

# A spontaneous genetically induced epiallele at a retrotransposon shapes host genome function

Tessa M Bertozzi<sup>†</sup>, Nozomi Takahashi, Geula Hanin, Anastasiya Kazachenka<sup>‡</sup>, Anne C Ferguson-Smith\*

Department of Genetics, University of Cambridge, Cambridge, United Kingdom

**Abstract** Intracisternal A-particles (IAPs) are endogenous retroviruses (ERVs) responsible for most insertional mutations in the mouse. Full-length IAPs harbour genes flanked by long terminal repeats (LTRs). Here, we identify a solo LTR IAP variant (*lap5-1<sup>solo</sup>*) recently formed in the inbred C57BL/6J mouse strain. In contrast to the C57BL/6J full-length IAP at this locus (*lap5-1<sup>full</sup>*), *lap5-1<sup>solo</sup>* lacks DNA methylation and H3K9 trimethylation. The distinct DNA methylation levels between the two alleles are established during preimplantation development, likely due to loss of KRAB zinc finger protein binding at the *lap5-1<sup>solo</sup>* variant. *lap5-1<sup>solo</sup>* methylation increases and becomes more variable in a hybrid genetic background yet is unresponsive to maternal dietary methyl supplementation. Differential epigenetic modification of the two variants is associated with metabolic differences and tissue-specific changes in adjacent gene expression. Our characterisation of *lap5-1* as a genetically induced epiallele with functional consequences establishes a new model to study transposable element repression and host-element co-evolution.

\*For correspondence:  
afsmith@gen.cam.ac.uk

Present address: <sup>†</sup>Whitehead Institute, Cambridge, United States; <sup>‡</sup>The Francis Crick Institute, London, United Kingdom

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 18

**Received:** 27 November 2020

**Accepted:** 22 March 2021

**Published:** 23 March 2021

**Reviewing editor:** Deborah Bourc'his, Institut Curie, France

© Copyright Bertozzi et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

More than 10% of the mouse genome is made up of endogenous retroviruses (ERVs) (Smit et al., 2015). ERVs are transposable elements (TEs) containing retrotransposition-enabling genes flanked by non-coding identical 5' and 3' long terminal repeats (LTRs). Although most ERVs have lost their mobilisation potential due to mutational decay, they retain the ability to modulate host genome function through the use of transcriptional regulation motifs contained in their sequences. These include transcription factor binding sites, polyadenylation signals, and splice acceptor sites (Maksakova et al., 2006). For instance, solo LTRs produced via inter-LTR homologous recombination make up a large fraction of mammalian ERV sequences and can influence host gene expression through their promoter activity (Nellåker et al., 2012; Ruda et al., 2004; Subramanian et al., 2011).

Mammals have evolved a range of mechanisms to mitigate the deleterious effects of ERV retrotransposition and transcriptional disruption. The vast majority of mouse ERVs exhibit high levels of DNA methylation, and loss of DNA methyltransferase activity causes an increase in ERV transcription in embryos and in the germline (Barau et al., 2016; Bourc'his and Bestor, 2004; Jain et al., 2017; Walsh et al., 1998). In addition, the deposition of histone H3 lysine nine trimethylation (H3K9me3) by the methyltransferase SETDB1 (or ESET) in early development is crucial for ERV repression (Matsui et al., 2010). SETDB1 is recruited to ERVs following the binding of KAP1 (or TRIM28) to KRAB zinc finger proteins (KZFPs) that recognise specific sequence motifs within ERV elements (Ecco et al., 2017). RNA-mediated mechanisms such as the piRNA pathway also play key roles in silencing mammalian ERVs, especially in the male germline (Aravin et al., 2007; Carmell et al., 2007).

**eLife digest** Our genome provides a complete set of genetic instructions for life. It begins by directing the growth and development of the embryo, and subsequently supports all the cells of the adult body in their daily routines. Yet approximately 10% of the DNA in mammalian genomes is made up of sequences originating from past retroviral infections, leaving a calling card in our genetic code.

While these segments of retroviral DNA can no longer produce new infectious viruses, some of them retain the ability to copy themselves and jump into new parts of the genome. This can be problematic if they jump into and disrupt an important piece of genetic code. To protect against this, our bodies have evolved the ability to chemically strap down retroviral sequences by adding methyl groups to them and by modifying the proteins they are wrapped around. However, some of these endogenous retroviruses can dodge such so-called epigenetic modifications and disrupt genome function as a result.

Studying a population of widely used inbred laboratory mice, Bertozzi et al. have identified a retroviral element that evades these epigenetic restraints. They discovered that some mice carry a full-length retroviral sequence while others have a shortened version of the same element. The shorter sequence lacked the repressive epigenetic marks found on the longer version, and this affected the expression of nearby genes. Moreover, the repressive marks could be partially restored by breeding the short-version mice with a distantly related mouse strain.

Bertozzi et al. highlight an important issue for research using mouse models. Inbred laboratory mouse strains are assumed to have a fixed genetic code which allows scientists to conclude that any observed differences in their experiments are not a product of background genetic variation. However, this study emphasizes that this assumption is not guaranteed, and that hidden genetic diversity may be present in ostensibly genetically identical mice, with important implications for experimental outcomes.

In addition, Bertozzi et al. provide a new mouse model for researchers to study the evolution and regulation of retroviral sequences and the impact of these processes on cell function.

Intracisternal A-particles (IAPs) are murine-specific ERVs. They are young and highly active elements, exhibiting extensive insertional polymorphism across inbred mouse strains and accounting for more than 10,000 ERVs in the C57BL/6J (B6) inbred strain (Nellåker et al., 2012; Smit et al., 2015). IAP insertions are responsible for the majority of documented insertional mutations in the mouse, most of which disrupt host gene function by generating aberrant or fusion gene transcripts (Gagnier et al., 2019). In some cases, the phenotypic severity of the mutation is associated with the methylation status of the IAP. For example, the IAP LTR promoter at the *Agouti viable yellow* ( $A^y$ ) allele drives ectopic expression of the downstream coat colour gene *Agouti* when unmethylated. By mechanisms not yet fully understood, the  $A^y$  IAP exhibits variable DNA methylation levels across genetically identical individuals, which results in inbred mice displaying a range of coat colours (Duhl et al., 1994; Morgan et al., 1999).

We previously carried out a genome-wide screen to probe the generalisability of inter-individual methylation variability at IAPs (Elmer et al., 2021; Kazachenka et al., 2018). We identified dozens of novel variably methylated IAPs (VM-IAPs) in the B6 genome and showed that their characteristic methylation variability is recapitulated from generation to generation irrespective of parental methylation level. Our screen for VM-IAPs identified the IAP-*Pgm2* element, named after its closest annotated coding gene *Phosphoglucosyltransferase-2* (*Pgm2*). It is a fully structured 7.5 kb IAP located on Chromosome 5 containing identical 5' and 3' LTRs of the IAPLTR2\_Mm subclass (mm10 coordinates: chr5:64,030,834–64,038,297; **Figure 1—figure supplement 1**).

Here we show that IAP-*Pgm2*, in sharp contrast to VM-IAPs, exhibits two distinct DNA methylation states which are stably inherited from parent to offspring within the B6 strain. Sequencing of the locus reveals that the epiallele is genetically conferred, where the methylated variant of IAP-*Pgm2* is a full-length IAP matching the B6 reference genome (renamed *lap5-1<sup>full</sup>*) and the unmethylated variant is a solo LTR produced from an inter-LTR recombination event (renamed *lap5-1<sup>solo</sup>*). The absence of DNA methylation at *lap5-1<sup>solo</sup>* is accompanied by a loss of H3K9me3 marks. We find

that differential modification of the two variants is established during early preimplantation development and identify candidate KZFPs responsible for the acquisition of contrasting epigenetic states. We report that *lap5-1<sup>solo</sup>* is unresponsive to dietary methyl supplementation but highly susceptible to genetic background, becoming a bona fide VM-IAP in an F1 hybrid context. In addition, we demonstrate that formation of the *lap5-1<sup>solo</sup>* allele is associated with tissue-specific changes in neighbouring gene expression and a decrease in fasting plasma glucose and triglyceride concentrations. Our study establishes the *lap5-1* locus as a naturally occurring and biologically relevant model in the widely studied reference mouse strain to investigate the mechanisms underlying TE repression, inter-individual methylation variability, and TE-induced disruptions to host genome function.

## Results

### IAP-*Pgm2* methylation is tri-modally distributed in inbred B6 mice

Following our genome-wide screen for VM-IAPs in the B6 genome, DNA methylation at the distal CpGs of the 5' LTR of each candidate was validated using genomic DNA (gDNA) extracted from adult inbred B6 mice (Elmer et al., 2021; Kazachenka et al., 2018). While VM-IAP DNA methylation levels display continuous probability distributions in the B6 population, the IAP-*Pgm2* 5' LTR exhibited three distinct states: high (>85%), low (<20%), and intermediate (60–70%) methylation (Figure 1A and B). This pattern was observed in both sexes (Figure 1—figure supplement 2). In addition, 5' and 3' LTR methylation levels were consistent with one another within an individual (Figure 1A and B).

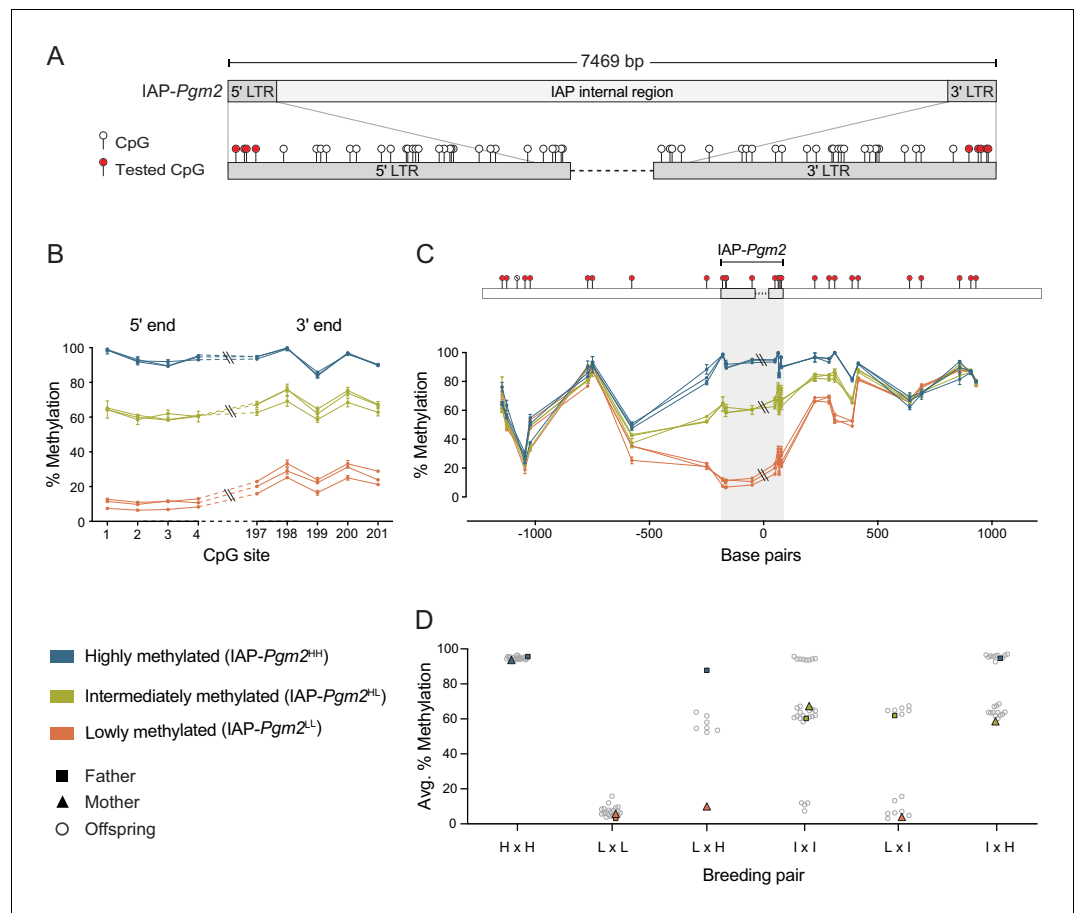
Methylation quantification of unique non-repetitive DNA immediately up- and downstream of IAP-*Pgm2* showed that the three distinct methylation states become less defined as the distance from the LTR borders increases, ultimately collapsing approximately 500 bp from either side of the IAP (Figure 1C). This provides evidence for short-distance spreading of DNA methylation levels from IAP-*Pgm2* into bordering DNA and suggests that the methylation differences observed between individuals are intrinsic to the IAP-*Pgm2* element rather than a reflection of differential methylation of the insertion site prior to integration.

