Title

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

Authors

Tessa M. Bertozzi[1], Geula Hanin[1†], Nozomi Takahashi[1†], Anastasiya Kazachenka[1#], Anne C. Ferguson-Smith[1*]

Affiliations

[1] Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK

[#]Present address: The Fancis Crick Institute, London, NW1 1AT, UK

[†]These authors contributed equally.

[*]Correspondence: afsmith@gen.cam.ac.uk

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 (*C57iap1*^{solo}) recently formed in the inbred C57BL/6J mouse strain. In contrast to the C57BL/6J full-length IAP at this locus (*C57iap1*^{full}), *C57iap1*^{solo} 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 *C57iap1*^{solo} variant. *C57iap1*^{solo} 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 *C57iap1* as a genetically-induced epiallele with functional consequences establishes a new model to study transposable element repression and host-element co-evolution.

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 9 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).

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^{vy}) allele drives ectopic expression of the downstream coat colour gene *Agouti* when unmethylated. By mechanisms not yet fully understood, the A^{vy} 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., 2020; 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-*Pgm1* element, named after its closest annotated coding gene *Phosphoglucomutase-1* (*Pgm1*). 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-*Pgm1*, 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-*Pgm1* is a full-length IAP matching the B6 reference genome (renamed $C57iap1^{full}$) and the unmethylated variant is a solo LTR produced from an inter-LTR recombination event (renamed $C57iap1^{solo}$). The absence of DNA methylation at $C57iap1^{solo}$ 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 *C57iap1*^{solo} 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 *C57iap1*^{solo} 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 *C57iap1* 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-*Pgm1* 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., 2020; Kazachenka et al., 2018). While VM-IAP DNA methylation levels display continuous probability distributions in the B6 population, the IAP-*Pgm1* 5′ LTR exhibited three distinct states: high (> 85%), low (< 20%), and intermediate (60-70%) methylation (Figure 1A). 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).

Methylation quantification of unique non-repetitive DNA immediately up- and downstream of IAP-*Pgm1* showed that the three distinct methylation states become less defined as the distance from the LTR borders increases, ultimately collapsing approximately 500 bp

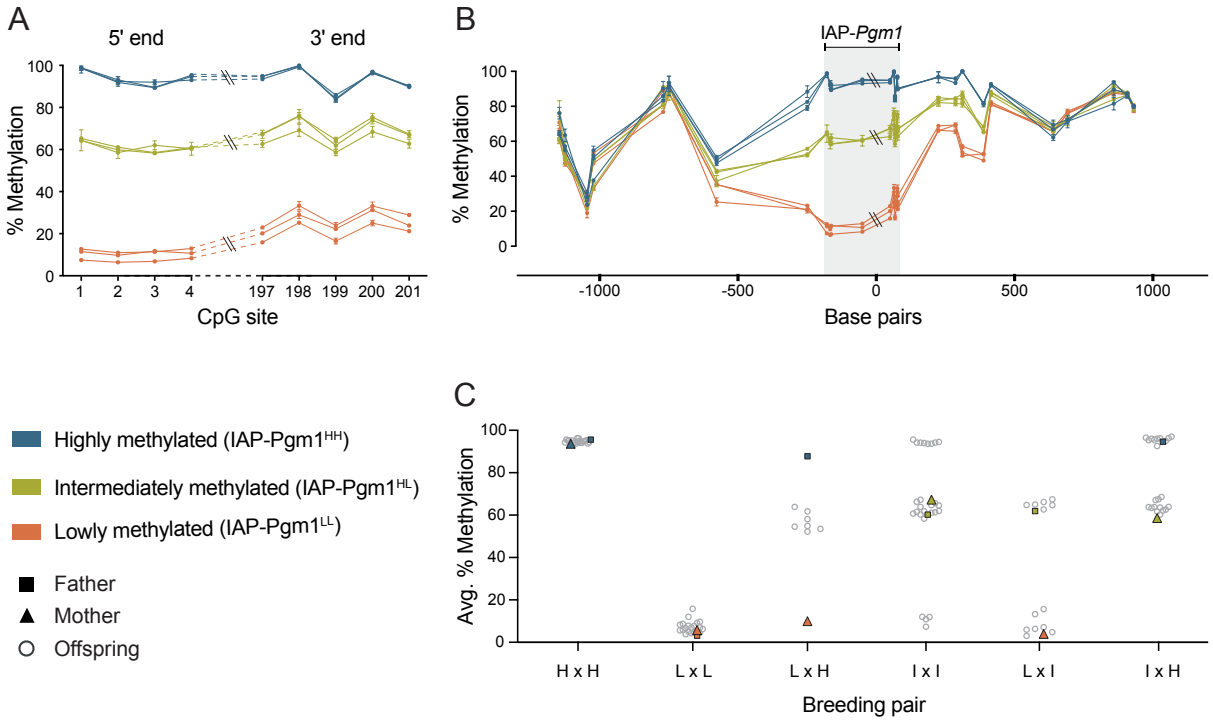

Figure 1. IAP-*Pgm1* methylation is tri-modally distributed and stably inherited in inbred B6 mice. (**A**) IAP-*Pgm1* methylation levels are consistent between the 5′ and 3′ ends. Methylation levels were quantified at the most distal CpGs of the IAP-*Pgm1* 5′ and 3′ LTRs (nearest to the boundary with unique DNA) using bisulphite pyrosequencing in ear DNA. Each point represents a CpG, each line represents an individual, and error bars represent standard deviations between technical triplicates. (**B**) Inter-individual methylation variation collapses within 500 bp on either side of IAP-*Pgm1*. Data presentation as in panel A. (**C**) Stable Mendelian inheritance of IAP-*Pgm1* methylation reveals that high (blue, H) and low (orange, L) methylation reflect two allelic states of IAP-*Pgm1*, with intermediate methylation (orange, I) representing heterozygosity. Each data point represents average % methylation across the four most distal CpGs of the 5′ LTR for one individual.

from either side of the IAP (Figure 1B). This provides evidence for short-distance spreading of DNA methylation levels from IAP-*Pgm1* into bordering DNA and suggests that the methylation differences observed between individuals are intrinsic to the IAP-*Pgm1* element rather than a reflection of differential methylation of the insertion site prior to integration.

IAP-*Pgm1* 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-*Pgm1*, specific parental combinations were set up for breeding and IAP-*Pgm1* 5′ LTR methylation levels were quantified in the offspring. In stark contrast to VM-IAPs, IAP-*Pgm1* 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 1C). When one parent was highly methylated and the other lowly methylated, all offspring were intermediately methylated (Figure 1C). This indicates that high and low methylation states are allelic variants of IAP-*Pgm1* (designated IAP-*Pgm1*^{HH} and IAP-*Pgm1*^{LL}), with intermediate methylation representing co-dominant epigenetic heterozygosity (IAP-*Pgm1*^{HL}). Additional crosses confirmed this inheritance pattern: an IAP-*Pgm1*^{HL} intercross produced IAP-*Pgm1*^{HH}, IAP-*Pgm1*^{LL}, and IAP-*Pgm1*^{HL} offspring; an IAP-*Pgm1*^{HH} x IAP-*Pgm1*^{HL} cross produced IAP-*Pgm1*^{HH} and IAP-*Pgm1*^{HL} offspring; and an IAP-*Pgm1*^{LL} x IAP-*Pgm1*^{HL} cross produced IAP-*Pgm1*^{LL} and IAP-*Pgm1*^{HL} offspring (Figure 1C). These results additionally demonstrate that the methylation state of one IAP-*Pgm1* variant does not influence the methylation state of the other in a heterozygous context.

The methylation state of IAP-*Pgm1* is genetically determined

The stable inheritance of IAP-*Pgm1* 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-*Pgm1* from IAP-*Pgm1*^{HH} and IAP-*Pgm1*^{LL} gDNA (Figure 2A). Nested primer pairs N1 and N2 amplified two overlapping fragments, each containing half of the IAP-*Pgm1* element (Figure 2A). Agarose gel electrophoresis of the PCR products revealed amplification of both fragments from IAP-*Pgm1*^{HH} gDNA but no amplification of either fragment from IAP-*Pgm1*^{LL} gDNA (Figure 2B and C), pointing to a substantial genetic difference between IAP-*Pgm1*^{HH} and IAP-*Pgm1*^{LL} 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-*Pgm1*^{HH} gDNA, a 1.5 kb band was amplified from both IAP-*Pgm1*^{LL} and IAP-*Pgm1*^{HL} gDNA (Figure 2D). The smaller amplicon is indicative of a 7 kb deletion on the IAP-*Pgm1*^{L} allele between the two primer annealing sites, while the lack of amplification from the IAP-*Pgm1*^{H} allele is likely due to the technical challenges associated with amplifying a large repetitive DNA fragment.

