OPEN ACCESS

Repository of the Max Delbrück Center for Molecular Medicine (MDC) in the Helmholtz Association

https://edoc.mdc-berlin.de/16686

How to tame an endogenous retrovirus: HERVH and the evolution of human pluripotency

Römer, C., Singh, M., Hurst, L.D., Izsvák, Z.

This is the final version of the accepted manuscript. The original article has been published in final edited form in:

Current Opinion in Virology
2017 AUG ; 25: 49-58
2017 JUL 24 (first published online: final publication)
doi: 10.1016/j.coviro.2017.07.001
Publisher: Elsevier

Copyright © 2017, Elsevier. This manuscript version is made available under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

How to tame an endogenous retrovirus: HERVH and the evolution of human pluripotency

Christine Römer, Manvendra Singh, Laurence D. Hurst[#], Zsuzsanna Izsvák

Mobile DNA, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

University of Bath, Department of Biology and Biochemistry, Bath, Somerset, UK, BA2 7AY

Abstract

HERVH is one of the most successful endogenous retrovirus in the human genome. Relative to other endogenous retroviruses, slower degradation of HERVH internal sequences indicates their potential relevance for the host. HERVH is transcriptionally active during human preimplantation embryogenesis. In this review, we focus on the role of HERVH in regulating human pluripotency. The HERVH-mediated pluripotency network has been evolved recently in primates. Nevertheless, it became an essential feature of human pluripotency. We discuss how HERVH modulates the human pluripotency network by providing alternative transcription factor binding sites, functioning as a long-range enhancer, and as being a major source for pluripotency specific long non-coding RNAs and chimeric transcripts.

Highlights

HERVH expression marks human pluripotent stem cells

HERVH provides binding site for key pluripotency transcription factors

HERVH gives rise to pluripotency specific lncRNAs, lincRNAs and chimeric transcripts

HERVH is needed for acquisition and maintenance of pluripotency

HERVH is required for self-renewal and inhibits differentiation

Introduction

Retroviruses have invaded the genome of human ancestors in several waves [1,2]. If a retrovirus successfully integrates in the germline, it can go through an endogenization process. The endogenization process enables their vertical transmission from parent to progeny via germline transmission [3]. The endogenized retroviruses (ERVs) eventually lose their ability to infect and leave the cells [3], and follow the life cycle of transposable elements (TEs) [4,5]. Following an active period of transposition and amplification, ERVs get inactivated transpositionally, and subjected to a degradation process [6]. As viral fossils, ERV sequences were long considered to have no functional relevance in the human genome and indeed many are probably functionless degrading fossils. Intriguingly, ERV-derived sequences can, however, be occasionally co-opted by the host for various cellular processes. As a classical example, ERV-derived fusogenic syncytin genes have important function in placenta development and embryo implantation in mammals [7,8]. Located in the vicinity of certain immune genes, ERV-derived sequences function as interferon-γ inducible enhancers and regulate the transcription of immune genes [9], underlying their involvement in fine-tuning the innate immune responses. Several ERV families show transcriptional activity during early embryogenesis [10,11]. Recently, two ERV families, HERVK and HERVH, have been implicated in a domestication process, being incorporated into the host defense against exogenous viral infection [11] and the regulatory network of pluripotency, respectively [10,12,13].

Endogenous retroviruses as transposable elements

TEs are almost certainly present in all species, with significant contributions to genome structure (in humans, for example, it is 45-65%) [14,15]. The most common TEs in mammalian genomes are the retrotransposons that use a reverse transcription step during transposition [16]. They are classified, based on the presence of a long terminal repeat (LTR), into non-LTR and LTR retrotransposons. [17]. ERVs belong to the latter. Transposition of an intact ERV is initiated by transcriptional activation. The mRNA is then translated into gag and pol polyproteins. The *pol* open reading frame includes a protease, a reverse transcriptase, an RnaseH, and an integrase coding sequence. The gag protein encapsulates these viral proteins and the ERV RNA, which is reverse-transcribed into DNA and is integrated into a target genomic location [4,5,17]. Although transposing ERVs can pose a great mutagenic potential [18,19], there are no confirmed transpositionally active ERVs in the human genome [20-22]. Nevertheless, one of the youngest ERV families, HERVK, has been active recently

61 enough to be polymorphic loci in the human genome [23,24]. By contrast, a few,
62 phylogenetically younger (< 7 million years old) subfamilies of the non-LTR
63 retrotransposons, including Line1 (L1), SVA and Alu elements are still actively jumping [22].

64

65 **HERVH invaded primates genomes in multiple waves**

66 Today, ERV-related sequences make up about 8% of the human genome [15,25]. These
67 sequences belong to different human ERV (HERV) families which are classified by their
68 tRNA primer binding sites (e.g lysine for HERVK, histidine for HERVH etc.) [26]. HERVH
69 integrated into the genome of human ancestors at the time of divergence of Old and New
70 World monkeys, 30-40 million years ago (MYA) [27,28], with an expansion at about 25
71 MYA [27]. The expansion of HERVH coincided with the loss of the *env* gene [29]. This
72 process resulted in high HERVH copy numbers in the Old World monkeys (approximately
73 900 [30]) compared with only a few (25-50 copies) in New World monkeys [28].
74 Interestingly, another wave of invasion of HERVH retrovirus [31] has occurred at the
75 divergence of hominoids between 10 and 17 million years ago, resulting in about 100
76 additional copies [27]. HERVH as retrovirus has been inactivated around 10 MYA [32].

77

78 **HERVH is the most abundant ERV in the human genome**

79 The first HERVH copy, a deleted version of the full-length elements, was identified from the
80 human genome in 1984 [33]. It was a 5-6 kb repetitive sequence that consisted of the 415 bp
81 LTRs, flanking the internal sequence and a histidine tRNA primer binding site that was
82 located directly after the 5'LTR. The regulatory LTR sequences which flank the HERVH
83 internal sequence (HERVH-int) are characteristic to HERVH elements and the HERVH
84 related LTR subtypes are LTR7, LTR7A, LTR7B, LTR7C and LTR7Y [34], representing
85 different evolutionary age. Other than their regulatory LTRs, HERVHs, are further
86 characterized based on their structure into complete, slimmed down, substituted and solo LTR
87 elements [16].

88 The human genome contains around 50-100 copies of the almost intact forms of HERVH with
89 a full-length size of 8.7 kb [35-37]. A few HERVH copies have an *env* open reading frame
90 which, however, has not been found to produce a protein [38]. Although the almost intact
91 HERVH copies have the full repertoire of retroviral genes, these carry several mutations or

92 deletions and are not replication competent [18,35,37]. The vast majority of the HERVH
93 integrations in the human genome originate from a common 5.8 kb form with a structure of
94 5'LTR-*gag*-*pol*-3'LTR [39] (Figure 1) that carries large deletions in its *pol* coding region [28]
95 and lacks the *env* coding region [28,35-37]. The integrity of an inactive ERV is typically not
96 protected, and their sequences are exposed to various degradation processes such as
97 homologous recombination that generates solitary LTR copies, and various insertions and
98 deletions (indels), resulting in fragmentation and even in complete 'stochastic loss'. During
99 the degradation process, the active full-length elements, operating as autonomous
100 retrotransposition machineries, might mobilize the more common and partially deleted non-
101 autonomous forms of ERVs [5,28].

102 What is noteworthy about HERVH compared to other ERV families is the unusually high
103 number of full-length and partially deleted insertions relative to solitary LTRs [40,41],
104 suggesting that the degradation of HERVH copies occur at a much slower rate compared to
105 other ERVs (Figure 2) [40]. With 1060 copies flanked by LTRs and another 1270 copies of
106 solitary LTR sequences per haploid genome [29], HERVH is the HERV family with the
107 highest copy number in the human genome [27,33]. Lack of the *env* coding region, as part of
108 the endogenization process, has been associated with the ability to more effectively amplify in
109 copy number [42], of this ERV within the human genome. The maintenance of HERVH full
110 length or partially deleted forms in the human genome gave rise to the hypothesis that some
111 HERVH sequences might have gained functional relevance for the host [40].

