# KRAB zinc finger protein diversification drives mammalian inter-individual methylation variability

Tessa M. Bertozzi[a], Jessica L. Elmer[a], Todd S. MacFarlan[b], Anne C. Ferguson-Smith[a,1]

**Affiliations:**

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

[b] The Eunice Kennedy Shriver National Institute of Child Health and Human Development, The National Institutes of Health, Bethesda, MD 20892, USA

[1] To whom correspondence may be addressed. Email: afsmith@gen.cam.ac.uk; Phone number: +44 (0)1223 333834

**Major classification:**

Biological Sciences

**Minor classification:**

Genetics

**Keywords:**

DNA methylation, endogenous retrovirus, KRAB zinc finger proteins, metastable epiallele, VM-IAP

## Abstract

Most transposable elements (TEs) in the mouse genome are heavily modified by DNA methylation and repressive histone modifications. However, a subset of TEs exhibit variable methylation levels in genetically identical individuals and this is associated with epigenetically-conferred phenotypic differences, environmental adaptability, and transgenerational epigenetic inheritance. The evolutionary origins and molecular mechanisms underlying inter-individual epigenetic variability remain unknown. Using a repertoire of murine variably methylated intracisternal A-particle (VM-IAP) epialleles as a model, we demonstrate that variable DNA methylation states at TEs are highly susceptible to genetic background effects. Taking a classical genetics approach coupled with genome-wide analysis, we harness these effects and identify a cluster of KRAB zinc finger protein (KZFP) genes that modifies VM-IAPs in *trans* in a sequence-specific manner. Deletion of the cluster results in decreased DNA methylation levels at these variably methylated alleles. An analysis of ChIP-seq and RNA-seq datasets generated from KZFP-cluster mutants reveals that the observed DNA methylation changes are accompanied by altered histone modifications and by dysregulation of neighbouring genes. We find that VM-IAPs cluster together phylogenetically and that this is linked to differential KZFP binding, suggestive of an ongoing evolutionary arms race between TEs and this large family of epigenetic regulators. These findings indicate that KZFP divergence and concomitant evolution of DNA binding capabilities are mechanistically linked to methylation variability in mammals, with implications for phenotypic variation and putative paradigms of mammalian epigenetic inheritance.

## Significance statement

Transposable elements (TEs) are repetitive sequences with potential to mobilise, causing genetic diversity. To restrict this, most TEs in the mouse are heavily epigenetically modified by DNA methylation. However, a few TEs exhibit variable methylation levels that differ between individuals and confer an epigenetic, rather than genetic, influence on phenotype. The mechanism underlying this remains unknown. We report the identification of a polymorphic cluster of KRAB zinc finger proteins (KZFPs) responsible for the epigenetic properties of these variably methylated TEs, with deletion of the cluster profoundly influencing their DNA methylation and expression of adjacent genes. We propose that rapid KZFP divergence underlies variable epigenetic states in mammals, with implications for epigenetically conferred phenotypic differences between individuals within and across generations.

## Main text

## Introduction

Complex genetic interactions contribute to evolutionary fitness, phenotypic variation, and disease risk. This is highlighted by comparative research across inbred mouse strains showing that genetic background not only influences basic fitness traits such as litter size and sperm count, but also modulates the penetrance and expressivity of numerous gene mutations (1, 2). Despite the extensive documentation of strain-specific epistatic effects in the mouse and their important implications for mechanistic insight and experimental reproducibility, the underlying modifier genes remain uncharacterised in most cases.

Studies on foreign DNA insertions in the mouse genome demonstrate that modifier genes can act via epigenetic pathways to drive genetic background-dependent phenotypes. A number of transgenes show predictable strain-specific DNA methylation patterns that are associated with transgene expression levels (3–6). Similar effects have been reported on the methylation state of endogenous retroviruses (ERVs), as exemplified by the MusD ERV insertion *Dac*$^{1J}$ which is methylated in mouse strains that carry the unlinked *Mdac* modifier gene (7, 8). In strains lacking the *Mdac* allele, *Dac*$^{1J}$ is unmethylated and the mice exhibit limb malformation.

Another example is provided by the *Agouti viable yellow* (*A*$^{vy}$) metastable epiallele, where a spontaneously inserted intracisternal A-particle (IAP) element influences the expression of the downstream coat-colour gene *Agouti* (9). IAPs are an evolutionarily young and highly active class of ERV (10). Variable DNA methylation of the *A*$^{vy}$ IAP is established early in development across genetically identical mice and is correlated with a spectrum of coat colour phenotypes, which in turn display transgenerational inheritance and environmental sensitivity (11–13). Both the distribution and heritability of *A*$^{vy}$ phenotypes are influenced by genetic background (14–16).

Therefore, the identification and characterisation of the responsible modifier genes can provide insight into the mechanisms governing the early establishment of stochastic methylation states at mammalian transposable elements.

We recently conducted a genome-wide screen for individual variably methylated IAPs (VM-IAPs) in the C57BL/6J (B6) inbred mouse strain (17, 18). The screen yielded a robust set of experimentally validated regions to use as a model to investigate inter-individual epigenetic variability. Most VM-IAPs belong to the IAPLTR1_Mm and IAPLTR2_Mm subclasses. Approximately half of them are full-length IAPs with an internal coding region flanked by near-identical long terminal repeats (LTRs); the other half are solo LTRs. While solo LTRs lack autonomous retrotransposition potential, they are rich in regulatory sequences and thus have the ability to affect host gene expression. As observed for $A^{vy}$, methylation variability is re-established at VM-IAPs from one generation to the next regardless of parental methylation states. Importantly, VM-IAP methylation levels are consistent across all tissues of a single mouse, suggesting that individual-specific methylation states are acquired in early development prior to tissue differentiation. The inter-individual variability suggests that the establishment of VM-IAP methylation levels involves an early stochastic phase.

Here, we introduce genetic variation to the study of VM-IAPs. We report that half of the IAPs found to be variably methylated in B6 are present in 129 sub-strains while the vast majority are absent from the CAST/EiJ (CAST) genome. We find that a subset of the shared loci between B6 and 129 display variable methylation in both stains; the remainder are hypermethylated in 129. Further methylation quantification in reciprocal B6 x CAST F1 hybrids reveals pervasive maternal and zygotic genetic background effects. Through backcrossing and genetic mapping experiments, we identify a cluster of KRAB zinc finger proteins (KZFPs) on Chromosome 4 responsible for the

strain-specific *trans*-acting hypermethylation of multiple B6 VM-IAPs. We show that deletion of the KZFP cluster leads to a decrease in DNA and H3K9 methylation, an increase in H3K4 trimethylation, and alterations in nearby gene expression at the targeted VM-IAPs. A phylogenetic sequence analysis demonstrates that genetic sequence plays a crucial role not only in the targeting of VM-IAPs by strain-specific KZFPs, but also in the establishment of methylation variability in a pure B6 context. Based on our findings, we propose that KZFP diversification is at the centre of the mechanism leading to variable epigenetic states within and across mouse strains.

## Results

### VM-IAPs exhibit strain-specific methylation states

To determine whether B6 VM-IAPs are variably methylated in other inbred mouse strains, we first catalogued their presence or absence in the 129S1/SvlmJ (129) and CAST strains based on a previous analysis of polymorphic ERVs (19). The classification was verified, and at times corrected, by visually assessing each locus in the 129 and CAST reference genomes (20). Out of 51 experimentally validated B6 VM-IAPs (18), 25 of the IAPs are present in 129 and 3 are present in CAST (Fig. 1*A*). These numbers are consistent with our previous work showing that VM-IAPs are evolutionarily young IAPs (17), and were expected given the evolutionary relationship between these three strains: B6 and 129 are classical inbred laboratory strains derived from several subspecies while CAST is wild-derived and evolutionarily more distant.

We compared the methylation level of 19 loci conserved between B6 and 129 by bisulphite pyrosequencing. The probed cytosine-guanine dinucleotide (CpG) sites are comparable across loci and located at the most distal end of the 5′ LTR of each element, close enough to the bordering

unique sequence to ensure amplification of a single product. As expected, all 19 regions exhibited methylation variability across inbred B6 mice (i.e. more than 10% variability across individuals). In contrast, only eight loci were variably methylated in 129 mice (Fig. 1*B*). Most of these displayed distinct methylation ranges compared to those observed in B6. The remaining 11 IAPs lacked inter-individual variability and are therefore not VM-IAPs in the 129 strain (Fig. 1*B*). For the most part, these elements were highly methylated, akin to the vast majority of the ~10,000 IAPs in the mouse genome. The susceptibility of VM-IAPs to genetic background effects provides an opportunity to map the genetic determinants of inter-individual methylation variability.

Due to the repetitive nature of IAPs, it is difficult to rule out the possibility that the differences in methylation between B6 and 129 are a result of sequence divergence within the elements themselves rather than a consequence of *trans*-acting modifiers. For example, a LINE element is embedded in IAP-*Rab6b* in the 129 genome that is absent in the B6 genome (***SI Appendix***, Fig. S3*A*). To avoid this confounder, we implemented a hybrid breeding scheme using B6 and CAST mice (Fig. 2*A*). Because B6 VM-IAPs are largely absent from the CAST genome, F1 hybrid offspring inherit a single allele from their B6 parent. This property allowed us to assess whether a haploid CAST genome is capable of inducing methylation changes at B6-specific VM-IAPs in *trans*. Maternal and paternal transmission of these alleles was followed by crossing B6 females to CAST males (BC) and CAST females to B6 males (CB), respectively. The reciprocal design enabled the exploration of parent-of-origin effects in addition to genetic background effects for the 12 B6-specific VM-IAPs examined in this experiment. Furthermore, we used large sample sizes to guarantee the detection of subtle shifts in the distribution of methylation levels at each locus, which additionally revealed that the frequency distributions of VM-IAP methylation levels in the

pure B6 population form skewed bell curves rather than normal distributions (*SI Appendix*, Fig. S1).

Two thirds of the assessed B6-specific VM-IAPs showed significant differences between BC and CB methylation distributions (Fig. 2*B* and *SI Appendix*, Fig. S2). These effects were not reciprocal, indicating they were not imprinting effects. For instance, at half of the loci, CB hybrids showed hypermethylation of the paternally inherited B6 VM-IAP while the BC hybrids exhibited levels comparable to pure B6. This suggests the presence of a CAST-specific maternally inherited modifier acting on the paternally inherited B6 allele (Fig. 2*B* and *SI Appendix*, Fig. S2). Paternal transmission of the CAST modifier had no effect on the maternally inherited B6 VM-IAP, consistent with a maternal effect. Additional experiments are required to better understand these maternal genetic background effects, but strain-specific factors derived from the oocyte are likely involved.

