UMI-s joined, reads trimmed and mapped to the reference mouse genome mm9 by TopHat (Kim et al., 2013). The resulting bam files were converted to tag directories employing Homer (Heinz et al., 2010) and were subsequently

used to estimate the reads in all annotated genes. Annotations were in GTF file format retrieved from UCSC and were concatenated to a GTF file with the ERCC annotations. Gene counts were then imported to R (R Development Core Team, 2010) and libraries with a median gene expression of log2 counts per million (cpm) under 0 were excluded from further analysis. Cell libraries, excluding the ERCC spike-in counts, were normalized with EdgeR (Robinson et al., 2010) using the TMM normalization method. The ERCC counts were used for normalization between the various embryonic cell stages by scaling the library sizes. EdgeR was also employed for the subsequent differential gene expression analysis, which was performed on genes that had 1 cpm in at least five or more samples and the rest of the genes were filtered out. After removing the low expressing/unexpressed genes, the gene counts were renormalized.

Principal component analysis was performed in R by using the genes that were significant in any of the comparisons between the genotypes. Heatmaps were plotted on TMM normalized counts exported from EdgeR and gene expression was standardized across all samples (mean=0 and sd=1). Samples and genes were clustered using hierarchical clustering in R and plotted employing the ComplexHeatmap library (Gu et al., 2016). The same gene set was used for the cell trajectory (pseudotime) analysis by the monocle package (Trapnell et al., 2014).

Gene ontology analysis was carried out in R (R Development Core Team, 2010) with the topGO library. To identify enriched gene ontology terms, the classic algorithm and Fisher statistic were used and analysis was carried out on up- and downregulated genes separately. Semantic similarity between the GO terms was calculated using the Wang algorithm in GOSemSim bioconductor package (Yu et al., 2010). The result is given as the best match average (BMA) score that ranges from 0 to 1.

Gene set enrichment analysis was conducted to test whether the mouse genes homologous to the human genes regulated between the 4c and 8c stage were also regulated in mouse development. To calculate the p-values, the geneSetTest function from the limma package (Ritchie et al., 2015) was used. The moving average of the enrichment was calculated with

509 the tricubeMovingAverage function and plotted with ggplot2. Homologene from NCBI was
510 used to convert the human genes to the homologous mouse genes. The significance in
511 overlap of the human genes with the mouse genes regulated by the *Plag1KO* at the 2c stage
512 was calculated with the Fisher test in R. Genes were also converted to protein families using
513 the bioconductor libraries for genome-wide annotation for human and mouse
514 (org.Hs.eg.db/org.Mm.eg.db). For genes that had more than one protein family annotated to
515 them, only one of the protein families was used in order not to inflate the number of
516 overlapping or non-overlapping families between the different gene groups.

517 ***Promoter analyses***

518 Human embryo promoter sequence analysis was performed as previously described
519 (Tohonen et al., 2015). In brief, the *de novo* motif was compared with known motifs by
520 TomTom (Gupta et al., 2007). We applied MEME (Bailey and Elkan, 1994) for motif analysis
521 within the upregulated promoters, and identified similar sequences with the PLAG1 motif
522 (MA0163.1 in JASPAR (Meng et al., 2005; Sandelin et al., 2004)) by MAST (Bailey and
523 Gribskov, 1998). The location of Alu elements within the promoters was based on the
524 RepeatMasker track in the UCSC Genome Browser.

525 Human and mouse SINE repetitive elements DF0000002 (AluY), DF0000051 (AluSz),
526 DF0000034 (AluJo), DF0000144 (FLAM_C), DF0000016 (7SLRNA), DF0003101 (PB1) and
527 DF0001733 (B1_Mm) were retrieved from the Dfam database (Hubley et al., 2016) and
528 aligned and highlighted according to the percent identity by JalView2.

529 Mouse embryo promoter analysis was carried out with Homer (Heinz et al., 2010). The
530 JASPAR (Sandelin et al., 2004) PLAG1 motif (MA0163.1) (Meng et al., 2005) and the *de*
531 *novo* motifs (Tohonen et al., 2015) were used, and all annotated transcriptional start sites
532 were scanned for the presence of the motifs from 2,000 bp upstream to 500 bp downstream
533 of the transcriptional start site. Enrichment was analysed using Fischer's exact test.

534 ***Statistical analysis***

Continuous data were analyzed with Student's t-test or one-way ANOVA (one categorical predictor) or two-way ANOVA (two categorical predictors) followed by Fischer LSD *post hoc* testing when necessary. Normality was tested with the Shaphiro Wilks test and homoscedasticity with Levene's test. Categorical data were analyzed with the χ^2 test. The exit time of embryos from each developmental stage was plotted as an empirical distribution function (ecdf) using ggplot2. To test the significance between the exit times of the WT and *matPlag1KO* embryos, the Kolmogorov-Smirnov test was used. All analyses were carried out using R (R Development Core Team, 2010). All p-values are two-tailed and were considered significant when $p<0.05$.

Data Availability

RNA-seq data have been deposited to Gene Expression Omnibus data repository as a SuperSeries record under the reference GSE111040.

Acknowledgements and funding

Prof Wim Van de Ven is particularly acknowledged for giving us access to the *Plag1KO* animal model, and Dr Schuurmans for sending the mice to us. Stanisław Wawrzyczek is thanked for assistance with X-gal staining and Ingegerd Fransson for mouse genotyping and library preparation. This study was performed in part at the Preclinical Laboratory (PKL) animal facility and the Live Cell Imaging unit, Department of Biosciences and Nutrition, Karolinska Institutet, Sweden, supported by grants from the Knut and Alice Wallenberg Foundation, the Swedish Research Council, the Centre for Innovative Medicine and the Jonasson donation to the School of Technology and Health, Royal Institute of Technology, Sweden. The morphological phenotype analysis (FENO) core facility and Tarja Schröder are thanked for the preparation of the histological specimens. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under projects b2010037 and snic2017-7-317. The study was supported by Knut and Alice Wallenberg Foundation (KAW2015.0096) and by the

562 Distinguished Professor Award by Karolinska Institutet to JK. Essential funding was also

563 provided by Jane & Aatos Erkko Foundation to PD.

