Title

ERH Enables Early Embryonic Differentiation and Overlays H3K9me3 Heterochromatin on a Cryptic Pluripotency H3K9me3 Landscape in Somatic Cells

Author list and affiliations

Andrew Katznelson[1,*], Blake Hernandez[1,*], Holly Fahning[2], Kylea Tapia[2], Adam Burton[3], Jingchao Zhang[1], Maria-Elena Torres-Padilla[3,4], Nicolas Plachta[1], Kenneth S. Zaret[1**], Ryan L. McCarthy[2,5,6**]

*These authors contributed equally

[1]Institute for Regenerative Medicine and Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[2]Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA 98101, USA

[3]Institute of Epigenetics and Stem Cells (IES), Helmholtz Zentrum München D-81377 München, Germany

[4]Faculty of Biology, Ludwig-Maximilians Universität, München, Germany

[5]Department of Pediatrics, University of Washington School of Medicine, Seattle, WA 98195, USA

[6]Lead contact

**co-corresponding authors

R. McCarthy, Ryan.McCarthy@seattlechildrens.org

K. Zaret, zaret@pennmedicine.upenn.edu

Key words: ERH, H3K9me3, heterochromatin, embryo, pre-implantation, differentiation, pluripotency, naive

Abstract

Enhancer of Rudimentary Homolog (ERH) is an evolutionarily conserved protein originally characterized in fission yeast[1] and recently shown to maintain H3K9me3 in human fibroblasts[2]. Here, we find that ERH depletion in fibroblasts reverts the H3K9me3 landscape to an embryonic stem cell (ESC) state and enables activation of naïve and pluripotency genes and transposable elements during induced pluripotent stem cell (iPSC) reprogramming. We find that ERH similarly represses totipotent and alternative lineage programs during mouse preimplantation development and is required for proper segregation of the inner cell mass and trophectoderm cell lineages. During human ESC differentiation into germ layer lineages, ERH silences naïve and pluripotency genes, transposable elements, and alternative lineage somatic genes. As in fission yeast, we find that mammalian ERH interacts with RNA-binding proteins to engage and repress its chromatin targets. Our findings reveal a fundamental role for ERH in cell fate specification via the initiation and maintenance of early developmental gene repression.

39 **Main text**

40 Heterochromatin marked by histone H3 lysine 9 trimethylation (H3K9me3) represses the
41 expression of many protein-coding genes and DNA repeat elements[3–5]. Yet despite its stability,
42 H3K9me3-heterochromatin is dynamically established and rearranged during embryogenesis[3,6].
43 Zygotic genome activation (ZGA), marked by initial transcription of key genes and transposable
44 elements[7], occurs at the 1-2 cell stage in mice and the 4-8 cell stage in humans[7] and requires
45 demethylation of maternal H3K9me3 marks in oocytes[8,9]. As the embryo loses totipotency and
46 differentiates into the first two developmentally-specified lineages, the inner cell mass (ICM) and
47 the trophectoderm of the blastocyst, ZGA-specific protein-coding genes and repeats gain
48 H3K9me3 and are subsequently silenced [6,10,11]. As development proceeds, H3K9me3 is
49 established on lineage-inappropriate genes and repeat elements [6,10,12], with the highest number
50 of genes marked by H3K9me3 at the germ-layer stage[3]. Throughout germ layer specification to
51 mature somatic lineages, H3K9me3 is removed from genes that become active and is deposited
52 onto early developmental genes that become repressed[3]. The mechanisms operating upstream
53 in the initial targeting of histone methyltransferases to deposit H3K9me3 and the maintenance of
54 the initially established H3K9me3 domains remain poorly understood. In this study, we compared
55 how *enhancer of rudimentary homolog* (ERH), a highly conserved heterochromatin protein distinct
56 from histone methyltransferases, contributes to H3K9me3 dynamics and gene repression in
57 differentiated somatic cells, pluripotent cell lines, and early mammalian development.

58 ERH is an evolutionarily conserved protein originally discovered in fission yeast to regulate
59 sporulation, the primary form of yeast cell differentiation, via H3K9 methylation [1,13]. In *S. pombe,*
60 Erh1 forms a complex with the RNA binding protein Mmi1, which is then recruited by binding of
61 nascent meiotic transcripts through its YTH domain[1,15]. Erh1-Mmi1 then causes the degradation
62 of its bound RNA transcripts, the establishment of H3K9 methylation, and consequent repression
63 of meiotic and developmental genes, transposons, and rDNA[1,13]. We recently found that in
64 terminally-differentiated primary human fibroblasts, ERH is highly enriched in H3K9me3-

65 heterochromatin, where it represses meiotic and germ cell genes, similar to that observed in
66 fission yeast[1,2,5]. We further found that ERH in mammals has evolutionarily expanded its function
67 to include the majority of regions that exhibit H3K9me3 heterochromatin in somatic cells, including
68 alternative-lineage protein coding genes and repeat elements[2]. While ERH is highly conserved
69 among eukaryotes, its fission yeast partner Mmi1 lacks a metazoan homolog[14]. Thus, how cell-
70 type specific cofactors recruit ERH to chromatin to regulate cell identity throughout mammalian
71 development is unknown.

72 Here, we find that upon ERH loss in fibroblasts, there is reversion to a pluripotency-like
73 H3K9me3 state which is permissive to binding of pluripotency transcription factors, which in turn
74 activate naive and alternative lineage genes and repeat elements. The activation of a cryptic naïve
75 state led us to investigate the role of ERH in the earliest developmental stages where repression
76 by H3K9me3 has been implicated: during trophectoderm and inner cell mass segregation in the
77 blastocyst[6,10,11] and the differentiation from pluripotency to the germ layer lineage[3]. By genetic
78 perturbation, chromatin and gene expression profiling, and imaging, we identify a role for ERH
79 across mouse and human systems in repressing early naïve pluripotency and alternative lineage
80 genes. We find that targeting of mammalian ERH to chromatin occurs through cell-type specific
81 interactions with RNA binding proteins, enabling context-dependent functions**.** Together, our
82 findings demonstrate a fundamental role for ERH during developmentally acquired H3K9me3-
83 mediated gene repression and maintenance of this repression in somatic cells.

84

85 **Results**

86

87 **ERH depletion in primary human fibroblasts causes H3K9me3 to revert to an ESC-like state**

88 We assessed how H3K9me3 is disrupted after ERH loss by CUT&RUN for H3K9me3 after siRNA
89 knockdown in human primary fibroblasts (Fig. 1a). Consistent with our previous study[2], we
90 confirmed that H3K9me3 is globally reduced in human fibroblasts after ERH knockdown (Fig. 1b).

91 Yet while large megabase-scale domains were diminished, 10,526 residual peaks were retained
92 in the ERH knockdown (Fig. 1c and Supplementary Table 1). Remarkably, the residual H3K9me3
93 signals resemble the H3K9me3 profile in H1 ESCs (Fig. 1c, see inset, 1d). Across the genome,
94 85% of the residual fibroblast siERH H3K9me3 sites overlap with H3K9me3 peaks in H1 hESCs
95 (Extended Data Fig. 1a). In contrast, only 30% of fibroblast siControl H3K9me3 sites overlap
96 hESC H3K9me3 (Extended Data Fig. 1a,1b). Pearson correlation of H3K9me3 peaks from
97 fibroblasts after siControl or siERH, as well as in naive hESCs, supports a shift in the H3K9me3
98 heterochromatin landscape to an embryonic state (Fig. 1e). We validated that the residual sites
99 are bona fide marked by H3K9me3 in ESCs, rather than another chromatin protein with antibody
100 cross-reactivity, by showing enrichment at these sites for ESC HP1 (Fig. 1d, f, and Extended Data
101 Fig.1c).
102 To identify the chromatin proteins that engage the residual peaks after ERH knockdown,
103 we performed de novo motif analysis. We find that the most highly enriched sequence motifs in
104 the residual peaks are for KRAB-ZNFs and SOX transcription factors (Extended Data Fig. 1d).
105 We next used the cistromeDB Toolkit to identify specific proteins enriched at residual peaks,
106 which compares these regions against published ChIP-seq data sets[16,17]. We find that in addition
107 to the core H3K9me3 machinery, including SETDB1 and HP1, most proteins enriched at the
108 residual peaks are KRAB-ZNFs (Fig. 1g). As KRAB-ZNFs establish H3K9me3 to repress
109 transposons, we hypothesized the residual peaks would be enriched over repeats. Repeat
110 enrichment analysis found that the residual peaks are enriched over canonical KRAB-ZNF
111 targets: LINEs, LTRs, and satellite repeats (Extended Data Fig. 1e). We conclude that ERH is
112 required to maintain the broad H3K9me3 domains established in development but does not
113 control H3K9me3 peaks that persist from pluripotency over constitutively repressed repeats.
114
115 **ERH-dependent H3K9me3 suppresses SOX2 binding and activation of naïve-specific**
116 **genes and repeats during OSKM reprogramming**

117 H3K9me3 impairs induced pluripotent stem cell (iPSC) reprogramming by blocking transcription
118 factor binding, and pluripotency genes marked by H3K9me3 show delayed activation compared
119 to genes in heterochromatin marked with H3K27me3 or neither mark (Extended Data Fig. 2a)[15].
120 We therefore assessed the impact of ERH depletion in the context of OSKM transcription factor
121 binding and gene activation during iPSC reprogramming. We focused on a 48 hour time point
122 after initiation of iPSC reprogramming (Fig. 1a and Supplementary Tables 2-4), as H3K9me3 is
123 inhibitory to transcription factor activity at this stage[15].

124 CUT&RUN of SOX2 after 48 hour OSKM reprogramming showed that while most binding
125 sites were shared between siControl and siERH fibroblasts, 2,538 binding sites were gained
126 (Extended Data Fig. 2b and Supplementary Table 2). We observed that ERH knockdown enabled
127 SOX2 to target loci that lose H3K9me3, including within heterochromatin domains previously
128 annotated as resistant to reprogramming factor binding in fibroblasts[15] (Fig. 2a,b, and Extended
129 Data Fig. 2c,d). We asked if the siERH-gained SOX2 targets occurred at normal SOX2 targets by
130 assessing the genomic locations of siERH-gained SOX2 peaks in the cistromeDB toolkit[16,17]. We
131 found that the newly gained SOX2 sites had the most similarity to ChIP-seq peak sets of SOX2
132 and its pluripotency partners NANOG and OCT4 in embryonic stem cells (Extended Data Fig.
133 2e). Inspection of genome browser tracks identified specific loci where loss of ERH-controlled
134 H3K9me3 enabled SOX2 binding over normal SOX2 and NANOG targets in ES cells (Fig. 2b).
135 Protein coding genes embedded in H3K9me3 domains, normally refractory to transcription factor
136 activation, were activated after siERH (136 up, 13 down; Fig. 2c). Significantly, genes associated
137 with the human 8-cell like cell (8CLC) state[18] were activated, including *ZSCAN4, DUXA, DPPA4,*
138 *ARGFX, TPRXL*[19]*, KLF5, TRIM49*, and the histone variant *H3.Y*[20] (Fig. 2d and Extended Data
139 Fig. 2f). We compared the activation levels of these genes in siERH+OSKM with existing RNA-
140 seq datasets[21] and found that *H3.Y, DUXA* and *TPRXL* do not normally get activated at similar
141 levels to siERH+OSKM at any stage during iPSC reprogramming (Extended Data Fig. 2g). We

142 conclude that ERH-dependent H3K9me3 safeguards early totipotent and pluripotent genes from
143 transcription factor binding and activation.

144 We observed that *LTR7* and *HERVH* repeats, which are expressed in and required to
145 maintain pluripotency[22,23], lose H3K9me3 after ERH depletion and are activated in siERH+OSKM
146 (Fig. 2e). *LTR7Y* repeats, which promote naïve pluripotency[24], exhibited similar ERH-dependent
147 H3K9me3 and repression (Fig. 2f,g and Extended Data Fig. 2h). Notably, residual H3K9me3
148 peaks retained at *LTR7Y* repeats were insufficient to maintain repression (Fig. 2f,g). Furthermore,
149 we observed siERH+OSKM activated repeat elements that function specifically during the human
150 8 cell-like cell stage, and are later marked by H3K9me3 and silenced in the human blastocyst[6]
151 (Extended Data Fig. 2l). These results demonstrate that ERH maintains H3K9me3 repression of
152 a subset of 8-cell stage and pluripotency repeats in terminally differentiated human cells.

153 Despite increased activation of pluripotency and 8-cell genes at 48 hours after initiation of
154 OSKM-based iPSC reprogramming, by 7 days we observed substantial increased cell death
155 during siERH (Extended Data Fig. 2j). This is in contrast to our previous published observations
156 where ERH depletion without the addition of reprogramming factors and with the addition of
157 hepatic reprogramming factors did not cause marked cell death or block successful
158 reprogramming[2]. Lower ERH siRNA decreased cell death, but did not change the emerging iPSC
159 colony number (Extended Data Fig. 2k), indicating that ERH is necessary for later stages of iPSC
160 reprogramming.

161

162 **ERH is necessary for trophectoderm development, the first embryonic differentiation**

163 Our finding that ERH represses naïve genes, normally expressed in 2-8 cell embryos, in fully
164 differentiated fibroblasts led us to investigate if ERH establishes silencing of 2-8 cell genes during
165 pre-implantation development. We first analyzed existing RNA-seq data of mouse embryogenesis
166 timepoints from zygote to blastocyst[25] and found that ERH mRNA is present across pre-
167 implantation development, and is significantly increased from the 8-cell stage through the

168 blastocyst stage (Extended Data Fig. 3a). Immunostaining mouse embryos for ERH protein at
169 the zygote, 2-cell, 8-cell and blastocyst stages revealed that ERH is expressed and nuclear-
170 localized at all stages analyzed, with increasing abundance in the early and late blastocyst stages
171 (Fig. 3a,b), consistent with its mRNA levels (Extended Data Fig. 3a). At the blastocyst stage,
172 ERH levels are similar between ICM and trophectoderm cells (Extended Data Fig. 3b).

173 To investigate if ERH is involved in pre-implantation lineage specification, we depleted
174 ERH levels by siRNA micro-injection at the 1-cell stage of mouse embryos, compared to a
175 nontargeting siRNA control, and assessed impacts on the blastocysts (Extended Data Fig. 3c).
176 ERH depletion did not cause major morphological changes in blastocyst structure, as both
177 siControl and siERH blastocysts form a cavity and segregate ICM and trophectoderm cells
178 (Extended Data Fig. 3c,d). Staining blastocysts for key transcription factors involved in lineage
179 specification revealed similar levels of the pluripotency marker OCT4 in ICM and TE cells between
180 siControl and siERH blastocysts (Extended Data Fig. 3e). However, the levels of the
181 trophectoderm-determining transcription factor CDX2[26] (Fig. 3b,c) and the trophectoderm marker
182 Keratin 8[27] (Fig. 3d,e) were significantly decreased in the ERH depleted blastocysts. Thus, ERH
183 is required for proper trophectoderm differentiation.

184 The role of ERH in H3K9me heterochromatin regulation in *S. pombe*[1,13] and human
185 fibroblasts[2] suggests that defects in trophectoderm differentiation may arise from alterations in
186 gene silencing mechanisms. Indeed, we observed a 32% decrease (p=0.0025) in H3K9me3 in
187 blastocysts of ERH-depleted embryos (Fig. 3f,g and Supplementary Table 5). We then performed
188 RNA-seq on mouse blastocysts injected at the 1 cell stage with siControl or siERH
189 (Supplementary Table 6), which confirmed an ~80% knockdown of ERH (Extended Data Fig. 3f).
190 We classified genes upregulated in ERH depleted blastocysts into 4 groups, based upon their
191 expression during normal preimplantation development (Fig. 3h). Group 1 consists of transcripts
192 present at the zygote stage, including maternally inherited transcripts, which become silenced or
193 degraded as development progresses. Group 2 consists of transcripts highly expressed in the 2

194 to 8 cell stages and low or absent in the zygote and blastocyst stage. Group 3 consists of
195 transcripts absent or lowly expressed at all stages measured. Group 4 consists of transcripts with
196 highest expression in the blastocyst stage. Of the ERH-repressed genes (i.e., genes whose
197 expression goes up upon ERH knockdown), 25 normally have peak expression during the 2 to 8
198 cell stage (group 2) and 54 are normally not expressed during preimplantation development
199 (group 3). ERH-repressed group 2 genes included the 2-cell specific *Zscan4d*[28] and the 4-cell
200 specific *D1Pas1*[29] (Fig. 3i). ERH-repressed group 3 alternative lineage genes included the brain-
201 specific *Cyp46a1* neuroprotective enzyme for cholesterol degradation[30] and the liver and
202 intestine-specific carboxylesterase *Ces2g*[31] (Fig. 3i).

203 Comparing our RNA-seq data with existing mouse preimplantation H3K9me3 ChIP-seq
204 [10], we found that the genes repressed by ERH at the 2-cell stage and later, alternative lineage
205 genes were initially marked by H3K9me3 (Fig. 3i). We further found that upon ERH depletion key
206 markers of primitive endoderm, including *Gata4* and *Sox17*[32], were significantly repressed
207 (Extended Data Fig 3g), while epiblast genes were unchanged (Extended Data Fig. 3h). We
208 conclude that ERH is necessary to silence 2-8 cell stage genes and prevent the spontaneous
209 activation of later developmental genes during the blastocyst stage, and that failure to silence
210 results in improper lineage specification in the blastocyst.

