Pleomorphic Adenoma Gene 1 Is Needed For Timely Zygotic Genome Activation and Early Embryo Development

Elo Madissoon1, Anastasios Damdimopoulos2, Shintaro Katayama3, Kaarel Krjutškov3,4, Elisabet Einarsson1,4, Katarina Mamia5, Bert De Groef6, Outi Hovatta5, Juha Kere1,7,8 & Pauliina Damdimopoulou1,5

Pleomorphic adenoma gene 1 (PLAG1) is a transcription factor involved in cancer and growth. We discovered a de novo DNA motif containing a PLAG1 binding site in the promoters of genes activated during zygotic genome activation (ZGA) in human embryos. This motif was located within an Alu element in a region that was conserved in the murine B1 element. We show that maternally provided Plag1 is needed for timely mouse preimplantation embryo development. Heterozygous mouse embryos lacking maternal Plag1 showed disrupted regulation of 1,089 genes, spent significantly longer time in the 2-cell stage, and started expressing Plag1 ectopically from the paternal allele. The de novo PLAG1 motif was enriched in the promoters of the genes whose activation was delayed in the absence of Plag1. Further, these mouse genes showed a significant overlap with genes upregulated during human ZGA that also contain the motif. By gene ontology, the mouse and human ZGA genes with de novo PLAG1 motifs were involved in ribosome biogenesis and protein synthesis. Collectively, our data suggest that PLAG1 affects embryo development in mice and humans through a conserved DNA motif within Alu/B1 elements located in the promoters of a subset of ZGA genes.

Early preimplantation embryo development is dependent on zygotic genome activation (ZGA)1,2. Transcription from the newly formed zygotic genome starts gradually already in the one-cell embryo, and a major increase in transcriptional output, known as major ZGA, takes place during the 2-cell (2c) stage in mice and during the 4c-to-8c transition in humans1,2. Both minor and major ZGA are essential for cleavage stage development in the mouse3,4 and thus needed for the formation of a blastocyst capable of implanting to the uterine endometrium. Therefore, knowledge about the gene expression program during the first stages of embryonic development and regulation of ZGA will help discover factors controlling pluripotency, lineage differentiation and fertility. This has prompted many studies to map transcriptional programs during preimplantation development5–10.

In order to understand regulation of gene expression, precise knowledge of transcription start sites (TSSs) is needed. We have used an RNA-seq technology based on the detection of the 5′ ends of transcripts to map active TSSs during the first three days of human preimplantation development11. De novo motif calling using the regions around the detected TSSs led to the identification of multiple significant motifs harboring known transcription factor binding sites5,11. Here, we studied a de novo motif discovered in these analyses containing a putative

1Department of Biosciences and Nutrition, Karolinska Institutet, SE-14186, Stockholm, Sweden. 2Bioinformatics and Expression Analysis core facility, Department of Biosciences and Nutrition, Karolinska Institutet, SE-14186, Stockholm, Sweden. 3Competence Centre on Health Technologies, 50410, Tartu, Estonia. 4Molecular Neurology Research Program, University of Helsinki and Folkhålsan Institute of Genetics, 00014, Helsinki, Finland. 5Department of Clinical Science, Intervention and Technology, Karolinska Institutet, SE-14186, Stockholm, Sweden. 6Department of Physiology, Anatomy and Microbiology, La Trobe University, Bundoora, Victoria, 3086, Australia. 7Research Programs Unit, Molecular Neurology, University of Helsinki, and Folkhålsan Institute of Genetics, 00014, Helsinki, Finland. 8School of Basic and Medical Biosciences, King’s College London, Guy’s Hospital, London, WC2R 2LS, UK. Elo Madissoon, Anastasios Damdimopoulou and Shintaro Katayama contributed equally. Correspondence and requests for materials should be addressed to E.M. (email: elo.madissoon@ebi.ac.uk) or J.K. (email: juha.kere@ki.se) or P.D. (email: pauliina.damdimopoulou@ki.se)
binding site for the pleomorphic adenoma gene 1 (PLAG1). PLAG1 encodes a C2H2 zinc finger transcription factor and an oncogene that was first characterized in pleomorphic adenomas of the salivary gland. It belongs to the same protein family as the functionally redundant proto-oncogene PLAG-like 2 (PLAGL2) and the imprinted tumor suppressor PLAG-like 1 (PLAGL1). Ectopic expression of PLAG1 and PLAGL2 resulting from chromosomal translocation events can be found in malignant tumors. Associations with cancer have been the focus of most studies on PLAG 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. By using our human embryo transcriptome data set, Plag1 knockout (KO) mice, breeding experiments, single-embryo RNA-seq, and time-lapse analysis of cleavage stage embryo development, we show that PLAG1 controls a subset of ZGA genes and is needed for normal cleavage stage embryo development.