564

Figure legends

Figure 1. A *de novo* DNA motif that harbors a known PLAG1 binding site is found in the promoters of genes upregulated during ZGA in humans.

a) Comparison of the JASPAR PLAG1 binding motif (MA0163.1, top) and the PLAG1 *de novo* motif (bottom). **b)** Location of source sites for the PLAG1 *de novo* motifs (red dots) and similar sites (yellow dots) in the promoters of human major ZGA genes. Promoters of the 93 genes containing the motifs are stacked and aligned ranging from 2,000 upstream bases to 500 downstream bases around transcription start site (TSS, dashed line). The genes are listed in File S1. Locations of AluJ, AluS and AluY elements are highlighted in grey. The curve on top of the figure illustrates the density of the PLAG1 *de novo* motif along the promoter. **c)** Comparisons of our earlier ZGA *de novo* motif (i) (Tohonen et al., 2015), the current PLAG1 *de novo* motif (ii), known PDR-like transcription factor binding site (iii) (Jolma et al., 2013) and PLAG1 JASPAR (iv) (Meng et al., 2005) and consensus (v) (Voz et al., 2000) binding sites with human and mouse short interspersed nuclear elements (SINE). The consensus motif (v) is described by IUPAC notation; R is purine (G or A), and K is keto (G or T). The internal RNA polymerase III promoter, A-box, is indicated. The *de novo* motifs and binding motifs are reverse-complemented. Sequences of AluY, AluSz, AluJo, FLAM_C, 7SLRNA, PB1 and B1_Mm were extracted from the Dfam database of repetitive DNA elements. ZGA, zygotic genome activation; TF, transcription factor

Figure 2. *Plag1* deficiency affects body weight, litter size, and breeding success.

a) Observed and expected numbers of pup genotypes in litters from heterozygous intercrosses. **b)** Body weights of female and male pups at weaning. **c)** Mean number of pups in litters from different parent genotypes. **d)** Mean number of implanted embryos on 7.5–8.5

dpc, and **e)** representative photos of uteri at dissection. Asterisks indicate implanted embryos. Scale bars are 1 cm. **f)** Frequency of litters from breeding pairs of different genotypes maintained in continuous breeding for 90 days. Every breeding pair is represented with a horizontal line and the birth of a litter is illustrated with a circle. Quantification of the data is shown on the right. The data are presented as means + SEM and the number of observations is shown inside the columns. Statistical analysis by χ^2 test (a) or one-way ANOVA followed by Fisher LSD post hoc test (b-f). * p<0.05, ** p<0.01, *** p<0.001. F, female; HET, heterozygous; KO, knockout; M, male; WT, wild type.

Figure 3. *Plag1* deficiency does not have significant effects on ovaries and uterus.

a) Number of oocytes collected after superovulation presented as violin plots. The median is marked as a red line. **b)** Absolute and relative (body weight-adjusted) weights of ovaries, means + SEM. **c)** Representative images of hematoxylin-eosin stained ovarian cross sections from WT and *Plag1KO* mice. The scale bar is 1 mm. **d)** Numbers of preantral, antral and atretic follicles as well as corpora lutea (CL) per mm^2 of ovary. Data are presented as median (red line), interquartile ranges (box), and non-outlier ranges (whisker). Outliers are depicted as dots. **e)** Representative photos and hematoxylin-eosin stained histological images of WT and *Plag1KO* uteri. The scale bar is 500 µm. **f)** Absolute and relative (body weight-adjusted) uterine weights presented as means + SEM (n=8). **g)** Principal component analysis of global gene expression in WT and *Plag1KO* uterus samples divided into “uterus” (a piece of uterine horn) (n=8), “mucosa” (cells scraped from the endometrial surface of the uterine horn) (n=8), and “myometrium” (tissue left after mucosa has been removed) (n=8). Statistical analysis by t-test (a, b, f) or two-way ANOVA (d), ***p<0.001. CL, corpus luteum; KO, knockout; PC, principal component; WT, wild type.

615 **Figure 4: *MatPlag1KO* embryos spend significantly more time in the 2c stage**
616 **compared to WT embryos.**

617 The time spent in each developmental stage for each embryo (exit time) was plotted as an
618 empirical distribution function. Plots show the cumulative proportion of embryos (y-axis, 0–1)
619 that have exited the corresponding developmental stage at each time point (x-axis, hours).
620 To test the significance between the exit times of the WT (n=53) and *matPlag1KO* (n=75)
621 embryos, the Kolmogorov-Smirnov test was used.
622

623 **Figure 5. Single-embryo RNA-seq distinguishes *matPlag1KO* embryos from WT**
624 **embryos at the 2c stage.**

625 **a)** Size of the spike-in normalized RNA-seq libraries by genotype and developmental stage.
626 The line represents the median, the box first and third quartiles, and the whisker non-outlier
627 range. Outliers are depicted as dots and defined as data points >1.5 times interquartile range
628 outside range. One sequencing reaction equals one embryo, and the number of embryos
629 (biological replicates) used for sequencing is given in the figure. **b)** Number of differentially
630 expressed genes from one developmental stage to another (horizontal arrows) in
631 *matPlag1KO* and WT embryos. **c)** Number of differentially expressed genes between
632 *matPlag1KO* and WT embryos at MII oocyte, 2c and 8c stage. **d)** Heat map, **e)** principal
633 component analysis, and **f)** pseudotime (cell trajectory) analysis based on all differentially
634 expressed genes in the libraries. 2c, 2-cell stage; 8c, 8-cell stage; DEG, differentially
635 expressed gene; KO, knockout; PC, principal component; WT, wild type.
636

637 **Figure 6. *MatPlag1KO* embryos have a delayed transcriptional program, start**
638 **expressing *Plag1* from the paternal allele at the 2c stage, and catch up with WT**
639 **embryos by the 8c stage.**

a) Mean normalized expression (Z) pattern of the 559 genes that are significantly upregulated ("delayed-degradation") and the 503 that are significantly downregulated ("delayed-activation") in *matPlag1KO* embryos compared to WT embryos at the 2c stage. **b)** Normalized expression of *Plag1* in the embryos at MII oocyte, 2c and 8c stages. **c)** Heat maps displaying semantic similarity among the top 150 significantly enriched GO terms associated to the delayed-degradation (left) and delayed-activation (right) genes. Five largest clusters are depicted (cl1–cl5), and common denominators among GO terms belonging to these clusters are shown. The full GO lists are provided in File S2. 2c, 2-cell stage; 8c, 8-cell stage; CPM, counts per million; GO, gene ontology; KO knockout; WT, wildtype.

Figure 7. *MatPlag1KO* delayed-activation genes show a significant overlap with human ZGA genes as well as enrichment of the PLAG1 *de novo* motif in their promoters.

a) Gene set enrichment analysis comparing gene expression changes from mouse 2c–8c stage (x-axis) with human 4c–8c gene expression changes (y-axis). **b)** Gene set enrichment analysis comparing genes that are expressed in mouse embryos (WT and *matPlag1KO*) at the 2c stage (x-axis) with human 4c–8c gene expression changes (y-axis). Black vertical lines shows the location of human genes among the ranked mouse genes, and curve depicts the enrichment. Red dotted line indicates "no enrichment" level. **c)** Venn diagram showing the overlap between *matPlag1KO* delayed-activation and delayed-degradation genes with the human major ZGA genes. **d)** Venn diagram showing the same overlaps based on protein families instead of genes. **e)** Presence of PLAG1 database binding sites and PLAG1 *de novo* motifs in the promoters of genes expressed in mouse 2c stage embryos. Total number of genes in different categories as well as number of unique genes with the motif are shown. Enrichment over not affected genes was analyzed with Fischer's exact test. ***p<0.001. **f)** Gene set enrichment analysis as in b) but only displaying those human and mouse genes with a PLAG1 *de novo* motif in their promoters. **g)** t-SNE plot demonstrating similarity among

666 the top 100 GO categories associated to human ZGA, mouse delayed-activation and mouse
667 delayed-degradation genes that contain PLAG1 *de novo* motifs. Eight largest clusters with
668 the most common words within the clusters are shown. 2c, 2-cell stage; 4c, 4-cell stage; 8c,
669 8-cell stage; KO, knockout; OR, odds ratio; WT, wild type.