211

212 **ERH is necessary to silence naïve and pluripotency genes during human ESC**
213 **differentiation**

214 Our finding that ERH is critical for blastocyst development led us to test its function in pluripotent
215 human ESCs (hESCs). ERH depletion in H1 hESCs cultured in pluripotency-maintaining
216 conditions caused no observable changes in morphology or abundance of pluripotency factors
217 (Fig. 4a). H3K9me3 (Fig. 4b) levels were not changed in siERH treated hESCs, consistent with
218 our observation that ERH-depleted fibroblasts retained hESC-like H3K9me3 (Fig. 1c-e). We also
219 observed no change in H3K27me3 (Extended Data Fig. 4a). Despite no apparent disruption of

220 the pluripotent state or global H3K9me3 levels, ERH-depleted hESCs show an increased nuclear
221 area and nuclear blebbing (Extended Data Fig. 4b), which are changes associated with chromatin
222 decompaction[33]. ERH depletion in self-renewing hESCs resulted in substantial differential gene
223 expression (Extended Data Fig. 4c and Supplementary Table 7) despite not disrupting
224 pluripotency per se (Fig. 4a). Analysis of genes upregulated in naïve vs. primed hESCs[34] showed
225 that ERH depletion in hESCs caused upregulation of genes associated with naïve pluripotency
226 (Extended Data Fig. 4d). Our results demonstrate that ERH represses naïve pluripotency genes
227 in primed H1 hESCs.

228 Next, we asked if ERH depletion impacted the ability of hESCs to differentiate. Activin A
229 and the GSK-3 Inhibitor CHIR99021 together differentiate hESCs to an endoderm identity[35]. As
230 expected, within 24 hours of treatment of undifferentiated H1 hESCs, we observed repression of
231 pluripotency factors and initial expression of endodermal markers in a subset of cells (Extended
232 Data Fig. 4e). Yet depletion of ERH in H1 hESCs followed by 24 hour endoderm differentiation
233 showed increased NANOG protein levels (Fig. 4c and Extended Data Fig. 4f) and a reduced
234 emergence of SOX17+ cells (Fig. 4c). To assess if this observation was specific to endoderm,
235 we also performed differentiation to ectoderm, with control and ERH depleted cells, for 2 days
236 with retinoic acid treatment. We observed sustained expression of the pluripotency factor OCT4
237 and decreased abundance of the ectodermal transcription factor PAX6 (Fig. 4d).

238 To investigate transcriptional changes caused by ERH depletion, we performed RNA-seq
239 of siRNA control and ERH depleted H1 hESCs (Extended Data Fig. 4g and Supplementary Table
240 7) after 1 day of endoderm or 2 days of ectoderm differentiation. Genes normally repressed in the
241 siControl endoderm and ectoderm differentiations were preferentially upregulated by ERH
242 depletion (endoderm, Extended Data Fig. 4h; ectoderm Extended Data Fig. 4i). During endoderm
243 differentiation we observed that ERH depletion preferentially upregulated genes identified as
244 higher in naïve vs primed hESCs[34] (Extended Data Fig. 4j), beyond what was observed in the
245 pluripotent state (Extended Data Fig. 4d). For example, the pluripotency factor NANOG,

246 previously demonstrated to be repressed by H3K9me3 during pluripotency exit[36], was significantly
247 upregulated by ERH depletion (Fig. 4e), consistent with our immunostaining (Fig. 4c and
248 Extended Data Fig. 4f). Additionally, genes found to be expressed in human pluripotent stem cells
249 with a high capacity for self-renewal[37] were upregulated during endoderm and ectoderm
250 differentiation after ERH depletion, including *DUSP5/6, NODAL, LEFTY1/2*, and *DACT1*
251 (Extended Data Fig. 4k and Supplementary Table 7). Genes associated with 2 cell stage and
252 naïve pluripotency including *ZSCAN4*[28], *KLF5*[24], *ESRRB*[38], *NR5A2*[39] and *NANOGP1*[40] were
253 upregulated during siERH treated hESC differentiation (Fig. 4f and Extended Data Fig. 4l). We
254 conclude that ERH is required to repress pluripotency and maintain repression of naïve genes
255 during hESC pluripotency exit.
256
257 **ERH regulates transposable elements with cis-regulatory function**
258 Transposable elements regulate developmental gene networks by acting as enhancers for
259 transcription factor binding[24,41–44]. We therefore assessed transposable elements in our RNA-seq
260 data from siControl and siERH treated H1 hESCs during self-renewal and differentiation
261 (Supplementary Table 8). Transposable elements upregulated by ERH knockdown in H1,
262 definitive endoderm, and ectoderm included several classes of L1 elements and the hominid-
263 specific[45] SINE VNTR and Alu (SVA) elements (Fig. 4g). Satellite repeats were also derepressed
264 (Fig. 4g). Differentiation-specific ERH-repressed transposable element classes, shared between
265 definitive endoderm and ectoderm conditions, included human morula-blastocyst specific
266 *LTR7Ys*[24] and human 8 cell stage expressed *HERVKs*[46] (Fig. 4g). After stratifying L1 elements
267 by their evolutionary age[47], we found that ERH preferentially represses evolutionarily young L1
268 repeats in pluripotent and differentiating hESCs, but broadly repressed L1 elements in terminally
269 differentiated fibroblasts (Fig. 4h). Also upregulated by siERH were 8 cell stage *HERVK9-int* and
270 *HERVK11* transposable elements and their LTR regulatory element *LTR5_Hs* (Extended Data

271 Fig. 5a,b) as well as *SVA* transposable elements (Extended Data Fig. 5c) which are specific to
272 the naïve state[41,48].

273 Next, we wished to identify instances of co-expression of genes and nearby repeat
274 elements. By filtering our data for genes expressed during pluripotency that exhibit ERH-
275 dependent repression during differentiation, we identified instances with proximal evolutionarily
276 young transposable elements (Fig. 3c and Extended Data Fig. 5d,e). Globally, the locus specific
277 expression changes in siERH endoderm and fibroblasts+OSKM of L1 repeats correlated with the
278 gene expression of the closest gene within 25 kb (Extended Data Fig. 5f). The ERH repressed
279 L1 elements gain H3K9me3 during pluripotency to definitive endoderm differentiation (Fig. 4i).
280 This contrasts with H3K9me3 levels at gene bodies of genes upregulated in siERH DE vs
281 siControl DE, which only gain a modest increase in H3K9me3 levels during pluripotency to
282 endoderm differentiation (Extended Data Fig. 5g); indicating the role of H3K9me3 and ERH at
283 this cell fate change occurs more at transposable element regions than gene bodies. We
284 conclude that ERH silences repeat elements that function as regulatory elements for promoting
285 pluripotency and naïve states.

286

287 **ERH restricts lineage commitment during human ESC differentiation**

288 During differentiation, a key function of H3K9me3 heterochromatin is to repress alternative
289 lineage genes[3,10]. Our finding that ERH depletion in blastocysts upregulates lineage-incompatible
290 genes (Fig. 3h,i) led us to investigate if a similar consequence occurs during hESC differentiation.
291 GO analysis of genes upregulated by ERH depletion during endoderm and ectoderm
292 differentiation confirmed activation of transcriptional networks typically not induced by the specific
293 differentiation treatment (Fig. 5a, Extended Data Fig. 6a and Supplementary Table 9). ERH
294 depletion upregulated meiotic and gametogenic genes in all tested developmental stages
295 (Extended Data Fig. 6b), consistent with previous findings in human fibroblasts[2]. To further assess
296 the cell lineages in which the upregulated genes function, we generated composite expression

297 profiles of the genes upregulated by ERH depletion during endoderm (Fig. 5b) and ectoderm
298 differentiation (Fig. 5c) from expression data downloaded from the BBI Descartes human single
299 cell atlas[49], and found that heart, muscle and liver genes are preferentially activated. Directed
300 differentiation of hESCs with Activin A and WNT activation, which typically induces definitive
301 endoderm at high efficiency[35], instead induced both early and late mesodermal genes[50] when
302 ERH was depleted (Fig. 5d,e). Notable precocious gene activation in siERH definitive endoderm
303 included *NXK2-5, NPPB* and *FLNC*, typically expressed in heart muscle, as well as *TAGLN,* a
304 canonical smooth muscle marker (Fig. 5e). Increased levels of BRACHYURY protein, a key
305 determinant of early mesoderm[51], was confirmed by immunofluorescence (Fig. 5f). Off target gene
306 activation during ectoderm differentiation included genes typically activated in liver development
307 (Extended Data Fig. 6c). Thus, while ERH appears dispensable to maintain hESC pluripotency,
308 it is essential to restrict cell identity for proper germ layer differentiation.

309 To assess the role of ERH in the repression of lineage-specific transposable elements
310 during differentiation, we used recent annotations of these elements for the early cell lineages[44]
311 (Fig. 5g). Notably, *MER11A* and *MER11C* elements, which are silenced by H3K9me3 during in
312 vitro endoderm differentiation[44], were significantly upregulated in siERH endoderm differentiation
313 (Fig. 5g and Extended Data 6d). Transposable elements activated by specific transcription
314 factors[42,44], such as *MER11C* by Brachyury and *LTR6B* by GATA6, demonstrated increased
315 activation in the siERH endoderm where the transcription factors are expressed, but were not
316 activated under pluripotent conditions or during ectoderm differentiation where the factors are
317 silent (Fig. 5g). Notably, we were able to map reads to specific repeat element sites and observed
318 instances where the repeat element and adjacent genes were upregulated following ERH
319 depletion in regions marked by H3K9me3 in hESC derived endoderm (Fig. 5h and Extended Data
320 6e; note different scales on subsets to right). ERH therefore represses evolutionarily young and
321 lineage-specific repeats, which in turn repress pluripotency and alternative lineage genes during
322 differentiation.

323 **ERH is recruited to chromatin by cell-specific cofactors**

324 Our observation that ERH is required for fibroblast H3K9me3 maintenance (Fig. 1b) but
325 dispensable for hESC H3K9me3 (Fig. 4b) led us to investigate potential ERH protein interactions
326 that may drive this difference in regulatory mechanism. In *S. pombe* Erh is recruited by the RNA
327 binding factor Mmi1[1]. Mmi1 is a YTH domain-containing RNA binding factor, so we investigated
328 potential interactions of ERH with human YTH domain proteins YTHDC1 and YTHDC2. YTHDC1
329 regulates H3K9me3 through SETDB1 recruitment and repression of repeat elements in mouse
330 ESCs[52,53] and YTHDC2 has been shown to positively regulate transposable elements and the
331 non-coding RNA LINC-ROR in human hESCs[54]. To specifically interrogate on chromatin ERH
332 interacting proteins, we employed a strategy of isolating soluble nucleoplasm (NPL) and insoluble
333 chromatin-associated fractions (CAF) and then performing immunoprecipitation of ERH (Fig. 6a).
334 In fibroblasts but not hESCs, we observed that ERH interacts with YTHDC1 in the chromatin
335 fraction (Fig. 6b,c). ERH did not interact with YTHDC2 in the chromatin fraction in either hESCs
336 or fibroblasts (Fig. 6b,c). We also probed for SAFB1, a known ERH interacting protein[55], and
337 observed consistent ERH-SAFB1 interactions in both fibroblasts and pluripotent hESCs (Fig.
338 6b,c), demonstrating that YTHDC1 but not SAFB1 interactions with ERH are cell type-dependent.

339 We next tested if chromatin association of ERH was RNA dependent and if either YTHDC1
340 or YTHDC2 facilitated this recruitment. We employed an immunostaining approach[56] in which
341 permeabilized nuclei were treated with RNase A, and then stained for ERH, and found that ERH
342 nuclear retention was significantly decreased (Fig. 6d). In contrast, RNase H treatment, which
343 digests DNA:RNA hybrids, did not release ERH from the nucleus (Fig. 6d). In an orthogonal
344 approach, RNase A treatment of fibroblasts during soluble nuclear protein isolation released a
345 portion of chromatin-bound ERH into the nucleoplasm, indicating that its chromatin retention is at
346 least partially RNA-dependent (Extended Data Fig. 7a). To determine if YTHDC1 or YTHDC2
347 facilitate ERH chromatin recruitment, we targeted each separately with siRNAs in human

348 fibroblasts and observed that YTHDC1 but not YTHDC2 depletion led to decreased levels of ERH
349 in the chromatin fraction and increased ERH in the soluble nuclear fraction (Fig. 7e). We
350 confirmed the role of YTHDC1 in ERH retention utilizing our cell permeabilization and
351 immunostaining assay (Extended Data Fig. 7b). From these experiments we concluded that ERH
352 associates with chromatin with YTHDC1 in fibroblasts, but not in hESCs, and that a subset of
353 ERH protein requires RNA and YTHDC1 for chromatin localization.

354 We next assessed for co-regulation of common targets by ERH, YTHDC1 and YTHDC2,
355 first by evaluating expression changes of published YTHDC1 and YTHDC2 targets, ERVK
356 elements repressed by YTHDC1[57], and HERVH elements activated by YTHDC2[54]. ERH
357 represses HERVK and HERVH elements in fibroblasts treated with OSKM factors but not in
358 pluripotent hESCs (Fig. 6f), matching the interaction pattern of ERH with YTHDC1 (Fig. 6b,c).
359 Major satellite repeats, which are regulated by SAFB1[58], are repressed by ERH in both fibroblasts
360 and pluripotent hESCs (Fig. 6f).

361 As we found that ERH interacts with YTHDC1 in differentiated cells but not pluripotent
362 cells, we investigated if YTHDC1 depletion recapitulates aspects of ERH depletion during hESC
363 differentiation. Depletion of YTHDC1 but not YTHDC2 during endoderm differentiation of hESCs
364 resulted in an increase of Brachyury staining (Fig. 6g), matching our findings from ERH depletion
365 during endoderm differentiation (Fig. 5f). We found that ERH repressed many genes not regulated
366 by its interacting partners YTHDC1 and SAFB1, which predominately repress different subsets of
367 ERH targets (Fig. 6h). Taken together, the results indicate that various RNA binding proteins
368 function non-redundantly to recruit ERH to different targets. Based upon these findings, we
369 propose that ERH interactions with RNA binding proteins, including YTHDC1 and SAFB1,
370 facilitate cell-specific repression of genes and repeat elements (Fig. 6i).

371

372 **Discussion**

373 We discovered that depletion of ERH reverts the H3K9me3 landscape in terminally differentiated
374 cells to an hESC-like state, permissive to activation of early embryonic genes. This surprising
375 result led us to investigate the role of ERH in establishing gene repression during the earliest
376 stages of mammalian lineage segregation, in maintaining repression of pluripotency, and enabling
377 germ layer specification. We found that ERH controls H3K9me3 for repression of naïve and
378 alternative lineage genes during the cell fate commitment of totipotent cells to ICM and
379 trophectoderm in mouse blastocysts as well as of hESCs to the endoderm and ectoderm cell
380 lineages, to enforce cell fate and prevent anomalous gene activation. Between the three systems
381 we employed, mouse preimplantation development, hESC differentiation, and fibroblast iPSC
382 reprogramming, we observe that ERH commonly repressed a shared set of naïve genes such as
383 *ZSCAN4* across all systems, indicating that ERH is involved in both the initial repression of this
384 early transcriptional program and in maintaining repression in terminally differentiated cells. We
385 find it remarkable that ERH, which is the sole apparent mechanism for repressing meiotic genes
386 in *S. pombe*, is conserved in repressing germ cell genes in mammals[2] and has expanded to
387 foundationally control diverse other types of developmental genes and stages.

388 In a developmental context, the failure to repress 2-cell, 8-cell and pluripotency genes
389 after ERH knockdown results in delays of specific-lineage gene expression, including primitive
390 endoderm genes in mouse blastocysts and various endoderm and ectoderm genes during hESC
391 differentiation. Furthermore, we demonstrated that ERH is necessary to maintain repression of
392 alternative lineage genes, as ERH depletion elicited precocious expression of genes typical of
393 later cell differentiation in mouse blastocysts and human hESCs, as well as heterochromatized
394 genes during initial iPSC reprogramming. ERH-dependent gene regulation likely occurs not only
395 through changes in direct repression of genes but also through the regulation of transposable
396 elements, which are canonically repressed by H3K9me3 and serve as developmental regulatory
397 elements[41,42,44]. We observed that evolutionarily young transposable elements, including the

398 human specific *LTR7Ys*[24], are especially upregulated by ERH depletion during hESC
399 differentiation and iPSC reprogramming. Much of the lineage specific expression changes upon
400 ERH depletion during hESC differentiation may be attributed to increased accessibility of
401 transcription factor binding sites among de-repressed repeats[24,41,44].

402 A key finding of our analysis is that ERH depletion in human fibroblasts reverted the
403 H3K9me3 chromatin landscape to an hESC-like state. The cryptic hESC-like H3K9me3
404 landscape functionally permits binding of the SOX2 factor in fibroblasts to endogenous hESC
405 SOX2 targets, activating embryonic genes. That ERH is necessary for maintaining the broad
406 H3K9me3 domains observed in somatic cells, but is dispensable for maintaining hESC H3K9me3
407 peaks, is consistent with our observation that ERH depletion in hESCs did not change global
408 H3K9me3 levels. We suggest that the residual H3K9me3 peaks in pluripotent cells, enriched for
409 zinc finger motifs, may be H3K9me3 nucleation sites and that ERH is essential for the spread of
410 H3K9me3 from these sites to establish the typical broad domains observed in human somatic
411 cells. That many residual H3K9me3 peaks in ERH-depleted fibroblasts are seen in the pluripotent
412 state suggests that many somatic H3K9me3 domains may be nucleated early in development,
413 only to be expanded to form repressive domains at developmentally appropriate times.
414 Importantly, when ERH is knocked down in fibroblasts, we observe increased activation of many
415 genes and transposable elements that harbor a residual H3K9me3 peak, yet lose broad
416 H3K9me3, indicating that residual H3K9me3 peaks are not sufficient to block gene activation. Our
417 knowledge of H3K9 methylation nucleation and spread has been primarily elucidated through
418 studies in *S. pombe*[59–61], leaving further investigation needed into how H3K9me3 nucleation and
419 spread occurs in multicellular organisms.

