Title: Evolution of imprinting via lineage-specific insertion of retroviral promoters

Authors: Aaron B. Bogutz[1*], Julie Brind'Amour[1*], Hisato Kobayashi[2*], Kristoffer N. Jensen[1], Kazuhiko Nakabayashi[3], Hiroo Imai[4], Matthew C. Lorincz[1#], and Louis Lefebvre[1#]

Author affiliations: [1]Department of Medical Genetics, Molecular Epigenetics Group, Life Sciences Institute, University of British Columbia, Vancouver, BC V6T 1Z3, Canada. [2]Department of Embryology, Nara Medical University, Kashihara, Nara 634-8521, Japan. [3]Division of Developmental Genomics, Research Institute, National Center for Child Health and Development, Setagaya, Tokyo, 157-8535, Japan. [4]Molecular Biology Section, Department of Cellular and Molecular Biology, Primate Research Institute, Kyoto University, Inuyama, Aichi, 484-8506, Japan.

* These authors contributed equally to this work.

[#] Authors for correspondence (matthew.lorincz@ubc.ca, louis.lefebvre@ubc.ca)

Summary

Imprinted genes are expressed from a single parental allele. In mammals, this unusual mode of transcription generally depends on the epigenetic silencing of one allele by DNA methylation (DNAme) established in the germline. While many species-specific imprinted orthologues have been documented in eutherians, the molecular mechanisms underlying the evolutionary switch from biallelic to imprinted expression are currently unknown. During mouse oogenesis, gametic differentially methylated regions (gDMRs) acquire DNAme in a process guided by transcription. Here we show that transcription initiating in proximal lineage-specific endogenous retroviruses (ERVs) is likely responsible for DNAme established in oocytes at 4/6 mouse-specific and 17/110 human-specific maternal imprinted gDMRs (igDMRs). The latter can be further divided into Catarrhini (Old World monkeys and apes)- or Hominoidea (ape)-specific igDMRs, which are embedded within transcription units initiating in ERVs specific to these primate lineages. Using CRISPR-Cas9 mutagenesis, we deleted the relevant murine-specific ERVs upstream of the maternally methylated genes *Impact* and *Slc38a4*. Strikingly, imprinting at these genes was lost in the offspring of females harboring these deletions and biallelic expression was observed. Our work reveals a novel evolutionary mechanism whereby maternally silenced genes arise from biallelically expressed progenitors.

During postnatal oocyte growth in mammals, transcribed genomic regions acquire two characteristic epigenetic marks: transcription-coupled trimethylation at lysine 36 of histone H3 (H3K36me3)[1] and DNAme[2-5]. In both human and mouse female germline, the overall levels of CpG methylation increase from less than 5% in post-migratory primordial germ cells (PGCs) to ~40-50% in fully-grown, germinal vesicle oocytes (GVOs)[4,6-8]Kobayashi:2013ia}. In growing murine oocytes, 85-90% of this DNAme is deposited over H3K36me3-marked transcribed regions of the genome[3]. Notably, we recently demonstrated that SETD2-dependent deposition of H3K36me3 is required for *de novo*

51 DNAme of transcribed gene bodies in mouse oocytes[9]. Such transcription-coupled *de novo* DNAme is
52 also responsible for the establishment of regions of differential DNAme between the mature gametes
53 (gametic differentially methylated regions, gDMRs), a subset of which are maintained throughout
54 preimplantation development. This unique class of gDMRs, the imprinted gDMRs (igDMRs), can
55 direct imprinted paternal allele-specific expression of the gene(s) under their regulation, the imprinted
56 genes[3,10,11]. Importantly, transcription initiating within upstream oocyte-specific promoters has been
57 reported to play a critical role in *de novo* DNAme at the maternal igDMRs of a number of mouse
58 imprinted genes[3,12-16], but the evolutionary origin of such promoters remains unexplored.
59
60 Specific families of ERVs, also known as long terminal repeat (LTR) retrotransposons, are
61 highly transcribed in mouse oocytes[3,17]. Such LTRs function as oocyte-specific promoters for both
62 novel protein-coding genes and non-coding transcripts, the latter of which are widely distributed in
63 intergenic regions[3,17,18]. We recently identified numerous LTR-initiated transcription units (LITs) in
64 mouse oocytes associated with H3K36me3-coupled *de novo* DNAme[19]. Notably, CpG islands (CGIs)
65 embedded within such LITs also become hypermethylated by *de novo* DNAme in oocytes[19] and the
66 majority of maternal igDMRs associated with paternally expressed genes in both mice and humans
67 overlap with CGIs[14,20]. Since ERVs are highly variable among mammalian species and *de novo*
68 DNAme at maternal igDMRs is transcription-coupled, we hypothesized that active LTRs and their
69 associated LITs may have played an important role in the genesis of lineage-specific maternal
70 imprinting in mammals.
71
72 **LITs and the establishment of species-specific imprints**
73 To investigate the contribution of LITs in maternal gametic imprinting in humans and mice, we first
74 curated a list that includes well characterized and putative maternally methylated igDMRs[21-27] by
75 interrogating published whole genome bisulfite sequencing (WGBS) data from gametes, placenta and

76 somatic tissues (detailed in **Supplementary Table 1**). In total, we identified 21 mouse and 125 human
77 igDMRs that were previously validated through Sanger bisulfite sequencing, pyrosequencing or whole
78 genome bisulfite sequencing (WGBS, see **Methods** and **Supplementary Table 2**). Amongst these, 6
79 and 110 maternal igDMRs are restricted to the mouse or the human lineage, respectively, while only 15
80 are conserved between the two species (**Fig. 1a**). We next applied *de novo* transcriptome assembly[28] to
81 published human[29] and mouse[19] oocyte RNA-sequencing (RNA-seq) datasets and identified the
82 transcript(s) and transcription start site (TSS) likely responsible for transcription-coupled deposition of
83 DNAme over 20/21 mouse and 90/125 human igDMRs (**Supplementary Table 2** and **Supplementary**
84 **Fig. 1a**; identified using LIONS or *de novo* transcript assembly, see **Methods**)[28,30,31]. Amongst these, 4
85 mouse and 17 human maternal igDMRs, all of which are specific to the mouse or human genome,
86 respectively, are embedded within or immediately downstream of LITs (**Fig. 1a**, **Supplementary Fig.**
87 **1a-b** and **Supplementary Table 2**). Thus, the presence of a lineage-specific proximal LTR that
88 initiates a transcript in oocytes that overlaps with a genic gDMR is associated exclusively with genes
89 showing evidence of a species-specific igDMR.
90
91 Notably, of the 17 LITs associated with human-specific igDMRs, 12 initiate within primate
92 (Hominoidea or Catarrhini)-specific ERV families (**Fig. 1a** and **Supplementary Fig. 1c**). Moreover,
93 the 4 LITs apparently responsible for transcription-coupled *de novo* DNAme of the mouse-specific
94 igDMRs, namely at *retro-Coro1c* (also known as *2010001K21Rik* and *AK008011*), *Cdh15*[26], *Slc38a4*
95 (also known as *Ata3*)[32,33] and *Impact*[34], all initiate in rodent-specific ERVs (**Fig. 1a** and
96 **Supplementary Fig. 1c**). While several such lineage-specific LTR families (i.e. LTR12C/MER51E or
97 MTC/RMER19B) are actively transcribed in oocytes[18,19], others are generally expressed at low levels
98 (**Supplementary Fig. 1d-e**), indicating that igDMR-coupled LITs do not necessarily initiate from LTR
99 families that are widely expressed in oocytes. As hypothesized, the igDMRs associated with lineage-
100 specific LITs also show species-specific hypermethylation in oocytes, which is retained (>35%

101 DNAme) in the blastocyst[8], the placenta[25] or at least one somatic tissue surveyed, indicating that these
102 genomic regions are indeed likely to carry imprinted DNAme marks (**Fig. 1b** and **Supplementary**
103 **Table 2**). In accordance with previous reports of preferential maintenance of maternal germline-
104 derived DNAme in human placenta[21,22,24,25,35], most of the LIT-associated human igDMRs retain
105 DNAme in purified cytotrophoblast (CT; cells isolated from first-trimester human placenta)[25] but are
106 hypomethylated in adult tissues (**Fig. 1b**). The remaining 5/17 human-specific maternal igDMRs are
107 associated with LITs driven by LTR families that colonized the common ancestor of the
108 Euarchontoglires (i.e. rodents and primates). These specific LTR insertions, however, are not detected
109 at the syntenic loci of the mouse genome and only 1/5 of the regions is methylated in mouse oocytes
110 (the region syntenic to the human igDMR at *HTR5A*) which may be due to a distinct non-LTR initiated
111 transcript (**Fig. 1b**, **Supplementary Fig. 1b** and **Supplementary Table 2**).
112
113 This species-specific pattern of transcription initiation from an upstream LTR element leading
114 to transcription-coupled establishment of a DNAme domain that includes the downstream igDMR/CGI
115 can be visualized in a genome-browser view of orthologous regions. Note that we observe both sense
116 and antisense configurations of the relevant LITs, driven by LTRs located either upstream or
117 downstream (within an intron or 3') of the regulated gene, respectively (**Supplementary Fig. 2a**). Of
118 the 17 LITs putatively implicated in the induction of human igDMRs, 12 are in the 5' sense
119 configuration and 5 in an antisense configuration (**Supplementary Table 3**). At the *RHOBTB3/*
120 *Rhobtb3* locus for example, transcription in human oocytes initiates within an unmethylated primate-
121 specific MSTA element located ~25 kb upstream of the promoter CGI/igDMR, forming a chimeric
122 transcript that splices to the downstream genic exons of *RHOBTB3* (**Fig. 1c**). Coincident with this LIT,
123 a large block of DNAme is deposited in oocytes over the promoter CGI and overlapping igDMR.
124 Importantly, these regions are hypomethylated in human female 11-week gonadal PGCs[7] and in sperm.
125 As previously documented for many human igDMRs[21,22,35], this imprint is maintained in the blastocyst

126 and cytotrophoblast[25], but is hypomethylated (< 2% DNAme) in adult tissues (**Fig. 1b-c** and
127 **Supplementary Table 2**). Notably, *RHOBTB3* is expressed predominantly from the paternal allele in
128 human placenta[23,25,36,37]. In contrast, in mouse oocytes, *Rhobtb3* transcription initiates at the promoter
129 CGI, which is unmethylated in oocytes, placenta and adult tissues (**Fig. 1b-c**). Similarly, at the *SCIN*
130 locus, a LIT initiates in an unmethylated LTR12C element ~14 kb upstream of the igDMR in human
131 oocytes and extends into the gene, concomitant with *de novo* DNAme of this region between the PGC
132 and mature oocyte stages (**Fig. 1d**). While the *SCIN* igDMR shows ~50% DNAme in human blastocyst
133 and placenta (CT), the syntenic region in mice, including the *Scin* CGI promoter, is hypomethylated in
134 each of these cell types and no upstream initiating transcript is observed in mouse oocytes (**Fig. 1b, d**).
135 Consistent with the DNAme status of the locus in each species, *SCIN/Scin* is expressed exclusively
136 from the paternal allele in human but not in mouse placenta[23,38]. Importantly, unlike in oocytes, 8 of the
137 9 genes that are associated with igDMRs and expressed in purified human cytotrophoblast (CT) show
138 no evidence of transcription initiating from the proximal LTR in this cell type. Rather, for all 6 genes
139 (*GLIS3*, *MCCC1*, *RHOBTB3*, *ZFP90*, *CLDN23*, *ZC3H12C*) that show clear paternally-biased
140 expression (>70%) in CT (**Supplementary Table 3**), transcription initiates predominantly within the
141 unmethylated igDMR promoter (**Supplementary Fig. 2b, 3**). Thus, imprinted expression of these
142 genes in the placenta is correlated with LIT-associated deposition of DNAme over the igDMR in the
143 oocyte and concomitant silencing of the maternal allele in the extraembryonic trophoblast lineage in the
144 offspring.
145

146 **Conservation of LTRs, LITs and igDMRs in apes versus old world monkeys**

147 Excluding the *SVOPL* igDMR, which is unmethylated in cytotrophoblast, we focused on the 16 human
148 loci with evidence for maintenance of the maternal igDMRs in this placental lineage (**Fig. 1b**).
149 Intriguingly, 5 of these, including *SCIN*, initiate in human ERV (HERV) families which colonized the
150 common ancestor of the Hominoidea[39] (LTR12C or LTR12E), while 7 of them initiate within families

