Mechanisms regulating interindividual epigenetic variability at transposable elements

Kevin R. Costello[1,2], Amy Leung[1], Candi Trac[1], Michael Lee[1,2], Mudasar Basam[1], J. Andrew Pospisilik[3] and Dustin E. Schones[1,2]*

[1]Department of Diabetes Complications and Metabolism, Beckman Research Institute

[2]Irell and Manella Graduate School of Biological Sciences, City of Hope, Duarte, CA 91010

[3]Van Andel Research Institute, Grand Rapids, MI, 49503, USA.

Abstract

Transposable elements (TEs) are mobile genetic elements that make up a large fraction of mammalian genomes. While select TEs have been co-opted in host genomes to have function, the vast majority of these elements are epigenetically silenced in somatic cells. However, some TEs in the mouse genome, including the Intracisternal A-particle (IAP) subfamily of retrotransposons, have been shown to display interindividual variation in DNA methylation. While recent work has identified that IAP sequence differences and strain-specific KRAB zinc finger proteins (KZFPs) may influence the methylation state of these IAPs, the mechanisms underlying the establishment and maintenance of interindividual variability in DNA methylation still remain unclear. Here we report that sequence divergence and genomic context influence the likelihood that IAPs become variably methylated. IAPs that have diverged from the consensus IAP sequence have altered KZFP recruitment that can lead to decreased KAP1 recruitment when in a euchromatic environment. These variably methylated loci have a high CpG density and are bound by ZF-CxxC proteins, providing a potential mechanism to maintain this permissive chromatin environment and protect from DNA methylation. These observations indicate that variably methylated IAPs escape silencing through both attenuation of KZFP binding and recognition by CxxC-containing proteins to maintain a hypomethylated state.

Keywords: DNA methylation, Epigenetic variability, Transposable elements, ZF-CxxC proteins, Krab zinc finger proteins, Metastable epialleles

Background

Sequences derived from transposable elements (TEs) make up a large fraction of mammalian genomes (Bourque et al., 2018). TEs are silenced by both histone modifications and DNA methylation to prevent these elements from having any deleterious impact on the host genome. Certain TEs that escape silencing and behave as sites of interindividual epigenetic variability across all tissues are known as metastable epialleles (Bakshi et al., 2016; Dolinoy et al., 2007a; Gunasekara et al., 2019; Hernando-Herraez et al., 2015; Oey et al., 2015; Reiss et al., 2010). These loci have been reported to be sensitive to environmental stimuli such as diet and exposure to toxins (Dolinoy et al., 2007b; Morgan et al., 1999; Rosenfeld et al., 2013; Waterland et al., 2007; Waterland and Jirtle, 2003). The most studied metastable epiallele is found in the Agouti viable yellow mouse, where genetically identical mice display phenotypic diversity that manifests as a range of coat colors and covariant susceptibility to obesity (Bernal et al., 2011; Rakyan et al., 2002). This variable coat color has been linked to the methylation state of an Intracisternal A-particle (IAP) retrotransposon (Dolinoy et al., 2010). When hypomethylated, this IAP acts as a novel promoter element that drives aberrant transcription of the *agouti* gene across all tissues, and subsequently, yellow coat color, hyperphagia and hyperinsulinemia. Elucidating the mechanisms underlying the establishment and maintenance of these variably methylated TEs is important to further our understanding of non-Mendelian phenotypic variability and epigenetic plasticity (Ecker et al., 2018, 2017; Tejedor and Fraga, 2017).

Recent work profiling interindividual epigenetic variability in mice identified roughly 50 loci that are variably methylated in a manner similar to the IAP that drives the agouti mouse phenotype (Elmer et al., 2020). They identified putative metastable epialleles as TEs that displayed 25% methylation

variation between the second-highest and second-lowest average methylation level between biological replicates using the WGBS-seq datasets from purified immune cells in mice (Adams et al., 2012). These metastable epialleles were later confirmed by targeted profiling in multiple tissues, indicating that they are true metastable epialleles (Elmer et al., 2020). The majority of these variably methylated loci are IAP retrotransposons (Bertozzi et al., 2020; Faulk et al., 2013; Kazachenka et al., 2018). IAPs are silenced in a sequence-specific manner by KRAB-domain containing zinc finger proteins (KZFPs) (Coluccio et al., 2018). KZFPs recruit KAP1, also known as TRIM28, which, in turn, recruits repressive protein complexes to deposit H3K9me3 and DNA methylation at these loci (Bulut-Karslioglu et al., 2014; Ecco et al., 2017; Schultz et al., 2002; Turelli et al., 2014). These KZFPs evolve in clusters through imperfect gene duplication and spontaneous mutations that can allow for the recognition of novel TEs (Imbeault et al., 2017; Kauzlaric et al., 2017). Over time, TEs can mutate and escape from KZFP targeted silencing (Imbeault et al., 2017; Kauzlaric et al., 2017). It has been proposed that these KZFPs are engaged in an evolutionary arms race with TEs to maintain their silencing (Jacobs et al., 2014).

IAPs have an unusually high CpG content that is more similar to CpG islands (CGIs) than to the rest of the mouse genome (Elmer et al., 2020; Kazachenka et al., 2018). CGIs are generally hypomethylated and bound by ZF-CxxC proteins, which help maintain a permissive chromatin environment (Clouaire et al., 2012; Gu et al., 2018; Long et al., 2013; van de Lagemaat et al., 2018). Notable CxxC domain containing proteins include CFP1, which is associated with protein complexes that deposit H3K4me3, as well as TET1/TET3, which are responsible for the active removal of methylation from CpG dinucleotides to maintain a hypomethylated state (Gu et al., 2018; Long et al., 2013). Mammalian genomes in general have an underrepresentation of CpG dinucleotides as a result of deamination at methylated cytosines over evolutionary time (Duncan and Miller, 1980; Lewis et al., 2016; Shen et al., 1994). The “richness” of CpG dinucleotides at IAPs is potentially due to the fact these elements are recently evolved TEs that have yet to

undergo ‘evolutionary’ deamination at methylated CpG dinucleotides. It has been shown that recently evolved CpG dense TEs have the potential to be hypomethylated when there is no targeted suppression of these elements (Jacobs et al., 2014; Long et al., 2016; Ward et al., 2013).

Based on this evidence, we postulated that variably methylated CpG dense IAPs can have sequence variants that allow for escape from KZFP-mediated silencing (Bertozzi et al., 2021; Faulk et al., 2013; Kazachenka et al., 2018) and subsequently be bound by ZF-CxxC proteins, thereby preventing stable repression. Focusing on the IAPLTR1_Mm subfamily, we performed multiple sequence alignment and hierarchical clustering of these elements and confirmed that the VM-IAPs have a unique sequence that is distinct from other IAPLTRs. These ‘divergent’ IAPLTRs have altered KZFP binding and diminished KAP1 recruitment. Relative to silenced IAPs - that are found in gene poor regions or proximal to tissue specific transcripts – we find that VM-IAPs are over-represented proximal to constitutively expressed genes. Importantly, we find that CpG dense TEs (regardless of subfamily) have increased recruitment of ZF-CxxC proteins such as CFP1 and TET1 in the absence of KZFP-mediated silencing. This work indicates that it is binding-site competition between repressive (KZFP/KAP1) and protective (ZF-CxxC) factors that leads to the establishment of variable methylation at TEs.

Results

IAPLTR sequence influences KZFP recruitment and VM-IAPs establishment

Recent work has identified novel VM loci, which are largely IAPLTR1_Mms (or IAPLTR1s) and IAPLTR2_Mms (or IAPLTR2s), across tissues in C57BL/6J mice (Bertozzi et al., 2020). As KZFPs are responsible for silencing IAP elements in a sequence-specific manner (Coluccio et al., 2018), we examined if the previously identified VM-IAPs have shared sequence variants that could allow for escape from KZFP-mediated silencing. We performed multiple sequence alignment and hierarchical clustering of all IAPLTR1s and IAPLTR2s greater than 300 bps long using the ete3

pipeline (Huerta-Cepas et al., 2016). This profiling resulted in the characterization of four sequence 'clades' for IAPLTR1s (Figure 1A and Figure 1 - figure supplement 1A), and two sequence 'clades' for the IAPLTR2s (Figure 1 - figure supplement 1B). VM-IAPLTR1s were enriched in IAPLTR1 clade 3 (Figure 1A and Figure 1 - figure supplement 1C), while all VM-IAPLTR2 elements were IAPLTR2 clade 1 elements (Figure 1 - figure supplement 1D). Consistent with previous work (Bertozzi et al., 2020; Faulk et al., 2013; Kazachenka et al., 2018), these results demonstrate that sequence can be a contributing factor in the establishment of VM-IAPs.

To determine whether these identified IAPLTR clades have altered KZFP recruitment, we used published KZFP ChIP-seq libraries (31) to survey the occupancy of KZFPs at IAPLTRs. We identified that IAPLTR1 elements that belong to clade 1 have evidence of binding by ZFP989 and Gm21082, while IAPLTR1 elements that belong to clade 2 have evidence of binding by ZFP429 (Figure 1A). However, IAPLTR1 clade 3 and clade 4 elements do not show evidence of binding by these KZFPs and have decreased KAP1 binding (Figure 1A,B). Similar analysis with IAPLTR2s demonstrated that clade 1 IAPLTR2s are bound by ZFP429, while clade 2 Figure 1 - figure supplement 2A) elements are bound by ZFP989 (Figure 1 - figure supplement 2A). Thus, IAPLTR sequence variants have altered KZFP recruitment and KAP1-mediated silencing.

Chromatin environment can influence establishment of a VM-IAP

While it is clear that sequence is a factor in the establishment of variable methylation, not all IAPLTRs with sequences identical to VM-IAPs are variably methylated (Figure 1A), as has been reported previously (Kazachenka et al., 2018). We therefore examined the potential influence of genomic context, such as being in a euchromatic environment, on the establishment of variable methylation. We used proximity of the IAP to constitutively expressed genes (FPKM > 5 in all tissues profiled (Li et al., 2017)) and ENCODE annotated enhancer elements (ENCODE Project Consortium et al., 2020) as our proxy for a euchromatic environment. We found that 83% of the

VM-IAPLTR1s were within 50kb of a constitutively expressed gene, while the remaining 17% of VM-IAPLTR1s were less than 1kb from an annotated enhancer element. In contrast, only 12% of the non-VM IAPs were proximal to a constitutively expressed gene and 7% were proximal to an enhancer (Figure 1C). This indicates that while sequence allows these elements to have the potential to become variably methylated, VM-IAPs must be in a permissive chromatin environment to escape being silenced, as heterochromatin spreading from neighboring loci could silence these elements.

