Prevalent binding of GATA2/3 and MSX2 on endogenous retrovirus-derived regulatory elements in human trophoblast stem cells

Cui Du[1,#], Jing Jiang[1,#], Yuzhuo Li[1], Miao Yu[2], Jian Jin[1], Todd S. Macfarlan[3], Bin Cao[2,*], Ming-an Sun[1,4,5,*]

[1] Institute of Comparative Medicine, College of Veterinary Medicine, Yangzhou University, Yangzhou, Jiangsu, China.

[2] Fujian Provincial Key Laboratory of Reproductive Health Research, School of Medicine, Xiamen University, Xiamen, Fujian 361102, China.

[3] The *Eunice Kennedy Shriver* National Institutes of Child Health and Human Development, NIH, Bethesda, MD 20892, USA.

[4] Joint International Research Laboratory of Important Animal Infectious Diseases and Zoonoses of Jiangsu Higher Education Institutions, Yangzhou, Jiangsu, China.

[5] Jiangsu Co-innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonosis, Joint International Research Laboratory of Agriculture and Agri-Product Safety of Ministry of Education of China, Yangzhou University, Yangzhou, Jiangsu, China

# These authors contributed equally to the manuscript.

* Corresponding authors

Ming-an Sun: mingansun@yzu.edu.cn; Bin Cao: caobin19@xmu.edu.cn.

Running Title: GATA2/3 & MSX2 bind ERV-derived TSC cis-elements

Abstract

The placenta is an organ with extraordinary phenotypic diversity in eutherian mammals. Recent evidence suggests that numerous human placental enhancers are evolved from lineage-specific insertions of endogenous retroviruses (ERVs), yet the transcription factors (TFs) underlying their regulation remain largely elusive. Here, by first focusing on MER41, a primate-specific ERV family previously linked to placenta and innate immunity, we uncover the binding motifs of multiple crucial trophoblast TFs (GATA2/3, MSX2, GRHL2) in addition to innate immunity TFs STAT1 and IRF1. Integration of ChIP-Seq data confirms the binding of GATA2/3 and MSX2 on the majority of MER41-derived enhancers in human trophoblast stem cells (TSCs). Notably, MER41-derived enhancers that are constitutively active in human TSCs are distinct from those activated upon interferon stimulation, which is determined by the binding of trophoblast TFs and their sub-family compositions. We further demonstrate that GATA2/3 and MSX2 have prevalent binding on numerous other ERV families – indicating their broad impact on ERV-derived placental enhancers. Functionally, the derepression of many syncytiotrophoblast genes after disruption of MSX2 is likely to be mediated by regulatory elements derived from ERVs – suggesting ERVs are also important for mediating transcriptional repression. Overall, this study characterized the prevalent binding of GATA2/3, MSX2 and their co-factors on ERV-derived regulatory elements in human TSCs and provided mechanistic insights into the importance of ERVs in human trophoblast regulatory network.

Key words:

Endogenous retrovirus; Regulatory element; Human trophoblast cell; MER41; GATA2/3; MSX2.

48 ## Introduction

49 Endogenous retroviruses (ERVs) make up ~8% of the human genome. ERVs originate from
50 germline infection of exogenous retroviruses, which can then be vertically inherited and expanded
51 in the host during evolution (Stoye 2012; Johnson 2019). The intact structure of ERVs includes
52 the internal protein-coding genes (eg. *gag*, *pol and env*) flanked by Long Terminal Repeats (LTRs)
53 at both ends, yet the internal genes frequently got lost via LTR recombination resulting in a high
54 proportion of "solo-LTRs" left in the genome (Johnson 2019). ERVs have long been ignored as
55 "junk DNA" - partly because most of them are epigenetically repressed (Sharif et al. 2016; Deniz
56 et al. 2019; Wolf et al. 2020), yet accumulating evidence suggests that specific ERV families could
57 be activated during embryonic development, tumorigenesis or immune response (Macfarlan et al.
58 2012; Chuong et al. 2016; Ito et al. 2020). Importantly, many ERV families are lineage-specific
59 and rich of transcription factor binding sites (TFBSs) in their LTRs, therefore can be co-opted as
60 cis-regulatory elements to boost the genetic novelty of the host (Chuong et al. 2016; Chuong et al.
61 2017; Senft and Macfarlan 2021; Sun et al. 2021; Buttler and Chuong 2022).

62 The placenta is a temporary organ crucial for nutrient/waste exchange, hormone secretion and
63 maternal-fetus immune tolerance during pregnancy (Maltepe and Fisher 2015; Ander et al. 2019).
64 Even though shared by all eutherian mammals, placentae are extraordinarily diversified in their
65 shape, structure and cellular composition (Ramsey et al. 1976; Mossman 1987; Hemberger et al.
66 2020). At the molecular level, numerous genes have altered expression patterns in human placenta
67 relative to rodents or even other primates (Rosenkrantz et al. 2021; Sun et al. 2021).
68 Correspondingly, the enhancers of placenta are the least conserved across species compared with
69 other tissues (Sun et al. 2021), with ERVs playing crucial roles for lineage-specific enhancer
70 evolution. For example, comparison between mouse and rat trophoblast stem cells (TSCs)

71 indicates that the mouse-specific RLTR13D5 family creates hundreds of enhancers co-bound by
72 the core trophoblast transcription factors (TFs) Cdx2, Eomes and Elf5 (Chuong et al. 2013). The
73 functional importance of ERVs in primate placentae has also been analyzed recently. For example,
74 one primate-specific THE1B element creates an enhancer driving *CRH* expression to influence
75 gestation length (Dunn-Fletcher et al. 2018). Recent transcriptomic and epigenomic comparisons
76 for human, macaque and mouse placentae further identified dozens of ERV families that are
77 significantly enriched in human placental enhancers (Sun et al. 2021).

78 Despite the well-recognized importance of ERVs in the placenta, the transcription factors that
79 regulate ERV activation in the human placenta remain largely obscure. Previously we
80 demonstrated that the primate-specific MER41 is amongst the top most enriched ERV families in
81 human placental enhancers (Sun et al. 2021). MER41 has six sub-families including A/B/C/D/E/G
82 (Kojima 2018), that have been shown to play important roles in innate immunity (Buttler and
83 Chuong 2022). Multiple studies suggest that MER41 elements facilitate primate-specific innate
84 immunity evolution by creating hundreds of interferon (IFN)-stimulated cis-elements bound by
85 STAT1 and IRF1 (Schmid and Bucher 2010; Chuong et al. 2016). This raises the question of which
86 TFs mediate the activation of MER41-associated enhancers in placenta. We previously identified
87 Serum response factor (SRF) as one TF that binds dozens of MER41-associated enhancers (Sun et
88 al. 2021), including one adjacent to *FBN2*, a human-placenta-expressed gene crucial for cell
89 invasion (Yu et al. 2020). However, SRF only binds a relatively small percentage (~6%) of
90 MER41-associated enhancers, indicating the existence of additional upstream TFs. Uncovering
91 the regulators of ERV-associated enhancers will be crucial for an in-depth understanding of the
92 gene regulatory networks in human placenta.

93 In this study, we first determined a list of candidate TFs that potentially regulate MER41-
94 associated enhancers, and then the binding of several crucial trophoblast TFs including GATA2/3
95 and MSX2 are validated by ChIP-seq data for human TSCs (Krendl et al. 2017; Hornbachner et
96 al. 2021). Importantly, GATA2/3 and MSX2 have prevalent binding on numerous other families
97 of ERV-derived enhancers – indicating their broad impact on ERV-associated enhancers. Overall,
98 this study improves our understanding about the regulatory mechanism and function of ERV-
99 associated enhancers in human placenta.

100 ## Results

101 **Tissue-specific activation of MER41-associated enhancers in human placenta and**
102 **trophoblast cells**

103 The primate-specific MER41 family creates numerous lineage-specific enhancers in human
104 placenta (Sun et al. 2021). To better characterize MER41-associated enhancers (hereafter
105 abbreviated as MER41-enhancer), we examined the epigenetically-annotated enhancers in
106 different human tissues and trophoblast cells. Re-analysis of ENCODE data (**Table S1**)
107 demonstrates that multiple subfamilies of MER41 (A/B/C/D) are specifically enriched in the
108 enhancers of placenta (as well as the chorion, a fetally derived extraembryonic membrane) relative
109 to other human tissues (**Figure 1A, S1**). Furthermore, genes adjacent to MER41-enhancers have
110 increased expression in human placenta relative to other tissues and mouse placenta (**Figure 1B,**
111 **S2**), and some enriched GO terms (eg. endocytosis, chordae embryonic development) for MER41-
112 enhancers are placenta-related (**Figure S3**). Through integration of published epigenomic data for
113 different types of human trophoblast cells (Okae et al. 2018), we further annotated the enhancers
114 for human TSC, syncytiotrophoblast (STB) and extravillous trophoblast (EVT) cells (**Figure S4**).
115 MER41 elements are enriched in the enhancers for all three trophoblast cell types (**Figure 1C**).