151 that colonized the primate lineage, including the common ancestors of the Catarrhini. The remaining 4
152 initiate in more ancient elements derived from LTR families common to both primates and rodents
153 (Euarchontoglires) (**Fig. 2a** and **Supplementary Fig. 1c)**. To characterize the relationship between
154 LITs and the imprinting status of these gDMRs in non-human primates, we first determined whether
155 the specific LTR insertions associated with these human placental igDMRs are annotated in the
156 genomes of chimpanzees (*Pan troglodytes*; Hominoidea lineage) and rhesus macaques (*Macaca*
157 *mulatta*; Catarrhini lineage). All 16 relevant LTR insertions are present in the chimpanzee genome. As
158 expected, all 4 insertions from LTR families that colonized the Euarchontoglires common ancestor and
159 6/7 LTR insertions from families that colonized the common ancestor of the Catarrhini are also present
160 in the orthologous loci in macaque. Similarly, the Hominoidea-specific LTR12C and LTR12E families,
161 which include LTRs driving transcripts overlapping the *ST8SIA1*, *SCIN*, *ZFP90*, *COL26A1* and *SORD*
162 igDMRs in human oocytes, are absent from the macaque genome (**Fig. 2a** and **Supplementary Fig.**
163 **1c)**.
164
165 To assess the conservation of LITs and DNA methylation of associated gDMRs in macaque
166 oocytes, we analyzed published RNA-sequencing[40] and genome wide DNA methylome data[41]. As
167 expected from the phylogeny of LTR12C elements, no LITs overlapping the region orthologous to the
168 human igDMRs at the *ST8SIA1*, *SCIN*, or *COL26A1* loci were detected in macaque oocytes, and their
169 CGIs remain hypomethylated (**Fig. 2a-c** and **Supplementary Fig. 4a**). Intriguingly, despite the absence
170 of an LTR12E element, the region syntenic to the human *SORD* igDMR exhibits high levels of
171 DNAme in macaque oocytes. Inspection of oocyte transcripts around the macaque *SORD* CGI reveals
172 the presence of a highly transcribed LTR12F element oriented towards the putative igDMR, which may
173 be responsible for deposition of DNAme at this locus (**Supplementary Fig. 4a, b**).
174

175 In contrast, LITs initiating in older LTR families that colonized the common ancestor of the
176 Catarrhini or Euarchontoglire lineages show greater conservation between human and macaque
177 oocytes, with 8/13 of such LTR insertions also transcriptionally active in macaque oocytes, a subset of
178 which display similar splicing events in both species (**Fig. 2b**). Remarkably, at least 6 of the associated
179 CGIs are also hypermethylated in macaque oocytes (**Supplementary Fig. 4a**). For example, *de novo*
180 methylation of the *RHOBTB3* CGI promoter/gDMR in both human and macaque oocytes appears to be
181 the result of a conserved LIT initiating in a MSTA situated >20 kb upstream of the promoter, which
182 clearly splices into the gene in both species (**Fig. 2d**). Other examples of LITs apparently conserved
183 between the human and macaque lineages include those associated with the *HECW1* (LTR12F),
184 *DNAH7* (THE1C), *HTR5A* (LTR53B), *AGBL3* (MER50) and *ZC3H12C* (MLT1A) putative igDMRs.
185 In contrast, no LITs were detected at 4/11 loci despite the presence of a conserved LTR insertion in the
186 orthologous macaque locus **(Supplementary Figs. 4-5)**. For example, while the igDMR at the 5' end
187 of *GLIS3* is embedded within a LIT initiating in an active upstream MSTA in human oocytes, no LIT is
188 detected in the orthologous region in the macaque locus nor is any RNA-seq coverage detected over the
189 MSTA itself (**Supplementary Fig. 5a-b**).
190
191 The lineage-specific expression of orthologous LTRs may be explained by the accumulation
192 over evolutionary time of indels and/or base substitutions that impact their promoter and/or splice
193 donor activity. Indeed, sequence alignment of the orthologous MSTA insertions reveals a number of
194 small deletions and single nucleotide polymorphisms (SNPs) which may impact transcription of the
195 macaque LTR (**Supplementary Fig. 5c**). Similarly, despite the presence of a MER51E upstream of the
196 *MCCC1* gene, no transcript initiating in this LTR is detected in macaque oocytes and no RNA-seq
197 coverage is detected over the element itself (**Supplementary Fig. 5d-e**). Closer inspection of the
198 MER51E insertions in the macaque locus reveals a number of SNPs and short INDELs relative to the
199 orthologous MER51E in chimp and human, including mutations that likely disrupt an otherwise

200 conserved PBX3 binding site that may render the LTR in the former inactive (**Supplementary Fig. 5f**).
201 Taken together, these data indicate that the establishment of DNAme in oocytes at the igDMRs of a
202 number of Hominoid- and primate-specific maternal igDMRs is likely induced by LITs originating in
203 proximal lineage-specific LTR elements actively transcribed during oogenesis.
204
205 **Oocyte LITs predict imprinting in primate placenta**
206 To establish the imprinting status of these gDMRs in chimp and macaque placenta, we determined their
207 methylation using a targeted high-throughput sodium bisulfite sequencing approach. We focused
208 initially on the *ST8SIA1* gene, shown previously to be maternally methylated and paternally expressed
209 in human placenta[23]. The genomic region in chimp syntenic to the human *ST8SIA1* igDMR is
210 hypomethylated in adult tissues (**Supplementary Fig. 4a**), but shows a bimodal distribution of
211 hypermethylated and hypomethylated sequenced reads in the placenta (**Fig. 2e** and **Supplementary**
212 **Fig. 6a**), indicative of conserved placental-specific *ST8SIA1* imprinting within the Hominoidea. In
213 contrast, the syntenic CGI in macaque, as in mouse, is hypomethylated in the placenta (**Fig. 2a, c, e**).
214 Therefore, the presence of an active LTR12C element ~40 kb upstream of the gDMR and its associated
215 LIT are correlated with the imprinting status of the *ST8SIA1* promoter. *SCIN*, which is imprinted and
216 paternally expressed in human placenta[23], also shows a bimodal distribution of DNAme in the
217 orthologous region in chimp placenta. In contrast, in the absence of an LTR12C insertion or a LIT, the
218 orthologous region in macaque placenta is hypomethylated (**Supplementary Fig. 6a**). On the other
219 hand, despite the presence of a proximal LTR12C element in the chimp (as in the human genome), the
220 regions syntenic to the human igDMRs at the *ZFP90* and *COL26A1* genes are unmethylated in chimp
221 placenta (**Fig. 2a**), suggesting that, similar to the LTR insertions proximal to the *GLIS3* and *MCCC1*
222 CGI in macaques, these orthologous LTR may be transcriptionally inert in chimp oocytes. This remains
223 to be determined, however, as RNA-seq data from chimp oocytes is currently not available. The
224 regions syntenic to the human *SORD* igDMR show bimodal DNAme in chimp as well as macaque,

225 consistent with the presence of proximal active LTRs in human and macaque oocytes. While the
226 methylation of the macaque *SORD* locus is likely explained by the alternative LTR12F-initiated
227 transcript described above (**Supplementary Fig. 4b**), it remains to be determined whether an antisense
228 LTR12E-initiated LIT orthologous to that observed in human oocytes is responsible for depositing
229 DNAme over the promoter CGI of this locus in chimpanzees.
230
231 As predicted by the presence of an orthologous LTR12F insertion oriented towards the *HECW1*
232 gene in both the chimp and macaque loci, the genomic region of shared synteny to the human *HECW1*
233 igDMR shows a clear bimodal distribution of hypermethylated and hypomethylated reads in the
234 placentae of both species, indicative of conservation of LIT-associated *HECW1* imprinting in
235 Catarrhines (**Fig. 2e** and **Supplementary Fig. 4c**). Moreover, exploiting a single SNP within the
236 *HECW1* CGI, we observed clear allelic methylation (**Fig. 2f** and **Supplementary Fig. 6b**), confirming
237 maternal allele-specific DNAme at the orthologous igDMR. Similarly, analysis of the chimp and
238 macaque genomes in the regions syntenic to the *RHOBTB3* igDMR, which is embedded within an
239 MSTA-initiated transcript in both human and macaque oocytes (**Fig. 2d**), reveals that the orthologous
240 locus is also likely imprinted in these species (**Fig. 2e**). Furthermore, analysis of DNAme data from
241 informative trios also reveals maternal-specific DNAme of the igDMRs at the *GLIS3* and *MCCC1*
242 genes in the chimp placenta (**Fig. 2e-f** and **Supplementary Fig. 6b**). In agreement with the cognate
243 LTR insertions being transcriptionally inactive in macaque oocytes (**Supplementary Fig. 5**), the
244 genomic regions orthologous to the human and chimp igDMRs at both of these loci are
245 hypomethylated in macaque placentae. Taken together, these observations reveal that transcription
246 initiating within lineage-specific LTR elements in the oocytes of primates likely plays a critical role in
247 the establishment of DNAme at proximal CGIs. The persistence of this oocyte-derived imprint in the
248 placenta of their progeny yields maternal igDMRs that can potentially direct imprinted expression in
249 this extraembryonic tissue, as shown for several human genes.

250

251 **Muroidea-specific LTRs and maternal igDMRs**

252 If the LTRs upstream of the "mouse-specific" igDMRs shown in **Fig. 1** are indeed responsible for the
253 establishment of maternal imprinting at these loci, then the orthologous genes should also be imprinted
254 in those species harboring the same active LTR insertions. As the *retro-Coro1c* retrogene is absent
255 from rats and more distantly related rodents (**Supplementary Fig. 7a**), we focused on the remaining
256 three genes, namely *Cdh15, Slc38a4* and *Impact* and their associated LTRs: MTD, MT2A, and MTC,
257 respectively. All three are present in the orthologous regions of the rat (*Rattus norvegicus*) genome
258 (**Fig. 3a**), consistent with the fact that these LTR families colonized the common ancestor of the
259 *Muroidea* (**Supplementary Fig. 1c, 7**)[18]. In contrast, while the relevant MTD and MT2A elements are
260 also present at the orthologous *Cdh15* and *Slc38a4* loci in the golden hamster (*Mesocricetus auratus)*,
261 the MTC element upstream of *Impact* is absent.

262

263 To assess whether the LITs identified at these imprinted loci in mice are also present in rat and
264 golden hamster, we mined published oocyte RNA-seq data[18,19]. Notably, LITs emanating from the
265 relevant LTRs were clearly detected at *Cdh15* and *Slc38a4* genes in rat oocytes, while in hamster, a
266 LIT was detected only at the *Slc38a4* locus (**Fig. 3a and Supplementary Fig. 7b-d**). In rat, as in
267 mouse[19], the *Slc38a4* LIT overlaps the DNAme block that extends into the 5' end of the gene. In
268 hamster, the MT2A-initiated LIT upstream of *Slc38a4* splices into exon 2 of the gene, creating a
269 chimaeric transcript covering the entire gene, including the region syntenic to the mouse igDMR
270 (**Supplementary Fig. 7c**). In contrast, in human oocytes, the *SLC38A4* locus is not transcribed and the
271 5' CGI is unmethylated (**Fig. 3b**). In the case of *Impact*, a LIT initiating in a co-oriented upstream
272 MTC element encompasses the entire annotated gene in mouse and demarcates a hypermethylated
273 domain, including the igDMR (**Fig. 3b**). Intriguingly, while the syntenic *Impact* CGI is
274 hypermethylated in rat oocytes, the orthologous rat MTC insertion is apparently oriented on the

275 opposite strand relative to the *Impact* gene (**Supplementary Fig. 7d**)[42]. Closer analysis of the locus
276 reveals that a transcript encompassing the *Impact* CGI may originate from a distinct (non-LTR)
277 upstream start site in rat oocytes, but a gap in the reference genomic sequence precludes detailed
278 analysis of the upstream transcript at this locus (**Supplementary Fig. 7d**). Regardless, as in the mouse,
279 *Impact* is expressed from the paternal allele in rat brain[42], consistent with epigenetic imprinting of its
280 CGI and imprinted expression. Given the probable conservation of imprinting at *Slc38a4* and *Impact* in
281 the Muroidea lineage, and the location of the intragenic igDMRs near their 5' ends, we focused on
282 these genes for further functional experiments in the mouse.
283
284 **An MT2A is required for *Slc38a4* imprinting**
285 In mouse oocytes, transcription over the *Slc38a4* igDMR originates in an H3K4me3 marked MT2A
286 LTR located ~14 kb upstream of the gene, in a sense configuration. The 5' end of the locus is
287 embedded within the LIT, enriched for H3K36me3, and overlaps with a *de novo* DNAme block that
288 encompasses the igDMR (**Fig. 3b**). In addition, our ChIP-seq analysis of RNA pol II in mouse GVOs
289 reveals enrichment consistent with transcription extending from the MT2A through intron 1 of *Slc38a4*
290 (**Fig. 3b**). Notably, MT2A elements are not generally transcribed at high levels in mouse oocytes, and
291 analysis of all insertions > 450 bp (1,458 in total) indicates that only 16 initiate transcripts in GVOs at
292 detectable levels (**Fig. 3c** and **Supplementary Fig. 1e**). As the MT2A upstream of *Slc38a4* has
293 significantly diverged from the consensus sequence of this relatively old LTR family (**Fig. 3d**) and this
294 LTR is not closely related to the few other MT2A elements active in oocytes (**Supplementary Fig. 8**),
295 its transcriptional activity in the female germline may be due to the acquisition of novel transcription
296 factor binding sites.
297
298 To test directly the role of the upstream MT2A insertion in *Slc38a4* imprinting, we generated a
299 638-bp deletion of the MT2A LTR in C57BL/6N (B6) embryonic stem cells using CRISPR-Cas9 and