### IAP-*Pgm2* methylation exhibits stable Mendelian inheritance

One of the characteristic properties of VM-IAPs is the reconstruction of inter-individual methylation variability from one generation to another regardless of parental methylation level (Kazachenka et al., 2018). To test whether this phenomenon occurs at IAP-*Pgm2*, specific parental combinations were set up for breeding and IAP-*Pgm2* 5' LTR methylation levels were quantified in the offspring. In stark contrast to VM-IAPs, IAP-*Pgm2* exhibited stable inheritance of methylation levels. Offspring born to highly methylated parents were all highly methylated and offspring born to lowly methylated parents were all lowly methylated (Figure 1D). When one parent was highly methylated and the other lowly methylated, all offspring were intermediately methylated (Figure 1D). This indicates that high and low methylation states are allelic variants of IAP-*Pgm2* (designated IAP-*Pgm2*<sup>HH</sup> and IAP-*Pgm2*<sup>LL</sup>), with intermediate methylation representing co-dominant epigenetic heterozygosity (IAP-*Pgm2*<sup>HL</sup>). Additional crosses confirmed this inheritance pattern: an IAP-*Pgm2*<sup>HL</sup> intercross produced IAP-*Pgm2*<sup>HH</sup>, IAP-*Pgm2*<sup>LL</sup>, and IAP-*Pgm2*<sup>HL</sup> offspring; an IAP-*Pgm2*<sup>HH</sup> × IAP-*Pgm2*<sup>HL</sup> cross produced IAP-*Pgm2*<sup>HH</sup> and IAP-*Pgm2*<sup>HL</sup> offspring; and an IAP-*Pgm2*<sup>LL</sup> × IAP-*Pgm2*<sup>HL</sup> cross produced IAP-*Pgm2*<sup>LL</sup> and IAP-*Pgm2*<sup>HL</sup> offspring (Figure 1D). These results additionally demonstrate that the methylation state of one IAP-*Pgm2* variant does not influence the methylation state of the other in a heterozygous context.

### The methylation state of IAP-*Pgm2* is genetically determined

The stable inheritance of IAP-*Pgm2* methylation is indicative of a spontaneous genetic mutation in the B6 population, either in the IAP element itself or in a gene involved in its epigenetic regulation. To investigate the former possibility, we designed PCR primers amplifying the entirety of IAP-*Pgm2* from IAP-*Pgm2*<sup>HH</sup> and IAP-*Pgm2*<sup>LL</sup> gDNA (Figure 2A). Nested primer pairs N1 and N2 amplified two overlapping fragments, each containing half of the IAP-*Pgm2* element (Figure 2A). Agarose gel electrophoresis of the PCR products revealed amplification of both fragments from IAP-*Pgm2*<sup>HH</sup>



**Figure 1.** Intracisternal A-particle (IAP)-*Pgm2* methylation is tri-modally distributed and stably inherited in inbred B6 mice. (A) Map of CpG positions in the IAP-*Pgm2* long terminal repeats (LTRs). CpGs assayed in panel B are shown in red. (B) IAP-*Pgm2* methylation levels are consistent between the 5' and 3' ends. Methylation levels were quantified at the most distal CpGs of the IAP-*Pgm2* 5' and 3' LTRs (nearest to the boundary with unique DNA) using bisulphite pyrosequencing in ear DNA. Each point represents a CpG, each line represents an individual, and error bars represent standard deviations of technical triplicates. (C) Inter-individual methylation variation collapses within 500 bp on either side of IAP-*Pgm2*. Data presentation as in panel A. Assayed CpGs are shown in red above the graph. (D) Stable Mendelian inheritance of IAP-*Pgm2* methylation reveals that high (blue, H) and low (orange, L) methylation reflect two allelic states of IAP-*Pgm2*, with intermediate methylation (orange, I) representing heterozygosity. Each data point represents average methylation of ear DNA across the four most distal CpGs of the 5' LTR for one individual.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

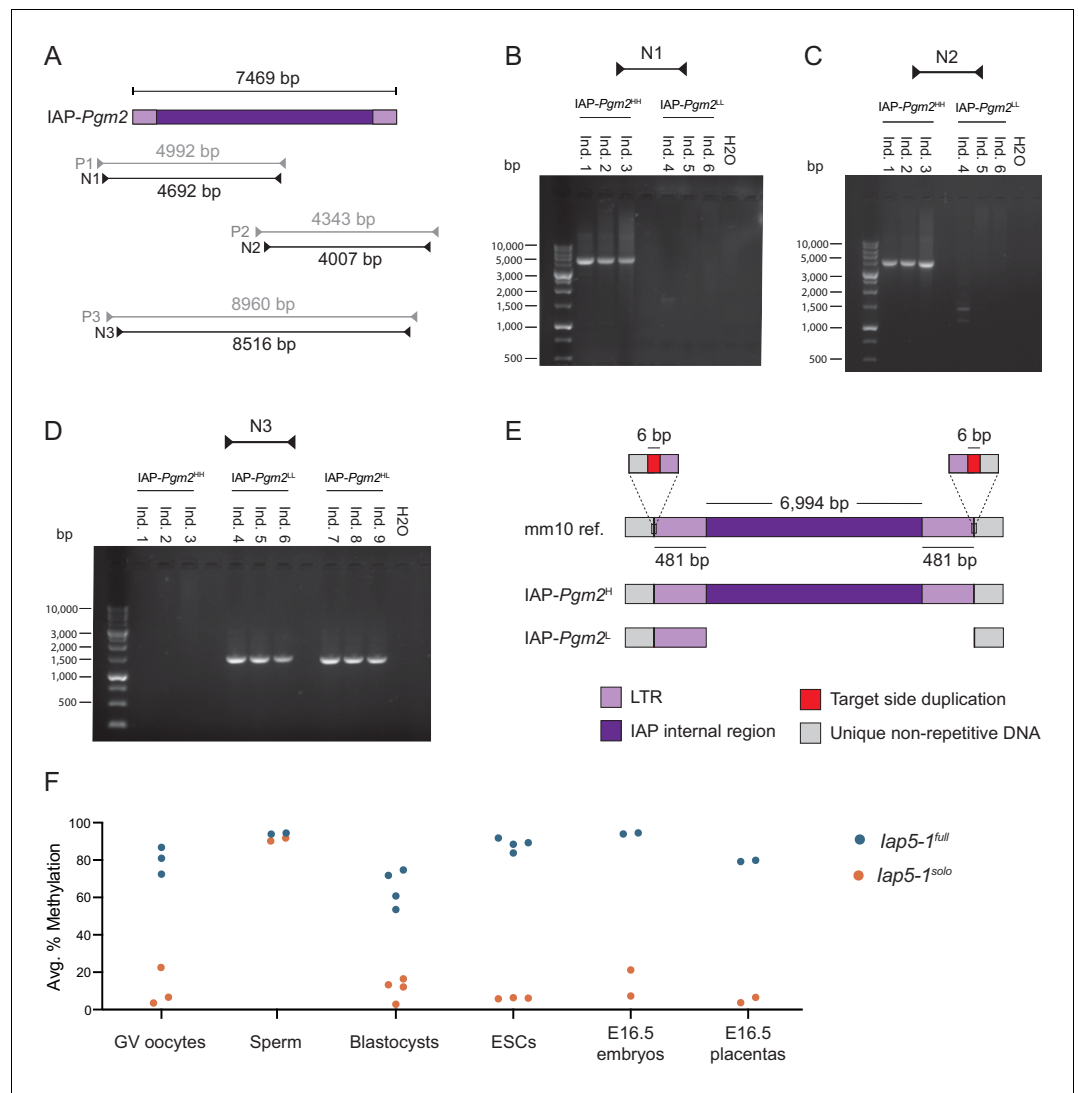
**Figure supplement 1.** Genomic structure and coding potential of intracisternal A-particle (IAP)-*Pgm2*.

**Figure supplement 2.** The intracisternal A-particle (IAP)-*Pgm2* methylation pattern is not sex-linked.

**Figure supplement 2—source data 1.** Numerical data represented in panels B-D, and **Figure 1—figure supplement 2.**

gDNA but no amplification of either fragment from IAP-*Pgm2*<sup>LL</sup> gDNA (**Figure 2B and C**), pointing to a substantial genetic difference between IAP-*Pgm2*<sup>HH</sup> and IAP-*Pgm2*<sup>LL</sup> individuals.

The nested primer pair N3 was designed to target the unique bordering regions on either side of the IAP element, amplifying an 8.5 kb fragment based on the GRC38/mm10 (mm10) genome assembly (**Figure 2A**). While no DNA bands were observed following the use of this primer pair on IAP-*Pgm2*<sup>HH</sup> gDNA, a 1.5 kb band was amplified from both IAP-*Pgm2*<sup>LL</sup> and IAP-*Pgm2*<sup>HL</sup> gDNA (**Figure 2D**). The smaller amplicon is indicative of a 7 kb deletion on the IAP-*Pgm2*<sup>L</sup> allele between the two primer annealing sites, while the lack of amplification from the IAP-*Pgm2*<sup>H</sup> allele is likely due to the technical challenges associated with amplifying a large repetitive DNA fragment.



**Figure 2.** The intracisternal A-particle (IAP)-*Pgm2<sup>L</sup>* allele is a solo long terminal repeat (LTR) formed via inter-LTR homologous recombination. (A) Map of PCR primer pairs P1–P3 and nested PCR primer pairs N1–N3 (not drawn to scale). (B and C) Agarose gel electrophoresis of PCR products amplified from three IAP-*Pgm2<sup>H/H</sup>* and three IAP-*Pgm2<sup>L/L</sup>* DNA samples using primer pairs N1 and N2. (D) As in panels B and C, but using primer pair N3 and including three IAP-*Pgm2<sup>H/L</sup>* DNA samples. (E) Schematic representation of the alignment of the IAP-*Pgm2* mm10 reference sequence and the assembled IAP-*Pgm2<sup>H</sup>* and IAP-*Pgm2<sup>L</sup>* sequences following Sanger sequencing (not drawn to scale). The IAP-*Pgm2<sup>L</sup>* solo LTR could have equivalently been shown aligned to the 3' LTR because the 5' and 3' LTRs have identical sequences. The base-resolution sequence of the new IAP-*Pgm2<sup>L</sup>* allele, renamed *lap5-1<sup>solo</sup>*, is available on GenBank (accession number: MW308129). (F) Both *lap5-1* variants are methylated in the male germline, with differential methylation re-established in early embryonic development. DNA methylation levels at the *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* alleles were quantified in oocytes, sperm, blastocysts, ESC lines, and E16.5 embryonic tail and placenta samples. Respective data points represent the following: 100 pooled oocytes, sperm collected from one male, pooled littermate blastocysts, one ESC line, and individual E16.5 embryos and placentas. All data points represent average DNA methylation across the four most distal CpGs of the 5' end of *lap5-1*. The online version of this article includes the following source data and figure supplement(s) for figure 2:

**Source data 1.** Numerical data represented in panel F.

**Figure supplement 1.** Possible mechanisms of homologous recombination between intracisternal A-particle (IAP) long terminal repeats (LTRs) leading to the formation of a solo LTR at the *lap5-1* locus.

**Figure supplement 2.** Differential DNA methylation between the *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* variants extends over the entire long terminal repeat(s) (LTR[s]).

To determine the location of the 7 kb deletion, we purified and sequenced the PCR products and aligned the assembled IAP-*Pgm2*<sup>H</sup> and IAP-*Pgm2*<sup>L</sup> sequences. The alignment revealed that the IAP-*Pgm2*<sup>H</sup> allele is an identical match to the full-length IAP-*Pgm2* sequence from the mm10 reference, while the IAP-*Pgm2*<sup>L</sup> allele is a solo LTR (**Figure 2E**). The IAP-*Pgm2*<sup>L</sup> solo LTR and the IAP-*Pgm2*<sup>H</sup> full-length IAP are both flanked by the same target site duplications (TSDs) and the IAP-*Pgm2*<sup>L</sup> solo LTR exhibits 100% sequence identity to both the 5' and 3' LTRs of the full-length IAP-*Pgm2*<sup>H</sup> (**Figure 2E**). Therefore, a recent inter-LTR homologous recombination event in the inbred B6 mouse strain gave rise to the solo LTR variant. In accordance with nomenclature guidelines from the International Committee on Standardized Genetic Nomenclature for Mice (ICSGNM), IAP-*Pgm2*<sup>H</sup> and IAP-*Pgm2*<sup>L</sup> are hereafter referred to as *lap5-1*<sup>full</sup> and *lap5-1*<sup>solo</sup>, respectively.