Notably, hundreds of MER41-enhancers are specifically activated in one or two trophoblast cell types (**Figure 1D,E**), and their activity correlates with the transcription of at least some adjacent genes, such as *FBN2* (**Figure 1F**). Together, these results confirm the tissue-specific activation of MER41-associated enhancers in human placenta and indicate their potential regulatory effects on some adjacent genes.

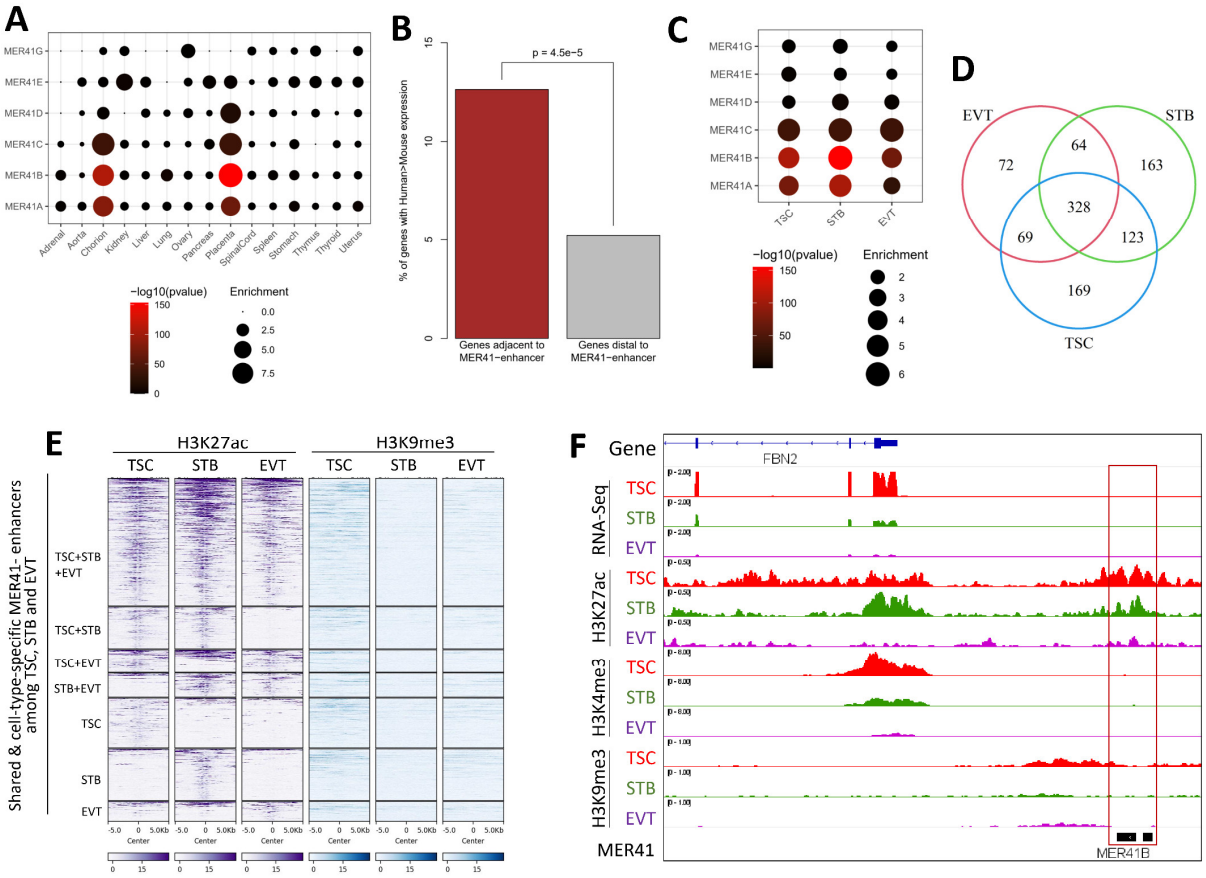

Figure 1. Characterization of MER41-enhancers in different human tissues and trophoblast cells

(A) Enrichment of MER41 subfamilies in the epigenetically-annotated enhancers for different human tissues. **(B)** Relative to distal genes, higher proportion of MER41-enhancer adjacent genes have increased expression in human placenta relative to mouse. **(C)** Enrichment of MER41 subfamilies in the enhancers annotated for human TSC, STB and EVT. **(D)** Overlapping of the MER41-enhancers annotated for human TSC, STB and EVT. **(E)** Heatmap shows the H3K27ac and H3K9me3 intensity for the shared or cell-specific MER41-enhancers for different human trophoblast cells. **(F)** IGV tracks showing the transcriptomic and epigenomic profiles surrounding the representative MER41-enhancer adjacent to *FBN2*.

130 **Identification of GATA2/3 and MSX2 as putative regulators of MER41-enhancers in human**
131 **TSCs**

132 To screen for candidate TFs that bind MER41-enhancers, we first determined the motif
133 composition of MER41B consensus sequence and one representative MER41-enhancer adjacent
134 to *FBN2*. Apart from several TFs (ie. STAT1, IRF1, SRF) already known to bind MER41 elements
135 (Chuong et al. 2016; Sun et al. 2021), the binding motifs for dozens of other TFs are predicted
136 (**Figure 2A,B).** Importantly, four of them, including GATA2/3, MSX2 and GRHL2, have
137 placenta-enriched expression (**Figure 2C, S5**). Particularly, GATA2/3 are key pioneer TFs for
138 human and mouse TSC specification (Home et al. 2017; Krendl et al. 2017; Paul et al. 2017) and
139 MSX2 is essential in restraining STB gene expression in human TSCs (Hornbachner et al. 2021).
140 Even though GRHL2 has not been analyzed in human placenta, it is known to regulate trophoblast
141 branching morphogenesis in mice (Walentin et al. 2015). Therefore, motif prediction and
142 transcriptomic analysis uncovered GATA2/3, MSX2 and GRHL2 as candidate trophoblast TFs
143 that bind MER41-enhancers in human placenta.

144 Recent studies have profiled the genome-wide binding of GATA2/3 (together with
145 TFAP2A/C which are collectively coined as the "trophectoderm four") and MSX2 (together with
146 its co-factors ARID1A and SMARCA4) in human TSCs by using ChIP-Seq (Krendl et al. 2017;
147 Hornbachner et al. 2021). Through re-analysis of these data, we demonstrate that GATA2/3 and
148 MSX2 have substantially overlapped binding (**Figure 2D**), and all these TFs and co-factors have
149 enriched binding on MER41-enhancers in human TSCs (**Figure 2E**). Notably, the majority (>60%)
150 of placental MER41-enhancers are bound by GATA3 and MSX2 (and its co-factors ARID1A and
151 SMARCA4), as indicated by both overlapping analysis and heatmap visualization (**Figure 2F,G**).
152 The remaining TFs also have widespread binding on MER41-enhancers, even though of lower