112

113 **HERVH gets transcriptionally activated during human preimplantation embryogenesis**
114 **following an exceptional pattern**

115 Mammalian preimplantation embryogenesis passes the following stages: the fertilized oocyte
116 forms a zygote, which develops into a 2-cell, 4-cell and 8-cell embryo, followed by the
117 morula and then the blastocyst. During these stages the totipotent cells of the zygote become
118 pluripotent within the epiblast of the blastocyst [43]. The pluripotent epiblast can still give
119 rise to ectoderm, mesoderm and endoderm, but no longer to extraembryonic tissues. The
120 process of preimplantation embryogenesis involves rapid, dynamic, and well-orchestrated
121 changes in the epigenome of the embryo. The global epigenetic changes gradually deactivate
122 the transcription of various TE families (including both transpositionally competent and
123 inactive) in a well-defined temporal order [10]. In fact, the transcriptome of TEs differs so

124 dramatically between the stages of embryogenesis, that the cell status can be identified solely
125 based on their expression pattern (Figures 3 and 4) [10,11]. For the sake of genome stability,
126 repression of actively transposing retrotransposons by the host is selectively favored during
127 embryogenesis [44]. Indeed, the host has evolved several layers of regulatory mechanisms to
128 control TE activities. Transcription from a TE locus is epigenetically regulated by DNA
129 methylation and histone modification. The KRAB-ZN finger proteins can specifically
130 recognize TE families and recruit the TRIM28/KAP1 repression complex and induce
131 heterochromatin formation [45]. At the portranscriptional level, TEs can be also controlled by
132 small RNA-mediated silencing or by APOBEC-mediated gene editing [46-49]. Although
133 these regulatory mechanisms were established during the 'arms race' between TEs and the
134 host [44], several of them are still functional (e.g. control the transcription).

135 Initial stages of embryonic development are governed by maternal effects [50], and
136 embryonic genome activation (EGA) in the human embryo occurs later than in mice (2 cell),
137 and only between the 4- and 8-cell stages of development (Figure 3A and 4) [51]. Notably, a
138 massive activation of TEs in both species occurs at the switch from maternal to embryonic
139 genome activation [10]. In human, DNA transposon-derived transcripts are relatively
140 abundant in zygotes and 2-cells stage, but their levels, together with other phylogenetically
141 Old (> 7 MY) TEs, gradually decline as development proceeds (Figure 3B). This might reflect
142 decay of remnant RNAs expressed in oocytes. In human, the transcriptional activation of the
143 Young (< 7 MY) elements, including L1 (L1_Hs) and SVA (SVA_D, E and F), capable of
144 retrotransposition [52] is substantial from 8-cell stage, peaking at morula with a contrasting
145 dynamic to DNMT3A and 3B, but declining in the blastocyst (Figure 3C). "Old" LTR7-
146 HERVH peaking at the blastocyst [10,11] is the one exception to the Old/Young difference,
147 and curiously opposing the expression pattern of Young TEs, including the mutagenic SVA or
148 the human specific HERVK_Hs elements (Figure 3C).

149 Contrary to other TEs, transcriptional activation of HERVH (with various intensity) occurs
150 throughout early human development. Based on the LTR type, the expression of LTR7- and
151 LTR7Y-driven HERVH peaks at the blastocyst [10,11], LTR7B-driven HERVH is
152 transcribed at the 8-cell stage, while HERVH associated LTR7 sequence with a certain 38 bp
153 deletion is activated even before the 8-cell stage of human preimplantation embryogenesis
154 (Figure 4) [10]. Thus, HERVH, driven by distinct LTR variants, is expressed during the entire
155 human preimplantation embryogenesis [10].

156

157 **HERVH expression in human pluripotent stem cells**

158 *In vitro* models of human pluripotency maintenance and differentiation are pluripotent stem
159 cells, including human embryonic stem cells (hESCs) [53] and induced pluripotent stem cells
160 (hiPSCs) [54]. While hESC cultures are established from the pluripotent blastocysts, iPSCs
161 are reprogrammed somatic cells that have regained a pluripotent state. Both, hESCs and
162 hiPSCs are enriched for LTR7-driven HERVH and associated transcripts (Figure 5A and B)
163 [10,12,13,55], indicating that HERVH-derived transcripts are hallmarks of human pluripotent
164 stem cells (hPSCs). HERVH transcripts are highly abundant and account for 2% of total RNA
165 in hPSCs [39] and are concentrated in the nucleus [56]. The transcription of HERVH is
166 supported by open chromatin [57,58], characterized by markers for transcriptionally active
167 promoters like H3K4me1/2/3, H3K9ac, H3K27ac, H3K36me3 and H3K79me2 [12,39,55,59].
168 By contrast, repressive chromatin marks of H3K27me3 and H3K9me3 at the transcriptionally
169 active HERVH loci are rare [12]. Thus, LTR7 has an active function as promoter and
170 enhancer in hESCs [10,58].

171

172 **HERVH provides a platform to key pluripotency transcription factors**

173 How is it that HERVH is transcriptionally active in pluripotent stem cells? In the human
174 genome, one fourth of the all the binding sites for transcription factors modulating
175 pluripotency are provided by TEs [60]. Active HERVH copies have binding sites for four key
176 transcription factors driving pluripotency, such as OCT4, SOX2, LBP9 (TFCP2L1) and
177 NANOG [12,61-63]. Specifically, NANOG has been shown to bind the 5'LTR of HERVH,
178 while OCT4 and SOX2 have binding sites in the HERVH internal sequence towards its 5' end
179 [39].

180 Functioning as a platform for alternative transcription factor binding sites and as long-range
181 enhancers, LTR7/HERVH is expected to affect the transcriptional regulatory network of
182 pluripotency (Figure 6A). In fact, HERVH is indispensable for pluripotency maintenance, as
183 knocking down of HERVH in hPSCs results in loss of pluripotency [12,56,64]. Curiously, the
184 global knockdown effect of HERVH and LBP9 are correlated, suggesting that the functions of
185 LBP9 binding to HERVH and modulating pluripotency might have evolved together [12].
186 Besides regulating pluripotency, the observed correlation argues for an additional function of
187 LBP9 in retroviral control present only in primates (hence the name LTR-binding protein 9).

188

189 HERVH produces pluripotency-specific ncRNAs

190 How might HERVH impact on pluripotency? Noncoding RNAs (ncRNAs) have been
191 frequently found to be involved in regulating developmental processes including pluripotency
192 and differentiation [65-67]. While TEs contribute very little to protein coding sequences [68],
193 they give rise to numerous ncRNAs, including long non-coding RNAs (lncRNAs), long
194 intergenic non-coding RNAs (lincRNAs), microRNAs, etc [69,70]. Intriguingly, TEs are
195 integral to 83% of lincRNAs [68] and thus likely shaped their evolution. Once upregulated,
196 HERVH affects the expression of genes within a 40 kb window [55]. In hPSCs, HERVH
197 serves as a major source of alternative transcripts, regulating pluripotency [58,60]. These
198 HERVH-derived transcripts were recruited in a primate [12,30] or even human-specific
199 manner [30]. Transcriptionally active HERVHs are driving the production of numerous
200 lncRNAs and lincRNAs [12,69-72]. LTR7/HERVH sequences alone account for more than
201 43% of hESC-specific lncRNA promoters [13] and give rise to over a hundred (127)
202 lincRNAs which are robustly and specifically transcribed in hPSCs [68,72].

203 Transcription of HERVH derived lincRNAs requires prior binding of SP1, OCT4 or NANOG
204 to the 5'LTR of HERVH [56,68]. Curiously, HERVH-lncRNAs share a conserved core
205 domain [12], which is capable of recruiting RNA-binding proteins, pluripotency factors and
206 histone modifiers [12,67]. As a scaffold, HERVH-derived nuclear lncRNA interacts with
207 pluripotency factors (e.g. OCT4) and transcriptional co-activators (e.g. p300, mediator
208 subunits MED6 and MED12) [66,67]. This scaffold acts as a feedback regulator, modulates
209 LTR7 enhancer function and the expression of neighboring genes that are essential for hPSCs
210 identity (Figure 6B) [56]. An interesting example of HERVH-lincRNA is the linc-Regulator
211 of Reprogramming (linc-RoR) that supports pluripotency by functioning as a miRNA sponge.
212 Linc-RoR shares miRNA response elements with core pluripotency transcription factors
213 OCT4, SOX2 and NANOG, thus protecting them from miRNA-mediated decay (Figure 6C)
214 [73].

215

216 HERVH produces chimeric transcripts between cellular and viral sequences

217 Another means by which HERVH regulates pluripotency maintenance in hPSCs is via the
218 generation of various chimeric transcripts thereby combining cellular and viral sequences.

219 HERVH contains a conserved splice donor site that can connect the retroviral element with
220 splice acceptor sites of cellular protein-coding genes (Figure 1) [10]. With a transcriptional
221 start site (TSS) between the 5'LTR and HERVH sequence, these chimeric transcripts often
222 lack their 5' exon(s) of the canonical version, while part of HERVH can be exonized.

223 Examples of HERVH-enforced chimeras include SCGB3A2, RPL39L, RP11-69I8.2NCR1,
224 OC90 and CALB-KLKB1 [12]. We assume that the function of novel HERVH-enforced
225 chimeras may be related to the original gene, but modified. However, in certain cases the
226 LTR7/HERVH-enforced chimeric gene model is so robustly altered that it is not even
227 possible to predict its new function.

