Supplemental Figure S1. Categorisation of VM-IAPs.

(A) Methylation level as assessed by bisulphite pyrosequencing of 10 tissue-specific VM-IAPs (tsVM-IAPs) that are variably methylated in multiple tissues. Each point represents an individual and is the average methylation level across the CpGs at the distal end of the IAP LTR. Tissues with a methylation range ≥10% and <9% are shown in red and black, respectively. Tissues with a methylation range ≥9% but <10% are shown in orange. Bars represent the median and range of methylation values between individuals.

(B) Tissue-specific VM-IAPs display methylation consistency across tissues within each individual at a given locus, similarly to cVM-IAPs. Each line represents an individual mouse; the individuals with the highest and lowest methylation in the most variable tissue of each IAP are highlighted for comparison.

(C) As in (A) for the 16 tsVM-IAPs with methylation variability restricted to B cells. n.d. = no data for this tissue.

(D) Methylation range of validated elements separated into constitutive and tissue-specific VM-IAPs. Each point represents an individual element. The methylation range in the most variable tissue is plotted for tsVM-IAPs. For cVM-IAPs, the methylation range in ear is shown. The dotted line represents the median methylation range.

(E) As in (A) for five false positive IAPs with no methylation variability in any tissue assayed.
Supplemental Figure S2

Acss2 mRNA

qPCR primers

Scale
chr2: 155,530,000 155,540,000 155,550,000 155,560,000 155,570,000 155,580,000

Exon number
1 2 3-7 8-13 14 15-18

IAP-Acss2
Acss2
Acss2

SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown

20 kb

ACSS2

1-2

3-7 4-7

ACSS2 OS

GSS

ACSS2 1-2
ACSS2 3-7
ACSS2 4-7

BRAIN

Relative Gene Expression

R² = 0.05935

p = 0.4454

KIDNEY

Methylation (%)
**B**

| Scale       | chr1: | mm10 |
|-------------|-------|------|
|             | 146,550,000 | 146,600,000 |
|             | 146,650,000 | 146,700,000 |
|             | 146,750,000 | 146,800,000 |
| IAP-FamSc   | 146,850,000 | 146,900,000 |

| Exon number | Brinp3 | Brinp3 |
|-------------|--------|--------|
| 1           |        |        |
| 2           |        |        |
| 3           |        |        |
| 4           |        |        |
| 5-6         |        |        |
| 7           |        |        |
| 8           |        |        |

| Simple      | SINE   | LINE  |
|-------------|--------|-------|
| Low Complexity | DNA   |       |
| Satellite    | RNA    | Other |
| Unknown      |        |       |

| Fam5C mRNA | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|------------|---|---|---|---|---|---|---|---|---|
| qPCR primers | 4-6 | 7-9 | 8-9 |

**Fam5C 4-6**

**Fam5C 7-9**

**Fam5C 8-9**

**Brain**

```
| Methylation (%) | Relative Gene Expression |
|-----------------|--------------------------|
| 50              | 1.0                      |
| 60              | 1.0                      |
| 70              | 1.0                      |
| 80              | 1.0                      |
| 90              | 1.0                      |
| 100             | 1.0                      |
```

```
p = 0.3895  
R² = 0.07486
```

**Testes**

```
| Methylation (%) | Relative Gene Expression |
|-----------------|--------------------------|
| 80              | 1.0                      |
| 90              | 1.0                      |
| 95              | 1.0                      |
| 100             | 1.0                      |
```

```
p = 0.4305  
R² = 0.16100
```
Supplemental Figure S2. Tissue-specific VM-IAPs rarely affect gene expression.

Screenshots of the UCSC genome browser in each top panel showing the position of the tsVM-IAP within the gene selected for expression analysis, as well as a diagram of locations of the qPCR primers relative to the exons. The relationship between gene expression and methylation of the tsVM-IAP is shown in each bottom panel. tsVM-IAPs shown are (A) IAP-\textit{Acss2}, (B) IAP-\textit{Fam5C}, (C) IAP-\textit{Rftn1}, and (D) IAP-\textit{Mtmr7}. Expression was quantified by qPCR, with input cDNA normalised to two housekeeping genes, \textit{Pgk1} and \textit{Gapdh}; p-values are from two-tailed Pearson test. n = 6 for testes and 12 for all other tissues.
Supplemental Figure S3

A

B
D

% CpG Methylation

ERV number

- Ear
- B cells
Supplemental Figure S3. Results and validations of screens for variable methylation in retrotransposons.