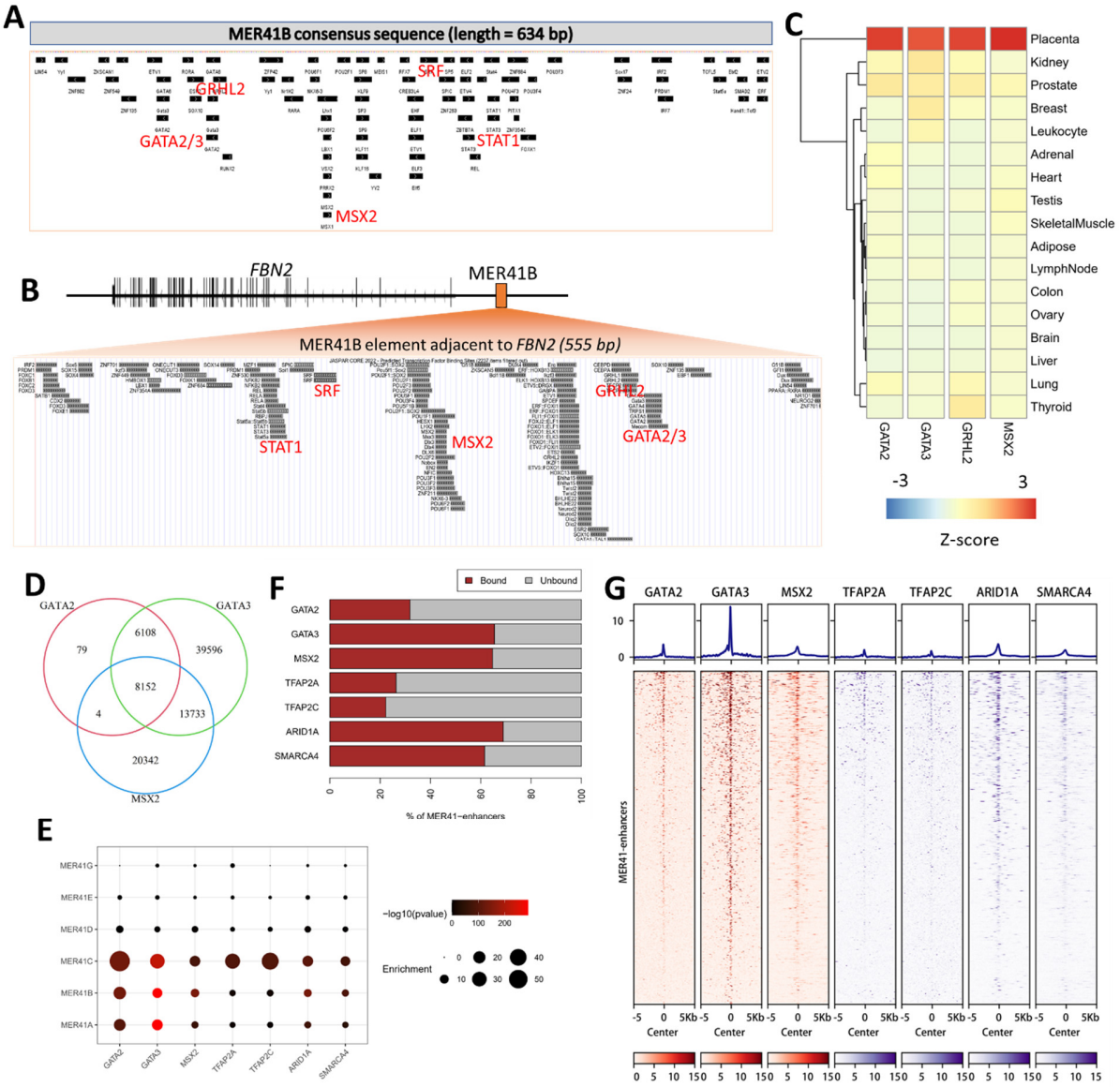

Figure 2. Integrative analysis identified GATA2/3 and MSX2 as candidate regulators of MER41-enhancers in human TSCs

(A,B) Motif configurations of MER41B consensus (A) and one representative MER41-enhancer adjacent to *FBN2* (B) The motifs for several known (STAT1, SRF) and novel (GATA2/3, MSX2, GRHL2) TFs of MER41 are labelled in red color. **(C)** Expression profiles of the four TFs, including *GATA2/3*, *MSX2* and *GRHL2*, that have placenta-enriched expression. The color gradient indicates the column Z-score based on normalized TPM values. **(D)** Venn diagram showing the peak overlapping of GATA2/3 and MSX2. **(E)** Enrichment of MER41 subfamilies in the peaks for GATA2/3, MSX2 and co-factors. **(F)** Bar plots demonstrate the percentages of MER41-enhancers that are bound by GATA2/3, MSX2 and co-factors. (**G**) Heatmaps demonstrate the prevalent binding of GATA2/3, MSX2 and co-factors on MER41-enhancers in human TSCs.

164 frequency (**Figure 2F,G**). These results were further confirmed by manual inspection of a few
165 representative loci, such as *FBN2*, and *C1QTNF6* (**Figure S6**). Together, our integrative analyses
166 identified several crucial trophoblast TFs, including GATA2/3 and MSX2, as putative regulators
167 of MER41-enhancers in human TSCs.

168 **Direct comparison of MER41-enhancers that are constitutively-active or IFN-stimulated in**
169 **human TSCs**

170 Intrigued by previous findings that MER41 elements facilitate the evolution of lineage-specific
171 enhancers that are either constitutively activated in normal placenta (Sun et al. 2021) or interferon-
172 stimulated in different human cells (Chuong et al. 2016), we directly compared between placenta-
173 and immune-related MER41-enhancers using human TSC as model. For this purpose, we
174 performed transcriptomic and epigenomic (H3K27ac) profiling of human TSCs with or without
175 IFN-γ stimulation (**Figure 3A, Table S1**). As expected, IFN-γ stimulates hundreds of genes in
176 human TSCs which are highly associated with immune response (**Figure S7**). We next determined
177 thousands of IFN-stimulated enhancers (**Figure S8**) and further examined their overlap with
178 MER41 subfamilies. While MER41A/B/C are enriched with similar fold within constitutively-
179 active enhancers, MER41B is enriched within IFN-stimulated enhancers to a higher degree
180 (**Figure 3B**). Impressively, the constitutively-active enhancers – in contrast to IFN-stimulated ones
181 – are indeed unresponsive to IFN-stimulation (**Figure 3C,D**), and closer inspection of
182 representative loci adjacent to *FBN2* and *AIM2* confirmed their distinct IFN-response (**Figure**
183 **3E,F**). As expected, GATA2/3, MSX2 and co-factors bind the constitutively-active group with
184 higher frequency and intensity (**Figure 3G,H)**, partly explains why these MER41-enhancers are
185 activated in human TSCs without IFN-treatment. Together, substantial differences regarding the
186 subfamilies and TF binding exist between the placenta- and immune-related MER41-enhancers,
187 suggesting the functional divergence of MER41 elements during evolution.

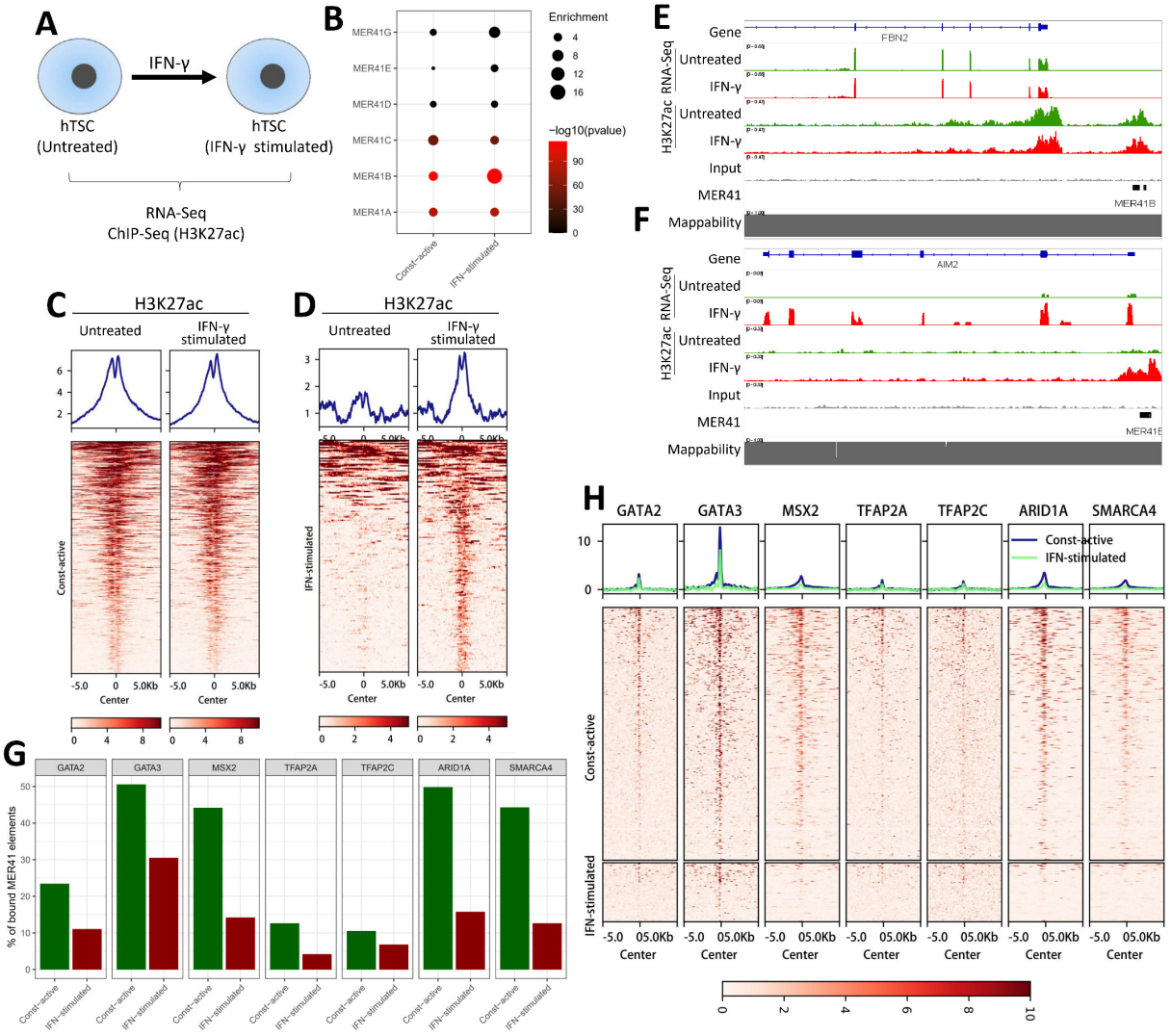

Figure 3. Characterization and comparison of the MER41-enhancers that are constitutively-active or IFN-stimulated in human TSCs

