HCFC1 and YY1 mediate recruitment of COMPASS and Integrator to initiate X chromosome inactivation.

Jeffrey Boeren[1*], Beatrice F. Tan[1*], Sarra Merzouk[1,2], Eveline Rentmeester[1], Rien van Haperen[1], Samuel J. Luchsinger-Morcelle[1], Cheryl Maduro[1], Richelle Rietdijk[1], Büşra Göynük[1], Wilfred F.J. van IJcken[3], Jeroen A.A. Demmers[4], Cristina Gontan[1,†], Hegias Mira Bontenbal[1,5,†,✉], Joost Gribnau[1,†,✉]

Affiliations

[1]Department of Developmental Biology, Erasmus MC, 3015CN, Rotterdam, The Netherlands,
[2]Current address: Broad Institute of MIT and Harvard, Cambridge, MA, United States,
[3]Center for Biomics, Erasmus MC, 3015CN, Rotterdam, The Netherlands,
[4]Center for Proteomics, Erasmus MC, PO Box 2040, 3000 CA, Rotterdam, The Netherlands,
[5]Current address: IGMM, University of Montpellier, CNRS, Montpellier, France.

*These authors contributed equally.
†These authors contributed equally.
✉Correspondence: hegias.mira-bontenbal@igmm.cnrs.fr, j.gribnau@erasmusmc.nl

Abstract

The evolution of mammalian sex chromosomes has driven the emergence of mechanisms that balance X-linked gene dosage between male (XY) and female (XX) cells. In females, dosage compensation is achieved through X chromosome inactivation (XCI), initiated by upregulation of the long non-coding RNA *Xist*, which spreads in *cis*, recruiting chromatin modifiers to silence gene expression on one X chromosome. Here, we conducted a forward genetic screen and identified X-encoded Host Cell Factor 1 (HCFC1), a member of the COMPASS H3K4 methyltransferase complex, as a dose-dependent XCI-activator. HCFC1 loss results in genome-wide reduction of H3K4me3 at specific regulatory elements and downregulation of nearby genes, including *Xist*. We show that HCFC1 and YY1 are co-recruited genome-wide to gene regulatory elements. Mass spectrometry analysis confirmed an interaction of HCFC1 and YY1 and uncovered the Integrator complex as another prominent YY1 partner. YY1 depletion results in genome-wide loss of Integrator recruitment at gene regulatory elements and reduced expression of nearby genes, including *Xist cis*-regulatory genes *Jpx* and *Ftx*. These results highlight a co-regulatory role for HCFC1 in COMPASS recruitment and *Xist* activation, alongside YY1-mediated recruitment of Integrator to *Xist* regulatory elements and genes to activate female-exclusive XCI.

Introduction

X chromosome inactivation (XCI) provides a paradigm for studying gene regulation in a chromosomal context. In mice and other mammals, random XCI is initiated in all female somatic cells shortly after implantation and can be modeled in female (XX) mouse embryonic stem cells (ESCs), which trigger XCI upon *in vitro* differentiation. This process is orchestrated by *Xist*, a long non-coding RNA (lncRNA) encoded within the X inactivation center (XIC), a genomic locus enriched for lncRNA genes that regulate *Xist* expression. Upon initiation of XCI, *Xist* coats the X chromosome in *cis* and recruits chromatin-modifying complexes to silence X-linked genes. Prior to and during early XCI, *Xist* transcription is repressed on the future active X by antisense transcription of the lncRNA *Tsix* (reviewed in [1,2]).

Female-exclusive XCI is regulated by X-encoded activators of the XCI process directing dose-dependent activation of *Xist*[3,4]. The X-encoded E3 ubiquitin-ligase RNF12 was identified as an XCI-activator through dose-dependent degradation of pluripotency factor REX1[5,6]. YY1 and its

homologue REX1 compete for binding at the same gene regulatory element of *Xist*, where REX1 recruitment prevents YY1-mediated activation of *Xist*[7]. Homozygous *Rnf12* knockout (KO) ESCs fail to initiate XCI during differentiation, whereas heterozygous loss of *Rnf12* leads to delayed XCI initiation with skewing towards inactivation of the X chromosome carrying the mutated *Rnf12* allele in *vitro*[8]. While the maternally inherited *Rnf12* allele is essential for imprinted XCI in extra-embryonic tissues in mice[9,10], a conditional *Rnf12* KO mouse model showed that *Rnf12* is dispensable for the initiation of random XCI in the post-implantation epiblast[11]. Additionally, compound homozygous KO ESCs and mice of *Rnf12* and its primary target *Zfp42* (encoding REX1) display a partial rescue of the XCI phenotype, restoring it to near wild-type levels[10].