228 An interesting example of a HERVH enforced novel transcript, part HERVH, part host DNA,
229 is ESRG that has a putative open reading frame (or frames) only in humans [12,74]. ESRG is
230 expressed in hPSCs [12,74,75], and has been shown to promote the reprogramming process
231 [12]. A knockdown of ESRG hampers the self-renewal potential and pluripotency of hPSCs
232 [12]. The case of the ESRG gene is extreme as it consists almost entirely of repetitive
233 sequences [12]. The non-HERVH sequence recruited is intronic DNA from the host gene
234 within which ESRG resides. The ability of HERVH to generate a great degree of diversity via
235 its heterogeneous transcripts can in part explain why selection may have favoured the
236 preservation of some HERVH-associated transcripts. With such a diversity it is more likely
237 that a few will evolve new functions in the cells in which they are easily expressed.

238

239 **HERVH promotes somatic cell reprogramming**

240 Human somatic cells can be reprogrammed to become pluripotent via different routes. The
241 classical way is the forced expression of pluripotency factors OCT4, SOX2, KLF4 and c-
242 MYC (referred to as the OSKM factors) [12,54]. Alternatively, hiPSCs can be generated by
243 ectopic expression of combinations of other pluripotency-specific transcription factors (e.g.
244 NANOG, LIN28), small molecules (e.g. VPA, BIX01294), and/or ectopic expression of
245 microRNAs (e.g. miR302/367) [76,77]. hiPSC generation is a slow, stochastic process that
246 can be accelerated by manipulating cell division rate through inhibition of the p53/p21
247 pathway [76] and ectopic expression of connexin 45 [78] or LIN28 [76].

248 By providing specific binding sites for pluripotency factors LTR7/HERVH is good source
249 material to evolve functions that influence the process of somatic cell reprogramming

250 [12,60,71,72]. Indeed, exogenous NANOG can activate HERVH transcription in fibroblasts
251 and promote the acquisition of pluripotency in somatic cells [71]. Furthermore, the ectopic
252 expression of OCT4 would only increase reprogramming efficiency when certain HERVH
253 transcripts are present [71]. Thus, besides serving as a binding platform for pluripotency
254 specific transcription factors, HERVH driven transcripts facilitate the hiPSC generation.

255 Indeed, certain LTR7/HERVH products (e.g ESRG, linc-RoR), implicated in promoting
256 reprogramming are expressed during the process (Figure 5B) [12,72]. Only when the
257 reprogramming is completed and cells have acquired the pluripotent state do the levels of
258 HERVH and its flanking LTR7 drop to those observed in hESCs [12]. In line with this, linc-
259 RoR, thought to be a direct target of pluripotency factors, is silenced when cells adapt to a
260 differentiated phenotype [56,67]. However, LTR7/HERVH expression stays high in some
261 hiPSC lines vs hESCs [55,56,64,67]. Failure to suppress LTR7 and pluripotency transcription
262 factor KLF4 after reprogramming is complete leads to a differentiation defective phenotype
263 [64]. In these cells, KLF4 binding is thought to antagonize TRIM28/KRAB-associated protein
264 1 (KAP-1) binding to LTR7, thereby preventing HERVH suppression [44,64].

265 hiPSCs, in comparison with hESCs, are more prone to genomic and epigenetic aberrations
266 [79-82]. These include random point-mutations, duplications and deletions in protein coding
267 and non-coding regions [82].The epigenetic turbulence during the reprogramming process
268 also activates TEs in a characteristic pattern. While young TEs, such as the HERVK_Hs is
269 preferentially active during the intermediate stage of the reprogramming process, LTR7-
270 driven HERVH expression increases and stays high in cultured hiPSCs [12,20], where
271 HERVK_HS expression is simply a read-through event [12]. Reprogramming has been
272 reported to activate the young mutagenic elements (e.g. L1 and SVA). Low activity of the L1
273 and SVA has been also reported to occur during hPSCs culturing in both hESC and hiPSC
274 cultures [83].

275 Collectively, in addition to the role in supporting self-renewal and inhibiting differentiation,
276 LTR7/HERVH plays a role in pluripotency acquisition, and perhaps in stabilization of the
277 pluripotent state, underlining the various cellular functions of HERVH that might affect
278 pluripotency.

279

280 **HERVH as a marker of naïve-like pluripotency**

Ground-state or naïve pluripotency is an ability of the pluripotent blastocyst for unbiased differentiation. *In vitro* hPSC cultures consist of a heterogeneous cell population [53] with only around 4% naïve-like cells exhibiting naive-like pluripotency [12]. Importantly, these naïve-like cells can be identified based on elevated LTR7/HERVH [12] expression compared to primed hESCs. Higher expression levels of HERVH in naïve-like hPSCs have been associated with a subset of HERVH possessing a binding site for the pluripotency transcription factor LBP9 [12,84]. This initiates transcription of specific lncRNAs and chimeric transcripts as a part of the primate-specific pluripotency network [12]. Others have reported LTR7Y/HERVH [10], HERVK or SVA as being a more precise marker for naïve-like hPSCs [11,85]. This discrepancy could be explained with the existence of various naïve-like cell lines. Which naïve-like cell line mimics real developmental conditions most precisely, remains currently a matter of debate. Importantly, however, SVA is potentially mutagenic, and in contrast to HERVH, to date there is no evidence of HERVK function in pluripotency.

Conclusions

Here, we have reviewed the recent literature on HERVH in regulating human pluripotency. The HERVH-mediated circuitry, modulating pluripotency has been evolved recently in primates. The conserved function of HERVH expressing ncRNA, estimated to arose before the divergence of gorilla [86]. Nevertheless, the HERVH-derived regulatory network has been incorporated, and appears to be an essential feature of human pluripotency. In what way pluripotency is specific to us humans, still needs to be deciphered. HERVH provides binding sites for pluripotency transcription factors, functions as a long-range enhancer and serves as a major source for ncRNAs and chimeric transcripts in hPSCs. A few of the many HERVH enforced pluripotent stem cell specific transcripts have been already characterized to have functionality. However, as the domestication process of HERVH is relatively young, it remains to be also established what proportion of the HERVH associated transcripts is functional. The diverse mechanisms, by which HERVH participates in modulating pluripotency are only starting to be elucidated. Finally, besides regulating pluripotency, HERVH might have been co-opted to other functions in other stages of the preimplantation embryogenesis.

313 **Acknowledgements**

314 Z.I. is funded by European Research Council, ERC Advanced [ERC-2011-ADG 294742].
315 The authors thank Katarina Stevanovic for proof-reading the manuscript.

316

317 **References and recommended reading**

318 1. Benveniste RE, Todaro GJ: **Evolution of type C viral genes: I. Nucleic acid from baboon type C**
319 **virus as a measure of divergence among primate species**. *Proc Natl Acad Sci U S A* 1974,
320 **71**:4513-4518.
321 2. Martin MA, Bryan T, Rasheed S, Khan AS: **Identification and cloning of endogenous retroviral**
322 **sequences present in human DNA**. *Proc Natl Acad Sci U S A* 1981, **78**:4892-4896.
323 3. Katzourakis A, Gifford RJ: **Endogenous viral elements in animal genomes**. *PLoS Genet* 2010,
324 **6**:e1001191.
325 4. Lower R, Lower J, Kurth R: **The viruses in all of us: characteristics and biological significance of**
326 **human endogenous retrovirus sequences**. *Proc Natl Acad Sci U S A* 1996, **93**:5177-5184.
327 5. Tchenio T, Heidmann T: **Defective retroviruses can disperse in the human genome by intracellular**
328 **transposition**. *J Virol* 1991, **65**:2113-2118.
329 6. Gifford R, Tristem M: **The evolution, distribution and diversity of endogenous retroviruses**. *Virus*
330 *Genes* 2003, **26**:291-315.
331 7. Blaise S, de Parseval N, Benit L, Heidmann T: **Genomewide screening for fusogenic human**
332 **endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate**
333 **evolution**. *Proc Natl Acad Sci U S A* 2003, **100**:13013-13018.
334 8. Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, et al.:
335 **Syncytin is a captive retroviral envelope protein involved in human placental**
336 **morphogenesis**. *Nature* 2000, **403**:785-789.
337 9. Chuong EB, Elde NC, Feschotte C: **Regulatory evolution of innate immunity through co-option of**
338 **endogenous retroviruses**. *Science* 2016, **351**:1083-1087.
339 10. Goke J, Lu X, Chan YS, Ng HH, Ly LH, Sachs F, Szczerbinska I: **Dynamic transcription of distinct**
340 **classes of endogenous retroviral elements marks specific populations of early human**
341 **embryonic cells**. *Cell Stem Cell* 2015, **16**:135-141.
342 ** The authors report on stage-specific transcription of various various endogenous
343 retroviruses during human preimplantation embryogenesis.
344 11. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA,
345 Chang HY, et al.: **Intrinsic retroviral reactivation in human preimplantation embryos and**
346 **pluripotent cells**. *Nature* 2015, **522**:221-225.
347 * Demonstration that a HERVK product induces a viral restriction pathway in certain human
348 naïve-like cells
349 12. Wang J, Xie G, Singh M, Ghanbarian AT, Rasko T, Szvetnik A, Cai H, Besser D, Prigione A, Fuchs NV,
350 et al.: **Primate-specific endogenous retrovirus-driven transcription defines naive-like stem**
351 **cells**. *Nature* 2014, **516**:405-409.
352 ** The authors have thoroughly investigated the role of HERVH in stem cells, and report on a
353 HERVH-based regulatory network in primates, consisting on LTR7/HERVH-enforced and
354 formation of chimeric transcripts as well as lncRNAs. They use a LTR7/HERVH driven reporter
355 to extact naïve-like cell from huamn cultures.
356 13. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D,
357 et al.: **Epigenomic analysis of multilineage differentiation of human embryonic stem cells**.
358 *Cell* 2013, **153**:1134-1148.