300 flanking guide RNAs (**Fig. 4a**, Δ; **Supplementary Fig. 9a**). Germline transmission from male
301 chimeras allowed us to establish the *Slc38a4*^{MT2AKO} mutant mouse line on a pure B6 background.
302 Heterozygous and homozygous animals were normal and fertile under standard husbandry conditions.
303 In reciprocal crosses involving a wild-type and a heterozygous parent, heterozygotes were recovered at
304 the expected frequency (**Supplementary Table 5**). We first assessed the effect of the MT2A deletion
305 on DNAme at the *Slc38a4* igDMR in fully-grown oocytes using sodium bisulfite sequencing of 11
306 CpG sites within the igDMR located at the beginning of intron 1 (**Fig. 4a**). Notably, whereas these sites
307 are hypermethylated in wild-type oocytes, they remain unmethylated in oocytes from females
308 homozygous for the MT2A deletion (**Fig. 4b**). To test for allele-specific transcription, we generated F1
309 progeny from reciprocal crosses with CAST/EiJ (CAST) mice, a hybrid strain background in which
310 *Slc38a4* was previously shown to be imprinted[33]. Sodium bisulfite analysis confirmed that the oocyte-
311 derived DNAme mark is absent in F1 embryos (E13.5) upon maternal transmission of the
312 *Slc38a4*^{MT2AKO} allele (**Fig. 4c**). Consistent with the absence of maternal DNAme, *Slc38a4* is expressed
313 from both alleles in E13.5 placenta when the deletion is maternally inherited, while monoallelic
314 imprinted expression from the paternal allele is maintained in paternal heterozygotes (**Fig. 4d-f**).
315 *Slc38a4* codes for an amino acid transporter (ATA3) highly expressed in the mouse placenta[33],
316 suggesting that variations in its overall expression levels might lead to placental and/or growth
317 abnormalities, as described for other imprinted genes. However, we did not observe either phenotype in
318 maternal heterozygotes (**Supplementary Fig. 9c-f**), despite the fact that *Slc38a4* mRNA levels increase
319 ~5-fold from E8.5 to E18.5 in wild-type placenta (**Fig. 4g**). Consistent with the absence of placental or
320 embryonic growth abnormalities, we found that loss of imprinting and biallelic expression at *Slc38a4* is
321 not accompanied by an increase in total mRNA levels in maternal heterozygotes (**Fig. 4g**). These
322 results, in accord with previously published findings[43], suggest that *Slc38a4* levels are normalized by
323 transcriptional or post-transcriptional mechanisms in the placenta. Nevertheless, the possibility that a
324 decrease in *Slc38a4* mRNA would bring its levels within the dosage-sensitive zone in which imprinting

325 of the locus might provide a selective advantage is supported by the abnormal embryonic growth
326 phenotype observed in *Slc38a4*-null conceptuses (Gavin Kelsey and Miguel Constancia, personal
327 communication).
328
329 **An upstream MTC directs imprinting at *Impact***
330 In mouse oocytes, the *Impact* gene is transcribed from an unmethylated upstream MTC element in the
331 sense configuration. This LIT splices onto canonical exon 2 of *Impact*, as supported by RNA-seq
332 data[3,19] and a 5' RACE product from GVOs[14] (**Fig. 3b**). As seen at the *Slc38a4* locus, the MTC itself is
333 marked by H3K4me3[44], and RNA pol II, H3K36me3 and DNAme are enriched over the transcribed
334 region, which encompasses the entire *Impact* locus, including its igDMR and intronic CGI. Using a
335 CRISPR-Cas9 mutagenesis approach similar to the one described for the MT2A at *Slc38a4*, we
336 generated a mouse line in which a ~3 kb region upstream of the *Impact* gene, including the full-length
337 MTC element, was deleted to generate the $Impact^{MTCKO}$ allele (**Fig. 5a, Δ; Supplementary Fig. 9b**).
338 Following germline transmission through male chimeras, heterozygous males and females were bred to
339 wild-type B6 mice and heterozygotes were recovered at the expected Mendelian ratios
340 (**Supplementary Table 5**). Maternal and paternal heterozygotes, as well as homozygous mice of both
341 sexes, appear normal and are fertile. As we observed for the MT2A deletion at *Slc38a4*, the igDMR at
342 *Impact* fails to acquire *de novo* DNAme in mature oocytes from $Impact^{MTCKO/MTCKO}$ homozygous
343 females (**Fig. 5b**).
344
345 To study the effect of the MTC deletion on imprinting of the downstream *Impact* gene,
346 reciprocal crosses were performed between $Impact^{+/MTCKO}$ and CAST mice and embryos were collected
347 at E13.5. DNAme analysis of the *Impact* igDMR confirmed that DNAme is absent in embryos carrying
348 a maternally-inherited KO allele ($Impact^{MTCKO/+}$, **Fig. 5c**). Futhermore, allele-specific RT-PCR of head
349 cDNA generated from the same E13.5 F1 progeny revealed that *Impact* is paternally expressed in all

350 samples analyzed, with the exception of embryos in which the MTCKO allele is maternally inherited,
351 in which *Impact* is expressed from both parental alleles (**Fig. 5d-f**). Consistent with biallelic expression
352 in these *Impact*$^{MTCKO/C+}$ embryos, quantitative analysis of total expression levels reveals that *Impact*
353 mRNA levels are nearly doubled relative to those measured in wild-type controls (**Fig. 5g**). To
354 determine whether this loss of imprinting at *Impact* influences the growth of mutant mice, we
355 measured the weights of males and female wild-type and *Impact*$^{MTCKO/+}$progeny. No significant weight
356 difference was detected through postnatal week 60 (**Supplementary Fig. 9g)**, indicating that under
357 standard husbandry conditions, increased *Impact* levels do not affect intrauterine or postnatal growth.
358 Regardless, our analyses clearly reveal that DNAme and expression imprinting of the *Impact* gene in
359 mice is dependent upon the presence of the upstream MTC element.
360

361 **Discussion**

362 Genomic imprinting is an epigenetic mechanism required for normal development and postnatal
363 survival in mammals. Although more than a hundred imprinted genes have been identified and much
364 has been learned about the epigenetic mechanisms regulating their monoallelic expression, it is still
365 unknown how an ancestral gene, expressed from both alleles, acquires an imprinted expression pattern
366 during evolution. Previous work has shown that through retrotransposition new retrogenes can acquire
367 a maternal igDMR and show paternal allele-specific expression when inserted within a host gene
368 expressed during oogenesis[45]. Although a number of such examples have been described in
369 mammals[46], this phenomenon involves the creation of a new gene. Thus, it does not provide a model
370 for the evolutionary switch from biallelic to imprinted expression manifest at species-specific imprinted
371 genes, as explored in this study.
372

373 Here, we uncovered a novel mechanism whereby lineage-specific insertions of LTR
374 retrotransposons, transcriptionally active during oocyte growth, can drive *de novo* DNAme and

375 imprinting at a nearby gene. CRISPR-Cas9 mutagenesis at two mouse-specific imprinted genes, *Impact*
376 and *Slc38a4*, confirmed the key role played by LTR-promoted transcription in guiding imprinting at
377 these two mouse loci. Interestingly, *Slc38a4* was previously reported to be imprinted in mouse
378 preimplantation embryos and early extraembryonic tissues in an H3K27me3-dependent manner[47].
379 However, the broad intragenic domain of H3K27me3 observed in oocytes and on the maternal *Slc38a4*
380 allele in early embryos[47] is mutually exclusive of the upstream H3K36me3- and DNAme-marked
381 domain described here (**Supplementary Fig. 10)**. This observation is consistent with a previous report
382 showing that H3K36me3 inhibits PRC2 activity[48]. Thus, as with canonical imprinted genes, the
383 *Slc38a4* igDMR in mice, which overlaps with an annotated TSS of this gene, is enriched for DNAme
384 but devoid of H3K27me3. In addition to our results on the MT2A deletion allele, the importance of the
385 LIT-induced igDMR for allelic usage is also supported by the observation of imprinted expression at
386 *Slc38a4* in embryonic and adult epiblast-derived tissues in which the igDMR, but not the intragenic
387 H3K27me3 domain, is present (**Supplementary Fig. 10**)[33,49]. The paradoxical critical role of two
388 different maternal epigenetic marks laid down in two different regions of the *Slc38a4* locus may be
389 reconciled by the fact that an alternative annotated TSS of the *Slc38a4* gene is located downstream of
390 the igDMR and is embedded within the intragenic region enriched for H3K27me3 described by Inoue
391 and colleagues[47]. This alternative TSS is used in adult liver, the tissue showing the highest *Slc38a4*
392 expression levels after placenta (**Supplementary Fig. 10**). Although the igDMR is present in adult
393 liver, *Slc38a4* expression is biallelic in this tissue, suggesting that tissue-specific loss of imprinted
394 expression occurs via promoter switching[33]. Thus, depending on the tissue, imprinted expression of this
395 gene likely requires one or the other epigenetic mark. Further studies are required to address this
396 intriguing possibility and to determine whether establishment or maintenance of the igDMR at *Slc38a4*
397 requires maternal H3K27me3. Notably, although preimplantation maintenance of the maternal DNAme
398 imprint at *Impact* requires the protective action of the KRAB-zinc finger proteins ZFP57 and ZFP445,
399 the igDMR at *Slc38a4* does not[50]. This observation is consistent with a report showing that DNAme at

400 the *Slc38a4* igDMR in postimplantation embryos is sensitive to loss of the H3K9 methyltransferase
401 EHMT2/G9A[51].
402
403 Our observation that 15.5% (17/110) of human-specific and 66.7% (4/6) of mouse-specific
404 maternal igDMRs are potentially induced by ERV promoters highlights the importance of this
405 mechanism and its impact on the evolution of species-specific imprinted genes in mammalian species.
406 Furthermore, our analysis also revealed evidence for convergent evolution of imprinting, where a
407 distinct LTR may be responsible for imprinting in different species (LTR12E/F at the *SORD* locus in
408 human/macaque for example). Further studies, such as those applied here in the mouse, will be required
409 to determine whether alternative LTRs play a role in the imprinting of such loci in other species. Those
410 species-specific maternal igDMRs not associated with a LIT are likely methylated as a consequence of
411 transcription initiating in novel, single-copy, TSSs active in oocytes (as seen at the rat *Impact* locus).
412
413 While disruption of imprinting at some maternally silenced genes is associated with severe
414 phenotypic outcomes, consistent with our observations at *Slc38a4* and *Impact*, loss of imprinting and
415 biallelic expression of imprinted genes is not necessarily associated with obvious abnormal phenotypes.
416 Previous studies of the paternally expressed genes *Plagl1*[3], *Peg3*[16], and *Zrsr1*[15] for example, revealed
417 only subtle impacts on reproductive fitness following loss of imprinting, although detailed analyses of
418 potential growth-related phenotypes have not always been reported.
419
420 Although the role of active transcription in guiding DNAme of the oocyte genome, including at
421 maternal igDMRs, is well established, the molecular mechanism involved still remains to be elucidated.
422 We recently found that transcription-coupled deposition of H3K36me3 by SETD2 plays a critical role
423 in this process, as DNAme at all maternal igDMRs, including at *Impact* and *Slc38a4*, is lost in mouse
424 oocytes in which *Setd2* is deleted[9]. Thus, transcription-coupled H3K36me3 deposition is likely the

425 critical common feature for the establishment of DNAme at maternal igDMRs in mammalian oocytes,
426 including at igDMRs embedded within lineage-specific active LTR-initiated transcription units and the
427 novel igDMRs created via retrotransposition within transcribed genes.
428
429 A link between genomic imprinting and repetitive elements was explored previously by several
430 authors, who proposed that the function of DNAme as a host defense mechanism might have been co-
431 opted for allelic silencing at imprinted genes[52-55]. However, this hypothesis is inconsistent with our
432 current understanding of *de novo* DNAme in the female germline, a process that is intimately
433 associated with transcribed regions, which include imprinted gDMRs. Indeed, the mechanism we
434 uncover here explains the establishment of DNAme imprints at single-copy CpG-rich regions that are
435 unremarkable with respect to their repetitive element content. Rather, deposition of DNAme is
436 dependent upon the transcriptional activity of nearby ERVs that evade silencing, a key aspect
437 mechanistically opposite to the one central to the host defense hypothesis. Furthermore, once the
438 specific requirements for post-fertilization imprint maintenance are met, active ERVs, which propagate
439 via retrotransposition, can theoretically induce imprinting of any downstream gene, with the
440 prerequisite that the LTR be active in growing oocytes, when DNAme is established *de novo* by the
441 DNMT3A/3L complex.
442
443

444 **Methods**

445

446 **Data reporting.** No statistical methods were used to predetermine sample size. The experiments were
447 not randomized and the investigators were not blinded to allocation during experiments and outcome
448 assessment.