670

671 **Movie 1: Time-lapse video of WT embryo development**


Movie 1.avi

672

673 A representative movie of WT mouse embryo development from zygote to blastocyst. The
674 video time line is indicated (0-89 h) as well as the time spent in 2c stage.

675

676 **Movie 2: Time-lapse video of *matPlag1KO* embryo development**


Movie 2.avi

677

678 A representative movie of *matPlag1KO* mouse embryo development from zygote to
679 blastocyst. The video time line is indicated (0-89 h) as well as the time spent in 2c stage.

680

References

Argente, M. J., Calle, E. W., Garcia, M. L. and Blasco, A. (2017). Correlated response in litter size components in rabbits selected for litter size variability. *J Anim Breed Genet* **134**, 505-511.

Bailey, T. L. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36.

Bailey, T. L. and Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**, 48-54.

Bruce, N. W. and Wellstead, J. R. (1992). Spacing of fetuses and local competition in strains of mice with large, medium and small litters. *J Reprod Fertil* **95**, 783-789.

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J. and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* **49**, 941-945.

Elbarbary, R. A., Lucas, B. A. and Maquat, L. E. (2016). Retrotransposons as regulators of gene expression. *Science* **351**, aac7247.

Fink, T., Tiplady, K., Lopdell, T., Johnson, T., Snell, R. G., Spelman, R. J., Davis, S. R. and Littlejohn, M. D. (2017). Functional confirmation of PLAG1 as the candidate causative gene underlying major pleiotropic effects on body weight and milk characteristics. *Sci Rep* **7**, 44793.

Ge, S. X. (2017). Exploratory bioinformatics investigation reveals importance of "junk" DNA in early embryo development. *BMC Genomics* **18**, 200.

Gu, Z., Eils, R. and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849.

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. and Noble, W. S. (2007). Quantifying similarity between motifs. *Genome Biol* **8**, R24.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589.

Hensen, K., Braem, C., Declercq, J., Van Dyck, F., Dewerchin, M., Fiette, L., Denef, C. and Van de Ven, W. J. (2004). Targeted disruption of the murine Plag1 proto-oncogene causes growth retardation and reduced fertility. *Dev Growth Differ* **46**, 459-470.

Hensen, K., Van Valckenborgh, I. C., Kas, K., Van de Ven, W. J. and Voz, M. L. (2002). The tumorigenic diversity of the three PLAG family members is associated with different DNA binding capacities. *Cancer Res* **62**, 1510-1517.

Hubley, R., Finn, R. D., Clements, J., Eddy, S. R., Jones, T. A., Bao, W., Smit, A. F. and Wheeler, T. J. (2016). The Dfam database of repetitive DNA families. *Nucleic Acids Res* **44**, D81-89.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. *Cell* **152**, 327-339.

Jouhilahti, E. M., Madissoon, E., Vesterlund, L., Tohonen, V., Krjutskov, K., Plaza Reyes, A., Petropoulos, S., Mansson, R., Linnarsson, S., Burglin, T., et al. (2016). The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation. *Development* **143**, 3459-3469.

Jukam, D., Shariati, S. A. M. and Skotheim, J. M. (2017). Zygotic Genome Activation in Vertebrates. *Dev Cell* **42**, 316-332.

Juma, A. R., Damdimopoulou, P. E., Grommen, S. V., Van de Ven, W. J. and De Groef, B. (2016). Emerging role of PLAG1 as a regulator of growth and reproduction. *J Endocrinol* **228**, R45-56.

Juma, A. R., Grommen, S. V. H., O'Bryan, M. K., O'Connor, A. E., Merriner, D. J., Hall, N. E., Doyle, S. R., Damdimopoulou, P. E., Barriga, D., Hart, A. H., et al. (2017). PLAG1 deficiency impairs spermatogenesis and sperm motility in mice. *Sci Rep* **7**, 5317.

Kas, K., Voz, M. L., Roijer, E., Astrom, A. K., Meyen, E., Stenman, G. and Van de Ven, W. J. (1997). Promoter swapping between the genes for a novel zinc finger protein and beta-catenin in pleiomorphic adenomas with t(3;8)(p21;q12) translocations. *Nat Genet* **15**, 170-174.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36.

Krjutskov, K., Katayama, S., Saare, M., Vera-Rodriguez, M., Lubenets, D., Samuel, K., Laisk-Podar, T., Teder, H., Einarsdottir, E., Salumets, A., et al. (2016). Single-cell transcriptome analysis of endometrial tissue. *Hum Reprod* **31**, 844-853.

Labuda, D., Sinnett, D., Richer, C., Deragon, J. M. and Striker, G. (1991). Evolution of mouse B1 repeats: 7SL RNA folding pattern conserved. *J Mol Evol* **32**, 405-414.

Ludwig, A., Rozhdestvensky, T. S., Kuryshev, V. Y., Schmitz, J. and Brosius, J. (2005). An unusual primate locus that attracted two independent Alu insertions and facilitates their transcription. *J Mol Biol* **350**, 200-214.

Madissoon, E., Jouhilahti, E. M., Vesterlund, L., Tohonen, V., Krjutskov, K., Petropoulous, S., Einarsdottir, E., Linnarsson, S., Lanner, F., Mansson, R., et al. (2016). Characterization and target genes of nine human PRD-like homeobox domain genes expressed exclusively in early embryos. *Sci Rep* **6**, 28995.

Meng, X., Brodsky, M. H. and Wolfe, S. A. (2005). A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat Biotechnol* **23**, 988-994.

Mouse Genome Sequencing, C., Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562.

Niakan, K. K., Han, J., Pedersen, R. A., Simon, C. and Pera, R. A. (2012). Human pre-implantation embryo development. *Development* **139**, 829-841.

Pelletier, J., Thomas, G. and Volarevic, S. (2018). Ribosome biogenesis in cancer: new players and therapeutic avenues. *Nat Rev Cancer* **18**, 51-63.

Petropoulos, S., Edsgard, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R. and Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165**, 1012-1026.