359 14. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD: **Repetitive elements may comprise over**
360 **two-thirds of the human genome**. *PLoS Genet* 2011, **7**:e1002384.
361 15. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M,
362 FitzHugh W, et al.: **Initial sequencing and analysis of the human genome**. *Nature* 2001,
363 **409**:860-921.
364 16. Mager DL, Stoye JP: **Mammalian Endogenous Retroviruses**. *Microbiol Spectr* 2015, **3**:MDNA3-
365 0009-2014.
366 * A comprehensive and recent review on diverse functions of endogenous retroviruses in
367 mammalian genome.
368 17. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M,
369 Panaud O, et al.: **A unified classification system for eukaryotic transposable elements**. *Nat*
370 *Rev Genet* 2007, **8**:973-982.
371 18. Jern P, Coffin JM: **Effects of retroviruses on host genome function**. *Annu Rev Genet* 2008,
372 **42**:709-732.
373 19. Volff JN: **Turning junk into gold: domestication of transposable elements and the creation of**
374 **new genes in eukaryotes**. *Bioessays* 2006, **28**:913-922.
375 20. Gerdes P, Richardson SR, Mager DL, Faulkner GJ: **Transposable elements in the mammalian**
376 **embryo: pioneers surviving through stealth and service**. *Genome Biol* 2016, **17**:100.
377 21. Huang CR, Burns KH, Boeke JD: **Active transposition in genomes**. *Annu Rev Genet* 2012, **46**:651-
378 675.
379 22. Mills RE, Bennett EA, Iskow RC, Devine SE: **Which transposable elements are active in the human**
380 **genome?** *Trends Genet* 2007, **23**:183-191.
381 23. Marchi E, Kanapin A, Magiorkinis G, Belshaw R: **Unfixed endogenous retroviral insertions in the**
382 **human population**. *J Virol* 2014, **88**:9529-9537.
383 24. Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JM: **Discovery of**
384 **unfixed endogenous retrovirus insertions in diverse human populations**. *Proc Natl Acad Sci*
385 *U S A* 2016, **113**:E2326-2334.
386 25. Paces J, Pavlicek A, Paces V: **HERVd: database of human endogenous retroviruses**. *Nucleic Acids*
387 *Res* 2002, **30**:205-206.
388 26. Larsson E, Kato N, Cohen M: **Human endogenous proviruses**. *Curr Top Microbiol Immunol* 1989,
389 **148**:115-132.
390 27. Goodchild NL, Wilkinson DA, Mager DL: **Recent evolutionary expansion of a subfamily of RTVL-H**
391 **human endogenous retrovirus-like elements**. *Virology* 1993, **196**:778-788.
392 28. Mager DL, Freeman JD: **HERV-H endogenous retroviruses: presence in the New World branch**
393 **but amplification in the Old World primate lineage**. *Virology* 1995, **213**:395-404.
394 29. Babaian A, Mager DL: **Endogenous retroviral promoter exaptation in human cancer**. *Mob DNA*
395 2016, **7**:24.
396 30. Izsvak Z, Wang J, Singh M, Mager DL, Hurst LD: **Pluripotency and the endogenous retrovirus**
397 **HERVH: Conflict or serendipity?** *Bioessays* 2016, **38**:109-117.
398 31. Weiss RA: **The discovery of endogenous retroviruses**. *Retrovirology* 2006, **3**:67.
399 32. Magiorkinis G, Blanco-Melo D, Belshaw R: **The decline of human endogenous retroviruses:**
400 **extinction and survival**. *Retrovirology* 2015, **12**:8.
401 33. Mager DL, Henthorn PS: **Identification of a retrovirus-like repetitive element in human DNA**.
402 *Proc Natl Acad Sci U S A* 1984, **81**:7510-7514.
403 34. Bao W, Kojima KK, Kohany O: **Repbase Update, a database of repetitive elements in eukaryotic**
404 **genomes**. *Mob DNA* 2015, **6**:11.
405 35. Hirose Y, Takamatsu M, Harada F: **Presence of env genes in members of the RTVL-H family of**
406 **human endogenous retrovirus-like elements**. *Virology* 1993, **192**:52-61.
407 36. Lindeskog M, Mager DL, Blomberg J: **Isolation of a human endogenous retroviral HERV-H**
408 **element with an open env reading frame**. *Virology* 1999, **258**:441-450.

409 37. Wilkinson DA, Goodchild NL, Saxton TM, Wood S, Mager DL: **Evidence for a functional subclass of**
410 **the RTVL-H family of human endogenous retrovirus-like sequences**. *J Virol* 1993, **67**:2981-
411 2989.
412 38. de Parseval N, Casella J, Gressin L, Heidmann T: **Characterization of the three HERV-H proviruses**
413 **with an open envelope reading frame encompassing the immunosuppressive domain and**
414 **evolutionary history in primates**. *Virology* 2001, **279**:558-569.
415 39. Santoni FA, Guerra J, Luban J: **HERV-H RNA is abundant in human embryonic stem cells and a**
416 **precise marker for pluripotency**. *Retrovirology* 2012, **9**:111.
417 40. Gemmell P, Hein J, Katzourakis A: **Phylogenetic Analysis Reveals That ERVs "Die Young" but**
418 **HERV-H Is Unusually Conserved**. *PLoS Comput Biol* 2016, **12**:e1004964.
419 ** In this paper the authors show that HERVH is deleted at a much slower rate compared to
420 other endogenous retroviruses.
421 41. Stankiewicz P, Lupski JR: **Genome architecture, rearrangements and genomic disorders**. *Trends*
422 *Genet* 2002, **18**:74-82.
423 42. Magiorkinis G, Gifford RJ, Katzourakis A, De Ranter J, Belshaw R: **Env-less endogenous**
424 **retroviruses are genomic superspreaders**. *Proc Natl Acad Sci U S A* 2012, **109**:7385-7390.
425 43. Stadtfeld M, Hochedlinger K: **Induced pluripotency: history, mechanisms, and applications**.
426 *Genes Dev* 2010, **24**:2239-2263.
427 44. Rowe HM, Trono D: **Dynamic control of endogenous retroviruses during development**. *Virology*
428 2011, **411**:273-287.
429 45. Imbeault M, Helleboid PY, Trono D: **KRAB zinc-finger proteins contribute to the evolution of**
430 **gene regulatory networks**. *Nature* 2017, **543**:550-554.
431 ** Authors demonstrate that KRAB zinc-finger proteins partner with transposable elements
432 to build a species-specific layer of their epigenetic regulation..
433 46. Marchetto MC, Narvaiza I, Denli AM, Benner C, Lazzarini TA, Nathanson JL, Paquola AC, Desai KN,
434 Herai RH, Weitzman MD, et al.: **Differential L1 regulation in pluripotent stem cells of**
435 **humans and apes**. *Nature* 2013, **503**:525-529.
436 47. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, Flory E, Schumann GG, Munk
437 C: **APOBEC3 proteins inhibit human LINE-1 retrotransposition**. *J Biol Chem* 2006, **281**:22161-
438 22172.
439 48. Muotri AR, Marchetto MC, Coufal NG, Oefner R, Yeo G, Nakashima K, Gage FH: **L1**
440 **retrotransposition in neurons is modulated by MeCP2**. *Nature* 2010, **468**:443-446.
441 49. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, Tieng V, Quenneville S, Trono
442 D: **Interplay of TRIM28 and DNA methylation in controlling human endogenous**
443 **retroelements**. *Genome Res* 2014, **24**:1260-1270.
444 50. Braude P, Bolton V, Moore S: **Human gene expression first occurs between the four- and eight-**
445 **cell stages of preimplantation development**. *Nature* 1988, **332**:459-461.
446 51. Latham KE, Schultz RM: **Embryonic genome activation**. *Front Biosci* 2001, **6**:D748-759.
447 52. Hancks DC, Kazazian HH, Jr.: **Active human retrotransposons: variation and disease**. *Curr Opin*
448 *Genet Dev* 2012, **22**:191-203.
449 53. Chenoweth JG, McKay RD, Tesar PJ: **Epiblast stem cells contribute new insight into pluripotency**
450 **and gastrulation**. *Dev Growth Differ* 2010, **52**:293-301.
451 54. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S: **Induction of**
452 **pluripotent stem cells from adult human fibroblasts by defined factors**. *Cell* 2007, **131**:861-
453 872.
454 55. Friedli M, Turelli P, Kapopoulou A, Rauwel B, Castro-Diaz N, Rowe HM, Ecco G, Unzu C, Planet E,
455 Lombardo A, et al.: **Loss of transcriptional control over endogenous retroelements during**
456 **reprogramming to pluripotency**. *Genome Res* 2014, **24**:1251-1259.
457 56. Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, Bourque G, Ng HH: **The retrovirus HERVH is a long**
458 **noncoding RNA required for human embryonic stem cell identity**. *Nat Struct Mol Biol* 2014,
459 **21**:423-425.