449

450 **Validation of previously identified mouse and human maternally methylated igDMRs.** To validate
451 previously identified maternally methylated imprinted gDMRs (igDMRs) [21-24], we interrogated
452 DNAme profiles from published whole genome bisulfite sequencing data from gametes, placenta and
453 somatic tissues (**Supplementary Table 1**) [56]. Specifically, we identified regions that are
454 hypermethylated (> 70% DNAme) in oocytes, hypomethylated (< 30% DNAme) in sperm, retain
455 DNAme (>25%) in the blastocyst, and show 35-65% DNAme in placenta (purified first-trimester
456 cytotrophoblast; CT) or at least one adult somatic tissue[4,5,7,8,18,19,29,40,41,44,47,49,57-69]. We further validated
457 and refined this list by only including gDMRs harboring either reported maternal, monoallelic or
458 bimodal DNAme in the placenta and/or at least one somatic tissue. A "bimodal" DNAme pattern is a
459 specific class of sequences methylated at ~50% for which individual DNA sequencing strands from
460 WGBS datasets, although not overlapping a SNP with known parental origin, are either fully- (hyper)
461 methylated or un- (hypo-) methylated. If the DNA sequence analyzed also encompasses a SNP, the
462 bimodal pattern can be described as "allelic" when all hypo- or all hyper-methylated strands contain the
463 same variant at the SNP: only one of the parental alleles is methylated. Only when the parental origin
464 of the sequence variants of this SNP is known in a given sample, can the allelic DMR be referred to as
465 imprinted (maternally or paternally methylated). In total, 125 human igDMRs (46 of which are
466 associated with reported paternal-biased transcription of the nearby gene) and 21 mouse igDMRs (all
467 with reported paternal-biased transcription) are presented and analyzed in **Supplementary Table 2**.
468 Syntenic regions between the mouse (mm10) and human (hg19), as well as chimpanzee (panTro4),

469 macaque (rheMac8), rat (rn6) and golden hamster (mesAur1) genomes were obtained using the
470 Liftover tool from the UCSC Genome Browser (http://genome.ucsc.edu/){Haeussler:2019jw}.
471
472 **Identification of LTR transcripts overlapping igDMRs.**
473 Oocyte RNA-seq libraries (see **Supplementary Table 1**) were aligned to the mm10 (mouse), hg19
474 (human), rhemac8 (macaque), mesAur1 (golden hamster) and rn6 (rat) assemblies using STAR
475 v.2.4.0.i42[70]. *De novo* transcriptome assemblies were produced using Stringtie v.1.3.5 (mouse, human,
476 golden hamster and rat) [31] or Cufflinks v.2.1.121 (macaque) [28] with default parameters. All *de novo*
477 transcripts overlapping or in proximity of putative imprinted gDMRs were identified, and their
478 transcription initiation site was confirmed by visual inspection of splice sites. Transcripts initiating in
479 transposable elements were identified as described previously using LIONS [19,30] The 5' ends of *de novo*
480 assembled transcripts were classified based on overlaps with UCSC Repeat Masker annotation and our
481 putative igDMR list, and transcripts with the Up or UpEdge classification (per LIONS raw output)
482 were taken into consideration. For manual inspection of LITs over specific CGI promoters, LITs were
483 either identified by manually validating transcripts with EInside transposable element contribution in
484 the LIONS raw output or by intersecting *de novo* transcripts with the boundaries of hypermethylated
485 domains.
486
487 **Evolutionary tree of LTR families.** Phylogenetic tree of species shown was generated from TimeTree
488 (http://timetreebeta.igem.temple.edu/). Integration of LTR families was imputed based on the presence
489 or absence of the family in the species investigated, and colonization time is indicated at the root of
490 those species in which it was found in common. For examining the evolutionary history of MT2A
491 family, including the MT2A apparently responsible for *de novo* methylation of the *Slc38a4* igDMR in
492 mouse oocytes, we identified all annotated (UCSC RepeatMasker) MT2A elements > 450 bp (the
493 MT2A consensus sequence is 533 bp). Multiple sequence analysis and phylogenetic tree construction

494 were carried out with MEGA X[71]using MUSCLE UPMGA algorithm and by Neighbour-Joining
495 method, respectively. Visualization and annotation of the tree were carried out with iTOL
496 (https://itol.embl.de/)[72].
497
498 **Ethical approval for animal work.** All mouse experiments were performed under certificates A15-
499 0291, and A15-0181 from the UBC Animal Care Committee and complied with the national Canadian
500 Council on Animal Care guidelines for the ethical care and use of experimental animals. All primate
501 experiments were approved by the animal experiment committee of Primate Research Institute (PRI) of
502 Kyoto University (Approval No: 2018-004). Three rhesus macaque (Macaca mulatta) placentas and
503 their parental blood DNA were collected and stored at -80°C before use. Three chimpanzee (Pan
504 troglodytes verus) placentas and their parental blood DNA were provided from Kumamoto Zoo and
505 Kumamoto Sanctuary via Great Ape Information Network (GAIN).
506
507 **SNP genotyping and targeted bisulfite analysis in primates.** Regions orthologous to human
508 placental igDMR were PCR-amplified with TaKaRa EX Taq HS (TaKaRa Bio) with primers specific
509 for chimpanzee and rhesus macaque CGIs (**Supplementary Table 4**). PCR were performed under the
510 following conditions: 1 min at 94 °C, then 40 cycles of 30 sec at 94 °C, 30 sec at 57 °C, 30 sec at 72
511 °C, and 5 min final elongation at 72 °C. PCR products were purified with QIAquick PCR Purification
512 Kit (Qiagen). Sanger sequencing was carried out for each amplicon using the BigDye Terminator v3.1
513 Cycle Sequencing Kit and the ABI 3130xl DNA Analyzer (Thermo Fisher Scientific). The sequencing
514 data were assembled using ATGC sequence assembly software (bundled with GENETYX Ver. 14,
515 Genetyx Corporation) to identify SNPs at a given loci and determine the genotypes for each individual
516 samples (**Supplementary Table 4**). Genomic DNA (100 ng per sample) was treated with sodium
517 bisulfite using the DNA Methylation Gold Kit (Zymo Research). The bisulfite-treated DNA was PCR-
518 amplified with TaKaRa EpiTaq HS (TaKaRa Bio) with primers specific for chimpanzee and rhesus

519 macaque CGIs (**Supplementary Table 4**). PCR were performed under the following conditions: 1 min
520 at 94 °C, then 35-37 cycles of 30 sec at 94 °C, 30 sec at 57 °C, 30 sec at 72 °C, and 5 min final
521 elongation at 72 °C. PCR products were purified with QIAquick PCR Purification Kit (Qiagen).
522 Amplicons were tagged using NEB Next Ultra II DNA Library Prep Kit for Illumina (NEB) with 5
523 thermal cycles for amplification and subjected to paired-end sequencing on a MiSeq platform using
524 MiSeq Reagent Nano Kit v2, 300 Cycles (Illumina). We used the QUMA website
525 (http://quma.cdb.riken.jp/top/index.html) to quantify CpG methylation levels of each CGI. From each
526 sequence data (fastq files), the top 400 forward reads (R1) were extracted and mapped to each CGI
527 with QUMA's default conditions and uniquely mapped reads were used to calculate DNA methylation
528 level per CpG dinucleotide site. Due to mappability issues, 20 mismatches were allowed for mapping
529 the amplicons over the chimpanzee *ZNF396* gDMR locus. When SNPs were available, reads perfectly
530 matching each parental genome were extracted and aligned separately for allele-specific analyses.
531
532 **ChIP-sequencing in mouse oocytes.** Germinal vesicle oocytes (GVOs) were isolated from 7-10
533 weeks-old C57BL/6J females. The zona pellucida was dissolved by passaging the oocytes through an
534 acid Tyrode's solution, followed by neutralization in M2 media. The oocytes were re-suspended in
535 nuclear isolation buffer (Sigma), flash-frozen in liquid nitrogen and stored at -80°C until usage. RNA
536 polymerase II ChIP-seq libraries were prepared from ~200 GVOs using a modified version of ULI-
537 NChIP-seq{Brindamour:2015kr} . Briefly, chromatin was fragmented with MNase (NEB), diluted in
538 native ChIP buffer (20 mM Tris-HCl pH 8.0; 2 mM EDTA; 150 mM NaCl; 0.1% Triton X-100)
539 containing 1 mM PMSF (Sigma) and EDTA-free protease inhibitor cocktail (Roche), and incubated
540 overnight at 4°C with 0.1 µg of anti-RNA polymerase II monoclonal antibodies (Abcam ab817 -
541 8WG16-) and 5 µl of protein A: protein G 1:1 Dynabeads (Thermofisher). Antibody-bond chromatin
542 was washed four times in low salt buffer (20 mM Tris-HCl pH 8.0; 2 mM EDTA; 150 mM NaCl; 1%
543 Triton X-100; 0.1% SDS) and eluted at 65°C for 1 hour in elution buffer (0.1 nM NaCO3; 1% SDS).

544 DNA was then extracted using phenol:chloroform and precipitated in 75% ethanol, followed by paired-
545 end library construction. Libraries were sequenced (75 bp paired-end) on a NextSeq 500 according to
546 manufacturer's protocols.
547
548 **ERV deletions using CRISPR-Cas9.** The ERVs at *Slc38a4* and *Impact* were deleted in mouse ESCs
549 using transient delivery of sgRNAs and Cas9 from an expression vector similar to pX330-U6-
550 Chimeric_BB-CBh-hSpCas9 (Addgene #42230) [73]. Plasmid phU6-gRNA-CBh-hCas9 was obtained
551 from Rupesh Amin and Mark Groudine and first modified by insertion of the PGK-puro-pA selectable
552 marker from pPGKpuro (Addgene #11349) [74] to obtain pPuro2-hU6-gRNA-CBh-Cas9. Individual
553 sgRNAs were designed from http://crispr.mit.edu/ and incorporated on a single forward oligonucleotide
554 (IDT) including a 5' SapI site and a 29-bp region of complementarity to a universal reverse primer
555 containing an XbaI site. All primer sequences are shown in **Supplementary Table 4**. For each
556 sgRNA, 10 pmoles of forward and universal reverse primers were annealed and extended with Q5
557 high-fidelity DNA polymerase (NEB) according to the recommended conditions and with the following
558 program: 3 min/96°C, 3 cycles of 3 sec/96°C-30 sec/50°C-3 min/72°C, and final extension for 5
559 min/72°C. The PCR reaction was cleaned with the QIAquick PCR Purification Kit (QIAGEN), and
560 cloned as a ~110-bp SapI-XbaI fragment into pPuro2-hU6-gRNA-CBh-Cas9. For each deletion, we
561 designed 4 different sgRNAs, 2 targeting each side of the ERV (Supplementary Fig. 9), and tested the
562 efficiency of each individual guide using an endonuclease assay. For these assays, 1 µg of each
563 sgRNA plasmid were transfected in C57BL/6N C2 ESCs[75] by lipofection (Lipofectamine 2000,
564 ThermoFisher). After 16 hours, puromycin selection (4 µg/ml) was started and continued for 48 hours,
565 after which the cells were grown without selection for 5 days. Whole cell populations were collected
566 and prepared for PCR by HotSHOT[76]. Genomic PCR was performed with primers flanking the sgRNA
567 site (**Supplementary Table 4**) and purified amplicons were melted, reannealed, and then digested
568 using T4 Endonuclease I (NEB). Cut/uncut ratios were calculated following agarose gel electrophoresis

569 and the most efficient sgRNA pairs were chosen for subsequent use. For each deletion, sgRNA plasmid
570 pairs (total of ~ 10 μg) were electroporated into C2 cells as previously described[75]. After 48h of
571 puromycin selection, cells were grown in ESM for 7 days. Single colonies were picked and expanded,
572 then screened for full-length LTR deletion by PCR, Sanger sequencing, and chromosome contents.
573 Euploid lines containing full-length deletion were selected for aggregation chimera production.
574
575 **Aggregation chimera production**. Blastocysts were collected at E3.5 from female albino C57Bl/6J-
576 TyrC2J mice following natural mating, and placed in a 200ul drop of KSOM overlaid with embryo-
577 tested mineral oil and incubated at 37°C until ready for injection. Pooled clones of *Slc38a4* LTR KO
578 (#1, #2 and #10) or *Impact* LTR KO (#3 and #5) ESCs were suspended in injection media (DMEM
579 with HEPS + 25% KO ES media without LIF/2i), and blastocysts were injected with 8-10 ES cells per
580 embryo. After injection, 10-20 embryos were implanted into E2.5 pseudo pregnant mice via uterine
581 transfer. In total, 28 and 7 chimeric pups were born for the *Slc38a4*^{MT2AKO} and *Impact*^{MTCKO} mutant
582 ESCs, respectively. From these, at least three chimeric males per line transmitted a deletion allele.
583 Transgenic mice were bred and maintained in the Centre for Disease Modelling, Life Sciences Institute,
584 University of British Columbia, under pathogen-free conditions. In all heterozygous genotypes, the
585 maternally-inherited allele is always presented first.
586
587 **Allele-specific expression analysis.** E13.5 embryos and placentae were dissected from reciprocal F1
588 crosses between heterozygous LTR KO animals on the C57BL/6 background and WT CAST mice.
589 Embryonic heads for the *Impact* MTCKO and placentae for the *Slc38a4* MT2AKO crosses were used
590 for RNA extraction using Trizol (ThermoFisher). RNA was DNase I-treated (Fisher) at 37 °C for 1 h,
591 then heat inactivated at 65 °C for 15 min. cDNA was synthesized with MMLV-RT (ThermoFisher)
592 using N15 oligonucleotides. RT-PCR was performed with primers described in Supplementary Table 4,
593 and purified PCR products were subjected to restriction enzyme digestion (PvuII for *Slc38a4* or MluCI