420 We identify that ERH interacts with and is partially recruited to the chromatin by the RNA
421 binding protein YTHDC1 in somatic cells, but not pluripotent hESCs. We find that ERH and
422 YTHDC1 repress similar targets in fibroblasts and differentiating hESCs, but not in homeostatic
423 pluripotent hESCs. In contrast, we observe that ERH interacts with the RNA binding protein

424 SAFB1 in both hESCs and human fibroblasts and likewise both repress the satellite repeat
425 expression in hESCs and fibroblasts. Our knockdowns of YTHDC1 and gene expression analysis
426 of YTHDC1 and SAFB1 fail to fully explain how ERH is recruited to chromatin to repress
427 transcription. We thus speculate that ERH is recruited to cell-type specific genes and repeats
428 through a diverse network of chromatin-associated RNA binding factors that are yet to be fully
429 identified.

430 Comparing the role of ERH in H3K9me3 and cell fate regulation in *S. pombe* and
431 mammalian systems provides insight into how H3K9me3 heterochromatin has been adapted
432 through evolution to the purpose of cell fate regulation in complex multicellular mammals. We
433 observe in mouse development and hESC systems that ERH is functionally required in multiple
434 developmental contexts to repress meiotic and gametogenic genes, as it is in *S. pombe*[1], although
435 the gene members targeted by ERH has greatly evolved and expanded in mammals. Our findings
436 on meiotic gene repression are supported by the recent identification of ERH in a screen for
437 proteins silencing germline genes in mESCs[62]. Other aspects of *Erh1* function in *S. pombe,* such
438 as the targeting of RNAs for degradation by the RNA exosome[1], appear to be conserved as we
439 observed upregulation of developmental genes typically downregulated by RNA exosome
440 degradation[63]. ERH was recently found by mass spectrometry to associate with the
441 Microprocessor and NEXT complexes in mouse ESCs[64]. Evolutionarily divergent functions of
442 ERH may be driven by evolutionarily young transposable elements that retain RNA expression,
443 similar to how transposable elements have influenced the evolution of KRAB Zinc finger proteins
444 DNA dependent targeting of H3K9me3[65,66].

445 We found that ERH enacts gene repression at the earliest stages of development and that
446 it is essential for maintaining this repression in terminally differentiated cells. Our findings advance
447 the understanding of how heterochromatic repression is established and maintained during
448 development. We propose that understanding mechanisms by which ERH is employed to direct

449 H3K9me3 to specific loci in a developmentally coordinated fashion will be crucial to modulating
450 developmental H3K9me3 rearrangements to achieve greater control of cell identity.

451 **References**

452

453 1. Sugiyama, T. *et al.* Enhancer of Rudimentary Cooperates with Conserved RNA-Processing Factors to
454 Promote Meiotic mRNA Decay and Facultative Heterochromatin Assembly. *Mol Cell* **61**, 747–759
455 (2016).

456 2. McCarthy, R. L. *et al.* Diverse heterochromatin-associated proteins repress distinct classes of genes
457 and repetitive elements. *Nat Cell Biol* **23**, 905–914 (2021).

458 3. Nicetto, D. *et al.* H3K9me3-heterochromatin loss at protein-coding genes enables developmental
459 lineage specification. *Science* **363**, 294–297 (2019).

460 4. Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. A. SetDB1 contributes to
461 repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes &*
462 *development* **23**, 2484–9 (2009).

463 5. Becker, J. S. *et al.* Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of
464 Alternate Fate Genes. *Mol Cell* **68**, 1023-1037.e15 (2017).

465 6. Xu, R. *et al.* Stage-specific H3K9me3 occupancy ensures retrotransposon silencing in human pre-
466 implantation embryos. *Cell Stem Cell* **29**, 1051-1066.e8 (2022).

467 7. Schulz, K. N. & Harrison, M. M. Mechanisms regulating zygotic genome activation. *Nat Rev Genet* **20**,
468 221–234 (2019).

469 8. Xu, R. *et al.* Unreprogrammed H3K9me3 prevents minor zygotic genome activation and lineage
470 commitment in SCNT embryos. *Nat Commun* **14**, 4807 (2023).

471 9. Sankar, A. *et al.* KDM4A regulates the maternal-to-zygotic transition by protecting broad H3K4me3
472 domains from H3K9me3 invasion in oocytes. *Nat Cell Biol* **22**, 380–388 (2020).

473 10. Wang, C. *et al.* Reprogramming of H3K9me3-dependent heterochromatin during mammalian
474 embryo development. *Nature cell biology* **20**, 620–631 (2018).

475 11. Yu, H. *et al.* Dynamic reprogramming of H3K9me3 at hominoid-specific retrotransposons during
476 human preimplantation development. *Cell Stem Cell* **29**, 1031-1050.e12 (2022).
477 12. Burton, A. *et al.* Heterochromatin establishment during early mammalian development is regulated
478 by pericentromeric RNA and characterized by non-repressive H3K9me3. *Nat Cell Biol* **22**, 767–778
479 (2020).
480 13. Xie, G. *et al.* A conserved dimer interface connects ERH and YTH family proteins to promote gene
481 silencing. *Nat Commun* **10**, 251 (2019).
482 14. Kuznetsov, D. *et al.* OrthoDB v11: annotation of orthologs in the widest sampling of organismal
483 diversity. *Nucleic Acids Res* **51**, D445–D451 (2023).
484 15. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency
485 reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
486 16. Taing, L. *et al.* Cistrome Data Browser: integrated search, analysis and visualization of chromatin
487 data. *Nucleic Acids Res* **52**, D61–D66 (2024).
488 17. Zheng, R. *et al.* Cistrome Data Browser: expanded datasets and new tools for gene regulatory
489 analysis. *Nucleic Acids Res* **47**, D729–D735 (2019).
490 18. Yu, X. *et al.* Recapitulating early human development with 8C-like cells. *Cell Rep* **39**, 110994 (2022).
491 19. Zou, Z. *et al.* Translatome and transcriptome co-profiling reveals a role of TPRXs in human zygotic
492 genome activation. *Science* **378**, abo7923 (2022).
493 20. Taubenschmid-Stowers, J. *et al.* 8C-like cells capture the human zygotic genome activation program
494 in vitro. *Cell Stem Cell* **29**, 449-459.e6 (2022).
495 21. Cacchiarelli, D. *et al.* Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of
496 Induced Pluripotency. *Cell* **162**, 412–424 (2015).
497 22. Wang, J. *et al.* Primate-specific endogenous retrovirus-driven transcription defines naive-like stem
498 cells. *Nature* **516**, 405–409 (2014).

499 23. Lu, X. *et al.* The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell
500 identity. *Nat Struct Mol Biol* **21**, 423–425 (2014).
501 24. Ai, Z. *et al.* Krüppel-like factor 5 rewires NANOG regulatory network to activate human naive
502 pluripotency specific LTR7Ys and promote naive pluripotency. *Cell Rep* **40**, 111240 (2022).
503 25. Gao, Y. *et al.* Protein Expression Landscape of Mouse Embryos during Pre-implantation
504 Development. *Cell reports* **21**, 3957–3969 (2017).
505 26. Ralston, A. & Rossant, J. Cdx2 acts downstream of cell polarization to cell-autonomously promote
506 trophectoderm fate in the early mouse embryo. *Dev Biol* **313**, 614–629 (2008).
507 27. Lim, H. Y. G. *et al.* Keratins are asymmetrically inherited fate determinants in the mammalian
508 embryo. *Nature* **585**, 404–409 (2020).
509 28. Falco, G. *et al.* Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic
510 stem cells. *Dev Biol* **307**, 539–550 (2007).
511 29. Hamatani, T., Carter, M. G., Sharov, A. A. & Ko, M. S. H. Dynamics of global gene expression changes
512 during mouse preimplantation development. *Dev Cell* **6**, 117–131 (2004).
513 30. Boussicault, L. *et al.* CYP46A1, the rate-limiting enzyme for cholesterol degradation, is
514 neuroprotective in Huntington's disease. *Brain* **139**, 953–970 (2016).
515 31. Chalhoub, G. *et al.* Carboxylesterase 2 proteins are efficient diglyceride and monoglyceride lipases
516 possibly implicated in metabolic disease. *J Lipid Res* **62**, 100075 (2021).
517 32. Artus, J., Piliszek, A. & Hadjantonakis, A.-K. The primitive endoderm lineage of the mouse blastocyst:
518 sequential transcription factor activation and regulation of differentiation by Sox17. *Dev Biol* **350**,
519 393–404 (2011).
520 33. Stephens, A. D. *et al.* Chromatin histone modifications and rigidity affect nuclear morphology
521 independent of lamins. *Mol Biol Cell* **29**, 220–233 (2018).

522 34. Messmer, T. *et al.* Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells
523 at Single-Cell Resolution. *Cell Rep* **26**, 815-824.e4 (2019).
524 35. Teo, A. K. K., Valdez, I. A., Dirice, E. & Kulkarni, R. N. Comparable generation of activin-induced
525 definitive endoderm via additive Wnt or BMP signaling in absence of serum. *Stem Cell Reports* **3**, 5–
526 14 (2014).
527 36. Dubois, A. *et al.* H3K9 tri-methylation at Nanog times differentiation commitment and enables the
528 acquisition of primitive endoderm fate. *Development* **149**, dev201074 (2022).
529 37. Lau, K. X. *et al.* Unique properties of a subset of human pluripotent stem cells with high capacity for
530 self-renewal. *Nat Commun* **11**, 2420 (2020).
531 38. Martello, G. *et al.* Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-
532 renewal. *Cell Stem Cell* **11**, 491–504 (2012).
533 39. Gassler, J. *et al.* Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science* **378**,
534 1305–1315 (2022).
535 40. Maskalenka, K. *et al.* NANOGP1, a tandem duplicate of NANOG, exhibits partial functional
536 conservation in human naïve pluripotent stem cells. *Development* **150**, dev201155 (2023).
537 41. Pontis, J. *et al.* Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic
538 Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell* **24**, 724-735.e5
539 (2019).
540 42. Wu, F. *et al.* Species-specific rewiring of definitive endoderm developmental gene activation via
541 endogenous retroviruses through TET1-mediated demethylation. *Cell Rep* **41**, 111791 (2022).
542 43. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity.
543 *Nature* **487**, 57–63 (2012).
544 44. Pontis, J. *et al.* Primate-specific transposable elements shape transcriptional networks during
545 human development. *Nat Commun* **13**, 7178 (2022).

546 45. Wang, H. *et al.* SVA elements: a hominid-specific retroposon family. *J Mol Biol* **354**, 994–1007
547 (2005).
548 46. Grow, E. J. *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent
549 cells. *Nature* **522**, 221–5 (2015).
550 47. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1
551 retrotransposons since the origin of primates. *Genome Res* **16**, 78–87 (2006).
552 48. Theunissen, T. W. *et al.* Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem*
553 *Cell* **19**, 502–515 (2016).
554 49. Domcke, S. *et al.* A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7612 (2020).
555 50. Evseenko, D. *et al.* Mapping the first stages of mesoderm commitment during differentiation of
556 human embryonic stem cells. *Proc Natl Acad Sci U S A* **107**, 13742–13747 (2010).
557 51. Bernardo, A. S. *et al.* BRACHYURY and CDX2 mediate BMP-induced differentiation of human and
558 mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell* **9**, 144–
559 155 (2011).
560 52. Liao, S., Sun, H. & Xu, C. YTH Domain: A Family of N6-methyladenosine (m6A) Readers. *Genomics*
561 *Proteomics Bioinformatics* **16**, 99–107 (2018).
562 53. Liu, J. *et al.* The RNA m(6)A reader YTHDC1 silences retrotransposons and guards ES cell identity.
563 *Nature* **591**, 322–326 (2021).
564 54. Sun, T. *et al.* Crosstalk between RNA m6A and DNA methylation regulates transposable element
565 chromatin activation and cell fate in human pluripotent stem cells. *Nat Genet* **55**, 1324–1335 (2023).
566 55. Drakouli, S., Lyberopoulou, A., Papathanassiou, M., Mylonis, I. & Georgatsou, E. Enhancer of
567 rudimentary homologue interacts with scaffold attachment factor B at the nuclear matrix to
568 regulate SR protein phosphorylation. *FEBS J* **284**, 2482–2500 (2017).

569 56. Maison, C. *et al.* Higher-order structure in pericentric heterochromatin involves a distinct pattern of
570 histone modification and an RNA component. *Nature genetics* **30**, 329–34 (2002).
571 57. Xu, W. *et al.* METTL3 regulates heterochromatin in mouse embryonic stem cells. *Nature* **591**, 317–
572 321 (2021).
573 58. Huo, X. *et al.* The Nuclear Matrix Protein SAFB Cooperates with Major Satellite RNAs to Stabilize
574 Heterochromatin Architecture Partially through Phase Separation. *Mol Cell* **77**, 368-383 e7 (2020).
575 59. Obersriebnig, M. J., Pallesen, E. M. H., Sneppen, K., Trusina, A. & Thon, G. Nucleation and spreading
576 of a heterochromatic domain in fission yeast. *Nat Commun* **7**, 11518 (2016).
577 60. Murawska, M. *et al.* The histone chaperone FACT facilitates heterochromatin spreading by
578 regulating histone turnover and H3K9 methylation states. *Cell Rep* **37**, 109944 (2021).
579 61. Greenstein, R. A. *et al.* Noncoding RNA-nucleated heterochromatin spreading is intrinsically labile
580 and requires accessory elements for epigenetic stability. *Elife* **7**, (2018).
581 62. Al Adhami, H. *et al.* Systematic identification of factors involved in the silencing of germline genes in
582 mouse embryonic stem cells. *Nucleic Acids Res* gkad071 (2023) doi:10.1093/nar/gkad071.
583 63. Wu, D. & Dean, J. RNA exosome ribonuclease DIS3 degrades Pou6f1 to promote mouse pre-
584 implantation cell differentiation. *Cell Rep* **42**, 112047 (2023).
585 64. Imamura, K. *et al.* A functional connection between the Microprocessor and a variant NEXT
586 complex. *Mol Cell* **84**, 4158-4174.e6 (2024).
587 65. Yang, P., Wang, Y. & Macfarlan, T. S. The Role of KRAB-ZFPs in Transposable Element Repression and
588 Mammalian Evolution. *Trends Genet* **33**, 871–881 (2017).
589 66. Imbeault, M., Helleboid, P.-Y. & Trono, D. KRAB zinc-finger proteins contribute to the evolution of
590 gene regulatory networks. *Nature* **543**, 550–554 (2017).

591 67. Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. TEtranscripts: a package for including transposable
592 elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593–3599
593 (2015).
594 68. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science* **370**, eaba7721 (2020).
595 69. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of
596 DNA binding sites. *Elife* **6**, e21856 (2017).
597 70. Katznelson, A. & Zaret, K. A fluorescence-based protocol to quantitatively titrate CUT&RUN buffer
598 components. *STAR Protoc* **5**, 102866 (2024).
599 71. Liu, N. *et al.* Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch.
600 *Cell* **173**, 430-442.e17 (2018).
601 72. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359
602 (2012).
603 73. Jin, H. *et al.* ChIPseqSpikeInFree: a ChIP-seq normalization approach to reveal global changes in
604 histone modifications without spike-in. *Bioinformatics* **36**, 1270–1272 (2020).
605 74. Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for
606 CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019).
607 75. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-
608 regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576–89
609 (2010).
610 76. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis.
611 *Nucleic Acids Res* **44**, W160-165 (2016).
612 77. Khan, A. & Mathelier, A. Intervene: a tool for intersection and visualization of multiple gene or
613 genomic region sets. *BMC Bioinformatics* **18**, 287 (2017).
614 78. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

615 79. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning
616 sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
617 80. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq
618 data with DESeq2. *Genome Biol* **15**, 550 (2014).
619 81. Filippi-Chiela, E. C. *et al.* Nuclear morphometric analysis (NMA): screening of senescence, apoptosis
620 and nuclear irregularities. *PLoS One* **7**, e42522 (2012).
621 82. Thomas, P. D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci* **31**,
622 8–22 (2022).
623

624 **Figure legends**

625 Figure 1. ERH depletion reverts fibroblast H3K9me3 patterns to an hESC-like state

626 **a.** Schematic of siRNA treatments and timing for H3K9me3 CUT&RUN experiments shown
627 in figure 1 and SOX2 CUT&RUN and RNA-seq experiments shown in figure 2.

628 **b.** Immunofluorescence of ERH (red), H3K9me3 (red) and DAPI (blue) in siControl and
629 siERH human fibroblasts. Scale bars, 50μm.

630 **c.** Browser track of H3K9me3 levels in siControl and siERH treated human fibroblasts and
631 H1 hESCs (GSM1003585) at the megabase-domain level and gene level. Example
632 H3K9me3 peaks common between siERH fibroblasts and H1 hESCs are indicated by red
633 triangles.