460 * In this paper the authors reveal a novel role role for HERVH RNA functioning as a scaffold in
461 stem cells.
462 57. Fort A, Yamada D, Hashimoto K, Koseki H, Carninci P: **Nuclear transcriptome profiling of induced**
463 **pluripotent stem cells and embryonic stem cells identify non-coding loci resistant to**
464 **reprogramming**. *Cell Cycle* 2015, **14**:1148-1155.
465 * The authors show the relevance of LTR7 and HERVH internal sequence contribution into
466 the stem cell nuclear transcriptome.
467 58. Jacques PE, Jeyakani J, Bourque G: **The majority of primate-specific regulatory sequences are**
468 **derived from transposable elements**. *PLoS Genet* 2013, **9**:e1003504.
469 59. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-**
470 **resolution profiling of histone methylations in the human genome**. *Cell* 2007, **129**:823-837.
471 60. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G: **Transposable**
472 **elements have rewired the core regulatory network of human embryonic stem cells**. *Nat*
473 *Genet* 2010, **42**:631-634.
474 ** Authors demonstrate that transposable elements provide up to 25% of transcription
475 factor binding sites in human pluripotent stem cells and wire novel genes into regulatory
476 network in these cells.
477 61. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray
478 HL, Jenner RG, et al.: **Core transcriptional regulatory circuitry in human embryonic stem**
479 **cells**. *Cell* 2005, **122**:947-956.
480 62. Chambers I, Smith A: **Self-renewal of teratocarcinoma and embryonic stem cells**. *Oncogene*
481 2004, **23**:7150-7160.
482 63. Niwa H: **How is pluripotency determined and maintained?** *Development* 2007, **134**:635-646.
483 64. Ohnuki M, Tanabe K, Sutou K, Teramoto I, Sawamura Y, Narita M, Nakamura M, Tokunaga Y,
484 Nakamura M, Watanabe A, et al.: **Dynamic regulation of human endogenous retroviruses**
485 **mediates factor-induced reprogramming and differentiation potential**. *Proc Natl Acad Sci U*
486 *S A* 2014, **111**:12426-12431.
487 * In this paper the authors show that HERVH expression promotes the process of
488 reprogramming somatic cells into iPSCs.
489 65. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R,
490 Bruhn L, et al.: **lincRNAs act in the circuitry controlling pluripotency and differentiation**.
491 *Nature* 2011, **477**:295-300.
492 66. Kretz M, Webster DE, Flockhart RJ, Lee CS, Zehnder A, Lopez-Pajares V, Qu K, Zheng GX, Chow J,
493 Kim GE, et al.: **Suppression of progenitor differentiation requires the long noncoding RNA**
494 **ANCR**. *Genes Dev* 2012, **26**:338-343.
495 67. Ng SY, Johnson R, Stanton LW: **Human long non-coding RNAs promote pluripotency and**
496 **neuronal differentiation by association with chromatin modifiers and transcription factors**.
497 *EMBO J* 2012, **31**:522-533.
498 68. Kelley D, Rinn J: **Transposable elements reveal a stem cell-specific class of long noncoding RNAs**.
499 *Genome Biol* 2012, **13**:R107.
500 69. Robertson HM, Martos R: **Molecular evolution of the second ancient human mariner**
501 **transposon, Hsmar2, illustrates patterns of neutral evolution in the human genome**
502 **lineage**. *Gene* 1997, **205**:219-228.
503 70. Wang Y, Pryputniewicz-Dobrinska D, Nagy EE, Kaufman CD, Singh M, Yant S, Wang J, Dalda A, Kay
504 MA, Ivics Z, et al.: **Regulated complex assembly safeguards the fidelity of Sleeping Beauty**
505 **transposition**. *Nucleic Acids Res* 2017, **45**:311-326.
506 71. Durruthy-Durruthy J, Sebastiano V, Wossidlo M, Cepeda D, Cui J, Grow EJ, Davila J, Mall M, Wong
507 WH, Wysocka J, et al.: **The primate-specific noncoding RNA HPAT5 regulates pluripotency**
508 **during human preimplantation development and nuclear reprogramming**. *Nat Genet* 2016,
509 **48**:44-52.
510 ** Authors have analyzed the importance of HERVH linc-RNAs in preimplantation

511 embryogenesis and found their relevance in pluripotent cell fate determination, formation of
512 inner cell mass, and pluripotency acquisition.
513 72. Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, Garber M, Curran M, Onder T,
514 Agarwal S, et al.: **Large intergenic non-coding RNA-RoR modulates reprogramming of**
515 **human induced pluripotent stem cells**. *Nat Genet* 2010, **42**:1113-1117.
516 73. Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, Wu M, Xiong J, Guo X, Liu H: **Endogenous miRNA**
517 **sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-**
518 **renewal**. *Dev Cell* 2013, **25**:69-80.
519 74. Zhao M, Ren C, Yang H, Feng X, Jiang X, Zhu B, Zhou W, Wang L, Zeng Y, Yao K: **Transcriptional**
520 **profiling of human embryonic stem cells and embryoid bodies identifies HESRG, a novel**
521 **stem cell gene**. *Biochem Biophys Res Commun* 2007, **362**:916-922.
522 75. Li G, Ren C, Shi J, Huang W, Liu H, Feng X, Liu W, Zhu B, Zhang C, Wang L, et al.: **Identification,**
523 **expression and subcellular localization of ESRG**. *Biochem Biophys Res Commun* 2013,
524 **435**:160-164.
525 76. Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, van Oudenaarden A, Jaenisch R:
526 **Direct cell reprogramming is a stochastic process amenable to acceleration**. *Nature* 2009,
527 **462**:595-601.
528 77. Zhu Z, Huangfu D: **Human pluripotent stem cells: an emerging model in developmental biology**.
529 *Development* 2013, **140**:705-717.
530 78. Ke Q, Li L, Yao X, Lai X, Cai B, Chen H, Chen R, Zhai Z, Huang L, Li K, et al.: **Enhanced generation of**
531 **human induced pluripotent stem cells by ectopic expression of Connexin 45**. *Sci Rep* 2017,
532 **7**:458.
533 79. Hussein SM, Batada NN, Vuoristo S, Ching RW, Autio R, Narva E, Ng S, Sourour M, Hamalainen R,
534 Olsson C, et al.: **Copy number variation and selection during reprogramming to**
535 **pluripotency**. *Nature* 2011, **471**:58-62.
536 80. Laurent LC, Ulitsky I, Slavin I, Tran H, Schork A, Morey R, Lynch C, Harness JV, Lee S, Barrero MJ, et
537 al.: **Dynamic changes in the copy number of pluripotency and cell proliferation genes in**
538 **human ESCs and iPSCs during reprogramming and time in culture**. *Cell Stem Cell* 2011,
539 **8**:106-118.
540 81. Ruiz S, Diep D, Gore A, Panopoulos AD, Montserrat N, Plongthongkum N, Kumar S, Fung HL,
541 Giorgetti A, Bilic J, et al.: **Identification of a specific reprogramming-associated epigenetic**
542 **signature in human induced pluripotent stem cells**. *Proc Natl Acad Sci U S A* 2012,
543 **109**:16196-16201.
544 82. Ruiz S, Gore A, Li Z, Panopoulos AD, Montserrat N, Fung HL, Giorgetti A, Bilic J, Batchelder EM,
545 Zaehres H, et al.: **Analysis of protein-coding mutations in hiPSCs and their possible role**
546 **during somatic cell reprogramming**. *Nat Commun* 2013, **4**:1382.
547 83. Klawitter S, Fuchs NV, Upton KR, Munoz-Lopez M, Shukla R, Wang J, Garcia-Canadas M, Lopez-
548 Ruiz C, Gerhardt DJ, Sebe A, et al.: **Reprogramming triggers endogenous L1 and Alu**
549 **retrotransposition in human induced pluripotent stem cells**. *Nat Commun* 2016, **7**:10286.
550 * This manuscript report on L1 and SVA activities in cultured pluripotent stem cells.
551 84. Martello G, Bertone P, Smith A: **Identification of the missing pluripotency mediator downstream**
552 **of leukaemia inhibitory factor**. *EMBO J* 2013, **32**:2561-2574.
553 85. Theunissen TW, Friedli M, He Y, Planet E, O'Neil RC, Markoulaki S, Pontis J, Wang H, Iouranova A,
554 Imbeault M, et al.: **Molecular Criteria for Defining the Naive Human Pluripotent State**. *Cell*
555 *Stem Cell* 2016, **19**:502-515.
556 * The authors claim that SVA and the HERVK-associated LTR, LTR5_Hs, are more readily
557 transcribed in naive cells.
558 86. Ramsay L, Marchetto MC, Caron M, Chen SH, Busche S, Kwan T, Pastinen T, Gage FH, Bourque G:
559 **Conserved expression of transposon-derived non-coding transcripts in primate stem cells**.
560 *BMC Genomics* 2017, **18**:214.