(A) Diagram shows the IFNγ-stimulation related experiments in human TSCs. (**B**) Enrichment of MER41 subfamilies in the two groups of MER41-enhancers that are constitutively-active or IFN-stimulated in human TSCs. (**C,D**) Heatmaps show the H3K27ac intensity flanking the two groups of MER41-enhancers with or without IFNγ-stimulation. **(E,F)** Representative IGV tracks show the alterations of gene expression and H3K27ac occupancy flanking the MER41-enhancers adjacent to *FBN2* and *AIM2*, respectively. **(G)** Bar plots for the binding frequency of GATA2/3, MSX2 and co-factors on the two groups of MER41-enhancers. **(H)** Heatmaps show the binding intensity of GATA2/3, MSX2 and co-factors on the two groups of MER41-enhancers.

200 **Prevalent binding of GATA2/3 and MSX2 on different families of ERV- enhancers in human**
201 **TSCs**

202 Given the enriched binding of the aforementioned trophoblast TFs on MER41-enhancers, we are
203 eager to check if other families of ERV-associated enhancers are also bound by these TFs in human
204 TSCs. Globally, GATA2/3, MSX2 and co-factors preferentially occupy ERVs; specifically, 30.0%
205 and 23.8% of the peaks for GATA2 and GATA3 overlap ERV elements, which is remarkably
206 higher than the 8.8% as expected in randomly shuffled peaks across genome (**Figure 4A**). We
207 further examined their binding on each ERV family, and as expected, MER41A/B/C are among
208 the top enriched (**Figure 4B**). Interestingly, dozens of other ERV families, such as LTR8,
209 MER11A, MER4B, LTR5_Hs and MER4A1/D1/D/E, are also significantly enriched (**Figure 4B**).
210 Notably, many of the enriched ERV families are primate- or even human-specific. Importantly,
211 the ERV families enriched in TSC enhancers and in the peaks for GATA2/3 and MSX2 correlate
212 well (**Figure 4C**), and correspondingly, 49.2% (7,637 out of 15,504) of the TSC enhancers that
213 overlap ERVs are bound by GATA2/3 or MSX2 (**Figure 4D**). Indeed, the top enriched ERV
214 families not only have global binding by GATA3 and MSX2, but also possess high intensity of
215 H3K27ac marks (**Figure 4E**) – suggesting these ERV elements also form enhancers. Closer
216 inspection further confirmed the binding of these TFs on one LTR8-associated enhancer adjacent
217 to the imprinted gene *H19* (**Figure 4F**) and one MER11A-associated enhancer adjacent to *FSTL3*
218 (**Figure 4G**) – these two genes are both tightly associated with placenta development and diseases
219 (Fowden et al. 2006; Xie et al. 2018; Gong et al. 2021). Together, these results suggest that apart
220 from MER41 elements, these core trophoblast TFs also have prevalent binding on many other
221 families of ERV-associated enhancers – suggesting the broad involvement of these TFs and the
222 bound ERV-associated enhancers in human TSCs.

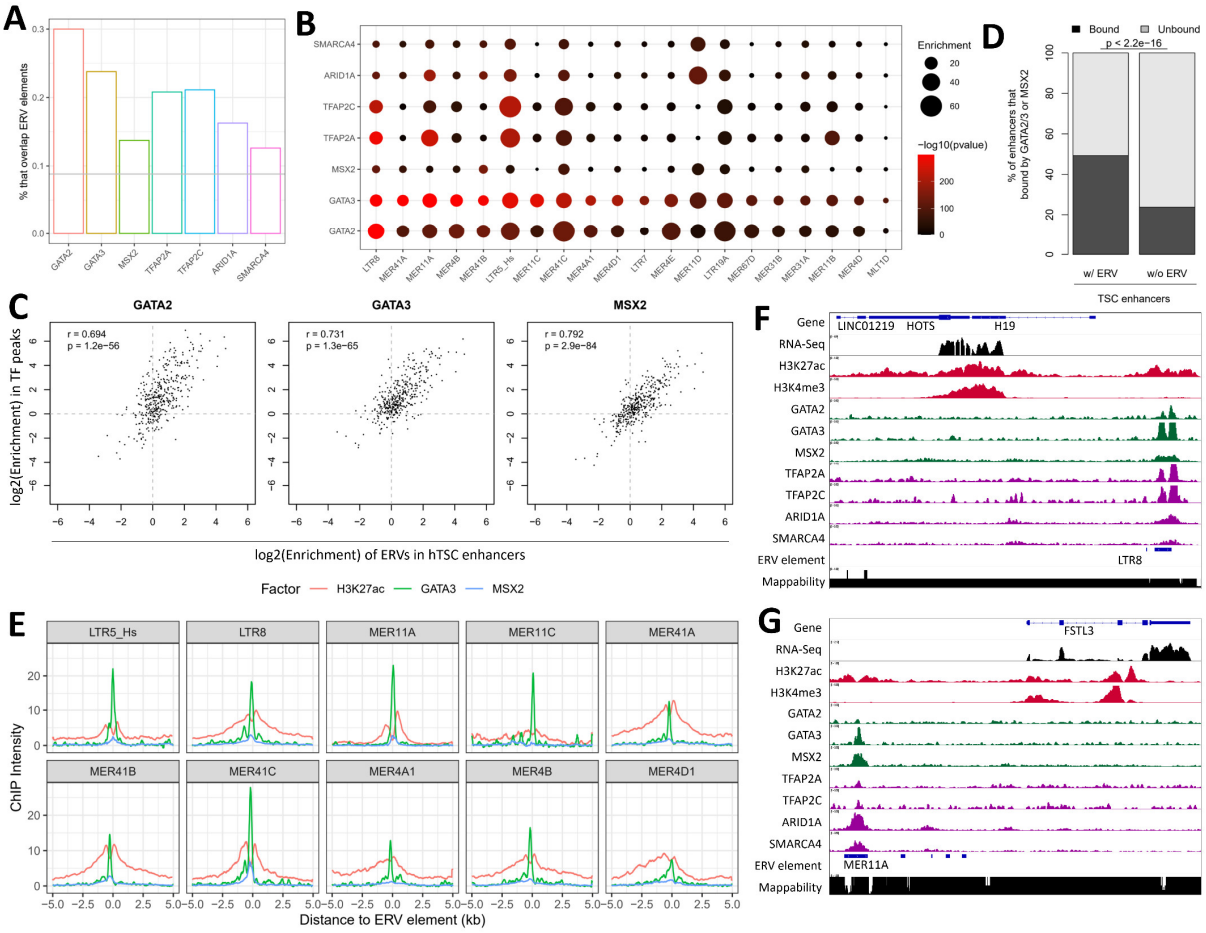

Figure 4. Binding of GATA2/3 and co-factors on different families of ERV-associated enhancers in human TSCs

(A) Percentages of the peaks that binding ERV elements for GATA2/3, MSX2 and co-factors. (**B**) Enrichment of the top twenty ERV families that have enriched binding by GATA2/3, MSX2 or co-factors. These ERV families are selected based on the ranking according to the enrichment p-values against the TF peaks. (**C**) Correlation for the ERV enrichment in annotated enhancers vs. GATA2/3 and MSX2 peaks. **(D)** Comparison of the percentages of enhancers (w/wo ERV overlapping) that are bound by at least one TFs among GATA2/3 and MSX2. P-value calculated by Fisher's Exact Test is denoted. **(E)** Averaged profiles show the intensity of H3K27ac, GATA3 and MSX2 occupancy on the top ten ERV families that have enriched binding by GATA2/3, MSX2 and co-factors. **(F,G)** Representative IGV tracks show the epigenetic profiles and trophoblast TF binding near the LTR8- and MER11A-associated enhancers adjacent to *H19* and *FSTL3*, respectively.