In addition to genetic background-specific maternal effects, a subset of VM-IAPs exhibited zygotic genetic background effects, defined as changes in VM-IAP methylation caused by the introduction of a CAST haploid genome regardless of parental origin. Four VM-IAPs displayed significant shifts in methylation when BC and CB were compared to the B6 population (Fig. 2*B* and *SI Appendix*, Fig. S2). IAP-*Marveld2* was alone in showing a reduction in methylation in F1 hybrids compared to B6 individuals. The other three (IAP-*Rab6b*, IAP-*Sema6d*, and IAP-*Fam78b*) were hypermethylated in F1 hybrids compared to B6 mice, suggesting that CAST-encoded modifier(s) may be targeting these loci for repression. We note that IAP-*Rab6b* and IAP-*Sema6d* displayed both maternal and zygotic genetic background effects, while IAP-*Gm13849* and IAP-*Slc15a2* displayed neither. The range of responses indicates that the mechanisms influencing variable methylation at IAPs are not common across all loci.

## Strain-specific IAP-*Rab6b* hypermethylation is driven by a single modifier locus

A successful genetic mapping experiment relies on an unambiguous phenotype. Unlike most of the VM-IAPs examined in hybrids, IAP-*Rab6b* (a solo LTR) exhibited non-overlapping B6 and F1 hybrid methylation distributions that were clearly distinguishable using a 60% methylation threshold (Fig. 2*B*). Due to the categorical nature of this 'methylation phenotype', IAP-*Rab6b* was selected to identify VM-IAP modifiers using B6/CAST hybrids.

We first investigated whether the low methylation state (<60%) could be rescued in a subsequent generation by backcrossing F1 hybrids to B6 mice. Low methylation was reacquired in approximately half of N1 backcrossed offspring, irrespective of parental origin (Fig. 3*A* and *SI Appendix*, Fig. S3*B*). This is indicative of limited redundancy in IAP-*Rab6b*-targeting modifiers in the CAST genome. The segregation of methylation states was not attributable to IAP-*Rab6b* copy number, as hemi- and homozygous individuals were represented in both the highly and lowly methylated groups (*SI Appendix*, Fig. S3*C*).

We conducted an additional round of B6 backcrossing to further characterise the inheritance pattern of IAP-*Rab6b* methylation states. N2 offspring generated from highly methylated N1 males recreated the 1:1 ratio of high-to-low methylation observed in N1 offspring, while N2 offspring generated from lowly methylated N1 males were all lowly methylated (Fig. 3*A* and *B*). This Mendelian inheritance pattern indicates that a single dominant CAST-derived locus causes the hypermethylation of B6-derived IAP-*Rab6b* in *trans*. This was confirmed by crossing highly methylated N2 males to B6 females, once again producing roughly equal numbers of highly and lowly methylated N3 offspring (Fig. 3*A* and *B*).

We next mapped the modifier locus using the Giga Mouse Universal Genotyping Array (GigaMUGA), a 141,090 single nucleotide polymorphism (SNP) microarray designed to capture

the genetic diversity found across mouse strains (21). Due to the evolutionary distance between B6 and CAST, a majority of the probed SNPs are informative between the two strains. DNA samples from 47 N3 individuals were analysed on the array. SNP calls were filtered to identify heterozygous SNPs shared by all 23 highly methylated individuals and absent from all 24 lowly methylated individuals. The resulting SNPs all mapped to a 7.3 Mb interval on distal Chromosome 4 (Fig. 3*C*; *SI Appendix*, Fig. S4; Dataset S1: Table S1). We separately analysed 22 N2 individuals using the lower-resolution MiniMUGA array and independently identified the same genomic region (Dataset S1: Table S2). Combining both mapping experiments delimited a 6.4 Mb window on Chromosome 4 containing the IAP-*Rab6b* modifier(s) (GRC38/mm10, chr4:141,964,197-148,393,136). Of note, this locus was found to exhibit high heterozygous SNP density in a study comparing the genomes of sixteen different laboratory mouse strains (20).

Assessment of the genes within the mapped interval revealed a cluster of KZFPs (Fig. 3*D*). Present in the hundreds, KZFPs make up the largest and most diverse transcription factor family in higher vertebrate genomes (22, 23). They are best known for their role in sequence-dependent transposable element repression. Their $C_2H_2$ zinc finger arrays recognize and bind DNA motifs with high specificity and their KRAB domain recruits the scaffold protein KAP1, which in turn induces heterochromatin (24). The rapid evolutionary expansion of murine KZFPs sets them apart from other more conserved epigenetic regulators. We therefore hypothesised that this KZFP cluster, designated Chr4-cl, contains the strain-specific modifier(s) of IAP-*Rab6b*. The B6, CAST, and 129 variants of this cluster are henceforth referred to as Chr4-cl$^{B6}$, Chr4-cl$^{CAST}$, and Chr4-cl$^{129}$, respectively.

## The CAST-derived modifier locus targets multiple VM-IAPs in a sequence-specific manner

IAP sequences are highly repetitive in the mouse genome due to the evolutionary youth and retrotransposition potential of IAP elements (10). In view of the sequence-specificity of KZFP-induced epigenetic repression, we reasoned that VM-IAPs with sequence similarity to IAP-*Rab6b* may also be targeted by the same modifier locus. Six solo LTR VM-IAPs with more than 90% sequence identity to IAP-*Rab6b* were selected as potential targets along with IAP-*Sema6d* and IAP-*Fam78b*, which had exhibited hypermethylation in F1 hybrids (Fig. 2*B*). Methylation was quantified in N2 individuals, half of which were highly methylated at IAP-*Rab6b* (i.e. heterozygous carriers of the CAST modifier locus) and half of which were lowly methylated at IAP-*Rab6b* (i.e. non-carriers of the CAST modifier locus). We found that individuals that were highly methylated at IAP-*Rab6b* were also highly methylated at six out of the eight assessed loci – IAP-*Tmprss11d*, IAP-*Pink1*, IAP-*Rps12*, IAP-*Trbv31*, IAP-*Ect2l*, and IAP-*Sema6d* – and vice versa for the lowly methylated individuals (Fig. 4*A*). This result suggests that these regions are additional modifier targets. In contrast, IAP-*Gm20110* and IAP-*Fam78b* methylation levels were not concordant with IAP-*Rab6b* methylation levels (Fig. 4*A*). A sequence alignment of the nine solo LTR VM-IAPs revealed a single region that distinguished IAP-*Gm20110* and IAP-*Fam78b* from the other six IAPs (Fig. 4*B* and *SI Appendix*, Fig. S5). The 28 bp DNA segment, containing an insertion and various SNPs in IAP-*Gm20110* and IAP-*Fam78b*, is a likely binding site for the CAST-specific modifier.

The cross-locus comparison highlights the sequence dependence of the modifiers of these epialleles and provides support for a KZFP-mediated mechanism. Interestingly, our earlier observations in pure 129 mice showed variable methylation at IAP-*Gm20110* and IAP-*Fam78b*,

and hypermethylation at the other six IAPs (Fig. 1*B*), suggesting that Chr4-cl$^{129}$ shares VM-IAP modifier allele(s) with Chr4-cl$^{CAST}$ that are absent from Chr4-cl$^{B6}$.

**The KZFP cluster on Chromosome 4 modifies VM-IAP methylation states**

The unique clustered organisation of KZFPs in the mouse genome stems from segmental duplications, resulting in high sequence similarity among adjacent KZFPs and low-quality cluster reference sequences (22). To circumvent the technical difficulties and potential functional redundancy associated with generating single-KZFP knockouts (KOs), we examined the consequences of deleting the entire Chr4-cl using a previously generated Chr4-cl KO mouse line (25).

We first assessed DNA methylation effects caused by the loss of Chr4-cl in a pure B6 genetic background. Compared to wild type (WT) mice, which exhibited the expected inter-individual methylation variability at all loci, Chr4-cl KO mice showed significantly lower methylation levels at IAP-*Rab6b*, IAP-*Pink1*, IAP-*Ect2l*, and IAP-*Rps12* (Fig. 4*C*). The effect was particularly pronounced at IAP-*Rab6b*, where all KO mice were completely unmethylated. This result shows that Chr4-cl$^{B6}$ is necessary for the acquisition of variable methylation at IAP-*Rab6b* and reveals an important mechanistic role for KZFPs in the stochastic methylation of retrotransposons. Of note, the other VM-IAPs targeted by the CAST-specific modifier did not show a reduction in methylation in the absence of Chr4-cl$^{B6}$. Given the extensive redundancy displayed KZFPs in the mouse genome (25), it is possible that the variable methylation observed at these regions in B6 mice is conferred by KZFP(s) located in other cluster(s).

We next asked whether Chr4-cl can mediate the strain-specific hypermethylation of VM-IAPs using the 129 Chr4-cl locus. Homozygous B6 Chr4-cl KO mice were crossed to WT 129 mice, which harbour Chr4-cl$^{129}$ as well as most VM-IAPs of interest (Fig. 1*A*, Fig. 4*D*). F1 mice were backcrossed to 129 followed by two rounds of heterozygous intercrosses (Fig. 4*D*). VM-IAP methylation was assessed in the resulting Chr4-cl KO and WT mice of mixed B6/129 genetic background. In this instance, all of the predicted Chr4-cl targets from our cross-locus comparison in Fig. 4*A* exhibited significantly lower methylation levels in Chr4-cl KO mice compared to their WT counterparts, often reflecting a return to the variable levels observed at these regions in pure B6 mice (Fig. 4*E*). WT methylation levels were largely consistent with the pure 129 data from Fig. 1*B* despite the use of different 129 sub-strains. These results indicate that Chr4-cl is the functionally relevant segment of the 6.4 Mb interval identified in our mapping experiment and demonstrate that KZFPs are strain-specific VM-IAP modifiers.