IAPEz-int sequence variants influence KZFP recruitment

As previously reported, 93% of VM-IAPLTR1s are full length IAPs, while 83% of the VM-IAPLTR2s are solo LTR elements (Kazachenka et al., 2018) (Figure 1 - figure supplement 2B). As the majority of IAPLTR2s exist as solo LTRs this indicates that these elements can be repressed through binding of KZFPs to the IAP internal sequence. The 5' end of IAP internal sequence IAPEz-int, which is flanked by IAPLTR1s and IAPLTR2s, has been shown to be bound by the KZFP Gm14419 (Wolf et al., 2020). In order to investigate the potential of sequence variants in the IAP internal element influencing variable methylation, we performed multiple sequence alignments and hierarchical clustering of the first 150 bps of IAPEz-ints that are flanked by IAPLTR1 or IAPLTR2 elements. We identified three sequence variants and observed that IAPEz-int clade3 elements have the weakest Gm14419 binding signal (**Figure 2*A*** and Figure 2 - figure supplement 1A). We observed that all full length IAPLTR2s only flank clade 1 or 2 IAPEz-ints, suggesting that they are repressed by the binding of Gm14419 to their internal element. However, full length IAPLTR1s can flank either a strong binding IAPEz-int variant or a weak binding IAPEz-int variant that can evade KZFP mediated silencing (Figure 2B). Additionally, nearly all full length VM-IAPs elements have the IAPEz-int clade 3 element (Figure 2 - figure supplement 1B). Together, this demonstrates a mechanism by which some VM-IAPLTRs exist as intact IAPs.

Conservation of IAPLTR1 elements between mouse strains

IAP elements are a potential source of novelty between mouse strains and VM-IAPs in particular have been shown to be highly polymorphic (Gagnier et al., 2019; Kazachenka et al., 2018; Nellåker et al., 2012; Rebollo et al., 2020). In order to investigate the relationship between IAP sequence, potential for variable methylation and conservation across strains, we leveraged structural variant differences between mouse strains identified by the Sanger mouse genome project (Keane et al., 2011). We found that clade 3-4 IAPLTR1s, which have the lowest KAP1 binding of the IAPLTR1 clades (Figure 1B), are the least conserved across mouse strains (Figure 2C). We furthermore found that clade1 IAPLTR1 elements containing the clade 3 IAPEz-int elements are more strain specific than those elements with the clade 2 IAPEz-ints (Figure 2C). Thus, IAPs that have diminished KZFP recruitment exhibit higher intra-strain polymorphisms

IAPs that have loss of KZFP binding have recruitment of the ZF-CxxC containing proteins TET1 and CFP1.

It has previously been reported that IAPLTR1s and IAPLTR2s have a significantly higher CpG density than the genomic average (Elmer et al., 2020; Kazachenka et al., 2018). Given the prevalence of variably methylation at these elements, we wanted to assess the relationship between CpG content and variable methylation directly. Utilizing a previously identified list of all VM-TEs in C57BL/6J mice (Kazachenka et al., 2018), we observed that TE subfamilies with a higher average CpG density had a larger percentage of elements that are variably methylated, with IAPLTRs being the most CpG dense and having the largest percentage of elements being variably methylated (Figure 3A). The CpG content of VM-IAPLTRs, however, is similar to non-VM-IAPLTRs, indicating that high CpG content alone is not sufficient to induce variable methylation (Figure 3B). As CpG dense loci are capable of being recognized and bound by ZF-CxxC proteins (Clouaire et al., 2012; Gu et al., 2018; Long et al., 2013; van de Lagemaat et al.,

2018), we utilized a publicly available ChIP-seq dataset for the ZF-CxxC-domain containing protein TET1 in ES cells (Gu et al., 2018) to examine ZF-CxxC binding across all IAPLTR1s. We identified increased TET1 binding at clade3 IAPLTR1s elements, which is enriched for VM-IAPs (Figure 3C). Additionally, as the ZF-CxxC-domain containing protein CFP1 is a subunit of the SET1 complex responsible for methylating H3K4, we examined available H3K4me3 ChIP-seq data from B cells (Adams et al., 2012) to determine if H3K4me3 is enriched at VM-IAPs. We found that VM-IAPs were more likely to have H3K4me3 signal compared to non-variable IAPs (Figure 3D,E and Figure 3 - figure supplement 1A). *De novo* motif discovery applied to VM-IAP sequences furthermore revealed known recognition sequences for ZF-CxxC proteins such as CFP1 (Figure 3 - figure supplement 1B). ZF-CxxC protein binding has been shown to protect the underlying DNA from methylation (Gu et al., 2018; Thomson et al., 2010). We profiled CpG methylation of VM-IAPs in a TET2/3 knockout cell line (López-Moyado et al., 2019) and found that the VM-IAPs displayed a dramatic increase in CpG methylation when TET2/3 are knocked out (Figure 3 - figure supplement 1C). Thus, VM-IAPs are capable of binding ZF-CxxC proteins to maintain a DNA hypomethylated state.

To determine if recently evolved CpG-dense TEs are also variably methylated in humans, we used a previously identified list of loci that display interindividual epigenetic variation in humans (Gunasekara et al., 2019). Profiling the observed distribution of variably methylated loci in humans compared to a random sampling of the genome, we observed that recently evolved TEs were overrepresented in the variably methylated dataset (Figure 4A). TE subfamily ages were obtained from DFAM (Hubley et al., 2016). We also found a correlation between extent of variable methylation and the average CpG density for in each TE subfamily, with the highly CpG dense transposable element subfamily LTR12C having both one of the highest CpG densities and one of highest percentage of elements that are variably methylated (Figure 4B and Figure 4 - figure supplement 1A). LTR12Cs are recently evolved TEs that have previously been shown to have

latent regulatory potential (Ward et al., 2013) and can act as promoter elements in cancer (Babaian and Mager, 2016). Together our results highlight the potentially conserved importance of high CpG density in the establishment of variable methylation.

CpG dense TEs are hypomethylated and recruit ZF-CxxC proteins in the absence of KZFP-mediated silencing

To further investigate the relationship between KZFP/KAP1 silencing, DNA methylation, and ZF-CxxC binding to CpG dense TEs, we profiled DNA methylation and ZF-CxxC binding in Tc1 mouse livers. This “transchromosomic” mouse carries the majority of human chromosome 21 (Hsa21) possesses human specific transposable elements without corresponding KZFPs (O’Doherty et al., 2005) and has previously been used to identify the latent regulatory potential of recently evolved TEs (Jacobs et al., 2014; Long et al., 2016; Ward et al., 2013). We performed nanopore sequencing to profile DNA methylation of human chromosome 21 in Tc1 mouse liver and compared this with existing DNA methylation data from HepG2 cells (WGBS from ENCODE, ENCSR881XOU (Davis et al., 2018)) (Figure 5A,B). We found that TE-derived CGIs were hypomethylated in the absence of KZFP mediated silencing, consistent with previous reports (Jacobs et al., 2014; Long et al., 2016; Ward et al., 2013). In contrast, TE-derived CGIs that can be targeted by KZFPs shared between mice and humans were largely methylated (Figure 5 - figure supplement 1A). These silenced TE-derived CGIs are largely Alu elements that are derived from the same ancestral sequence as mouse B2 SINE elements and can be silenced by the KZFP ZNF182, which is present in both mice and humans (Imbeault et al., 2017). To profile ZF-CxxC occupancy at these hypomethylated TEs, we performed CUT&RUN for CFP1 in Tc1 mice and compared this with existing CFP1 ChIP-seq data from human erythroblasts (van de Lagemaat et al., 2018). We observed that CpG dense TEs that are not targeted by KZFPs displayed CFP1 recruitment, while the silenced TEs, regardless of CpG content, were largely unbound (Figure 5C). This CFP1 binding was also unique to the Tc1 mice as there was no CFP1 binding at these

loci in humans (Figure 5D), while non-TE CGIs in both mice and humans have CFP1 binding (Figure 5 - figure supplement 1B). Together this work demonstrates that in the absence of targeted KZFP mediated silencing, TEs with high CpG content can be hypomethylated and bound by ZF-CxxC proteins.

Trim28 haploinsufficiency activates evolutionary young and CpG dense TEs

As Trim28/KAP1 is a key protein involved in the establishment of heterochromatin at TEs, we profiled Trim28 haploinsufficient mice to determine if decreased abundance of Trim28 allows for CpG dense TEs to escape silencing. Trim28 haploinsufficiency has previously been shown to impact the Agouti viable yellow mouse coat color (Daxinger et al., 2016) and drive a bimodal obesity phenotype in FVB/NJ mice (Dalgaard et al., 2016). We performed both H3K4me3 ChIP-seq and CFP1 CUT&RUN in the livers of Trim28 haploinsufficient (*Trim28*$^{+/D9}$) mice and their wild type littermates (Methods; Figure 6A). We identified 69 loci that had novel H3K4me3 signal in the *Trim28*$^{+/D9}$ mice compared to control (Methods; Figure 6 - figure supplement 1A). The majority of these loci with novel H3K4me3 recruitment were recently evolved transposable elements (Figure 6B). These loci also had a significant increase in CFP1 signal in *Trim28*$^{+/D9}$ mice relative to control mice (Figure 6C and Figure 6 - figure supplement 1B). Additionally, we profiled changes in expression of TE subfamilies in the *Trim28*$^{+/D9}$ mice compared to wild type mice using our previously generated RNA-seq datasets (Dalgaard et al., 2016). We identified that TEs which have a significant increase in gene expression in the *Trim28*$^{+/D9}$ mice compared to wild type mice have significantly higher CpG density than expected when compared to non-responsive TEs (Figure 6D). Together this work further demonstrates that loss of KAP1 mediated silencing allows for CpG dense transposable elements to be bound by ZF-CxxC proteins and escape targeted silencing.