Inter-LTR recombination events occur when identical or near-identical LTR sequences engage in ectopic homologous recombination, resulting in the formation of solo LTRs (*Jern and Coffin, 2008*). These can occur both intra- and inter-chromosomally (**Figure 2—figure supplement 1**). Although we identified both *lap5-1* allelic variants in our own B6 colony, we also detected them in B6 samples sent to us from mainland Europe and North America (data not shown), so the recombination event most likely occurred at a B6-distributing facility. This type of proviral excision event is common: approximately half of the IAPs in the mouse genome are solo LTRs (*Nellåker et al., 2012*). However, the vast majority of the ~5,000 solo LTRs in the B6 genome are highly methylated (*Shimosuga et al., 2017*), making *lap5-1*<sup>solo</sup> unique from a regulatory perspective and raising questions regarding the functional consequences of a solo LTR left unmodified. Using clonal bisulphite sequencing, we confirmed that the entirety of the *lap5-1*<sup>solo</sup> solo LTR is unmethylated (bar the occasional methylated CpGs at the 5' end, consistent with the pyrosequencing data) while all CpGs in both the 5' and 3' LTRs of *lap5-1*<sup>full</sup> are highly methylated (**Figure 2—figure supplement 2**).

## Developmental dynamics of *lap5-1* methylation states

The experiments described thus far have focused on adult somatic *lap5-1* DNA methylation within and across generations. Given the dynamic nature of DNA methylation during mammalian development and considering previous reports on the resistance of IAPs to genome-wide methylation erasure (*Lane et al., 2003; Seisenberger et al., 2012*), we sought to examine and compare DNA methylation levels of the *lap5-1*<sup>full</sup> and *lap5-1*<sup>solo</sup> variants in the germline and during early embryonic development.

Germlinal vesicle (GV) oocytes and mature sperm were collected from *lap5-1*<sup>full</sup> and *lap5-1*<sup>solo</sup> adult B6 females and males. Oocytes collected from *lap5-1*<sup>full</sup> and *lap5-1*<sup>solo</sup> females were highly and lowly methylated at the *lap5-1* locus, respectively (**Figure 2F**). Thus, oocyte *lap5-1* methylation levels are reflective of somatic *lap5-1* methylation levels. In contrast, both variants were hypermethylated in sperm (**Figure 2F**), indicating that both alleles are targeted for repression during spermatogenesis by mechanism(s) that are distinct from those operating on *lap5-1* in the soma. Together, these experiments indicate that the maternal and paternal *lap5-1*<sup>solo</sup> alleles are differentially methylated upon fertilisation in the early zygote.

We examined the behaviour of *lap5-1* methylation levels during early development in blastocysts (embryonic day 3.5, E3.5) collected from the uteri of *lap5-1*<sup>full</sup> and *lap5-1*<sup>solo</sup> B6 females bred to *lap5-1*<sup>full</sup> and *lap5-1*<sup>solo</sup> B6 males, respectively. Blastocysts generated from *lap5-1*<sup>solo</sup> parents exhibited low methylation levels at *lap5-1*, suggesting that the paternally inherited hypermethylated *lap5-1*<sup>solo</sup> allele becomes demethylated shortly after fertilisation (**Figure 2F**). By comparison, blastocysts generated from *lap5-1*<sup>full</sup> parents exhibited methylation levels around 70% (**Figure 2F**). It is unclear whether *lap5-1*<sup>full</sup> is demethylated after fertilisation and rapidly methylated again by the blastocyst stage, or whether *lap5-1*<sup>full</sup> is generally resistant to epigenetic reprogramming during preimplantation development. Nonetheless, these results demonstrate that methylation patterns at *lap5-1*<sup>full</sup> and *lap5-1*<sup>solo</sup> are specified prior to implantation.

Despite the marked distinction in methylation states between *lap5-1*<sup>full</sup> and *lap5-1*<sup>solo</sup> blastocysts, *lap5-1*<sup>full</sup> methylation levels are lower at the blastocyst stage than in adult somatic tissue. It is possible that this incomplete methylation is symptomatic of lower methylation levels in the developing trophectoderm which counteract the higher methylation levels in the inner cell mass (ICM). In support of this, DNA methylation levels in embryonic stem (ES) cell lines derived from the ICM of *lap5-1*<sup>full</sup> and *lap5-1*<sup>solo</sup> blastocysts closely matched those observed in adult somatic tissues, with ES cells derived from *lap5-1*<sup>full</sup> blastocysts nearing 100% methylation (**Figure 2F**). In addition, we found that

*lap5-1<sup>full</sup>* methylation levels in E16.5 placentas were lower than those in E16.5 embryonic tissue (Figure 2F), consistent with reduced global DNA methylation in the placenta (Ehrlich et al., 1982; Schroeder et al., 2015). These results provide evidence for differential methylation between ICM- and trophoctoderm-derived lineages at the *lap5-1* locus. Of note, even though placental *lap5-1* methylation levels are less methylated compared to their embryonic counterparts, the two variants retain a pronounced difference in DNA methylation levels in this tissue.

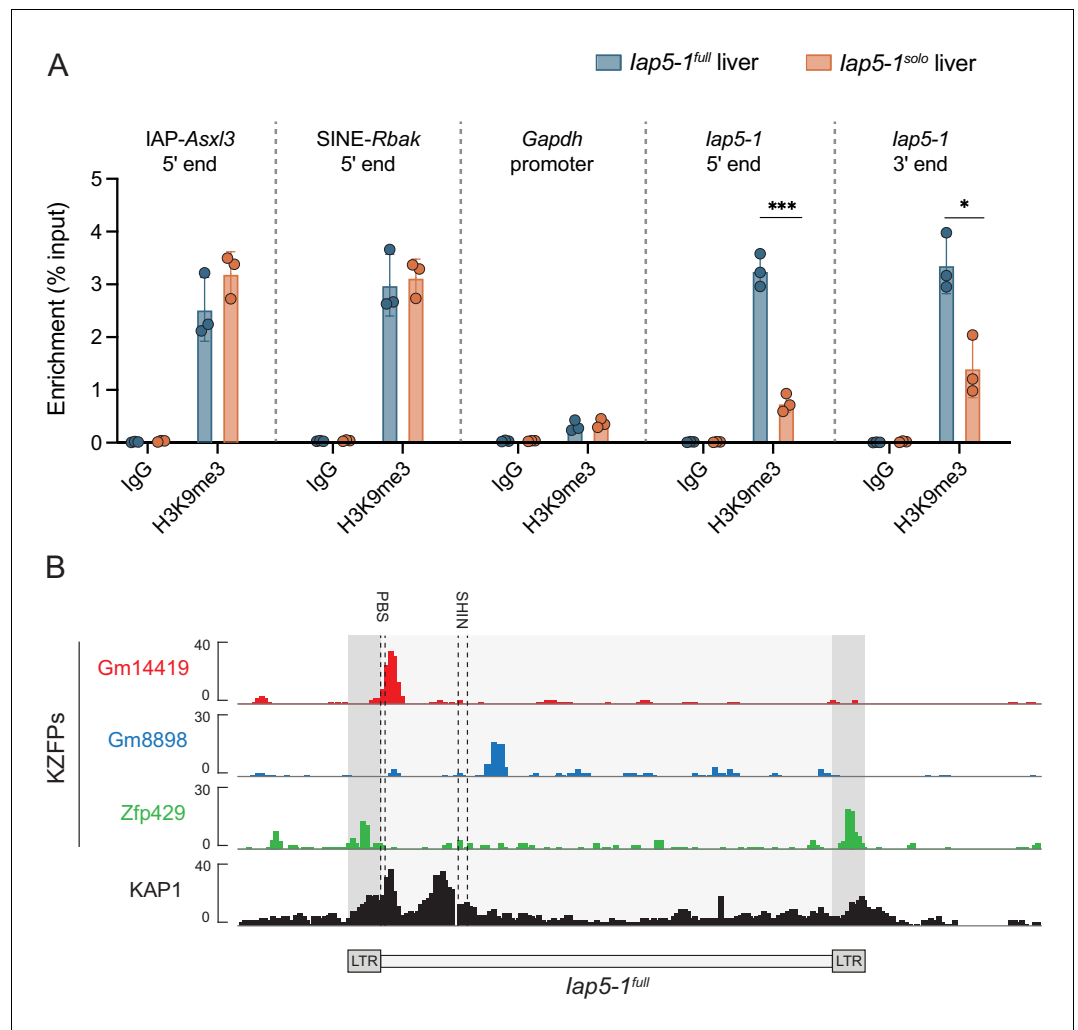
### Loss of H3K9me3 occupancy at *lap5-1<sup>solo</sup>*

To determine whether loss of DNA methylation at *lap5-1<sup>solo</sup>* is associated with changes in histone modifications, we quantified H3K9me3 enrichment in *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* adult liver samples via ChIP-qPCR. The retrotransposons IAP-*Asx13* and SINE-*Rbak*, used as positive controls, exhibited equivalent H3K9me3 enrichment between *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* individuals (Figure 3A). The *Gapdh* promoter was used as a negative control. Because the sequence of *lap5-1<sup>solo</sup>* is identical to that of the 5' and 3' LTRs of *lap5-1<sup>full</sup>*, the same primers were used to probe H3K9me3 enrichment at the borders of both variants. H3K9me3 enrichment at the 5' and the 3' borders was significantly decreased in *lap5-1<sup>solo</sup>* samples compared to *lap5-1<sup>full</sup>* samples (Figure 3A). *lap5-1<sup>full</sup>* showed comparable H3K9me3 levels to those observed at the positive controls (Figure 3A). We suggest that the slightly higher H3K9me3 enrichment at the 3' end compared to the 5' end in *lap5-1<sup>solo</sup>* samples is due to the presence of an ERV element immediately downstream of *lap5-1*.

Heterochromatin formation at mammalian TEs occurs in early development following their sequence-specific recognition by KZFPs. KZFPs recruit the scaffold protein KAP1, which in turn recruits the H3K9 methyltransferase SETDB1 as well as de novo DNA methyltransferases (Ecco et al., 2017). We reasoned that *lap5-1<sup>full</sup>* may be targeted for repression in the early embryo by KZFP(s) whose binding sites are located in the internal proviral portion of the IAP which is no longer present in the *lap5-1<sup>solo</sup>* variant. To explore this hypothesis, we analysed previously published ChIP-seq (chromatin immunoprecipitation followed by high-throughput sequencing) binding profiles of more than 60 murine KZFPs (Wolf et al., 2020) and identified two *lap5-1<sup>full</sup>*-binding candidates, Gm14419 and Gm8898 (Figure 3B). The ChIP-seq profiles for Gm14419 and Gm8898 in B6 ES cells displayed peaks immediately downstream and 1.5 kb downstream of the 5' LTR, respectively (Figure 3B). We observed KAP1 occupancy at the Gm14419 binding site, rendering Gm14419 a particularly promising candidate for future mechanistic research (Figure 3B). The Gm14419 peak overlaps with the primer binding site (PBS) just downstream of the 5' LTR (Figure 3B), a retroviral sequence often bound by KZFPs to effectuate repression (Wolf and Goff, 2007). Furthermore, it is likely that additional as yet unidentified KZFP(s) bind to *lap5-1<sup>full</sup>*, as evidenced by a second KAP1 peak in the *lap5-1<sup>full</sup>* internal region which does not overlap with the binding site of any of the KZFPs identified in our analysis (Figure 3B). This is consistent with the redundant nature of KZFP-mediated ERV repression (Imbeault et al., 2017; Wolf et al., 2020). In addition, we detected ZFP429 binding peaks at the 5' and 3' LTRs of *lap5-1<sup>full</sup>* (Figure 3B). Given the shared sequence between the *lap5-1<sup>full</sup>* LTRs and *lap5-1<sup>solo</sup>*, this observation suggests that ZFP429 does not recruit heterochromatin factors as effectively as other KZFPs. This is in line with our previous finding that ZFP429 is enriched at VM-IAPs compared to fully methylated IAPs of the same subclass (Bertozzi et al., 2020).