**Loss of Chr4-cl alters the chromatin and transcriptional landscape near targeted VM-IAPs**

The recruitment of KAP1 and subsequent H3K9 trimethylation by the methyltransferase SETDB1 are characteristic of epigenetic silencing by KZFPs. To gain insight into the mechanism by which strain-specific Chr4-cl KZFP(s) target VM-IAPs, we analysed previously generated ChIP-seq datasets that profiled histone modifications in Chr4-cl WT and KO embryonic stem (ES) cells of mixed B6/129 background (25). Visual inspection of H3K9me3 ChIP-seq tracks at Chr4-cl-targeted VM-IAPs revealed a modest decrease in H3K9me3 enrichment upon loss of Chr4-cl (Fig.5*A* and *SI Appendix*, Fig. S6*A* and *B*). More striking, however, was a marked increase in H3K4me3 in KO cells at Chr4-cl targets, with levels equivalent to those observed at neighbouring gene promoters (Fig. 5*A*, *C* and *SI Appendix*, Fig. S6*A*). No increase in H3K4me3 was observed

at non-targets IAP-*Gm20110* and IAP-*Fam78b* (Fig. 5*B*). Therefore, the acquisition of H3K4me3 represents the default chromatin state at these loci, which is partially impeded either directly or indirectly by Chr4-cl KZFPs in early development.

The H3K4me3 mark is associated with transcriptional activity and localises to gene promoters, with greatest enrichment in the region immediately downstream of the transcription start site (TSS) (26). In line with this, the increase in H3K4me3 in Chr4-cl KO cells at targeted VM-IAPs was exclusively found at their 3′ end, downstream of the TSS embedded in the solo LTRs (Fig. 4*B*, Fig. 5*A*, *C* and *SI Appendix*, Fig. S6*A*).

Next, we explored whether the remodelled chromatin landscape at VM-IAPs in Chr4-cl KO ES cells was associated with altered expression of neighbouring genes. RNA-seq datasets generated from the same Chr4-cl WT and KO ES cells revealed differences in gene expression near IAP-*Pink1* and IAP-*Rab6b*. *Pink1* and *Rab6b*, located 1 kb upstream and 3 kb downstream of IAP-*Pink1* and IAP-*Rab6b*, respectively, were upregulated in Chr4-cl KO ES cells (but only *Pink1* reached statistical significance) (Fig. 5*D* and *SI Appendix*, Fig. S6*C*). *Slco2a1*, located 20 kb upstream of IAP-*Rab6b*, was significantly downregulated in Chr4-cl KO ES cells. These data indicate that the loss of Chr4-cl results in a range of transcriptional disruptions. While transcriptional changes were not observed near the other Chr4-cl-targeted VM-IAPs, our analysis does not rule out longer range transcriptional effects.

**Methylation variability at IAPLTR2_Mm elements is sequence dependent**

The seven VM-IAPs that we identified as Chr4-cl targets are all solo LTRs of the IAPLTR2_Mm subclass. To determine how VM-IAPs compare to other solo LTRs of this subclass from an

evolutionary perspective, we built a neighbour-joining tree of all solo IAPLTR2_Mm elements in the B6 genome. Consistent with a KZFP-mediated mechanism, we found that VM-IAPs of this subclass cluster together phylogenetically (Fig. 6*A*). This is in agreement with our previous analysis on IAPs of the IAPLTR1_Mm subclass (17) and reinforces the concept that genetic sequence is instructive in the establishment of inter-individual methylation variability. We selected five epigenetically uncharacterised IAPs in the VM-IAP-enriched subtree to test whether members of this clade are in fact unidentified VM-IAPs. All five candidates failed to display methylation variability, highlighting that other determinants such as genomic context likely also play a role in the acquisition of methylation variability at IAPs (Fig. 6*B*). Given that murine IAPs are almost invariably highly methylated, it follows that the VM-IAP-enriched subtree has escaped epigenetic repression, at least partially. Notably, this clade was enriched in H3K4me3 in WT ES cells and showed a greater increase in H3K4me3 in Chr4-cl KO ES cells compared to other IAPLTR2_Mm solo LTRs (Fig. 7*A*).

**Chr4-cl KZFPs target IAPLTR2_Mm elements**

To determine whether Chr4-cl KZFPs are capable of recognising IAPLTR2_Mm elements, the IAPLTR2_Mm consensus sequence was queried for binding motifs previously assigned to 16 Chr4-cl KZFPs (25). Two of the query hits, ZFP989 and Gm21082, exhibited ChIP-seq enrichment at IAPLTR2_Mm solo LTRs (Fig. 7*B*). Both ZFP989 and Gm21082 appear to bind the same region of the LTR and are thus likely the product of a gene duplication event within Chr4-cl. Gm21082 was previously reported to target IAPLTR2_Mm elements along with one other KZFP, ZFP429, which is located in a KZFP cluster on Chromosome 13 (Chr13-cl) (25). Interestingly, ZFP989 and Gm21082 showed reduced enrichment at VM-IAPs compared to other IAPLTR2_Mm solo LTRs,

whereas ZFP429 exhibited strong preferential binding at IAPs in the VM-IAP-enriched subtree (Fig. 7*B*).

It is feasible that the CAST and 129 alleles of ZFP989 or Gm21082 are responsible for the strain-specific hypermethylation of VM-IAPs, while ZFP429 may be involved in the establishment of inter-individual methylation variability in B6 mice. This would explain why DNA methylation at some of the VM-IAPs targeted by Chr4-cl$^{129}$ were unaffected by the loss of Chr4-cl in a pure B6 background (Fig. 4*C*). We note that while we have shown that multiple KZFPs are capable of binding LTRs of the IAPLTR2_Mm subtype, none of the candidate Chr4-cl KZFPs appear to recognise the predicted binding site identified in Fig. 4*E*. This is unsurprising considering that the B6 allele(s) of the Chr4-cl KZFP modifier(s) are not expected to strongly bind VM-IAPs, and the ChIP-seq datasets used for this analysis were generated through stable expression of epitope-tagged B6 KZFPs.

## Discussion

Variable methylation of murine IAPs across genetically identical individuals was reported more than two decades ago (27), yet the underlying mechanisms and evolutionary origins of this phenomenon have remained elusive. In this study, we identify widespread genetic background-specific modification of VM-IAPs and exploit these to investigate the genetic determinants of mammalian epigenetic stochasticity. We demonstrate that a polymorphic KZFP cluster on Chromosome 4 promotes the sequence- and strain-specific hypermethylation of multiple VM-IAPs in *trans*, the loss of which alters the chromatin and transcriptional landscape of the modified loci and their surrounding genetic environment. We expect our classical genetics approach using inbred mouse strains to be generalisable to other variably methylated regions in the mouse genome.

The identification of Chr4-cl KZFP(s) as strain-specific VM-IAP modifier(s) is consistent with the literature documenting KZFPs as products of rapidly evolving genes with critical functions in transposable element repression (reviewed in Ecco et al., 2017). The large number of species-specific KZFPs in the mouse compared to most other higher vertebrates suggests that murine KZFPs have undergone particularly rapid amplification (23). It is thought that this expansion reflects an active evolutionary arms race following ERV invasion events (22, 28). Our finding that Chr4-cl contains strain-specific modifier(s) of certain IAP elements indicates that murine KZFPs are evolving rapidly enough to detect significant divergence *within* the mouse species with important epigenetic and transcriptional ramifications. IAPs, which are murine-specific and represent the most mutagenic ERV class in the mouse (10), have likely played a major role in this process. Thus, comparative research across mouse strains is uniquely suited to the study of KZFP gene evolution.

A significant technical hindrance in taking full advantage of cross-strain mouse genetics in this context relates to the extensive redundancy exhibited by KZFPs both within and across clusters (22, 23, 25). The current mouse strain reference genomes have large gaps in KZFP clusters, rendering a cross-strain comparison of Chr4-cl sequences currently unfeasible (20). In fact, while we expect significant KZFP sequence differences in Chr4-cl across mouse strains, we acknowledge that we have not excluded the possibility that the strain-specific effects we have identified are driven by differences in KZFP gene regulation rather than allelic variation of KZFPs themselves. Indeed, a recent study showed that the NOD and B6 mouse strains exhibit differential T cell gene expression and 3D chromatin organisation in Chr4-cl, with implications for diabetes phenotypes (29). The improvement of genetic engineering tools for repetitive gene families and

the generation of high-quality mouse strain reference genomes with full coverage over KZFP clusters will be crucial in addressing this issue.

We have shown that in addition to mediating the strain-specific hypermethylation of VM-IAPs, KZFPs play an important role in the establishment of inter-individual methylation variability in a pure B6 background, as evidenced by the complete loss of DNA methylation at IAP-*Rab6b* in B6 Chr4-cl KO mice. The specificity of KZFP binding relies on four amino acids within each zinc finger, so mutations in these key residues or in the DNA sequence of their binding sites have important implications for target site binding kinetics (30). We propose that stochastic methylation arises when VM-IAP sequences are weakly recognised by KZFPs during early preimplantation development. Consistent with this model, a large number of murine KZFPs are highly expressed in ES cells, including most of the KZFPs in Chr4-cl (22, 23, 25, 31). Furthermore, we previously documented a lack of methylation co-variation across VM-IAPs within an individual mouse (17), which is in agreement with low-affinity binding interactions occurring independently between a KZFP and multiple VM-IAP targets. The phylogenetic clustering of VM-IAPs provides further support for this mechanism given that KZFP function is driven by DNA sequence recognition. It is noteworthy that ZFP429 preferentially binds solo LTRs in the VM-IAP-enriched subtree compared to other IAPLTR2_Mm solo LTRs, perhaps reflecting a host adaptive response to elements that have escaped epigenetic repression. Alternatively, it is possible that inter-individual methylation variability is an early sign of transposable element domestication.

While this framework predicts that the sequence of an IAP and that of its KZFP modifier(s) are the prime drivers of inter-individual methylation variability, additional factors are expected to influence the probability of a binding event occurring. This is illustrated by the presence of highly methylated IAPLTR2_Mm elements within the VM-IAP-enriched subtree. Potential influencing

factors include chromatin accessibility of VM-IAP insertion sites and the number and expression level of KZFPs targeting a particular locus. Importantly, we envisage that other IAP-binding proteins interfere with binding kinetics between KZFPs and VM-IAPs. In fact, VM-IAPs are enriched for CCCTC-binding factor (CTCF), a methylation-sensitive DNA binding protein that may act as an antagonist to methylation-promoting KZFP modifiers (17, 18).

The most prominent difference in chromatin structure that we observed between Chr4-cl WT and KO ES cells was a substantial increase in H3K4me3 in Chr4-cl KO cells at the 3′ end of targeted VM-IAPs. This suggests that KZFP binding in early development prevents the accumulation of H3K4me3 at VM-IAP TSSs, potentially enabling subsequent DNA methylation. In line with this, the ADD domain of *de novo* DNA methyltransferases DNMT3A and DNMT3B (and of their cofactor DNMT3L) specifically binds unmethylated H3K4 (32, 33). In the absence of modifying KZFPs, other transcription factors and histone modifying enzymes have increased access to VM-IAP sequences, which may in turn contribute to the dysregulation of VM-IAP-neighbouring genes. Interestingly, a recent study using the BXD recombinant inbred mouse panel identified six major *trans*-acting dominant suppressors of H3K4me3 in male germ cells, all of which were mapped to KZFP clusters (34).