To determine the location of the 7 kb deletion, we purified and sequenced the PCR products and aligned the assembled IAP-*Pgm1*^{H} and IAP-*Pgm1*^{L} sequences. The alignment revealed that the IAP-*Pgm1*^{H} allele is an identical match to the full-length IAP-*Pgm1* sequence from the mm10 reference, while the IAP-*Pgm1*^{L} allele is a solo LTR (Figure 2E). The IAP-*Pgm1*^{L} solo LTR and the IAP-*Pgm1*^{H} full-length IAP are both flanked by the same target site duplications (TSDs) and the IAP-*Pgm1*^{L} solo LTR exhibits 100% sequence identity to both the 5′ and 3′ LTRs of the full-length IAP-*Pgm1*^{H} (Figure 2E). Therefore, a recent inter-LTR

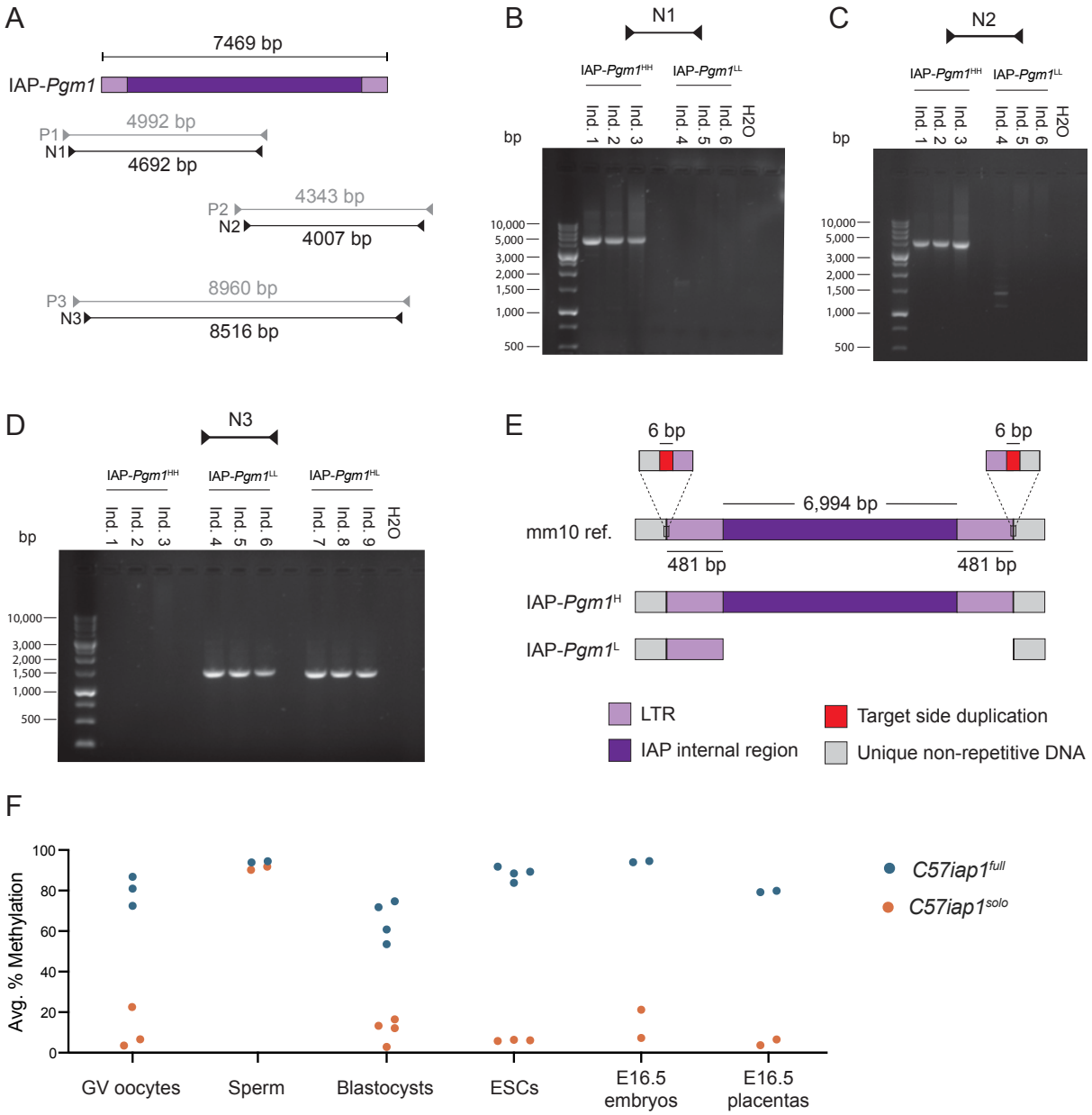

Figure 2. The IAP-*Pgm1*^{L} allele is a solo LTR formed via inter-LTR homologous recombination. (**A**) Map of PCR primer pairs P1-P3 and nested PCR primer pairs N1-N3 (not drawn to scale). (**B** and **C**) Agarose gel electrophoresis of PCR products amplified from three IAP-*Pgm1*^{HH} and three IAP-*Pgm1*^{LL} DNA samples using primer pairs N1 and N2. (**D**) As in panels B and C, but using primer pair N3 and including three IAP-*Pgm1*^{HL} DNA samples. (**E**) Schematic representation of the alignment of the IAP-*Pgm1* mm10 reference sequence and the assembled IAP-*Pgm1*^{H} and IAP-*Pgm1*^{L} sequences following Sanger sequencing (not drawn to scale).The IAP-*Pgm1*^{L} solo LTR could have equivalently been shown aligned to the 3′ LTR because the 5′ and 3′ LTRs have identical sequences. (**F**) Both *C57iap1* variants are methylated in the male germline, with differential methylation re-established in early embryonic development. DNA methylation levels at the *C57iap1*^{full} and *C57iap1*^{solo} 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 *C57iap1*.

166 homologous recombination event in the inbred B6 mouse strain gave rise to the solo LTR
167 variant. IAP-*Pgm1*^{H} and IAP-*Pgm1*^{L} are hereafter referred to as *C57BL/6J IAP 1*^{full}
168 (*C57iap1*^{full}) and *C57BL/6J IAP 1*^{solo} (*C57iap1*^{solo}), respectively.

169 Inter-LTR recombination events occur when identical or near-identical LTR sequences
170 engage in ectopic homologous recombination, resulting in the formation of solo LTRs (Jern
171 and Coffin, 2008). These can occur both intra- and inter-chromosomally (Figure 2—figure
172 supplement 1). Although we identified both IAP-*Pgm1* allelic variants in our own B6 colony,
173 we also detected them in B6 samples sent to us from mainland Europe and North America (data
174 not shown), so the recombination event most likely occurred at a B6-distributing facility. This
175 type of proviral excision event is common: approximately half of the IAPs in the mouse
176 genome are solo LTRs (Nellåker et al., 2012). However, the vast majority of the ~5,000 solo
177 LTRs in the B6 genome are highly methylated (Shimosuga et al., 2017), making *C57iap1*^{solo}
178 unique from a regulatory perspective and raising questions regarding the functional
179 consequences of a solo LTR left unmodified.

180

181 **Developmental dynamics of *C57iap1* methylation states**

182 The experiments described thus far have focused on adult somatic *C57iap1* DNA methylation
183 within and across generations. Given the dynamic nature of DNA methylation during
184 mammalian development and considering previous reports on the resistance of IAPs to
185 genome-wide methylation erasure (Lane et al., 2003; Seisenberger et al., 2012), we sought to
186 examine and compare DNA methylation levels of the *C57iap1*^{full} and *C57iap1*^{solo} variants in
187 the germline and during early embryonic development.

188 Germinal vesicle (GV) oocytes and mature sperm were collected from *C57iap1*^{full} and
189 *C57iap1*^{solo} adult B6 females and males. Oocytes collected from *C57iap1*^{full} and *C57iap1*^{solo}

females were highly and lowly methylated at the *C57iap1* locus, respectively (Figure 2F). Thus, oocyte *C57iap1* methylation levels are reflective of somatic *C57iap1* 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 *C57iap1* in the soma. Together, these experiments indicate that the maternal and paternal *C57iap1*^{solo} alleles are differentially methylated upon fertilisation in the early zygote.

We examined the behaviour of *C57iap1* methylation levels during early development in blastocysts (embryonic day 3.5, E3.5) collected from the uteri of *C57iap1*^{full} and *C57iap1*^{solo} B6 females bred to *C57iap1*^{full} and *C57iap1*^{solo} B6 males, respectively. Blastocysts generated from *C57iap1*^{solo} parents exhibited low methylation levels at *C57iap1*, suggesting that the paternally inherited hypermethylated *C57iap1*^{solo} allele becomes demethylated shortly after fertilisation (Figure 2F). By comparison, blastocysts generated from *C57iap1*^{full} parents exhibited methylation levels around 70% (Figure 2F). It is unclear whether *C57iap1*^{full} is demethylated after fertilisation and rapidly methylated again by the blastocyst stage, or whether *C57iap1*^{full} is generally resistant to epigenetic reprogramming during preimplantation development. Nonetheless, these results demonstrate that methylation patterns at *C57iap1*^{full} and *C57iap1*^{solo} are specified prior to implantation.