236 **MSX2 represses numerous STB-genes in human TSCs through the binding on ERV-**
237 **associated regulatory elements**

238 Despite the prevalent binding of GATA2/3 and MSX2 on ERV-associated enhancers, it should be
239 noted that these TFs have distinct functions: while GATA2/3 are recognized as pioneer TFs in
240 human TSCs for transcriptional activation (Krendl et al. 2017), MSX2 was recently identified as
241 a transcriptional repressor to restrain the undesired expression of STB-genes in human TSCs
242 (Hornbachner et al. 2021). Given that how ERV elements mediate the function of transcription
243 repressors is rarely reported, in-depth analysis on the binding and function of MSX2 on ERV-
244 associated enhancers is highly desirable. We found that MSX2 has enriched binding on dozens of
245 ERV families, including MER41A/B/C, LTR8, LTR8B, MER21A, and LTR3A (**Figure 5A**). To
246 clarify the function of MSX2 on ERVs, we compared the H3K27ac marks on the top ten enriched
247 ERV families after the knockdown of MSX2. As expected, knockdown of MSX2 causes increased
248 H3K27ac levels on the MSX2-bound ERV elements, yet the degrees of changes differ among ERV
249 families (**Figure 5B**). For example, the H3K27ac level for LTR8B is increased more than three
250 fold, while LTR3A is increased more moderately. These results suggest that MSX2 has a global
251 repressive effect on the bound ERV-enhancers in human TSCs.

252 MSX2 is known to restrict the expression of STB-genes in human TSCs (Hornbachner et
253 al. 2021). To uncover if ERV-associated cis-elements participate in this process, we determined
254 the loci with significantly altered H3K27ac levels after knockdown of MSX2 (MSX2KD) for
255 further analysis. The majority (2,519 out of 2,706) of these differential loci have increased
256 H3K27ac levels (**Figure 5C**), and MSX2 peaks overlap 54.4% of those with increased H3K27ac
257 – significantly higher than those with decreased H3K27ac (**Figure 5D**). Several ERV families are
258 also significantly overrepresented within the MSX2-regulated cis-elements (defined as genomic

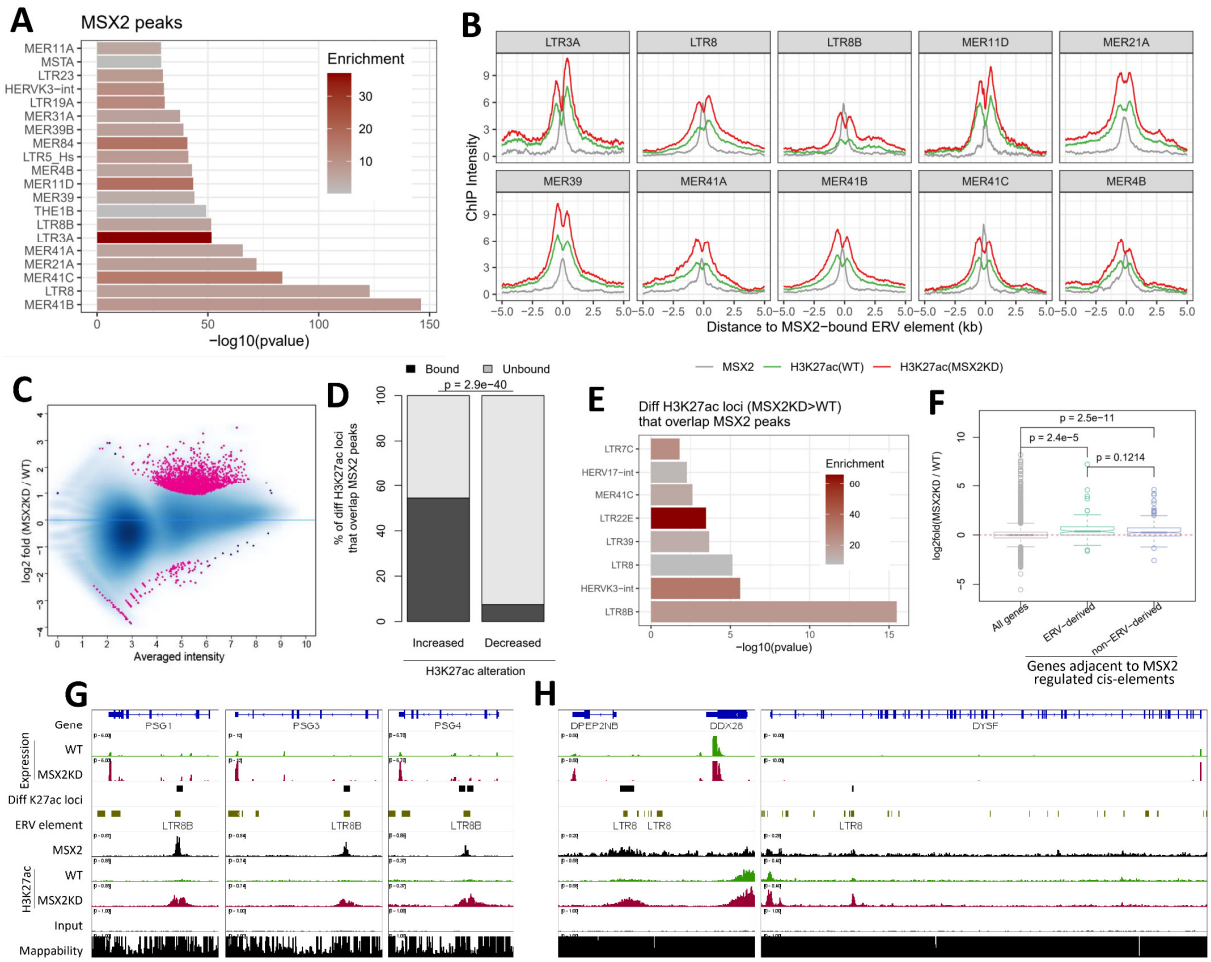

Figure 5. MSX2 restricts the expression of some STB-genes in human TSCs through the binding of ERV-derived enhancers

(A) Enrichment of ERV families within MSX2 peaks. The top 20 families as ranked by p-values are presented. **(B)** Averaged curves show the ChIP intensity of MSX2 and H3K27ac (WT vs. MSX2KD) flanking the MSX2-bound elements from the top ten ERV families. **(C)** MA-plot shows the differential H3K27ac peaks between WT and MSX2KD human TSCs. **(D)** Comparison of the MSX2-bound percentages between genomic loci with increased or decreased H3K27ac levels after MSX2KD. P-value calculated with Fisher's Exact Test is denoted. **(E)** Enrichment of ERV families within MSX2-bound loci that have increased H3K27ac levels after MSX2KD. **(F)** Comparison of the altered expression (MSX2KD vs. WT) across different groups of genes defined based on their association with MSX2-regulated enhancers. A threshold of less than 10 kb from TSSs was used to group genes. MSX2-regulated enhancers are further classified as two groups (ERV-derived or non-ERV-derived) based on their overlapping with ERV elements. P-values calculated from Two-sided Student's t-test are denoted. **(G,H)** Representative IGV tracks showing several LTR8B- or LTR8-derived enhancers that are likely mediating the MSX2-dependent repression of STB-genes, including *PSG1/3/4*, *DPEP2NB*, *DDX28* and *DYSF*.

274 loci that overlap MSX2 peaks and have increased H3K27ac levels in MSX2KD), including LTR8B,
275 HERVK3-int and LTR8 (**Figure 5E**). We further examined the genes adjacent to MSX2-regulated
276 cis-elements which are further classified as ERV-derived or non-ERV-derived, and as expected,
277 they both have increased expression after MSX2KD (**Figure 5F**). Manual inspection of the
278 canonical STB-genes repressed by MSX2 indicates that some are likely to be regulated through
279 ERV-derived cis-elements. For example, multiple *PSG* genes (PSG1/3/4/6/8/9) that get
280 derepressed in MSX2KD harbor LTR8B-derived cis-elements that underlie the regulation of
281 MSX2 (**Figure 5G, S9**). In addition, one LTR8-derived cis-elements between *DPEP2NB* and
282 *DDX28*, and another within *DYSF* are also likely regulated by MSX2 (**Figure 5H**). Together, these
283 results suggest that MSX2 restricts the undesired expression of many STB-genes in human TSCs
284 through ERV-associated regulatory elements.