### Genetic background influences *lap5-1<sup>solo</sup>* methylation levels

Previous studies have shown that in some cases IAP methylation is modulated by genetic background (Bertozzi et al., 2020; Elmer and Ferguson-Smith, 2020; Rakyen et al., 2003; Wolff, 1971). To assess whether this is the case for *lap5-1* methylation, we carried out reciprocal crosses between B6 and wild-derived CAST/EiJ (CAST) mice (BC, B6 female × CAST male; CB, CAST female × B6 male). F1 hybrid offspring carry a single copy of *lap5-1* inherited from their B6 parent because the *lap5-1* insertion is absent from the CAST genome (Figure 4A). In line with our breeding experiments in pure B6 mice, hemizygous offspring born to a *lap5-1<sup>full</sup>* B6 parent were highly methylated and those born to a *lap5-1<sup>solo</sup>* B6 parent were lowly methylated (Figure 4B). However, although *lap5-1<sup>solo</sup>* methylation levels remained low in BC and CB F1 hybrids, they were significantly higher compared to those in pure B6 individuals (Figure 4C), suggesting that CAST-derived modifier(s) act on *lap5-1<sup>solo</sup>* in trans. In addition, CB offspring displayed higher and more variable *lap5-1<sup>solo</sup>*



**Figure 3.** Lack of DNA methylation at *lap5-1<sup>solo</sup>* is accompanied by a loss of H3K9me3 marks. (A) H3K9me3 ChIP-qPCR on *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* adult male livers. IAP-Asx1/3 and SINE-Rbak are positive control loci, the *Gapdh* promoter is a negative control locus, and the Rabbit IgG antibody serves as a negative isotype control in the two B6 populations. H3K9me3 enrichment was calculated using the per cent input method and compared between genotypes using unpaired t-tests (\* $p < 0.05$ ; \*\*\* $p < 0.0005$ ). Error bars represent standard deviations of biological replicates. (B) *lap5-1<sup>full</sup>* is bound by multiple KZFPs. Publicly available ChIP-seq data sets indicate that KZFPs Gm14419 and Gm8898 are capable of binding the internal region of *lap5-1<sup>full</sup>* and that KZFP Zfp429 is capable of binding the long terminal repeat (LTR) sequence shared by both *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>*. The primer binding site (PBS) and the short heterochromatin inducing sequence (SHIN), both frequently bound by KZFPs, are shown with dashed lines. ChIP-seq data sets were downloaded from the GEO database (accession numbers: Gm14419, GSM3173720; Gm8898, GSM3173728; Zfp429, GSM3173732; KAP1, sum of GSM3173661, GSM3173662, and GSM3173663).

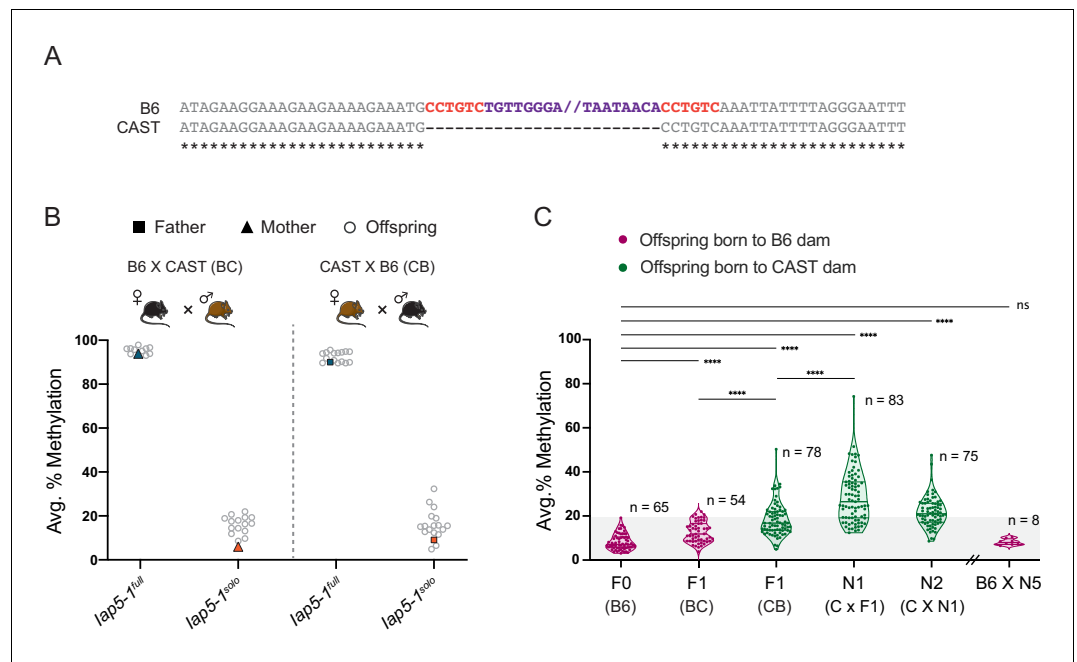
The online version of this article includes the following source data for figure 3:

**Source data 1.** Numerical data represented in panel A.

methylation levels compared to BC offspring, indicative of a genetic background-specific maternal effect similar to those recently reported at VM-IAPs (Bertozzi et al., 2020).

We further interrogated the strain-specific maternal effect by backcrossing *lap5-1<sup>solo</sup>* CB hybrid males to pure CAST females. This resulted in a cumulative effect, whereby N1 *lap5-1<sup>solo</sup>* mice showed even higher and more variable methylation levels compared to F1 *lap5-1<sup>solo</sup>* mice (Figure 4C). Therefore, the CAST content of the inherited paternal genome and passage through a CAST egg may compound each other. The subsequent N2 backcrossed generation did not cause a further increase in methylation, and backcrossing N5 males to B6 females resulted in a reversion to





**Figure 4.** Genetic background influences *lap5-1<sup>solo</sup>* methylation levels. (A) B6 and CAST sequence alignment of the *lap5-1* insertion site reveals that the *lap5-1* element (purple) and its associated target site duplication (TSD) sequence (red) are not present in the CAST genome. Asterisks indicate nucleotide conservation. The internal portion of the B6 sequence was omitted; nothing was omitted from the CAST sequence. (B) Reciprocal F1 hybrids uncover a variant-specific genetic background effect. *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* B6 females were crossed to CAST males (BC, left); *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* B6 males were crossed to CAST females (CB, right). DNA methylation levels were quantified in ear samples collected from hemizygous offspring. Each data point represents average DNA methylation across the four distal CpGs of the 5' end of *lap5-1* for one individual. (C) The genetic background-specific maternal effect at *lap5-1<sup>solo</sup>* is strengthened upon backcrossing to a CAST female and lost following backcross to a B6 female. The N1 generation was produced by crossing F1 CB males to CAST females, and the N2 generation was generated by crossing N1 males harbouring the *lap5-1<sup>solo</sup>* allele to CAST females. After five generations of backcrossing to CAST, N5 males were crossed to B6 females. Offspring that did not inherit the *lap5-1<sup>solo</sup>* allele were not included in the analysis. Thick and thin lines in the violin plots designate the median and distribution quartiles, respectively. Sample sizes are shown on the graph. Grey shaded area represents the range of methylation levels observed in B6 individuals. Statistics: Welch's ANOVA test followed by Games-Howell's post hoc multiple comparison test (\*\*\*\*p<0.0001; ns: not significant).

The online version of this article includes the following source data and figure supplement(s) for figure 4:

**Figure supplement 1.** DNA methylation at *lap5-1<sup>solo</sup>* is unresponsive to maternal dietary methyl supplementation.  
**Figure supplement 1—source data 1.** Numerical data represented in panels B, C, and **Figure 4—figure supplement 1**.

the original B6 methylation state (**Figure 4C**). These results are reminiscent of strain-specific behaviours of transgene methylation reported decades ago and support a role for strain-specific oocyte factors (possibly polymorphic KZFPs) in driving these genetic-epigenetic interactions (**Allen et al., 1990; Bertozzi et al., 2020; Kearns et al., 2000**).

### *lap5-1<sup>solo</sup>* methylation is not affected by maternal dietary methyl supplementation

The IAPs located at the *A<sup>vy</sup>* and *Axin<sup>Fu</sup>* VM-IAPs exhibit increased DNA methylation following gestational methyl supplementation, resulting in shifts in the associated coat colour and tail morphology phenotypes, respectively (**Waterland et al., 2006; Waterland and Jirtle, 2003**). However, it remains unclear whether susceptibility to this environmental exposure is conferred by *variable* or by *incomplete* DNA methylation at these epialleles (or neither). The unmethylated *lap5-1<sup>solo</sup>* variant provides an opportunity to test the latter.

To examine whether *lap5-1<sup>solo</sup>* is responsive to methyl supplementation, *lap5-1<sup>solo</sup>* females were put on a methyl-supplemented diet 2 weeks prior to mating and were kept on the same diet throughout pregnancy and lactation. Control females were fed standard chow throughout and all pups in the experiment were weaned onto standard chow. DNA methylation levels at the *lap5-1<sup>solo</sup>* LTR were quantified in 8-week-old offspring liver samples. Unlike *A<sup>vy</sup>* and *Axin<sup>Fu</sup>* loci, *lap5-1<sup>solo</sup>* remained unmethylated in the methyl-supplemented offspring (**Figure 4—figure supplement 1**), indicating that complete lack of methylation at IAPs does not go hand-in-hand with susceptibility to dietary methyl supplementation.

### ***lap5-1* variants influence adjacent gene expression in a tissue-specific manner**

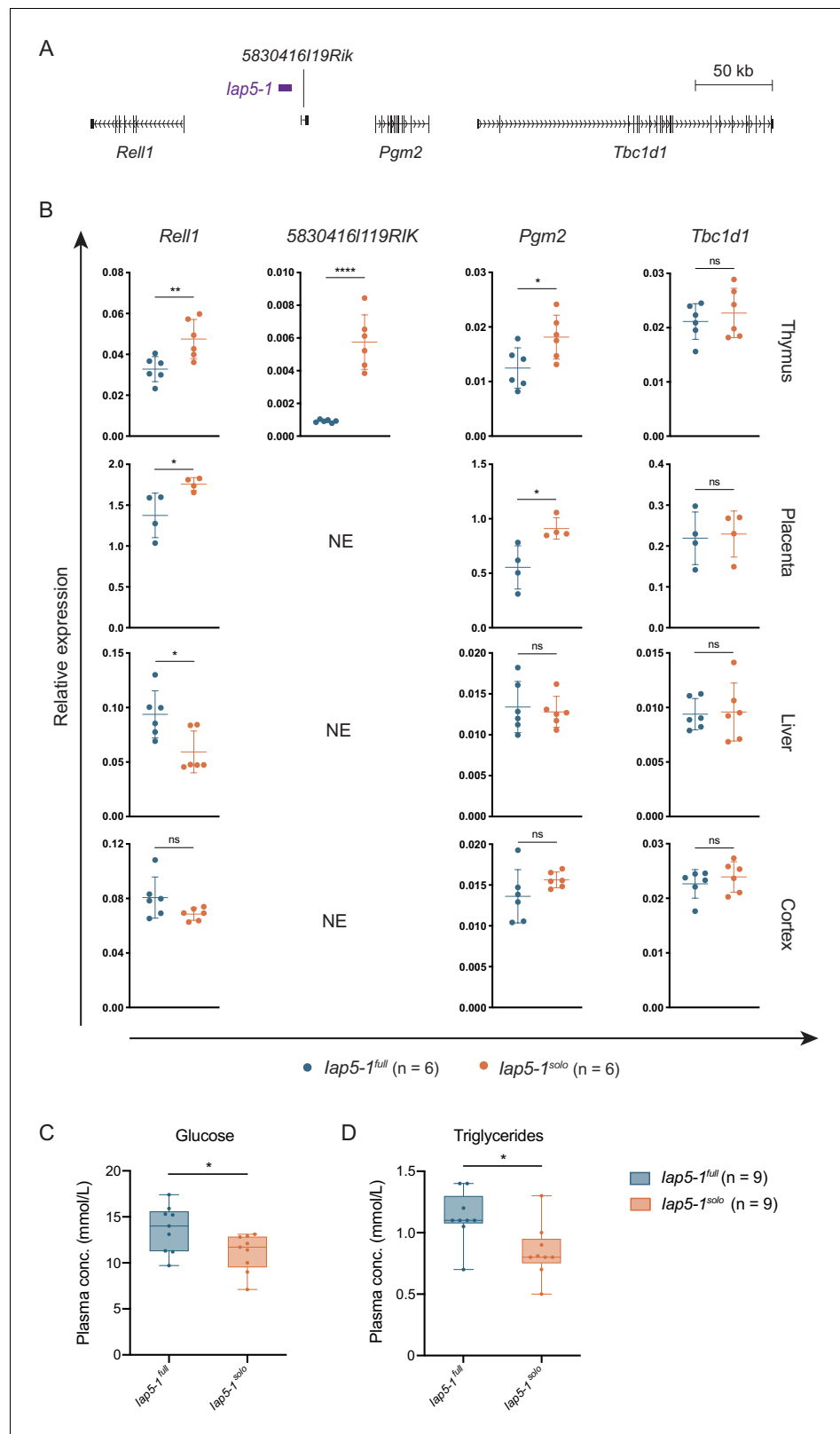
IAP insertions can influence neighbouring gene expression. Intergenic IAPs can induce the formation of chimeric transcripts initiated at the promoter in the IAP LTR and may also act as enhancers (**Gagnier et al., 2019**). These effects are sometimes dependent on the epigenetic properties of the IAP, as illustrated by the *A<sup>vy</sup>* and *Axin<sup>Fu</sup>* IAPs which modulate *Agouti* or *Axin* expression in a methylation-dependent manner.