Results
A de novo motif containing a PLAG1 binding site is found in the promoters of human ZGA genes. Analysis of the TSSs upregulated during major ZGA in human embryos revealed a significant 31-bp de novo DNA motifs harboring a known PLAG1 binding site (Fig. 1a). This de novo PLAG1 motif was present in 74 of the 129 genes upregulated during ZGA and highly similar sites were found in additional 19 promoters (Fig. 1b and File S1). The de novo PLAG1 motif was located within a conserved region of an Alu element in close vicinity to an RNA polymerase III promoter A-box, and it also partially overlapped with the PRD-like transcription factor motif that we identified in our previous study (Fig. 1c; File S1). Alu elements are primate-specific retrotransposable short interspersed nuclear elements (SINEs) that evolved from a duplication of the 7SL RNA gene. 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.
Plag1 deficiency affects reproductive success but not ovarian or uterine function. To study the role of Plag1 in fertility and ZGA, Plag1 KO mice were obtained. The original phenotype was described on the Swiss Webster background whereas our colony was on the CD-1 strain, so we started by studying the reproductive phenotype. Pups from heterozygous intercrosses (HET × HET) did not significantly deviate from the expected Mendelian distribution, although 43% fewer KO pups were born than expected (Fig. 2a). KO pups were 41% smaller than wildtype (WT) at weaning, as reported before (Fig. 2b)\(^1\).\(^4\). Litter frequency over a three-month continuous breeding period did not differ between HET and KO females (Fig. 2c). However, KO males with HET females did not manage to maintain the approximate one litter–per–month rate (Fig. 2c). KO × KO intercrosses had significantly reduced litter frequency: the three KO × KO breeding pairs produced only two litters in total during the entire three-month test period (Fig. 2c).

Litter size was affected in couples with KO mothers (Fig. 2d)\(^4\). Breeding pairs with a KO female and a HET male produced significantly fewer pups per litter, whereas litter size was normal if the parental genotypes were reversed (HET female × KO male) (Fig. 2d). Homozygous KO × KO crosses produced the smallest litters (Fig. 2d). The significant reduction in litter size was seen as early as 7.5–8.5 days post coitum (dpc), when fewer implantation marks were observed in KO uteri as compared to HET and WT mothers (Fig. 2e). This did not depend on ovarian function, as neither oocyte yield in response to gonadotropin-induced superovulation or ovarian follicle counts differed between WT and KO females (Fig. S2a–d).

Maternal Plag1 knockout leads to delayed 2-cell stage embryo development, disrupted gene expression, and ectopic expression of paternal Plag1. We next studied embryos. Since Plag1 KO females had small litters regardless of the paternal genotype, we focused on the maternal Plag1 effect and studied embryos derived from Plag1 KO females crossed with WT males. These breeding pairs produced HET embryos that lack the maternal Plag1 allele, and will hereafter be referred to as matPlag1KO embryos.

Preimplantation development of 53 WT and 75 matPlag1KO embryos was analyzed by time-lapse microscopy from zygote to late morula during three independent imaging sessions, each of which contained both genotypes.
earlier human embryo transcriptome data and converted the mouse genes to human orthologues. The gene motif (31.5%) compared to “expressed but not affected” (21.5%) and delayed-degradation (19.3%) genes (Fig. 5d, A significantly higher proportion of the delayed-activation promoters contained at least one \( \text{PLAG1} \) de novo motif (31.5%) compared to “expressed but not affected” (21.5%) and delayed-degradation (19.3%) genes (Fig. 5d). However, the results were the opposite when we used our results were the opposite when we used our

In addition to occurrence (present or not), we also considered the frequency (how many) of the

The delayed-activation genes overlap with human ZGA genes and have enriched \( \text{PLAG1} \) de novo motif in their promoters.