Discussion

This work provides the foundation for a deeper understanding of how mammalian genomes are shaped during evolution. It has been proposed that genomes are locked in an “evolutionary arms race” between endogenous retrovirus and KZFPs to maintain proper silencing of the genome. It appears as though the variable methylation observed at IAPs could be the result of the IAPs winning this “arms race”. We observed that VM-IAPs have unique sequence variants that allow them to have diminished KZFP recruitment. This provides an explanation as to why previous works stated there is a sequence bias for variably methylated IAPs in mice (Bertozzi et al., 2020; Faulk et al., 2013; Kazachenka et al., 2018). We also found that IAPs with the same sequence variants as the VM-IAPs are more likely to be polymorphic between individual mouse strains. Interestingly, we find that the IAPLTRs flanking the “master” IAP that has been proposed to be responsible for an expansion of IAPs in C3H/HeJ mice (Rebollo et al., 2020) has the same sequence as the clade 3 IAPLTR1s (Figure 7 - figure supplement 1). Together this implies that these VM-IAPs may be more active in certain mouse strains and may be propagating due to a lack of repression.

Our work also highlights the importance of CpG density to the activation of the endogenous retroviruses. We find that ZF-CxxC recruitment can play a role in the protection from DNA methylation at CpG-dense TEs, while CpG poor TEs remain silenced. Additionally, we observed that recently evolved CpG-dense TEs were most likely to be variably methylated in humans. In the absence of targeted silencing, these CpG dense TEs have the potential to become hypomethylated and recruit CxxC domain containing proteins. These results illustrate how deamination of methylated loci could be a mechanism for driving permanent silencing of transposable elements over time. Overall, this work profiles the mechanisms involved in regulating variable methylation in mammalian genomes.

Based on our results, we propose a model that VM-IAP loci are established as a result of partial KZFP mediated silencing at CpG dense TEs, which are protected from silencing through recruitment of ZF-CxxC containing proteins such as TET1 (Figure 7). This potentially results in a stochastically silenced locus, which is then inherited by all tissues during development. Together our work has significant implications in furthering our understanding of how these recently evolved transposable elements can alter both the genomes and epigenomes of mammals.

Methods

Multiple sequence alignment and hierarchical clustering

IAPLTR1_Mm, IAPLTR2_Mm, and IAPEz-int elements were defined by RepeatMasker (SMIT and A., 2004) in the mm10 genome. Intact elements were selected as elements at least 75% the length of the consensus IAP. IAPs were processed using the ete3 pipeline (Huerta-Cepas et al., 2016) using mafft with default settings (Katoh and Standley, 2013), and bases that were found in less than 10% of the samples were trimmed from the multiple sequence alignment (Capella-Gutiérrez et al., 2009). The multiple sequence alignments were then hierarchically clustered using phyml (Guindon et al., 2010). Sequence clades were empirically selected. For IAPEz-ints, only IAPEz-int elements flanked by either IAPLTR1_Mm or IAPLTR2_Mm elements were selected for profiling. The first 150bps of the 5' end of the IAPEz-int were extracted due and processed using the ete3 pipeline as previously stated.

ChIP-seq analysis of available data

ChIP-seq reads were downloaded from GEO using SRA-toolkit https://sbgrid.org/software/titles/sra-toolkit. Reads were aligned to the mm10 genome using bowtie2 with the --end-to-end --very-sensitive options (Langmead and Salzberg, 2012). Aligned reads were sorted using samtools (Li et al., 2009) and filtered for duplicate reads using the MarkDuplicates function of Picardtools http://broadinstitute.github.io/picard/. Bedgraph files were

generated using bedtools genomecov with the -bg -ibam options (Quinlan and Hall, 2010). BigWigs were generated using the UCSC tool bedGraphToBigWig (Karolchik et al., 2004). Heatmaps across the IAP multiple sequence alignments by using a custom script to find the read coverage at each base and were visualized using pheatmap https://cran.r-project.org/package=pheatmap. Heatmaps and aggregate plots of loci that extend beyond the IAP element were generated using deeptools (Ramírez et al., 2016). Regions of enrichment were called using MACS2 callpeaks (Zhang et al., 2008) with a *q*-val threshold of 0.001.

Chromatin context profiling

We used proximity to constitutively expressed genes or annotated enhancer elements as a proxy for profiling a euchromatic environment. Constitutive expressed genes were identified as genes that have a fragments per kilobase of transcript per million mapped reads (FPKM) greater than 5 in more than 90% of the samples profiled in the mouse gTEX project (Li et al., 2017), while enhancers were all ENCODE annotated enhancer elements (ENCODE Project Consortium et al., 2020).

Polymorphism analysis between mouse strains

Coordinates of the structural variants between mouse strains were obtained from the Sanger mouse genome project (Keane et al., 2011). Deletions between mouse strains were extracted for each of the profiled mouse strains. Bedtools intersect (Quinlan and Hall, 2010) with the -f 1 option was used to determine if an IAP was fully deleted in each mouse strain relative to C57BL/6J. Phylogeny of the mouse strains was obtained from previous work (Nellåker et al., 2012). For branches where multiple mouse strains are similarly diverged from C57BL6/J, the IAP was considered to be present if the IAP existed in any of profiled mice at that branch.

MinION sequencing and alignment

Livers from B6129S-Tc(HSA21)1TybEmcf/J mice were purchased from JAX Laboratory. Purified DNA from one Tc1 mouse liver sample was prepared using the SQK-LSK109 ligation sequencing kit and run on R9 flow cells. 1ug of DNA was end repaired with NEBNext FFPE DNA repair (NEB) and Ultra II End-prep (NEB). The repaired DNA was cleaned with KAPA Pure Beads (KAPA). Adapters were ligated using NEBNext quick T4 ligase (NEB). Prepared libraries were mixed with Sequencing Buffer from the SQK-LSK109 kit and approximately 200ng were loaded onto the MinION flowcell. Reads were aligned using bwa mem with the options -x ont2d -t 100 (Li, 2013) to a custom assembly that included all of the mm10 genome and human chromosome 21, and then sorted using samtools sort (Li et al., 2009). Reads were processed using minimap2 with the options -a -x map-ont (Li, 2018). Methylation state of CpGs was called using nanopolish with the options call-methylation -t 100 (Loman et al., 2015). Only loci with greater than 5x coverage were considered in the analysis. CpG islands were obtained from the UCSC table browser (Gardiner-Garden and Frommer, 1987; Karolchik et al., 2004). Methylation percentage was averaged across CpG islands.

CFP1 CUT&RUN sequencing and alignment

CUT&RUN was performed as previously described (Skene and Henikoff, 2017) using the Millipore CFP1 monoclonal antibody (ABE211). Briefly, unfixed permeabilized cells are incubated with the CFP1 antibody fused to A-Micrococcal Nuclease. Fragmented DNA was isolated and sequenced on an Illumina HiSeq 2500 System. We assessed standard QC measures on the FASTQ file using FASTQC (Andrews and Others, 2010) and adapters were trimmed using Trimgalore ("Babraham Bioinformatics - Trim Galore!," n.d.; Martin, 2011). Trimmed reads were aligned to either the mm10 genome or a custom assembly that included both the mm10 genome and human chromosome 21 using bowtie1 retaining only reads that could be mapped to unambiguously to a single loci using the -m 1 option (Langmead et al., 2009). Aligned reads were sorted using samtools (Li et al., 2009) and filtered for duplicate reads using the MarkDuplicates function of

Picardtools http://broadinstitute.github.io/picard/. Regions of histone enrichment were called using MACS2 callpeaks (Zhang et al., 2008) with a *q*-val threshold of 0.001. Only peaks found in two animals were retained to remove biological noise. Additionally, due to small fragment size, loci that could not be mapped unambiguously by 90 bp fragments were removed from consideration using the deadzones function from RSEG (Song and Smith, 2011).

Animals

The generation of Trim28+/D9 has been described elsewhere (Blewitt et al., 2005). Animals were kept on a 12 hr light/dark cycle with free access to food and water and housed in accordance with international guidelines. Livers were isolated at 2 weeks of age and immediately snap-frozen for further processing.

ChIP-seq analysis of H3K4me3 ChIP-seq

Chromatin immunoprecipitation was performed with an H3K4me3 antibody (ab8580, Abcam) as previously described (Leung et al., 2013). Isolated DNA was sequenced on an Illumina HiSeq 2500 System. We assessed standard QC measures on the FASTQ file using FASTQC (Andrews and Others, 2010) and adapters were trimmed using Trimgalore ("Babraham Bioinformatics - Trim Galore!," n.d.; Martin, 2011). Reads were aligned to the mm10 genome using bowtie1 retaining only reads that could be mapped to unambiguously to a single locus using the -m 1 option (Langmead et al., 2009). Aligned reads were sorted using samtools (Li et al., 2009) and filtered for duplicate reads using the MarkDuplicates function of Picardtools http://broadinstitute.github.io/picard/. Bedgraph files were generated using bedtools genomecov with the -bg -ibam options (Quinlan and Hall, 2010). BigWigs were generated using the UCSC tool bedGraphToBigWig (Karolchik et al., 2004). Regions of enrichment were called using MACS2 callpeaks (Zhang et al., 2008) with a *q*-val threshold of 0.001. Loci that displayed enrichment of H3K4me3 in at least two animals were retained to remove biological noise. The loci

with increased H3K4me3 as shown in Fig 6*B-C* were identified by finding peaks present in *Trim28*$^{+/D9}$ mice that were absent from WT mice and also had 3x more H3K4me3 signal in *Trim28*$^{+/D9}$ mice compared to WT. Heatmaps and aggregate plots were generated using deeptools (Ramírez et al., 2016).

Differential expression analysis of repetitive element subfamilies.

RNA-seq data was downloaded from the european nucleotide archive. Adapters were trimmed using Trimgalore ("Babraham Bioinformatics - Trim Galore!," n.d.; Martin, 2011). Trimmed reads were aligned to the mm10 genome using Bowtie2 (Langmead and Salzberg, 2012), and further processed using the RepEnrich2 package (Criscione et al., 2014). This package totals the number of reads mapping to each repetitive element subfamily. Differentially expressed transcripts were identified using DEseq2 (Love et al., 2014) as with a p-val less than 0.05 between *Trim28*$^{+/D9}$ and wild type mice.