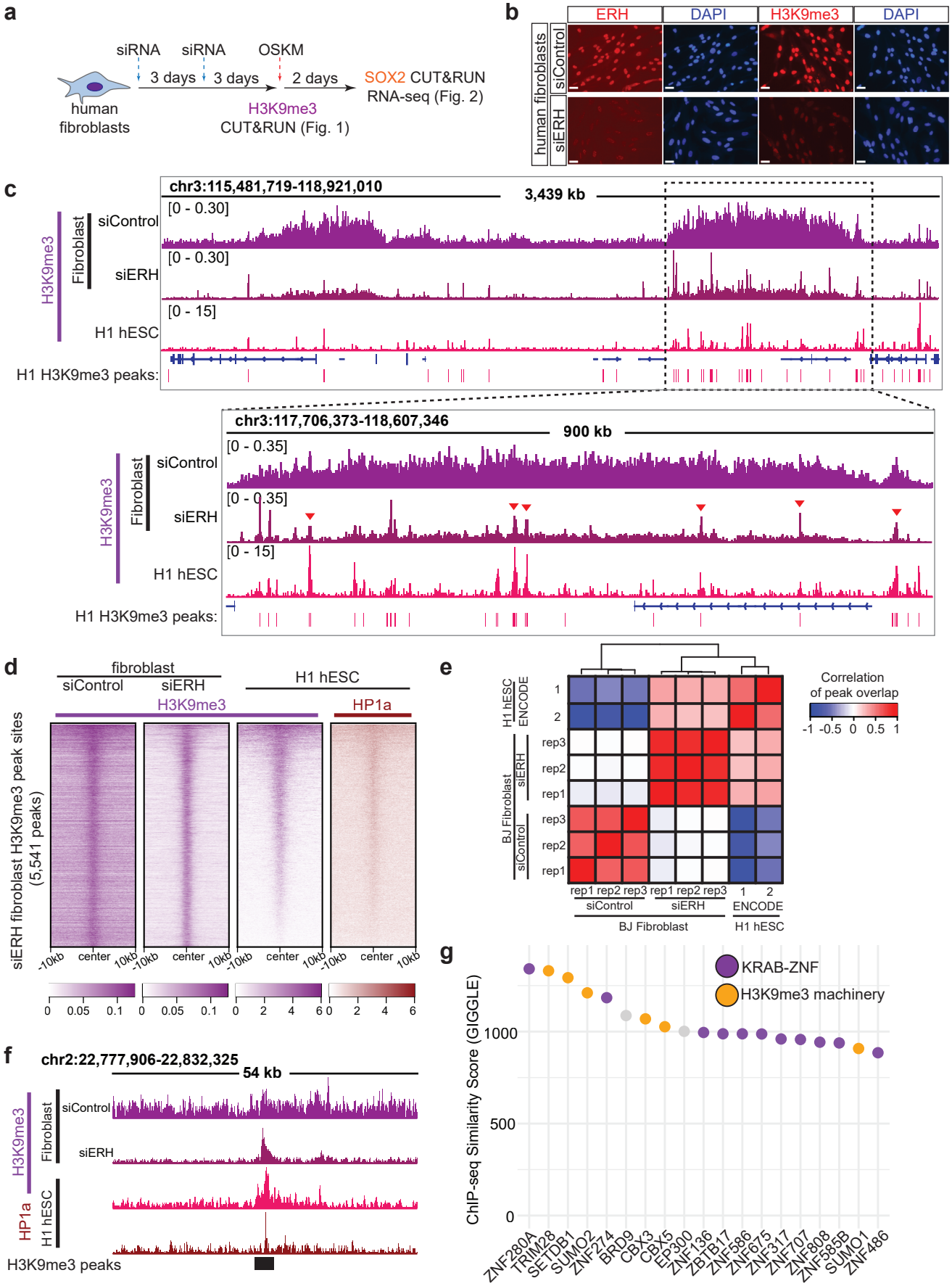
634 **d.** Heatmap of fibroblast siControl and siERH H3K9me3, H1 hESC H3K9me3
635 (GSM1003585) and H1 hESC HP1α (GSE123229) levels centered at siERH retained
636 H3K9me3 peaks located within fibroblast H3K9me3 domains +/-10kb.

637 **e.** Correlation of H3K9me3 from BJ fibroblasts after siControl or siERH with H1 hESC
638 H3K9me3.

639 **f.** Browser track comparing siControl and siERH fibroblast H3K9me3 with H1 hESC
640 H3K9me3 (GSM1003585) and HP1a (GSM1701825) patterns. Peak locations called by
641 SEACR analysis on siERH H3K9me3 data are shown.

642 **g.** CistromeDB Toolkit enrichment analysis of fibroblast siERH H3K9me3 peaks. The top 20
643 proteins scored by the analysis are visualized. Proteins colored by characterization as
644 H3K9me3-associated (orange), KRAB zinc-finger transcription factors (purple).
645 See also Extended Data Fig. 1 and Supplementary Table 1.

Katznelson et al., Figure 1



646 Figure 2. ERH depletion enhances gene activation of pluripotency and alternative lineage genes
647 during iPSC reprogramming

648 **a.** Heatmap of H3K9me3 CUT&RUN from siControl and siERH human fibroblasts and SOX2
649 CUT&RUN from siControl and siERH fibroblasts + OSKM at siERH-gained SOX2 peaks.

650 **b.** Browser track of example hESC SOX2 and NANOG peak coincident with fibroblast siERH
651 H3K9me3 loss and SOX2 peak gain.

652 **c.** Volcano plot of expression changes (DESeq2) in siERH vs siControl fibroblasts+OSKM of
653 iPSC activated genes located in fibroblast H3K9me3 domains.

654 **d.** Expression of ERH and example naïve pluripotency genes activated in 48 hours OSKM
655 by ERH depletion, transcripts per million (TPM) values from RNA-seq replicates shown
656 (DESeq2 adjusted p-value).

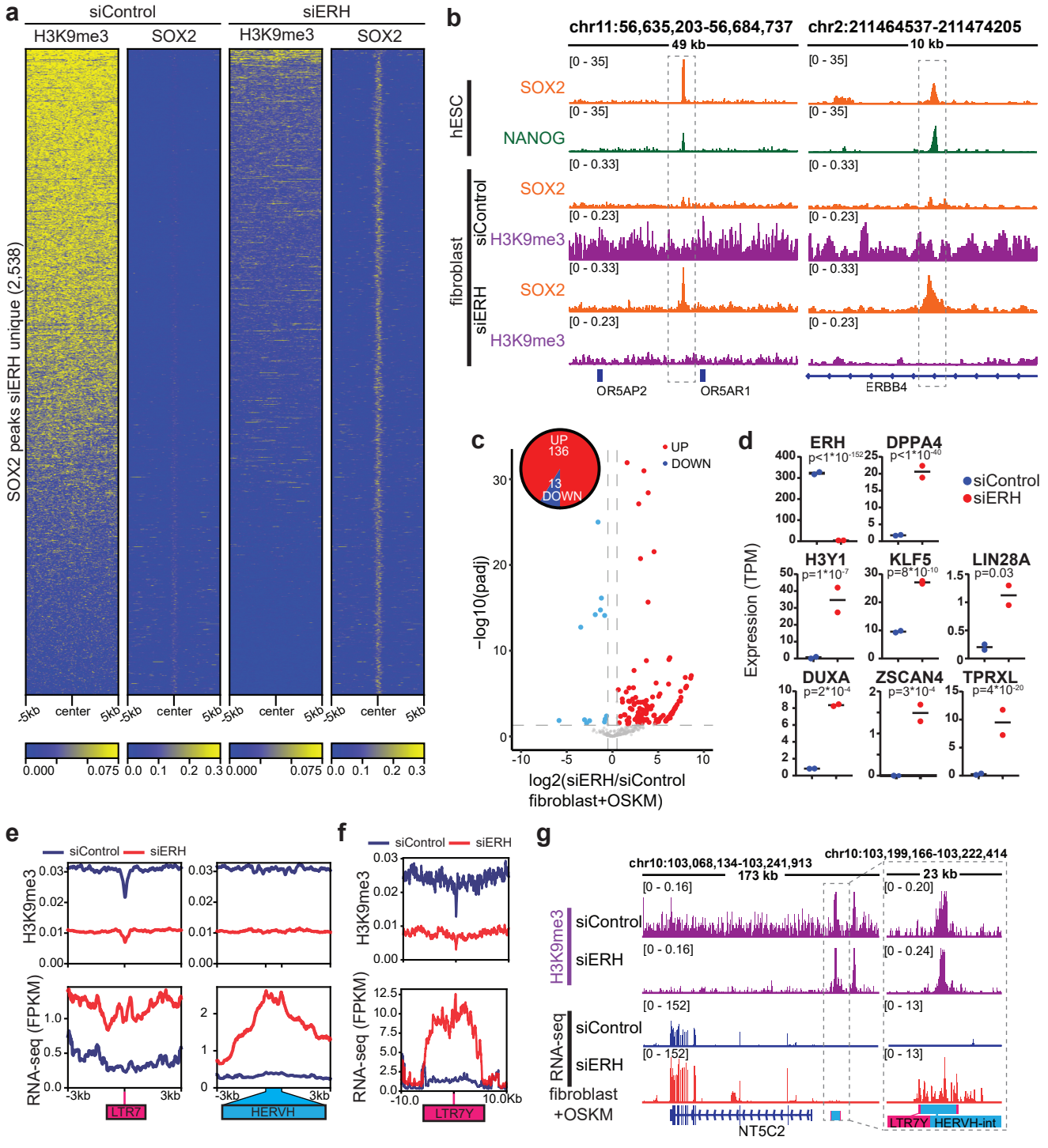
657 **e.** Profile plots of H3K9me3 and RNA levels from LTR7 and HERVH repeats in siERH vs
658 siControl fibroblasts+OSKM.

659 **f.** Profile plots of H3K9me3 and RNA levels from LTR7Y repeats in siERH vs siControl
660 fibroblasts+OSKM.

661 **g.** Browser track showing H3K9me3 in siControl and siERH fibroblasts and RNA levels in
662 siControl+OSKM and siERH+OSKM for gene and adjacent LTR7Y naïve pluripotency
663 repeat.

664 See also Extended Data Fig. 2 and Supplementary Tables 2-4.

Katznelson et al., Figure 2



665 Figure 3. ERH is expressed during pre-implantation mouse development and is critical for normal
666 blastocyst development by repressing 2-cell and alternative lineage genes.

667 **a.** ERH (white) and DAPI (blue) staining during mouse pre-implantation development.
668 Representative confocal z-sections are shown (zygote n=15, N=3, 2-cell stage n=18, N=3,
669 8-cell stage n=6, N=2, blastocyst n=10, N=3). Scale bars, 10μm.

670 **b.** Representative images of trophectoderm marker Cdx2 (green) in siControl and siERH
671 blastocysts with Phalloidin (orange) marking cell boundaries. Scale bars, 10μm.

672 **c.** Quantification of CDX2 intensity in siControl vs siERH blastocysts (unpaired t-test, n=8
673 blastocysts for each siRNA).

674 **d.** Representative image of trophectoderm marker KRT8 (orange) in siControl and siERH
675 blastocysts with DAPI (blue). Scale bars, 10μm.

676 **e.** Quantification of KRT8 intensity in siControl vs siERH blastocysts (unpaired t-test, n=9
677 blastocysts, n=8 blastocysts for each siRNA).

678 **f.** Representative images of H3K9me3 (magenta), DAPI (teal) in siControl vs siERH
679 blastocysts. Scale bars, 10μm.

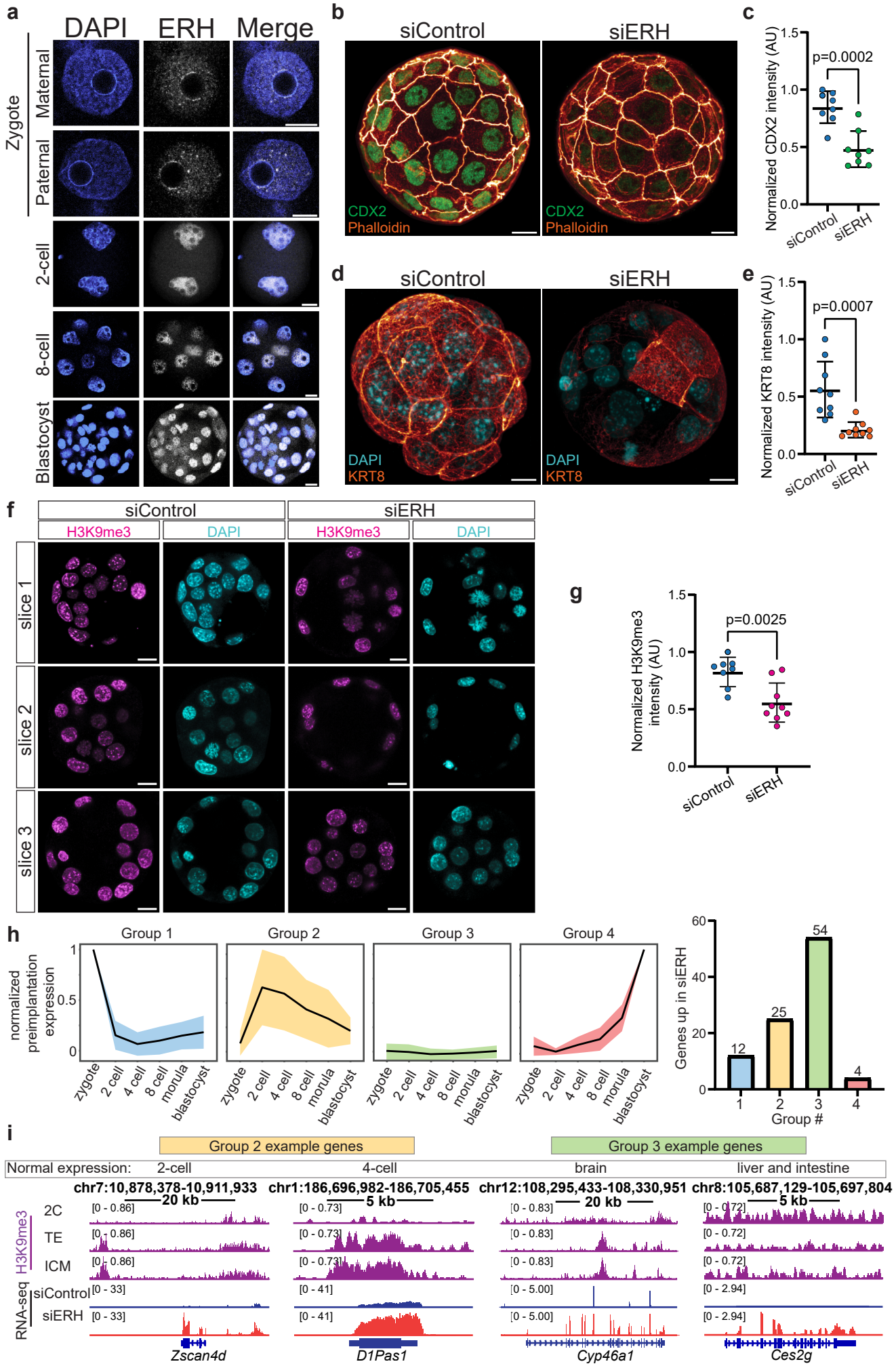
680 **g.** Quantification of H3K9me3 level in siControl vs siERH blastocysts. (unpaired t-test, n=8
681 for siControl and n=9 for siERH blastocysts).

682 **h.** ERH repressed blastocyst genes grouped based upon normal expression patterns
683 exhibited during mouse pre-implantation development re-analyzed from [25].

684 **i.** Example gene tracks of genes significantly upregulated in siERH vs siControl blastocysts
685 (adjusted p-value<=0.05, log2fc>=0.05) with H3K9me3 tracks for 2 cell, trophectoderm
686 and ICM cells from[10].

687 See also Extended Data 3 and Supplementary Tables 5,6.

Katznelson et al., Figure 3



688 Figure 4. ERH is dispensable for pluripotency maintenance but is involved in silencing of
689 pluripotency and naïve genes during differentiation

690 **a.** Representative images showing hESC morphology and SOX2 (green) pluripotency
691 marker staining in siControl and siERH H1 hESCs under pluripotency maintaining
692 conditions. Scale bars, 50μm.

693 **b.** Representative images of H3K9me3 (red) and DAPI (blue) in siControl and siERH H1
694 hESCs under pluripotency maintaining conditions. Scale bars, 50μm.

695 **c.** Representative images of NANOG (red), SOX17 (red) and DAPI (blue) during 1 day of
696 endoderm differentiation of siControl and siERH H1 hESCs. Scale bars, 50μm.

697 **d.** Representative images of OCT4 (red), PAX6 (green) and DAPI (blue) during 2 days of
698 ectoderm differentiation of siControl and siERH H1 hESCs. Scale bars, 50μm.

699 **e.** Plot of NANOG expression in transcripts per million (TPM) by RNA-seq in siControl and
700 siERH treated H1 hESCs cultured in pluripotency (H1), 1 day endoderm differentiation
701 (DE) and 2 day ectoderm differentiation (EC). (DESeq2 adjusted p-value reported).

702 **f.** Plot of KLF5 and ZSCAN4 expression in transcripts per million (TPM) by RNA-seq in
703 siControl and siERH treated H1 hESCs cultured in pluripotency (H1), 1 day endoderm
704 differentiation (DE) and 2 day ectoderm differentiation (EC). (DESeq2 adjusted p-value
705 reported).