Our work is consistent with previous research on strain-specific modifiers (35). A growing number of ERV-derived mouse mutations whose phenotypic penetrance is dependent on genetic background have been associated with modifier genes located in an interval on Chromosome 13 containing a KZFP cluster (Chr13-cl). This region harbours *Mdac,* modifier of the dactylaplasia-causing $Dac^{1J}$ insertion, as well as modifiers of IAPs shown to mediate a range of strain-specific phenotypes (8, 36–38). Moreover, two polymorphic KZFPs in Chr13-cl, SNERV1 and SNERV2, were recently found to influence ERV expression in lupus-susceptible mouse strains (39).

Interestingly, the Chr4-cl KZFP *Zfp979* is responsible for the strain-specific methylation of the HRD transgene (40), indicating that polymorphic Chr4-cl KZFPs target a variety of foreign DNA sequences. Notably, the genes whose expression we found to be altered in Chr4-cl KO mice near Chr4-cl-targeted VM-IAPs have been implicated in human disease. Mutations in *PINK1*, which codes for a kinase involved in mitochondrial quality control, are associated with Parkinson's disease (41); mutations in *SLCO2A1*, a prostaglandin transporter, cause chronic enteropathy (42). It is possible that Chr4-cl KZFP diversification influences related susceptibilities in mice. Taken together, a picture emerges of KZFPs as fine-tuners of evolution within the mouse species, whereby KZFP divergence across strains leads to strain-specific epigenetic landscapes with important phenotypic implications. Moreover, it is possible that KZFP divergence across human populations contributes to variable phenotypic outcomes via the differential recognition of endogenous (and perhaps even exogenous) retroviruses.

We have shown that most VM-IAPs are susceptible to maternal effects and posit that these too are driven by strain-specific KZFPs. A number of KZFPs have been characterised as maternal-effect genes expressed in the oocyte (43–45). Maternally-derived ZFP57, for example, is required for the maintenance of DNA methylation at imprinted regions during global post-fertilisation methylation erasure (43). It is intriguing that, on a B6 background, $A^{vy}$ and IAP-*Gm13849* only exhibit epigenetic inheritance upon maternal transmission (12, 17). We speculate that certain paradigms regarded as mammalian transgenerational epigenetic inheritance may in actual fact be the product of post-fertilisation re-targeting of epigenetic states by germline-derived KZFPs.

## Materials and Methods

All mouse work was conducted in compliance 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). DNA methylation was quantified using bisulphite pyrosequencing and genetic mapping was carried using the GigaMUGA SNP genotyping array (21). Data availability and details of mouse experiments, molecular techniques, and computational analyses performed in this study are described in *SI Appendix*, Materials and Methods.

## Supporting information

SI Appendix: Materials and Methods; Supplemental Figures S1-S5.

Dataset S1: Supplemental Tables S1-S6.

Dataset S2: Unprocessed MiniMUGA SNP calls.

Dataset S3: Unprocessed GigaMUGA SNP calls.

## Acknowledgements

This study was supported by the Wellcome Trust, United Kingdom (210757/Z/18/Z to A.C.F.-S.), the BBSRC, United Kingdom (BB/R009996/1 to A.C.F.-S.), the Medical Research Council, United Kingdom (MR/R009791/1 to A.C.F.-S.), the National Institutes of Health, U.S. (1ZIAHD008933 to T.S.M.), and by PhD scholarships from the Cambridge Trust and the BBSRC to T.M.B. and J.L.E., respectively. We thank Michael Imbeault for valuable advice on our work

and Noah Kessler, Amir Hay, Jessi Becker, and other members of the Ferguson-Smith lab for useful discussions. We are grateful to Jesse Loilargosain for technical assistance, Ali Robinson for mouse husbandry support, Gernot Wolf for communications on NGS datasets, Melania Bruno and Sherry Ralls for sample preparation and shipment, and Anastasiya Kazachenka for preliminary data on VM-IAP methylation levels in 129 mice.

## References

1. J. R. Shorter, *et al.*, A diallel of the mouse collaborative cross founders reveals strong strain-specific maternal effects on litter size. *G3 Genes, Genomes, Genet.* **9**, 1613–1622 (2019).
2. F. Odet, *et al.*, The founder strains of the collaborative cross express a complex combination of advantageous and deleterious traits for male reproduction. *G3 Genes, Genomes, Genet.* **5**, 2671–2683 (2015).
3. C. Sapienza, J. Paquette, T. H. Tran, A. Peterson, Epigenetic and genetic factors affect transgene methylation imprinting. *Development* **107**, 165–168 (1989).
4. N. D. Allen, M. L. Norris, M. A. Surani, Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. *Cell* **61**, 853–861 (1990).
5. P. Engler, *et al.*, A strain-specific modifier on mouse chromosome 4 controls the methylation of independent transgene loci. *Cell* **65**, 939–947 (1991).
6. A. Schumacher, P. A. Koetsier, J. Hertz, W. Doerfler, Epigenetic and genotype-specific effects on the stability of de novo imposed methylation patterns in transgenic mice. *J. Biol. Chem.* **275**, 37915–37921 (2000).
7. K. R. Johnson, P. W. Lane, P. Ward-Bailey, M. T. Davisson, Mapping the mouse dactylaplasia mutation, Dac, and a gene that controls its expression, mdac. *Genomics* **29**, 457–464 (1995).
8. H. Kano, H. Kurahashi, T. Toda, Genetically regulated epigenetic transcriptional activation of retrotransposon insertion confers mouse dactylaplasia phenotype. *Proc. Natl. Acad. Sci.* **104**, 19034–19039 (2007).
9. D. M. J. Duhl, H. Vrieling, K. A. Miller, G. L. Wolff, G. S. Barsh, Neomorphic agouti mutations in obese yellow mice. *Nat. Genet.* **8**, 59–65 (1994).
10. L. Gagnier, V. P. Belancio, D. L. Mager, Mouse germ line mutations due to retrotransposon insertions. *Mob. DNA* **10**, 15 (2019).
11. G. L. Wolff, R. L. Kodell, S. R. Moore, C. A. Cooney, Maternal epigenetics and methyl

supplements affect agouti gene expression in Avy/a mice. *FASEB J.* **12**, 949–957 (1998).

12. H. D. Morgan, H. G. Sutherland, D. I. Martin, E. Whitelaw, Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* **23**, 314–318 (1999).

13. T. M. Bertozzi, A. C. Ferguson-Smith, Metastable epialleles and their contribution to epigenetic inheritance in mammals. *Semin. Cell Dev. Biol.* **97**, 93–105 (2020).

14. G. L. Wolff, Genetic modification of homeostatic regulation in the mouse. *Am. Nat.* **105**, 241–252 (1971).

15. G. L. Wolff, Influence of maternal phenotype on metabolic differentiation of agouti locus mutants in the mouse. *Genetics* **88**, 529–39 (1978).

16. V. K. Rakyan, *et al.*, 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 (2003).

17. A. Kazachenka, *et al.*, Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell* **175**, 1259–1271 (2018).

18. J. L. Elmer, *et al.*, Genomic properties of variably methylated retrotransposons in mouse. *bioRxiv*, 2020.10.21.349217 (2020).

19. C. Nellåker, *et al.*, The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol.* **13**, R45 (2012).

20. J. Lilue, *et al.*, Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci. *Nat. Genet.* **50**, 1574–1583 (2018).

21. A. P. Morgan, *et al.*, The mouse universal genotyping array: From substrains to subspecies. *G3 Genes, Genomes, Genet.* **6**, 263–279 (2016).

22. A. Kauzlaric, *et al.*, The mouse genome displays highly dynamic populations of KRAB-zinc finger protein genes and related genetic units. *PLoS One* **12**, e0173746 (2017).

23. M. Imbeault, P. Y. Helleboid, D. Trono, KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**, 550–554 (2017).

24. G. Ecco, M. Imbeault, D. Trono, KRAB zinc finger proteins. *Development* **144**, 2719–2729 (2017).

25. G. Wolf, *et al.*, Krab-zinc finger protein gene expansion in response to active retrotransposons in the murine lineage. *Elife* **9**, 1–22 (2020).

26. N. D. Heintzman, *et al.*, Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).

27. E. J. Michaud, *et al.*, Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. *Genes Dev.* **8**, 1463–1472 (1994).

28. M. Bruno, M. Mahgoub, T. S. Macfarlan, The Arms Race Between KRAB–Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. *Annu. Rev. Genet.* **53**, 393–416 (2019).

29. M. Fasolino, *et al.*, Genetic Variation in Type 1 Diabetes Reconfigures the 3D Chromatin Organization of T Cells and Alters Gene Expression. *Immunity* **52**, 257–274 (2020).

30. M. Elrod-Erickson, T. E. Benson, C. O. Pabo, High-resolution structures of variant Zif268–DNA complexes: implications for understanding zinc finger–DNA recognition. *Structure* **6**, 451–464 (1998).

31. A. Corsinotti, *et al.*, Global and Stage Specific Patterns of Krüppel-Associated-Box Zinc Finger Protein Gene Expression in Murine Early Embryonic Cells. *PLoS One* **8**, e56721 (2013).

32. S. K. T. Ooi, *et al.*, DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).

33. M. Morselli, *et al.*, In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *Elife* **4**, e06205 (2015).

34. C. L. Baker, *et al.*, Tissue-specific trans regulation of the mouse epigenome. *Genetics* **211**, 831–845 (2019).

35. J. L. Elmer, A. C. Ferguson-Smith, Strain-Specific Epigenetic Regulation of Endogenous Retroviruses: The Role of Trans-Acting Modifiers. *Viruses* **12**, 1–21 (2020).

36. J. A. Plamondon, M. J. Harris, D. L. Mager, L. Gagnier, D. M. Juriloff, The clf2 gene has an epigenetic role in the multifactorial etiology of cleft lip and palate in the A/WySn mouse strain. *Birth Defects Res. Part A - Clin. Mol. Teratol.* **91**, 716–727 (2011).

37. C. J. Krebs, *et al.*, Regulator of sex-limitation (Rs1) encodes a pair of KRAB zinc-finger genes that control sexually dimorphic liver gene expression. *Genes Dev.* **17**, 2664–2674 (2003).