594 for *Impact*) and resolved on a 2% agarose gel. Uncut PCR products were sent for Sanger sequencing,
595 and .ab1 files were analyzed using the PHRED software
596 (http://www.phrap.org/phredphrapconsed.html).
597
598 **Bisulfite sequencing.** E13.5 embryos and placentae were dissected from reciprocal F1 crosses between
599 heterozygous LTR KO animals on the C57BL/6 background and wild-type CAST/EiJ mice obtained
600 from the Jackson Laboratory (stock number 000735). Bodies were minced and gDNA was extracted by
601 proteinase K digestion followed by Phenol:Chloroform extraction. 100ng of gDNA was vortexed to
602 shear the DNA, denatured in 0.2M NaOH, bisulfite converted in 0.225M NaOH; 0.0125%
603 hydroquinone; 4M $NaHSO_3$ overnight at 50°C, cleaned using the Wizard DNA cleanup kit (Promega)
604 and desulfonated in 0.3M NaOH at 37°C for 30 minutes. Oocytes were harvested from females of the
605 appropriate genotype at 21-28 days of age. GVOs were isolated by passing dissected ovaries through a
606 100µm filter using a blunt tool, applying the filtrate to a 35µm filter, then backwashing the trapped
607 GVOs from the filter. GVOs were further purified by manual collection. >200 Oocytes were pooled
608 from multiple females (typically 4-5 females total). Oocyte DNA was harvested by incubation in 0.1%
609 SDS 1 µg/µl Proteinase K in the presence of 1µg lambda DNA at 37°C for 60 minutes followed by
610 98°C for 15 minutes. DNA was converted using EZ DNA Methylation Gold Kit (Zymo) per
611 manufacturer's protocols. 3 biological replicates for oocytes and 3 independent SBS conversions for
612 embryos were amplified using semi-nested primers, using touchdown-PCR conditions. Purified PCR
613 products were TA-cloned into pGEM-T (Promega), and sequenced at the McGill/Genome Quebec
614 Innovation Centre. Sequences were analyzed using BiQ Software (https://biq-analyzer.bioinf.mpi-
615 inf.mpg.de/). Informative SNPs were identified in final sequences and used to identify individual
616 C57BL/6 or CAST strands.
617

618 **Quantitative RT-PCR.** To measure total levels of expression, quantitative PCR (RT-qPCR) was
619 performed on cDNA (described above) using Eva Green (Biotium) on a Step-One Plus Real time PCR
620 system (Applied Biosystems). All reactions were run as follows: 2 min at 95 °C, then 40 cycles of 30
621 sec at 95 °C, 30 sec at 60 °C, 30 sec at 72 °C, and fluorescence was read at 80 °C. Ct values of six
622 biological replicates from two litters for each stage were averaged and used to calculate relative
623 amounts of transcripts, normalized to levels of the housekeeping gene *Ppia*. Primer sequences are
624 available in **Supplementary Table 4**.
625
626 **Data Availability.** Data generated for this manuscript were deposited at GEO datasets under the
627 accession GSE126363. Tracks for oocyte data analyzed here can be accessed through the data hub
628 https://datahub-d85hei26.udes.genap.ca/NatComm2019/hub.txt. Datasets analyzed for this manuscript
629 are detailed in **Supplementary Table 1**.
630
631 **Acknowledgements**
632 We thank Rupesh Amin and Mark Groudine for their CRISPR-Cas9 vector, Hiromi Kamura, Prof.
633 Kenichiro Hata (National Center for Child Health and Development), Tomoya Takashima, Dr. Keisuke
634 Tanaka, and Prof. Tomohiro Kono (Tokyo University of Agriculture) for their technical assistance with
635 targeted bisulfite sequencing analysis, Julien Richard Albert and Amanda Ha for technical assistance,
636 and Dixie Mager and Carolyn Brown for helpful discussions. M.C.L. is supported by CIHR Grant
637 MOP-133417 and NSERC Discovery Grant 2015-05228. L.L is supported by CIHR Grant MOP-
638 119357 and NSERC Discovery Grant 386979-12 H.K. is supported by MEXT KAKENHI Grant
639 Numbers 15H05579 and JP18H05553. H.I is supported by MEXT KAKENHI Grant Numbers
640 15H05242 and 18H04005. A portion of primate samples were provided through the Great Ape
641 Information Network (supported by The Cooperative Research Programs of the Genome Research for
642 BioResource, NODAI Genome Research Center, Tokyo University of Agriculture, and Primate

643 Research Institute and Wildlife Research Center, Kyoto University). A portion of primate samples were
644 provided through the Great Ape Information Network. H.K. and H.I. are supported by The Cooperative
645 Research Programs of the Genome Research for BioResource, NODAI Genome Research Center,
646 Tokyo University of Agriculture, and Primate Research Institute and Wildlife Research Center, Kyoto
647 University.
648

649 **Author Contributions**

650 L.L., M.C.L., H.K., J.B., and A.B.B. conceived the project and designed the experiments. L.L., A.B.B.,
651 K.N.J., and J.B. performed murine experiments. H.K., K.N., and H.I. performed primate experiments.
652 A.B.B., J.B., and H.K. performed data analyses. L.L., M.C.L., H.K., J.B., and A.B.B. wrote the
653 manuscript.
654
655

656 **Author Information**

657 The authors declare no competing interests. Correspondence and requests for materials should be
658 addressed to L.L (louis.lefebvre@ubc.ca) or M.C.L. (matthew.lorincz@ubc.ca).
659
660

661 **References**

662

663

664 1. Stewart, K. R. *et al.* Dynamic changes in histone modifications precede de novo DNA
665 methylation in oocytes. *Genes Dev.* **29,** 2449–2462 (2015).
666 2. Gahurova, L. *et al.* Transcription and chromatin determinants of de novo DNA methylation
667 timing in oocytes. *Epigenetics Chromatin* **10,** 25 (2017).
668 3. Veselovska, L. *et al.* Deep sequencing and de novo assembly of the mouse oocyte transcriptome
669 define the contribution of transcription to the DNA methylation landscape. *Genome Biol.* **16,** 209
670 (2015).
671 4. Shirane, K. *et al.* Mouse oocyte methylomes at base resolution reveal genome-wide
672 accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet.* **9,**
673 e1003439 (2013).
674 5. Kobayashi, H. *et al.* Contribution of intragenic DNA methylation in mouse gametic DNA
675 methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* **8,** e1002440 (2012).
676 6. Tang, W. W. C. *et al.* A Unique Gene Regulatory Network Resets the Human Germline
677 Epigenome for Development. *Cell* **161,** 1453–1467 (2015).
678 7. Guo, F. *et al.* The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ
679 Cells. *Cell* **161,** 1437–1452 (2015).
680 8. Okae, H. *et al.* Genome-wide analysis of DNA methylation dynamics during early human
681 development. *PLoS Genet.* **10,** e1004868 (2014).
682 9. Xu, Q. *et al.* SETD2 regulates the maternal epigenome, genomic imprinting and embryonic
683 development. *Nat. Genet.* **51,** 844–856 (2019).
684 10. Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and
685 preimplantation embryos. *Nat. Genet.* **43,** 811–814 (2011).
686 11. Hiura, H., Obata, Y., Komiyama, J., Shirai, M. & Kono, T. Oocyte growth-dependent
687 progression of maternal imprinting in mice. *Genes Cells* **11,** 353–361 (2006).
688 12. Singh, V. B. *et al.* Blocked transcription through KvDMR1 results in absence of methylation and
689 gene silencing resembling Beckwith-Wiedemann syndrome. *Development* **144,** 1820–1830
690 (2017).
691 13. Smith, E. Y., Futtner, C. R., Chamberlain, S. J., Johnstone, K. A. & Resnick, J. L. Transcription
692 is required to establish maternal imprinting at the Prader-Willi syndrome and Angelman
693 syndrome locus. *PLoS Genet.* **7,** e1002422 (2011).

694 14. Chotalia, M. *et al.* Transcription is required for establishment of germline methylation marks at
695 imprinted genes. *Genes Dev.* **23,** 105–117 (2009).
696 15. Joh, K. *et al.* Growing oocyte-specific transcription-dependent de novo DNA methylation at the
697 imprinted Zrsr1-DMR. *Epigenetics Chromatin* **11,** 28 (2018).
698 16. Bretz, C. L. & Kim, J. Transcription-driven DNA methylation setting on the mouse Peg3 locus.
699 *Epigenetics* **12,** 945–952 (2017).
700 17. Peaston, A. E. *et al.* Retrotransposons regulate host genes in mouse oocytes and preimplantation
701 embryos. *Dev. Cell* **7,** 597–606 (2004).
702 18. Franke, V. *et al.* Long terminal repeats power evolution of genes and gene expression programs
703 in mammalian oocytes and zygotes. *Genome Res.* **27,** 1384–1394 (2017).
704 19. Brind'Amour, J. *et al.* LTR retrotransposons transcribed in oocytes drive species-specific and
705 heritable changes in DNA methylation. *Nature Communications* **9,** 3331–3344 (2018).
706 20. Schulz, R. *et al.* The parental non-equivalence of imprinting control regions during mammalian
707 development and evolution. *PLoS Genet.* **6,** e1001214 (2010).
708 21. Hanna, C. W. *et al.* Pervasive polymorphic imprinted methylation in the human placenta.
709 *Genome Res.* **26,** 756–767 (2016).
710 22. Court, F. *et al.* Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies
711 of human imprinting and suggests a germline methylation-independent mechanism of
712 establishment. *Genome Res.* **24,** 554–569 (2014).
713 23. Sanchez-Delgado, M. *et al.* Absence of Maternal Methylation in Biparental Hydatidiform Moles
714 from Women with NLRP7 Maternal-Effect Mutations Reveals Widespread Placenta-Specific
715 Imprinting. *PLoS Genet.* **11,** e1005644 (2015).
716 24. Cuellar Partida, G. *et al.* Genome-wide survey of parent-of-origin effects on DNA methylation
717 identifies candidate imprinted loci in humans. *Hum. Mol. Genet.* **27,** 2927–2939 (2018).
718 25. Hamada, H. *et al.* Allele-Specific Methylome and Transcriptome Analysis Reveals Widespread
719 Imprinting in the Human Placenta. *Am. J. Hum. Genet.* **99,** 1045–1058 (2016).
720 26. Proudhon, C. *et al.* Protection against de novo methylation is instrumental in maintaining parent-
721 of-origin methylation inherited from the gametes. *Mol Cell* **47,** 909–920 (2012).
722 27. Wang, L. *et al.* Programming and inheritance of parental DNA methylomes in mammals. *Cell*
723 **157,** 979–991 (2014).
724 28. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated
725 transcripts and isoform switching during cell differentiation. *Nat Biotech* **28,** 511–515 (2010).

726 29. Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating
727 cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **487,** 57 (2017).
728 30. Babaian, A. *et al.* LIONS: Analysis Suite for Detecting and Quantifying Transposable Element
729 Initiated Transcription from RNA-seq. *Bioinformatics* **7,** 24. (2019).
730 31. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq
731 reads. *Nat Biotech* **33,** 290–295 (2015).
732 32. Mizuno, Y. *et al.* Asb4, Ata3, and Dcn are novel imprinted genes identified by high-throughput
733 screening using RIKEN cDNA microarray. *Biochem Biophys Res Commun* **290,** 1499–1505
734 (2002).
735 33. Smith, R. J., Dean, W., Konfortova, G. & Kelsey, G. Identification of novel imprinted genes in a
736 genome-wide screen for maternal methylation. *Genome Res.* **13,** 558–569 (2003).
737 34. Hagiwara, Y. *et al.* Screening for imprinted genes by allelic message display: identification of a
738 paternally expressed gene impact on mouse chromosome 18. *Proc Natl Acad Sci USA* **94,** 9249–
739 9254 (1997).
740 35. Sanchez-Delgado, M., Vidal, E., PLoS, J. M.2016. Human oocyte-derived methylation
741 differences persist in the placenta revealing widespread transient imprinting. *PLoS Genet.* **12,**
742 e1006427 (2016).
743 36. Metsalu, T. *et al.* Using RNA sequencing for identifying gene imprinting and random
744 monoallelic expression in human placenta. *Epigenetics* **9,** 1397–1409 (2014).
745 37. Yuen, R. K., Jiang, R., Peñaherrera, M. S., Mcfadden, D. E. & Robinson, W. P. Genome-wide
746 mapping of imprinted differentially methylated regions by DNA methylation profiling of human
747 placentas from triploidies. *Epigenetics Chromatin* **4,** 10 (2011).
748 38. Okae, H. *et al.* Re-investigation and RNA sequencing-based identification of genes with
749 placenta-specific imprinted expression. *Hum. Mol. Genet.* **21,** 548–558 (2012).
750 39. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in
751 eukaryotic genomes. *Mob DNA* **6,** 11 (2015).
752 40. Ruebel, M. L. *et al.* Transcriptome analysis of rhesus monkey failed-to-mature oocytes:
753 deficiencies in transcriptional regulation and cytoplasmic maturation of the oocyte mRNA
754 population. *Molecular Human Reproduction* **24,** 478–494 (2018).
755 41. Gao, F. *et al.* De novo DNA methylation during monkey pre-implantation embryogenesis. *Cell*
756 *Res* **27,** 526–539 (2017).