We asked whether allelic variation at the *lap5-1* locus influences the expression of neighbouring genes by quantifying expression of the four closest genes (**Figure 5A**) in liver, cortex, thymus, and placental tissues collected from *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* individuals. Protein-coding genes *Pgm2*, *TBC1 domain family, member 1* (*Tbc1d1*), and *RELT-like protein 1* (*Rell1*) were expressed in all tissues examined; long non-coding RNA (lncRNA) *5830416119Rik* transcripts were only detected in the thymus. *Pgm2* and *Rell1* expression levels in the thymus and placenta were significantly higher in *lap5-1<sup>solo</sup>* than in *lap5-1<sup>full</sup>* individuals, indicating that in these tissues the unmethylated solo LTR is associated with increased expression (**Figure 5B**). *Tbc1d1*, the furthest in distance from *lap5-1*, did not display significant differences in expression in any of the tested tissues, suggesting that proximity to the *lap5-1* locus is predictive of its transcriptional effect (**Figure 5B**). *5830416119Rik* expression in the thymus was barely detected in *lap5-1<sup>full</sup>* samples, showing a highly significant increase in *lap5-1<sup>solo</sup>* samples (**Figure 5B**). *Rell1* was the only gene to show a significant difference in expression in liver tissue and the directionality of the effect was inverted, with lower expression levels observed in *lap5-1<sup>solo</sup>* individuals (**Figure 5B**). No significant differences in expression were observed for any of the genes in cortex samples (**Figure 5B**). In thymus and cortex, however, we detected transcription of the unique non-coding regions bordering *lap5-1<sup>solo</sup>*, which was not observed for the equivalent regions bordering *lap5-1<sup>full</sup>* (**Figure 5—figure supplement 1**). Together, these data show that the *lap5-1* polymorphism is associated with tissue-specific altered adjacent gene expression.

### **The formation of the *lap5-1<sup>solo</sup>* variant is associated with metabolic phenotypes**

Little is known about the biological functions of *Rell1* and *5830416119Rik*, but *Pgm2* is better characterised. *Pgm2* is the closest coding gene to *lap5-1*, lying 55 kb downstream of the 3' LTR. PGM2 catalyses the interconversion between glucose 1-phosphate and glucose 6-phosphate and has a secondary role to that of the predominant PGM isozyme, PGM1 (**Geer et al., 2010**). PGM1 deficiency in humans is associated with glycogen storage disease and a congenital disorder of glycosylation (**Beamer, 2015**). PGM2 has been associated with metabolic disease in GWAS studies (**Timmons et al., 2018**).

Considering the established role of *Pgm2* in metabolic pathways, we investigated potential metabolic effects resulting from the formation of the *lap5-1<sup>solo</sup>* allele. We screened plasma samples collected from *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* adult males for a range of metabolic biomarkers. While most assays revealed no difference between the two variants (**Figure 5—figure supplement 2A**), we found that fasting plasma glucose and triglyceride concentrations were significantly lower in *lap5-1<sup>solo</sup>* mice (**Figure 5C**). Glucose tolerance tests did not show additional metabolic disparities between the two variants (**Figure 5—figure supplement 2B**). These findings suggest that the emergence of a derepressed solo LTR near a gene involved in glucose metabolism has phenotypic implications, providing a basis for further functional characterisation of this model.



**Figure 5.** Functional consequences of the genetically induced epiallele at the *lap5-1* locus. **(A)** Scaled map of the genes surrounding *lap5-1*. Gene transcripts were extracted from the UCSC Genome Browser ([Haussler et al., 2019](#)). **(B)** The *lap5-1* polymorphism influences neighbouring gene expression in a tissue-specific manner. Expression of *Rell1* (exon 7), 5830416/119RIK (exon 3), *Pgm2* (exon 9), and *Tbc1d1* (exon 3) was quantified in *lap5-*

Figure 5 continued on next page

Figure 5 continued

$1^{full}$  and  $lap5-1^{solo}$  thymus, placenta, liver, and brain tissues via RT-qPCR (NE: not expressed). Relative expression was normalised to *Hprt1* and  $\beta$ -actin expression and calculated using the  $\Delta$ Ct method. Means and standard deviations of biological replicates are shown in the graphs. (C and D) Plasma glucose (C) and triglyceride (D) concentrations in  $lap5-1^{full}$  and  $lap5-1^{solo}$  adult males. Box plots show the distribution quartiles and median. Statistics: unpaired t-tests (\* $p < 0.05$ ; \*\* $p < 0.005$ ; \*\*\* $p < 0.0001$ ; ns: not significant).

The online version of this article includes the following source data and figure supplement(s) for figure 5:

**Source data 1.** Numerical gene expression data represented in **Figure 5B** and **Figure 5—figure supplement 1B**.

**Source data 2.** Numerical data for metabolic experiments shown in **Figure 5C and D** and **Figure 5—figure supplement 2**.

**Figure supplement 1.** Bordering genomic regions are transcribed in  $lap5-1^{solo}$  tissues.

**Figure supplement 2.** Metabolic phenotyping of  $lap5-1^{full}$  and  $lap5-1^{solo}$  individuals.

## Discussion

Historically, inter-LTR recombination events have been identified due to phenotypic reversions of ERV-induced mutations, whereby the phenotypic effect of the insertion is reversed following proviral excision and solo LTR formation (Bultman et al., 1994; Seperack et al., 1988; Stoye et al., 1988). In this study, we identified one such chromosomal event based on its epigenetic rather than phenotypic outcome, revealing *lap5-1* as a genetically induced epiallele in the ostensibly isogenic B6 mouse strain. We demonstrated that the  $lap5-1^{full}$  and  $lap5-1^{solo}$  variants display distinct DNA and H3K9 methylation profiles which are associated with differential adjacent gene expression and altered metabolism, establishing the *lap5-1* locus as a valuable endogenous system to study the mechanisms, evolution, and functional implications of TE repression in both the germline and soma.

The unmethylated status of  $lap5-1^{solo}$  sets it apart from other solo LTRs in the B6 genome, the vast majority of which (>93%) are highly methylated (Shimosuga et al., 2017). Shimosuga and colleagues quantified DNA methylation at more than 8000 B6 IAP LTRs and only 14 exhibited less than 20% methylation in tail samples. Ten of these were solo LTRs, mostly of the same subtype as the *lap5-1* LTR (IAPLTR2\_mm). While the authors classified the *lap5-1* LTR as a hypomethylated 3' LTR of a full-length IAP, in all likelihood they unknowingly detected the hypomethylated  $lap5-1^{solo}$  allele in their particular samples. It is conceivable that some of the other documented hypomethylated LTRs of full-length IAPs are in fact solo LTRs formed from recent recombination events. In comparing tail and sperm DNA methylation, the same study showed that individual hypomethylated IAPs in either tail or sperm were non-overlapping. This is consistent with the hypermethylation of  $lap5-1^{solo}$  that we observed in sperm and highlights the divergence of TE silencing mechanisms between the soma and the germline.

Ancestral variants of spontaneous mutations in inbred mouse strains are easily lost from the population. The fortuitous timing of our discovery and the resultant availability of both  $lap5-1^{full}$  and  $lap5-1^{solo}$  mice allowed us to quantitatively explore the functional repercussions of a derepressed solo LTR. We showed that the  $lap5-1^{solo}$  variant is associated with increased expression of adjacent genes in thymic and placental tissue. This may reflect the use of newly accessible regulatory sequences in the solo LTR or, alternatively, a secondary effect stemming from heterochromatin loss at a neighbouring region. The transcriptional consequences were tissue-specific: *Rel1* expression was lower in  $lap5-1^{solo}$  compared to  $lap5-1^{full}$  livers, and no gene expression differences were observed in cortex samples. The lncRNA *5830416119Rik*, identified from the sequencing of adult male thymus and aorta cDNA libraries (Kawai et al., 2001), was only expressed in thymus samples collected from  $lap5-1^{solo}$  individuals; it was nearly undetectable in  $lap5-1^{full}$  thymic samples and not expressed at all in any of the other tested tissues. We speculate that the formation of  $lap5-1^{solo}$  gave rise to *5830416119Rik* as a novel tissue-specific (perhaps immune-specific) lncRNA. Indeed, IAP transcripts are rarely detected in somatic tissues outside of tumours and early embryogenesis, but thymic and activated splenic cells are notable exceptions (Kuff and Lueders, 1988). In addition, young ERVs have recently been implicated in shaping the evolution of transcriptional networks underlying innate immunity (Chuong et al., 2016; Tie et al., 2018; Ye et al., 2020), and our group recently identified individual IAP elements that exhibit inter-individual methylation variability exclusively in B cells (Elmer et al., 2021).

Along with the transcriptional differences observed at the metabolic gene *Pgm2*, we found that *lap5-1<sup>solo</sup>* individuals exhibit lower plasma glucose and triglyceride concentrations compared to *lap5-1<sup>full</sup>* individuals. While the effect sizes were small, we note that these measurements were taken in an unaltered environment. Administration of a long-term metabolic challenge such as a high-fat diet and subsequent metabolic phenotyping will aid in further elucidating the biological significance of these results. Nevertheless, our work adds to the body of work suggesting that TE insertions and their epigenetic modification can have quantifiable consequences on host metabolic pathways (Du et al., 2016; Kuehnen et al., 2012; Scherneck et al., 2009; Yen et al., 1994).

We have shown that *lap5-1<sup>solo</sup>* is susceptible to genetic background effects, whereby passage through a CAST egg promotes methylation of the B6-derived solo LTR. In a recent study we reported the same effect at a number of VM-IAPs and showed that KZFP diversification is associated with strain-specific IAP methylation (Bertozzi et al., 2020). This combined with the identification of candidate KZFPs capable of binding the *lap5-1* variants suggests that the differential recruitment of heterochromatin factors by KZFPs is involved in both driving the epigenetic differences between *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>*, and for the strain-specific methylation of *lap5-1<sup>solo</sup>*. In fact, by definition, the wide range of methylation levels observed at the *lap5-1<sup>solo</sup>* allele across genetically identical CB individuals renders this locus a bona fide VM-IAP (or *metastable epiallele*) in this hybrid context (Bertozzi and Ferguson-Smith, 2020; Rakyen et al., 2002). Therefore, we expect the new *lap5-1<sup>solo</sup>* variant to be a useful tool for the study of epigenetic stochasticity in the absence of genetic variation, a phenomenon for which the underlying mechanisms remain poorly understood.

The identification of *lap5-1<sup>solo</sup>* demonstrates that cryptic genetic diversity in an inbred mouse population can have functional repercussions with important implications for experimental outcomes. It serves as a cautionary tale for researchers working with inbred mouse colonies, particularly in the field of epigenetics where ruling out genetic effects is often paramount. We are reminded that the current mouse reference genome harbours large gaps and inaccuracies, largely due to mapping difficulties associated with the repeat genome. Inter-LTR recombination events are surely an underappreciated source of genetic variation considering that a lack of uniquely mapped reads in internal portions of TEs is more likely to be attributed to technical rather than biological limitations. A recent study in humans developed a computational pipeline to capture dimorphic human ERVs (HERVs) resulting from inter-LTR recombination events and detected dozens of previously unidentified candidates (Thomas et al., 2018). The advent of such analytical tools as well as routine long-read sequencing of whole genomes will be highly beneficial in addressing these issues.

In summary, we have identified and characterised a recent spontaneous inter-LTR recombination event at the *lap5-1* locus, introducing a genetic variant in the commonly investigated and supposedly inbred B6 mouse strain. The ancestral variant, *lap5-1<sup>full</sup>*, is a full-length IAP repressed by DNA methylation and H3K9 trimethylation, as is typical for this class of evolutionary young ERV. The recently formed solo LTR variant, *lap5-1<sup>solo</sup>*, lacks these silencing marks and is associated with metabolic changes and differential neighbouring gene expression. Our study lays the foundation for further comparative studies between the *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* variants aimed at better understanding the epigenetic and functional consequences of TE structural variation and evolution.