38. N. Maeda-Smithies, *et al.*, Ectopic expression of the Stabilin2 gene triggered by an intracisternal A particle (IAP) element in DBA/2J strain of mice. *Mamm. Genome* **31**, 2–16 (2020).

39. R. S. Treger, *et al.*, The Lupus Susceptibility Locus Sgp3 Encodes the Suppressor of Endogenous Retrovirus Expression SNERV. *Immunity* **50**, 334–347 (2019).

40. S. Ratnam, *et al.*, Identification of Ssm1b, a novel modifier of DNA methylation, and its expression during mouse embryogenesis. *Development* **141**, 2024–2034 (2014).

41. E. M. Valente, *et al.*, Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science* **304**, 1158–1160 (2004).

42. J. Umeno, *et al.*, A Hereditary Enteropathy Caused by Mutations in the SLCO2A1 Gene, Encoding a Prostaglandin Transporter. *PLOS Genet.* **11**, e1005581 (2015).

43. X. Li, *et al.*, A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. *Dev. Cell* **15**, 547–557 (2008).

44. S. B. V Ramos, *et al.*, The CCCH tandem zinc-finger protein Zfp36l2 is crucial for female fertility and early embryonic development. *Development* **131**, 4883–93 (2004).

45. M. K. Y. Seah, *et al.*, The KRAB-zinc-finger protein ZFP708 mediates epigenetic repression at RMER19B retrotransposons. *Development* **146**, dev.170266 (2019).

46. W. J. Kent, BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).

47. P. Stempor, J. Ahringer, SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics. *Wellcome Open Res.* **1**, 14 (2016).

## Figure legends

**Fig. 1.** Inter-individual methylation variability at IAPs is strain-specific. (*A*) B6 VM-IAPs are polymorphic across inbred mouse strains. All experimentally validated B6 VM-IAPs (18) were scored for presence (navy blue rectangles) or absence (light blue rectangles) in the 129S1/SvlmJ and CAST/EiJ reference genomes. Instances in which a classification could not be made with confidence due to gaps in the reference sequences are shown in white. VM-IAPs are colour-coded according to their structure (full-length IAPs, blue; truncated IAPs, orange; solo LTRs, pink). LTR subclass annotation, as defined by RepeatMasker, is indicated above each VM-IAP. VM-IAPs are named based on their closest coding gene. (*B*) DNA methylation levels in B6 (grey) and 129S2/Sv (purple) inbred mice of IAPs shared between the two strains. Some IAPs exhibit variable methylation (>10 % variance across individuals) in both strains (left of dotted line); others are only variably methylated in B6 mice (right of dotted line). Methylation levels of the distal-most CpGs of the IAP 5′ LTRs were quantified from genomic DNA using bisulphite pyrosequencing. Each dot represents the average methylation level across CpGs for one individual.

**Fig. 2.** VM-IAP methylation levels are subject to maternal and zygotic genetic background effects (GBEs). (*A*) Reciprocal hybrid breeding scheme. BC F1 hybrids (green diamonds) were generated by breeding B6 (black) females with CAST (brown) males. CB F1 hybrids (yellow diamonds) were produced from the reciprocal cross of CAST females and B6 males. (*B*) VM-IAPs classified based on their susceptibility to maternal GBEs (upper left), maternal zygotic GBEs (upper right), zygotic GBEs (lower left), or neither (lower right). Violin plots represent the B6 (grey), BC (green), and CB (yellow) F1 offspring methylation distributions. Dotted and dashed lines show the distribution quartiles and median, respectively. Faint hollow circles represent individual-specific

methylation levels, quantified from genomic DNA and averaged across the distal CpGs of the VM-IAP 5′ LTR. B6, BC, and CB methylation levels were compared for each VM-IAP using the Kruskal-Wallis test followed by Dunn's post hoc multiple comparison test (**** $p<0.0001$; ns: not significant). Sample sizes are shown below each graph. Graphs for the 8 additional VM-IAPs analysed in this experiment can be found in *SI Appendix,* Fig. S2.

**Fig. 3.** Strain-specific IAP-*Rab6b* hypermethylation is driven by a single dominant modifier locus on Chromosome 4. (*A*) Genetic backcrossing uncovers Mendelian inheritance pattern of IAP-*Rab6b* methylation states. F1 BC males were backcrossed to B6 females to produce the first backcrossed generation (N1). Three highly methylated (red) and three lowly methylated (grey) N1 males were backcrossed to B6 females to produce the N2 generation, and highly methylated N2 males were once again backcrossed to B6 females to produce the N3 generation. Average percent CAST DNA remaining in the genome at each generation is indicated under the graph. 60% methylation was used as the cut-off value to classify individuals as highly (red) or lowly (grey) methylated. (*B*) Pedigree illustrating the inheritance patterns of IAP-*Rab6b* methylation states. Percentages reflect the data in panel A. (*C*) Genetic mapping of the modifier locus to a 7.3 Mb interval on distal Chromosome 4 using the GigaMUGA SNP microarray. A map is shown of heterozygous SNPs along the chromosome that are informative between B6 and CAST in 20 N3 individuals (full set of individuals shown in *SI Appendix*, Fig. S4). Heterozygous SNPs shared amongst all highly methylated N3 individuals and absent from all lowly methylated N3 individuals are shown in blue. The corresponding mapped region is highlighted in yellow. (*D*) Expanded view of the 2.5 Mb KZFP cluster located within the mapped interval. Sequence gaps in the current reference genome (GRCm38/mm10) are displayed as black boxes above the annotated genes.

673 Stripped region represents the portion of DNA excluded by our independent analysis of N2
674 individuals using the MiniMUGA SNP microarray. KZFP genes are bolded. Annotations were
675 lifted from the UCSC Gencode V24 track.

676

677 **Fig. 4.** The KZFP cluster Chr4-cl modulates the methylation state of multiple VM-IAPs. (*A*) Cross-
678 locus comparison of N2 methylation states. Methylation levels were quantified at IAP-*Tmprss11d*,
679 IAP-*Pink1*, IAP-*Ect2l*, IAP-*Rps12*, IAP-*Trbv31*, IAP-*Sema6d*, IAP-*Gm20110*, and IAP-*Fam78b*
680 in 20 N2 individuals. IAP-*Rab6b* methylation level had previously been determined to be high
681 (>60 %, red) or low (< 60%, grey). Red dashed and grey dotted lines connect the average
682 methylation values of N2 individuals across regions. Percent sequence identity to IAP-*Rab6b*, as
683 determined by the BLAT alignment tool (46), is shown for each locus above the x-axis. (*B*)
684 Alignment of VM-IAP LTR sequences in the region displaying divergence between Chr4-cl
685 targets and non-targets. Contraoriented elements were reverse-complemented (rc) prior to
686 generating the alignment. Dots represent conserved bases, dashes indicate lack of sequence, and
687 divergent bases are shown in blue. The full-length alignment can be found in *SI Appendix*, Fig. S5.
688 (*C*) Methylation quantification of genomic DNA extracted from Chr4-cl WT mice (grey circles)
689 and Chr4-cl KO mice (hollow grey circles) on a pure B6 genetic background. (*D*) Diagram of
690 Chr4-cl KO location (Chr4:145383918-147853419, GRCm38/mm10) and breeding scheme used
691 for the data presented in panels D and E. The Chr4-cl KO was generated in B6 mice, which were
692 subsequently backcrossed to the 129X1/SvJ strain. (*E*) Methylation quantification of genomic
693 DNA extracted from Chr4-cl WT mice (purple circles) and Chr4-cl KO mice (hollow purple
694 circles) on a mixed B6/129 F2 genetic background. IAP-*Sema6d* was excluded from this analysis
695 because it is absent from the 129 genome (Fig. 1*A*).

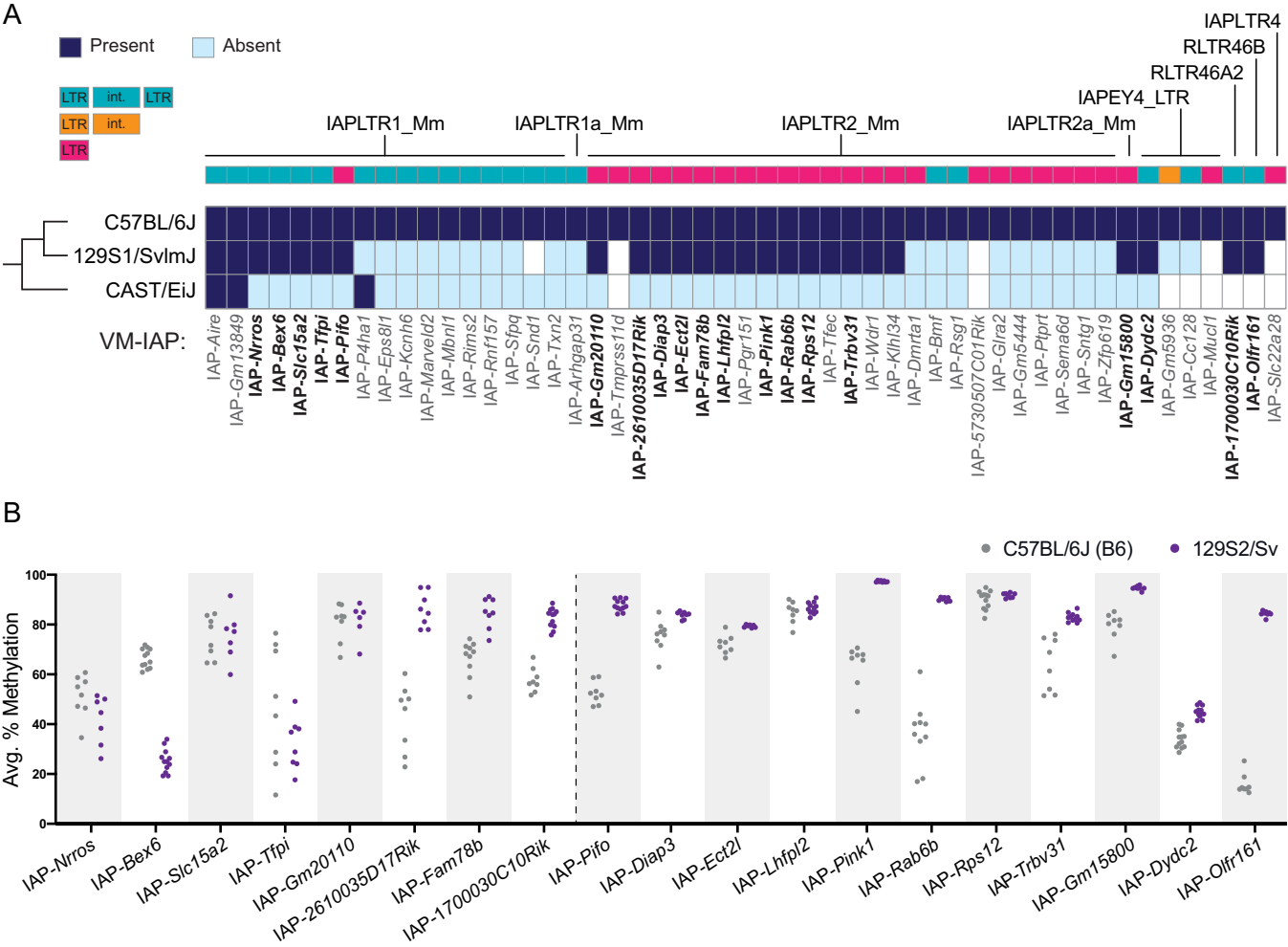
Statistics for panels D and E: unpaired t tests with False Discovery Rate of 5% computed using the two-stage step-up method of Benjamini, Krieger and Yekutieli (* $q < 0.05$; ** q value $< 0.01$; *** $q < 0.001$; **** $q < 0.0001$).