Despite the marked distinction in methylation states between *C57iap1*^{full} and *C57iap1*^{solo} blastocysts, *C57iap1*^{full} 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 are countering 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 *C57iap1*^{full} and *C57iap1*^{solo} blastocysts closely matched those observed in adult somatic tissues, with ES cells derived from *C57iap1*^{full}

blastocysts nearing 100% methylation (Figure 2F). In addition, we found that *C57iap1*^{full} methylation levels in E16.5 placentas were lower than those in E16.5 embryonic tissue (Figure 2F), consistent with previous reports on global placental hypomethylation (Ehrlich et al., 1982; Schroeder et al., 2015). These results provide evidence for differential methylation between ICM- and trophectoderm-derived lineages at the *C57iap1* locus. Of note, even though placental *C57iap1* methylation levels are hypomethylated compared to their embryonic counterparts, the two variants retain a pronounced difference in DNA methylation levels in this tissue.

Loss of H3K9me3 occupancy at *C57iap1*^{solo}

To determine whether loss of DNA methylation at *C57iap1*^{solo} is associated with changes in chromatin organisation, we quantified H3K9me3 enrichment in *C57iap1*^{full} and *C57iap1*^{solo} adult liver samples via ChIP-qPCR. The retrotransposons IAP-*Asxl3* and SINE-*Rbak*, used as positive controls, exhibited equivalent H3K9me3 enrichment between *C57iap1*^{full} and *C57iap1*^{solo} individuals (Figure 3A). The *Gapdh* promoter was used as a negative control. Because the sequence of *C57iap1*^{solo} is identical to that of the 5′ and 3′ LTRs of *C57iap1*^{full}, 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 *C57iap1*^{solo} samples compared to *C57iap1*^{full} samples (Figure 3A). *C57iap1*^{full} 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 *C57iap1*^{solo} samples is due to the presence of an ERV element immediately downstream of *C57iap1*.

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

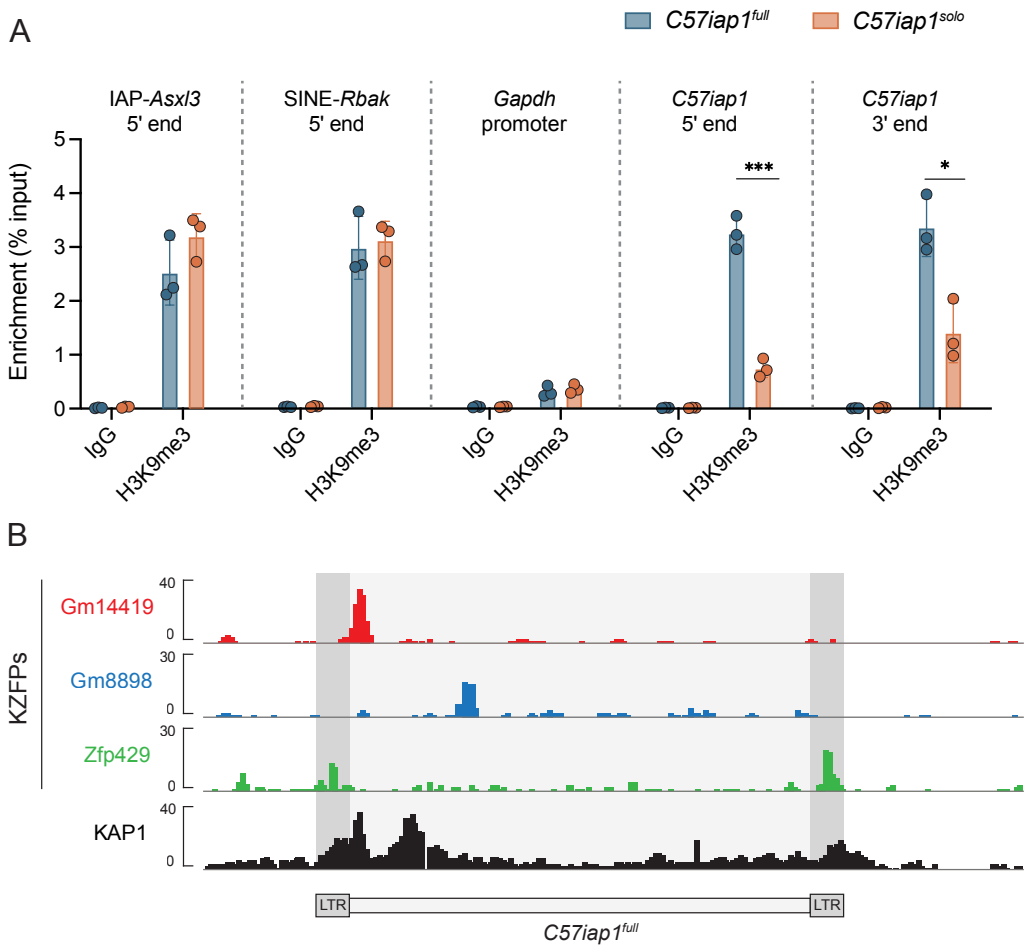

Figure 3. Lack of DNA methylation at *C57iap1*^{solo} is accompanied by a loss of H3K9me3 marks. (**A**) H3K9me3 ChIP-qPCR on *C57iap1*^{full} and *C57iap1*^{solo} adult male livers. IAP-*Asxl3* and SINE-*Rbak* are positive control loci, the *Gapdh* promoter is a negative control locus, and the Rabbit IgG antibody serves as a negative isotype control. H3K9me3 enrichment was calculated using the percent input method and compared between genotypes using unpaired t tests (* $p < 0.05$; *** $p < 0.0005$). (**B**) *C57iap1*^{full} is bound by multiple KZFPs. Publicly available ChIP-seq datasets indicate that KZFPs Gm14419 and Gm8898 are capable of binding the internal region of *C57iap1*^{full} and that KZFP Zfp429 is capable of binding the LTR sequence shared by both *C57iap1*^{full} and *C57iap1*^{solo}. ChIP-seq datasets were downloaded from the GEO database (accession numbers: Gm14419, GSM3173720; Gm8898, GSM3173728; Zfp429, GSM3173732; KAP1, sum of GSM3173661, GSM3173662, and GSM3173663).

239 methyltransferases (Ecco et al., 2017). We reasoned that *C57iap1*^{full} may be targeted for
240 repression in the early embryo by KZFP(s) whose binding sites are located in the internal
241 proviral portion of the IAP which is no longer present in the *C57iap1*^{solo} variant. To explore
242 this hypothesis, we analysed previously published ChIP-seq (chromatin immunoprecipitation
243 followed by high-throughput sequencing) binding profiles of more than 60 murine KZFPs
244 (Wolf et al., 2020) and identified two *C57iap1*^{full}-binding candidates, Gm14419 and Gm8898
245 (Figure 3B). The ChIP-seq profiles for Gm14419 and Gm8898 in B6 ES cells displayed peaks
246 immediately downstream and 1.5 kb downstream of the 5′ LTR, respectively (Figure 3B). We
247 observed KAP1 occupancy at the Gm14419 binding site, rendering Gm14419 a particularly
248 promising candidate for future mechanistic research (Figure 3B). Furthermore, it is likely that
249 additional as-yet-unidentified KZFP(s) bind to *C57iap1*^{full}, as evidenced by a second KAP1
250 peak in the *C57iap1*^{full} internal region which does not overlap with the binding site of any of
251 the KZFPs identified in our analysis (Figure 3B). This is consistent with the redundant nature
252 of KZFP-mediated ERV repression (Imbeault et al., 2017; Wolf et al., 2020). In addition, we
253 detected ZFP429 binding peaks at the 5′ and 3′ LTRs of *C57iap1*^{full} (Figure 3B). Given the
254 shared sequence between the *C57iap1*^{full} LTRs and *C57iap1*^{solo}, this observation suggests that
255 ZFP429 does not recruit heterochromatin factors as effectively as other KZFPs. This is in line
256 with our previous finding that ZFP429 is enriched at VM-IAPs compared to fully methylated
257 IAPs of the same subclass (Bertozzi et al., 2020).

258

259 **Genetic background influences *C57iap1*^{solo} methylation levels**

260 Previous studies have shown that in some cases IAP methylation is modulated by genetic
261 background (Bertozzi et al., 2020; Elmer and Ferguson-Smith, 2020; Rakyan et al., 2003;
262 Wolff, 1971). To assess whether this is the case for *C57iap1* methylation, we carried out
263 reciprocal crosses between B6 and wild-derived CAST/EiJ (CAST) mice (BC, B6 female x

CAST male; CB, CAST female x B6 male). F1 hybrid offspring carry a single copy of *C57iap1* inherited from their B6 parent because the *C57iap1* insertion is absent from the CAST genome (Figure 4A). In line with our breeding experiments in pure B6 mice, hemizygous offspring born to a *C57iap1*^{full} B6 parent were highly methylated and those born to a *C57iap1*^{solo} B6 parent were lowly methylated (Figure 4B). However, although *C57iap1*^{solo} 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 *C57iap1*^{solo} in *trans*. In addition, CB offspring displayed higher and more variable *C57iap1*^{solo} 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 *C57iap1*^{solo} CB hybrid males to pure CAST females. This resulted in a cumulative effect, whereby N1 *C57iap1*^{solo} mice showed even higher and more variable methylation levels compared to F1 *C57iap1*^{solo} 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 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).