## Materials and methods

### Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain, strain background ( <i>Mus musculus</i> )	C57BL/6J	The Jackson Laboratory via Charles River	000664 (JAX) 632 (Charles River)	
Strain, strain background ( <i>Mus musculus</i> )	CAST/EIJ	MRC Harwell Institute	FESA:773	

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Cell line ( <i>Mus musculus</i> )	C57BL6/J ES cell lines	This paper		See Materials and methods, Generation of ESC lines.
Antibody	Rabbit polyclonal anti-H3K9me3	Active Motif	RRID:AB_2532132	ChIP (2.5 µl/30 mg liver tissue)

## Mice

Mouse work was carried out in accordance with the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body (Home Office project license # PC213320E). C57BL/6J (RRID:IMSR\_JAX:000664) and CAST/EiJ (RRID:IMSR\_JAX:000928) mice were obtained from Charles River and the MRC Harwell Institute, respectively, and maintained under a 12 hr light–dark cycle in temperature- and humidity-controlled conditions. Mice were fed a standard chow diet (RM3(E); Special Diet Services) ad libitum unless otherwise noted. For pure and hybrid breeding experiments, mice were mated at 8–12 weeks of age and 10 day old pups were ear notched for DNA methylation quantification. Male and female offspring from multiple litters per breeding pair were included in the analyses.

## Dietary methyl supplementation

Adult *lap5-1<sup>solo</sup>* females were randomly placed on control diet (RM3(E); Special Diet Services) or methyl supplemented diet (RM3(E) supplemented with 15 g Choline, 15 g Betaine, 15 mg Folic acid, 1.5 mg Vitamin B12, 7.5 g L-methionine, 150 mg Zinc; Special Diet Services) 2 weeks prior to mating with *lap5-1<sup>solo</sup>* males and kept on their respective diets throughout pregnancy and lactation. Eight dams were used for each dietary group. The methyl supplemented diet recipe matches the 3SZM diet in *Wolff et al., 1998*. All offspring were weaned onto the control diet and culled at 8 weeks of age. DNA methylation was quantified at the distal CpGs of the 5' end of *lap5-1<sup>solo</sup>* in all offspring livers (control diet: 38 born from eight litters; methyl supplemented diet: 41 born from eight litters). Statistics were carried out on litter averages because the dietary intervention was done on the dams, not the offspring.

## Metabolic biomarker assays

Blood was sampled from 4-month-old *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* males via cardiac puncture following isoflurane-induced deep terminal anaesthesia. Blood samples were placed in EDTA-coated tubes and centrifuged at 1500 × g for 15 min at 4°C. Plasma was collected from the separated upper phase and stored at –80°C before sending to the Core biochemical assay laboratory (CBAL) at the Cambridge University Hospitals for analysis. Assays performed on the plasma samples included adiponectin (ng/ml), albumin (g/l), ALT (U/l), cholesterol (mmol/l), corticosterone (ng/ml), glucagon (pg/ml), glucose (mmol/l), HDL (mmol/l), insulin (µg/l), LDL (mmol/l), leptin (pg/ml), NEFA (µmol/l), and triglycerides (mmol/l).

## Glucose tolerance tests

Glucose tolerance tests were conducted on *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* 16-week-old males (n = 12 per genotype). Following an overnight fast of 15 hr, mice were weighed and baseline blood glucose levels (time 0) were measured using the Glucomen Areo glucometer. A glucose dosage of 2 g/kg was calculated for each mouse based on weight and administered via intraperitoneal injection. Tail blood glucose levels were measured 15, 30, 45, 60, 90, and 120 min post injection. Blood glucose levels were compared between *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* mice at each time point using multiple unpaired t-tests corrected for multiple comparisons using the Holm-Šidák method.

## Isolation and storage of biological samples

### Somatic tissues

Liver, cortex, and thymus tissues dissected from adult *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* males were snap frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  before use. Ear notches were stored at  $-20^{\circ}\text{C}$ .

GV oocytes. *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* females were injected interperitoneally with 5 IU of gonadotropin. Ovaries were dissected 40 hr later and placed in M2 medium supplemented with 0.06 g/l potassium penicillin-G, 0.05 g/l streptomycin sulphate, and 240  $\mu\text{M}$  dbcAMP. Ovary follicles were punctured and GV oocytes were detached from their associated cumulus cells using a mouth aspirator and a pulled capillary tube. GV oocytes were stored at  $-80^{\circ}\text{C}$  in pools of 100 oocytes collected from multiple females with matching *lap5-1* genotype.

### Sperm

Mature sperm was isolated from the cauda epididymides of fertile adult *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* males as previously described (Sharma et al., 2015) and stored in  $1\times$  PBS at  $-80^{\circ}\text{C}$ . Sperm samples from different males were analysed separately.

### Blastocysts

*lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* females were mated with *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* males, respectively. The uteri and oviducts of pregnant females were dissected 3 days after the identification of a vaginal plug and flushed with M2 medium (Sigma-Aldrich) supplemented with 0.06 g/l potassium penicillin-G and 0.05 g/l streptomycin sulphate. Embryos developed to the blastocyst stage were washed in M2 medium and stored at  $-80^{\circ}\text{C}$ . Blastocysts collected from each female were pooled prior to freezing.

### E16.5 embryos and placentas

*lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* females were mated with *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* males, respectively. Sixteen days after the identification of a vaginal plug, the uterus was excised from the abdominal cavity and individual E16.5 embryos were removed and rinsed in cold  $1\times$  PBS. After removal of the yolk sac and an additional rinse in cold  $1\times$  PBS, embryonic tails and full placentas were flash frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

## DNA extraction

For somatic tissues and ES cells, samples were treated with RNase A at  $37^{\circ}\text{C}$  for 60 min and digested with Proteinase K at  $55^{\circ}\text{C}$  overnight in lysis buffer (10 mM EDTA, 150 mM NaCl, 10 mM Tris-HCl pH 8, 0.1% SDS). gDNA was isolated the next day using a standard phenol-chloroform extraction and ethanol precipitation protocol. The same protocol was followed for sperm gDNA extraction except that the Proteinase K digestion was carried out in equal volumes of Solution A (75 mM NaCl pH 8; 25 mM EDTA) and Solution B (10 mM Tris-HCl pH 8; 10 mM EDTA; 1% SDS; 80 mM DTT) following centrifugation and removal of PBS from thawed sperm. To extract oocyte and blastocyst gDNA, pooled GV oocytes or blastocysts were incubated for 1 hr at  $37^{\circ}\text{C}$  in 14  $\mu\text{l}$  of ddH<sub>2</sub>O, 1  $\mu\text{l}$  of 1 mg/ml Carrier RNA (QIAGEN), 1  $\mu\text{l}$  of 10% SDS, and 1  $\mu\text{l}$  of 10 mg/ml Proteinase K. After a 15 min incubation at  $98^{\circ}\text{C}$ , samples were directly bisulphite-converted as described below.

## Generation of ESC lines

ESC lines were generated as previously described (Nichols and Jones, 2017) with the following modifications. The concentration of MEK inhibitor PDO325901 used in the KSOM+2i and N2B27+2i +LIF media was reduced to 0.2  $\mu\text{M}$  for the initial stages of the protocol and increased to 1  $\mu\text{M}$  following disaggregation of the ICM. The zona pellucide was removed from unhatched embryos using a 10 min pronase digestion at  $37^{\circ}\text{C}$  rather than using acidic Tyrode's solution, and rat serum was used as a source of complement instead of guinea pig serum. Laminin-coated wells were used to ensure proper cell attachment and Accutase solution was used to detach cells prior to passaging. Only male embryos were kept for the generation of ESC lines. Established lines were screened for mycoplasma contamination using the PCR Mycoplasma Test Kit I/C (PromoCell).

## Genotyping

Ear notches were used for *lap5-1* allelic variant genotyping. Ear notch gDNA was extracted using the PCR BIO Rapid Extract lysis kit (PCR Biosystems) and 1  $\mu$ l of 1:10 diluted DNA was used as a template for PCR using the REDTaq ReadyMix PCR Reaction Mix (Sigma-Aldrich). The PCR conditions were as follows: (1) 95°C for 4 min 30 s; (2) 94°C for 30 s, optimised T°C for 30 s, 72°C for 30 s, 40 cycles; (3) 72°C for 5 min. For trophoctoderm sex-genotyping, trophoctoderm lysates were placed in PCR buffer with Proteinase K (50 mM KCl, 10 mM Tris-HCl pH 8.3, 2.5 mM MgCl<sub>2</sub>, 0.1 mg/ml gelatin, 0.45% NP40, 0.45% Tween 20, 200  $\mu$ g/ml Pro K) and incubated at 55°C for 1 hr followed by 95°C for 10 min. Trophoctoderm gDNA samples were sex-genotyped by PCR using HotStarTaq DNA Polymerase (QIAGEN) and in the following conditions: (1) 95°C for 3 min; (2) 94°C for 30 s, 56°C for 30 s, 72°C for 55 s, 40 cycles; (3) 72°C for 5 min. Amplified DNA was evaluated by agarose gel electrophoresis. Genotyping primers for *lap5-1* allelic variants and the Y-linked *Sry* gene are listed in [Supplementary file 1](#).

## Sanger sequencing

The PCR primer pairs P1, P2, and P3 were designed to amplify the 5' half, 3' half, and the entirety of IAP-*Pgm2* prior to Sanger sequencing, respectively ([Figure 2A](#), [Supplementary file 1](#)). PCR amplification was carried out using the Expand Long Template PCR System (Roche) with the following thermocycler conditions: (1) 94°C for 2 min; (2) 94°C for 10 s, optimised T°C for 30 s, 68°C for 6 min, 10 cycles; (3) 94°C for 15 s, optimised T°C for 30 s, 68°C for 6 min + 20 s each successive cycle, 20 cycles; (4) 68°C for 7 min. Nested PCRs with primer pairs N1, N2, and N3 were performed using HotStarTaq DNA Polymerase (QIAGEN) and a 1:20 dilution of the first PCR products as templates ([Figure 2A](#), [Supplementary file 1](#)). Conditions for the nested PCRs were as follows: (1) 95°C for 3 min; (2) 94°C for 30 s, optimised T°C for 30 s, 72°C for 55 s, 40 cycles; (3) 72°C for 5 min. Nested PCR products were purified by gel extraction using the QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions and DNA was eluted in ddH<sub>2</sub>O. Sanger sequencing was carried out by Source BioScience. Sequencing primers were interspersed regularly across the IAP element ([Supplementary file 1](#)). Sequence traces were visually examined, and reliable sequences were merged using the EMBOSS *merger* tool. The resulting sequences for the two *lap5-1* alleles were aligned to the GRCm38/mm10 reference sequence using CLC Sequence Viewer 6. The base-resolution sequence for the *lap5-1<sup>olo</sup>* allele has been uploaded to GenBank (accession number: MW308129).

## Bisulphite pyrosequencing

Bisulphite conversions were carried out using the two-step modification procedure of the Imprint DNA Modification Kit (Sigma-Aldrich) according to the manufacturer's instructions. 1  $\mu$ g gDNA was used per conversion with the exception of the oocyte and blastocyst experiments. PyroMark Assay Design SW 2.0 software (QIAGEN) was used to design the pyrosequencing assays (primers listed in [Supplementary file 1](#)). Target regions were PCR-amplified in technical triplicates from bisulphite-converted DNA using a biotinylated forward or reverse primer and HotStarTaq DNA Polymerase (QIAGEN). PCR conditions were as follows: (1) 95°C for 3 min; (2) 94°C for 30 s, optimised T°C for 30 s, 72°C for 55 s, 40 cycles; (3) 72°C for 5 min. For low-input pyrosequencing of oocyte and blastocyst DNA, two rounds of PCRs were performed: the first PCR used non-biotinylated primers and 20 amplification cycles (other conditions remained the same); the second PCR used 1  $\mu$ l of the product from the first PCR as template as well as a biotinylated forward or reverse primer. Following PCR, Streptavidin Sepharose High Performance beads (GE healthcare) were bound to the product in binding buffer (10 mM Tris-HCl pH 7.6, 2 M NaCl, 1 mM EDTA, 0.1% Tween-20) at 1400 rpm for 5 min. The bead-bound biotinylated strands were washed consecutively in 70% ethanol, denaturation solution (0.2 M NaOH), and wash buffer (10 mM Tris-acetate, pH 7.6) using the PyroMark Q96 Vacuum Workstation (QIAGEN). Purified DNA resuspended in annealing buffer (20 mM Tris-acetate pH 7.6, 2 mM magnesium acetate) was incubated with the sequencing primer at 85°C for 4 min. Pyrosequencing was performed on the PyroMark Q96 MD pyrosequencer (QIAGEN) with PyroMark Gold Q96 Reagents and HS Capillary Tips (QIAGEN) according to the manufacturer's instructions. Per cent CpG methylation was calculated by Pyro Q-CpG 1.0.9 software (Biotage) using the ratio of C-to-T at each site. Technical triplicates were averaged, and samples were kept for subsequent analysis if the



standard deviation of technical triplicates did not exceed 5%. Where indicated, methylation levels were averaged across CpGs for each individual at each locus.