Funding and Acknowledgements

Funding was provided by the National Institutes of Health, grants R01DK112041 and R01CA220693 (D.E.S.). Research reported in this publication included work performed by Integrative Genomics Cores of the City of Hope supported by National Institutes of Health under NCI Award P30CA33572. The authors declare that they have no conflicts of interest with the contents of this article. We thank Vanessa Wegert for valuable technical support.

Availability of data and materials

All datasets generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.mln.nih.gov/geo/) under accession no. XXXXXXXX.

Public data used:

Mouse KZFP and KAP1 ChIP-seq GEO: GSE115291. Mouse TET1 ChIP-seq GEO:GSE100957. Mouse H3K4me3 ChIP-seq GEO: GSM2480410. Mouse CFP1 ChIP-seq GEO:GSE18578. Mouse TET2/3 DKO WGBS GEO:GSE134396. Human CFP1 ChIP-seq GEO:GSM3132538. HepG2 WGBS ENCODE, ENCSR881XOU. Trim28 haploinsufficient and wild type mouse RNA-seq was obtained from ENA:PRJEB11740. List of VM-loci in humans was obtained from published work (Gunasekara et al., 2019).

Figures Legends

Figure 1: **Sequence and chromatin context influence the establishment of a VM-IAPLTRs.** A) Hierarchical clustering of VM-IAPLTR1 elements based on sequence. Major sequence variants were separated into the separate clades with clade1 (blue), clade 2 (Green), clade 3 (red) and clade 4 (yellow). ZFP989 and Zfp429 ChIP-seq signal mapped across the aligned IAPLTR1 sequence. loc with no homologous sequence across the IAPs are displayed as grey. B) Average KAP1 signal across all IAPLTR1 clades. D) Percent of VM-IAPs and silenced IAP with the same sequence as VM-IAPs that are proximal to a constitutively expressed gene or enhancer element, as a proxy for constitutive euchromatin environment.

Figure 1 - Supplemental 1: A) Multiple sequence alignments and hierarchical clustering of IAPLTR1 and B) IAPLTR2 C) Distribution and observed over expected frequencies of VM-IAPLTR1s and D) VM-IAPLTR2s.

Figure 1 - Supplemental 2: A) Gm21082 and KAP1 ChIP-seq signal mapped across all IAPLTR elements. B) KZFP ChIP signal at all VM-IAPLTR1 and as well as the average KZFP signal across all VM-IAPLTR2 elements.

Figure 2: **IAPEz-int sequence influences the establishment of a variable methylated loci**. A) Hierarchical clustering and Gm14419 binding at the 5' end of IAPEZ-int elements that are flanked by IAPLTR1 and IAPLTR2 elements. B) Percent of IAPLTRs elements that are proximal to each IAPEZ-int variant. C) Conservation of IAPLTR1 and IAPEZ-int elements across all mouse strains as determined by structural variants identified from the Sanger mouse genome project.

Figure 3 - Supplemental 1: A) Multiple sequence alignment and clustering of the first 150 bps of 5' end of IAPEZ-ints. B) Percent of VM-IAPLTR1 elements proximal to each of the Gm14419

clades. C) Aggregate plots of Gm14419 signal across IAPEz-int variants. D) Hierarchical clustering of IAPLTR and 5' end IAPEz-int together with IAPLTR and IAPEz-int clades labeled.

Figure 3: **Divergent VM-IAPs elements have high CpG density and recruitment of CxxC proteins.** A) Percent of LTR elements which display VM for a given TE subfamily and average CpG score of the TE subfamily. The size of each dot is determined by the number of VM-loci for each subfamily. B) Average CpG density of silenced and variably methylated IAPLTRs. C) Heatmaps of TET1 occupation at IAPLTR1 elements. D) The percent of both silenced and variable methylated IAPLTR1s that have H3K4me3 signal.

Figure 3 - Supplemental 1: A) Heatmap of TET1 and H3K4me3 signal across all IAPLTR1s. B) MEME identified motifs present in VM-IAPLTRs. C) Average methylation percentage of non-variable and VM-IAPLTR elements in wild type and TET2/3 double knockout cells.

Figure 4: **Variably methylated loci in humans are enriched for evolutionarily young CpG-dense TEs.** A) Observed over expected distribution of VM-TE elements in humans divided by the evolutionary age of the TE. Evolutionary age was obtained from Dfam. B) Scatterplot showing the percent of LTR elements which display VM for a given TE subfamily and average CpG score of the TE subfamily. The size of each dot is determined by the number of VM-loci for each subfamily.

Figure 4 - Supplemental 1: A) Scatter plot showing the percent of SINE elements that display VM for a given TE subfamily and average CpG score of the TE subfamily in humans.

Fig 5: **CpG dense TEs are hypomethylated and recruit ZF-CxxC proteins in the absence of KZFP mediated silencing**. A) Screen-shot of an LTR12C element which is hypomethylated in the Tc1 mice and has novel CFP1 recruitment. B) Methylation status of non-KZFP targeted CpG

islands in both the Tc1 mice and humans. C) The percent of loci that have CFP1 enrichment and can be uniquely mapped to in the Tc1 mouse. D) CFP1 signal at hypomethylated CpG Islands on human chromosome 21 in Tc1 mice and humans.

Figure 5 - Supplemental 1: A) CpG methylation of all primate specific TEs elements on human chromosome 21. TE elements were separated by CpG density and their capability of being silenced by a KZFP element shared between mouse and humans. B) CFP1 signal at all non-repeat derived CpG Islands on human chromosome 21 in Tc1 mice and humans.

Fig 6: **Trim28 haploinsufficiency leads to activation of evolutionary young and CpG dense TEs.** A) Genome screenshot of an IAPLTR2 element with novel H3K4me3 enrichment in Trim28 haploinsufficient mice. B) Breakdown of loci with novel H3K4me3 in Trim28 haploinsufficient mice. Age of each TE was determined by DFAM. C) CpG density of TE subfamilies with increased expression in Trim28 haploinsufficient mice and a random selection of non-responsive TEs subfamilies, p-values was calculated using a t-test.

Figure 6 - Supplemental 1: A) Heatmap and aggregate plots of H3K4me3 and CFP1 at loci with novel H3K4me3 signal in Trim28 haplo mice. B) Coefficient of variation of H3K4me3 signal at IAP element and CGIs with H3K4me3 signal as determined by MACS2.

Fig 7: **Model for variable methylated transposable elements.** Elements with high CpG density and loss of KZFP binding have the potential to recruit ZF-CxxC proteins to protect these TEs from being silenced. However elements with high CpG density but strong KZFP recruitment or loss of KZFP binding and low CpG density will remain methylated.

Figure 7 - Supplemental 1: Comparison of the IAPLTR associated with the “master” IAP element

identified in the C3H mice compared to the consensus sequence for each clade. Only a section of the IAPLTR that contains identifiable sequence variants for each clade is shown.

References

Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, Bock C, Boehm B, Campo E, Caricasole A, Dahl F, Dermitzakis ET, Enver T, Esteller M, Estivill X, Ferguson-Smith A, Fitzgibbon J, Flicek P, Giehl C, Graf T, Grosveld F, Guigo R, Gut I, Helin K, Jarvius J, Küppers R, Lehrach H, Lengauer T, Lernmark Å, Leslie D, Loeffler M, Macintyre E, Mai A, Martens JHA, Minucci S, Ouwehand WH, Pelicci PG, Pendeville H, Porse B, Rakyan V, Reik W, Schrappe M, Schübeler D, Seifert M, Siebert R, Simmons D, Soranzo N, Spicuglia S, Stratton M, Stunnenberg HG, Tanay A, Torrents D, Valencia A, Vellenga E, Vingron M, Walter J, Willcocks S. 2012. BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol* **30**:224–226.

Andrews S, Others. 2010. FastQC: a quality control tool for high throughput sequence data.

Babaian A, Mager DL. 2016. Endogenous retroviral promoter exaptation in human cancer. *Mob DNA* **7**:24.

Babraham Bioinformatics - Trim Galore! n.d. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Bakshi A, Herke SW, Batzer MA, Kim J. 2016. DNA methylation variation of human-specific Alu repeats. *Epigenetics* **11**:163–173.

Bernal AJ, Murphy SK, Jirtle RL. 2011. Chapter 15 - Mouse Models of Epigenetic Inheritance In: Tollefsbol T, editor. Handbook of Epigenetics. San Diego: Academic Press. pp. 233–249.

Bertozzi TM, Elmer JL, Macfarlan TS, Ferguson-Smith AC. 2020. KRAB zinc finger protein diversification drives mammalian interindividual methylation variability. *Proc Natl Acad Sci U S A* **117**:31290–31300.

Bertozzi TM, Takahashi N, Hanin G, Kazachenka A, Ferguson-Smith AC. 2021. A spontaneous genetically induced epiallele at a retrotransposon shapes host genome function. *Elife* **10**. doi:10.7554/eLife.65233

Blewitt ME, Vickaryous NK, Hemley SJ, Ashe A, Bruxner TJ, Preis JI, Arkell R, Whitelaw E. 2005. An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc Natl Acad Sci U S A* **102**:7629–7634.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, Mager DL, Feschotte C. 2018. Ten things you should know about transposable elements. *Genome Biol* **19**:199.

Bulut-Karslioglu A, De La Rosa-Velazquez IA, Ramirez F, Barenboim M, Onishi-Seebacher M, Arand J, Galan C, Winter GE, Engist B, Gerle B, O'Sullivan RJ, Martens JH, Walter J, Manke T, Lachner M, Jenuwein T. 2014. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol Cell* **55**:277–290.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973.

Clouaire T, Webb S, Skene P, Illingworth R, Kerr A, Andrews R, Lee J-H, Skalnik D, Bird A. 2012. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev* **26**:1714–1728.

Coluccio A, Ecco G, Duc J, Offner S, Turelli P, Trono D. 2018. Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naive embryonic stem cells. *Epigenetics Chromatin* **11**:7.