285 **Discussion**

286 In eutherian mammals, the placenta is a transient organ characterized with extraordinary
287 phenotypic diversity (Gerri et al. 2020; Hemberger et al. 2020). At the molecular level, hundreds
288 of placental genes have lineage-specific expression (Dunn-Fletcher et al. 2018; Soncin et al. 2018;
289 Rosenkrantz et al. 2021; Sun et al. 2021), and many are driven by the fast-evolving enhancers
290 (Chuong et al. 2013; Sun et al. 2021). This study focused on endogenous retroviruses, a class of
291 transposons that are known to facilitate the lineage-specific evolution of placental enhancers in
292 both rodents and humans (Chuong et al. 2013; Sun et al. 2021). We recently demonstrated that
293 dozens of ERV families – many are primate-specific – create numerous human placental enhancers
294 (Sun et al. 2021), yet the underlying regulatory mechanism remains to be explored. Furthermore,
295 it is also highly desirable to delineate how ERV-associated enhancers participate in the trophoblast
296 gene network, which is under the cooperative regulation of many TFs – including some with

297 lineage-specific expression (Papuchova and Latos 2022). Starting with in-depth inspection of the
298 primate-specific MER41 family followed by comprehensive analysis on candidate TFs and the
299 entire ERVome, this study aims to better understand the regulation and function of ERV-associated
300 enhancers in the human trophoblast regulatory network.

301 Our study was initiated by focusing on the primate-specific MER41 family, which is
302 among the top enriched in human placental enhancers (Sun et al. 2021). MER41 elements also
303 known to create thousands of interferon-stimulated enhancers to drive innate immunity evolution
304 (Schmid and Bucher 2010; Chuong et al. 2016). Yet unlike immune-related MER41 elements
305 which have already been well characterized, how MER41-associated enhancers are regulated in
306 human placenta is still mystery. Interestingly, motif scanning indicates that MER41 harbors the
307 binding motifs for multiple trophoblast TFs, including GATA2/3, MSX2 and GRHL2 which are
308 important regulators for trophoblast lineage maintenance and development in human and/or mice
309 (Ralston et al. 2010; Walentin et al. 2015; Home et al. 2017; Krendl et al. 2017; Rhee et al. 2017;
310 Hornbachner et al. 2021). Impressively, integration of public data (Krendl et al. 2017; Hornbachner
311 et al. 2021) verified that the majority (>60%) of MER41-enhancers are bound by GATA3 and
312 MSX2 (and its co-factors ARID1A and SMARCA4), much more frequent than SRF which also
313 binds dozens of MER41-enhancers (Sun et al. 2021). Therefore, the trophoblast TFs, particularly
314 GATA2/3 and MSX2, are identified as promising regulators of MER41-enhancers in human TSCs.
315 Functionally, it is tempting to speculate that, through the binding of GATA2/3, MSX2 and co-
316 factors, MER41-enhancers particulate in the human trophoblast regulatory network.

317 While MER41 is among the top enriched in human placental enhancers, dozens of other
318 ERV families are also highly enriched – therefore it is eager to see if they are regulated through
319 the same set of TFs. Strikingly, we found that GATA2/3 and MSX2 have prevalent binding on

320 dozens of other families of ERV-associated enhancers, many are primate-specific (eg. LTR8,
321 MER4 and MER11) or even human-specific (eg. LTR5_Hs). Indeed, the ERV families frequently
322 bound by these TFs are well correlated with those enriched in TSC enhancers, suggesting
323 GATA2/3 and MSX2 have broad effect on ERV-associated TSC enhancers. Notably, GATA2/3
324 are generally recognized as transcriptional activators (Takaku et al. 2016; Krendl et al. 2017),
325 while MSX2 as a transcriptional repressor for STB genes (Hornbachner et al. 2021). In addition,
326 GATA2 and GATA3 are known to have functional redundancy in mouse placenta (Home et al.
327 2017). Therefore, the cooperation or even competition of these TFs in regulating ERV-associated
328 enhancers is expected. Currently, the molecular functions of these TFs have still not been
329 systematically compared. We expect that the combinatorial function of these TFs on the co-binding
330 loci – including ERV-derived ones – will be clarified through the knockout of each TF and the
331 manipulation of their binding motifs in representative cis-elements.

332 Even though GATA2/3 and MSX2 are most abundant in placenta, they also express and
333 function in a few other tissues or cell types. For example, GATA3 is also expressed in T cells (Wei
334 et al. 2011; Van de Walle et al. 2016) and mammary gland (Theodorou et al. 2013). So, why the
335 ERV-associated enhancers – which are frequently bound by these TFs – are preferably active in
336 placenta instead of other tissues? One possible reason is the unique epigenetic and chromatin
337 signatures owned by placenta. Unlike most other tissues which have global hyper-methylation,
338 placenta is abundant with partially methylated domains (Schroeder et al. 2013; Decato et al. 2017)
339 which are also featured by many tumors (Hansen et al. 2011). Mechanistically, it is speculated that
340 placenta and tumor share converged DNA methylation pathways that mediate the establishment of
341 such epigenetic features (Lorincz and Schubeler 2017). Given the well-recognized function of
342 DNA methylation for epigenetic repression, it is possible that many genomic loci (particularly

343 ERV-derived ones) are more accessible in the placenta relative to other tissues – thus could be
344 bound and activated easier in placenta. The derepression of ERVs has also been observed in
345 numerous types of tumors, presumably due to the attenuation of the epigenetic silencing (Chuong
346 et al. 2017; Ito et al. 2020). Therefore, it is possible that the prevalent binding of trophoblast TFs
347 on ERV-associated enhancers may also reply on the relatively relaxed epigenetic environment in
348 trophoblast cells. Apart from the epigenetic/chromatin environment, additional mechanisms, such
349 as the existence of specific co-factors, may also contribute to the tissue-specific binding and
350 regulation of ERV-associated enhancers by these TFs in placenta.

351 Most previous studies about ERV-derived enhancers focused on their activation, yet how
352 they mediate the function of transcription repressors is rarely reported. Aiming at mechanistic
353 insights into the involvement of ERVs for transcriptional repression, we performed in-depth
354 analysis on MSX2, which is known to be crucial in restraining the undesired expression of STB-
355 genes in human TSCs (Hornbachner et al. 2021). Interestingly, the knockdown of MSX2 resulted
356 in globally increased H3K27ac levels on the ERV-associated enhancers it bound, which suggests
357 that it has global repressive effect on the bound ERV-enhancers. Importantly, for the genomic loci
358 under the repression of MSX2, several ERV families are also significantly overrepresented. Genes
359 adjacent to MSX2-bound enhancers – no matter they are ERV-derived or not - usually have
360 increased expression after MSX2KD. Furthermore, some STB-genes are likely to be regulated
361 through ERV-derived enhancers, such as LTR8B-derived enhancers for multiple *PSG* genes, an
362 LTR8-derived adjacent to *DPEP2NB*, *DDX28* and *DYSF*. Therefore, MSX2 restricts the undesired
363 expression of many STB-genes in human TSCs through ERV-associated enhancers. These results
364 further suggest that through the recruiting of different transcriptional activators and repressors,

365 ERVs can be co-opted as a cis-regulatory module to mediate the accurate expression of associated
366 genes in the host.

367 Taken together, this study uncovered the prevalent binding of multiple trophoblast TFs,
368 particularly GATA2/3, MSX2 and their co-factors, on numerous ERV-derived enhancers in human
369 TSCs. Through comprehensive analysis on the links between ERV-derived enhancers, upstream
370 regulators and downstream targets, this study provides novel mechanistic insights into the
371 functional involvement of ERVs in the human trophoblast regulatory network.

372 Materials and methods

373 Human TSC culture and treatment

374 Human TSCs derived from human cytotrophoblast cells (Okae et al. 2018) were a gift from the
375 Okae lab. They were cultured in trophoblast stem cell medium (TSM), as described previously
376 (Sun et al. 2021). Interferon stimulation was performed by adding 1000U/mL of IFN-γ (PBL,
377 11500-2) into the culturing medium of human TSCs. After 24 hours, untreated and IFN-γ treated
378 samples were both collected for subsequent experiments.

379 RNA-Seq

380 Total RNA for human TSCs was extracted using RNeasy Micro kit (Qiagen, 74004) with on-
381 column DNase digestion, and then submitted for library construction by TruSeq stranded mRNA
382 sample preparation kit (Illumina). RNA-Seq libraries were sequenced as 75 bp paired-end reads
383 with HiSeq2500 (Illumina) platform.

384 Raw reads were trimmed with Trim Galore v0.6.4
385 (https://github.com/FelixKrueger/TrimGalore). Transcript Per Million (TPM) values were

386 calculated with RSEM v1.3.2 (Li and Dewey 2011). To perform differential expression analysis,
387 we aligned trimmed reads to the reference genome (GRCh38 for human) using STAR v2.7.3
388 (Dobin et al. 2013), and then obtained gene-level read counts using the *featureCount* function from
389 subread v2.0.0 (Liao et al. 2013). At last, differentially expressed genes were identified using
390 DESeq2 v1.30.1 (Love et al. 2014) with the cutoff: FDR<0.05 and |log2Foldchange|>1.