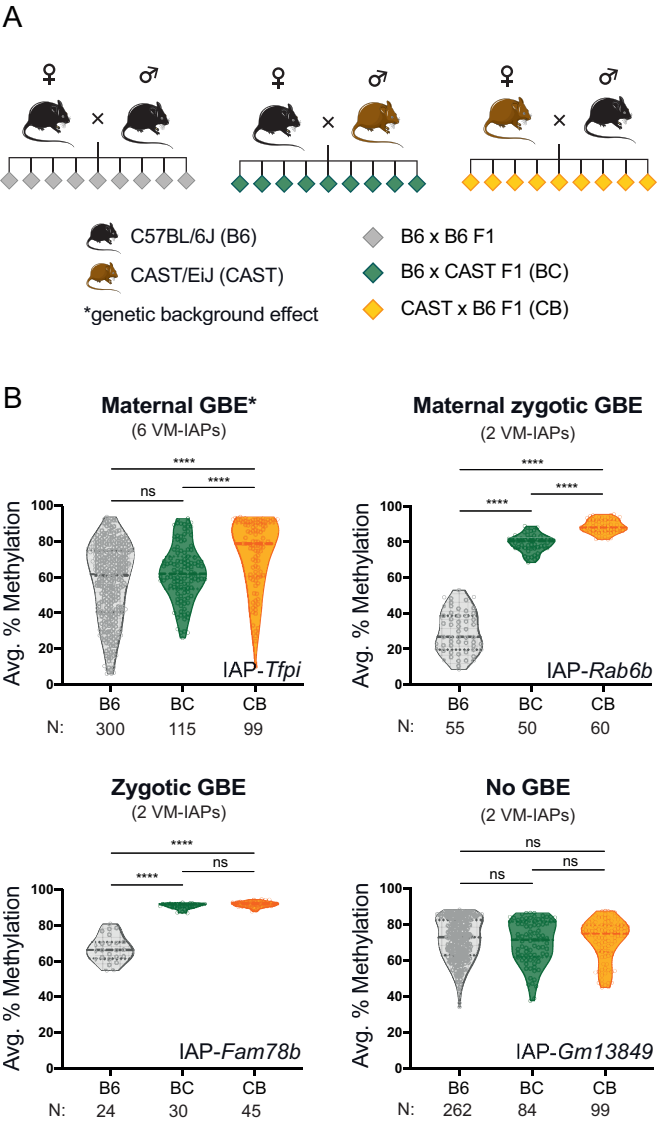
**Fig. 5.** Chr4-cl influences the chromatin and transcriptional landscape at and near targeted VM-IAPs. (*A*) H3K9me3 and H3K4me3 ChIP-seq signal at Chr4-cl target VM-IAPs in Chr4-cl WT (black) and KO (blue) ES cells of mixed B6/129 genetic background. BAM coverage tracks were generated and visualised in IGV. VM-IAPs are shown in red and directionality is indicated with a white arrow. NCBI37/mm9 genome coordinates and neighbouring annotated genes are displayed above and below the ChIP-seq tracks, respectively. (*B*) As in panel *A*, but for non-target VM-IAPs. (*C*) Mean H3K4me3 ChIP-seq signal over the seven confirmed Chr4-cl targets (upper panel) and over all solo LTRs of the IAPLTR2_Mm subclass in the mouse genome (lower panel). Dotted lines represent mean signal and shaded regions represent error estimates (standard error and 95% confidence interval). Plots were generated using SeqPlots software (47). (*D*) RNA-seq signal from Chr4-cl WT (black) and KO (blue) ES cells of mixed B6/129 genetic background for the genes *Pink1* (upstream of IAP-*Pink1*), *Slco2A1* (upstream of IAP-*Rab6b*), and *Rab6b* (downstream of IAP-*Rab6b*). Two biological replicates per genotype are shown. Datasets were downloaded from the GEO database (accession numbers listed in Dataset S1: Table S6).

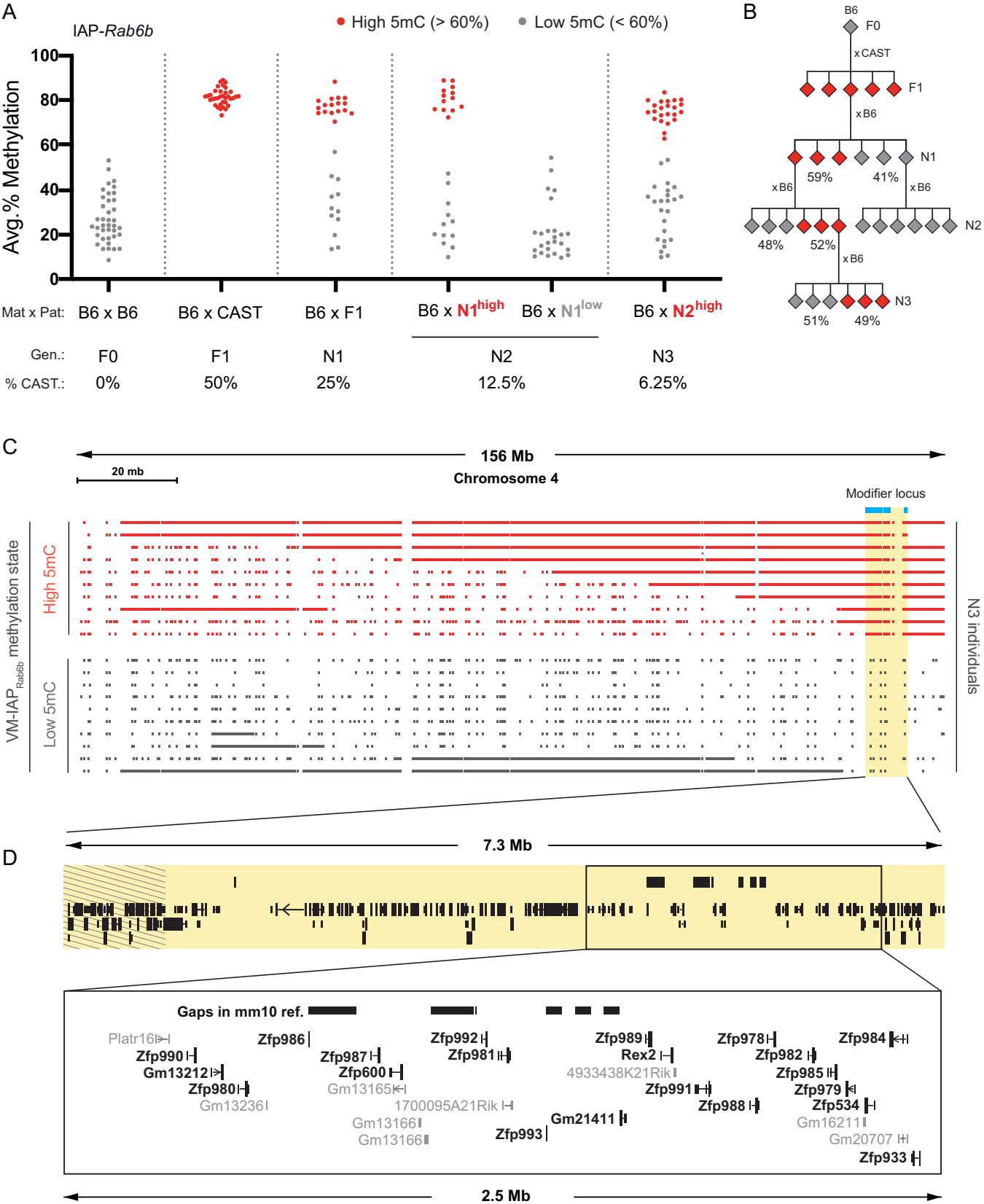
**Fig. 6.** Methylation variability at IAPLTR2_Mm elements is sequence driven. (*A*) Neighbour-joining tree of all solo LTR IAPs of the IAPLTR2_Mm subclass in the B6 genome between 200 and 800 bp in length. Solo LTR sequences were aligned using MUSCLE software and the neighbour-joining tree was built using Geneious Prime software. Navy blue and orange nodes