757 42. Okamura, K., Sakaki, Y. & Ito, T. Comparative genomics approach toward critical determinants
758 for the imprinting of an evolutionarily conserved gene Impact. *Biochem Biophys Res Commun*
759 **329,** 824–830 (2005).
760 43. Okae, H. *et al.* RNA sequencing-based identification of aberrant imprinting in cloned mice.
761 *Hum. Mol. Genet.* **23,** 992–1001 (2014).
762 44. Hanna, C. W. *et al.* MLL2 conveys transcription-independent H3K4 trimethylation in oocytes.
763 *Nat Struct Mol Biol* **25,** 73–82 (2018).
764 45. Wood, A. J. *et al.* A screen for retrotransposed imprinted genes reveals an association between X
765 chromosome homology and maternal germ-line methylation. *PLoS Genet.* **3,** e20 (2007).
766 46. Cowley, M. & Oakey, R. J. Retrotransposition and genomic imprinting. *Brief Funct Genomics* **9,**
767 340–346 (2010).
768 47. Inoue, A., Jiang, L., Lu, F., Suzuki, T. & Zhang, Y. Maternal H3K27me3 controls DNA
769 methylation-independent imprinting. *Nature* **547,** 419–424 (2017).
770 48. Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks.
771 *Mol Cell* **42,** 330–341 (2011).
772 49. Matoba, S. *et al.* Loss of H3K27me3 Imprinting in Somatic Cell Nuclear Transfer Embryos
773 Disrupts Post-Implantation Development. *Cell Stem Cell* **23,** 343–354.e5 (2018).
774 50. Takahashi, N. *et al.* ZNF445 is a primary regulator of genomic imprinting. *Genes Dev.* **33,** 49–
775 54 (2019).
776 51. Auclair, G. *et al.* EHMT2 directs DNA methylation for efficient gene silencing in mouse
777 embryos. *Genome Res.* **26,** 192–202 (2016).
778 52. Barlow, D. P. Methylation and imprinting: from host defense to gene regulation? *Science* **260,**
779 309–310 (1993).
780 53. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic
781 parasites. *Trends Genet* **13,** 335–340 (1997).
782 54. McDonald, J. F., Matzke, M. A. & Matzke, A. J. Host defenses to transposable elements and the
783 evolution of genomic imprinting. *Cytogenet Genome Res* **110,** 242–249 (2005).
784 55. Pask, A. J. *et al.* Analysis of the platypus genome suggests a transposon origin for mammalian
785 imprinting. *Genome Biol.* **10,** R1 (2009).
786 56. Song, Q. *et al.* A reference methylome database and analysis pipeline to facilitate integrative and
787 comparative epigenomics. *PLoS ONE* **8,** e81148 (2013).
788 57. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human
789 genome. *Nature* **489,** 57–74 (2012).

790 58. Hahn, O. *et al.* Dietary restriction protects from age-associated DNA methylation and induces
791 epigenetic reprogramming of lipid metabolism. *Genome Biol.* **18,** 56 (2017).
792 59. Harten, S. K. *et al.* The recently identified modifier of murine metastable epialleles, Rearranged
793 L-Myc Fusion, is involved in maintaining epigenetic marks at CpG island shores and enhancers.
794 *BMC Biol* **13,** 21 (2015).
795 60. Hodges, E. *et al.* Directional DNA methylation changes and complex intermediate states
796 accompany lineage specificity in the adult hematopoietic compartment. *Mol Cell* **44,** 17–28
797 (2011).
798 61. Hon, G. C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA methylation
799 maps from adult mouse tissues. *Nat. Genet.* **45,** 1198–1206 (2013).
800 62. Mendizabal, I. *et al.* Comparative Methylome Analyses Identify Epigenetic Regulatory Loci of
801 Human Brain Evolution. *Molecular Biology and Evolution* **33,** 2947–2959 (2016).
802 63. Molaro, A. *et al.* Sperm methylation profiles reveal features of epigenetic inheritance and
803 evolution in primates. *Cell* **146,** 1029–1041 (2011).
804 64. Okae, H. *et al.* Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell* **22,** 50–63.e6
805 (2018).
806 65. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human
807 epigenomes. *Nature* **518,** 317–330 (2015).
808 66. Schroeder, D. I. *et al.* Early Developmental and Evolutionary Origins of Gene Body DNA
809 Methylation Patterns in Mammalian Placentas. *PLoS Genet.* **11,** e1005442 (2015).
810 67. Schroeder, D. I., Lott, P., Korf, I. & LaSalle, J. M. Large-scale methylation domains mark a
811 functional subset of neuronally expressed genes. *Genome Res.* **21,** 1583–1591 (2011).
812 68. Tung, J. *et al.* Social environment is associated with gene regulatory variation in the rhesus
813 macaque immune system. *Proc Natl Acad Sci USA* **109,** 6490–6495 (2012).
814 69. Zeng, J. *et al.* Divergent whole-genome methylation maps of human and chimpanzee brains
815 reveal epigenetic basis of human regulatory evolution. *Am J Hum Genet* **91,** 455–465 (2012).
816 70. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2012).
817 71. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary
818 Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* **35,** 1547–
819 1549 (2018).
820 72. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and
821 annotation of phylogenetic and other trees. *Nucleic Acids Res* **44,** W242–W245 (2016).

822 73. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–
823 823 (2013).
824 74. Tucker, K. L. *et al.* Germ-line passage is required for establishment of methylation and
825 expression patterns of imprinted but not of nonimprinted genes. *Genes Dev.* **10,** 1008–1020
826 (1996).
827 75. Gertsenstein, M. *et al.* Efficient Generation of Germ Line Transmitting Chimeras from
828 C57BL/6N ES Cells by Aggregation with Outbred Host Embryos. *PLoS ONE* **5,** e11260 (2010).
829 76. Truett, G. E. *et al.* Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide
830 and tris (HotSHOT). *Biotech.* **29,** 52–54 (2000).
831
832
833 **Figure legends**
834
835 **Figure 1. Identification of human and mouse maternal igDMRs embedded within lineage-specific**
836 **LITs. a**, Venn diagram showing the intersection of known maternal igDMRs in mouse and human,
837 along with the subset of igDMRs in each species embedded within a LIT. For each LIT-associated
838 igDMR, the family of the LTR in which transcription initiates in oocytes is shown on the right. The
839 presence of each LTR family in relevant mammalian lineages is color-coded as in Supplementary Fig.
840 1c. **b**, List of imprinted genes/igDMRs associated with LITs. Maternal igDMRs unique to mouse (4) or
841 human (17) are shown, along with DNAme levels (heat map) for each igDMR in syntenic regions in
842 the gametes, blastocyst, placenta and adult tissues in human and mouse. The retrogene *retro-Coro1c* is
843 absent in the syntenic human region on chromosome 6p22.3 (*no orthologue*). *ZNF396* does not have a
844 syntenic CGI in mice (*no synteny*). CT: cytotrophoblast; PBMC: peripheral blood mononuclear cell. **c-**
845 **d**, Screenshots of the human and mouse *RHOBTB3*/*Rhobtb3* and *SCIN*/*Scin* loci, including locations of
846 annotated genes, LTR retrotransposons, and regions of syntenic homology. The relevant CGI, igDMR,
847 and upstream LTR in human are highlighted in green, blue, and red respectively. For each species,
848 RNA-seq data from GVOs are shown, along with assembled transcripts, including LITs and their 5'
849 LTR exons (red) for the human genes. DNAme levels in gametes, blastocyst, placenta, and liver are
850 shown across each locus in both species. For the human DNAme data, profiles from female 11-week
851 primordial germs cells are also shown (11W PGC) and oocyte DNAme is from a mixture of GVO and
852 MII oocytes. Details of all the datasets used in this study are presented in Supplementary Table 1.
853

854 **Figure 2. Conservation of oocyte LTR-initiated transcription and gametic imprinting in**
855 **primates. a**, Table of 16 human genes with igDMRs embedded within LITs active in oocytes and
856 showing maternal/allelic DNAme in blastocyst and cytotrophoblast. The family of the initiating LTR is
857 shown on the left, color-coded according to the phylogenetic distribution of the ERV family (top), as in
858 Fig. 1a. For each species, the presence of the initiating LTR at each locus is indicated by a matching
859 colored box and the igDMR DNAme status in human or the syntenic region in chimp, macaque and
860 mouse placenta is shown. An empty box indicates no data. Arrows indicate genes for which evidence
861 of allelic transcription has been published (Supplementary Table 2). Gene names in bold were analyzed
862 in greater detail. **b**, Conservation of LITs in human and macaque oocytes for the 18 igDMRs from
863 panel **a**. Solid boxes indicate LITs discovered by LIONS, boxed hatches indicate LITs with evidence of
864 splicing from the LTR over the igDMR, and unboxed hatches indicate evidence of transcription from
865 the LTR. Dashes indicate LTRs from which no transcription is seen, and loci for which the relevant
866 LTR is absent from the macaque genome are also shown (X: No LTR). *: An LTR12F may initiate a
867 LIT in macaque oocytes (Supplementary Fig. 3a). **c**, Screenshot of the human *ST8SIA1* locus, showing
868 its CGI promoter (green), annotated gene (blue), oocyte LIT (red), and DNAme in human and macaque
869 oocytes (black). Highlighted are the location of the upstream LTR12C in human and chimp (red) and
870 the human igDMR (blue). **d**, Screenshot of the human *RHOBTB3* locus annotated as in **c** and
871 highlighting the upstream MSTA LTR promoter (pink). **e**, Violin plots of the distribution of mean
872 DNAme levels in placenta for individual bisulphite-sequencing reads covering the igDMRs of the
873 orthologous chimp and macaque *ST8SIA1*, *HECW1*, *RHOBTB3* and *MCCC1* genes. Coloured boxes
874 indicate the presence of the proximal LTR. For each gene, the mean DNAme level at each of the CpGs
875 surveyed is shown below. Symbols for DNAme are as in **a**. Note that the MER51E in the macaque
876 *MCCC1* locus is transcriptionally inert, likely due to macaque-specific SNPs rendering it
877 transcriptionally inactive (Supplementary Fig. 4d-f). **f**, Placental DNAme data at the *HECW1, GLIS3,*
878 and *MCCC1* loci for informative samples heterozygous at SNPs of known parental origin, are shown
879 for the species indicated.
880
881 **Figure 3. Phylogenetic relationship between LTR, LIT, and maternal DNAme at mouse-specific**
882 **igDMRs. a**, The presence or absence of a LIT overlapping the CGI at *Cdh15*, *Slc38a4*, and *Impact* are
883 indicated with closed boxes and dashes, respectively. X: species in which the LTR is absent. **b**,
884 Screenshots of *Slc38a4*/*SLC38A4* and *Impact*/*IMPACT* loci showing GVO RNA-seq data (and the
885 associated *de novo* Cufflinks transcript assembly), DNAme profiles in NGO, GVO and sperm, and

886 H3K4me3, PolII, and H3K36me3 ChIP-seq tracks for mouse GVO. Highlighted are the upstream
887 initiating LTR (red), CGI promoter (green), and mouse igDMR (blue). The syntenic region in human is
888 also shown, including GVO RNA-seq and DNAme from oocytes and sperm. A 5' RACE gene model
889 initiating within the MTC element at the *Impact* locus is included in the right panel[14]. **c**, Scatter plot of
890 oocyte H3K4me3 and transcription levels for all mouse MT2A elements, with those acting as
891 transcription start sites in oocytes highlighted in blue. The remaining transcribed elements reflect
892 exonization events. The MT2A element initiating the LIT at *Slc38a4* (red) is amongst the most active
893 elements of this family. **d**, Histogram of the distribution of mouse MT2A LTRs as a function of
894 divergence from the consensus sequence. The LTR driving expression at *Slc38a4* is amongst the most
895 highly diverged.
896
897 **Figure 4. Loss of imprinting at *Slc38a4* upon maternal transmission of the MT2A KO allele. a**,
898 Genome browser screenshot of the mouse *Slc38a4* promoter and upstream region, including the MT2A
899 LTR (red), annotated *Slc38a4* exon 1, CGI (green), and igDMR (blue). GVO RNA-seq as well as RNA
900 pol II, H3K4me3 and H3K36me3 ChIP-seq tracks are shown, along with DNAme data for GVO, sperm
901 and adult liver. The region within the igDMR analyzed by sodium bisulfite sequencing (SBS), which
902 includes 11 CpG sites, is shown at the bottom. Δ: extent of the MT2AKO deletion allele. **b**, DNAme of
903 the *Slc38a4* igDMR in GVO from wild-type and $Slc38a4^{MT2AKO/MT2AKO}$ females determined by SBS. **c**,
904 DNAme of the *Slc38a4* igDMR in E13.5 ($Slc38a4^{+/MT2AKO}$ × CAST)F1 embryos determined by SBS. Data
905 for control ($+/+^{c}$) and heterozygous ($KO/+^{c}$) littermates with a maternally inherited MT2AKO are
906 shown. $+^{c}$: wild-type CAST allele; KO: $Slc38a4^{MT2AKO}$. A polymorphic insertion in the amplified region
907 allows for discrimination of maternal (Mat) and paternal (Pat) strands. Allele-specific expression
908 analyses of F1 E13.5 placental RNA by (**d**) RT-PCR followed by PvuII RFLP analysis (RT: reverse
909 transcriptase), and (**e-f**) Sanger sequencing of a T↔C transition in the 3'UTR of the *Slc38a4* cDNA
910 (maternal B6: T allele; and paternal CAST: C allele). Each bar in **f** shows the mean of individual
911 samples, and error bars show S.D. of two SNPs analyzed. **g**, Total *Slc38a4* mRNA levels in E8.5,
912 E13.5, E16.5 and E18.5 placentae, as determined by RT-qPCR (n=6 for each datapoint). Expression
913 levels are relative to the housekeeping gene *Ppia*. Graph shows mean ± S.D.
914
915 **Figure 5. Loss of imprinting at *Impact* upon maternal transmission of the MTC KO allele.**
916 **a**, Genome browser screenshot of the mouse *Impact* locus, including the upstream MTC LTR (red),
917 CGI (green), and igDMR (blue). GVO RNA-seq as well as RNA pol II, H3K4me3, and H3K36me3