The delayed-activation genes overlap with human ZGA genes and have enriched \( \text{PLAG1} \) de novo motif in their promoters. We next wanted to compare our current mouse embryo data with our earlier human embryo transcriptome data, and converted the mouse genes to human orthologues. The gene expression changes during major ZGA in humans (4c–8c transition) and WT mice (2c–8c transition) were highly similar in general between our datasets (Fig. S6). A gene set analysis showed that the genes upregulated during human major ZGA were significantly enriched among the mouse delayed-activation genes (Fig. S5a). The same result was obtained using the Chi-squared test: a significant overlap was found between the human major ZGA genes with the \( \text{matPlag1KO} \) delayed-activation genes but not with the delayed-degradation genes (Fig. S5b). Comparison of protein families yielded similar results (Fig. S5c).

We then studied the occurrence of the database \( \text{PLAG1} \) binding sites and the \( \text{de novo} \) \( \text{PLAG1} \) motifs in the affected mouse gene promoters. Database \( \text{PLAG1} \) binding sites were found in less than 10% of the promoters in general, and surprisingly the occurrence was significantly lower in the promoters of the delayed-activation genes (5.5%) (Fig. S4d). However, the results were the opposite when we used our \( \text{de novo} \) \( \text{PLAG1} \) motif instead. A significantly higher proportion of the delayed-activation promoters contained at least one \( \text{de novo} \) \( \text{PLAG1} \) motif (31.5%) compared to “expressed but not affected” (21.5%) and delayed-degradation (19.3%) genes (Fig. S4d, File S4). We also studied the connection of these \( \text{de novo} \) \( \text{PLAG1} \) motifs to B1 elements and obtained virtually identical results, showing that the vast majority of the \( \text{de novo} \) \( \text{PLAG1} \) motifs reside within B1 elements (Fig. S5d).

In addition to occurrence (present or not), we also considered the frequency (how many) of the \( \text{de novo} \) \( \text{PLAG1} \) motifs. A significantly higher frequency was found in the promoters of delayed-activation genes compared to delayed-degradation genes (Fig. S7).

Finally, we compared the homologous mouse delayed-activation and human ZGA genes that contained at least one \( \text{de novo} \) \( \text{PLAG1} \) motif and found a significant overlap (Fig. 5e). We GO-annotated the three gene groups, i.e.
Figure 3. Maternal Plag1 Knockout Leads to Delayed 2-cell stage Embryo Development and Disrupted Embryonic Gene Expression. (a) WT and matPlag1KO zygotes were collected for ex vivo imaging through time-lapse microscopy and cleavage stage developmental timing was recorded manually for each embryo. Plots show the empirical cumulative distribution function (y-axis, 0–1) of embryos that have exited the corresponding developmental stage at each time point (x-axis, hours). Significance between WT (N = 53) and matPlag1KO (N = 75) embryo developmental timing was tested using the Kolmogorov-Smirnov test and p-values are shown in the plots. Average time spent in the stage (SEM) is also shown.

(b) WT and matPlag1KO MII oocytes, 2-cell stage and 8-cell stage embryos were collected for single-embryo RNA-seq. The timing of hormonal treatments, WT embryo development and embryo collection time points are shown together with the numbers of embryos sequenced (numbers of donor females in brackets).

(c) Number of differentially expressed genes from one developmental stage to another (horizontal arrows) in matPlag1KO and WT embryos.

(d) Number of differentially expressed genes between matPlag1KO and WT embryos at the MII oocyte, 2c and 8c stage.
human ZGA, mouse delayed-activation and mouse-delayed degradation genes. 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, whereas mouse delayed-degradation genes showed little overlap with the human ZGA clusters (Fig. 5f). These data suggest that mouse and human embryos have a functionally conserved set of genes that is activated during ZGA and contain PLAG1 binding sites within repetitive elements in their promoters.