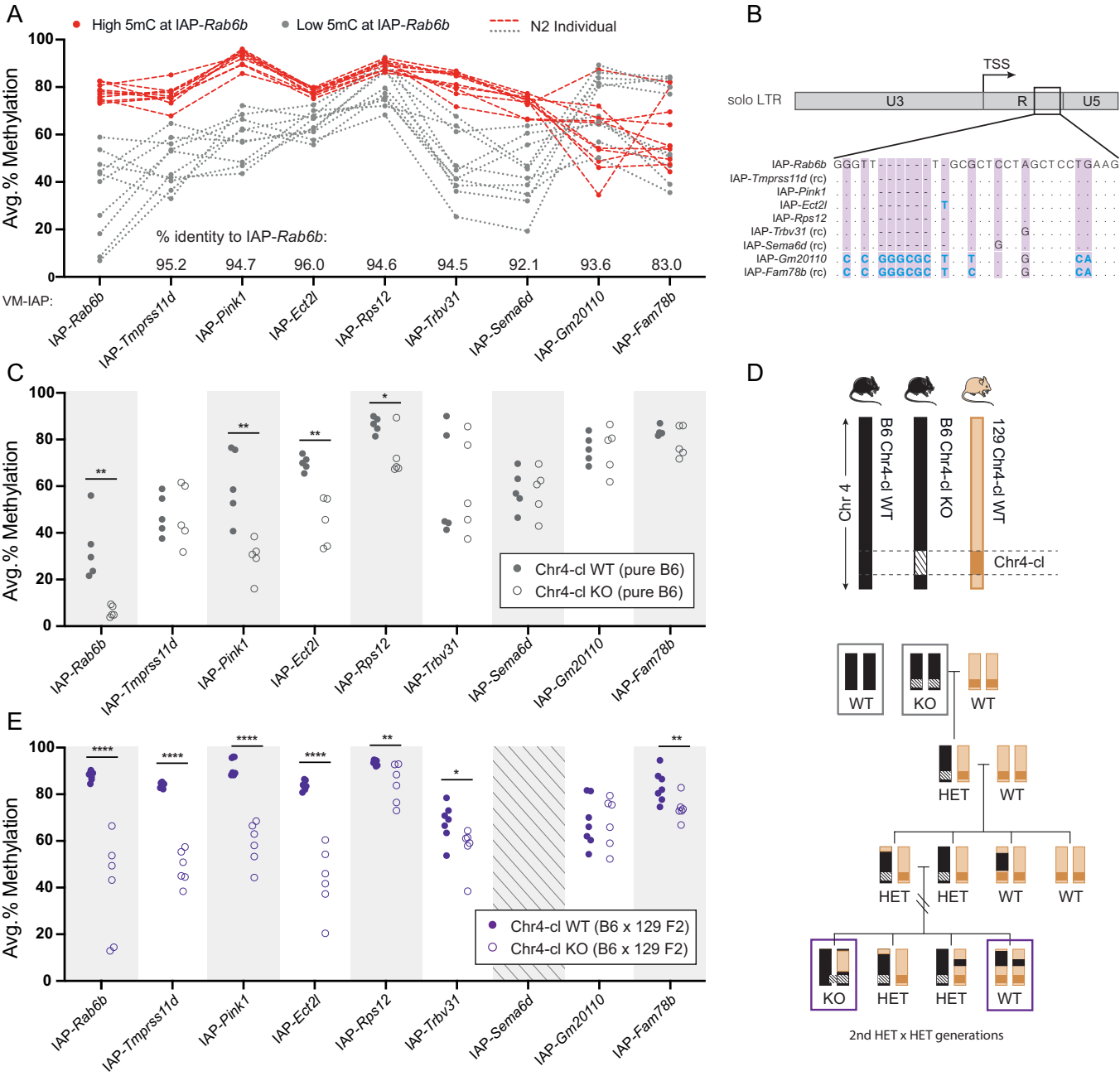
represent experimentally validated VM-IAPs and non-variable IAPs, respectively. The VM-IAP-enriched subtree, containing all known IAPLTR2_Mm VM-IAPs (navy), is shown in greater resolution and labelled with GRC38/mm10 genomic coordinates and strandedness. (*B*) Methylation quantification of genomic DNA from eight B6 individuals at five solo LTRs in the VM-IAP-enriched subtree (orange). Percent sequence identity to IAP-*Rab6b* is shown above the x-axis for each IAP.

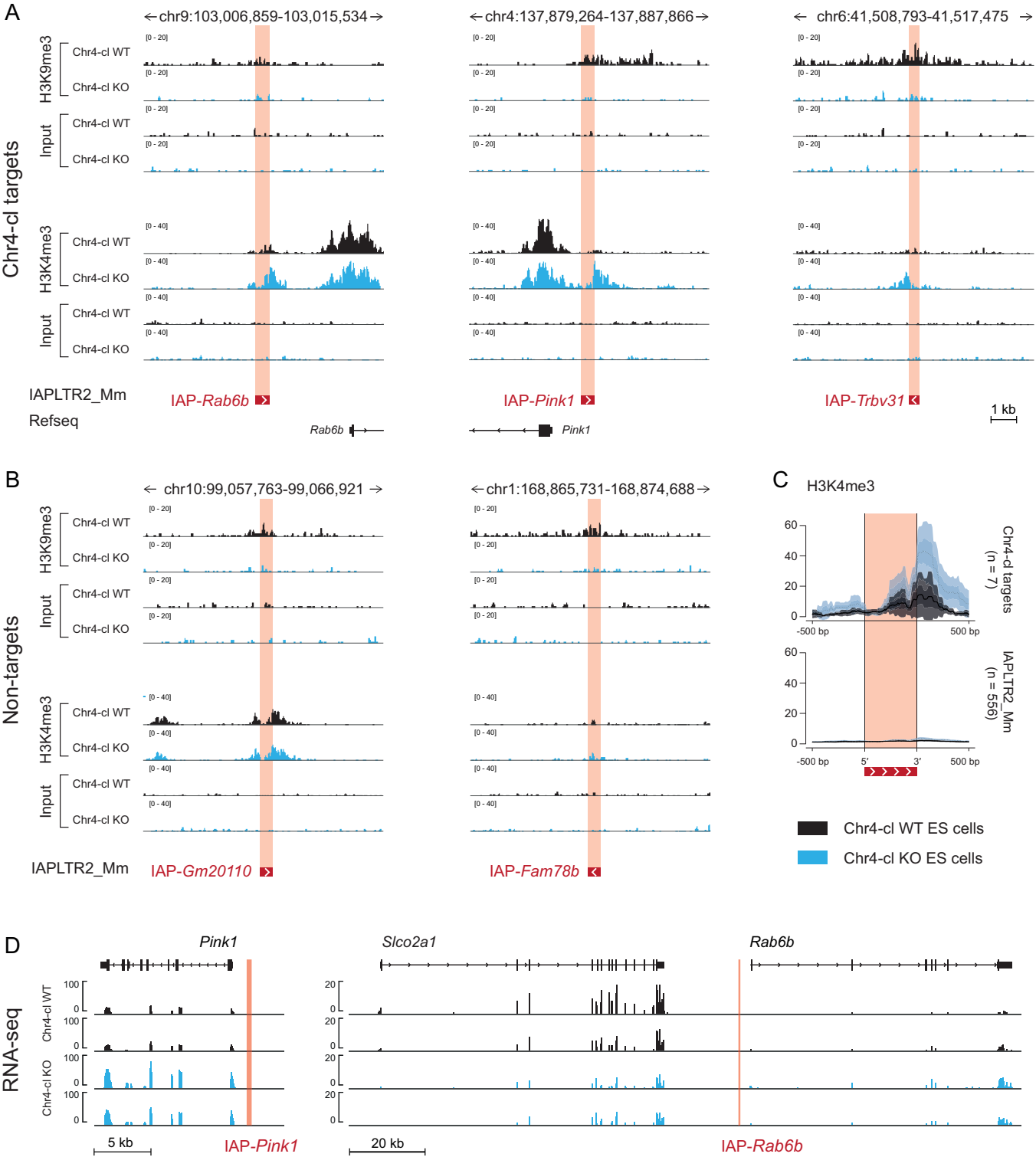
**Fig. 7.** The VM-IAP-enriched subtree exhibits H3K4 trimethylation and distinct KZFP binding. (*A*) Heatmaps of H3K4me3 ChIP-seq coverage in Chr4-cl WT (left) and KO (right) ES cells of mixed B6/129 genetic background over all solo LTRs of the IAPLTR2_Mm subclass (n = 556). VM-IAPs and IAPs belonging to the VM-IAP-enriched subtree were clustered for the analysis. All solo LTRs are anchored from their 5′ start to their 3′ end, with a pseudo-length of 500 bp. The analysis was extended to 500 bp up- and downstream of each element. Average read coverage is plotted above each heatmap. Dotted lines represent mean signal and shaded regions represent error estimates (standard error and 95% confidence interval). Plots and heatmaps were created using SeqPlots (47). (*B*) Heatmaps of overexpressed ZFP989-HA (left), Gm21082-FLAG (middle), and ZFP429-HA (right) ChIP-seq coverage in F9 EC cells over all solo LTRs of the IAPLTR2_Mm subclass (n = 556). Plotting settings as in panel *A*.