561 87. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J, et al.: **Single-cell RNA-Seq**
562 **profiling of human preimplantation embryos and embryonic stem cells**. *Nat Struct Mol Biol*
563 2013, **20**:1131-1139.

564

565

566 **Figure legends**

567 **Figure 1**. Structure of the most common HERVH sequence in the human genome. Internal
568 sequence of HERVH consists of viral genes *gag* and *pol* carrying mutations and deletions (not
569 shown) and is flanked by regulatory long-terminal repeats (LTRs) which in case of HERVH
570 are classified as type 7 LTR. Near the 5'LTR7 is the histidine tRNA binding site (PBS_{His}),
571 hence the name HERVH, and near the 3'LTR7 is the polypurine tail (PPT). SD and SA
572 indicate splice donor and splice acceptor sites, respectively.

573 **Figure 2. Predicted deletion dynamics of full-length ERVs in the next 25 million years.**
574 HERVHs would degrade more slowly compared to three other HERVs, including HERVKs
575 (estimates on deletion dynamics of full length elements: under a Weibull model, thick
576 black/thin grey lines; under an exponential model, black dotted lines). Adapted from [40]..

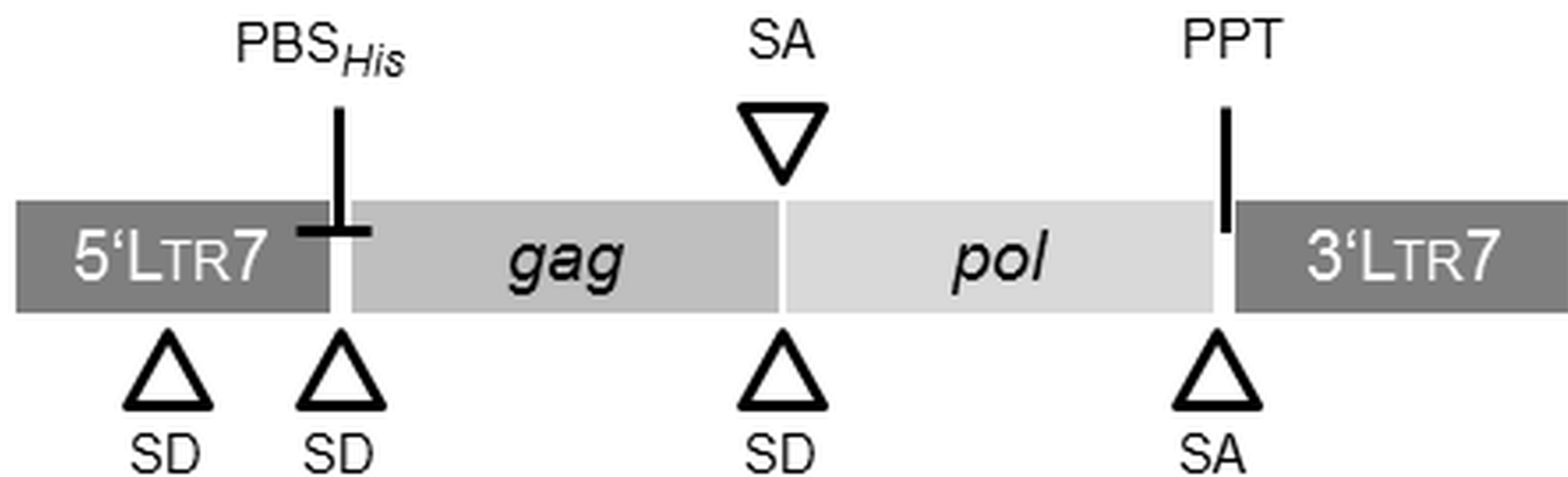
577 **Figure 3.** Expression dynamic of selected enzymes affecting genomic DNA methylation
578 during preimplantation embryogenesis in human vs mouse (data minding of single cell
579 transcriptome data. Lines on the plot are connecting medians of single cell expression levels
580 (log2 TPM) from the depicted stages of development (data mining of [87]). 'Early' includes
581 oocytes, zygote, 2-cell, 4-cell stages in human, while zygote, early 2-cell, late 2-cell, 4-cell in
582 mouse. 'Mid' represents 8-cell, 16-cell/morula stages, and 'Late' denotes blastocysts in both
583 human and mouse. Lines are smoothed by spline function (spar=45) (**A**). Expressional
584 dynamics of Old (> 7MY) and Young (< 7MY) TEs during human preimplantation stages
585 (**B**). Expressional dynamics of Young (< 7MY) TEs during human pre-implantation-stages.
586 TEs marked with * were demonstrated to retrotranspose in human [52]. Note that the
587 youngest LTR7, LTR7Y is (> 7MY) is included in the analysis (**C**).

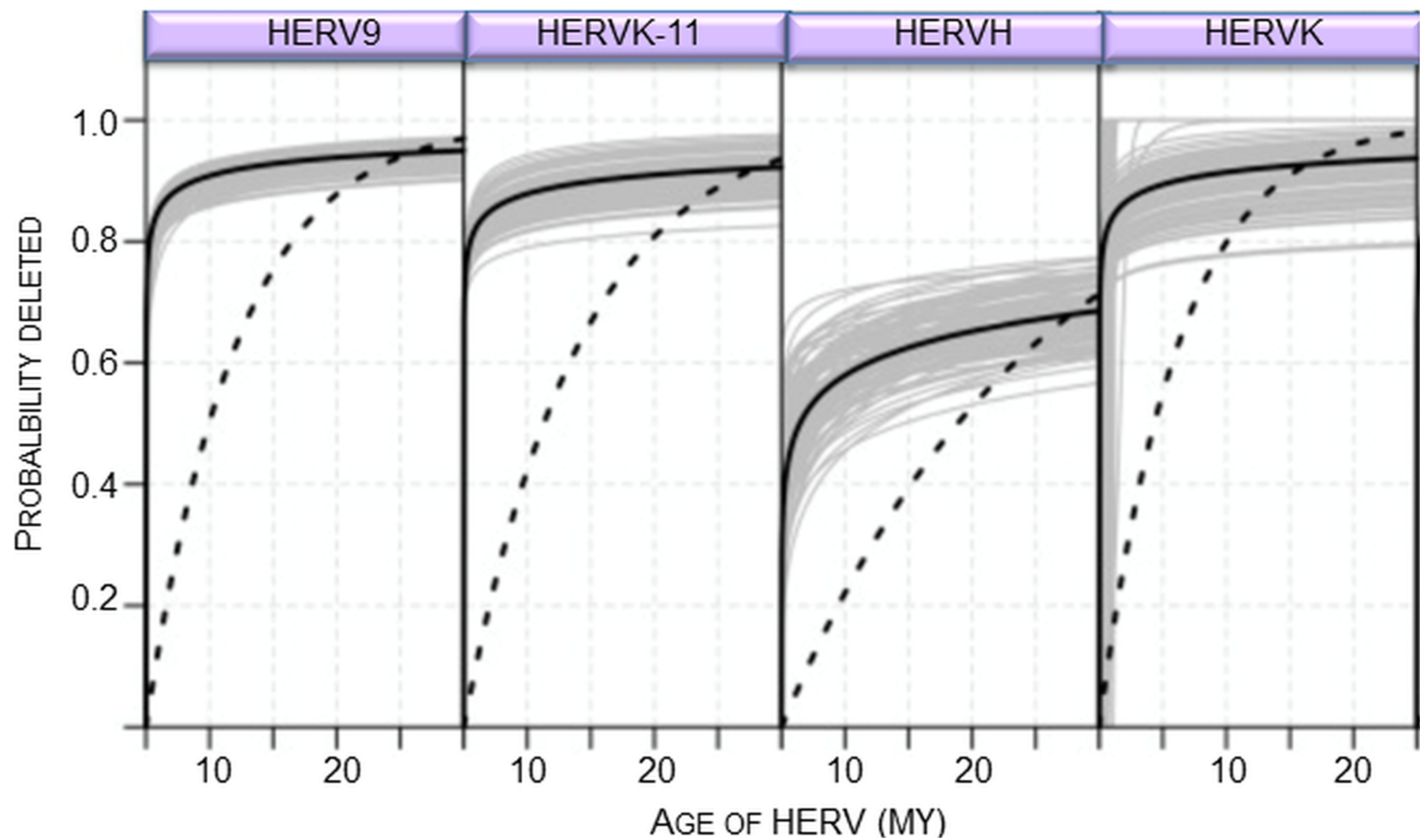
588 **Figure 4.** HERVH expression driven by LTR7 (LTR7, LTR7 with 38 base pair deletion,
589 LTR7B and LTR7Y) during early human embryogenesis. Untypical to other retroelements,
590 LTR7/HERVH expression peaks in human blastocyst. HERVH elements driven by LTR7
591 carrying a 38 base pair deletion is expressed in four cells stage, LTR7B in 8 cells/morula.
592 HERVH driven by the youngest LTR7Y is expressed from 8 cells to blastocyst.