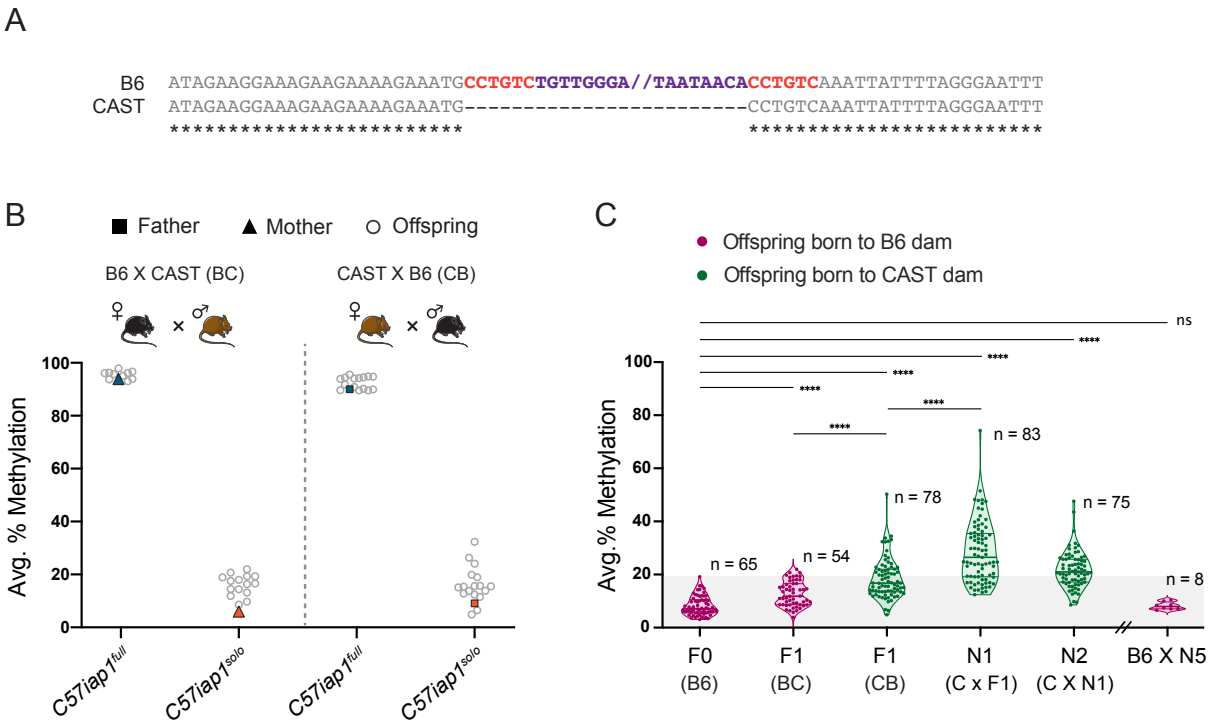

Figure 4. Genetic background influences *C57iap1*^{solo} methylation levels. (**A**) B6 and CAST sequence alignment of the *C57iap1* insertion site reveals that the *C57iap1* element (purple) and its associated target site duplication (TSD) sequence (red) are not present in the CAST genome. Asterisks 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. *C57iap1*^{full} and *C57iap1*^{solo} B6 females were crossed to CAST males (BC, left); *C57iap1*^{full} and *C57iap1*^{solo} 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 *C57iap1* for one individual. (**C**) The genetic background-specific maternal effect at *C57iap1*^{solo} 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, the N2 generation was generated by crossing N1 males harbouring the *C57iap1*^{solo} allele to CAST females. After 5 generations of backcrossing to CAST, N5 males were crossed to B6 females. Offspring that did not inherit the *C57iap1*^{solo} allele were not included in the analysis. Sample sizes are shown on the graph. 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).

C57iap1^{solo} methylation is not affected by maternal dietary methyl supplementation

The IAPs located at the A^{vy} and $Axin^{Fu}$ 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 *C57iap1*^{solo} variant provides an opportunity to test the latter.

To examine whether *C57iap1*^{solo} is responsive to methyl supplementation, *C57iap1*^{solo} females were put on a methyl-supplemented diet two 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 *C57iap1*^{solo} LTR were quantified in eight-week old offspring liver samples. Unlike A^{vy} and $Axin^{Fu}$ loci, *C57iap1*^{solo} 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.

C57iap1 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^{vy} and $Axin^{Fu}$ IAPs which modulate *Agouti* or *Axin* expression in a methylation-dependent manner.

We asked whether allelic variation at the *C57iap1* locus influences the expression of neighbouring genes by quantifying expression of the four closest genes (Figure 5A) in liver,

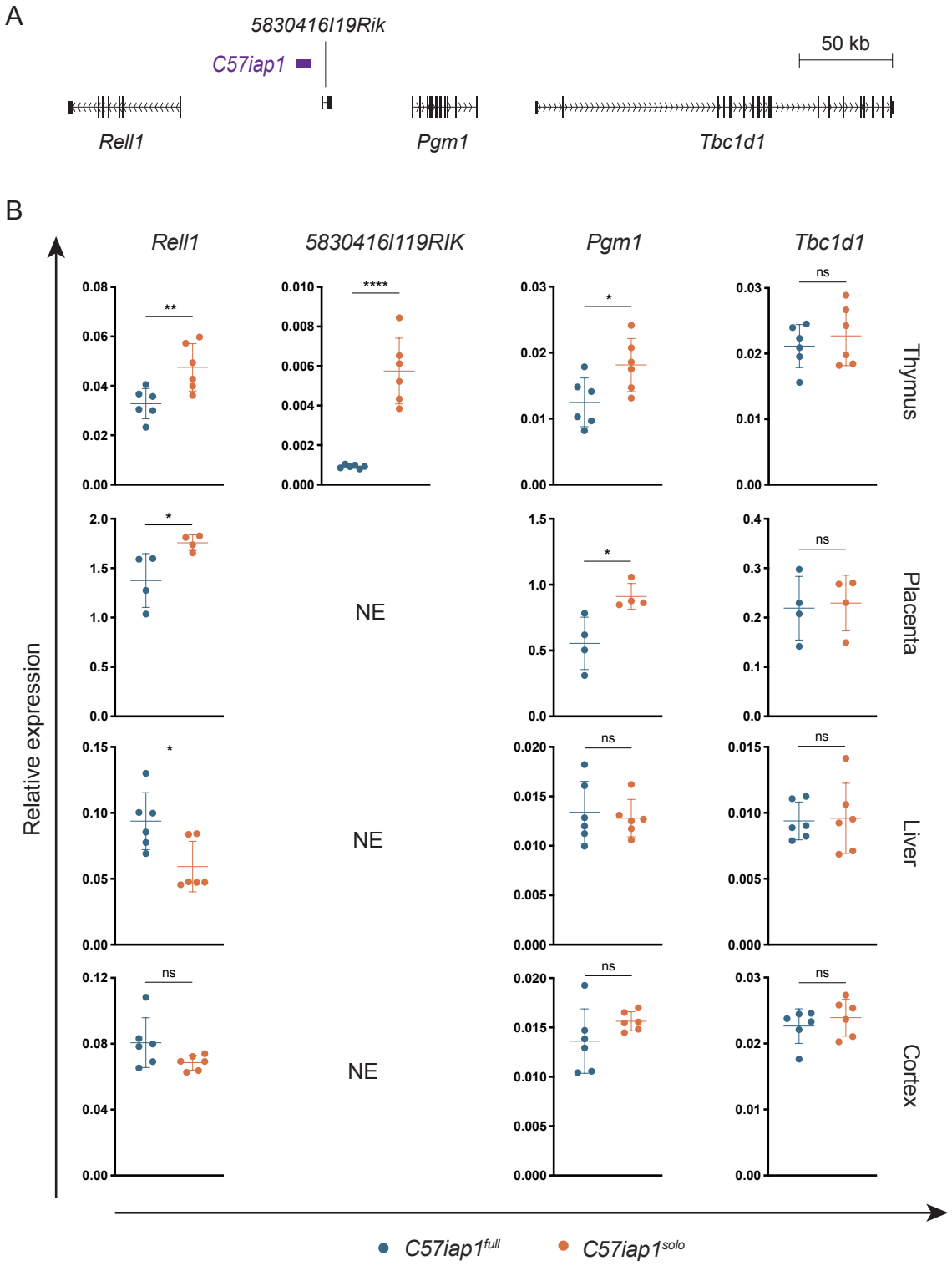

Figure 5. Functional consequences of the genetically-induced epiallele at the *C57iap1* locus. (**A**) Scaled map of the genes surrounding *C57iap1*. Gene transcripts were extracted from the UCSC Genome Browser (Haeussler et al., 2019). (**B**) The *C57iap1* polymorphism influences neighbouring gene expression in a tissue-specific manner. Expression of *Rell1* (exon 7), *5830416I19Rik* (exon 3), *Pgm1* (exon 9), and *Tbc1d1* (exon 3) was quantified in $C57iap1^{full}$ and $C57iap1^{solo}$ thymus, placenta, liver, and brain tissues via RT-qPCR (NE : not expressed). Relative expression was normalised to *Hprt1* and *β-actin* expression and calculated using the ΔCt method. (**C** and **D**) Plasma glucose (C) and triglyceride (D) concentrations in $C57iap1^{full}$ and $C57iap1^{solo}$ adult males. Statistics: unpaired t tests (*p < 0.05; ** p < 0.005; **** p < 0.0001; ns: not significant).