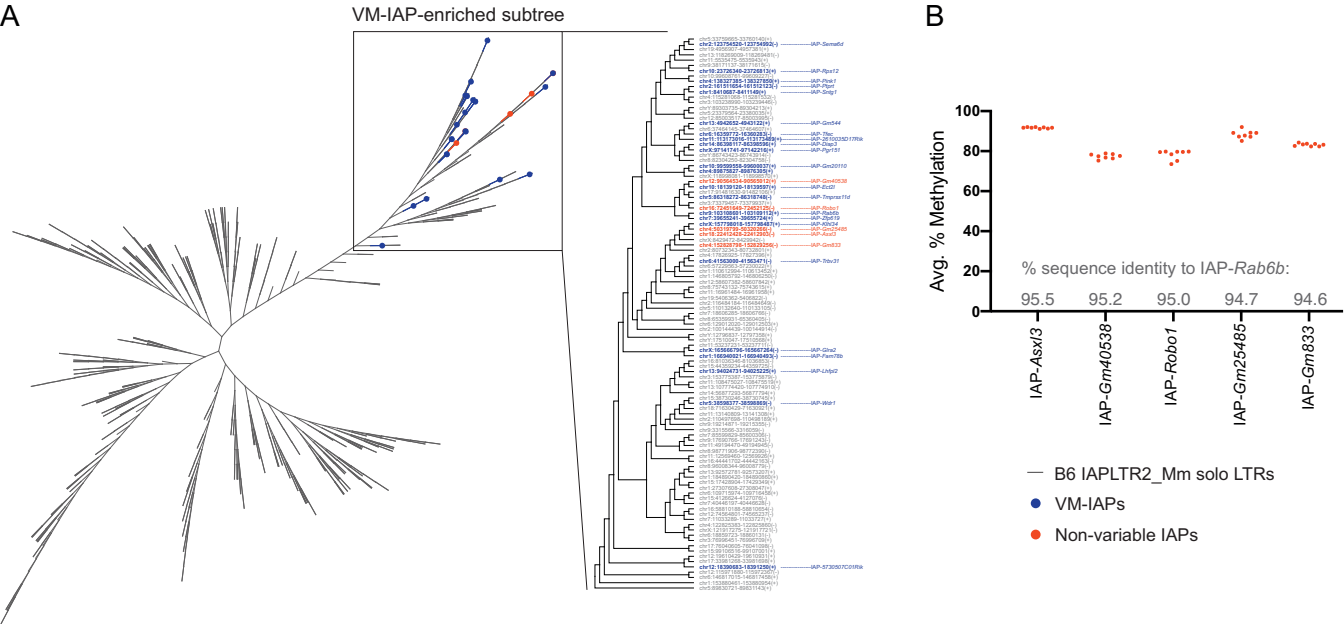

A
Present
Absent
LTR
int.
LTR
LTR
int.
LTR
IAPLTR1_Mm
IAPLTR1a_Mm
IAPLTR2_Mm
IAPLTR2a_Mm
IAPEY4_LTR
RLTR46A2
RLTR46B
IAPLTR4
C57BL/6J
129S1/SvImJ
CAST/EiJ
VM-IAP:
IAP-Aire
IAP-Gm13849
IAP-Nrros
IAP-Bex6
IAP-Slc15a2
IAP-Tfpi
IAP-Pifo
IAP-P4ha1
IAP-Eps8l1
IAP-Kcnh6
IAP-Marveld2
IAP-Mbnl1
IAP-Rims2
IAP-Rnf157
IAP-Sfpq
IAP-Snd1
IAP-Txn2
IAP-Arhgap31
IAP-Gm20110
IAP-Tmprss11d
IAP-2610035D17Rik
IAP-Diap3
IAP-Ect2l
IAP-Fam78b
IAP-Lhfpl2
IAP-Pgr151
IAP-Pink1
IAP-Rab6b
IAP-Rps12
IAP-Tfec
IAP-Trbv31
IAP-Wdr1
IAP-Klhl34
IAP-Dmrta1
IAP-Bmf
IAP-Rsg1
IAP-5730507C01Rik
IAP-Glra2
IAP-Gm5444
IAP-Ptprt
IAP-Sema6d
IAP-Sntg1
IAP-Zfp619
IAP-Gm15800
IAP-Dydc2
IAP-Gm5936
IAP-Cc128
IAP-Mucl1
IAP-1700030C10Rik
IAP-Olfr161
IAP-Slc22a28
B
C57BL/6J (B6)
129S2/Sv
Avg. % Methylation
100
80
60
40
20
0
IAP-Nrros
IAP-Bex6
IAP-Slc15a2
IAP-Tfpi
IAP-Gm20110
IAP-2610035D17Rik
IAP-Fam78b
IAP-1700030C10Rik
IAP-Pifo
IAP-Diap3
IAP-Ect2l
IAP-Lhfpl2
IAP-Pink1
IAP-Rab6b
IAP-Rps12
IAP-Trbv31
IAP-Gm15800
IAP-Dydc2
IAP-Olfr161

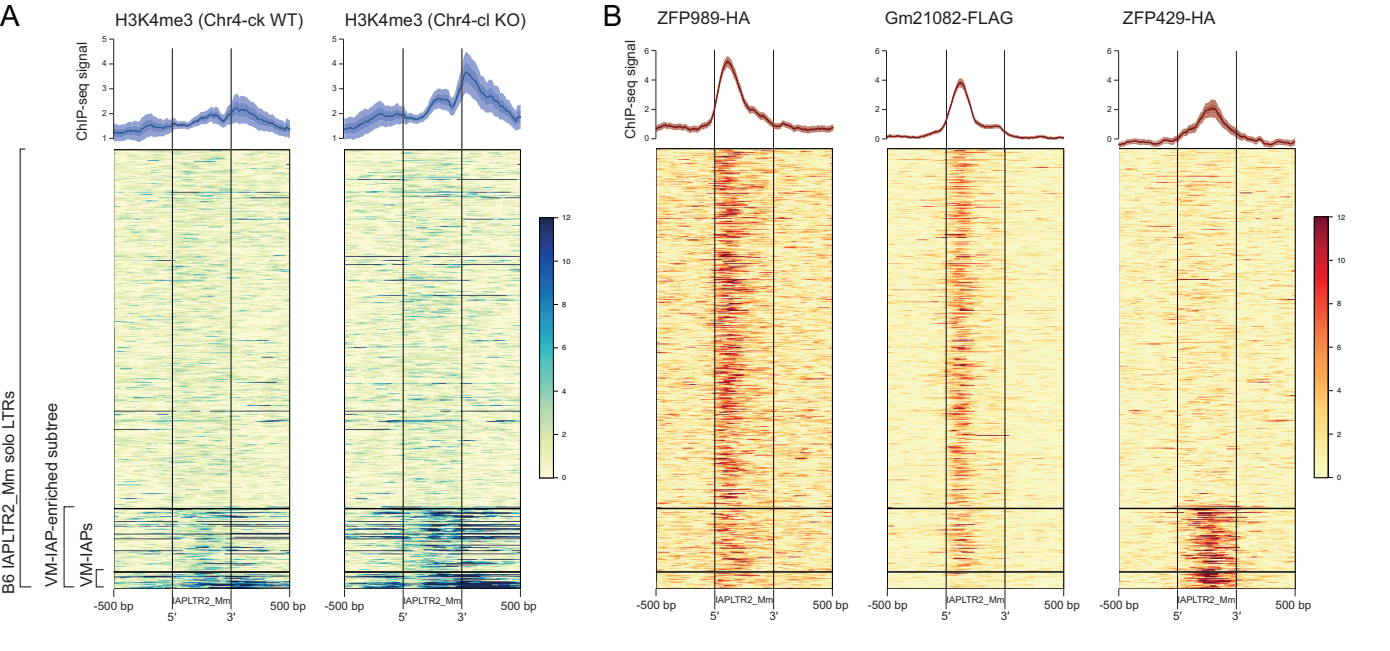

A
C57BL/6J (B6)
CAST/EiJ (CAST)
*genetic background effect
B6 x B6 F1
B6 x CAST F1 (BC)
CAST x B6 F1 (CB)
B
Maternal GBE*
(6 VM-IAPs)
Avg. % Methylation
IAP-Tfpi
B6 BC CB
N: 300 115 99
Maternal zygotic GBE
(2 VM-IAPs)
IAP-Rab6b
N: 55 50 60
Zygotic GBE
(2 VM-IAPs)
IAP-Fam78b
N: 24 30 45
No GBE
(2 VM-IAPs)
IAP-Gm13849
N: 262 84 99


A
IAP-Rab6b
High 5mC (> 60%)
Low 5mC (< 60%)
Avg.% Methylation
100
80
60
40
20
0
Mat x Pat:
B6 x B6
B6 x CAST
B6 x F1
B6 x N1high
B6 x N1low
B6 x N2high
Gen.:
F0
F1
N1
N2
N3
% CAST.:
0%
50%
25%
12.5%
6.25%
B
B6
F0
x CAST
F1
x B6
N1
59%
41%
x B6
x B6
N2
48%
52%
x B6
N3
51%
49%
C
156 Mb
20 mb
Chromosome 4
Modifier locus
VM-IAP Rab6b methylation state
High 5mC
Low 5mC
N3 individuals
7.3 Mb
D
Gaps in mm10 ref.
Platr16
Zfp990
Gm13212
Zfp980
Gm13236
Zfp986
Zfp987
Zfp600
Gm13165
1700095A21Rik
Gm13166
Gm13166
Zfp992
Zfp981
Zfp993
Gm21411
Zfp989
Rex2
4933438K21Rik
Zfp991
Zfp988
Zfp978
Zfp982
Zfp985
Zfp979
Zfp534
Gm16211
Gm20707
Zfp933
Zfp984
2.5 Mb


A
High 5mC at IAP-Rab6b
Low 5mC at IAP-Rab6b
N2 Individual
Avg.% Methylation
% identity to IAP-Rab6b:
95.2
94.7
96.0
94.6
94.5
92.1
93.6
83.0
VM-IAP:
IAP-Rab6b
IAP-Tmprss11d
IAP-Pink1
IAP-Ect2l
IAP-Rps12
IAP-Trbv31
IAP-Sema6d
IAP-Gm20110
IAP-Fam78b
B
TSS
solo LTR
U3
R
U5
IAP-Rab6b GGGTT - - - - - T - GCGCT CCT AGCT CCT GAAG
IAP-Tmprss11d (rc)
IAP-Pink1
IAP-Ect2l
IAP-Rps12
IAP-Trbv31 (rc)
IAP-Sema6d (rc)
IAP-Gm20110 C C GGGCGC T T G CA
IAP-Fam78b (rc) C C GGGCGC T C G CA
C
Avg.% Methylation
Chr4-cl WT (pure B6)
Chr4-cl KO (pure B6)
D
Chr 4
B6 Chr4-cl WT
B6 Chr4-cl KO
129 Chr4-cl WT
Chr4-cl
WT
KO
WT
HET
WT
HET
HET
WT
WT
KO
HET
HET
WT
2nd HET x HET generations
E
Avg.% Methylation
Chr4-cl WT (B6 x 129 F2)
Chr4-cl KO (B6 x 129 F2)


A
chr9:103,006,859-103,015,534
chr4:137,879,264-137,887,866
chr6:41,508,793-41,517,475
Chr4-cl targets
H3K9me3
H3K4me3
Input
Chr4-cl WT
Chr4-cl KO
IAPLTR2_Mm
IAP-Rab6b
IAP-Pink1
IAP-Trbv31
Refseq
Rab6b
Pink1
1 kb
B
chr10:99,057,763-99,066,921
chr1:168,865,731-168,874,688
Non-targets
IAP-Gm20110
IAP-Fam78b
C
H3K4me3
Chr4-cl targets (n = 7)
IAPLTR2_Mm (n = 556)
-500 bp
500 bp
5'
3'
Chr4-cl WT ES cells
Chr4-cl KO ES cells
D
RNA-seq
Pink1
Slco2a1
Rab6b
5 kb
20 kb
IAP-Pink1
IAP-Rab6b


A
VM-IAP-enriched subtree
B
Avg. % Methylation
100
80
60
40
20
0
% sequence identity to IAP-Rab6b:
95.5
95.2
95.0
94.7
94.6
IAP-Asxl3
IAP-Gm40538
IAP-Robo1
IAP-Gm25485
IAP-Gm833
B6 IAPLTR2_Mm solo LTRs
VM-IAPs
Non-variable IAPs


A
H3K4me3 (Chr4-ck WT)
H3K4me3 (Chr4-cl KO)
ChIP-seq signal
B6 IAPLTR2_Mm solo LTRs
VM-IAP-enriched subtree
VM-IAPs
-500 bp
IAPLTR2_Mm
500 bp
5′
3′
B
ZFP989-HA
Gm21082-FLAG
ZFP429-HA