918 ChIP-seq tracks are shown, along with DNAme data for GVO, sperm, and adult liver. The region
919 within the igDMR analyzed by sodium bisulfite sequencing (SBS), which includes 10 CpG sites, is
920 shown at the bottom. Δ: extent of the upstream MTCKO deletion allele. **b**, DNAme of the *Impact*
921 igDMR in GVO from wildtype and *Impact*$^{MTCKO/MTCKO}$ females determined by SBS. **c**, DNAme of the *Impact*
922 igDMR in E13.5 (*Impact*$^{+/MTCKO}$ × CAST)F1 embryos determined by SBS. Data for control (+/+^{c}) and
923 heterozygous (KO/+^{c}) littermates with a maternally inherited MTCKO are shown. +^{c}: wild-type CAST
924 allele; KO: *Impact*^{MTCKO}. A polymorphic insertion and a SNP in the amplified region allow for
925 discrimination of maternal (Mat) and paternal (Pat) strands. **d-e**, Allele-specific expression analysis of
926 F1 E13.5 embryonic head RNA by (**d**) RT-PCR followed by MluCI RFLP analysis, and (**e**) Sanger
927 sequencing of a A⟷G transition in the 3'UTR of the *Impact* mRNA (maternal B6: A allele; and
928 paternal CAST: G allele). RT: reverse transcriptase. **f**, Quantification of relative levels of expression
929 from the paternal *Impact* allele based on the analysis of embryos as in **e**. Graph shows mean ±S.D of
930 three SNPs. **g**, *Impact* mRNA levels analyzed by RT-qPCR on E13.5 embryonic RNA, from six
931 embryos of each genotype from two litters. Expression levels are relative to those for the wild-type
932 allele. Graph shows mean ± S.E.
933

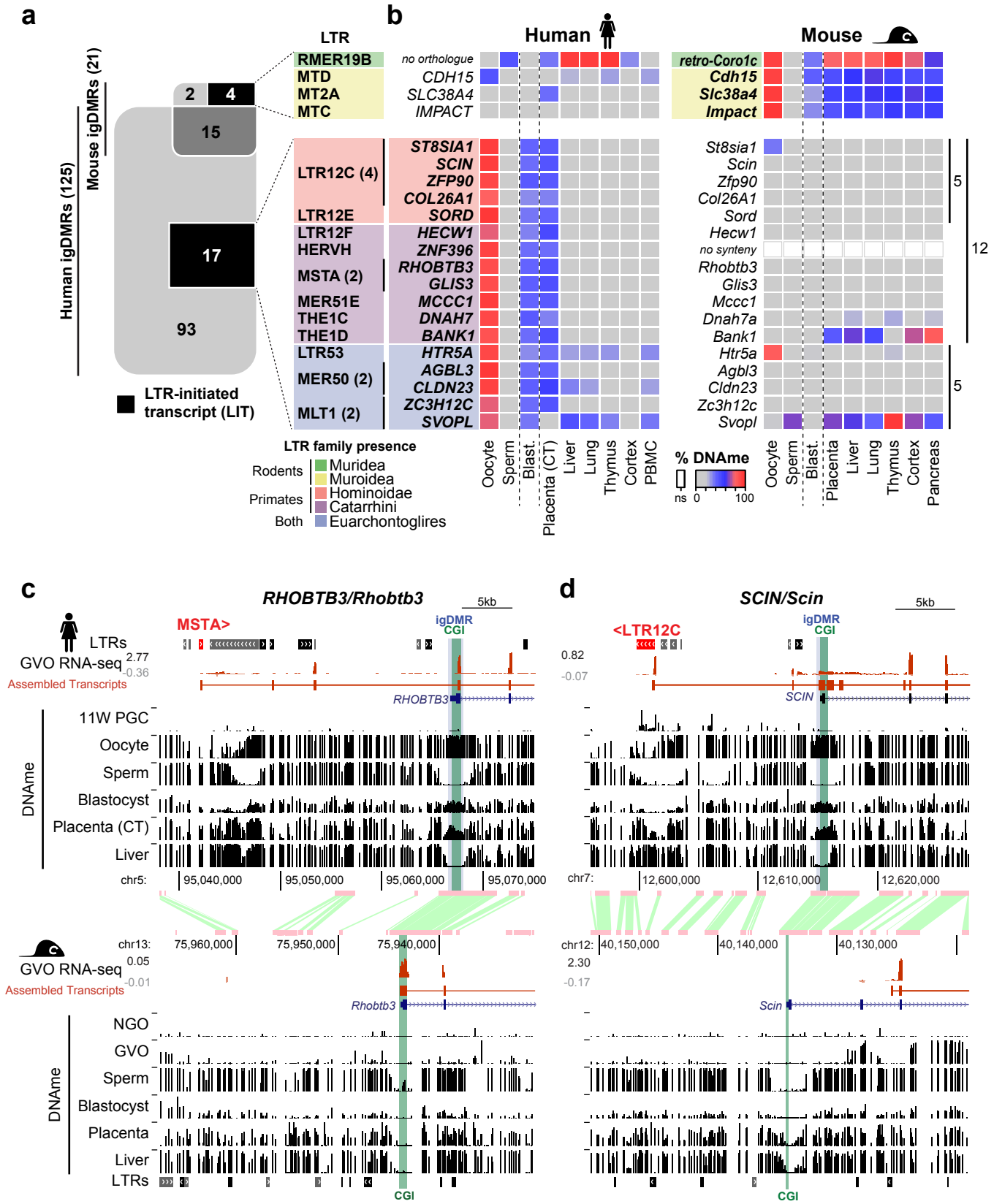
a
b
Human igDMRs (125)
Mouse igDMRs (21)
2
4
15
17
93
LTR-initiated transcript (LIT)
LTR
Human
Mouse
RMER19B
MTD
MT2A
MTC
no orthologue
CDH15
SLC38A4
IMPACT
retro-Coro1c
Cdh15
Slc38a4
Impact
LTR12C (4)
LTR12E
LTR12F
HERVH
MSTA (2)
MER51E
THE1C
THE1D
LTR53
MER50 (2)
MLT1 (2)
ST8SIA1
SCIN
ZFP90
COL26A1
SORD
HECW1
ZNF396
RHOBTB3
GLIS3
MCCC1
DNAH7
BANK1
HTR5A
AGBL3
CLDN23
ZC3H12C
SVOPL
St8sia1
Scin
Zfp90
Col26A1
Sord
Hecw1
no synteny
Rhobtb3
Glis3
Mccc1
Dnah7a
Bank1
Htr5a
Agbl3
Cldn23
Zc3h12c
Svopl
5
12
5
LTR family presence
Rodents
Muridea
Muroidea
Primates
Hominoidae
Catarrhini
Both
Euarchontoglires
Oocyte
Sperm
Blast.
Placenta (CT)
Liver
Lung
Thymus
Cortex
PBMC
Placenta
Pancreas
% DNAme
ns
0
100
c
RHOBTB3/Rhobtb3
5kb
MSTA>
igDMR
CGI
LTRs
GVO RNA-seq
2.77
-0.36
Assembled Transcripts
RHOBTB3
DNAme
11W PGC
Oocyte
Sperm
Blastocyst
Placenta (CT)
Liver
chr5:
95,040,000
95,050,000
95,060,000
95,070,000
chr13:
75,960,000
75,950,000
75,940,000
0.05
-0.01
Rhobtb3
NGO
GVO
Sperm
Blastocyst
Placenta
Liver
LTRs
CGI
d
SCIN/Scin
5kb
<LTR12C
igDMR
CGI
0.82
-0.07
SCIN
chr7:
12,600,000
12,610,000
12,620,000
chr12:
40,150,000
40,140,000
40,130,000
2.30
-0.17
Scin
CGI

Figure 1

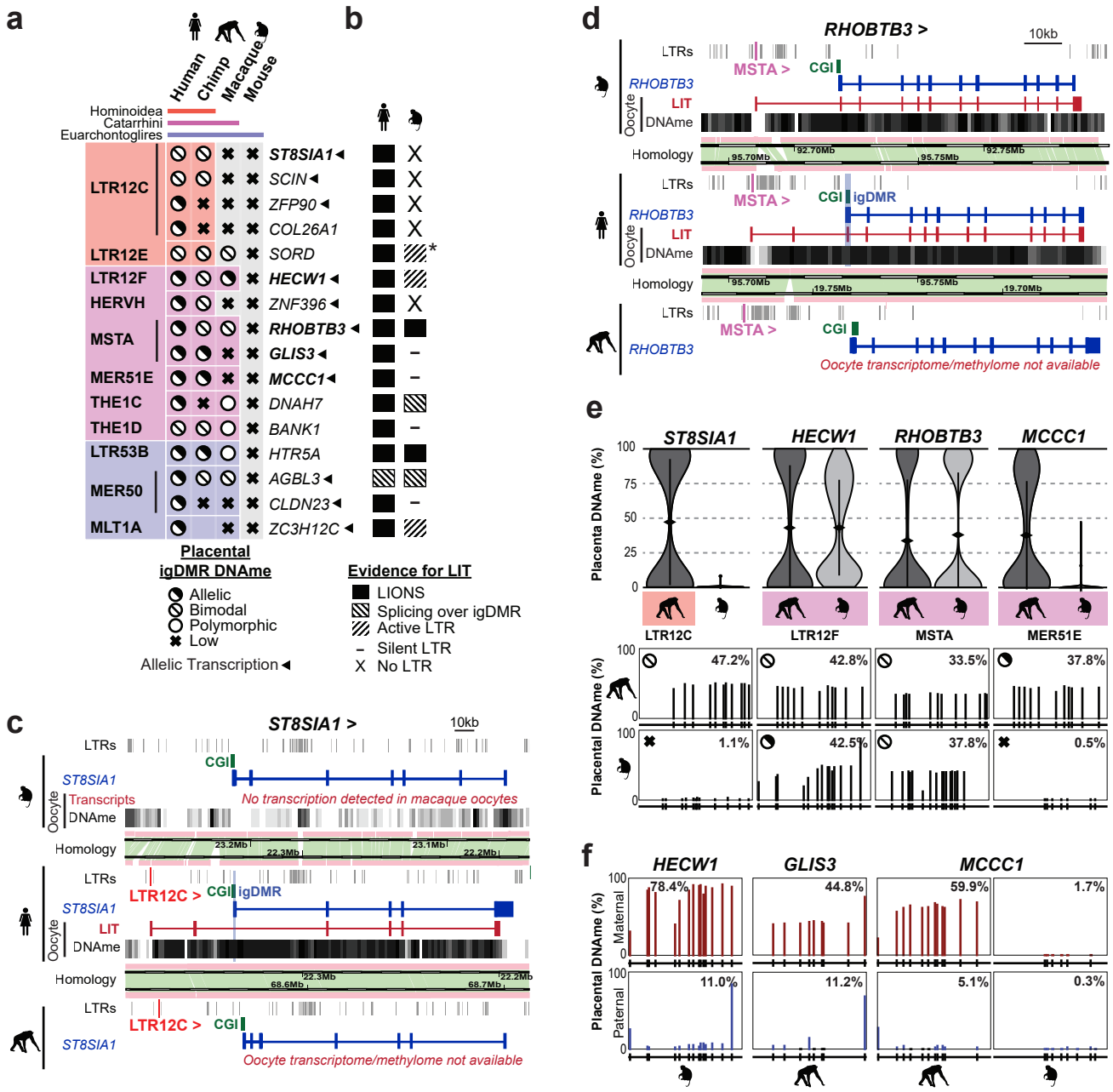
a
b
Human
Chimp
Macaque
Mouse
Hominoidea
Catarrhini
Euarchontoglires
LTR12C
LTR12E
LTR12F
HERVH
MSTA
MER51E
THE1C
THE1D
LTR53B
MER50
MLT1A
ST8SIA1
SCIN
ZFP90
COL26A1
SORD
HECW1
ZNF396
RHOBTB3
GLIS3
MCCC1
DNAH7
BANK1
HTR5A
AGBL3
CLDN23
ZC3H12C
Placental igDMR DNAme
Allelic
Bimodal
Polymorphic
Low
Allelic Transcription
Evidence for LIT
LIONS
Splicing over igDMR
Active LTR
Silent LTR
No LTR
c
ST8SIA1 >
10kb
LTRs
CGI
ST8SIA1
Transcripts
No transcription detected in macaque oocytes
Oocyte
DNAme
Homology
LTR12C >
igDMR
LIT
Oocyte transcriptome/methylome not available
d
RHOBTB3 >
MSTA >
RHOBTB3
Oocyte transcriptome/methylome not available
e
ST8SIA1
HECW1
RHOBTB3
MCCC1
Placenta DNAme (%)
LTR12C
LTR12F
MSTA
MER51E
47.2%
42.8%
33.5%
37.8%
1.1%
42.5%
37.8%
0.5%
f
HECW1
GLIS3
MCCC1
Placental DNAme (%)
Maternal
Paternal
78.4%
44.8%
59.9%
1.7%
11.0%
11.2%
5.1%
0.3%