cortex, thymus, and placental tissues collected from *C57iap1*^{full} and *C57iap1*^{solo} individuals. Protein-coding genes *Pgm1*, *TBC1 domain family, member 1* (*Tbc1d1*), and *RELT-like protein 1* (*Rell1*) were expressed in all tissues examined; long non-coding RNA (lncRNA) *5830416l19Rik* transcripts were only detected in the thymus. *Pgm1* and *Rell1* expression levels in the thymus and placenta were significantly higher in *C57iap1*^{solo} than in *C57iap1*^{full} individuals, indicating that in these tissues the unmethylated solo LTR is associated with increased expression (Figure 5B). *Tbc1d1*, the furthest in distance from *C57iap1*, did not display significant differences in expression in any of the tested tissues, suggesting that proximity to the *C57iap1* locus is predictive of its transcriptional effect (Figure 5B). *5830416l19Rik* expression in the thymus was barely detected in *C57iap1*^{full} samples, showing a highly significant increase in *C57iap1*^{solo} 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 inversed, with lower expression levels observed in *C57iap1*^{solo} individuals (Figure 5B). No significant differences in expression were observed for any of the genes in cortex samples (Figure 5B). Together, these data show that the *C57iap1* polymorphism is associated with tissue-specific altered adjacent gene expression.

The formation of the *C57iap1*^{solo} variant is associated with metabolic phenotypes

Little is known about the biological functions of *Rell1* and *5830416l19Rik*, but *Pgm1* is better characterised. *Pgm1* is the closest coding gene to *C57iap1*, lying 55 kb downstream of the 3′ LTR. PGM1 catalyses the interconversion between glucose 1-phosphate and glucose 6-phosphate and has a secondary role to that of the predominant PGM isozyme, PGM2 (Geer et al., 2009). The nomenclature of *Pgm1* and *Pgm2* is inversed in the mouse compared to other vertebrate species (the human orthologue of mouse *Pgm1* is PGM2, and vice versa). PGM1 deficiency in humans is associated with glycogen storage disease and a congenital disorder of

337 glycosylation (Beamer, 2015). PGM2 has been associated with metabolic disease in GWAS
338 studies (Timmons et al., 2018).

339 Considering the established role of *Pgm1* in metabolic pathways, we investigated potential
340 metabolic effects resulting from the formation of the *C57iap1*^{solo} allele. We screened plasma
341 samples collected from *C57iap1*^{full} and *C57iap1*^{solo} adult males for a range of metabolic
342 biomarkers. While most assays revealed no difference between the two variants (Figure 5—
343 figure supplement 1A), we found that fasting plasma glucose and triglyceride concentrations
344 were significantly lower in *C57iap1*^{solo} mice (Figure 5C). Glucose tolerance tests did not show
345 additional metabolic disparities between the two variants (Figure 5—figure supplement 1B).
346 These findings suggest that the emergence of a derepressed solo LTR near a gene involved in
347 glucose metabolism has phenotypic implications, providing a basis for further functional
348 characterisation of this model.

349

350 **Discussion**

351 Historically, inter-LTR recombination events have been identified due to phenotypic
352 reversions of ERV-induced mutations, whereby the phenotypic effect of the insertion is
353 reversed following proviral excision and solo LTR formation (Bultman et al., 1994; Seperack
354 et al., 1988; Stoye et al., 1988). In this study, we identified one such chromosomal event based
355 on its epigenetic rather than phenotypic outcome, revealing *C57iap1* as a genetically-induced
356 epiallele in the ostensibly isogenic B6 mouse strain. We demonstrated that the *C57iap1*^{full} and
357 *C57iap1*^{solo} variants display distinct DNA and H3K9 methylation profiles which are associated
358 with differential adjacent gene expression and altered metabolism, establishing the *C57iap1*
359 locus as a valuable endogenous system to study the mechanisms, evolution, and functional
360 implications of TE repression in both the germline and soma.

The unmethylated status of *C57iap1*^{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 8,000 B6 IAP LTRs and only 14 exhibited less than 20% methylation in tail samples. 10 of these were solo LTRs, mostly of the same subtype as the *C57iap1* LTR (IAPLTR2_mm). While the authors classified the *C57iap1* LTR as a hypomethylated 3′ LTR of a full-length IAP, in all likelihood they unknowingly detected the hypomethylated *C57iap1*^{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 *C57iap1*^{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 *C57iap1*^{full} and *C57iap1*^{solo} mice allowed us to quantitatively explore the functional repercussions of a derepressed solo LTR. We showed that the *C57iap1*^{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: *Rell1* expression was lower in *C57iap1*^{solo} compared to *C57iap1*^{full} livers, and no gene expression differences were observed in cortex samples. The lncRNA *5830416I19Rik*, identified from the sequencing of adult male thymus and aorta cDNA libraries (Kawai et al., 2001), was only expressed in thymus samples collected from *C57iap1*^{solo} individuals; it was nearly undetectable in *C57iap1*^{full} thymic samples and not expressed at all

386 in any of the other tested tissues. We speculate that the formation of *C57iap1*^{solo} gave rise to
387 *5830416I19Rik* as a novel tissue-specific (perhaps immune-specific) lncRNA. Indeed, IAP
388 transcripts are rarely detected in somatic tissues outside of tumours and early embryogenesis,
389 but thymic and activated splenic cells are notable exceptions (Kuff and Lueders, 1988). In
390 addition, young ERVs have recently been implicated in shaping the evolution of transcriptional
391 networks underlying innate immunity (Chuong et al., 2016; Tie et al., 2018; Ye et al., 2020),
392 and our group recently identified individual IAP elements that exhibit inter-individual
393 methylation variability exclusively in B cells (Elmer et al., 2020).

394 Along with the transcriptional differences observed at the metabolic gene *Pgm1*, we found
395 that *C57iap1*^{solo} individuals exhibit lower plasma glucose and triglyceride concentrations
396 compared to *C57iap1*^{full} individuals. While the effect sizes were small, we note that these
397 measurements were taken in an unaltered environment. Administration of a long-term
398 metabolic challenge such as a high-fat diet and subsequent metabolic phenotyping will aid in
399 further elucidating the biological significance of these results. Nevertheless, our work adds to
400 the body of work suggesting that TE insertions and their epigenetic modification can have
401 quantifiable consequences on host metabolic pathways (Du et al., 2016; Kuehnen et al., 2012;
402 Scherneck et al., 2009; Yen et al., 1994).

403 We have shown that *C57iap1*^{solo} is susceptible to genetic background effects, whereby
404 passage through a CAST egg promotes methylation of the B6-derived solo LTR. In a recent
405 study we reported the same effect at a number of VM-IAPs and showed that KZFP
406 diversification is associated with strain-specific IAP methylation (Bertozzi et al., 2020). This
407 combined with the identification of candidate KZFPs capable of binding the *C57iap1* variants
408 suggests that the differential recruitment of heterochromatin factors by KZFPs is involved in
409 both driving the epigenetic differences between *C57iap1*^{full} and *C57iap1*^{solo}, and for the strain-
410 specific methylation of *C57iap1*^{solo}. In fact, by definition, the wide range of methylation levels

411 observed at the *C57iap1*^{solo} allele across genetically identical CB individuals renders this locus
412 a bona fide VM-IAP (or *metastable epiallele*) in this hybrid context (Bertozzi and Ferguson-
413 Smith, 2020; Rakyan et al., 2002). Therefore, we expect the new *C57iap1*^{solo} variant to be a
414 useful tool for the study of epigenetic stochasticity in the absence of genetic variation, a
415 phenomenon for which the underlying mechanisms remain poorly understood.

416 The identification of *C57iap1*^{solo} demonstrates that cryptic genetic diversity in an inbred
417 mouse population can have functional repercussions with important implications for
418 experimental outcomes. It serves as a cautionary tale for researchers working with inbred
419 mouse colonies, particularly in the field of epigenetics where ruling out genetic effects is often
420 paramount. We are reminded that the current mouse reference genome harbours large gaps and
421 inaccuracies, largely due to mapping difficulties associated with the repeat genome. Inter-LTR
422 recombination events are surely an underappreciated source of genetic variation considering
423 that a lack of uniquely mapped reads in internal portions of TEs is more likely to be attributed
424 to technical rather than biological limitations. A recent study in humans developed a
425 computational pipeline to capture dimorphic human ERVs (HERVs) resulting from inter-LTR
426 recombination events and detected dozens of previously unidentified candidates (Thomas et
427 al., 2018). The advent of such analytical tools as well as routine long-read sequencing of whole
428 genomes will be highly beneficial in addressing these issues.