Polak, P. and Domany, E. (2006). Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* **7**, 133.

R Development Core Team (2010). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.

Rubin, C. J., Megens, H. J., Martinez Barrio, A., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, O., Jern, P., Jorgensen, C. B., et al. (2012). Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci U S A* **109**, 19529-19536.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**, D91-94.

Su, M., Han, D., Boyd-Kirkup, J., Yu, X. and Han, J. D. (2014). Evolution of Alu elements toward enhancers. *Cell Rep* **7**, 376-385.

Tohonen, V., Katayama, S., Vesterlund, L., Jouhilahti, E. M., Sheikhi, M., Madissoon, E., Filippini-Cattaneo, G., Jaconi, M., Johnsson, A., Burglin, T. R., et al. (2015). Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nat Commun* **6**, 8207.

Tong, Z. B., Gold, L., Pfeifer, K. E., Dorward, H., Lee, E., Bondy, C. A., Dean, J. and Nelson, L. M. (2000). Mater, a maternal effect gene required for early embryonic development in mice. *Nat Genet* **26**, 267-268.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. and Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386.

Tsirigos, A. and Rigoutsos, I. (2009). Alu and b1 repeats have been selectively retained in the upstream and intronic regions of genes of specific functional classes. *PLoS Comput Biol* **5**, e1000610.

Ullu, E. and Tschudi, C. (1984). Alu sequences are processed 7SL RNA genes. *Nature* **312**, 171-172.

Utsunomiya, Y. T., do Carmo, A. S., Carvalheiro, R., Neves, H. H., Matos, M. C., Zavarez, L. B., Perez O'Brien, A. M., Solkner, J., McEwan, J. C., Cole, J. B., et al. (2013). Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. *BMC Genet* **14**, 52.

Van Dyck, F., Braem, C. V., Chen, Z., Declercq, J., Deckers, R., Kim, B. M., Ito, S., Wu, M. K., Cohen, D. E., Dewerchin, M., et al. (2007). Loss of the PlagL2 transcription factor affects lacteal uptake of chylomicrons. *Cell Metab* **6**, 406-413.

Wang, H. and Dey, S. K. (2006). Roadmap to embryo implantation: clues from mouse models. *Nat Rev Genet* **7**, 185-199.

Vassena, R., Boue, S., Gonzalez-Roca, E., Aran, B., Auer, H., Veiga, A. and Izpisua Belmonte, J. C. (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development* **138**, 3699-3709.

Vogt, E. J., Meglicki, M., Hartung, K. I., Borsuk, E. and Behr, R. (2012). Importance of the pluripotency factor LIN28 in the mammalian nucleolus during early embryonic development. *Development* **139**, 4514-4523.

Wong, C. C., Loewke, K. E., Bossert, N. L., Behr, B., De Jonge, C. J., Baer, T. M. and Reijo Pera, R. A. (2010). Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nat Biotechnol* **28**, 1115-1121.

Voz, M. L., Agten, N. S., Van de Ven, W. J. and Kas, K. (2000). PLAG1, the main translocation target in pleomorphic adenoma of the salivary glands, is a positive regulator of IGF-II. *Cancer Res* **60**, 106-113.

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C. Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y. E., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593-597.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* **20**, 1131-1139.

Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y. and Wang, S. (2010). GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**, 976-978.

Zhang, W., Li, J., Guo, Y., Zhang, L., Xu, L., Gao, X., Zhu, B., Gao, H., Ni, H. and Chen, Y. (2016). Multi-strategy genome-wide association studies identify the DCAF16-NCAPG region as a susceptibility locus for average daily gain in cattle. *Sci Rep* **6**, 38073.