391 **ChIP-Seq**

392 ChIP-Seq for human TSCs was performed following our previous study (Sun et al. 2021).
393 Chromatin fragmentation was performed using Diagenode Bioruptor Plus Sonicator. The antibody
394 for H3K27ac (Abcam, ab4729) is used. The amount of chromatin is 20 μg per reaction. ChIP-Seq
395 libraries were constructed using Takara SMARTer ThruPLEX DNA-Seq Kit (Takara, R400674),
396 and sequenced as 50 bp paired-end reads with HiSeq2500 (Illumina) platform.

397 Reads were trimmed with TrimGalore v0.6.4 and then aligned to the corresponding reference
398 genome (GRCh38 for human) using Bowtie v2.3.5 (Langmead and Salzberg 2012) with default
399 settings. PCR duplicates were removed using the *rmdup* function of samtools v1.13 (Li et al. 2009).
400 After confirming the data reproducibility, reads from biological replicates were pooled together
401 for further analysis. Peak calling was performed with MACS v2.2.6 (Zhang et al. 2008). The peaks
402 were further cleaned by removing those that overlap ENCODE Blacklist V2 regions (Amemiya et
403 al. 2019). Differential binding analysis was performed using DiffBind v3.4.11 (Ross-Innes et al.
404 2012) with settings: minOverlap = 1, summits = 400, method = DBA_EDGER.

405 **Reference genome and annotation**

406 Reference genome and gene annotation for human (GRCh38) were downloaded from the
407 ENSEMBL database (release 102) (Yates et al. 2020). Transposable element annotations were

408 downloaded from the RepeatMasker website (http://www.repeatmasker.org/) on May 27, 2016.
409 The clades for ERV families were obtained from the Dfam database (Hubley et al. 2016). Genome
410 mappability along the reference genome was calculated using the GEM-mappability program from
411 GEM (GEnome Multitool) suite (Derrien et al. 2012).

412 **Gene ontology enrichment analysis**

413 Gene ontology enrichment analyses for differentially expressed genes were performed using
414 Metascape (Zhou et al. 2019). Gene Ontology enrichment analyses for genomic regions (eg. peaks
415 and putative enhancers) were performed with GREAT (McLean et al. 2010).

416 **Motif analysis**

417 Motif scanning on MER41 consensus was performed using FIMO (Grant et al. 2011). Motif
418 occurrence on representative MER41 elements was retrieved directly from UCSC Genome
419 Browser (Lee et al. 2020).

420 **Epigenetic-annotation of regulatory elements**

421 Putative regulatory elements are defined based on histone modifications and genomic distribution.
422 Promoters are defined as H3K4me3 occupied regions, and enhancers as H3K27ac peaks that are
423 more than 500 bp from TSSs.

424 **ERV enrichment analysis**

425 To determine if certain ERV families are overrepresented within given genomic regions (eg.
426 enhancers or peaks), we adopted the *window* function of BEDtools v2.29.2 (Quinlan and Hall 2010)
427 to determine the enrichment fold and p-values by using Fisher's Exact Test. To control for Family-
428 Wise Error Rate, the calculated p-values were further adjusted with Bonferroni method.

429 **Statistical analysis and data visualization**

430 All statistical analyses were performed with R statistical programming language (Team 2020).
431 Heatmaps for ChIP-Seq data were generated using DeepTools v3.5.1 (Ramirez et al. 2014).
432 Heatmaps from gene expression clustering analysis were generated using pheatmap
433 (https://github.com/raivokolde/pheatmap). RNA-Seq and ChIP-Seq tracks were visualized using
434 IGV v2.11.1 (Thorvaldsdottir et al. 2013).

435 Data access

436 All raw and processed sequencing data generated in this study have been submitted to the NCBI
437 Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number
438 GSE209541.

439 Competing interest statement

440 The authors declare no competing interests.

441 Acknowledgements

442 This study was supported by the grants from the National Natural Science Foundation of China
443 (31900422 to MAS; 82130047 & 81971414 to BC), the National Key Research and Development
444 Program of China (2018YFC1004400 to BC), the Intramural Research Program of the NICHD,
445 NIH (TSM), the Priority Academic Program Development of Jiangsu Higher Education
446 Institutions (PAPD), and the Natural Sciences Foundation of Fujian Province of China
447 (2020J06003 to BC). We thank Dr. Hiroaki Okae and Dr. Takahiro Arima for kindly providing
448 human trophoblast stem cells. This study utilized the computational resources of Yangzhou
449 University College of Veterinary Medicine High-Performance Computing cluster.

450 # References

451 Amemiya HM, Kundaje A, Boyle AP. 2019. The ENCODE Blacklist: Identification of Problematic
452 Regions of the Genome. *Sci Rep* **9**: 9354.
453 Ander SE, Diamond MS, Coyne CB. 2019. Immune responses at the maternal-fetal interface. *Sci*
454 *Immunol* **4**.
455 Buttler CA, Chuong EB. 2022. Emerging roles for endogenous retroviruses in immune epigenetic
456 regulation. *Immunol Rev* **305**: 165-178.
457 Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-
458 option of endogenous retroviruses. *Science* **351**: 1083-1087.
459 Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from
460 conflicts to benefits. *Nat Rev Genet* **18**: 71-86.
461 Chuong EB, Rumi MA, Soares MJ, Baker JC. 2013. Endogenous retroviruses function as species-
462 specific enhancer elements in the placenta. *Nat Genet* **45**: 325-329.
463 Decato BE, Lopez-Tello J, Sferruzzi-Perri AN, Smith AD, Dean MD. 2017. DNA Methylation
464 Divergence and Tissue Specialization in the Developing Mouse Placenta. *Mol Biol Evol*
465 **34**: 1702-1712.
466 Deniz O, Frost JM, Branco MR. 2019. Regulation of transposable elements by DNA
467 modifications. *Nat Rev Genet* **20**: 417-431.
468 Derrien T, Estelle J, Marco Sola S, Knowles DG, Raineri E, Guigo R, Ribeca P. 2012. Fast
469 computation and applications of genome mappability. *PLoS One* **7**: e30377.
470 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR.
471 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.
472 Dunn-Fletcher CE, Muglia LM, Pavlicev M, Wolf G, Sun MA, Hu YC, Huffman E, Tumukuntala S,
473 Thiele K, Mukherjee A et al. 2018. Anthropoid primate-specific retroviral element THE1B
474 controls expression of CRH in placenta and alters gestation length. *PLoS Biol* **16**:
475 e2006337.
476 Fowden AL, Sibley C, Reik W, Constancia M. 2006. Imprinted genes, placental development and
477 fetal growth. *Horm Res* **65 Suppl 3**: 50-58.
478 Gerri C, McCarthy A, Alanis-Lobato G, Demtschenko A, Bruneau A, Loubersac S, Fogarty NME,
479 Hampshire D, Elder K, Snell P et al. 2020. Initiation of a conserved trophectoderm
480 program in human, cow and mouse embryos. *Nature* **587**: 443-447.
481 Gong S, Gaccioli F, Dopierala J, Sovio U, Cook E, Volders PJ, Martens L, Kirk PDW, Richardson S,
482 Smith GCS et al. 2021. The RNA landscape of the human placenta in health and disease.
483 *Nat Commun* **12**: 2639.
484 Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif.
485 *Bioinformatics* **27**: 1017-1018.
486 Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y,
487 Diep D et al. 2011. Increased methylation variation in epigenetic domains across cancer
488 types. *Nat Genet* **43**: 768-775.
489 Hemberger M, Hanna CW, Dean W. 2020. Mechanisms of early placental development in
490 mouse and humans. *Nat Rev Genet* **21**: 27-43.
491 Home P, Kumar RP, Ganguly A, Saha B, Milano-Foster J, Bhattacharya B, Ray S, Gunewardena S,
492 Paul A, Camper SA et al. 2017. Genetic redundancy of GATA factors in the