429 In summary, we have identified and characterised a recent spontaneous inter-LTR
430 recombination event at the *C57iap1* locus, introducing a genetic variant in the commonly
431 investigated and supposedly inbred B6 mouse strain. The ancestral variant, *C57iap1*^{full}, is a
432 full-length IAP repressed by DNA methylation and H3K9 trimethylation, as is typical for this
433 class of evolutionary young ERV. The recently formed solo LTR variant, *C57iap1*^{solo}, lacks
434 these silencing marks and is associated with metabolic changes and differential neighbouring
435 gene expression. Our study lays the foundation for further comparative studies between the

436 *C57iap1*^{full} and *C57iap1*^{solo} variants aimed at better understanding the epigenetic and functional
437 consequences of TE structural variation and evolution.

438

439 **Materials and Methods**

440 **Mice**

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

451

452 **Dietary methyl supplementation**

453 Adult *C57iap1*^{solo} females were placed on control diet (RM3(E); Special Diet Services) or
454 methyl supplemented diet (RM3(E) supplemented with 15 g Choline, 15 g Betaine, 15 mg Folic
455 acid, 1.5 mg Vitamin B12, 7.5 g L-methionine, 150 mg Zinc; Special Diet Services) two weeks
456 prior to mating with *C57iap1*^{solo} males and kept on their respective diets throughout pregnancy
457 and lactation. Eight dams were used for each dietary group. The methyl supplemented diet
458 recipe matches the 3SZM diet in Wolff et al., 1998. All offspring were weaned onto the control
459 diet and culled at eight weeks of age. DNA methylation was quantified at the distal CpGs of

460 the 5′ end of *C57iap1*^{solo} in all offspring livers (control diet: 38 born from 8 litters; methyl
461 supplemented diet: 41 born from 8 litters). Statistics were carried out on litter averages because
462 the dietary intervention was done on the dams, not the offspring.

463

464 **Metabolic biomarker assays**

465 Blood was sampled from 4-month old *C57iap1*^{full} and *C57iap1*^{solo} males via cardiac puncture
466 following Isoflurane-induced deep terminal anaesthesia. Blood samples were placed in EDTA-
467 coated tubes and centrifuged at 1,500 x g for 15 min at 4°C. Plasma was collected from the
468 separated upper phase and stored at -80°C before sending to the Core biochemical assay
469 laboratory (CBAL) at the Cambridge University Hospitals for analysis. Assays performed on
470 the plasma samples included Adiponectin (ng/ml), Albumin (g/L), ALT (U/L), Cholesterol
471 (mmol/L), Corticosterone (ng/ml), Glucagon (pg/ml), Glucose (mmol/L), HDL (mmol/L),
472 Insulin (ug/L), LDL (mmol/L), Leptin (pg/ml), NEFA (umol/L), and Triglycerides (mmol/L).

473

474 **Glucose tolerance tests**

475 Glucose tolerance tests were conducted on *C57iap1*^{full} and *C57iap1*^{solo} 16-week-old males (n
476 = 12 per genotype). Following an overnight fast of 15 hours, mice were weighed and baseline
477 blood glucose levels (time 0) were measured using the Glucomen Areo glucometer. A glucose
478 dosage of 2 g/kg was calculated for each mouse based on weight and administered via
479 intraperitoneal injection. Tail blood glucose levels were measured 15, 30, 45, 60, 90, and 120
480 minutes post injection. Blood glucose levels were compared between *C57iap1*^{full} and
481 *C57iap1*^{solo} mice at each time point using multiple unpaired t tests corrected for multiple
482 comparisons using the Holm-Šidák method.

483

Isolation and storage of biological samples

Somatic tissues. Liver, cortex, and thymus tissues dissected from adult *C57iap1*^{full} and *C57iap1*^{solo} males were snap frozen in liquid nitrogen and stored at -80°C before use. Ear notches were stored at -20°C.

GV oocytes. C57iap1^{full} and *C57iap1*^{solo} females were injected interperitoneally with 5 IU of gonadotropin. Ovaries were dissected 40 hours later and placed in M2 medium supplemented with 0.06 g/L potassium penicillin-G, 0.05 g/L streptomycin sulfate, and 240 uM dbcAMP. Ovary follicles were punctures 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°C in pools of 100 oocytes collected from multiple females with matching *C57iap1* genotype.

Sperm. Mature sperm was isolated from the cauda epididymides of fertile adult *C57iap1*^{full} and *C57iap1*^{solo} males as previously described (Sharma et al., 2015) and stored in 1X PBS at -80°C. Sperm samples from different males were analysed separately.

Blastocysts. *C57iap1*^{full} and *C57iap1*^{solo} females were mated with *C57iap1*^{full} and *C57iap1*^{solo} males, respectively. The uteri and oviducts of pregnant females were dissected three 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 sulfate. Embryos developed to the blastocyst stage were washed in M2 medium and stored at -80°C. Blastocysts collected from each female were pooled prior to freezing.

E16.5 embryos and placentas. C57iap1^{full} and *C57iap1*^{solo} females were mated with *C57iap1*^{full} and *C57iap1*^{solo} males, respectively. 16 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 1X PBS. After removal of the yolk sac and an additional rinse in cold 1X PBS, embryonic tails and full placentas were flash frozen in liquid nitrogen and stored at -80°C.

508 **DNA extraction**

509 For somatic tissues and ES cells, samples were treated with RNase A at 37°C for 60 min and
510 digested with Proteinase K at 55°C overnight in lysis buffer (10 mM EDTA, 150 mM NaCl,
511 10 mM Tris-HCl pH 8, 0.1% SDS). gDNA was isolated the next day using a standard phenol-
512 chloroform extraction and ethanol precipitation protocol. The same protocol was followed for
513 sperm gDNA extraction except that the Proteinase K digestion was carried out in equal volumes
514 of Solution A (75 mM NaCl pH 8; 25 mM EDTA) and Solution B (10 mM Tris-HCl pH 8; 10
515 mM EDTA; 1% SDS; 80 mM DTT) following centrifugation and removal of PBS from thawed
516 sperm. To extract oocyte and blastocyst gDNA, pooled GV oocytes or blastocysts were
517 incubated for 1 hr at 37°C in 14 µl ddH_2O, 1 µl 1 mg/ml Carrier RNA (QIAGEN), 1 µl 10%
518 SDS, and 1 µl 10 mg/ml Proteinase K. After a 15 min incubation at 98°C, samples was directly
519 bisulphite-converted as described below.

520

521 **Generation of ESC lines**

522 ESC lines were generated as previously described (Nichols and Jones, 2017) with the following
523 modifications. The concentration of MEK inhibitor PDO325901 used in the KSOM + 2i and
524 N2B27+2i+LIF media was reduced to 0.2 µM for the initial stages of the protocol and increased
525 to 1 µM following disaggregation of the ICM. The zona pellucide was removed from unhatched
526 embryos using a10 min pronase digestion at 37°C rather than using acidic Tyrode's solution,
527 and rat serum was used as a source of complement instead of guinea pig serum. Laminin-coated
528 wells were used to ensure proper cell attachment and Accutase solution was used to detach
529 cells prior to passaging. Only male embryos were kept for the generation of ESC lines.

530

531

Genotyping

Ear notches were used for *C57iap1* allelic variant genotyping. Ear notch gDNA was extracted using the PCRBIO Rapid Extract lysis kit (PCR Biosystems) and 1 µl 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 trophectoderm sex-genotyping, trophectoderm lysates were placed in PCR buffer with Proteinase K (50 mM KCl, 10 mM TrisHCl pH 8.3, 2.5 mM MgCl2, 0.1 mg/mL gelatin, 0.45% NP40, 0.45% Tween 20, 200 µg/mL Pro K) and incubated at 55°C for one hour followed by 95°C for 10 min. Trophectoderm gDNA samples were sex-genotyped by PCR using HotStarTaq DNA Polymerase (QIAGEN) and 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 *C57iap1* allelic variants and the Y-linked *Sry* gene are listed in Supplementary Table 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-*Pgm1* prior to Sanger sequencing, respectively (Figure 2A, Supplementary Table 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 Table 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_2O. Sanger sequencing was carried out by Source BioScience. Sequencing primers were interspersed regularly across the IAP element (Supplementary Table 2). Sequence traces were visually examined, and reliable sequences were merged using the EMBOSS *merger* tool. The resulting sequences for the two *C57iap1* alleles were aligned to the GRCm38/mm10 reference sequence using CLC Sequence Viewer 6. The base-resolution sequence for the *C57iap1*^{solo} allele has been uploaded to GenBank.

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 µg 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 Table 3). Target regions were PCR-amplified in technical triplicates from bilsulphite-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 µL 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 pH7.6, 2M NaCl, 1 mM EDTA, 0.1% Tween-20) at 1,400 rpm for 5 min. The bead-bound biotinylated strands were washed consecutively in 70% Ethanol, denaturation solution (0.2M 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 pH7.6, 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. Percent 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.