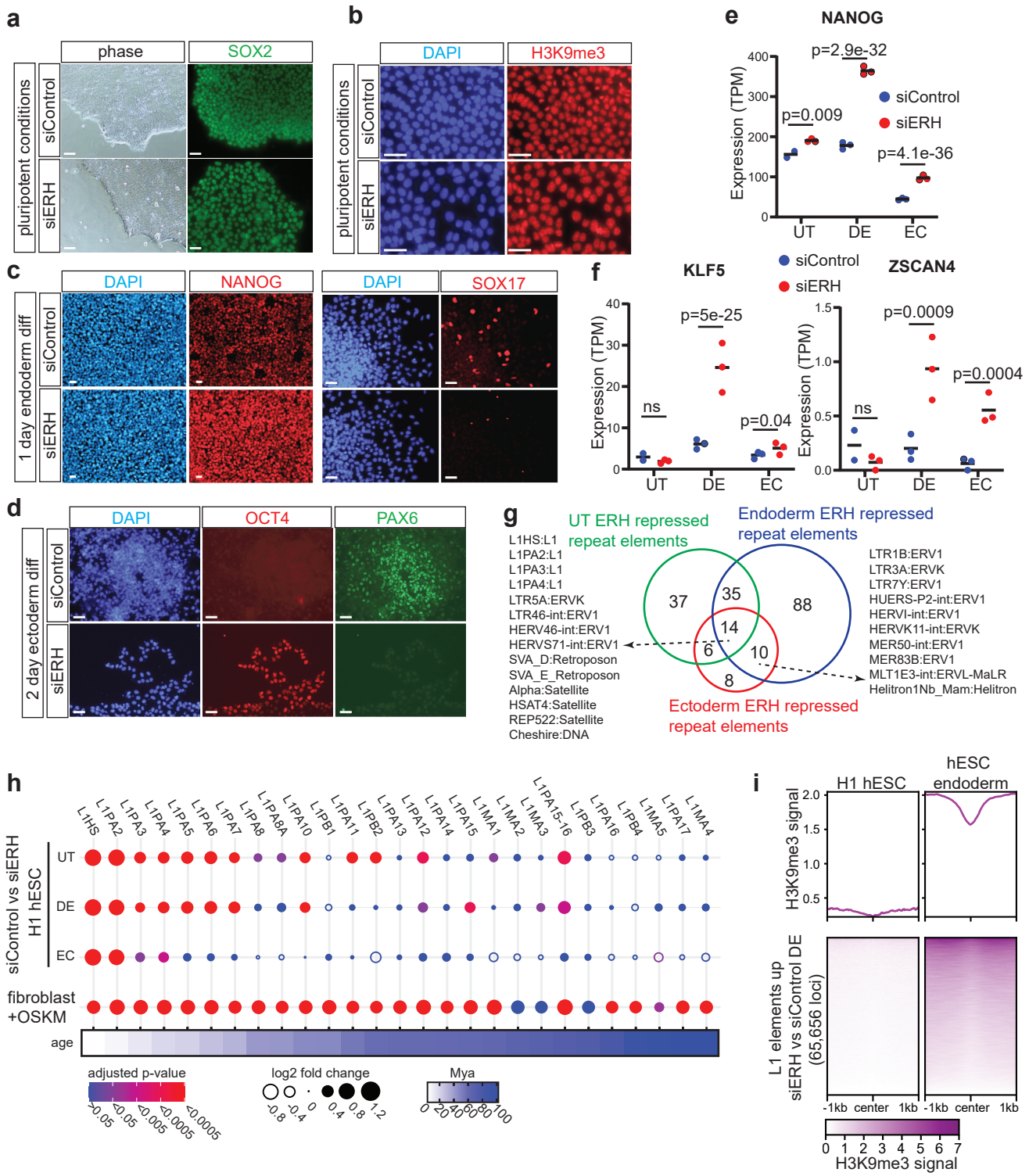
706 **g.** Venn diagram showing intersect of repeat classes upregulated (TETranscripts, padj<0.05,
707 log2fc>0.5) by ERH depletion in H1 hESCs untreated (UT), 1 day endoderm (DE) or 2 day
708 ectoderm (EC) conditions.

709 **h.** Expression change and significance, TEtranscripts[67] analysis, of L1 LTR transposable
710 elements in ERH depletion in H1 hESCs untreated (UT), 1 day endoderm (DE) or 2 day
711 ectoderm (EC) listed in order of evolutionary age as determined in[47].

712 **i.** H3K9me3 data from H1 hESCs (GSM1003585) and hESC derived endoderm
713 (GSM916057) at 65,656 L1 element loci +/- 1kb upregulated (adjusted pvalue<=0.05;
714 log2foldchange>=0.5) by ERH depletion during 1 day endoderm differentiation.

715 See also Extended Data Figs. 4,5 and Supplementary Tables 7,8.

Katznelson et al., Figure 4



716 Figure 5. ERH represses alternative lineage gene activation during hESC differentiation

717 **a.** GO analysis (PantherDB, statistical overrepresentation) of genes upregulated
718 (padj<=0.05, log2 fold change >= 0.5) in siERH vs siControl H1 hESCs 1 day endoderm
719 differentiation.

720 **b.** Radar plot showing cumulative expression profile across indicated human tissues, data
721 downloaded from BBI Descartes database[68], for genes upregulated (padj<=0.05, log2 fold
722 change >= 0.5) in siERH vs siControl H1 hESCs 1 day endoderm differentiation.

723 **c.** Radar plot showing cumulative expression profile across indicated human tissues, data
724 downloaded from BBI Descartes database[68], for genes upregulated (padj<=0.05, log2 fold
725 change >= 0.5) in siERH vs siControl H1 hESCs 2 day ectoderm differentiation.

726 **d.** Example early mesodermal genes activated during endoderm differentiation by ERH
727 depletion, transcripts per million (TPM) values from RNA-seq replicates shown (DESeq2
728 adjusted p-value).

729 **e.** Example late mesodermal genes activated during endoderm differentiation by ERH
730 depletion, transcripts per million (TPM) values from RNA-seq replicates shown DESeq2
731 adjusted p-value).

732 **f.** Representative image of DAPI (blue), BRACHYURY (red) of siControl and siERH treated
733 H1 hESCs after 1 day of endoderm differentiation. Scale bars, 50μm.

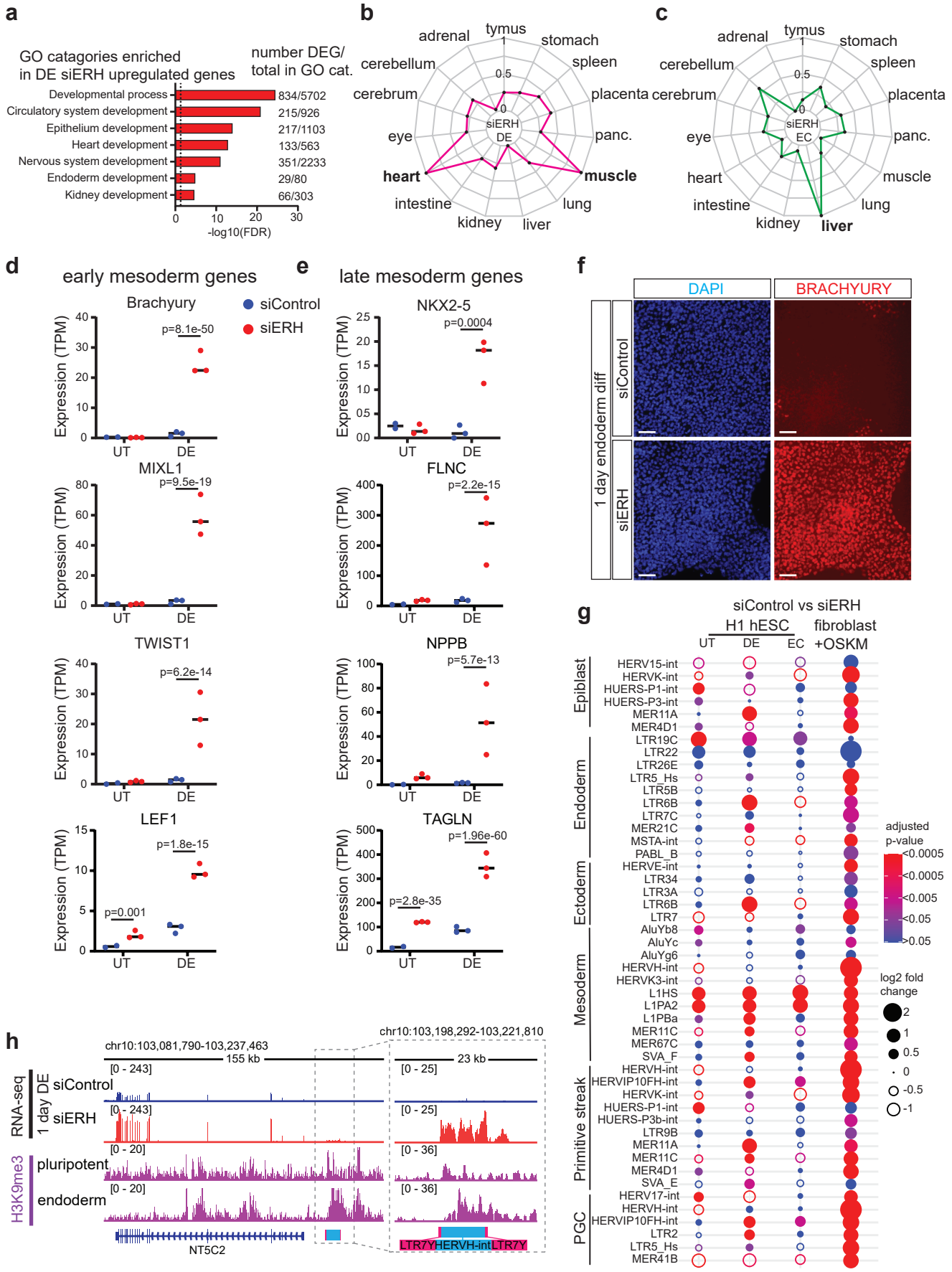
734 **g.** Expression change and significance, TEtranscripts[67] analysis, of transposable elements
735 by ERH depletion in H1 hESCs untreated (UT), 1 day endoderm (DE) or 2 day ectoderm
736 (EC) grouped by lineage/stage at which they are typically expressed as determined in[44].
737 Negative log2 fold changes are represented as unfilled circles.

738 **h.** Browser track showing RNA-seq of example repeat element and nearby genes in siControl
739 and siERH treated H1 hESCs after 1 day of endoderm differentiation and H3K9me3 tracks
740 from H1 hESCs (GSM1003585) and hESC derived endoderm (GSM916057).

741 See also Extended Data Fig. 6 and Supplementary Table 9.

742

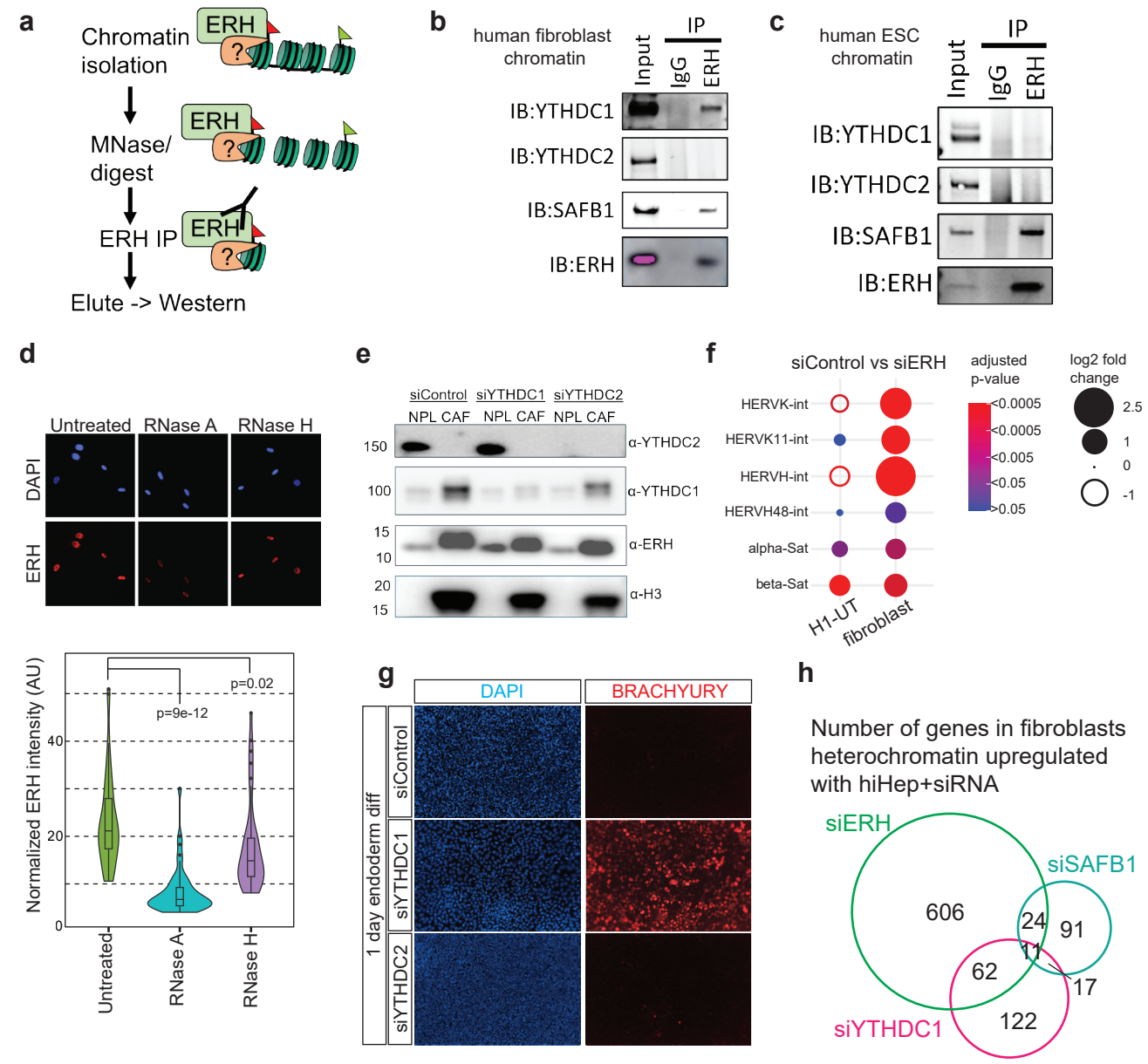
Katznelson et al., Figure 5



743 Figure 6. Cell type specific and independent recruitment of ERH to chromatin by RNA binding
744 proteins regulates gene and repeat repression

745 **a.** Strategy for identifying chromatin-bound ERH interaction partners by isolation of soluble
746 nucleoplasm and chromatin-associated protein fractions prior to immunoprecipitation.
747 **b.** Immunoprecipitation (IP) of ERH from the chromatin fraction of human fibroblasts.
748 Western blot analysis of YTHDC1, YTHDC2, SAFB1, and ERH in the input and IP
749 samples.
750 **c.** Immunoprecipitation (IP) of ERH from the chromatin fraction of human ESCs. Western
751 blot analysis of YTHDC1, YTHDC2, SAFB1, and ERH in the input and IP samples.
752 **d.** Immunofluorescence (IF) of ERH after RNase A and RNase H treatment.
753 **e.** Western blot of ERH, YTHDC1, and YTHDC2 in nucleoplasm and chromatin-associated
754 protein fractions of human fibroblasts after siYTHDC1 and siYTHDC2 treatment.
755 **f.** Expression change and significance, TEtranscripts[67] analysis, of HERVK HERVH and
756 satellite repeats in hESCs and human fibroblasts after siERH.
757 **g.** Immunofluorescence (IF) of BRACHYURY in siCTRL, siYTHDC1 and siYTHDC2 after 1
758 day endoderm differentiation.
759 **h.** Venn diagram of gene coregulation after siERH, siYTHDC1 and siSAFB1 in human
760 fibroblasts after addition of hepatic reprogramming factors, FOXA3, HNF1A, HNF4A.
761 **i.** Proposed model for cell-type specific, and cell-type agnostic gene and repeat regulation
762 by ERH, YTHDC1, and SAFB1.
763 See also Extended Data Fig. 7

Katznelson et al., Figure 6



764 **Methods**

765 **Cell lines culture and condition**

766 Human H1 ES cells were obtained from WiCell (WA01), grown on Matrigel coated wells and
767 cultured in mTeSR1 (Stem Cell Tech, 85850) medium at 37°C and 5% CO_2. Medium was
768 changed daily. Human BJ foreskin fibroblasts were obtained from ATCC (CRL-2522) and
769 cultured in Eagle's Minimum Essential Medium (EMEM, Sigma-Aldrich M2279) supplemented
770 with 10% fetal bovine serum (FBS, Hyclone SH30071) and 2mM Glutamax (ThermoFisher
771 35050061) at 37°C and 5% CO_2.

772

773 **Mouse embryo work**

774 Mouse embryo experimentation was approved by the Biological Resource Center Institutional
775 Animal Care and Use Committee (IACUC), Agency for Science, Technology and Research
776 (IACUC Protocol 806983). C57BL/6 wild-type 3–4-week-old female mice were superovulated
777 using 5 IU of pregnant mare serum (PMS, National Hormone and Peptide Program)
778 gonadotropin given intraperitoneally and 5 IU of recombinant chorionic gonadotrophin (CG,
779 National Hormone and Peptide Program) given 48 h after and immediately before mating,
780 according to animal ethics guidelines of the University of Pennsylvania. Embryos were flushed
781 from oviducts of plugged females using M2 medium (Merck) and cultured in KSOM+AA (Merck)
782 covered by mineral oil (Sigma, M-5310), at 37 °C and 5% CO2.