Criscione SW, Zhang Y, Thompson W, Sedivy JM, Neretti N. 2014. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics* **15**:583.

Dalgaard K, Landgraf K, Heyne S, Lempradl A, Longinotto J, Gossens K, Ruf M, Orthofer M, Strogantsev R, Selvaraj M, Lu TT-H, Casas E, Teperino R, Surani MA, Zvetkova I, Rimmington D, Tung YCL, Lam B, Larder R, Yeo GSH, O'Rahilly S, Vavouri T, Whitelaw E, Penninger JM, Jenuwein T, Cheung C-L, Ferguson-Smith AC, Coll AP, Körner A, Pospisilik JA. 2016. Trim28 Haploinsufficiency Triggers Bi-stable Epigenetic Obesity. *Cell* **164**:353–364.

Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, Onate KC, Graham K, Miyasato SR, Dreszer TR, Strattan JS, Jolanki O, Tanaka FY, Cherry JM. 2018. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **46**:D794–D801.

Daxinger L, Oey H, Isbel L, Whitelaw NC, Youngson NA, Spurling A, Vonk KKD, Whitelaw E. 2016. Hypomethylation of ERVs in the sperm of mice haploinsufficient for the histone methyltransferase Setdb1 correlates with a paternal effect on phenotype. *Sci Rep* **6**:25004.

Dolinoy DC, Das R, Weidman JR, Jirtle RL. 2007a. Metastable Epialleles, Imprinting, and the Fetal Origins of Adult Diseases. *Pediatr Res* **61**:30R.

Dolinoy DC, Huang D, Jirtle RL. 2007b. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proc Natl Acad Sci U S A* **104**:13056–13061.

Dolinoy DC, Weinhouse C, Jones TR, Rozek LS, Jirtle RL. 2010. Variable histone modifications at the A(vy) metastable epiallele. *Epigenetics* **5**:637–644.

Duncan BK, Miller JH. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature* **287**:560–561.

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

Ecker S, Chen L, Pancaldi V, Bagger FO, Fernández JM, Carrillo de Santa Pau E, Juan D, Mann AL, Watt S, Casale FP, Sidiropoulos N, Rapin N, Merkel A, BLUEPRINT Consortium, Stunnenberg HG, Stegle O, Frontini M, Downes K, Pastinen T, Kuijpers TW, Rico D, Valencia A, Beck S, Soranzo N, Paul DS. 2017. Genome-wide analysis of differential transcriptional and epigenetic variability across human immune cell types. *Genome Biol* **18**:18.

Ecker S, Pancaldi V, Valencia A, Beck S, Paul DS. 2018. Epigenetic and Transcriptional Variability Shape Phenotypic Plasticity. *Bioessays* **40**. doi:10.1002/bies.201700148

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

ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, Dobin A, Kaul R, Halow J, Van Nostrand EL, Freese P, Gorkin DU, Shen Y, He Y, Mackiewicz M, Pauli-Behn F, Williams BA, Mortazavi A, Keller CA, Zhang X-O, Elhajjajy SI, Huey J, Dickel DE, Snetkova V, Wei X, Wang X, Rivera-Mulia JC, Rozowsky J, Zhang J, Chhetri SB, Zhang J, Victorsen A, White KP, Visel A, Yeo GW, Burge CB, Lécuyer E, Gilbert DM, Dekker J, Rinn J, Mendenhall EM, Ecker JR, Kellis M, Klein RJ, Noble WS, Kundaje A, Guigó R, Farnham PJ, Cherry JM, Myers RM, Ren B, Graveley BR, Gerstein MB, Pennacchio LA, Snyder MP, Bernstein BE, Wold B, Hardison RC, Gingeras TR, Stamatoyannopoulos JA, Weng Z. 2020. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**:699–710.

Faulk C, Barks A, Dolinoy DC. 2013. Phylogenetic and DNA methylation analysis reveal novel regions of variable methylation in the mouse IAP class of transposons. *BMC Genomics* **14**:48.

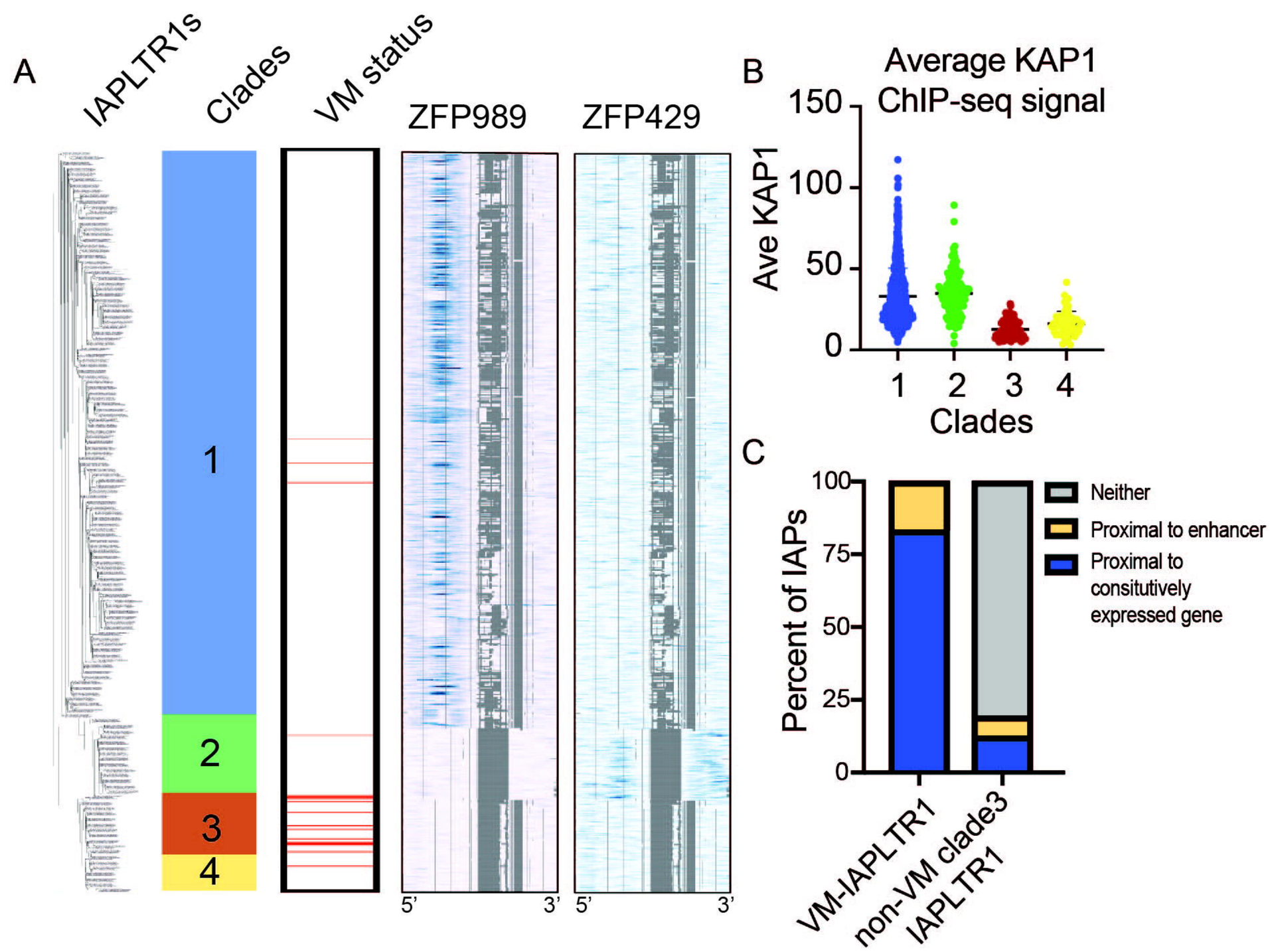
Gagnier L, Belancio VP, Mager DL. 2019. Mouse germ line mutations due to retrotransposon insertions. *Mob DNA* **10**:15.

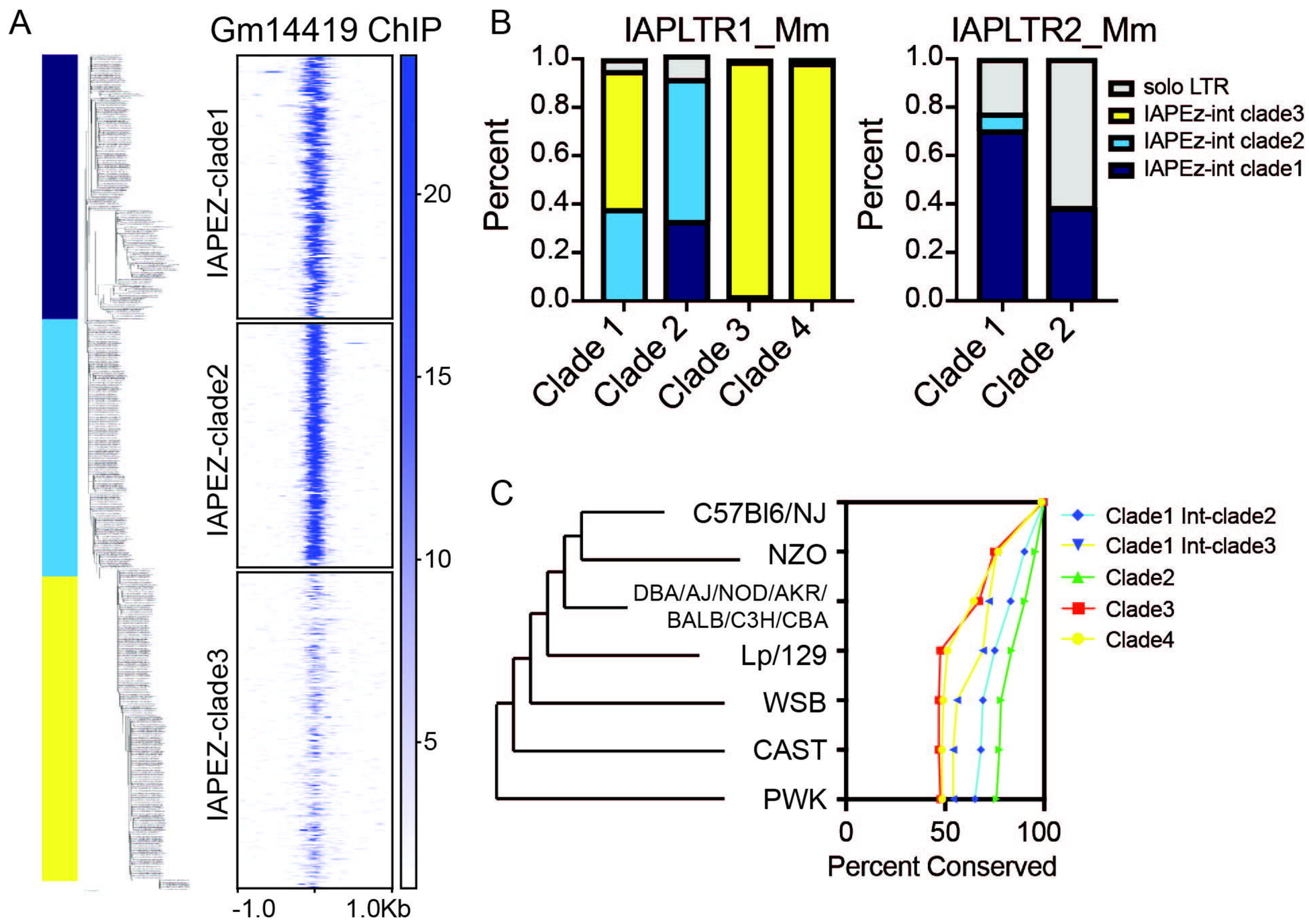
Gardiner-Garden M, Frommer M. 1987. CpG islands in vertebrate genomes. *J Mol Biol* **196**:261–282.