Figure 1

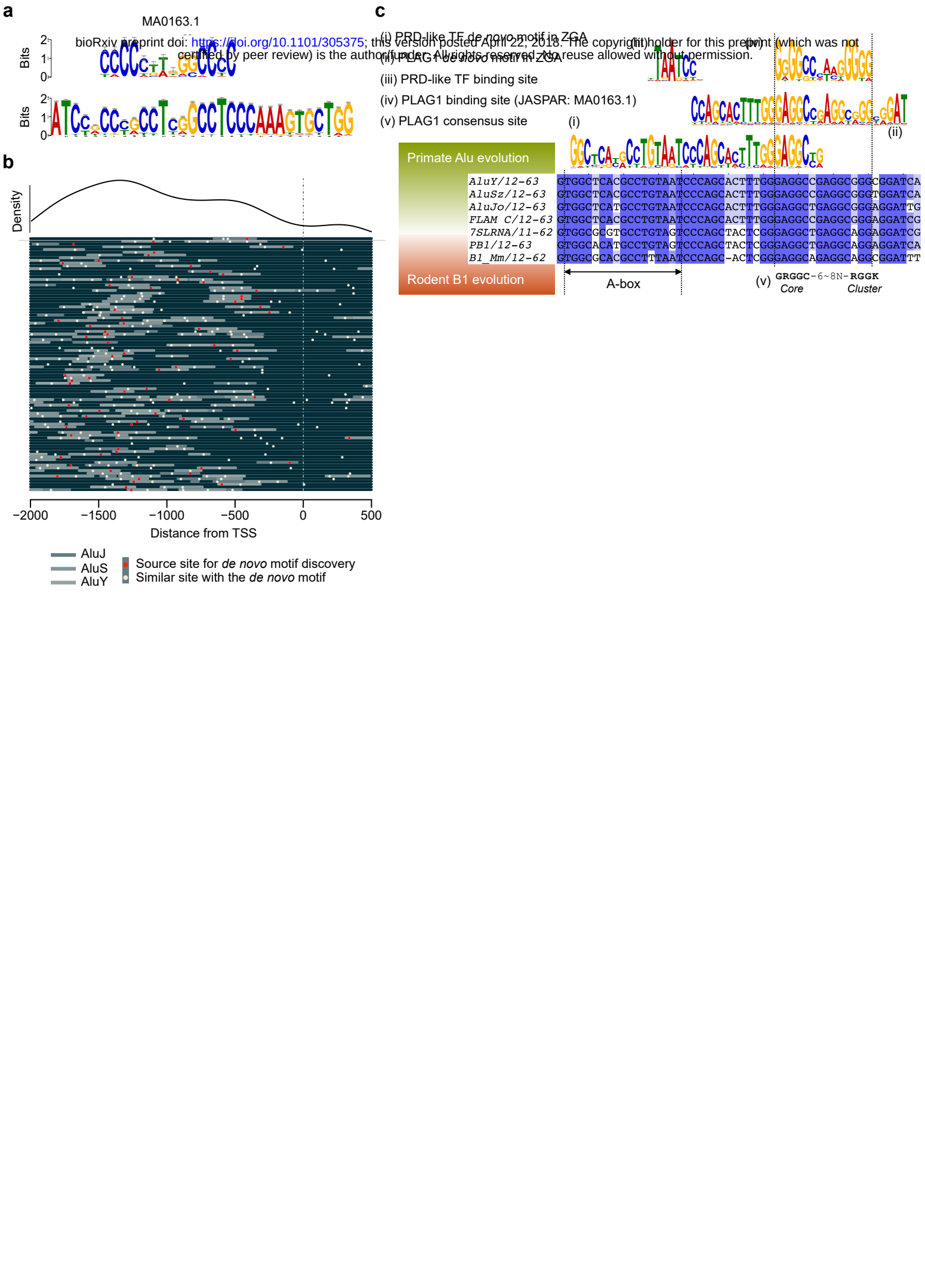


Figure 2

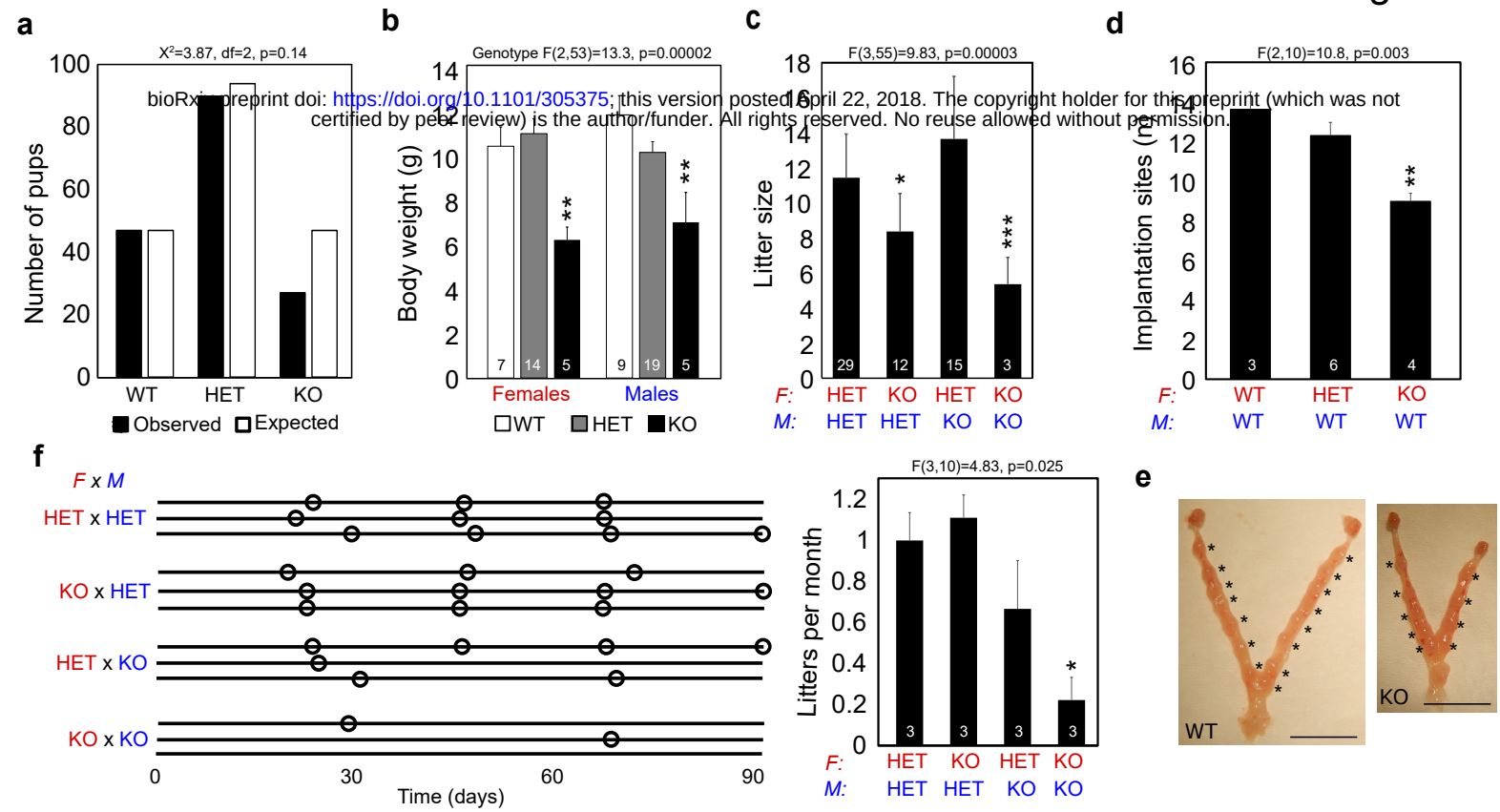


Figure 3