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 *C57iap1*^{full} and *C57iap1*^{solo} males was crosslinked in 1% formadehyde 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 1X PBS supplemented with EDTA-free protease inhibitor cocktail cOmplete™ (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, 1X EDTA-free cOmpleteTM; one wash), LB2 buffer (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA,

606 0.5 mM EGTA, 1X EDTA-free cOmpleteTM; one wash), and SDS shearing buffer (10 mM
607 Tris-HCl pH 8, 1 mM EDTA, 0.15% SDS, 1X EDTA-free cOmpleteTM; three washes).
608 Samples were centrifuged at 1,700 x g for 5 min at 4°C after every wash. The resulting
609 chromatin was sonicated for 8 cycles (one cycle: 30 s on and 30 s off) at 4°C using a Bioruptor.
610 The sonicated lysate was cleared by centrifugation at maximum speed for 10 min at 4°C and
611 10% of the input for each ChIP was stored at -20°C. 50 µL magnetic beads (Protein G
612 Dynabeads, Invitrogen) were pre-blocked in fresh blocking buffer (0.5% BSA in PBS) and
613 incubated with 2.5 µL of polyclonal H3K9me3 antibody (AB_2532132, Active Motif) or
614 Rabbit IgG (negative control) for 4 hrs at 4°C. Antibody-bound beads were washed twice in
615 blocking buffer using a magnetic stand at 4°C. The cleared lysate was topped up to 1 mL SDS
616 shearing buffer + 150 mM NaCl and 1% Triton-X-100) and incubated with the antibody-bound
617 beads on a rotating wheel overnight at 4°C. Non-specifically bound proteins were removed
618 with the following sequential washes at 4°C: low salt buffer (10 mM Tris-HCl pH 8.0, 150 mM
619 NaCl, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM PMSF; two washes), high salt buffer
620 (10 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.15% SDS, 1 mM
621 PMSF; one wash), LiCl buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 250
622 mM LiCl, 1% NP40, 1% Na- deoxycholate, 1 mM PMSF; one wash), and 10 mM Tris pH 8.0
623 (one wash). ChIP samples were resuspended in elution buffer (10 mM Tris pH 8.0, 1 mM
624 EDTA, 1% SDS, 150 mM NaCl), treated with RNaseA at 37°C for 1 hr at 1,100 rpm, and
625 reverse crosslinked overnight at 65°C at 1,100 rpm. The eluted samples were treated with
626 Proteinase K and DNA was purified using the Monarch PCR and DNA Cleanup Kit (NEB).

627

628 **RNA extraction and cDNA synthesis**

629 20-30 mg of thymus, liver, cortex, or placenta tissues was homogenised using the MagNA
630 Lyser (Roche) at 6,000 x g for 40 s. 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)

qPCR primers were designed using Primer3 software (Supplementary Table 4). Each reaction were 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 10s, 72°C for 10s, 45 cycles; (3) melting curve analysis of 65°C to 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 *β-actin* expression and calculated using the ΔCt 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 percent input.

Computational and statistical analyses

ChIP-seq datasets 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 Table 5. B6 and CAST DNA sequences at the *C57iap1* insertion site were extracted from the GRCm38/mm10 and CAST_EiJ_v1 assemblies, respectively, accessed through the UCSC genome browser. The sequence alignment was

655 generated using CLUSTAL OMEGA. All statistical tests in this study were carried out using
656 GraphPad Prism 8 software as indicated in the figure legends.

657

658 **Acknowledgements**

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

666

667 **Competing interests**

668 No competing interests to declare.

669

670 **References**

671 Allen ND, Norris ML, Surani MA. 1990. Epigenetic control of transgene expression and
672 imprinting by genotype-specific modifiers. *Cell* **61**:853–861. doi:10.1016/0092-
673 8674(90)90195-K

674 Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally
675 regulated piRNA clusters implicate MILI in transposon control. *Science* **316**:744–747.
676 doi:10.1126/science.1142612

677 Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Hérault Y, Guillou F, Bourc'his D.
678 2016. The DNA methyltransferase DNMT3C protects male germ cells from transposon
679 activity. *Science* **354**:909–912. doi:10.1126/science.aah5143

680 Beamer LJ. 2015. Mutations in hereditary phosphoglucomutase 1 deficiency map to key
681 regions of enzyme structure and function. *J Inherit Metab Dis* **38**:243–256.
682 doi:10.1007/s10545-014-9757-9

683 Bertozzi TM, Elmer JL, Macfarlan TS, Ferguson-Smith AC. 2020. KRAB zinc finger protein
684 diversification drives mammalian interindividual methylation variability. *Proc Natl*
685 *Acad Sci* 202017053. doi:10.1073/pnas.2017053117

686 Bertozzi TM, Ferguson-Smith AC. 2020. Metastable epialleles and their contribution to
687 epigenetic inheritance in mammals. *Semin Cell Dev Biol* **97**:93–105.
688 doi:10.1016/j.semcdb.2019.08.002

689 Bourc'his D, Bestor TH. 2004. Meiotic catastrophe and retrotransposon reactivation in male
690 germ cells lacking Dnmt3L. *Nature* **431**:96–99. doi:10.1038/nature02886

691 Bultman SJ, Klebig ML, Michaud EJ, Sweet HO, Davisson MT, Woychik RP. 1994.
692 Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (a(t)) and
693 white-bellied agouti (Aw) reveals alternative forms of agouti transcripts. *Genes Dev*
694 **8**:481–490. doi:10.1101/gad.8.4.481

695 Carmell MA, Girard A, van de Kant HJG, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ.
696 2007. MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the
697 Mouse Male Germline. *Dev Cell* **12**:503–514. doi:10.1016/J.DEVCEL.2007.03.001

698 Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through
699 co-option of endogenous retroviruses. *Science* **351**:1083–1087.
700 doi:10.1126/science.aad5497

701 Du J, Leung A, Trac C, Lee M, Parks BW, Lusis AJ, Natarajan R, Schones DE. 2016.
702 Chromatin variation associated with liver metabolism is mediated by transposable
703 elements. *Epigenetics and Chromatin* **9**:1–16. doi:10.1186/s13072-016-0078-0

704 Duhl DMJ, Vrieling H, Miller KA, Wolff GL, Barsh GS. 1994. Neomorphic agouti mutations
705 in obese yellow mice. *Nat Genet* **8**:59–65. doi:10.1038/ng0994-59

706 Ecco G, Imbeault M, Trono D. 2017. KRAB zinc finger proteins. *Development* **144**:2719–
707 2729. doi:10.1242/dev.132605

708 Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, Mccune RA, Gehrke C.
709 1982. Amount and distribution of 5-methylcytosine in human DNA from different types
710 of tissues or cells. *Nucleic Acids Res* **10**:2709–2721. doi:10.1093/nar/10.8.2709

711 Elmer JL, Ferguson-Smith AC. 2020. Strain-Specific Epigenetic Regulation of Endogenous
712 Retroviruses: The Role of Trans-Acting Modifiers. *Viruses* **12**:1–21.
713 doi:10.3390/v12080810

714 Elmer JL, Hay AD, Kessler NJ, Bertozzi TM, Ainscough E, Ferguson-smith AC. 2020.
715 Genomic properties of variably methylated retrotransposons in mouse. *bioRxiv*
716 2020.10.21.349217. doi:10.1101/2020.10.21.349217

717 Gagnier L, Belancio VP, Mager DL. 2019. Mouse germ line mutations due to retrotransposon
718 insertions. *Mob DNA* **10**:15. doi:10.1186/s13100-019-0157-4

719 Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH. 2009.
720 The NCBI BioSystems database. *Nucleic Acids Res* **38**:D492–D496.
721 doi:10.1093/nar/gkp858

722 Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT,
723 Hinrichs AS, Gonzalez JN, Gibson D, Diekhans M, Clawson H, Casper J, Barber GP,
724 Haussler D, Kuhn RM, Kent WJ. 2019. The UCSC Genome Browser database: 2019
725 update. *Nucleic Acids Res* **47**:D853–D858. doi:10.1093/nar/gky1095

726 Imbeault M, Helleboid PY, Trono D. 2017. KRAB zinc-finger proteins contribute to the

727 evolution of gene regulatory networks. *Nature* **543**:550–554. doi:10.1038/nature21683

728 Jain D, Meydan C, Lange J, Claeys Bouuaert C, Lailler N, Mason CE, Anderson K V.,
729 Keeney S. 2017. rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase
730 homolog required for meiosis and transposon repression in the mouse male germline.
731 *PLOS Genet* **13**:e1006964. doi:10.1371/journal.pgen.1006964

732 Jern P, Coffin JM. 2008. Effects of Retroviruses on Host Genome Function. *Annu Rev Genet*
733 **42**:709–732. doi:10.1146/annurev.genet.42.110807.091501