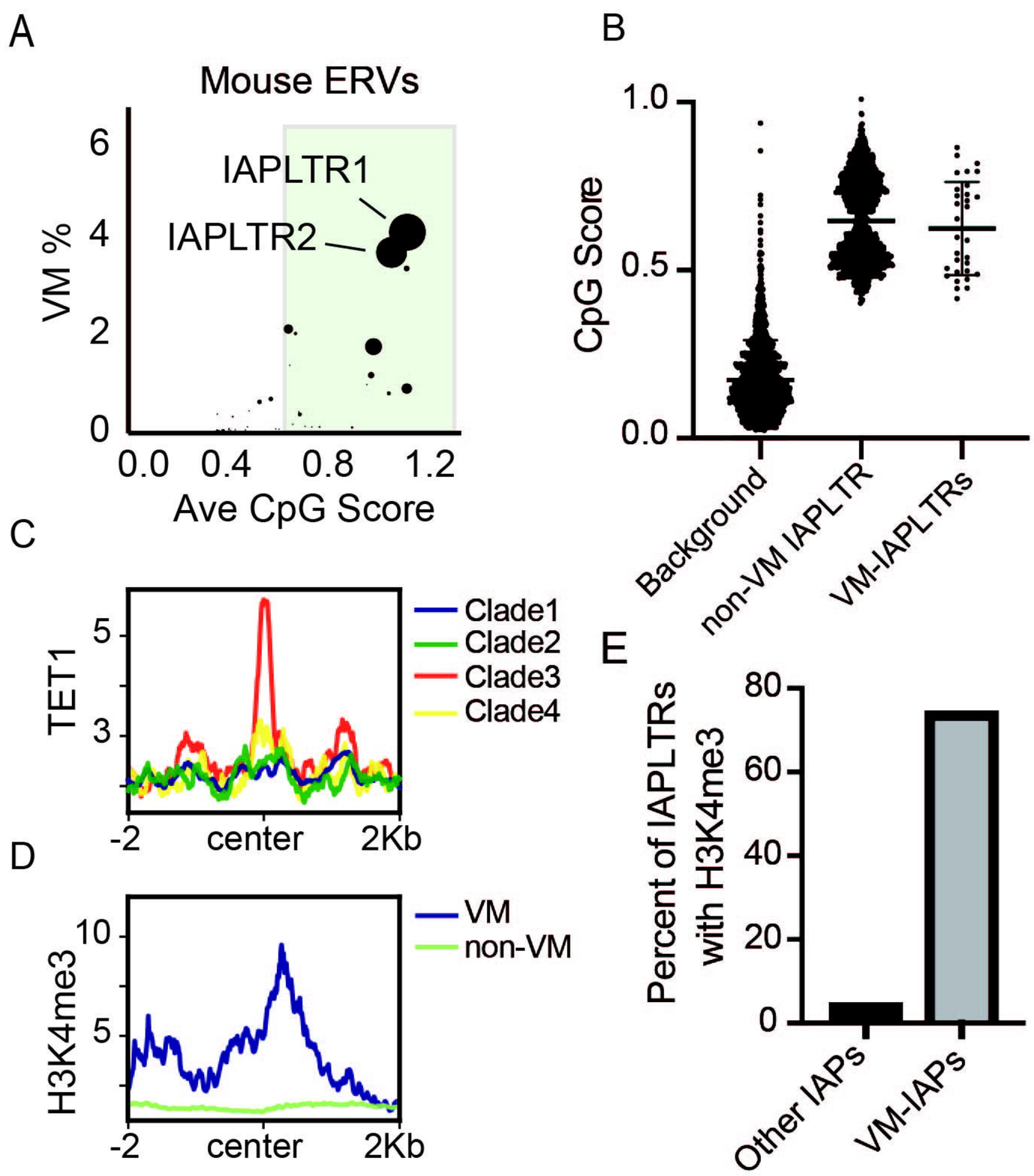
Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**:307–321.
Gunasekara CJ, Scott CA, Laritsky E, Baker MS, MacKay H, Duryea JD, Kessler NJ, Hellenthal G, Wood AC, Hodges KR, Gandhi M, Hair AB, Silver MJ, Moore SE, Prentice AM, Li Y, Chen R, Coarfa C, Waterland RA. 2019. A genomic atlas of systemic interindividual epigenetic variation in humans. *Genome Biol* **20**:105.
Gu T, Lin X, Cullen SM, Luo M, Jeong M, Estecio M, Shen J, Hardikar S, Sun D, Su J, Rux D, Guzman A, Lee M, Qi LS, Chen J-J, Kyba M, Huang Y, Chen T, Li W, Goodell MA. 2018. DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. *Genome Biol* **19**:88.
Hernando-Herraez I, Heyn H, Fernandez-Callejo M, Vidal E, Fernandez-Bellon H, Prado-Martinez J, Sharp AJ, Esteller M, Marques-Bonet T. 2015. The interplay between DNA methylation and sequence divergence in recent human evolution. *Nucleic Acids Res* **43**:8204–8214.
Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, Smit AFA, Wheeler TJ. 2016. The Dfam database of repetitive DNA families. *Nucleic Acids Res* **44**:D81–9.
Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol* **33**:1635–1638.
Imbeault M, Helleboid PY, Trono D. 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**:550–554.
Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**:242.
Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**:D493–6.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**:772–780.
Kauzlaric A, Ecco G, Cassano M, Duc J, Imbeault M, Trono D. 2017. The mouse genome displays highly dynamic populations of KRAB-zinc finger protein genes and related genetic units. *PLoS One* **12**:e0173746.
Kazachenka A, Bertozzi TM, Sjoberg-Herrera MK, Walker N, Gardner J, Gunning R, Pahita E, Adams S, Adams D, Ferguson-Smith AC. 2018. Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell* **175**:1259–1271.e13.
Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, Furlotte NA, Eskin E, Nellåker C, Whitley H, Cleak J, Janowitz D, Hernandez-Pliego P, Edwards A, Belgard TG, Oliver PL, McIntyre RE, Bhomra A, Nicod J, Gan X, Yuan W, van der Weyden L, Steward CA, Bala S, Stalker J, Mott R, Durbin R, Jackson IJ, Czechanski A, Guerra-Assunção JA, Donahue LR, Reinholdt LG, Payseur BA, Ponting CP, Birney E, Flint J, Adams DJ. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**:289–294.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–359.
Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**:R25.
Leung A, Trac C, Jin W, Lanting L, Akbany A, Sætrom P, Schones DE, Natarajan R. 2013. Novel long noncoding RNAs are regulated by angiotensin II in vascular smooth muscle cells. *Circ Res* **113**:266–278.
Lewis CA Jr, Crayle J, Zhou S, Swanstrom R, Wolfenden R. 2016. Cytosine deamination and the precipitous decline of spontaneous mutation during Earth's history. *Proc Natl Acad Sci*