(A-B) Methylation ranges in B and T cell WGBS data for (A) IAP elements and (B) ERVs. Each point represents an individual element: cVM-IAPs, tsVM-IAPs and false positives are coloured in (A); validated ERVs (15 VM-ERVs and 18 false positives) are coloured in (B).

(C) Bisulfite pyrosequencing of candidate ERVs assayed in ear samples. Only those passing the ≥10% methylation range threshold are shown. Each line represents an individual mouse. Samples were run in triplicate; the error bar shows the standard deviation of technical replicates. CpGs are assayed from the distal end of the ERV.

(D) Methylation level of individual ERVs which did not cross the ≥10% methylation range threshold, assayed in ear and B cells. Each point represents an individual and is the average methylation level across the CpGs located at the distal end of the ERV. Ear and B cell samples are shown in black and blue, respectively. ‘ERV numbers’ are shown, as gene-based identifiers were only generated for VM-ERVs.

(E) Same as panel (C), showing pyrosequencing of LINE, SINE, and non-ERV LTR candidates. All tested elements are shown (except the two SINEs in (F)), including those that did not pass the ≥10% methylation range threshold.

(F) Two tested SINEs are bistable epialleles due to their location on the X chromosome. Female and male individuals are coloured in red and black, respectively. Methylation was assayed in ear.
**Supplemental Figure S4. Increased CpG density in IAPLTR2_Mm cVM-IAPs.**

(A) The two IAP LTR types over-represented in cVM-IAPs, LTR1_Mm and LTR2_Mm, have higher average CpG density compared to all IAP LTRs (means: LTR1_Mm=5.496, LTR2_Mm=4.993, all IAP LTRs=4.345; p<2.2e-16 for both). LTR1_Mm cVM-IAPs have significantly lower CpG density compared to non-variable LTR1_Mm elements (difference of means=-0.348, Cohen’s d=-0.891, p<0.003), while LTR2_Mm cVM-IAPs have significantly higher CpG density compared to non-variable LTR2_Mm elements (difference of means=1.44, Cohen’s d=1.431, p<1.3e-12). The effect size measured by Cohen’s d is much larger for LTR2_Mm elements than LTR1_Mm elements. tsVM-IAPs are not enriched for a particular LTR type and do not show any difference in CpG density compared to all IAP LTRs (p= 0.18). The number of elements analysed per IAP LTR type is shown in parentheses to the right of the IAP LTR type names. CpG density was calculated by normalising the number of CpGs to the LTR length in base pairs for each IAP LTR; p-values are from Welch’s t test. Cohen’s d is an unbiased measure of effect size. Analyses only included LTRs; for fully-structured IAPs, only the 5′ LTR was used.

(B) VM-ERVs have higher average CpG density compared to all other ERVs (means: VM-ERVs=2.968, all ERVs=0.804; p=5.99e-5), but lower average CpG density compared to cVM-IAPs and tsVM-IAPs (means: cVM-IAPs=5.259, tsVM-IAPs=4.628). CpG density was calculated by normalising the number of CpGs to the LTR length in base pairs for each ERV LTR; p-values are from Welch’s t test.
Supplemental Figure S5. 18 IAP elements containing sequences over-represented in cVM-IAPs are hypermethylated.

Each point represents an individual mouse and is the average methylation level across the CpGs at the distal end of the IAP LTR, n = 8; methylation was assayed in kidney.
Supplemental Figure S6

A

log2 ratio of CTCF versus Input at IAP LTR consensus sequences

IAPLTR1_Mm

IAPLTR2b

IAPLTR2_Mm

IAPY5_LTR

IAPA_MM

IAP1-MM_LTR

IAPLTR2a

IAPLTR2a2_Mm

IAPY3C_LTR

IAPY3_LTR

IAPY_LTR

IAPLTR4

IAPLTR3

Distance from 5’ of LTR (bp)

B

CTCF binding site in IAP LTRs

CTCF binding site in constitutive VM-IAP LTRs

CTCF binding site in tissue-specific VM-IAP LTRs
Supplemental Figure S6. CTCF preferentially binds at a subset of IAP LTR types.