Discussion

In the present study, we have discovered a new role for the oncogene PLAG1 in the regulation of ZGA. We identified a de novo assembled motif containing a PLAG1 binding site among the promoters upregulated during ZGA in human embryos, and showed that the lack of maternally loaded Plag1 in mouse oocytes lead to a significant delay in ZGA on a transcriptional level with consequences for the timing of cleavage-stage development. Restoration of gene expression levels and embryo developmental speed coincided with ectopic expression of Plag1 from the paternal allele. These data imply a functional role for PLAG1 in the regulation of ZGA. Our data further propose that PLAG1 target genes have roles in central cellular processes that relate to ribosomes, RNA and protein metabolism, which undoubtedly have an essential function during early embryo growth.

We show that matPlag1KO embryos that lack the Plag1 transcript only during the first hours following fertilization spent significantly longer time in the 2c stage compared to WT embryos. The 2c stage is the developmental stage when the major ZGA takes place in the mouse, and failure to activate transcription leads to developmental arrest1. Embryo development can be affected by single genes. For example, knockdown of the pluripotency factor Lin28 in mouse embryos leads to arrest at the 2c–4c stage14, and KO of the maternal-effect gene Mater to arrested embryonic development at the 2c stage14. At the 2c stage, when the development of every studied matPlag1KO embryo was delayed compared to WT, these embryos started expressing Plag1 from the paternal allele. This upregulation was followed by regaining the normal developmental pace as well as normal transcriptomic program by the 8c stage, suggesting that the ectopic paternal expression of Plag1 rescued the embryo phenotype. Although these data argue that Plag1 is essential for embryo development at the 2c stage, we noted that even homozygous Plag1 KO intercrosses occasionally produced litters, showing that a complete lack of Plag1 is not strictly lethal to the embryos. Analysis of embryos from KO × KO breedings would help to understand the preimplantation phenotype. Unfortunately, the extremely poor breeding success of the KO × KO couples renders these studies unpractical. We further hypothesize that Plag2, the Plag1 -family member with redundant functions to Plag1 that is also maternally provided, might rescue embryo development in some cases. Testing this hypothesis would require generation of Plag1/Plag2 double KO mice, which is impossible due to the severe phenotype of Plag2 KO mice; KO pups die shortly after birth to starvation due to their inability to absorb chylomicrons20.

Although matPlag1KO embryos regained a normal developmental pace after exiting the 2c stage, their developmental success was not the same as that of WT embryos, as evidenced by the reduced number of implantation sites and smaller litter size. Our embryo time-lapse data do show that matPlag1KO embryos developed to the blastocyst stage with equal efficiency as WT embryos. Since KO uteri already showed less implantation marks at 7.5 – 8.5 dpc, the embryo loss must occur sometime between 4.5 and 7.5 dpc, i.e., when the blastocyst normally implants or shortly thereafter. Developmental competence of 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 viability19. Following this, we hypothesize that the delayed 2c stage development and associated dysregulation of over 1,000 genes in matPlag1KO embryos could have adverse effects on the overall developmental potential of the blastocysts. In addition, 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 contribute to some blastocysts missing this critical window22.

The de novo PLAG1 motif frequently localized within repetitive elements in the promoters of the ZGA genes. Alu elements are transposable elements ubiquitously present in primate genomes with involvement in gene regulation through various mechanisms23 and their counterparts in the mouse genome are B1 elements16. Despite independent evolution for over 80 million years, the densities of Alu and B1 elements in promoters of orthologous genes in humans and mice are surprisingly correlated15,24,25. In particular, Alu and B1 elements are highly enriched in promoters that are activated during ZGA26. The fact that the region that contains the PLAG1 binding site is conserved within the Alu and B1 elements suggests positive evolutionary selection. Collectively, these findings may suggest that Alu and B1 play a role in ZGA in humans and mice by attracting transcription factors such as PLAG1 to gene promoters.