Figure 2

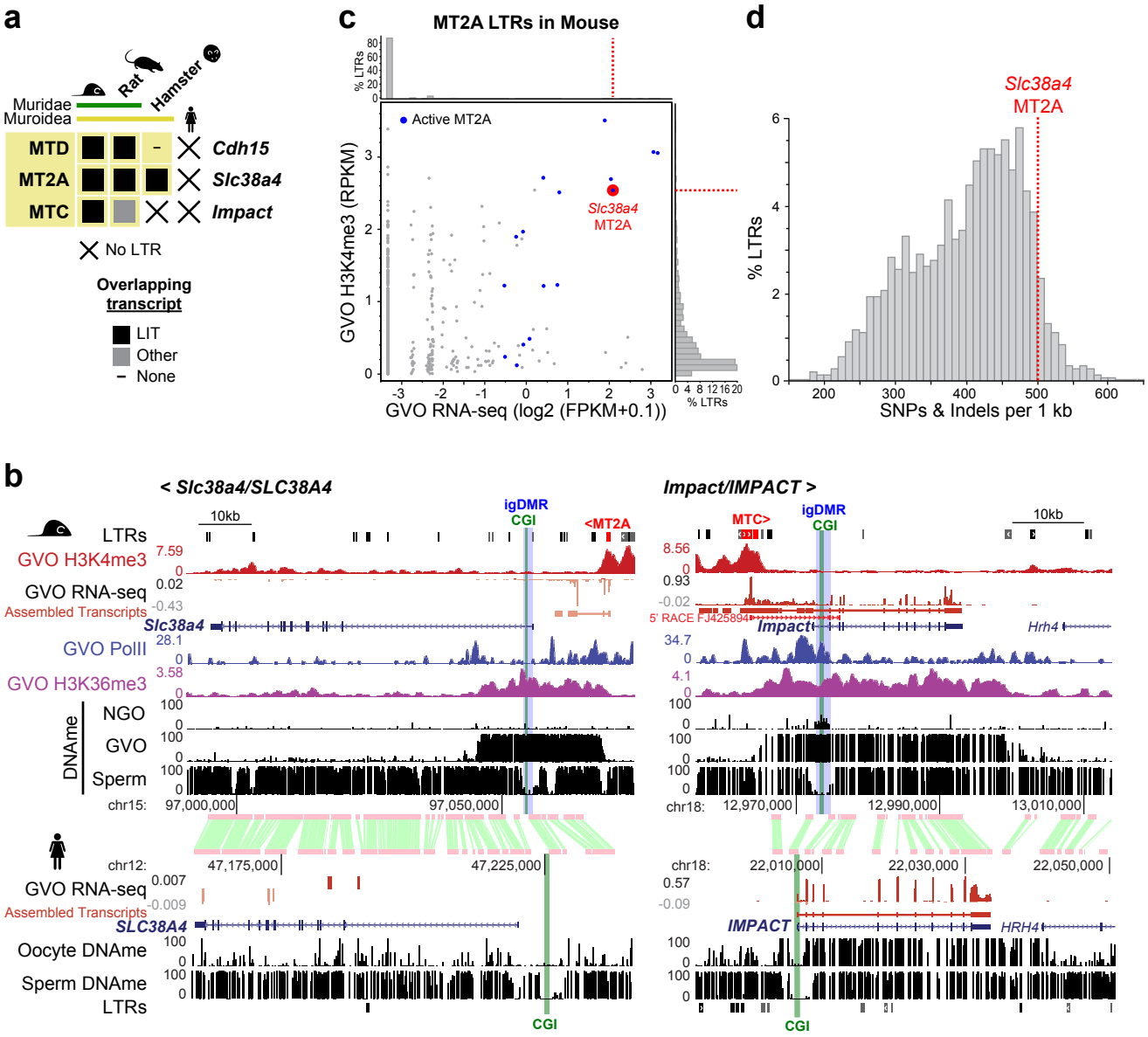
a
Rat
Hamster
Muridae
Muroidea
MTD
MT2A
MTC
Cdh15
Slc38a4
Impact
No LTR
Overlapping transcript
LIT
Other
None
c
MT2A LTRs in Mouse
% LTRs
Active MT2A
Slc38a4
MT2A
GVO H3K4me3 (RPKM)
GVO RNA-seq (log2 (FPKM+0.1))
d
Slc38a4
MT2A
% LTRs
SNPs & Indels per 1 kb
b
< Slc38a4/SLC38A4
Impact/IMPACT >
LTRs
GVO H3K4me3
GVO RNA-seq
Assembled Transcripts
GVO PolII
GVO H3K36me3
DNAme
NGO
GVO
Sperm
igDMR
CGI
<MT2A
MTC>
10kb
5' RACE FJ425894
Slc38a4
Impact
Hrh4
chr15:
97,000,000
97,050,000
chr18:
12,970,000
12,990,000
13,010,000
chr12:
47,175,000
47,225,000
22,010,000
22,030,000
22,050,000
SLC38A4
IMPACT
HRH4
Oocyte DNAme
Sperm DNAme
LTRs

Figure 3

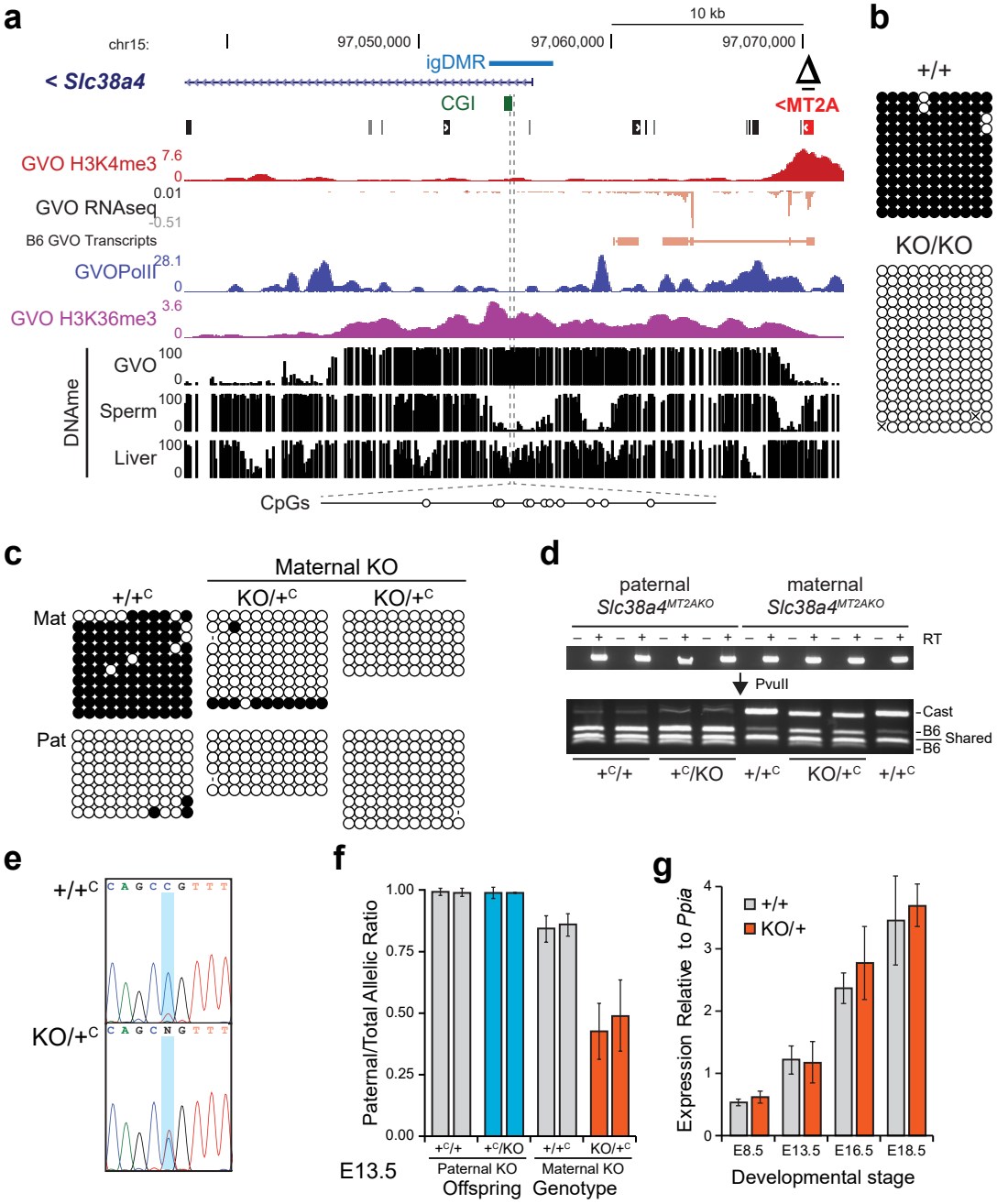
a
10 kb
chr15:
97,050,000
97,060,000
97,070,000
igDMR
< Slc38a4
CGI
<MT2A
GVO H3K4me3
GVO RNAseq
B6 GVO Transcripts
GVOPolII
GVO H3K36me3
DNAme
GVO
Sperm
Liver
CpGs
b
+/+
KO/KO
c
Maternal KO
+/+C
KO/+C
KO/+C
Mat
Pat
d
paternal
Slc38a4MT2AKO
maternal
Slc38a4MT2AKO
RT
PvuII
Cast
B6
B6
Shared
+C/+
+C/KO
+/+C
KO/+C
+/+C
e
+/+C
KO/+C
f
Paternal/Total Allelic Ratio
E13.5
Paternal KO
Maternal KO
Offspring Genotype
g
+/+
KO/+
Expression Relative to Ppia
E8.5
E13.5
E16.5
E18.5
Developmental stage

Figure 4

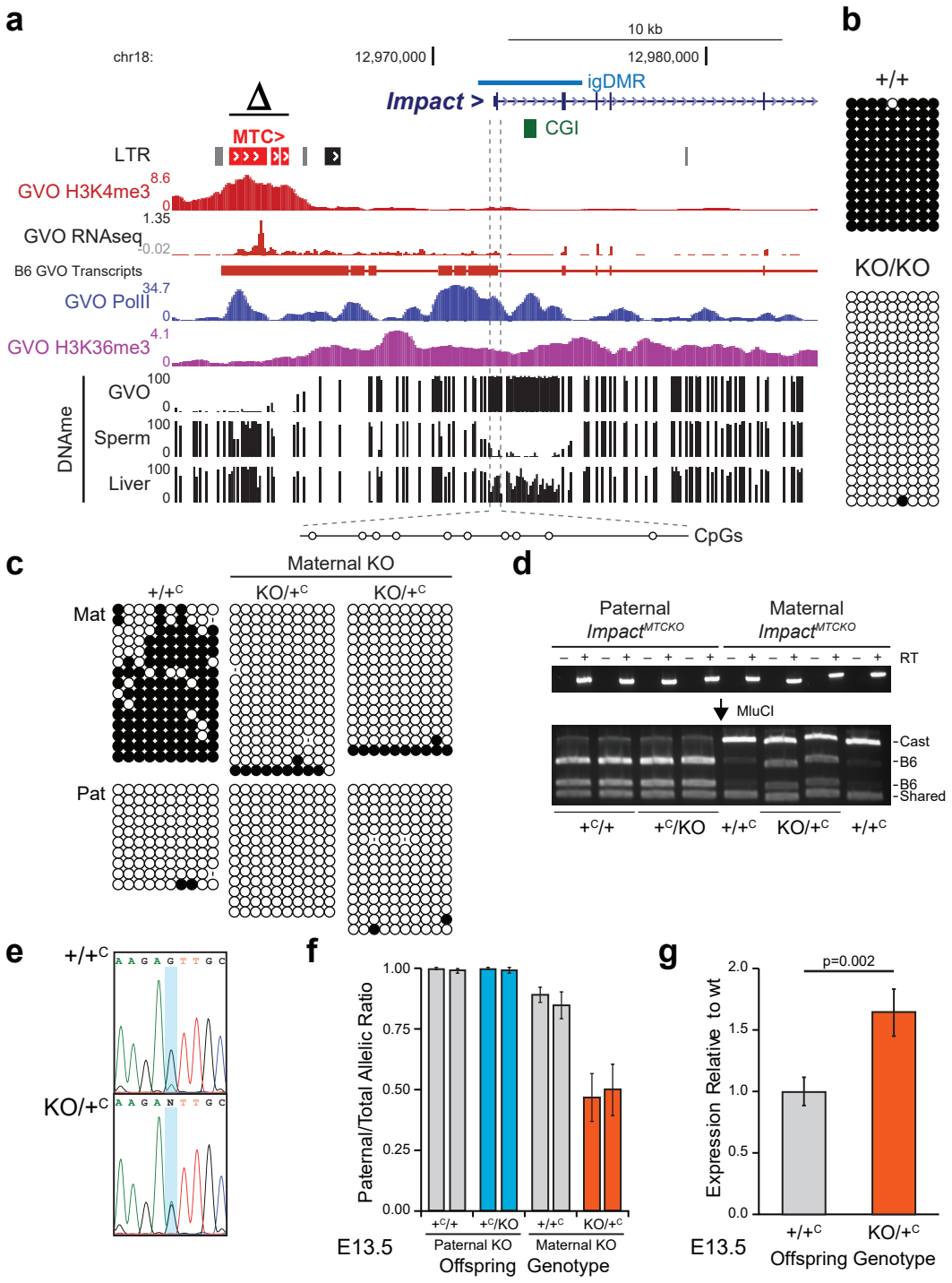
a
chr18:
12,970,000
12,980,000
10 kb
igDMR
Impact >
CGI
Δ
MTC>
LTR
GVO H3K4me3
GVO RNAseq
B6 GVO Transcripts
GVO PolII
GVO H3K36me3
DNAme
GVO
Sperm
Liver
CpGs
b
+/+
KO/KO
c
Maternal KO
+/+C
KO/+C
KO/+C
Mat
Pat
d
Paternal
Impact MTCKO
Maternal
Impact MTCKO
RT
MluCI
Cast
B6
B6
Shared
+C/+
+C/KO
+/+C
KO/+C
+/+C
e
+/+C
KO/+C
f
Paternal/Total Allelic Ratio
E13.5
Paternal KO
Maternal KO
Offspring Genotype
g
p=0.002
Expression Relative to wt
E13.5
+/+C
KO/+C
Offspring Genotype

Figure 5