(A) CTCF ChIP-seq data mapped to consensus sequences of each type of IAP LTR. CTCF appears to bind at the LTR types left of the dashed line, including the two types for which there is enrichment in cVM-IAPs, and not at other types (right of the dashed line).

(B) Motifs generated from CTCF binding sites located within LTRs of all IAPs, cVM-IAPs, and tsVM-IAPs. CTCF binding sites at these regions were originally identified by mapping the motif generated in Figure 3A to the genome.
Supplemental Figure S7

A

IAP-Marveld2

% CpG Methylation

\[ p = 0.0051 \]
\[ R^2 = 0.2331 \]

\[ 0 \] \[ 200 \] \[ 400 \] \[ 600 \] \[ 800 \]
\[ 0 \] \[ 25 \] \[ 50 \] \[ 75 \] \[ 100 \]

IAP-Tfpi

% CpG Methylation

\[ p < 0.0001 \]
\[ R^2 = 0.5920 \]

\[ 0 \] \[ 200 \] \[ 400 \] \[ 600 \] \[ 800 \]
\[ 0 \] \[ 25 \] \[ 50 \] \[ 75 \] \[ 100 \]

IAP-Rnf157

% CpG Methylation

\[ p = 0.0077 \]
\[ R^2 = 0.2142 \]

\[ 0 \] \[ 200 \] \[ 400 \] \[ 800 \] \[ 1200 \]
\[ 0 \] \[ 25 \] \[ 50 \] \[ 75 \] \[ 100 \]

IAP-Eps8l1

% CpG Methylation

\[ p = 0.0001 \]
\[ R^2 = 0.3963 \]

\[ 0 \] \[ 100 \] \[ 200 \] \[ 300 \] \[ 400 \]
\[ 0 \] \[ 25 \] \[ 50 \] \[ 75 \] \[ 100 \]

IAP-Rab6b

% CpG Methylation

\[ p = 0.0014 \]
\[ R^2 = 0.2918 \]

\[ 0 \] \[ 200 \] \[ 400 \] \[ 600 \] \[ 800 \]
\[ 0 \] \[ 25 \] \[ 50 \] \[ 75 \] \[ 100 \]

IAP-Mbnl1

% CpG Methylation

\[ p = 0.0446 \]
\[ R^2 = 0.1277 \]

\[ 0 \] \[ 100 \] \[ 200 \] \[ 300 \] \[ 400 \]
\[ 0 \] \[ 25 \] \[ 50 \] \[ 75 \] \[ 100 \]

IAP-Pink1

% CpG Methylation

\[ p = 0.1452 \]
\[ R^2 = 0.0939 \]

\[ 0 \] \[ 100 \] \[ 200 \] \[ 300 \] \[ 400 \]
\[ 0 \] \[ 25 \] \[ 50 \] \[ 75 \] \[ 100 \]
### IAP-Tfpi

| Scale chr2: | 84,505,200 | 84,505,250 | 84,505,300 | 84,505,350 | 84,505,400 |
|------------|------------|------------|------------|------------|------------|
| BS Pyro    |            |            |            |            |            |
| CTCF Binding Site | |            |            |            |            |

- **Gm13710**
  - SINE
  - LINE
  - LTR
  - DNA
  - Simple
  - Low Complexity
  - Satellite
  - RNA
  - Other
  - Unknown

>>> IAP-Tfpi >>>

### IAP-Marveld2

| Scale chr13: | 100,594,100 | 100,594,150 | 100,594,200 | 100,594,250 | 100,594,300 | 100,594,350 |
|--------------|------------|------------|------------|------------|------------|------------|
| BS Pyro      |            |            |            |            |            |            |
| CTCF Binding Site | |            |            |            |            |            |
| ChIP-qPCR    |            |            |            |            |            |            |

<<< IAP-Marveld2 <<<
Supplemental Figure S7. Methylation at many cVM-IAPs correlates inversely with CTCF binding.