The delayed-activation genes were mainly related to ribosome biogenesis, maturation and function by ontology. Even when we restricted the analysis to those genes that contained the de novo PLAG1 motif, GO analyses suggested roles in ribosome biogenesis and RNA and protein metabolism, both in mice and humans. It has been shown that Alu elements are enriched in promoters of genes involved in ribosome biogenesis, protein biosynthesis and RNA metabolism22, and one function could be to attract transcription factors such as PLAG1. Many oncogenes have effects on ribosome biogenesis, which enables cancerous cells to increase protein synthesis and grow rapidly27. Although PLAG1 is an oncogene, its potential role in ribosome biogenesis and protein synthesis has not been investigated. One of the most striking phenotypes of the Plag1 KO mice is their small size44. In
addition, PLAG1 polymorphisms associate with body size and growth in farm animals and humans. Based on our data, we present a hypothesis that PLAG1-associated growth phenotypes, such as growth retardation in the KO mice, results from modulation of protein synthesis that affects cell size and division rate. We conclude that the lack of maternally provided Plag1 leads to a delay in ZGA manifested as prolonged 2c stage development, which is rescued by ectopic paternal Plag1 expression. This delay and associated dysregulation of genes needed for ribosome biogenesis, RNA and protein metabolism could lead to reduced embryo competence for implantation, explaining the reduced litter size in KO mothers. We further propose that the effect of PLAG1 on ZGA genes arose through retrotransposition of Alu and B1 elements, where the PLAG1 binding sites have been under positive selection. Follow-up studies should focus on 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
Detailed materials and methods are provided in the Supplemental Information.

Mouse studies. All experiments were approved by the Swedish Board of Agriculture (#S5–14) and performed in accordance with the ethical licence. Plag1KO mice in CD-1 strain were a kind gift from Prof. Wim Van de Ven (University of Leuven, Belgium) and Dr. Carol Schuurmans (University of Calgary, Canada). The colony was established via embryo transfers,
maintained as HET breedings, and genotyped using ear punches. To generate oocytes and zygotes for experiments, sexually mature young females were superovulated. Their ovaries and uteri were collected for histology and gene expression analysis. For time-lapse microscopy, zygotes (n = 103 matPlag1KO and n = 89 WT) were placed into a live-cell imaging incubator under a Nikon Ti-E spinning disk wide-field microscope and imaged every 30 min with an Andor EM-CCD camera. The experiment was carried out three times with both
To count implantation marks, KO and WT females were mated with trained WT studs and their uteri were collected 7.5–8.5 dpc for visual inspection. The experiment was carried out twice, with a total of 8 KO and 8 WT females.

RNA expression in uterus samples (n = 8 Plag1KO, n = 8 WT) was measured using STRT RNA-seq protocol.42 Uterine horns were cut longitudinally, mucosa and myometrium separated by scraping, and the samples stored in RNAlater (Ambion, Foster City, CA, USA) until RNA extraction (RNeasy Mini kit, Qiagen, Hilden, Germany). Ten nanograms of RNA (RIN > 8) was used for RNA-seq as described.42

RNA expression in single manually picked oocytes and embryos was detected by analyzing two independent libraries prepared on three occasions with both genotypes present both times. Modified STRT protocol was used.

**RNA-seq data analysis.** RNA-seq data was analyzed as described previously.42 Libraries with a median gene expression of log2 counts per million (cpm) under 0 were excluded, and cell libraries were normalized with the TMM methods using EdgeR.33 Differential gene expression analysis using EdgeR was performed on genes that had 1 cpm in at least five or more samples. Principal component analysis, heatmaps, hierarchical clustering, cell trajectory and GO analyses were carried out using DEGs in R.43 Enriched GO terms were identified using Fisher statistics and compared via semantic similarity analysis using the Wang algorithm.34 Homologene from NCBI was used to convert the human genes to the homologous mouse genes, and gene enrichment analyses were carried out using geneSetTest function from the limma package.

**Promoter analyses.** Human embryo promoter analysis was performed as previously described.43 SINE repetitive elements were retrieved from the Dfam database and aligned in JalView. Mouse embryo promoter analysis was carried out with Homer.44 Distance of repetitive elements and de novo PLAG1 motifs to the nearest TSS was plotted using ggplot2. Enrichment was analyzed using Fisher’s exact test.

**Statistical analysis.** Continuous data were analyzed with Student’s t-test, one-way or two-way ANOVA, followed by Fisher LSD post-hoc test. Normality was tested with the Shaphiro-Wilks test and homoscedasticity with Levene's test. Categorical data were analyzed with the χ² test. The exit time of embryos from each developmental stage was plotted as an empirical distribution function (ecdf) using ggplot2 and significance tested using the Kolmogorov-Smirnov test. All analyses were carried out using R, the p values are two-tailed and considered significant if 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.