### Clonal bisulphite sequencing

Bisulphite-converted liver DNA was amplified by PCR using the primers listed in **Supplementary file 1** and HotStarTaq DNA Polymerase (QIAGEN). PCR conditions were as follows: (1) 95°C for 10 min; (2) 94°C for 30 s, 60°C for 30 s, 72°C for 30 s, 40 cycles; (3) 72°C for 5 min. PCR products were purified by gel extraction using the QIAquick Gel Extraction Kit (QIAGEN) and cloned into the pGEM-T Easy vector (Promega) using Stellar Competent Cells (Takara Bio). Sanger sequencing of individual clones was carried out by Source BioScience. QUMA software ([Kumaki et al., 2008](#)) was used to analyse the sequencing data and generate figures.

### Chromatin immunoprecipitation (ChIP)

ChIP was carried out as previously described with modifications for use on frozen tissue ([Imbeault et al., 2017](#)). 100 mg of manually powdered frozen liver tissue dissected from adult *lap5-1<sup>full</sup>* and *lap5-1<sup>solo</sup>* males was cross-linked in 1% formaldehyde for 10 min at room temperature (RT) and quenched in 250 mM Tris-HCl pH 8 for 10 min at RT on a rotating wheel. Quenched samples were washed twice in 1× PBS supplemented with EDTA-free protease inhibitor cocktail cComplete (Sigma Aldrich), flash frozen in liquid nitrogen, and stored at –80°C. Fixed liver cells were thawed on ice and sequentially lysed in the following buffers for 10 min at 4°C: LB1 buffer (50 mM HEPES-KOH pH 7.4, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10% glycerol, 0.5% NP-40, 0.25% Triton-X-100, 1× EDTA-free cCompleteTM; one wash), LB2 buffer (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1× EDTA-free cCompleteTM; one wash), and SDS shearing buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 0.15% SDS, 1× EDTA-free cCompleteTM; three washes). Samples were centrifuged at 1700 × g for 5 min at 4°C after every wash. The resulting chromatin was sonicated for eight cycles (one cycle: 30 s on and 30 s off) at 4°C using a Bioruptor. The sonicated lysate was cleared by centrifugation at maximum speed for 10 min at 4°C and 10% of the input for each ChIP was stored at –20°C. 50 µl magnetic beads (Protein G Dynabeads, Invitrogen) were pre-blocked in fresh blocking buffer (0.5% BSA in PBS) and incubated with 2.5 µl of polyclonal H3K9me3 antibody (RRID:AB\_2532132, Active Motif) or Rabbit IgG (negative control) for 4 hr at 4°C. Antibody-bound beads were washed twice in blocking buffer using a magnetic stand at 4°C. The cleared lysate was topped up to 1 ml SDS shearing buffer + 150 mM NaCl and 1% Triton-X-100 and incubated with the antibody-bound beads on a rotating wheel overnight at 4°C. Non-specifically bound proteins were removed with the following sequential washes at 4°C: low salt buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF; two washes), high salt buffer (10 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF; one wash), LiCl buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 250 mM LiCl, 1% NP40, 1% Na-deoxycholate, 1 mM PMSF; one wash), and 10 mM Tris pH 8.0 (one wash). ChIP samples were resuspended in elution buffer (10 mM Tris pH 8.0, 1 mM EDTA, 1% SDS, 150 mM NaCl), treated with RNaseA at 37°C for 1 hr at 1100 rpm, and reverse cross-linked overnight at 65°C at 1100 rpm. The eluted samples were treated with Proteinase K and DNA was purified using the Monarch PCR and DNA Cleanup Kit (NEB).

### RNA extraction and cDNA synthesis

20–30 mg of thymus, liver, cortex, or placenta tissues was homogenised using the MagNA Lyser (Roche) at 6000 × g for 40 s. Thymus and placenta tissues were selected due to the specific expression of *5830416/19Rik* in the thymus and the high expression of *Rell1* and *Pgm2* in the placenta, as reported by NCBI ([Geer et al., 2010](#)). Liver and cortex were randomly selected to increase the number of tested tissue types. Total RNA was extracted using the AllPrep DNA/RNA Mini Kit (QIAGEN). Tissue DNA was digested on the RNeasy spin column membrane using the RNase-Free DNase Set (QIAGEN) and RNA integrity was confirmed by agarose gel electrophoresis. cDNA was synthesised from 5 µg RNA using random hexamer primers and the RevertAid H Minus First Strand cDNA Synthesis kit (Thermo Scientific).

Quantitative PCR (qPCR) primers were designed using Primer3 software (**Supplementary file 1**). Each reaction was carried out in technical triplicates with Brilliant III Ultra-Fast SYBR Green QPCR

Master Mix (Agilent) on the LightCycler 480 Instrument (Roche) under the following conditions: (1) 95°C for 5 min; (2) 95°C for 10 s, 60°C for 10 s, 72°C for 10 s, 45 cycles; (3) melting curve analysis of 65–95°C. For RT-qPCR, minus RT and no template controls were run for each sample and each primer pair, respectively. Relative expression was normalised to *Hprt1* and  $\beta$ -actin expression and calculated using the  $\Delta$ Ct method. For ChIP-qPCR, no template controls and 10% ChIP input were run alongside the H3K9me3 and Rabbit IgG ChIP samples for each primer pair. Enrichment was calculated as per cent input.

## Computational and statistical analyses

ChIP-seq data sets were downloaded in Bigwig format from the GEO database and visualised in the Integrative Genomics Viewer using the NCBI37/mm9 mouse reference genome (Thorvaldsdóttir et al., 2013). The three KAP1 ChIP-seq biological replicate tracks were summed in IGV before importing into Adobe Illustrator CC 2020 v24.0 for figure design. GEO accession numbers are listed in **Supplementary file 1**. B6 and CAST DNA sequences at the *lap5-1* insertion site were extracted from the GRCm38/mm10 and CAST\_EiJ\_v1 assemblies, respectively, accessed through the UCSC genome browser. The sequence alignment was generated using CLUSTAL OMEGA. All statistical tests in this study were carried out using GraphPad Prism 8 software as indicated in the figure legends.

## Acknowledgements

This work was supported by the Wellcome Trust, United Kingdom (210757/Z/18/Z to ACF-S), the Medical Research Council, United Kingdom (MR/R009791/1 to ACF-S), the Biotechnology and Biological Sciences Research Council, United Kingdom (BB/R009996/1 to ACF-S), and a Cambridge Trust PhD scholarship to TMB. We are grateful to Michael Imbeault, Todd Macfarlan, and Felipe Karam Teixeira for useful discussions, to Amir Hay and Athina Triantou for advice on ChIP, to Rahia Mashoodh for technical assistance during sperm isolation, and to members of the Ferguson-Smith lab for valuable feedback on our work.

---

## Additional information

### Funding

Funder	Grant reference number	Author
Wellcome Trust	210757/Z/18/Z	Anne C Ferguson-Smith
Medical Research Council	MR/R009791/1	Anne C Ferguson-Smith
Biotechnology and Biological Sciences Research Council	BB/R009996/1	Anne C Ferguson-Smith
Cambridge Trust	PhD scholarship	Tessa M Bertozzi

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Tessa M Bertozzi, Conceptualization, Formal analysis, Investigation, Visualization, Writing - original draft; Nozomi Takahashi, Geula Hanin, Investigation, Writing - review and editing; Anastasiya Kazachenka, Investigation; Anne C Ferguson-Smith, Conceptualization, Supervision, Funding acquisition, Writing - review and editing

### Author ORCIDs

Tessa M Bertozzi  <https://orcid.org/0000-0003-2900-6740>

Anne C Ferguson-Smith  <https://orcid.org/0000-0002-7608-5894>

### Ethics

Animal experimentation: Mouse work was carried out in accordance with the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body (Home Office project license # PC213320E).

### Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.65233.sa1>

Author response <https://doi.org/10.7554/eLife.65233.sa2>

## Additional files

### Supplementary files

- Supplementary file 1. Primer sequences and GEO accession numbers.
- Transparent reporting form

### Data availability

Sequencing data have been deposited in GenBank under accession number MW308129. GEO accession codes for publicly available ChIP-seq datasets analysed in this study are listed in Supplementary File 1.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Bertozzi TM, Hanin G, Takahashi N, Kazachenka A, Ferguson-Smith AC	2020	<i>Mus musculus</i> domesticus strain C57BL/6J retrotransposon C57iap1_solo, complete sequence	<a href="https://www.ncbi.nlm.nih.gov/nucleotide/MW308129.1/">https://www.ncbi.nlm.nih.gov/nucleotide/MW308129.1/</a>	NCBI GenBank, MW308129

The following previously published dataset was used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Wolf G, de Iaco A, Sun MA, Bruno M	2019	Retrotransposon reactivation and mobilization upon deletions of megabase-scale KRAB zinc finger gene clusters in mice	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115291">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115291</a>	NCBI Gene Expression Omnibus, GSE115291

## References

- Allen ND, Norris ML, Surani MA. 1990. Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. *Cell* **61**:853–861. DOI: [https://doi.org/10.1016/0092-8674\(90\)90195-K](https://doi.org/10.1016/0092-8674(90)90195-K), PMID: 2111735
- Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally regulated piRNA clusters implicate MIL1 in transposon control. *Science* **316**:744–747. DOI: <https://doi.org/10.1126/science.1142612>, PMID: 17446352
- Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Héroult Y, Guillou F, Bourc'his D. 2016. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **354**:909–912. DOI: <https://doi.org/10.1126/science.aah5143>, PMID: 27856912
- Beamer LJ. 2015. Mutations in hereditary phosphoglucomutase 1 deficiency map to key regions of enzyme structure and function. *Journal of Inherited Metabolic Disease* **38**:243–256. DOI: <https://doi.org/10.1007/s10545-014-9757-9>, PMID: 25168163
- Bertozzi TM, Elmer JL, Macfarlan TS, Ferguson-Smith AC. 2020. KRAB zinc finger protein diversification drives mammalian interindividual methylation variability. *PNAS* **117**:31290–31300. DOI: <https://doi.org/10.1073/pnas.2017053117>, PMID: 33239447
- Bertozzi TM, Ferguson-Smith AC. 2020. Metastable epialleles and their contribution to epigenetic inheritance in mammals. *Seminars in Cell & Developmental Biology* **97**:93–105. DOI: <https://doi.org/10.1016/j.semcdb.2019.08.002>, PMID: 31551132
- Bourc'his D, Bestor TH. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**:96–99. DOI: <https://doi.org/10.1038/nature02886>, PMID: 15318244