493 extraembryonic trophoblast lineage ensures the progression of preimplantation and
494 postimplantation mammalian development. *Development* **144**: 876-888.
495 Hornbachner R, Lackner A, Papuchova H, Haider S, Knofler M, Mechtler K, Latos PA. 2021. MSX2
496 safeguards syncytiotrophoblast fate of human trophoblast stem cells. *Proc Natl Acad Sci*
497 *U S A* **118**.
498 Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, Smit AF, Wheeler TJ. 2016. The Dfam
499 database of repetitive DNA families. *Nucleic Acids Res* **44**: D81-89.
500 Ito J, Kimura I, Soper A, Coudray A, Koyanagi Y, Nakaoka H, Inoue I, Turelli P, Trono D, Sato K.
501 2020. Endogenous retroviruses drive KRAB zinc-finger protein family expression for
502 tumor suppression. *Sci Adv* **6**.
503 Johnson WE. 2019. Origins and evolutionary consequences of ancient endogenous retroviruses.
504 *Nat Rev Microbiol* **17**: 355-370.
505 Kojima KK. 2018. Human transposable elements in Repbase: genomic footprints from fish to
506 humans. *Mob DNA* **9**: 2.
507 Krendl C, Shaposhnikov D, Rishko V, Ori C, Ziegenhain C, Sass S, Simon L, Muller NS, Straub T,
508 Brooks KE et al. 2017. GATA2/3-TFAP2A/C transcription factor network couples human
509 pluripotent stem cell differentiation to trophectoderm with repression of pluripotency.
510 *Proc Natl Acad Sci U S A* **114**: E9579-E9588.
511 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:
512 357-359.
513 Lee CM, Barber GP, Casper J, Clawson H, Diekhans M, Gonzalez JN, Hinrichs AS, Lee BT, Nassar
514 LR, Powell CC et al. 2020. UCSC Genome Browser enters 20th year. *Nucleic Acids Res* **48**:
515 D756-D761.
516 Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or
517 without a reference genome. *BMC Bioinformatics* **12**: 323.
518 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
519 Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and
520 SAMtools. *Bioinformatics* **25**: 2078-2079.
521 Liao Y, Smyth GK, Shi W. 2013. The Subread aligner: fast, accurate and scalable read mapping by
522 seed-and-vote. *Nucleic Acids Res* **41**: e108.
523 Lorincz MC, Schubeler D. 2017. Evidence for Converging DNA Methylation Pathways in Placenta
524 and Cancer. *Dev Cell* **43**: 257-258.
525 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
526 RNA-seq data with DESeq2. *Genome Biol* **15**: 550.
527 Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono
528 D, Pfaff SL. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus
529 activity. *Nature* **487**: 57-63.
530 Maltepe E, Fisher SJ. 2015. Placenta: the forgotten organ. *Annu Rev Cell Dev Biol* **31**: 523-552.
531 McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010.
532 GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**:
533 495-501.
534 Mossman HW. 1987. *Vertebrate fetal membranes : comparative ontogeny and morphology,*
535 *evolution, phylogenetic significance, basic functions, research opportunities*. New
536 Brunswick, N.J. : Rutgers University Press.

537 Okae H, Toh H, Sato T, Hiura H, Takahashi S, Shirane K, Kabayama Y, Suyama M, Sasaki H, Arima
538 T. 2018. Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell* **22**: 50-63 e56.
539 Papuchova H, Latos PA. 2022. Transcription factor networks in trophoblast development.
540 *Cellular and Molecular Life Sciences* **79**.
541 Paul S, Home P, Bhattacharya B, Ray S. 2017. GATA factors: Master regulators of gene
542 expression in trophoblast progenitors. *Placenta* **60 Suppl 1**: S61-S66.
543 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
544 *Bioinformatics* **26**: 841-842.
545 Ralston A, Cox BJ, Nishioka N, Sasaki H, Chea E, Rugg-Gunn P, Guo G, Robson P, Draper JS,
546 Rossant J. 2010. Gata3 regulates trophoblast development downstream of Tead4 and in
547 parallel to Cdx2. *Development* **137**: 395-403.
548 Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. 2014. deepTools: a flexible platform for
549 exploring deep-sequencing data. *Nucleic Acids Res* **42**: W187-191.
550 Ramsey EM, Houston ML, Harris JW. 1976. Interactions of the trophoblast and maternal tissues
551 in three closely related primate species. *Am J Obstet Gynecol* **124**: 647-652.
552 Rhee C, Lee BK, Beck S, LeBlanc L, Tucker HO, Kim J. 2017. Mechanisms of transcription factor-
553 mediated direct reprogramming of mouse embryonic stem cells to trophoblast stem-like
554 cells. *Nucleic Acids Res* **45**: 10103-10114.
555 Rosenkrantz JL, Gaffney JE, Roberts VHJ, Carbone L, Chavez SL. 2021. Transcriptomic analysis of
556 primate placentas and novel rhesus trophoblast cell lines informs investigations of
557 human placentation. *BMC Biol* **19**: 127.
558 Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, Brown GD, Gojis O,
559 Ellis IO, Green AR et al. 2012. Differential oestrogen receptor binding is associated with
560 clinical outcome in breast cancer. *Nature* **481**: 389-393.
561 Schmid CD, Bucher P. 2010. MER41 repeat sequences contain inducible STAT1 binding sites.
562 *PLoS One* **5**: e11425.
563 Schroeder DI, Blair JD, Lott P, Yu HO, Hong D, Crary F, Ashwood P, Walker C, Korf I, Robinson
564 WP et al. 2013. The human placenta methylome. *Proc Natl Acad Sci U S A* **110**: 6037-
565 6042.
566 Senft AD, Macfarlan TS. 2021. Transposable elements shape the evolution of mammalian
567 development. *Nat Rev Genet* **22**: 691-711.
568 Sharif J, Endo TA, Nakayama M, Karimi MM, Shimada M, Katsuyama K, Goyal P, Brind'Amour J,
569 Sun MA, Sun Z et al. 2016. Activation of Endogenous Retroviruses in Dnmt1(-/-) ESCs
570 Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to
571 Hemimethylated DNA. *Cell Stem Cell* **19**: 81-94.
572 Soncin F, Khater M, To C, Pizzo D, Farah O, Wakeland A, Arul Nambi Rajan K, Nelson KK, Chang
573 CW, Moretto-Zita M et al. 2018. Comparative analysis of mouse and human placentae
574 across gestation reveals species-specific regulators of placental development.
575 *Development* **145**.
576 Stoye JP. 2012. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat*
577 *Rev Microbiol* **10**: 395-406.
578 Sun MA, Wolf G, Wang Y, Senft AD, Ralls S, Jin J, Dunn-Fletcher CE, Muglia LJ, Macfarlan TS.
579 2021. Endogenous retroviruses drive lineage-specific regulatory evolution across
580 primate and rodent placentae. *Mol Biol Evol* **38**: 4992-5004.

581 Takaku M, Grimm SA, Shimbo T, Perera L, Menafra R, Stunnenberg HG, Archer TK, Machida S,
582 Kurumizaka H, Wade PA. 2016. GATA3-dependent cellular reprogramming requires
583 activation-domain dependent recruitment of a chromatin remodeler. *Genome Biol* **17**:
584 36.
585 Team RC. 2020. R: A language and environment for statistical computing. *R Foundation for*
586 *Statistical Computing, Vienna, Austria*.
587 Theodorou V, Stark R, Menon S, Carroll JS. 2013. GATA3 acts upstream of FOXA1 in mediating
588 ESR1 binding by shaping enhancer accessibility. *Genome Res* **23**: 12-22.
589 Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-
590 performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178-192.
591 Van de Walle I, Dolens AC, Durinck K, De Mulder K, Van Loocke W, Damle S, Waegemans E, De
592 Medts J, Velghe I, De Smedt M et al. 2016. GATA3 induces human T-cell commitment by
593 restraining Notch activity and repressing NK-cell fate. *Nat Commun* **7**: 11171.
594 Walentin K, Hinze C, Werth M, Haase N, Varma S, Morell R, Aue A, Potschke E, Warburton D,
595 Qiu A et al. 2015. A Grhl2-dependent gene network controls trophoblast branching
596 morphogenesis. *Development* **142**: 1125-1136.
597 Wei G, Abraham BJ, Yagi R, Jothi R, Cui K, Sharma S, Narlikar L, Northrup DL, Tang Q, Paul WE et
598 al. 2011. Genome-wide analyses of transcription factor GATA3-mediated gene
599 regulation in distinct T cell types. *Immunity* **35**: 299-311.
600 Wolf G, de Iaco A, Sun MA, Bruno M, Tinkham M, Hoang D, Mitra A, Ralls S, Trono D, Macfarlan
601 TS. 2020. KRAB-zinc finger protein gene expansion in response to active
602 retrotransposons in the murine lineage. *Elife* **9**.
603 Xie J, Xu Y, Wan L, Wang P, Wang M, Dong M. 2018. Involvement of follistatin-like 3 in
604 preeclampsia. *Biochem Biophys Res Commun* **506**: 692-697.
605 Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov
606 AG, Bennett R et al. 2020. Ensembl 2020. *Nucleic Acids Res* **48**: D682-D688.
607 Yu Y, He JH, Hu LL, Jiang LL, Fang L, Yao GD, Wang SJ, Yang Q, Guo Y, Liu L et al. 2020. Placensin
608 is a glucogenic hormone secreted by human placenta. *EMBO Rep* **21**: e49530.
609 Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown
610 M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.
611 Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK.
612 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level
613 datasets. *Nat Commun* **10**: 1523.

614