**References**

1. Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic Genome Activation in Vertebrates. *Dev Cell* **42**, 316–332, https://doi.org/10.1016/j.devcel.2017.07.026 (2017).
2. Niakan, K. K., Han, J., Pedersen, R. A., Simon, C. & Perera, R. A. Human pre-implantation embryo development. *Development* **139**, 829–841, https://doi.org/10.1242/dev.060428 (2012).
3. Warner, C. M. & Verstreck, L. R. In vivo and in vitro effect of alpha-amanitin on preimplantation mouse embryo RNA polymerase. *Nature* **248**, 678–680 (1974).
4. Abe, K. I. et al. Minor zygotic gene activation is essential for mouse preimplantation development. *Proc Natl Acad Sci USA* **115**, E6780–E6788, https://doi.org/10.1073/pnas.1504390115 (2018).
5. De lao, A. et al. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* **49**, 941–945, https://doi.org/10.1038/ng.3858 (2017).
6. Metz, P. E. et al. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165**, 1012–1028, https://doi.org/10.1016/j.cell.2016.03.023 (2016).
7. Tohonen, V. et al. Transcription Activation Of Early Human Development Suggests DUX4 As An Embryonic Regulator. *bioRxiv*, https://doi.org/10.1101/123208 (2017).
8. Vassena, R. et al. Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development* **138**, 3699–3709, https://doi.org/10.1242/dev.064741 (2011).
9. Xue, Z. et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597, https://doi.org/10.1038/nature12364 (2013).
10. Yan, L. et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* **20**, 1131–1139, https://doi.org/10.1038/nsmb.2660 (2013).
11. Tohonen, V. et al. Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nat Commun* **6**, 8207, https://doi.org/10.1038/ncomms9207 (2015).
12. Kas, K. et al. Promoter swapping between the genes for a novel zinc finger protein and beta-catenin in pleomorphic adenomas with t(3;8)(p21;q12) translocations. *Nat Genet* **15**, 170–174, https://doi.org/10.1038/ng297-170 (1997).
13. Juma, A. R., Damdimopoulou, P. E., Grommen, S. V., Van de Ven, W. J. & De Groef, B. Emerging role of PLAG1 as a regulator of growth and reproduction. *J Endocrinol* **228**, R45–56, https://doi.org/10.1530/JOE-15-0449 (2016).
14. Hensen, K. et al. Targeted disruption of the murine Plag1 proto-oncogene causes growth retardation and reduced fertility. *Dev Growth Differ* **46**, 459–470, https://doi.org/10.1111/j.1440-2036.2004.00762.x (2004).
15. Imai, T. et al. Alu sequences are processed 7SL RNA genes. *Nature* **312**, 171–172 (1984).
16. Labuda, D., Sinnett, D., Richer, C., Deracon, J. M. & Striker, G. Evolution of mouse B1 repeats: 7SL RNA folding pattern conserved. *J Mol Evol* **32**, 405–414 (1991).
17. Voz, M. L., Agten, N. S., Van de Ven, W. J. & Kas, K. PLAG1, the main translocation target in pleomorphic adenoma of the salivary glands, is a positive regulator of IGF-II. *Cancer Res* **60**, 106–113 (2000).
18. Vogt, E. J., Miegisch, M., Hartung, K. I., Borsuk, E. & Behr, R. Importance of the pluri potency factor LIN28 in the mammalian nucleus during early embryonic development. Development 139, 4514–4523, https://doi.org/10.1242/dev.083279 (2012).

19. Tong, Z. B. et al. Mater, a maternal effect gene required for early embryonic development in mice. Nat Genet 26, 267–268, https://doi.org/10.1038/ng.1547 (2000).

20. Van Dyck, E. F. et al. Loss of the Plag1.2 transcription factor affects lactel uptake of chylomicrons. Cell Metab 6, 406–413, https://doi.org/10.1016/j.cmet.2007.09.010 (2007).

21. Wong, C. C. et al. Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. Nat Biotechnol 28, 1115–1121, https://doi.org/10.1038/nbt.1686 (2010).