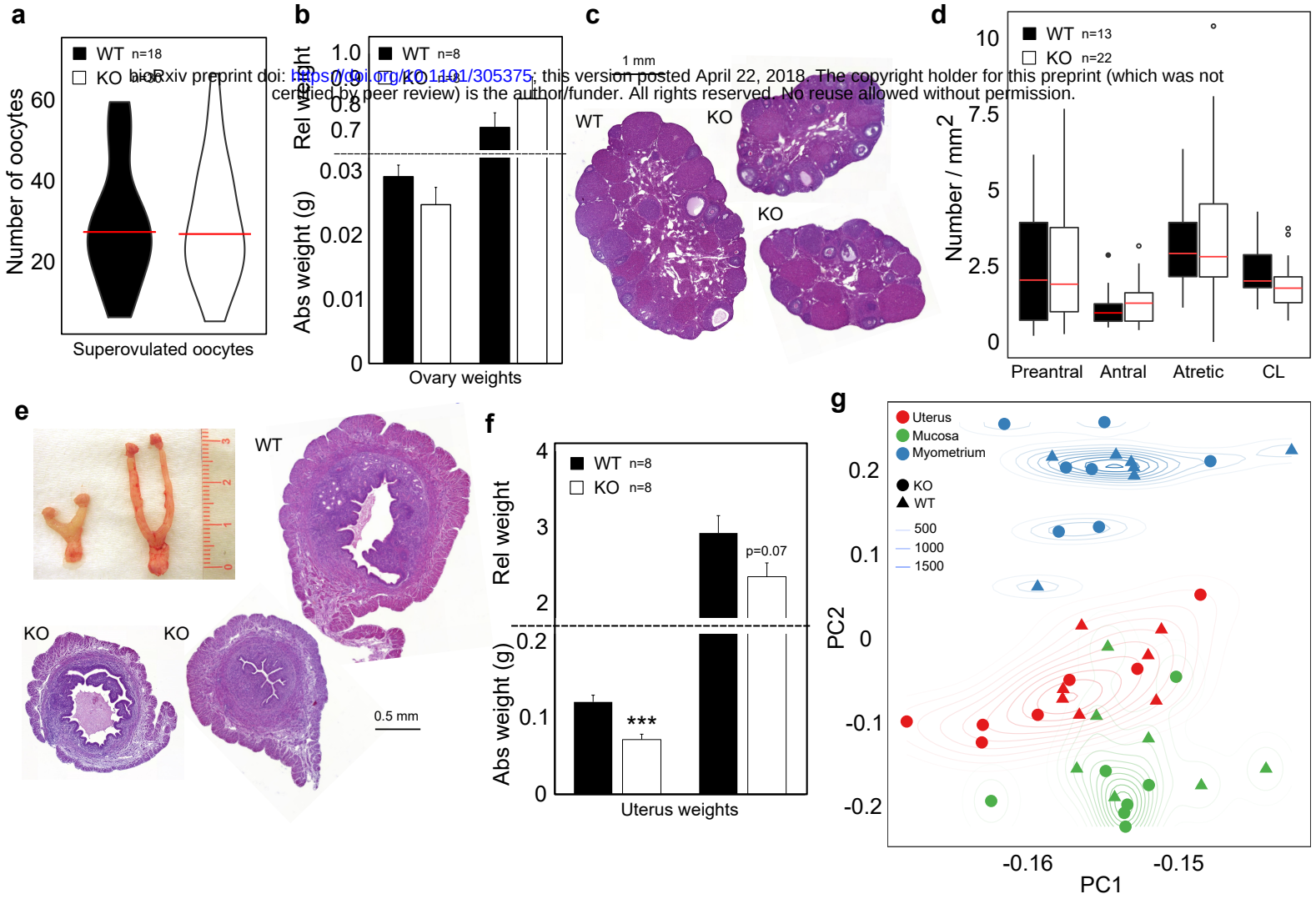


Figure 4

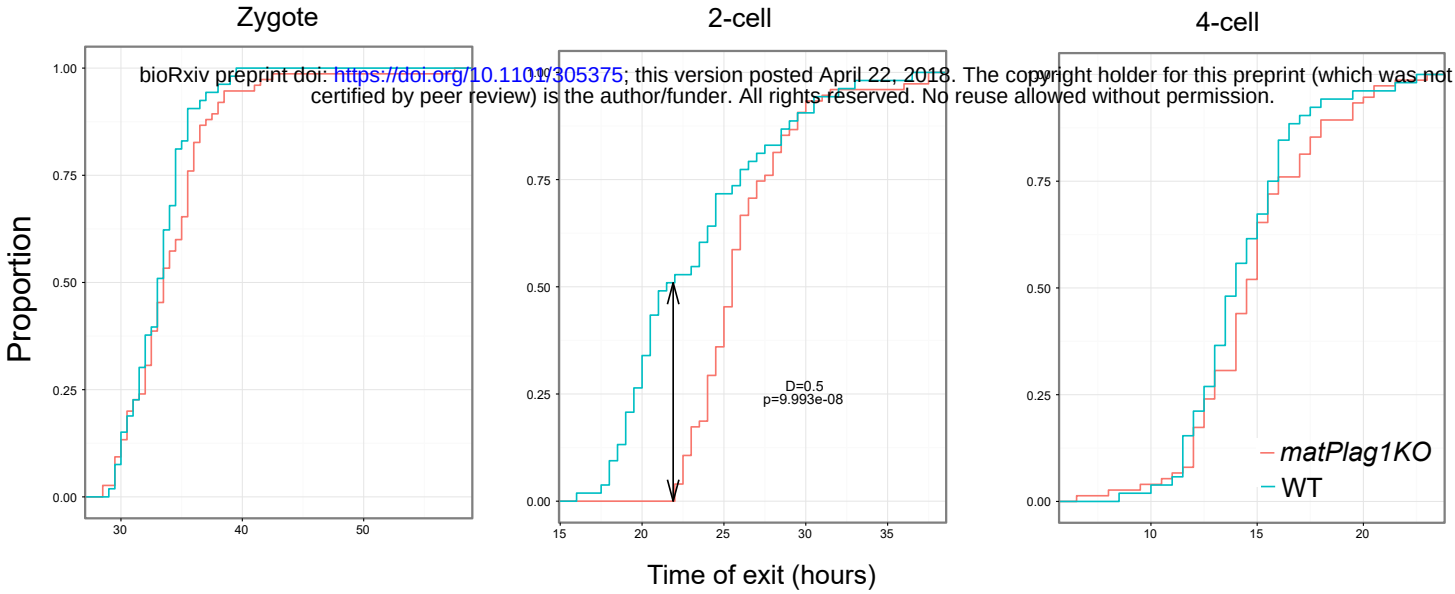