783

784 **Mouse embryo immunofluorescence and microscopy**

785 Embryos were fixed in 4% paraformaldehyde (Electron Microscopy Sciences, 15710) for 30 min
786 at room temperature washed twice in PBS with 0.1% Triton X-100 (Millipore Sigma, T8787),
787 permeabilized for 20 min in PBS with 0.5% Triton X-100, and blocked in 3% bovine serum
788 albumin (Gemini Bio, 700-100P) for 30 min. Embryos were then incubated at 4 °C overnight in
789 primary antibodies diluted in blocking solution at the following concentrations: ERH (Abcam,

790 ab166620) at 1:200, Keratin-8 (DSHB, THROMA-I) at 1:20, H3K9me3 (Abcam, ab8898) at
791 1:1000, and CDX2 (Invitrogen, ZC007) at 1:200. After primary antibody incubation, embryos
792 were washed 5 times for 20 min in PBS with 0.1% Triton X-100 and incubated for 30 min in
793 Alexa Fluor-conjugated secondary antibodies (Invitrogen, A32731, A32723, A32733, A32728
794 and A11006) diluted in blocking solution to 1:500. Phalloidin-Alexa Fluor 555 (Invitrogen,
795 A34055) diluted to 1:500 and DAPI (Sigma, 10236276001) diluted to 1:500 in blocking solution
796 were also used to label F-actin and DNA, respectively. Embryos were washed three times in
797 PBS with 0.1% Triton X-100 before mounting in PBS covered with mineral oil in an 8-well μ-
798 Slide (Ibidi, #80827). Embryos were imaged using a laser-scanning confocal microscope (Leica
799 SP8) with a water-immersion Apochromat 40× 1.1 NA objective and HyD detectors. Samples
800 were excited with standard laser lines (405-nm, 488-nm, 552-nm, 638-nm).
801
802 **Mouse embryo siRNA injection experiments**
803 Microinjections were performed using a FemtoJet (Eppendorf). For ERH knockdown
804 experiments, mouse embryos were injected with two siRNAs (Qiagen, S104403294 &
805 S104403287) against ERH at a total concentration of 400 nM. For knockdown experiments
806 AllStars negative control siRNA (Qiagen, 1027281) was injected in embryos at a concentration
807 of 400nM.
808
809 **Differentiation of hESCs**
810 Prior to differentiation, for samples to be analyzed by RNA-seq, cells grown to a density of 80%
811 were digested with Accutase (Stem Cell Technologies, 07920) for 7 min. Then, DMEM/F12 was
812 added and gently pipetted to generate a cell suspension. After centrifugation, cells were
813 resuspended in mTeSR1 medium with 5uM ROCK inhibitor Y-27632 (Cayman Chemical,
814 10005583) and seeded at $8X10^5$-$1X10^6$ cells per well in a 6 well plate coated with Matrigel
815 (Millipore Sigma, CLS354277) 24 hours prior to differentiation treatment. For definitive

816 endoderm differentiation of hESCs, mTeSR1 medium was removed and cells were cultured in
817 DMEM/F12 (Thermo Fisher, 11320033) supplemented with 10% (v/v) Fetal Bovine Serum
818 (Hyclone, SH30071) with 100ng/ml Activin A (Peprotech, 120-14P) and 3uM CHIR99021
819 (Cayman Chemical, 13122). For ectoderm differentiation of hESCs, mTeSR medium was
820 removed and cells were cultured in DMEM/F12 with 5uM Retinoic Acid (Cayman Chemical,
821 11017).
822
823 **Cell culture siRNA transfection experiments**
824 All ERH knockdown experiments were performed with two cycles of siRNA transfection
825 (Silencer Select siRNA, Thermo Fisher, s4815), three days apart. Transfections were performed
826 in fibroblasts using 5nM final concentration of siRNA and in hESCs using a 10nM final
827 concentration of siRNA and 2.5ul/ml final concentration of Lipofectamine RNAiMAX (Thermo
828 Fisher #31985070).
829
830 **Reprogramming factor expression experiments**
831 We utilized the CytoTune-iPS 2.0 Sendai reprogramming kit (ThermoFisher, A16517) to express
832 Oct4, SOX2, KLF4 and cMyc in BJ fibroblasts. Cells were treated according to the
833 manufacturer's instructions three days after the second of two cycles of siRNA transfection.
834
835 **Cell culture immunofluorescence and microscopy**
836 For immunofluorescence of human fibroblasts and hESCs cells were washed briefly with PBS,
837 and fixed in 4% paraformaldehyde in PBS for 10 minutes. Fixed cells were washed 3 times with
838 PBS, permeabilized with ice-cold 0.1% Triton-X in PBS for 10 minutes, and washed twice with
839 PBS-T (PBS, 0.05% Tween-20). Samples were blocked with 3% donkey serum (Millipore
840 Sigma, D9663) in PBS for 2 hours at room temperature and incubated overnight at 4°C using
841 the following concentrations: anti-H3K9me3 (1:300, Abcam ab8898), anti-H3K27me3 (1:300,

842 Active Motif, 39155), anti-SOX2 (1:200, Cell Signaling, 23064), anti-OCT4 (1:200, Abcam,
843 ab19857), anti-PAX6 (1:200, BD Biosciences, 561462), anti-Brachyury (1:200, Cell Signaling,
844 81694) in 3% donkey serum in PBS. Cells were washed 3 times with PBS-T and then incubated
845 with AlexaFluor 488- or 594-conjugated secondary antibodies raised in donkey (Thermo Fisher,
846 See Methods) at a 1:500 dilution in 3% donkey serum in PBS.
847

848 **Total RNA isolation and sequencing**

849 BJ fibroblasts grown in a 6-well dish were washed twice with cold PBS, followed by addition of 1
850 mL TRIzol reagent (ThermoFisher 15596026) and transferred to a chilled 1.5 mL tube using a
851 cell scraper. The cells were homogenized by pipetting through a 25G needle attached to a 2 mL
852 syringe 10 times. 200 µl of chloroform was added to the cells and vortexed for 15 seconds,
853 followed by a room temperature incubation of 3 minutes. Phase separation of nucleic acids and
854 protein was performed by centrifugation at 12,000 x g for 15 minutes at 4C. The upper aqueous
855 phase containing RNA was then transferred to a fresh chilled tube. RNA was precipitated from
856 the aqueous phase by adding an equal volume of isopropyl alcohol, followed by a 10 minute
857 incubation at room temperature. The RNA was pelleted by centrifugation at 12,000 x g for 10
858 minutes at 4C. After removal of supernatant, the pellet was washed twice with 1 mL 75%
859 ethanol, and resuspended in nuclease free water. RNA quality was verified using a high-
860 sensitivity RNA screentape (Agilent 5067-5576). 1ug of total RNA was sent to Novogene, who
861 assembled and sequenced bulk RNA-seq libraries. Two to three biological replicates were
862 prepared for each condition.
863

864 **CUT&RUN experiments**

865 CUT&RUN was performed as described in[69], with digitonin and NaCl in wash buffers optimized
866 for permeabilization of BJ fibroblasts and retainment of SOX2 on chromatin as performed in[70].
867 Next, BJ fibroblasts were detached from cell culture plates with Accutase (STEMCELL

868 Technologies 07920), washed, bound to magnetic concanavalin A beads, and permeabilized with
869 a dig-wash buffer (0.1% digitonin, 20 mM HEPES pH 7.5, 75 mM NaCl, 0.5 mM Spermidine,
870 EDTA-free protease inhibitor). Bead-bound cells were incubated with a SOX2 (CST, clone
871 D9B8N, 23064, 1:100) or H3K9me3 antibody (abcam ab8898, 1:100) at 4C overnight. The
872 following morning, the cells were washed twice with dig-wash buffer, incubated with pA/G-MNase
873 for an hour at 4C, and washed twice more. After chilling cells on an ice block for 5 minutes, MNase
874 was activated for 30 minutes by addition of 2mM $CaCl_2$. The reaction was stopped by adding an
875 equal volume of 2X STOP buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.1% Digitonin, 0.05
876 mg/mL glycogen, 5 mg/mL RNase A). Digested chromatin was extracted from permeabilized cells
877 at 37C for 30 minutes, and DNA was purified by phenol-chloroform. Libraries were prepared as
878 previously described[71]. Two to three biological replicates were prepared for each condition.
879

880 **CUT&RUN data processing**

881 Paired-end CUT&RUN reads were mapped to the human genome build hg38 using bowtie2[72]
882 (v2.3.4.1) with run parameters --local --very-sensitive --no-unal --no-mixed --no-discordant -I 10 -
883 X 1000. A quality-filtered BAM file was generated with the command samtools (v1.16.1) view -q
884 5 -f 2 -bS. Global scaling factors (SF) for siERH and siControl H3K9me3 data sets were calculated
885 using the default parameters of ChIPSeqSpikeInFree[73] (R v3.6.3). Bigwig files for visualization
886 were generated from indexed BAM files with command bamCoverage --smoothLength 10 --
887 normalizeUsing CPM --scaleFactor 1/SF --ignoreForNormalization chrM (deeptools 3.5.0). BED
888 files were generated from BAMs using the bedtools command bamToBed (bedtools v2.27.1). BED
889 files were converted to bedgraph format using the bedtools command genomecov. Peaks were
890 called on the bedgraph files with SEACR[74] (v1.3), using the stringent setting and selecting the top
891 0.01% of regions by AUC. Peak lists for downstream analyses were derived from individual
892 replicates or the union of overlapping sites from biological triplicates, as indicated in the text and
893 figures. Enrichment analysis of known DNA binding motifs was performed using HOMER[75].

894

895 **Heatmaps**

896 Heatmaps were generated with deeptools[76] version 3.5.0. Counts matrices were generated with
897 command computeMatrix reference-point --missingDataAsZero --referencePoint center. Figures
898 were produced with the plotHeatmap command.

899

900 **Correlation plots**

901 To assess genome wide similarity between hBJ siERH H3K9me3, hBJ siControl H3K9me3, and
902 H1-hESC H3K9me3 data sets, the Pearson correlation of peak overlap was calculated using the
903 Intervene tool[77]. In brief, peak files for individual H3K9me3 replicates were processed with the
904 command intervene pairwise --I *H3K9me3.bed --compute fract --htype color. Visualization of
905 Pearson correlation of fractional overlap between peak sets were generated on
906 https://intervene.shinyapps.io/intervene.

907

908 **RNA-seq data analysis**

909 RNA-seq reads (paired-end) were mapped to the mm10 or hg38 genome for mouse and human
910 experiments respectively using STAR[78] (2.7.10b). The mapped reads to genes were counted
911 using featureCounts[79]. For counting repeat mapping reads the TEcount function from
912 TEtranscripts[67] was run on STAR aligned bam files using the hg38 repeatmasker gtf file
913 annotation. The output count files were analyzed by DESeq2[80] to detect differentially expressed
914 genes (Benjamini adjusted p-value < 0.05, log2 fold change > 1 or log2 fold change > 1.5).
915 Genes and repeats with fewer than 10 total mapped reads were excluded from DESeq2 and
916 TETranscripts analysis. TPM, FPKM and RPM were calculated using an in-house R script.

917

918 **Mouse blastocyst image analysis**

919 3D visualizations and analysis of embryos were performed using Imaris 8.2 or Imaris 9.7
920 software (Bitplane). The manual surface rendering module was used for cellular and nuclear
921 segmentation, using phalloidin and DAPI signals to define cellular and nuclear boundaries,
922 respectively. Quantifications of immunofluorescence intensity were performed by measuring the
923 mean fluorescence intensity of voxels within the segmented object, standardized DAPI to
924 account for light scattering.
925
926 **Cell culture image analysis**
927 Quantification of nuclear area and signal intensity was performed using the Nuclear
928 Morphometric Analysis (NMA) plugin for ImageJ[81]. In ImageJ saturated pixels was set to 0.35%
929 prior to processing. NMA plugin was run using the following settings: show ellipse and use
930 watershed options selected, minimum nuclei area=150, maximum nuclei area=6000, nuclei
931 intensity segmentation threshold=95, nuclei insertion intensity=93%.
932
933 **Gene ontology enrichment analysis**
934 Gene ontology enrichment analysis was performed using PANTHER[82] (v18.0) with settings,
935 statistical overrepresentation test: GO biological process complete, all genes in the database
936 were used as the background reference list, and statistical significance was determined using
937 Fisher's Exact test with false discovery rate.
938
939 **Statistical analysis**
940 Statistical analyses were performed in GraphPad Prism, the R software for statistical computing
941 and with DESeq2. Pairwise comparisons were analyzed using an unpaired, two-tailed Student's
942 t-test. Correlation was analyzed with a two-tailed Pearson's Correlation test. Significant overlap
943 between gene sets was analyzed with the hypergeometric test. Differential gene expression
944 analysis was performed using DESeq2 which calculates p-values by the Wald test and adjusted

945 p-values corrected for multiple testing using the Banjamini and Hochberg method. Reproducibility
946 was confirmed by independent experiments.
947
948 **Acknowledgements**
949 R.L.M. was supported by an NIH NIDDK K01 postdoctoral fellowship DK117970-01. N.P. was
950 supported by NIGMS (GM139970-01) and NICHD (HD102013-01A1). NIHGM36477 and
951 R35GM153180 supported work in the lab of K.S.Z.
952
953 **Author contributions**

954 Conceptualization, R.L.M. and K.Z.; Formal analysis, A.K., B.H., H.F., K.T. and R.L.M.; Funding
955 acquisition, K.Z; Investigation, A.K., B.H., H.F., K.T., J.Z., A.B. and R.L.M.; Methodology, A.K.,
956 B.H., R.L.M. and K.Z.; Project administration, R.L.M and K.Z.; Resources, K.Z,; Supervision, M.T-
957 P., N.P., R.L.M. and K.Z.; Writing – Original Draft, A.K., R.L.M. and K.Z.
958
959 **Declaration of Interests**
960 The authors declare no competing interests.
961
962 **Corresponding author**
963 Correspondence to Ryan McCarthy (ryan.mccarthy@seatttlechildrens.org) and Kenneth Zaret
964 (zaret@pennmedicine.upenn.edu).
965
966 **Data availability**
967 • RNA-seq and CUT&RUN datasets collected in this study have been deposited in the
968 GEO of NCBI and are publicly available as of the date of publication. Accession
969 numbers are listed in the Methods. GEO reviewer tokens for accessing the data

970 while it remains in private status are for RNA-seq GSE268901 (wbsvkukqhzobpan),
971 and for CUT&RUN GSE268961 (qngpsiywdvyfjyj).
972
973 **Code availability**
974 - This paper does not report original code.
975 - Any additional information required to reanalyze the data reported in this paper is
976 available from the lead contact upon request.
977
978
979 **Supplementary Tables**
980 Supplementary Table 1. H3K9me3 peak coordinates in siERH human fibroblasts.
981 Supplementary Table 2. Fibroblast +OSKM 48 hours, SOX2 CUT&RUN peaks in siControl and
982 siERH.
983 Supplementary Table 3. Fibroblast +OSKM 48 hours, siERH/siControl RNA-seq DESeq2
984 differential gene expression analysis results.
985 Supplementary Table 4. Fibroblast +OSKM 48 hours, siERH/siControl TETranscripts differential
986 repeat expression analysis results.
987 Supplementary Table 5. Blastocyst immunofluorescence quantification values.
988 Supplementary Table 6. Blastocyst RNA-seq DESeq2 siERH/siControl differential gene
989 expression analysis results.
990 Supplementary Table 7. H1 hESC untreated, endoderm and ectoderm RNA-seq DESeq2
991 siERH/siControl differential gene expression analysis results.
992 Supplementary Table 8. H1 hESC untreated, endoderm, and ectoderm TETranscripts
993 siERH/siControl differential repeat expression analysis
994 Supplementary Table 9. H1 hESC untreated, endoderm and ectoderm GO analysis results.
995

996 **Extended data**

997 Extended Data Fig. 1. ERH depletion reverts fibroblast H3K9me3 patterns to an hESC-like state

998 **a.** Venn diagram of H3K9me3 peaks called using SEACR from fibroblast siControl and
999 siERH H3K9me3 CUT&RUN and H1 hESC ChIP-seq (GSM1003585).

1000 **b.** Browser track showing SEACR peaks called in siControl H3K9me3 domain.

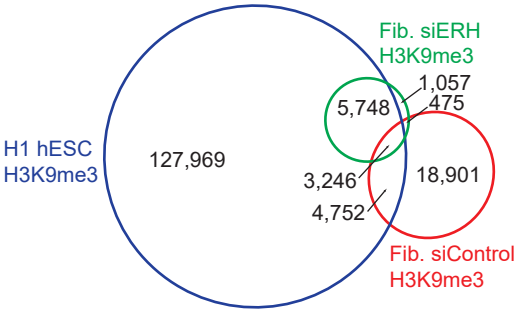
1001 **c.** Browser tracks showing H3K9me3 (GSM1003585) and HP1a (GSM1701825) in H1
1002 hESCs and H3K9me3 in siControl and siERH human fibroblasts.

1003 **d.** Top 10 statistically enriched transcription factor binding motifs at H3K9me3 peaks in
1004 siERH treated human fibroblasts (homer findMotifs).