(A) There is an inverse relationship between CTCF ChIP-seq narrowPeak scores and DNA methylation (measured by bisulfite pyrosequencing) in eight liver samples at six out of seven assayed cVM-IAPs: IAP-Marveld2, IAP-Tfpi, IAP-Rnf157, IAP-Eps8l1, IAP-Rab6b, and IAP-Mbnl1; inverse correlation was not observed for IAP-Pink1. NarrowPeak scores were generating using a p-value of 0.1 to ensure that peaks were called in all individuals in regions with low CTCF occupancy. The correlations are consistent using a q-value of 0.01 (data only shown for IAP-Pink1), but at some loci, peaks were unable to be called for all individuals if using this stricter threshold. Each point represents the average methylation of the three to four distal CpGs in the LTR of an individual. P-values of the correlations are from two-tailed Pearson test. For IAP-Marveld2 and IAP-Tfpi, CTCF enrichment of each individual at the 5′ edge of the IAP is shown in decreasing order of methylation in a screenshot from IGV. The location of the element is indicated by the box at the top of the plot; strandedness is indicated by the chevrons within the box.

(B) Screenshots of the UCSC genome browser showing the span of the CpGs assayed by bisulfite pyrosequencing (BS Pyro), the CTCF binding site, and the ChIP-qPCR assay (primers represented by boxes) relative to the 5′ edges of the cVM-IAPs, IAP-Tfpi and IAP-Marveld2.

(C-D) CTCF ChIP-qPCR fold enrichment versus DNA methylation at several IAPs in the same eight individual liver samples as in (A). Each coloured point represents the average methylation of the four distal CpGs in the LTR of an individual. The colours represent the same individuals across all plots (including those in Fig. 4) and show that the relative order of CTCF binding is not the same in each element. P-values are from two-tailed Pearson test. (C) CTCF ChIP-qPCR fold enrichment calculated relative to IAP-Dst at cVM-IAPs, IAP-Marveld2 and IAP-Tfpi, and
non-variably methylated IAPs, IAP-\textit{Earl} and IAP-\textit{Ell2}. (D) There is no significant correlation between CTCF ChIP-qPCR and DNA methylation at the non-variably methylated IAPs, IAP-\textit{Earl} and IAP-\textit{Dst}. 
Supplemental Figure S8. cVM-IAP methylation levels correlate with H3K9me3 enrichment.

H3K9me3 ChIP-qPCR fold enrichment (relative to H3K9me3 enrichment at non-variable IAP-Asxl3) is correlated with DNA methylation at IAP-Tfpi, IAP-Marveld2, and IAP-Rab6b in adult liver. IAP-Pink1 and IAP-Mbnl1 do not exhibit significant correlations. Each point represents the average methylation level of the three or four distal CpGs in the 5’ LTR of an individual. P-values are from two-tailed Pearson test.
Supplemental Figure S9
IAP-Marveld2 (12.8% - 59.1% methylation range)
**IAP-Pink1 (73.4% - 86.0% methylation range)**

**ChromHMM states**
- Poised Promoter
- Active Promoter
- Transcription
- Transcription Transition
- Weak Transcription
- Transcription Elongation
- Poised Enhancer
- Strong Enhancer
- Insulator
- Repressed
- Heterochromatin
- Promoter
- Transcription
- Enhancer
- Negative

**Genes and Features**
- <Pink1>
- <Mir7019>
- <Gm27646>
- <Fam43b>
- <Cda>
- <Mul1>
- <Ddost>
- <AB041806>
- <Camk2n1>
- Mir6399 > Hsp1bp3 > Kif17 > 700095J12Rik 2310026L22Rik > Eif4g3 > Sh2d5 > <Gm27646> <Fam43b> <Cda> <Mul1> <Ddost> <AB041806> <Camk2n1>
Indiv. 1 (74.6% met hy.)

Indiv. 3 (75.4% met hy.)

Indiv. 2 (81.3% met hy.)