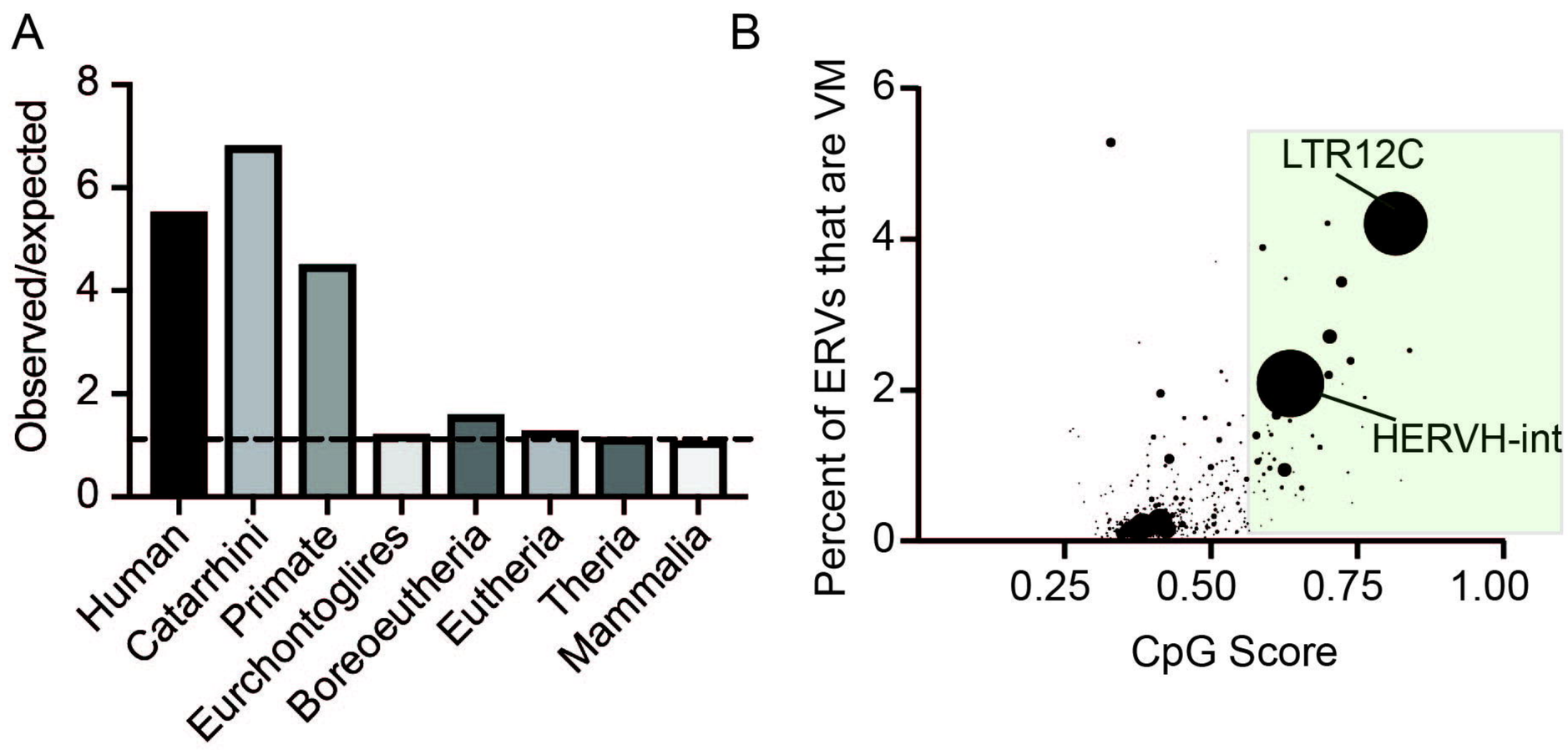
U S A **113**:8194–8199.
Li B, Qing T, Zhu J, Wen Z, Yu Y, Fukumura R, Zheng Y, Gondo Y, Shi L. 2017. A Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. *Sci Rep* **7**:4200.
Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bioGN]*.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079.
Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods* **12**:733–735.
Long HK, Blackledge NP, Klose RJ. 2013. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem Soc Trans* **41**:727–740.
Long HK, King HW, Patient RK, Odom DT, Klose RJ. 2016. Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved. *Nucleic Acids Res* **44**:6693–6706.
López-Moyado IF, Tsagaratou A, Yuita H, Seo H, Delatte B, Heinz S, Benner C, Rao A. 2019. Paradoxical association of TET loss of function with genome-wide DNA hypomethylation. *Proc Natl Acad Sci U S A* **116**:16933–16942.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**:10–12.
Morgan HD, Sutherland HG, Martin DI, Whitelaw E. 1999. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* **23**:314–318.
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.
O'Doherty A, Ruf S, Mulligan C, Hildreth V, Errington ML, Cooke S, Sesay A, Modino S, Vanes L, Hernandez D, Linehan JM, Sharpe PT, Brandner S, Bliss TVP, Henderson DJ, Nizetic D, Tybulewicz VLJ, Fisher EMC. 2005. An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes. *Science* **309**:2033–2037.
Oey H, Isbel L, Hickey P, Ebaid B, Whitelaw E. 2015. Genetic and epigenetic variation among inbred mouse littermates: identification of inter-individual differentially methylated regions. *Epigenetics Chromatin* **8**:54.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842.
Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E. 2002. Metastable epialleles in mammals. *Trends Genet* **18**:348–351.
Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**:W160–5.
Rebollo R, Galvão-Ferrarini M, Gagnier L, Zhang Y, Ferraj A, Beck CR, Lorincz MC, Mager DL. 2020. Inter-Strain Epigenomic Profiling Reveals a Candidate IAP Master Copy in C3H Mice. *Viruses* **12**. doi:10.3390/v12070783
Reiss D, Zhang Y, Rouhi A, Reuter M, Mager DL. 2010. Variable DNA methylation of transposable elements: the case study of mouse Early Transposons. *Epigenetics* **5**:68–79.
Rosenfeld CS, Sieli PT, Warzak DA, Ellersieck MR, Pennington KA, Roberts RM. 2013. Maternal exposure to bisphenol A and genistein has minimal effect on A(vy)/a offspring

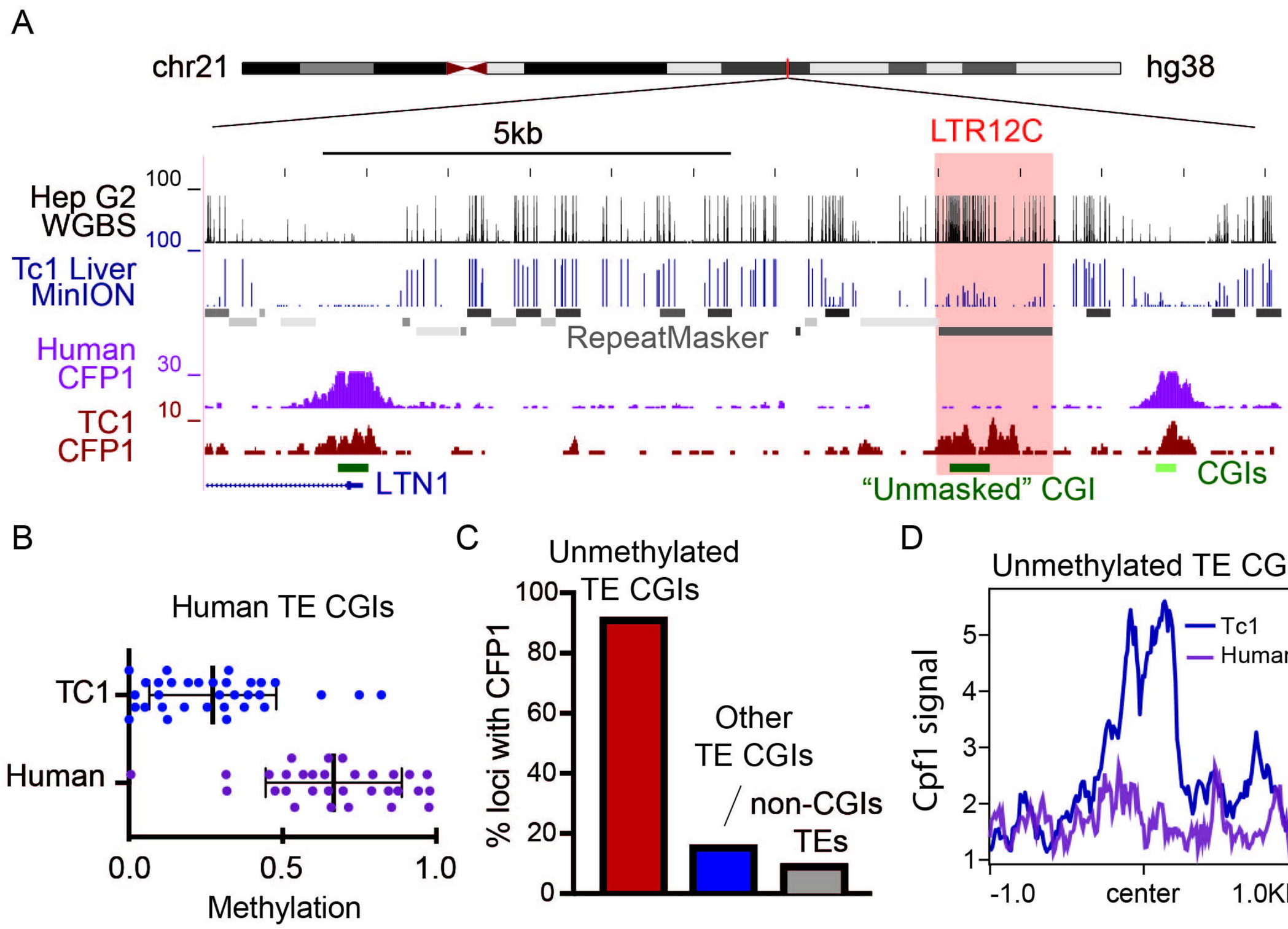
coat color but favors birth of agouti over nonagouti mice. *Proc Natl Acad Sci U S A* **110**:537–542.
Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ. 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* **16**:919–932.
Shen JC, Rideout WM 3rd, Jones PA. 1994. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res* **22**:972–976.
Skene PJ, Henikoff S. 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**. doi:10.7554/eLife.21856
SMIT, A. FA. 2004. Repeat-Masker Open-3.0. *http://www.repeatmasker.org*.
Song Q, Smith AD. 2011. Identifying dispersed epigenomic domains from ChIP-Seq data. *Bioinformatics* **27**:870–871.
Tejedor JR, Fraga MF. 2017. Interindividual epigenetic variability: Sound or noise? *Bioessays* **39**. doi:10.1002/bies.201700055
Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr ARW, Deaton A, Andrews R, James KD, Turner DJ, Illingworth R, Bird A. 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**:1082–1086.
Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, Tieng V, Quenneville S, Trono D. 2014. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res* **24**:1260–1270.
van de Lagemaat LN, Flenley M, Lynch MD, Garrick D, Tomlinson SR, Kranc KR, Vernimmen D. 2018. CpG binding protein (CFP1) occupies open chromatin regions of active genes, including enhancers and non-CpG islands. *Epigenetics Chromatin* **11**:59.
Ward MC, Wilson MD, Barbosa-Morais NL, Schmidt D, Stark R, Pan Q, Schwalie PC, Menon S, Lukk M, Watt S, Thybert D, Kutter C, Kirschner K, Flicek P, Blencowe BJ, Odom DT. 2013. Latent Regulatory Potential of Human-Specific Repetitive Elements. *Mol Cell* **49**:262–272.
Waterland RA, Jirtle RL. 2003. Transposable Elements: Targets for Early Nutritional Effects on Epigenetic Gene Regulation. *Mol Cell Biol* **23**:5293–5300.
Waterland RA, Travisano M, Tahiliani KG. 2007. Diet-induced hypermethylation at agouti viable yellow is not inherited transgenerationally through the female. *FASEB J* **21**:3380–3385.
Wolf G, de Iaco A, Sun M-A, Bruno M, Tinkham M, Hoang D, Mitra A, Ralls S, Trono D, Macfarlan TS. 2020. KRAB-zinc finger protein gene expansion in response to active retrotransposons in the murine lineage. *Elife* **9**. doi:10.7554/eLife.56337
Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**:R137.