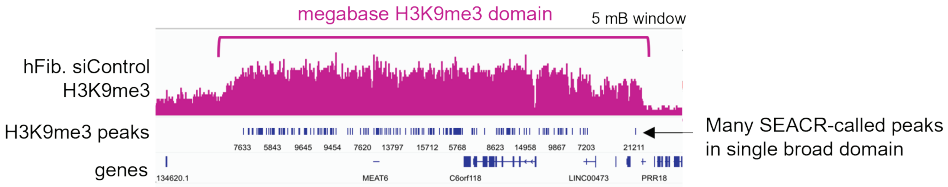
1005 **e.** Occurrence of repeats from RepEnrich database and repeat occurrence at sites of siERH
1006 fibroblasts H3K9me3 peaks.

Katznelson et al., Extended Data Figure 1

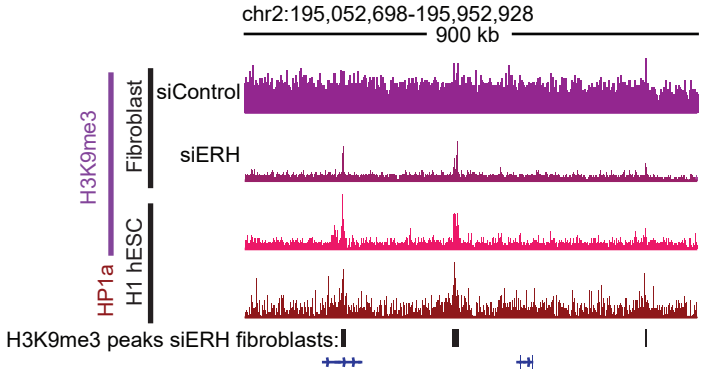
a Venn diagram of overlap of H3K9me3 peaks called by SEACR in H1 hESCs, and siControl and siERH treated fibroblasts



b Browser track showing limitations of peak calling to capture domain architecture of H3K9me3



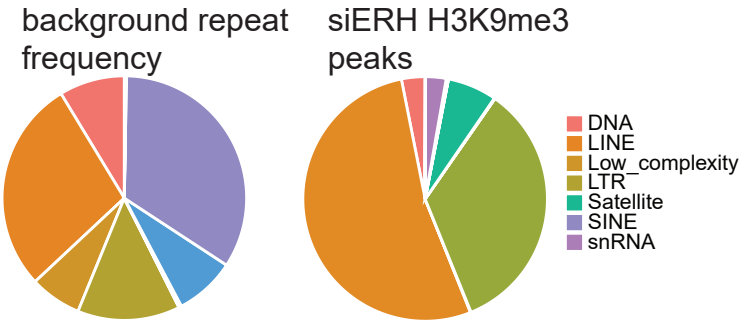
c Browser track of H3K9me3 in siControl and siERH compared to H3K9me3 and Hp1a in H1 hESC



d Top 10 statistically enriched transcription factor motifs at siERH fibroblast H3K9me3 peaks

	p-adj	Motif
ZNF136	$<10^{-5}$	
SOX10	$<10^{-5}$	
ZFX	$<10^{-5}$	
SOX21	$<10^{-5}$	
SOX3	$<10^{-5}$	
SOX4	$<10^{-5}$	
ZNF711	$<10^{-5}$	
ZNF317	$<10^{-5}$	
SOX15	$<10^{-5}$	
SOX9	$<10^{-5}$	

e Proportion of repeat class at siERH fibroblast H3K9me3 peaks vs background abundance



1007 Extended Data Fig. 2. ERH depletion enhances gene activation of pluripotency

1008 **a.** Expression of genes activated during fibroblasts to iPSC reprogramming timepoints[21] that
1009 are located in unmarked, H3K27me3 or H3K9me3 marked heterochromatin as annotated
1010 by[5].

1011 **b.** Venn diagram showing overlap between SOX2 peak calls in siControl and siERH
1012 fibroblasts+OSKM.

1013 **c.** Browser track of example hESC SOX2 and NANOG peak coincident with fibroblast siERH
1014 H3K9me3 loss and SOX2 peak gain.

1015 **d.** SOX2 CUT&RUN average profile in siControl and siERH fibroblasts+OSKM for genome
1016 regions identified as being differentially bound by reprogramming transcription factors [15].

1017 **e.** Rank ordered GIGGLE enrichment scores. Each circle indicates similarity between siERH
1018 SOX2-gained sites and a ChIP-seq peak set from cistromeDB. Blue circles are peak sets
1019 profiled in hESCs, white circles are peaks from a non-embryonic origin.

1020 **f.** Browser track of H3K9me3 in siControl and siERH fibroblasts and RNA in siControl and
1021 siERH fibroblasts+OSKM at DPPA4.

1022 **g.** Example 8 cell genes activated in siERH+OSKM but not activated during eight time points
1023 of iPSC reprogramming time course[21].

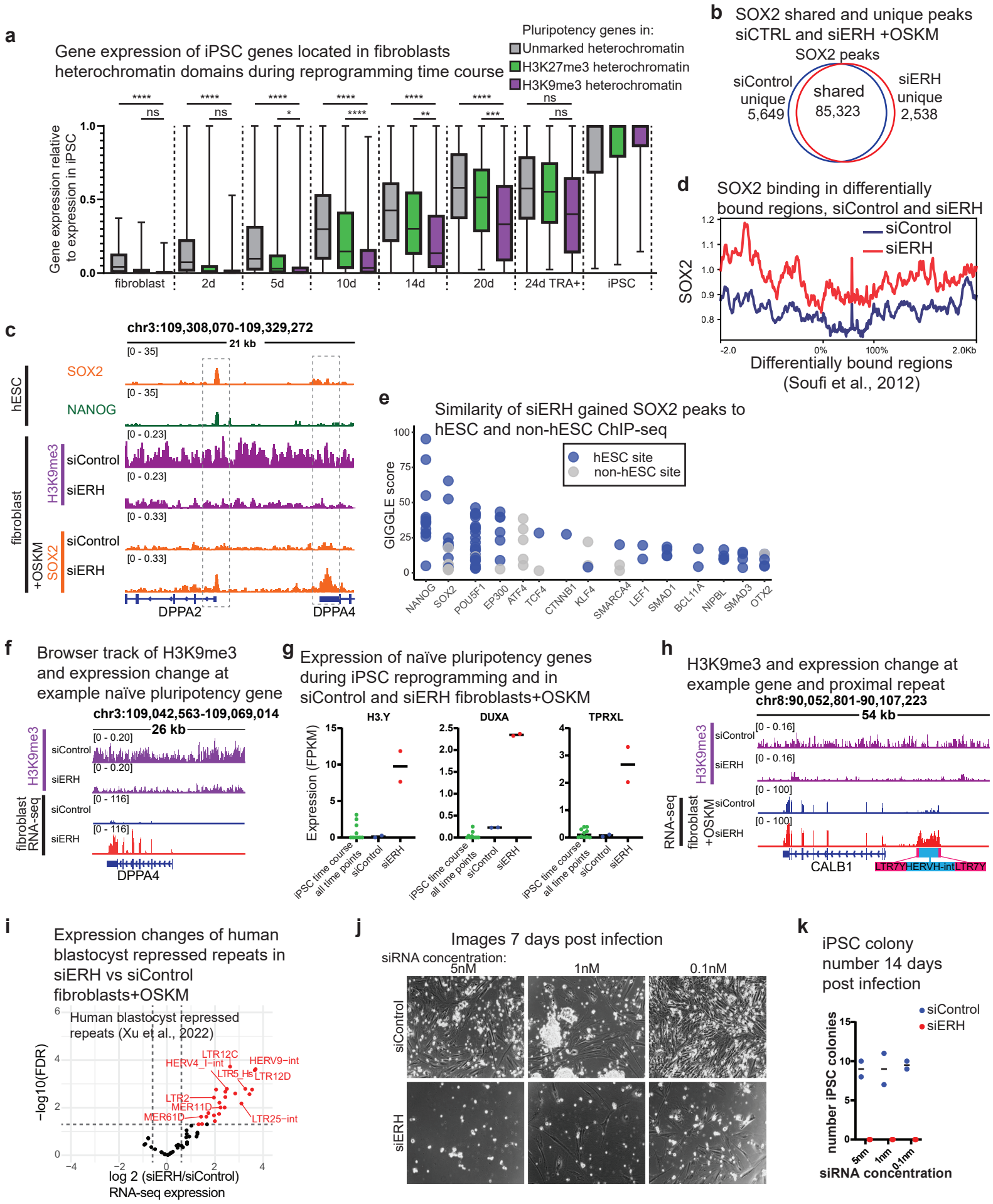
1024 **h.** Browser tracks of H3K9me3 in siControl and siERH fibroblasts and RNA in siControl and
1025 siERH fibroblasts+OSKM, at a locus of an LTR7Y-HERHV transposable element and
1026 nearby gene.

1027 **i.** Volcano plot showing differential repeat expression in siERH vs siControl
1028 fibroblast+OSKM for repeats annotated in[6] as repressed by H3K9me3 in human
1029 blastocysts. (TETranscripts, DESeq2, padj<=0.05, log2fc>0.5).

1030 **j.** Representative phase image showing cell confluency of siControl and siERH
1031 fibroblasts+OSKM cells 2 weeks after initial OSKM infection with titrated siRNA
1032 concentration. All treatments were plated at identical cell numbers 1 day prior to infection.

1033 **k.** Plot of iPSC colony count from siControl and siERH fibroblasts+OSKM cells 3 weeks after
1034 initial OSKM infection with titrated siRNA concentration. Each value represents the
1035 number of colonies generated in 1 well of a 24 well plate.

Katznelson et al., Extended Data Figure 2



1036 Extended Data Fig. 3. Characterization of ERH and the effects of ERH depletion in mouse
1037 blastocysts

1038 **a.** ERH RNA expression in reads per kilobase million (RPKM) during stages of mouse pre-
1039 implantation development, RNA-seq data from[25]. unpaired two tail t-test; n=6.

1040 **b.** Fluorescent imaging of ERH (magenta) and quantification in ICM and trophectoderm cells
1041 (unpaired t-test, ns=not significant).

1042 **c.** ERH (magenta) and DAPI (teal) staining in mouse blastocysts injected at the 1 cell stage
1043 with siControl and siERH siRNAs showing protein depletion. Scale bars, 10μm.

1044 **d.** Representative DIC images of mouse blastocysts injected at the 1 cell stage with siControl
1045 and siERH siRNAs, two examples for each shown.

1046 **e.** Fluorescent imaging of trophectoderm marker CDX2 (green), pluripotency marker OCT4
1047 (magenta), with Phalloidin (orange) to define cell boundary and quantification of OCT4
1048 levels in siControl siERH blastocysts with DAPI (blue). (t-test, ns=not significant).

1049 **f.** Expression of Erh in siControl and siERH treated mouse blastocysts by RNA-seq. (t-test,
1050 $p>0.0001$).

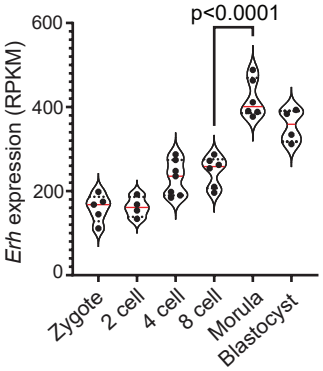
1051 **g.** TPM expression values from RNA-seq for primitive endoderm genes. (t-test)

1052 **h.** TPM expression values from RNA-seq for epiblast genes. (t-test, ns=not significant)

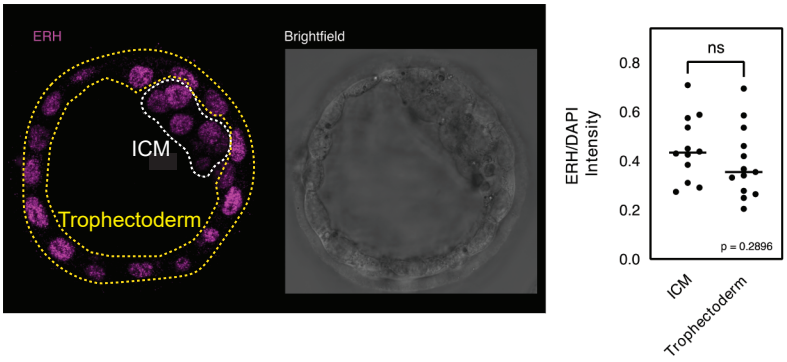
1053

Katznelson et al., Extended Data Figure 3

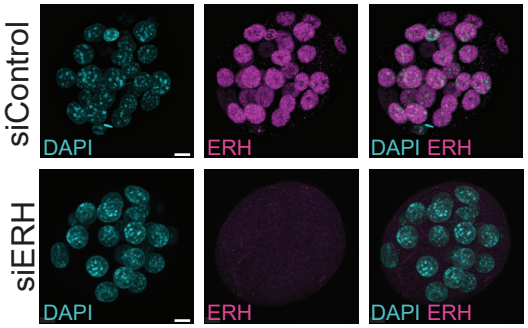
a *Erh* expression during mouse preimplantation



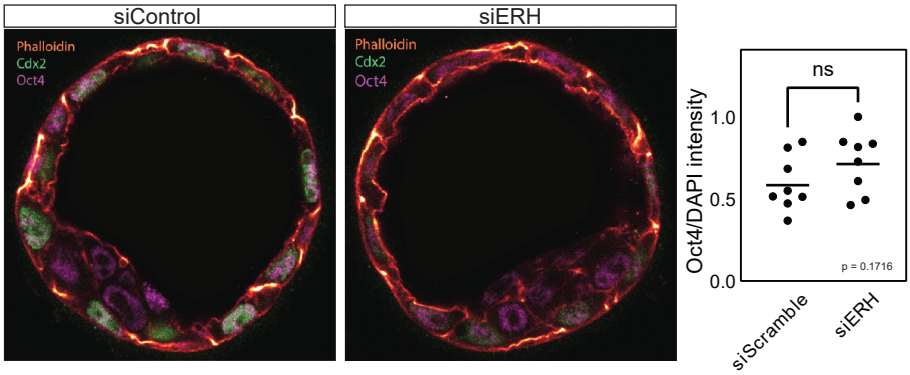
b ERH immunofluorescence in mouse blastocyst comparing protein levels in trophectoderm and ICM cells



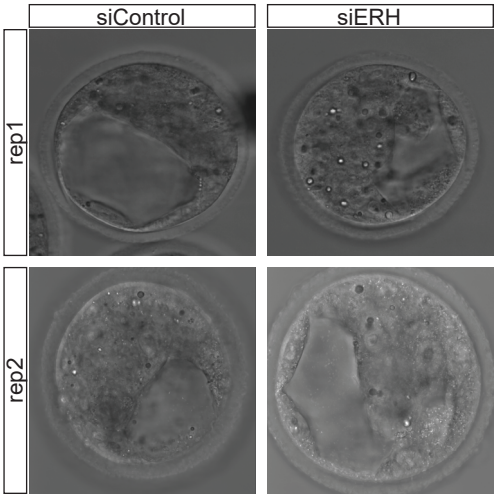
c ERH protein staining in mouse blastocysts injected with siRNA at the 1-cell stage



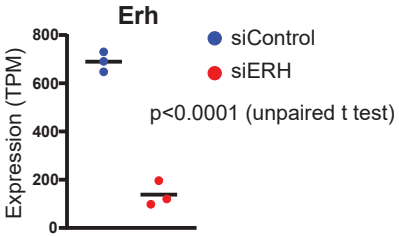
e Comparison of OCT4 protein levels between siControl and siERH mouse blastocysts



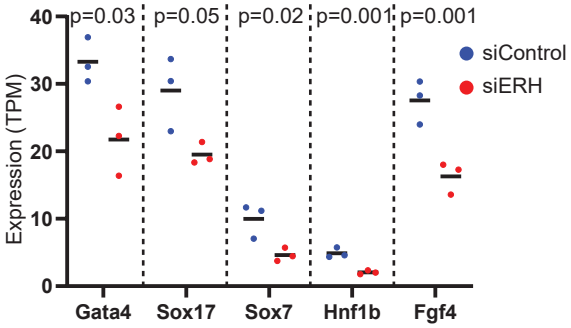
d Example brightfield images of mouse blastocysts injected at the 1 cell stage with siControl or siERH siRNAs



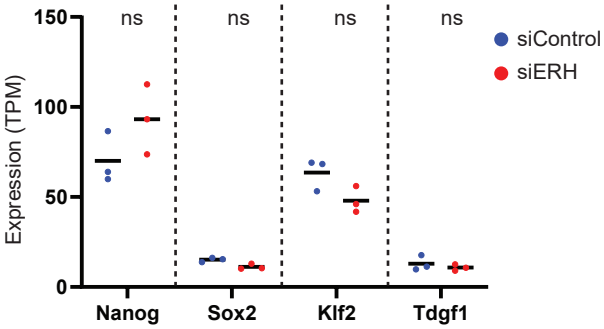
f *Erh* expression in siControl and siERH mouse blastocysts



g Expression of primitive endoderm genes in siControl and siERH mouse blastocysts



h Expression of epiblast genes in siControl and siERH mouse blastocysts



1054 Extended Data Fig. 4. ERH is dispensable for pluripotency maintenance but is involved in
1055 silencing of pluripotency and naïve genes during differentiation

1056 **a.** Immunofluorescence for DAPI (blue) and H3K27me3 (red) in siControl and siERH H1
1057 hESCs in pluripotent conditions. Scale bars, 50µm.

1058 **b.** Nuclear area, white arrows indicate nuclear blebs. $p=5.6*10^{-16}$, welch's two sample t-test.
1059 Scale bars, 50µm.