- Bultman SJ**, Klebig ML, Michaud EJ, Sweet HO, Davisson MT, Woychik RP. 1994. Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (a(t)) and white-bellied agouti (Aw) reveals alternative forms of agouti transcripts. *Genes & Development* **8**:481–490. DOI: <https://doi.org/10.1101/gad.8.4.481>
- Carmell MA**, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental Cell* **12**:503–514. DOI: <https://doi.org/10.1016/j.devcel.2007.03.001>, PMID: 17395546
- Chuong EB**, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**:1083–1087. DOI: <https://doi.org/10.1126/science.aad5497>, PMID: 26941318
- Du J**, Leung A, Trac C, Lee M, Parks BW, Lusic AJ, Natarajan R, Schones DE. 2016. Chromatin variation associated with liver metabolism is mediated by transposable elements. *Epigenetics & Chromatin* **9**:1–16. DOI: <https://doi.org/10.1186/s13072-016-0078-0>, PMID: 27398095
- Duhl DM**, Vrieling H, Miller KA, Wolff GL, Barsh GS. 1994. Neomorphic Agouti mutations in obese yellow mice. *Nature Genetics* **8**:59–65. DOI: <https://doi.org/10.1038/ng0994-59>, PMID: 7987393
- Ecco G**, Imbeault M, Trono D. 2017. KRAB zinc finger proteins. *Development* **144**:2719–2729. DOI: <https://doi.org/10.1242/dev.132605>, PMID: 28765213
- Ehrlich M**, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, Gehrke C. 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Research* **10**:2709–2721. DOI: <https://doi.org/10.1093/nar/10.8.2709>, PMID: 7079182
- Elmer JL**, Hay AD, Kessler NJ, Bertozzi TM, Ainscough E, Ferguson-Smith AC. 2021. Genomic properties of variably methylated repetitive elements in mouse. *Mobile DNA* **12**:6. DOI: <https://doi.org/10.1186/s13100-021-00235-1>
- Elmer JL**, Ferguson-Smith AC. 2020. Strain-Specific epigenetic regulation of endogenous retroviruses: the role of Trans-Acting modifiers. *Viruses* **12**:810–821. DOI: <https://doi.org/10.3390/v12080810>
- Gagnier L**, Belancio VP, Mager DL. 2019. Mouse germ line mutations due to retrotransposon insertions. *Mobile DNA* **10**:15. DOI: <https://doi.org/10.1186/s13100-019-0157-4>, PMID: 31011371
- Geer LY**, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH. 2010. The NCBI BioSystems database. *Nucleic Acids Research* **38**:D492–D496. DOI: <https://doi.org/10.1093/nar/gkp858>, PMID: 19854944
- Haeussler M**, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, Gibson D, Diekhans M, Clawson H, Casper J, Barber GP, Haussler D, Kuhn RM, Kent WJ. 2019. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Research* **47**:D853–D858. DOI: <https://doi.org/10.1093/nar/gky1095>
- Imbeault M**, Helleboid P-Y, Trono D. 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**:550–554. DOI: <https://doi.org/10.1038/nature21683>
- Jain D**, Meydan C, Lange J, Claeys Bouuaert C, Lailier N, Mason CE, Anderson KV, Keeney S. 2017. Rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. *PLOS Genetics* **13**:e1006964. DOI: <https://doi.org/10.1371/journal.pgen.1006964>, PMID: 28854222
- Jern P**, Coffin JM. 2008. Effects of retroviruses on host genome function. *Annual Review of Genetics* **42**:709–732. DOI: <https://doi.org/10.1146/annurev.genet.42.110807.091501>, PMID: 18694346
- Kawai J**, Shinagawa A, Shibata K, Yoshino M, Itoh M, Ishii Y, Arakawa T, Hara A, Fukunishi Y, Konno H, Adachi J, Fukuda S, Aizawa K, Izawa M, Nishi K, Kiyosawa H, Kondo S, Yamanaka I, Saito T, Okazaki Y, et al. 2001. Functional annotation of a full-length mouse cDNA collection. *Nature* **409**:685–689. DOI: <https://doi.org/10.1038/35055500>, PMID: 11217851
- Kazachenka A**, Bertozzi TM, Sjoberg-Herrera MK, Walker N, Gardner J, Gunning R, Pahita E, Adams S, Adams D, Ferguson-Smith AC. 2018. Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell* **175**:1259–1271. DOI: <https://doi.org/10.1016/j.cell.2018.09.043>
- Kearns M**, Preis J, McDonald M, Morris C, Whitelaw E. 2000. Complex patterns of inheritance of an imprinted murine transgene suggest incomplete germline erasure. *Nucleic Acids Research* **28**:3301–3309. DOI: <https://doi.org/10.1093/nar/28.17.3301>
- Kuehnen P**, Mischke M, Wiegand S, Sers C, Horsthemke B, Lau S, Keil T, Lee YA, Grueters A, Krude H. 2012. An alu element-associated hypermethylation variant of the POMC gene is associated with childhood obesity. *PLOS Genetics* **8**:e1002543. DOI: <https://doi.org/10.1371/journal.pgen.1002543>, PMID: 22438814
- Kuff EL**, Lueders KK. 1988. The intracisternal A-particle gene family: structure and functional aspects. *Advances in Cancer Research* **51**:183–276. DOI: [https://doi.org/10.1016/s0065-230x\(08\)60223-7](https://doi.org/10.1016/s0065-230x(08)60223-7), PMID: 3146900
- Kumaki Y**, Oda M, Okano M. 2008. QUMA: quantification tool for methylation analysis. *Nucleic Acids Research* **36**:W170–W175. DOI: <https://doi.org/10.1093/nar/gkn294>, PMID: 18487274
- Lane N**, Dean W, Erhardt S, Hajkova P, Surani A, Walter J, Reik W. 2003. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* **35**:88–93. DOI: <https://doi.org/10.1002/gene.10168>, PMID: 12533790
- Maksakova IA**, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL. 2006. Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLOS Genetics* **2**:e2. DOI: <https://doi.org/10.1371/journal.pgen.0020002>, PMID: 16440055
- Matsui T**, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**:927–931. DOI: <https://doi.org/10.1038/nature08858>, PMID: 20164836

- Morgan HD**, Sutherland HG, Martin DI, Whitelaw E. 1999. Epigenetic inheritance at the Agouti locus in the mouse. *Nature Genetics* **23**:314–318. DOI: <https://doi.org/10.1038/15490>, PMID: 10545949
- Nellåker C**, Keane TM, Yalcin B, Wong K, Agam A, Belgard TG, Flint J, Adams DJ, Frankel WN, Ponting CP. 2012. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biology* **13**:R45. DOI: <https://doi.org/10.1186/gb-2012-13-6-r45>, PMID: 22703977
- Nichols J**, Jones K. 2017. Derivation of mouse embryonic stem (ES) Cell lines using Small-Molecule inhibitors of erk and Gsk3 signaling (2i). *Cold Spring Harbor Protocols* **2017**:pdb.prot094086. DOI: <https://doi.org/10.1101/pdb.prot094086>
- Rakyan VK**, Blewitt ME, Druker R, Preis JI, Whitelaw E. 2002. Metastable epialleles in mammals. *Trends in Genetics* **18**:348–351. DOI: [https://doi.org/10.1016/S0168-9525\(02\)02709-9](https://doi.org/10.1016/S0168-9525(02)02709-9), PMID: 12127774
- Rakyan VK**, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KV, Whitelaw E. 2003. Transgenerational inheritance of epigenetic states at the murine axin(Fu) allele occurs after maternal and paternal transmission. *PNAS* **100**:2538–2543. DOI: <https://doi.org/10.1073/pnas.0436776100>, PMID: 12601169
- Ruda VM**, Akopov SB, Trubetskoy DO, Manuylov NL, Vetchinova AS, Zavalova LL, Nikolaev LG, Sverdlov ED. 2004. Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR. *Virus Research* **104**:11–16. DOI: <https://doi.org/10.1016/j.virusres.2004.02.036>, PMID: 15177887
- Scherneck S**, Nestler M, Vogel H, Blüher M, Block MD, Berriel Diaz M, Herzig S, Schulz N, Teichert M, Tischer S, Al-Hasani H, Kluge R, Schürmann A, Joost HG. 2009. Positional cloning of zinc finger domain transcription factor Zfp69, a candidate gene for obesity-associated diabetes contributed by mouse locus nidd/SJL. *PLOS Genetics* **5**:e1000541. DOI: <https://doi.org/10.1371/journal.pgen.1000541>, PMID: 19578398
- Schroeder DI**, Jayashankar K, Douglas KC, Thirkill TL, York D, Dickinson PJ, Williams LE, Samollow PB, Ross PJ, Bannasch DL, Douglas GC, LaSalle JM. 2015. Early developmental and evolutionary origins of gene body DNA methylation patterns in mammalian placentas. *PLOS Genetics* **11**:e1005442. DOI: <https://doi.org/10.1371/journal.pgen.1005442>, PMID: 26241857
- Seisenberger S**, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W. 2012. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Molecular Cell* **48**:849–862. DOI: <https://doi.org/10.1016/j.molcel.2012.11.001>, PMID: 23219530
- Seperack PK**, Strobel MC, Corrow DJ, Jenkins NA, Copeland NG. 1988. Somatic and germ-line reverse mutation rates of the retrovirus-induced dilute coat-color mutation of DBA mice. *PNAS* **85**:189–192. DOI: <https://doi.org/10.1073/pnas.85.1.189>, PMID: 3422417
- Sharma U**, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, Kucukural A, Serra RW, Sun F, Song L, Carone BR, Emiliano P, Xz L, Fauquier L, Moore MJ, Sullivan R, Mello CC, Rando OJ. 2015. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **6780**:1–9. DOI: <https://doi.org/10.1126/science.aad6780>
- Shimosuga K**, Fukuda K, Sasaki H, Ichianagi K. 2017. Locus-specific hypomethylation of the mouse IAP retrotransposon is associated with transcription factor-binding sites. *Mobile DNA* **8**:1–12. DOI: <https://doi.org/10.1186/s13100-017-0105-0>
- Smit AFA**, Hubley R, Green P. 2015. RepeatMasker Open. 4.0. <http://www.repeatmasker.org/>
- Stoye JP**, Fenner S, Greenoak GE, Moran C, Coffin JM. 1988. Role of endogenous retroviruses as mutagens: the hairless mutation of mice. *Cell* **54**:383–391. DOI: [https://doi.org/10.1016/0092-8674\(88\)90201-2](https://doi.org/10.1016/0092-8674(88)90201-2), PMID: 2840205
- Subramanian RP**, Wildschutte JH, Russo C, Coffin JM. 2011. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology* **8**:90. DOI: <https://doi.org/10.1186/1742-4690-8-90>, PMID: 22067224
- Thomas J**, Perron H, Feschotte C. 2018. Variation in proviral content among human genomes mediated by LTR recombination. *Mobile DNA* **9**:36. DOI: <https://doi.org/10.1186/s13100-018-0142-3>
- Thorvaldsdóttir H**, Robinson JT, Mesirov JP. 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**:178–192. DOI: <https://doi.org/10.1093/bib/bbs017>, PMID: 22517427
- Tie CH**, Fernandes L, Conde L, Robbez-Masson L, Sumner RP, Peacock T, Rodriguez-Plata MT, Mickute G, Gifford R, Towers GJ, Herrero J, Rowe HM. 2018. KAP 1 regulates endogenous retroviruses in adult human cells and contributes to innate immune control. *EMBO Reports* **19**:e45000. DOI: <https://doi.org/10.15252/embr.201745000>, PMID: 30061100
- Timmons JA**, Atherton PJ, Larsson O, Sood S, Blokhin IO, Brogan RJ, Volmar CH, Josse AR, Slentz C, Wahlestedt C, Phillips SM, Phillips BE, Gallagher IJ, Kraus WE. 2018. A coding and non-coding transcriptomic perspective on the genomics of human metabolic disease. *Nucleic Acids Research* **46**:7772–7792. DOI: <https://doi.org/10.1093/nar/gky570>, PMID: 29986096
- Walsh CP**, Chaillet JR, Bestor TH. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genetics* **20**:116–117. DOI: <https://doi.org/10.1038/2413>, PMID: 9771701
- Waterland RA**, Dolinoy DC, Lin JR, Smith CA, Shi X, Tahiliani KG. 2006. Maternal methyl supplements increase offspring DNA methylation at axin fused. *Genesis* **44**:401–406. DOI: <https://doi.org/10.1002/dvg.20230>, PMID: 16868943
- Waterland RA**, Jirtle RL. 2003. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Molecular and Cellular Biology* **23**:5293–5300. DOI: <https://doi.org/10.1128/MCB.23.15.5293-5300.2003>, PMID: 12861015

- Wolf G**, de Iaco A, Sun MA, Bruno M, Tinkham M, Hoang D, Mitra A, Ralls S, Trono D, Macfarlan TS. 2020. KRAB-zinc finger protein gene expansion in response to active retrotransposons in the murine lineage. *eLife* **9**: e56337. DOI: <https://doi.org/10.7554/eLife.56337>, PMID: 32479262
- Wolf D**, Goff SP. 2007. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131**:46–57. DOI: <https://doi.org/10.1016/j.cell.2007.07.026>, PMID: 17923087
- Wolff GL**. 1971. Genetic modification of homeostatic regulation in the mouse. *The American Naturalist* **105**:241–252. DOI: <https://doi.org/10.1086/282721>
- Wolff GL**, Kodell RL, Moore SR, Cooney CA. 1998. Maternal epigenetics and methyl supplements affect Agouti gene expression in avy/a mice. *FASEB J* **12**:949–957. DOI: <https://doi.org/10.1096/fasebj.12.11.949>, PMID: 9707167
- Ye M**, Goudot C, Hoyler T, Lemoine B, Amigorena S, Zueva E. 2020. Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers. *PNAS* **117**:7905–7916. DOI: <https://doi.org/10.1073/pnas.1912008117>, PMID: 32193341
- Yen TT**, Gill AM, Frigeri LG, Barsh GS, Wolff GL. 1994. Obesity, diabetes, and neoplasia in yellow A(vy)/- mice: ectopic expression of the Agouti gene. *FASEB J* **8**:479–488. DOI: <https://doi.org/10.1096/fasebj.8.8.8181666>, PMID: 8181666