Indiv. 4 (86.4% met hy.)

Indiv. 5 (93.8% met hy.)

chr3

Promoter

Transcription

Enhancer

Negative

ChromHMM states

Active Promoter

Weak Txn

Poised Enhancer

Strong Enhancer

Insulator

Heterochrom

RP23–349H13.2 > RP24–303N8.3 >

Mbnl1 >
IAP-Dst (95.4% - 97.1% methylation range)

Indiv. 1 (95.4% methy.)
Indiv. 2 (96.5% methy.)
Indiv. 3 (96.6% methy.)
Indiv. 4 (97.1% methy.)
Indiv. 5 (96.5% methy.)

chr1

Promoter
Transcription
Enhancer
Negative

ChromHMM states
Poised Promoter
Active Promoter
Weak Txn
Txn Elongation
Strong Enhancer
Insulator
Repressed
Heterochrom

< mmu-mir-5103
< Mir5103
< Imp4
Prss39
< Prss40
< Gm26788
< RP24–422C24.8
< RP24–152H15.5
< Gm28417
< Gm28306
< Gm28419
< Gm28416
< Gm5266
< Gm3052
< 4930568A12Rik
Cfc1
Supplemental Figure S9. Genomic interactions of VM-IAPs.

Above the dashed horizontal line are chromatin interaction plots generated from 4C-seq of the following five regions: two cVM-IAPs with low methylation values (IAP-Marvelsd2 and IAP-Tfec), two cVM-IAPs with high methylation values (IAP-Pink1 and IAP-Mbnl1), and one non-variable highly methylated IAP (IAP-Dst). The interaction plots for each region are shown for five individuals and are organised from highest to lowest methylation; each plot shows a 400kb region centred on the 4C viewpoint. The dashed vertical lines at the centre of the plots denote the viewpoint of the 4C-seq experiment and the grey dots represent fragments that interact with this viewpoint – the height of each dot reflects the intensity of the interaction. The trend line (blue) shows the general interaction profile of the region; the pink ribbon delimits the 95% confidence interval of the trend line. Immediately below the dashed horizontal line are the chromHMM chromatin states of the region (largely derived from histone tail modifications) followed by gene annotations.
B (continued)

| Gene | Methylation | Distance from LTR edge [kbp] |
|------|-------------|-----------------------------|
| New_9 |            |                             |
| P4ha1 |            |                             |
| Pink1 |            |                             |
| Rims2 |            |                             |
| Rsg1  |            |                             |
| Slc15a2 |          |                             |
| Tfec  |            |                             |
| Trbv31 |           |                             |
| Tfp619|            |                             |

Cell type | B | T
---|---|---

C (continued)

IAP-P4ha1

Base pairs from 5' IAP (bp)

IAP-Bex6

Base pairs from 5' or 3' end of IAP (bp)
Supplemental Figure S10. Methylation variability exists beyond the edges of VM-IAPs, but the surrounding methylation landscape varies with each element.

(A) Bisulfite pyrosequencing of liver (5 individuals; IAP-Rab6b, IAP-Marveld2, and IAP-Rnf157) and kidney (4 individuals; IAP-Bex6 and IAP-P4ha1) shows inter-individual methylation variation within 500-1000bp beyond the edge of the cVM-IAP LTRs.

(B) WGBS data from B and T cells (light red and blue) in the 10kb window surrounding the 5′ and 3′ ends of all validated cVM-IAPs. Methylation within the IAP (black line) is not shown. The y-axis shows DNA methylation from 0 to 1.

(C) WGBS of B and T cells (light red and blue) validated by bisulfite pyrosequencing (Pyro) in liver (black, n=5 for all; IAP-Rab6b, IAP-Rnf157, IAP-Eps8l1), kidney (black, n=4 for both; IAP-P4ha1, IAP-Bex6), or B and T cells (red and blue, n=4; IAP-P4ha1) extracted from individual mice at CpGs several kilobases beyond the 5′ and 3′ edges of cVM-IAPs. IAP length is not to scale. For the bisulfite pyrosequencing data, each dot represents a CpG. The dashed vertical lines represent the 5′ or 3′ ends of the IAPs.