22. Wang, H. & Dey, S. K. Roadmap to embryo implantation: clues from mouse models. Nat Rev Genet 7, 185–199, https://doi.org/10.1038/nrg1808 (2006).

23. Elbarbary, R. A., Lucas, B. A. & Maquat, L. E. Retrotransposons as regulators of gene expression. Science 351, aac7247, https://doi.org/10.1126/science.aac7247 (2016).

24. Mouse Genome Sequencing, C. Exploratory bioinformatics investigation reveals importance of “junk” DNA in early embryo development. BMC Genomics 18, 200, https://doi.org/10.1186/s12864-017-3566-0 (2017).

25. Polak, P. & Domany, E. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. BMC Genomics 7, 133, https://doi.org/10.1186/1471-2164-7-133 (2006).

26. Ge, S. X. Exploratory bioinformatics investigation reveals importance of “junk” DNA in early embryo development. BMC Genomics 18, 200, https://doi.org/10.1186/s12864-017-3566-0 (2017).

27. Pelletier, J., Thomas, G. & Volarevic, S. Ribosome biogenesis in cancer: new players and therapeutic avenues. Nat Rev Cancer 18, 51–63, https://doi.org/10.1038/nrc.2017.104 (2018).

28. Fink, T. et al. Functional confirmation of PLAG1 as the candidate causative gene underlying major pleiotropic effects on body weight and milk characteristics. Sci Rep 7, 44793, https://doi.org/10.1038/s41598-019-44882-0 (2019).

29. Rubín, C. J. et al. Strong signatures of selection in the domestic pig genome. Proc Natl Acad Sci USA 109, 19529–19536, https://doi.org/10.1073/pnas.1217149109 (2012).

30. Utsumoizumi, Y. T. et al. 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, https://doi.org/10.1186/1471-2164-14-52 (2013).

31. Zhang, W. et al. 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, https://doi.org/10.1038/srep38073 (2016).

32. Krutskov, K. et al. Single-cell transcriptome analysis of endometrial tissue. Hum Reprod 31, 844−853, https://doi.org/10.1093/humrep/dev008 (2016).

33. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139−140, https://doi.org/10.1093/bioinformatics/btp616 (2010).

34. R. A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria 2010).

35. Yu, G. et al. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics 26, 976−978, https://doi.org/10.1093/bioinformatics/btp064 (2010).

36. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, 447, https://doi.org/10.1093/nar/gkv007 (2015).

37. Hubley, R. et al. The Dfam database of repetitive DNA families. Nucleic Acids Res 44, D81–89, https://doi.org/10.1093/nar/gkv1272 (2016).

38. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576−589, https://doi.org/10.1016/j.molcel.2010.05.004 (2010).

Acknowledgements

Plag1 KO mice were generated in the Laboratory for Molecular Oncology of Prof. Wim Van de Ven, Center for Human Genetics, Catholic University of Leuven, Belgium. The authors thank Prof. Wim Van de Ven for access to the Plag1KO mouse line and Dr Carol Schuurmans (University of Calgary, Canada) for providing the mice. Stanislaw Wawrzyczek is thanked for assistance with X-gal staining and Ingegerd Fransson for RNA-seq library preparation. The morphological phenotype analysis (PEN0) core facility and Tarja Schröder are thanked for the preparation of the histological specimens. 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, that are 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 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 the Knut and Alice Wallenberg Foundation (KAW2015.0096) and by a Distinguished Professor Award from Karolinska Institutet to J.K. Essential funding was also provided by the Jane & Aatos Erkko Foundation to PD.

Author Contributions

P.D., E.M. and J.K. planned and designed the study. P.D. and E.M. carried out experiments. P.D., A.D. and E.M. analyzed data; A.D. and S.K. designed and carried out bioinformatic analyses; K.K. and E.E. performed RNA-seq analyses; K.M. conducted histomorphometric measurements and embry bioinformatic analyses; and B.D.G. did X-gal stainings. J.K. and O.H. provided essential resources. P.D., E.M., A.D., B.D.G. and J.K. wrote the paper. All authors commented on the manuscript, have approved the final version, and have agreed both to be personally accountable for the author’s own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-44882-0.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