1060 **c.** Volcano plot showing differential gene expression in siERH vs siControl H1 in pluripotency
1061 conditions. Inset pie chart indicates the number of genes upregulated and down regulated
1062 in siERH vs siControl H1 (DESeq2, padj<=0.05, log2fc>0.5).

1063 **d.** Imaging of endoderm differentiation showing changes in colony morphology (phase),
1064 DAPI (blue), silencing of SOX2 (green) and induction of endoderm marker SOX17 (red).
1065 Scale bars, 100µm.

1066 **e.** Quantification of nuclear NANOG immunofluorescence intensity in siControl vs siERH 1
1067 day endoderm differentiated H1 hESCs. (unpaired t-test).

1068 **f.** ERH transcripts per million (TPM) values from RNA-seq in siControl and siERH H1 hESCs
1069 in untreated, endoderm (DE) and ectoderm (EC) differentiation conditions.

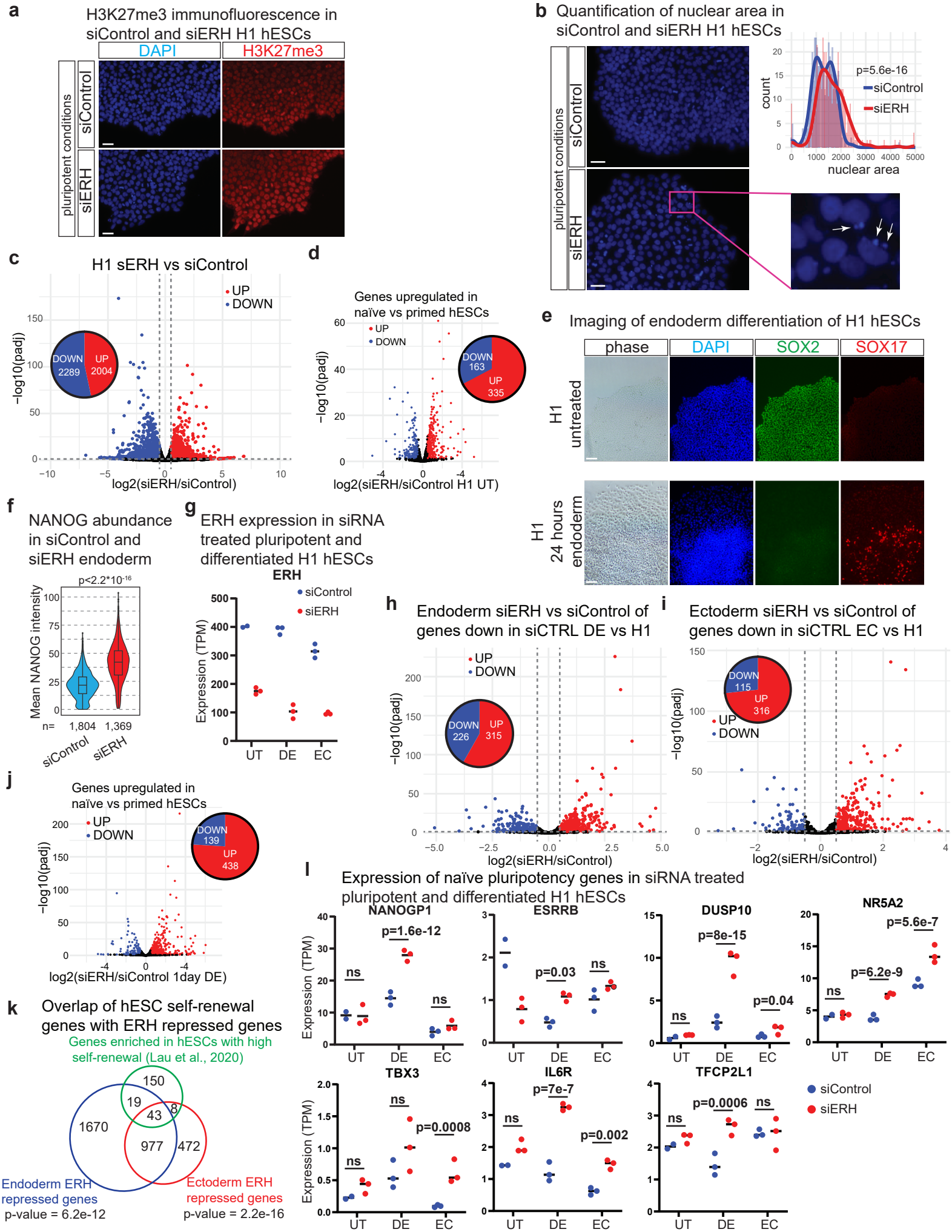
1070 **g.** Volcano plot showing differential gene expression in siERH vs siControl 1 day endoderm
1071 for genes down regulated in siControl DE vs siControl H1. Inset pie chart indicates the
1072 number of genes upregulated and down regulated in siERH vs siControl DE (DESeq2,
1073 padj<=0.05, log2fc>0.5).

1074 **h.** Volcano plot showing differential gene expression in siERH vs siControl 2 day ectoderm
1075 for genes down regulated in siControl EC vs siControl H1. Inset pie chart indicates the
1076 number of genes upregulated and down regulated in siERH vs siControl EC (DESeq2,
1077 padj<=0.05, log2fc>0.5).

1078 **i.** Intersection of genes enriched in hESCs with high self-renewal[37] with genes repressed by
1079 ERH during endoderm and ectoderm differentiation. (hypergeometric test)

1080 **j.** Graphs of naïve gene RNA-seq, transcripts per million (TPM) values from RNA-seq
1081 replicates shown (DESeq2 adjusted p-value).

Katznelson et al., Extended Data Figure 4



1082 Extended Data Fig. 5. ERH represses transposable elements that act as gene regulatory regions

1083 **a.** LTR7Y expression plot from RNA-seq data analyzed by TETranscripts in untreated (UT)
1084 and endoderm (DE) differentiated H1 hESCs treated with siControl or siERH. (DESeq2
1085 adjusted p-value).

1086 **b.** LTR5_Hs and HERVK expression plots in reads per million (RPM) from RNA-seq data
1087 analyzed by TETranscripts. (DESeq2 adjusted p-value).

1088 **c.** SVA class element expression plots in reads per million (RPM) from RNA-seq data
1089 analyzed by TETranscripts. (DESeq2 adjusted p-value).

1090 **d.** Browser track showing RNA-seq of example repeat element and nearby gene that is
1091 expressed in pluripotency, silenced in siControl DE and EC but expressed in siERH DE
1092 and EC conditions.

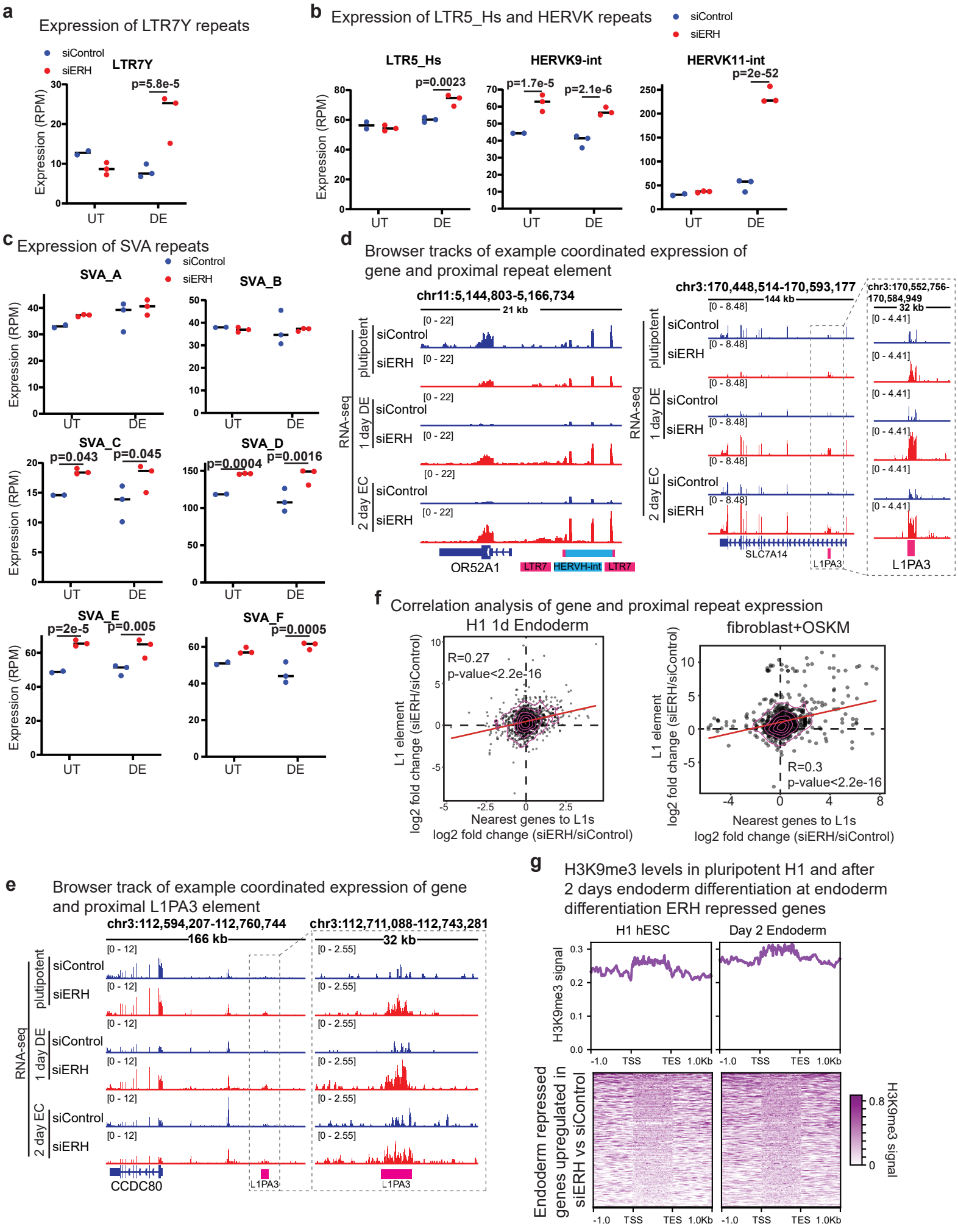
1093 **e.** Browser track showing RNA-seq of example repeat element and nearby gene that is
1094 expressed in pluripotency, silenced in siControl DE and EC but expressed in siERH DE
1095 condition.

1096 **f.** Plot of expression change of specific L1 element loci vs expression change of
1097 corresponding nearest gene within 25kb in 1 day endoderm. R, Pearson correlation.

1098 **g.** H3K9me3 data from H1 hESCs (GSM1003585) and hESC derived endoderm
1099 (GSM916057) at gene bodies +/- 1kb for upregulated (adjusted pvalue<=0.05;
1100 log2foldchange>=0.5) by ERH depletion during 1 day endoderm differentiation.

1101

Katznelson et al., Extended Data Figure 5



1102 Extended Data Fig. 6. ERH represses alternative lineage gene activation during hESC
1103 differentiation

1104 **a.** GO analysis (PantherDB, statistical overrepresentation) of genes upregulated
1105 (padj<=0.05, log2 fold change >= 0.5) in siERH vs siControl H1 hESCs 2 day ectoderm
1106 differentiation.

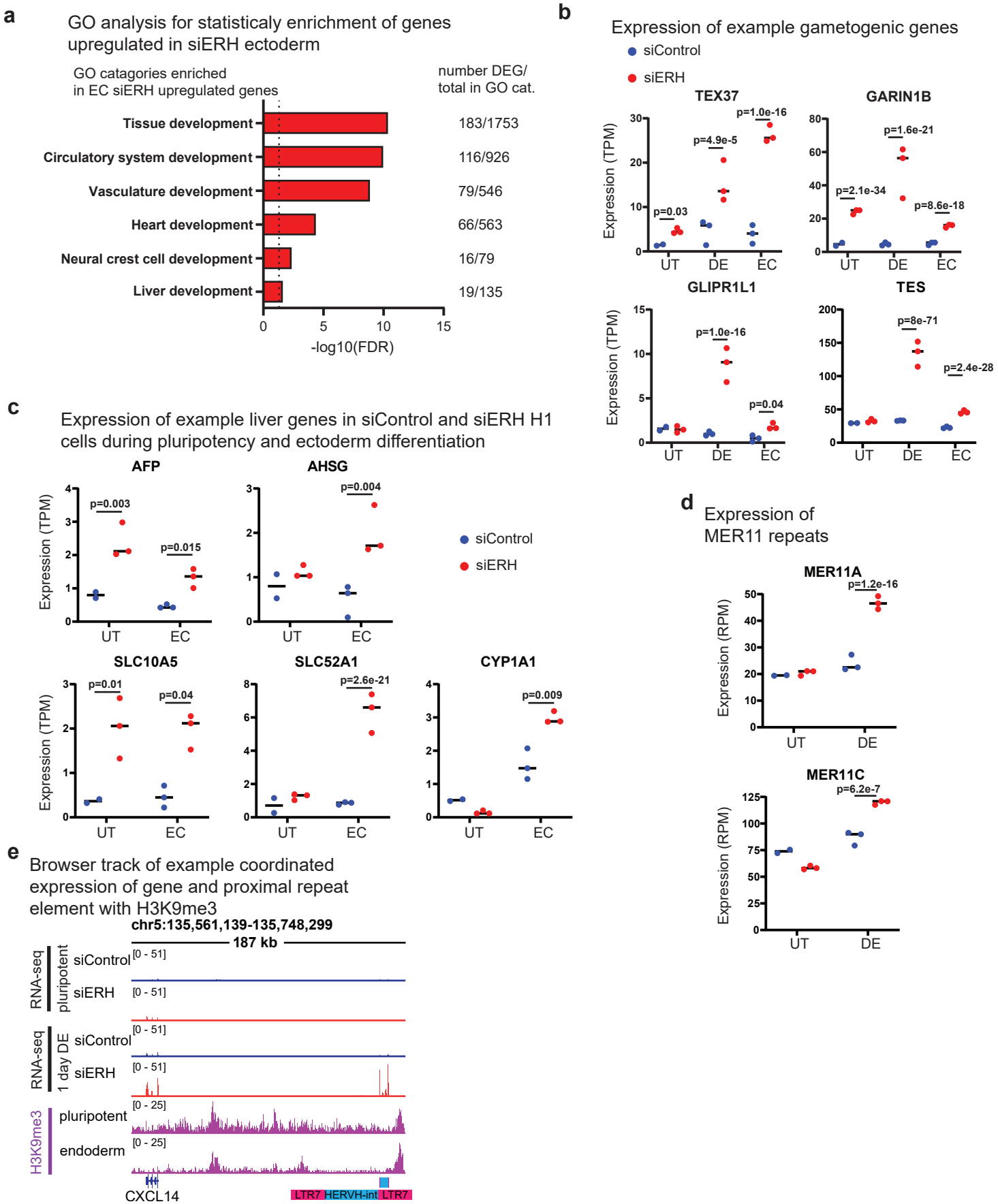
1107 **b.** Example meiotic and gametogenic genes activated during endoderm and ectoderm
1108 differentiation by ERH depletion, transcripts per million (TPM) values from RNA-seq
1109 replicates shown (DESeq2 adjusted p-value).

1110 **c.** Example alternative lineage genes activated during ectoderm differentiation by ERH
1111 depletion, transcripts per million (TPM) values from RNA-seq replicates shown (DESeq2
1112 adjusted p-value).

1113 **d.** MER11A and MER11C expression plots from RNA-seq data analyzed by TETranscripts.
1114 (DESeq2 adjusted p-value).

1115 **e.** Browser track showing RNA-seq of example repeat element and nearby gene in siControl
1116 and siERH treated H1 hESCs after 1 day of endoderm differentiation and H3K9me3 tracks
1117 from H1 hESCs (GSM1003585) and hESC derived endoderm (GSM916057).

Katznelson et al., Extended Data Figure 6



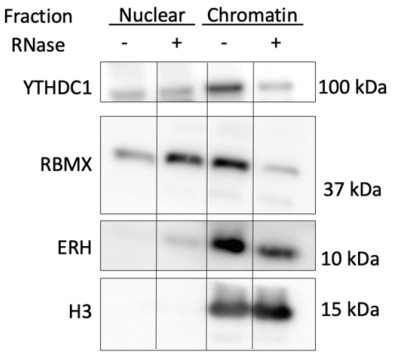
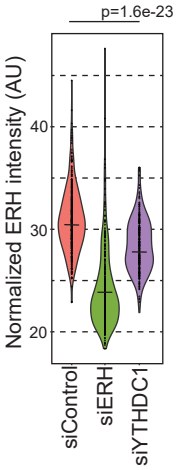
1118 Extended Data Fig. 7. ERH is recruited to the chromatin by RNA and RNA binding proteins.

1119 **a.** Western blot after RNase A treatment during soluble nuclear protein isolation. Western for
1120 ERH and YTHDC1 in nuclear and chromatin fractions with and without RNase. RBMX and
1121 H3 included as known RNA-sensitive and RNA-insensitive, chromatin-associated
1122 proteins.

1123 **b.** Quantification of immunofluorescence (IF) of nuclear ERH signal after siYTHDC1 followed
1124 by washout, in human fibroblasts.

Katznelson et al., Extended Data Figure 7

a RNase A treatment of soluble nuclear and chromatin fractions



b Quantification of ERH intensity by immunofluorescence after permeabilization and washout for indicated siRNA