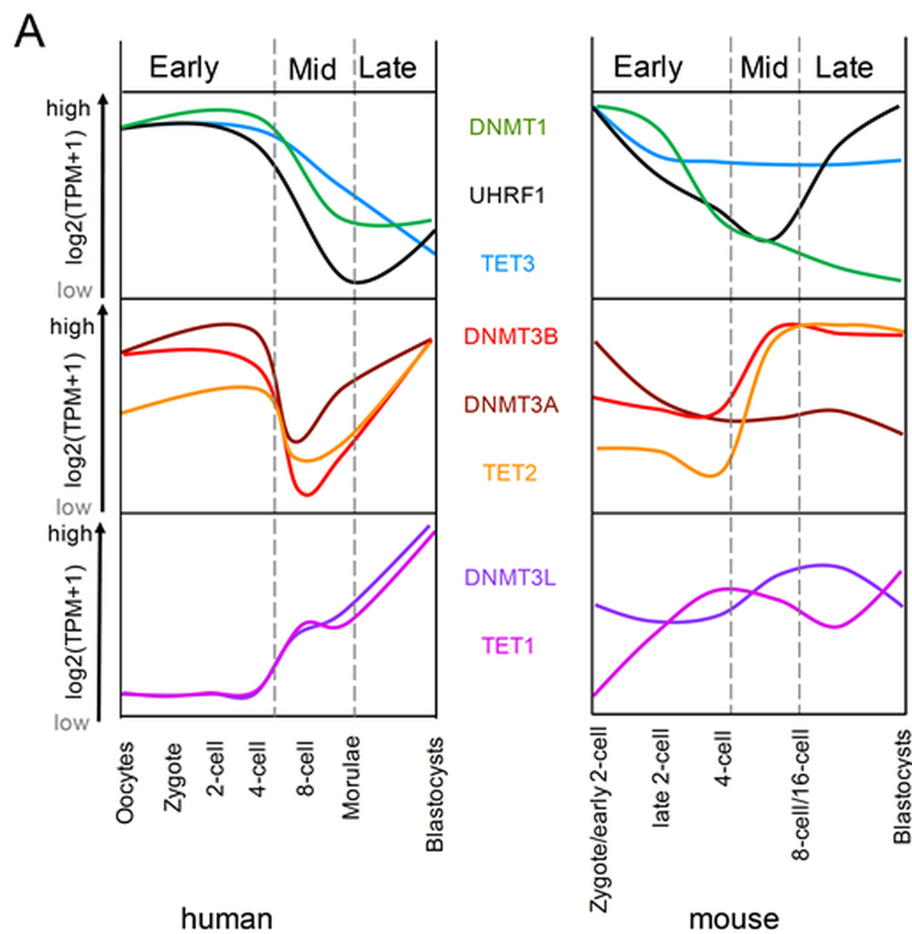
593 **Figure 5.** HERVH expression driven by LTR7 in models of pluripotency: human embryonic
594 stem cells (**A**) and induced pluripotent stem cells (iPSCs) (**B**). Harvested from the pluripotent
595 epiblast of the blastocyst, hESCs are enriched for HERVH driven by LTR7/7Y. In the
596 heterogeneous population of hPSCs, naïve-like hPSCs can be identified by LTR7/7Y
597 expression (**A**). Certain HERVH driven products (ESRG, linc-RoR) promote the acquisition

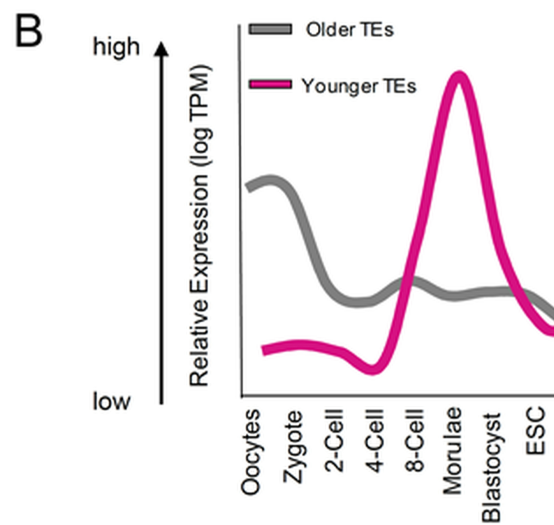
598 of pluripotency during somatic cell reprogramming and increases the efficiency of iPSC
599 generation (**B**).

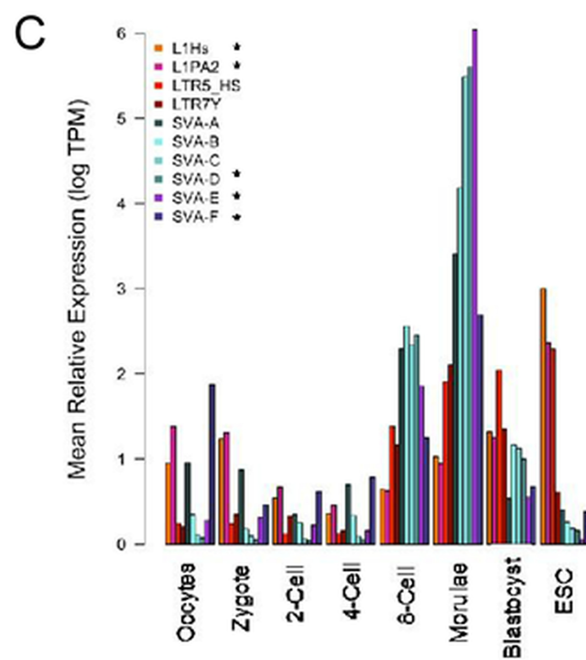
600 **Figure 6.** LTR7/HERVH modulates the human regulator network of pluripotency in various
601 ways. HERVH provides binding sites for pluripotency associated transcription factors
602 (triangles), such as NANOG, LBP9, OCT4 and KLF4. Besides transcribing HERVH-derived
603 products, HERVH also acts as a long-range enhancer, modulating the expression of
604 pluripotency genes (e.g. Gene B) (**A**). HERVH is a source for pluripotency-specific chimeric
605 transcripts (example ESRG which has a putative open reading frame only in humans) as well
606 as long non-coding RNAs (lncRNAs; example linc-RoR). HERVH derived linc-RoR
607 functions as a microRNA sponge and protects pluripotency associated transcription factors
608 from microRNA-mediated suppression (**B**). Certain HERVH derived transcripts act as a
609 scaffold, recruiting transcription factors and co-activators, regulating LTR7 enhancer function
610 and expression of certain genes (example Gene B) (**C**).