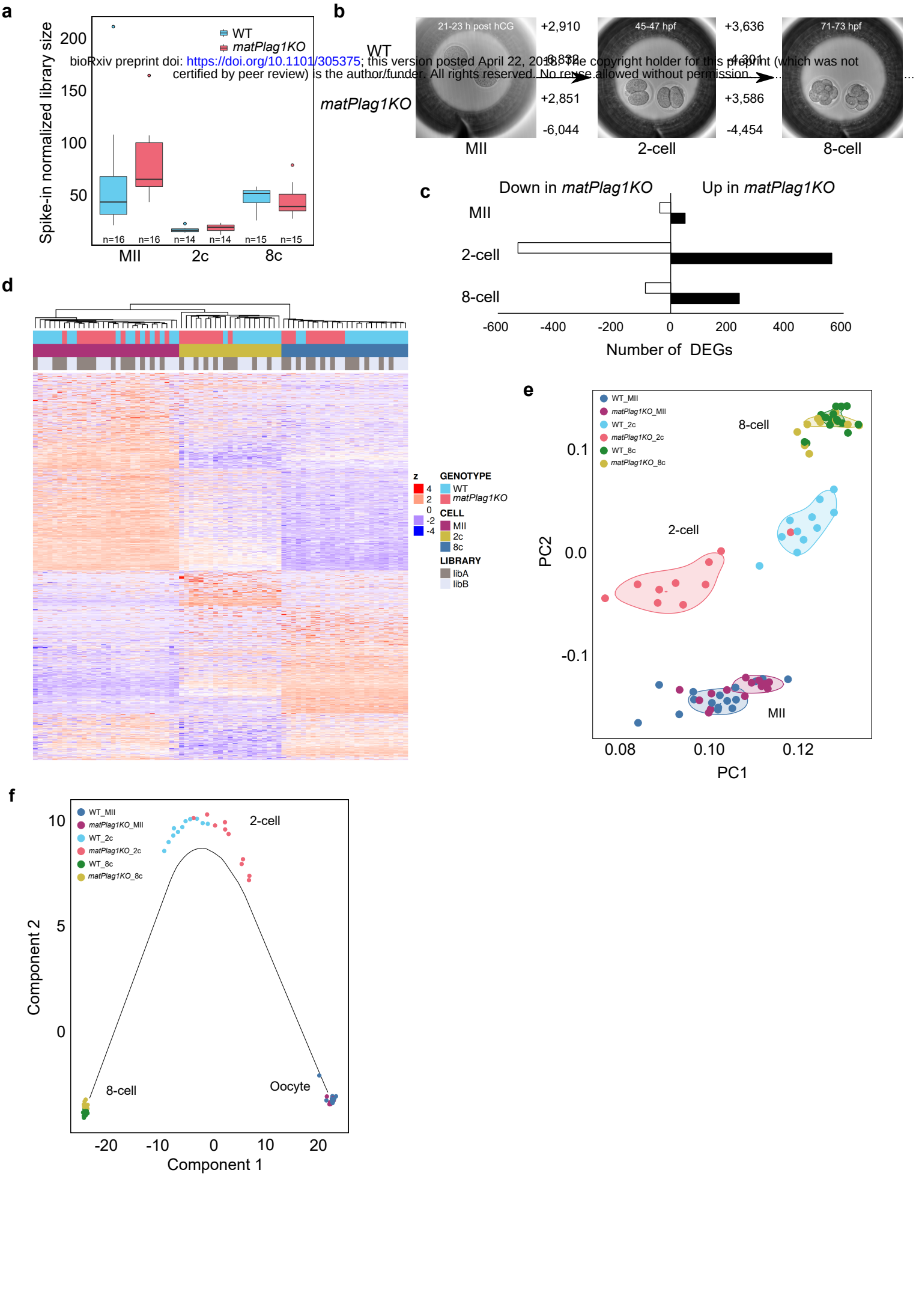
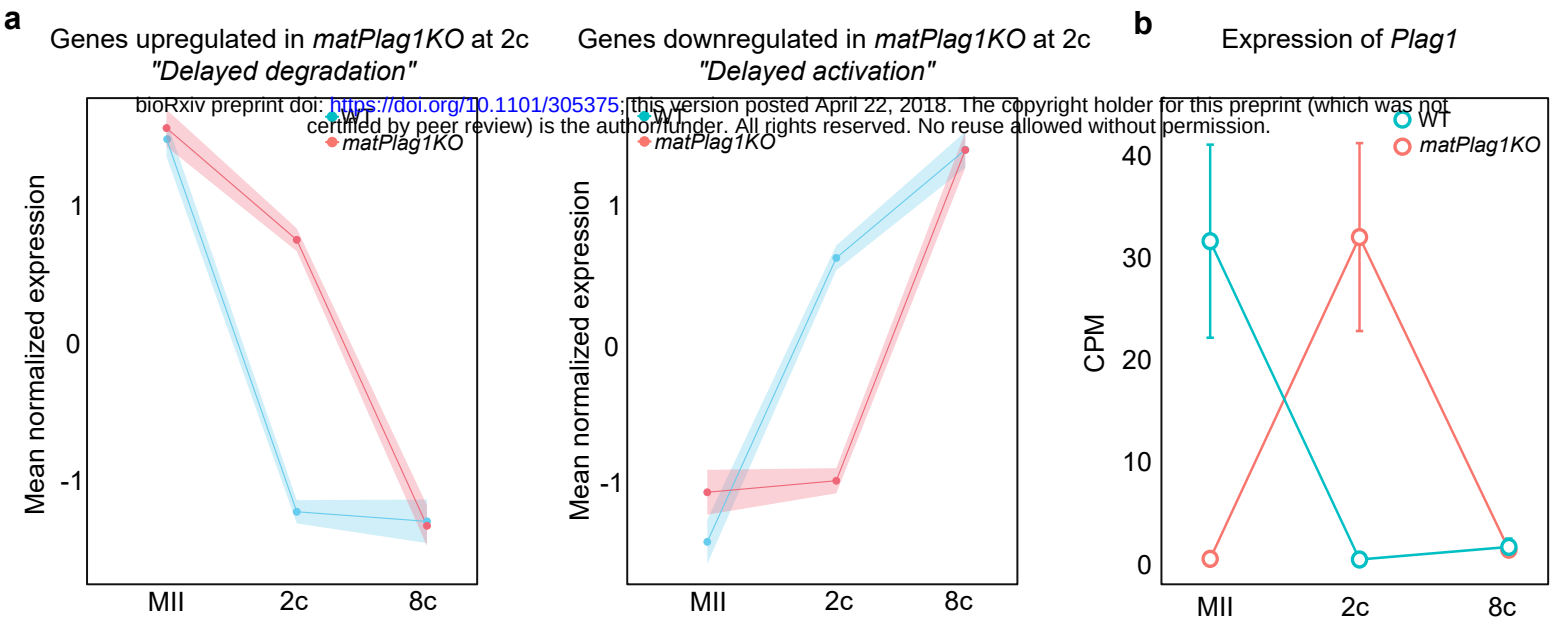
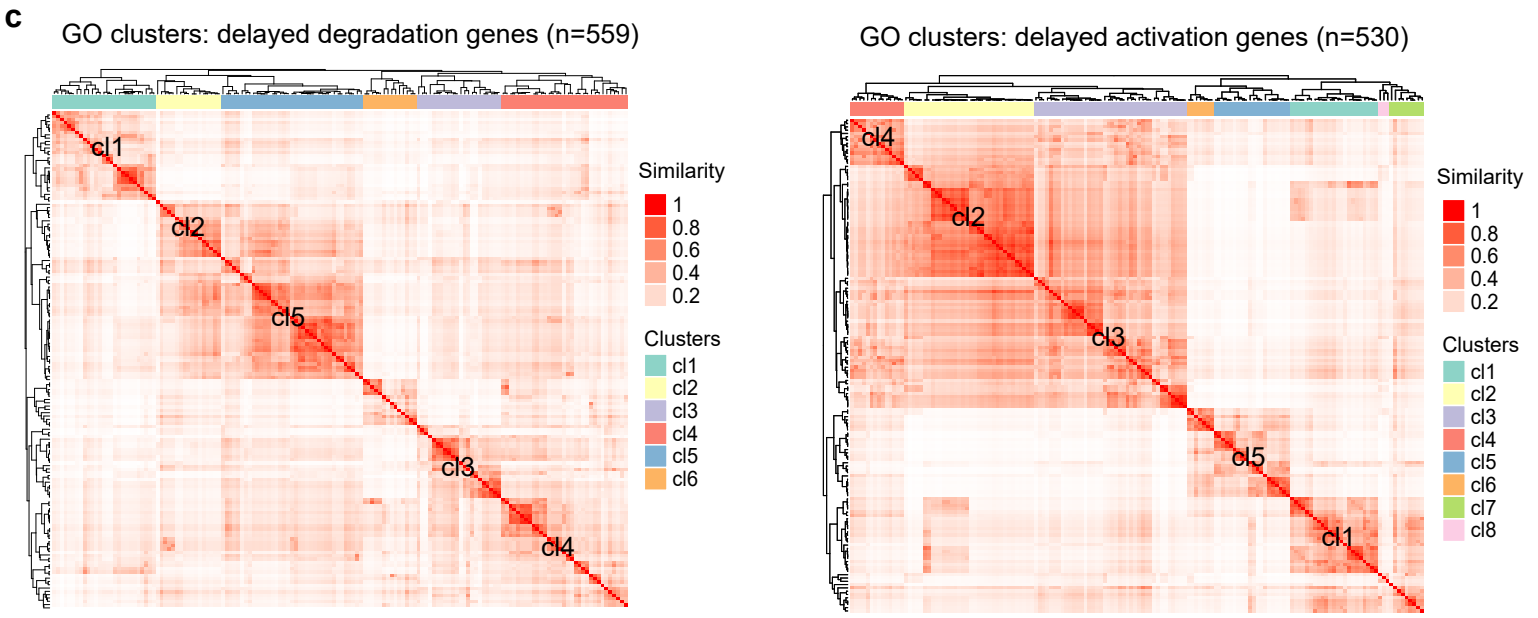