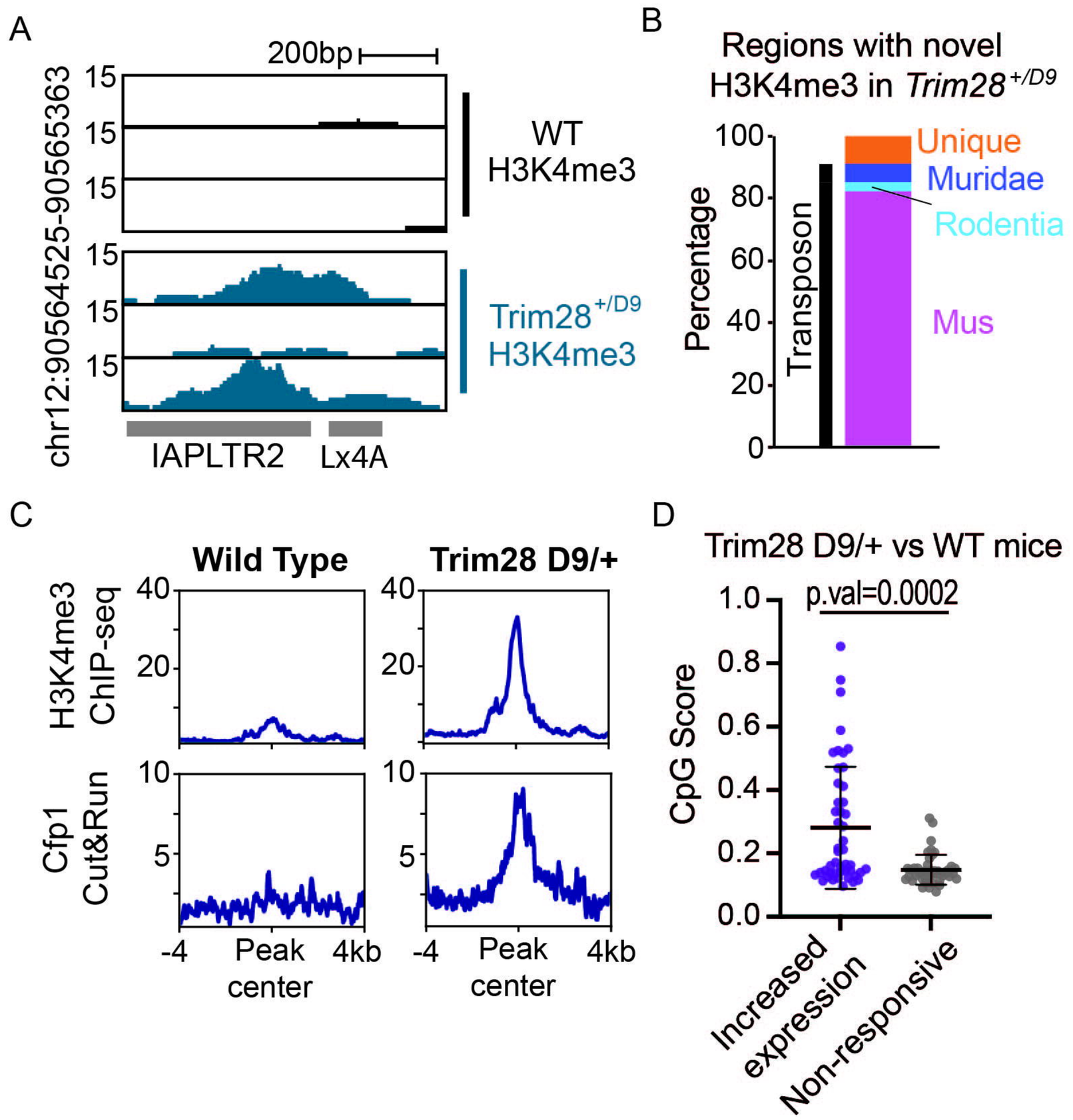

A
IAPLTR1s
Clades
VM status
ZFP989
ZFP429
1
2
3
4
5'
3'
5'
3'
B
Average KAP1
ChIP-seq signal
Ave KAP1
150
100
50
0
1
2
3
4
Clades
C
Percent of IAPs
100
75
50
25
0
Neither
Proximal to enhancer
Proximal to
consitutively
expressed gene
VM-IAPLTR1
non-VM clade3
IAPLTR1

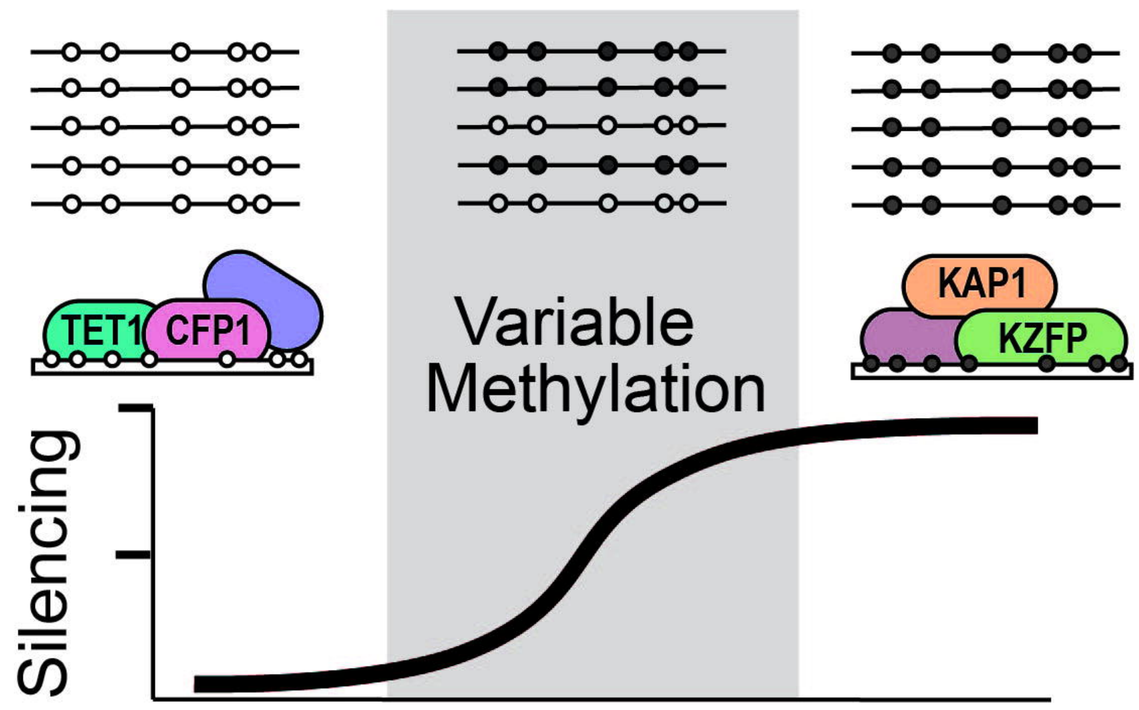

A
Gm14419 ChIP
IAPEZ-clade1
IAPEZ-clade2
IAPEZ-clade3
20
15
10
5
-1.0
1.0Kb
B
IAPLTR1_Mm
IAPLTR2_Mm
Percent
1.0
0.8
0.6
0.4
0.2
0.0
Clade 1
Clade 2
Clade 3
Clade 4
solo LTR
IAPEz-int clade3
IAPEz-int clade2
IAPEz-int clade1
C
C57Bl6/NJ
NZO
DBA/AJ/NOD/AKR/
BALB/C3H/CBA
Lp/129
WSB
CAST
PWK
0
50
100
Percent Conserved
Clade1 Int-clade2
Clade1 Int-clade3
Clade2
Clade3
Clade4


A
Mouse ERVs
IAPLTR1
IAPLTR2
VM %
Ave CpG Score
B
CpG Score
Background
non-VM IAPLTR
VM-IAPLTRs
C
TET1
Clade1
Clade2
Clade3
Clade4
-2
center
2Kb
D
H3K4me3
VM
non-VM
E
Percent of IAPLTRs with H3K4me3
Other IAPs
VM-IAPs


A
Observed/expected
8
6
4
2
0
Human
Catarrhini
Primate
Eurchontoglires
Boreoeutheria
Eutheria
Theria
Mammalia
B
Percent of ERVs that are VM
6
4
2
0
LTR12C
HERVH-int
0.25
0.50
0.75
1.00
CpG Score


A
chr21
hg38
5kb
LTR12C
Hep G2
WGBS
100
100
Tc1 Liver
MinION
RepeatMasker
Human
CFP1
30
10
TC1
CFP1
LTN1
"Unmasked" CGI
CGIs
B
Human TE CGIs
TC1
Human
0.0
0.5
1.0
Methylation
C
Unmethylated
TE CGIs
% loci with CFP1
100
80
60
40
20
0
Other
TE CGIs
non-CGIs
TEs
D
Unmethylated TE CGI
Cpf1 signal
5
4
3
2
1
Tc1
Human
-1.0
center
1.0Kb


A
200bp
chr12:90564525-90565363
15
15
15
15
15
15
WT
H3K4me3
Trim28+/D9
H3K4me3
IAPLTR2
Lx4A
B
Regions with novel
H3K4me3 in Trim28+/D9
Percentage
Transposon
100
80
60
40
20
0
Unique
Muridae
Rodentia
Mus
C
Wild Type
Trim28 D9/+
H3K4me3
ChIP-seq
Cfp1
Cut&Run
40
20
10
5
-4
Peak
center
4kb
D
Trim28 D9/+ vs WT mice
p.val=0.0002
CpG Score
1.0
0.8
0.6
0.4
0.2
0.0
Increased
expression
Non-responsive


TET1
CFP1
Variable
Methylation
KAP1
KZFP
Silencing