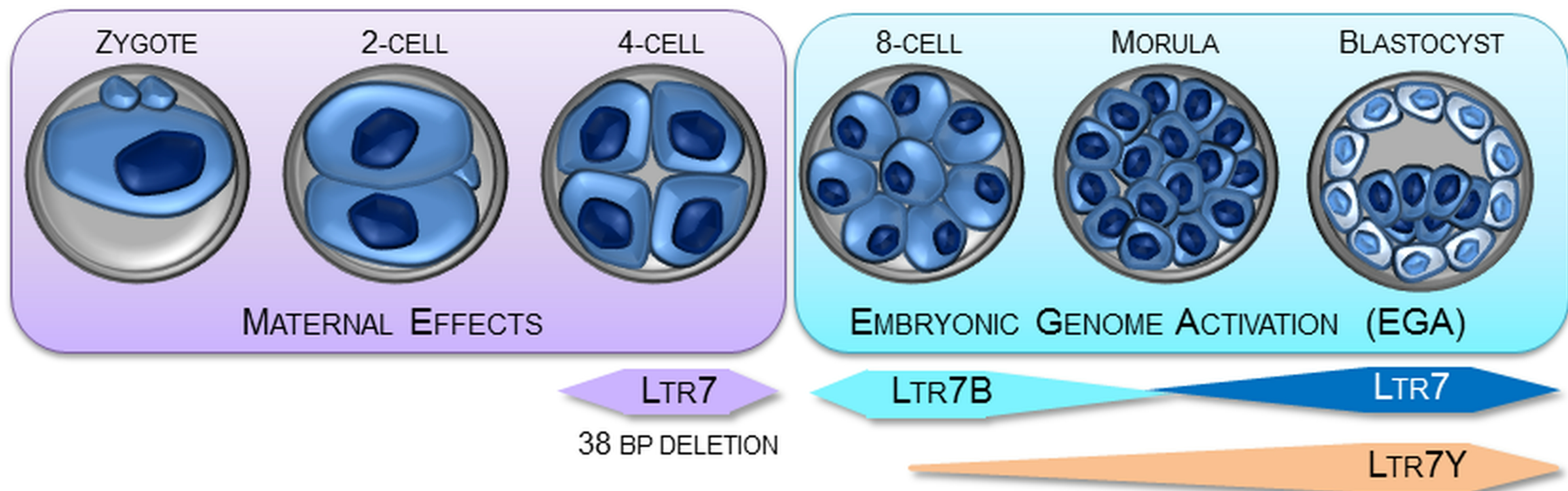

PBS$_{His}$
SA
PPT
5'LTR7
gag
pol
3'LTR7
SD
SD
SD
SA


HERV9
HERVK-11
HERVH
HERVK
1.0
0.8
0.6
0.4
0.2
Probablility deleted
10
20
Age of HERV (MY)

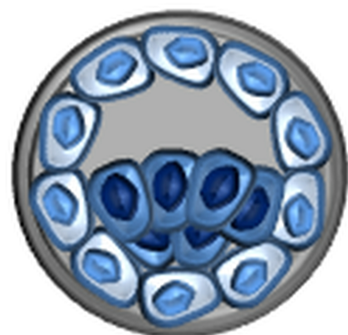

A
Early
Mid
Late
high
low
log2(TPM+1)
DNMT1
UHRF1
TET3
DNMT3B
DNMT3A
TET2
DNMT3L
TET1
Oocytes
Zygote
2-cell
4-cell
8-cell
Morulae
Blastocysts
human
Zygote/early 2-cell
late 2-cell
4-cell
8-cell/16-cell
Blastocysts
mouse

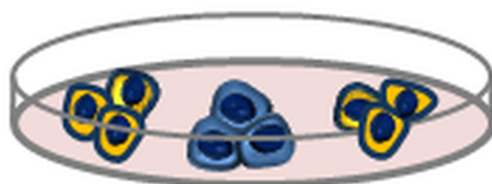

B
high
low
Relative Expression (log TPM)
Older TEs
Younger TEs
Oocytes
Zygote
2-Cell
4-Cell
8-Cell
Morulae
Blastocyst
ESC

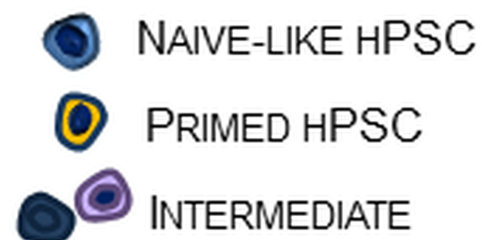

C
Mean Relative Expression (log TPM)
L1Hs *
L1PA2 *
LTR5_HS
LTR7Y
SVA-A
SVA-B
SVA-C
SVA-D *
SVA-E *
SVA-F *
Oocytes
Zygote
2-Cell
4-Cell
8-Cell
Morulae
Blastocyst
ESC



Zygote
2-cell
4-cell
8-cell
Morula
Blastocyst
Maternal Effects
Embryonic Genome Activation (EGA)
Ltr7
38 bp deletion
Ltr7B
Ltr7
Ltr7Y

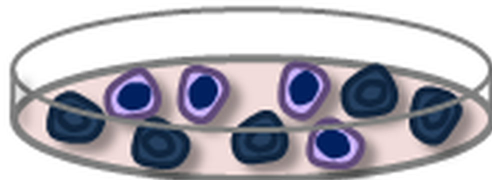

A
Blastocyst
Ltr7/HervH
Ltr7Y/HervH


hESC
Ltr7/HervH


Naive-like hPSC
Primed hPSC
Intermediate

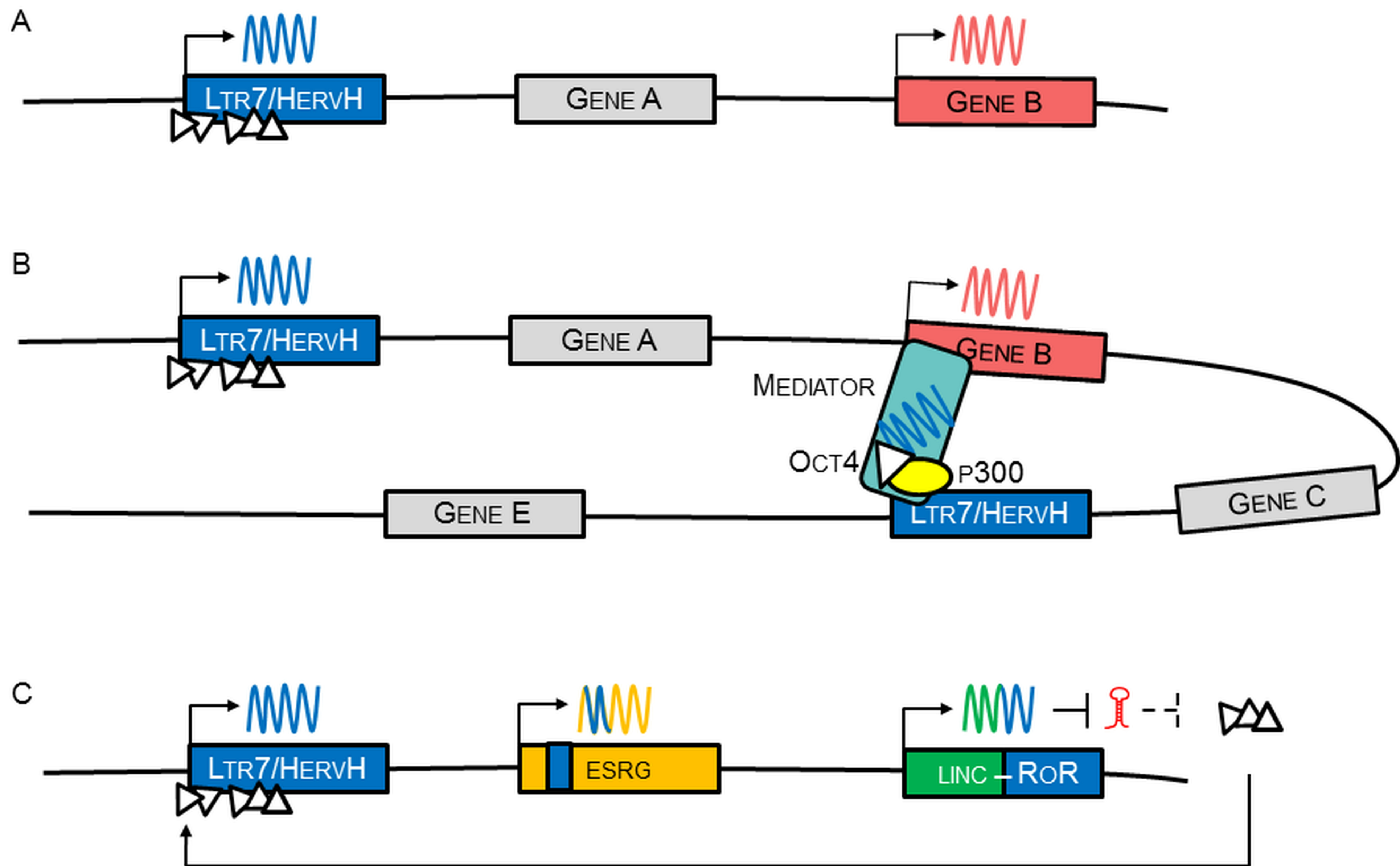
B
Fibroblast

Linc-RoR
ESRG


Intermediate

Linc-RoR
ESRG

hiPSC
Ltr7/HervH


A
LTR7/HERVH
GENE A
GENE B
B
LTR7/HERVH
GENE A
GENE B
MEDIATOR
OCT4
P300
GENE E
LTR7/HERVH
GENE C
C
LTR7/HERVH
ESRG
LINC – ROR