Figure 5



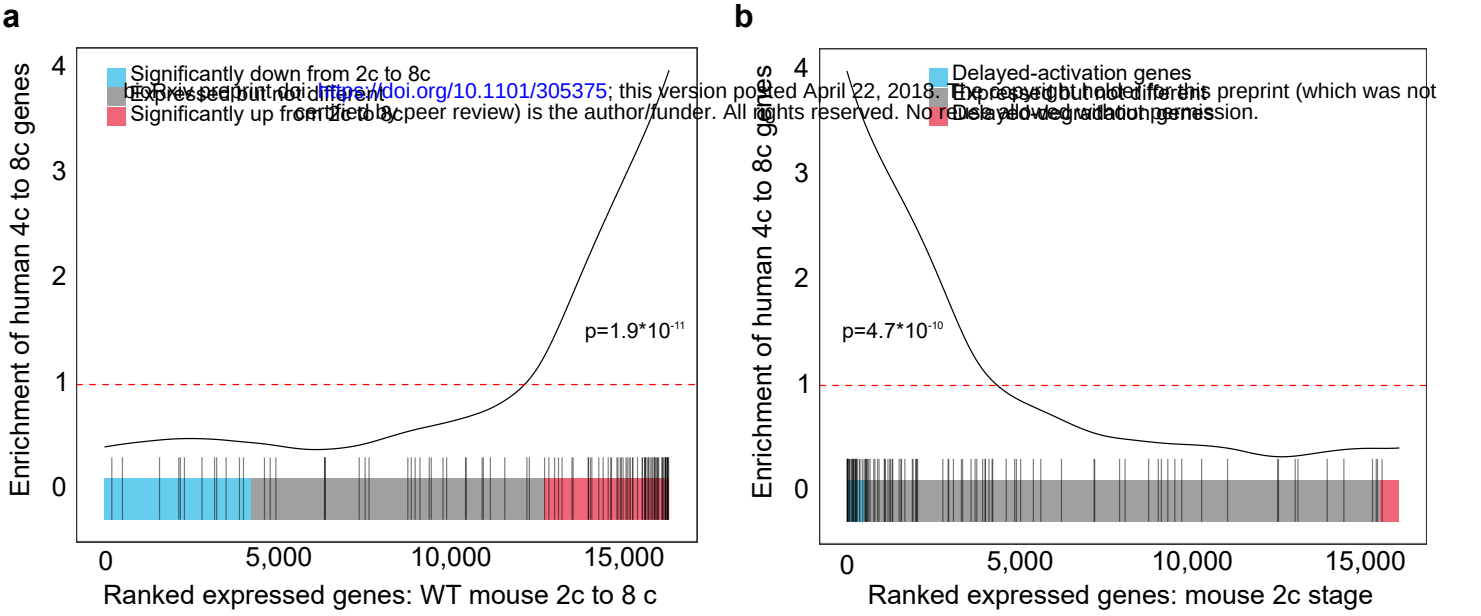


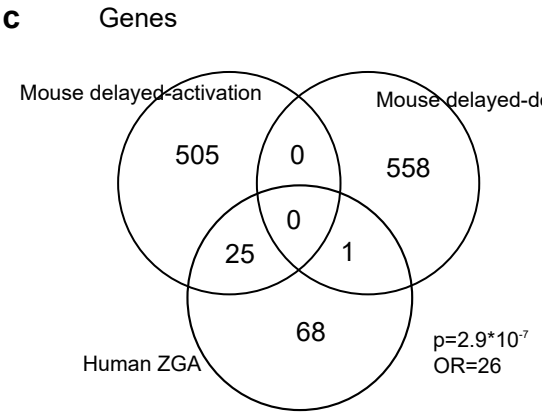


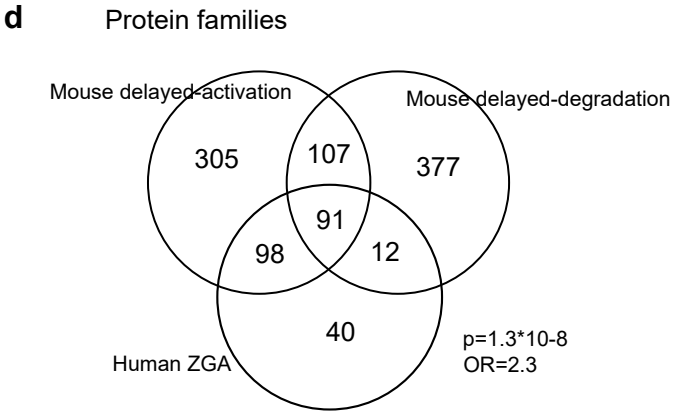
cl1: ribonucleoside metabolism, negative regulation of transcription and metabolism
cl2: development and morphogenesis, cell proliferation
cl3: protein localization, hormone secretion and transport, calcium transport
cl4: regulation of cell proliferation, hormone and second messenger signalling
cl5: regulation of neurogenesis, cell communication

cl1: ribosome biogenesis, cellular organelle assembly
cl2: rRNA maturation, RNA processing and slicing
cl3: cellular macromolecule biosynthesis, nucleoside metabolism, translation
cl4: nitrogen compound metabolism, organic substance biosynthesis
cl5: RNA nuclear export, establishment of protein localization

Figure 7







e

		2c: not affected (15,432 genes)	2c: delayed-degradation (559 genes)	2c: delayed-activation (530 genes)
PLAG1 site JASPAR		1,169 (7.6%)	52 (9.3%)	27 (5.1%)
PLAG1 *de novo* motif		2,509 (16.3%)	81 (14.5%)	**132 (24.9%)*** **