734 Kawai J, Shinagawa A, Shibata K, Yoshino M, Itoh M, Ishii Y, Arakawa T, Hara A,
735 Fukunishi Y, Konno H, Adachi J, Fukuda S, Aizawa K, Izawa M, Nishi K, Kiyosawa H,
736 Kondo S, Yamanaka I, Saito T, Okazaki Y, Gojobori T, Bono H, Kasukawa T, Saito R,
737 Kadota K, Matsuda H, Ashburner M, Batalov S, Casavant T, Fleischmann W,
738 Gaasterland T, Gissi C, King B, Kochiwa H, Kuehl P, Lewis S, Matsuo Y, Nikaido I,
739 Pesole G, Quackenbush J, Schriml LM, Staubli F, Suzuki R, Tomita M, Wagner L,
740 Washio T, Sakai K, Okido T, Furuno M, Aono H, Baldarelli R, Barsh G, Blake J,
741 Boffelli D, Bojunga N, Carninci P, De Bonaldo MF, Brownstein MJ, Bult C, Fletcher C,
742 Fujita M, Gariboldi M, Gustincich S, Hill D, Hofmann M, Hume DA, Kamiya M, Lee
743 NH, Lyons P, Marchionni L, Mashima J, Mazzarelli J, Mombaerts P, Nordone P, Ring
744 B, Ringwald M, Rodriguez I, Sakamoto N, Sasaki H, Sato K, Schönbach C, Seya T,
745 Shibata Y, Storch KF, Suzuki H, Toyo-Oka K, Wang KH, Weitz C, Whittaker C,
746 Wilming L, Wynshaw-Boris A, Yoshida K, Hasegawa Y, Kawaji H, Kohtsuki S,
747 Hayashizaki Y. 2001. Functional annotation of a full-length mouse cDNA collection.
748 *Nature* **409**:685–689. doi:10.1038/35055500

749 Kazachenka A, Bertozzi TM, Sjoberg-Herrera MK, Walker N, Gardner J, Gunning R, Pahita
750 E, Adams S, Adams D, Ferguson-Smith AC. 2018. Identification, Characterization, and
751 Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance.
752 *Cell* **175**:1259–1271. doi:10.1016/J.CELL.2018.09.043

753 Kearns M, Preis J, Mcdonald M, Morris C, Whitelaw E. 2000. Complex patterns of
754 inheritance of an imprinted murine transgene suggest incomplete germline erasure.
755 *Nucleic Acids Res* **28**:3301–3309. doi:10.1093/nar/28.17.3301

756 Kuehnen P, Mischke M, Wiegand S, Sers C, Horsthemke B, Lau S, Keil T, Lee YA, Grueters
757 A, Krude H. 2012. An alu element-associated hypermethylation variant of the POMC
758 gene is associated with childhood obesity. *PLoS Genet* **8**:1–12.
759 doi:10.1371/journal.pgen.1002543

760 Kuff EL, Lueders KK. 1988. The Intracisternal A-Particle Gene Family: Structure and
761 Functional Aspects. *Adv Cancer Res* **51**:183–276. doi:10.1016/S0065-230X(08)60223-7

762 Lane N, Dean W, Erhardt S, Hajkova P, Surani A, Walter J, Reik W. 2003. Resistance of
763 IAPs to methylation reprogramming may provide a mechanism for epigenetic
764 inheritance in the mouse. *Genesis* **35**:88–93. doi:10.1002/gene.10168

765 Maksakova IA, Romanish MT, Gagnier L, Dunn CA, Van De Lagemaat LN, Mager DL.
766 2006. Retroviral elements and their hosts: Insertional mutagenesis in the mouse germ
767 line. *PLoS Genet* **2**:1–10. doi:10.1371/journal.pgen.0020002

768 Matsui T, Leung D, Miyashita H, Maksakova I a, Miyachi H, Kimura H, Tachibana M,
769 Lorincz MC, Shinkai Y. 2010. Proviral silencing in embryonic stem cells requires the
770 histone methyltransferase ESET. *Nature* **464**:927–931. doi:10.1038/nature08858

771 Morgan HD, Sutherland HG, Martin DI, Whitelaw E. 1999. Epigenetic inheritance at the

agouti locus in the mouse. *Nat Genet* **23**:314–318. doi:10.1038/15490

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 Biol* **13**:R45. doi:10.1186/gb-2012-13-6-r45

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 Harb Protoc* **2017**:379–386. doi:10.1101/pdb.prot094086

Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. 2002. Metastable epialleles in mammals. *Trends Genet* **18**:348–351. doi:10.1016/S0168-9525(02)02709-9

Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KVK, Whitelaw E. 2003. Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proc Natl Acad Sci U S A* **100**:2538–2543. doi:10.1073/pnas.0436776100

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 Res* **104**:11–16. doi:10.1016/j.virusres.2004.02.036

Scherneck S, Nestler M, Vogel H, Blüher M, Block MD, Diaz MB, 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 Genet* **5**:e1000541. doi:10.1371/journal.pgen.1000541

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 Genet* **11**:1–20. doi:10.1371/journal.pgen.1005442

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. *Mol Cell* **48**:849–862. doi:10.1016/j.molcel.2012.11.001

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. *Proc Natl Acad Sci U S A* **85**:189–192. doi:10.1073/pnas.85.1.189

Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, Kucukural A, Serra RW, Sun F, Song L, Carone BR, Emiliano P, Li XZ, Fauquier L, Moore MJ, Sullivan R, Mello CC, Rando OJ. 2015. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals **6780**:1–9.

Shimosuga KI, Fukuda K, Sasaki H, Ichiyanagi K. 2017. Locus-specific hypomethylation of the mouse IAP retrotransposon is associated with transcription factor-binding sites. *Mob DNA* **8**:1–12. doi: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:10.1016/0092-8674(88)90201-2

816 Subramanian RP, Wildschutte JH, Russo C, Coffin JM. 2011. Identification, characterization,
817 and comparative genomic distribution of the HERV-K (HML-2) group of human
818 endogenous retroviruses. *Retrovirology* **8**:90. doi:10.1186/1742-4690-8-90

819 Thomas J, Perron H, Feschotte C. 2018. Variation in proviral content among human genomes
820 mediated by LTR recombination. *Mob DNA* **9**:36. doi:10.1186/s13100-018-0142-3

821 Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV):
822 High-performance genomics data visualization and exploration. *Brief Bioinform* **14**:178–
823 192. doi:10.1093/bib/bbs017

824 Tie CH, Fernandes L, Conde L, Robbez-Masson L, Sumner RP, Peacock T, Rodriguez-Plata
825 MT, Mickute G, Gifford R, Towers GJ, Herrero J, Rowe HM. 2018. KAP 1 regulates
826 endogenous retroviruses in adult human cells and contributes to innate immune control.
827 *EMBO Rep* **19**:e45000. doi:10.15252/embr.201745000

828 Timmons JA, Atherton PJ, Larsson O, Sood S, Blokhin IO, Brogan RJ, Volmar C-H, Josse
829 AR, Slentz C, Wahlestedt C, Phillips SM, Phillips BE, Gallagher IJ, Kraus WE. 2018. A
830 coding and non-coding transcriptomic perspective on the genomics of human metabolic
831 disease. *Nucleic Acids Res* **46**:7772–7792. doi:10.1093/nar/gky570

832 Walsh CP, Chaillet JR, Bestor TH. 1998. Transcription of IAP endogenous retroviruses is
833 constrained by cytosine methylation. *Nat Genet* **20**:116–117. doi:10.1038/2413

834 Waterland RA, Dolinoy DC, Lin JR, Smith CA, Shi X, Tahiliani KG. 2006. Maternal Methyl
835 Supplements Increase Offspring DNA Methylation at Axin Fused. *Genesis* **44**:401–406.
836 doi:10.1002/dvg

837 Waterland RA, Jirtle RL. 2003. Transposable Elements : Targets for Early Nutritional Effects
838 on Epigenetic Gene Regulation. *Mol Cell Biol* **23**:5293–5300.
839 doi:10.1128/MCB.23.15.5293

840 Wolf G, de Iaco A, Sun MA, Bruno M, Tinkham M, Hoang D, Mitra A, Ralls S, Trono D,
841 Macfarlan TS. 2020. Krab-zinc finger protein gene expansion in response to active
842 retrotransposons in the murine lineage. *Elife* **9**:1–22. doi:10.7554/eLife.56337

843 Wolff GL. 1971. Genetic modification of homeostatic regulation in the mouse. *Am Nat*
844 **105**:241–252. doi:10.1086/282721

845 Wolff GL, Kodell RL, Moore SR, Cooney CA. 1998. Maternal epigenetics and methyl
846 supplements affect agouti gene expression in Avy/a mice. *FASEB J* **12**:949–957.
847 doi:10.1096/fasebj.12.11.949

848 Ye M, Goudot C, Hoyler T, Lemoine B, Amigorena S, Zueva E. 2020. Specific subfamilies
849 of transposable elements contribute to different domains of T lymphocyte enhancers.
850 *Proc Natl Acad Sci U S A* **117**:7905–7916. doi:10.1073/pnas.1912008117

851 Yen TT, Gill AM, Frigeri LG, Barsh GS, Wolff GL. 1994. Obesity, diabetes, and neoplasia in
852 yellow A(vy)/- mice: ectopic expression of the agouti gene. *FASEB J* **8**:479–488.
853 doi:10.1096/fasebj.8.8.8181666

854
855