These findings highlight the important role RNF12 plays in XCI regulation. Nonetheless, the fact that XCI is still initiated in *Rnf12*$^{+/-}$ and *Rnf12*$^{-/-}$:*Zfp42*$^{-/-}$ differentiating female ESCs, but not in male ESCs, points to the presence of additional XCI activators. The lncRNAs *Jpx* and *Ftx*, respectively located 10 kb and 90 kb upstream of *Xist*, have been implicated as XCI-activators, playing important roles in *Xist* regulation (reviewed in[1,2]) However, deletion studies indicate that most of this activity is mediated in *cis*[8,12,13]. Also, the X-linked demethylase KDM5C has been described as an XCI-activator by catalyzing the conversion of H3K4me2/3 to H3K4me1 at the YY1-REX1 regulatory element of *Xist*[14]. Although *Kdm5c* KO female embryos are non-viable, *Xist* expression is unaffected in *Kdm5c* heterozygous KO females supporting the hypothesis that multiple XCI activators are at play[14]. Finally, the GATA transcription factor family has been implicated in XCI regulation; however, the only X-linked family member, *Gata1,* is not expressed in ESCs or the epiblast, indicating that other XCI-activators remain to be identified[15].

Here, we used a forward genetic screen deleting large segments of the X chromosome to identify novel regulators of XCI and uncover the X-linked Host Cell Factor 1 (HCFC1) as a potent XCI activator. HCFC1, a component of the COMPASS H3K4 methyltransferase complex, promotes XCI by recruiting COMPASS to deposit H3K4me3 at the *Xist* regulatory element. Acute HCFC1 depletion during mouse ESC differentiation abolishes *Xist* and *Jpx* expression, preventing XCI. We found that genome-wide recruitment of HCFC1 is partly mediated by the transcription factor YY1, which we confirmed to physically interact with HCFC1. YY1 interactome analysis further identified the Integrator complex, a key mediator of transcriptional pause release, as a major YY1-associated partner. Loss of YY1 reduces Integrator binding at thousands of regulatory sites,

including *Xist cis*-regulatory elements, without affecting H3K4me3 deposition. Together, these findings reveal a dual role for HCFC1 and YY1 in XCI initiation: HCFC1-mediated recruitment of COMPASS to establish H3K4me3 and YY1-mediated recruitment of Integrator to promote transcriptional activation of *Xist*.

Results

A forward genetic screen to identify novel XCI-activators

To identify new XCI-activators that act in parallel with *Rnf12* in the activation of XCI, we performed a CRISPR/Cas9-based screen deleting large regions of the X chromosome (Fig. 1a). We made use of female heterozygous *Rnf12*$^{+/-}$ F1 hybrid *Mus musculus castaneus* (Cast): *Mus musculus musculus* 129/Sv (129) ESCs providing a powerful SNP-based readout to differentiate the two X chromosomes. These cells also display complete skewed XCI of the *Rnf12* mutant X_{129} chromosome to retain one active copy of *Rnf12* X_{cast} throughout SC differentiation[8]. Due to this phenotype, large heterozygous deletions of the X_{129} chromosome are tolerated during differentiation except for deletions that affect XCI (Fig. 1a). The deletions generated covered all regions of the early evolutionary strata, which are expected to retain key regulatory elements and genes involved in XCI (Fig. 1b)[16–18]. Proper deletion of regions from the X_{129} chromosome was confirmed by PCR of allele-specific repeat amplifications, whole-genome sequencing (WGS) detecting SNPs between the X_{129} and X_{Cast} chromosomes, and DNA-FISH (Extended Data Fig. 1a-e and Extended Data Fig. 2a). Differentiation followed by *Xist* quantitative PCR (qPCR) analyses of at least two biological replicates per deletion indicated significant loss of *Xist* expression at day 8 of differentiation in deletion E (ΔE), located 29.5Mb centromeric to *Xist* (Fig. 1b,c). ΔE's inability to upregulate *Xist* was subsequently confirmed by RNA fluorescent in situ hybridization (RNA-FISH) (Extended Data Fig. 2b-d). Deletion of regions harboring putative activators *Gata1* (located in ΔA and ΔAB) and *Kdm*5c (located in ΔK) did not result in a significant loss of *Xist* expression (Fig. 1c)[14,15]. To identify the gene(s) involved in XCI activation, we generated incremental and convergent deletions within ΔE (Fig. 1d) and determined the effect on *Xist* expression at day 4 and day 8 of ESC differentiation. ΔE2 and ΔE3 showed a reduction of *Xist* expression, unlike ΔE1, implying that our target gene(s) must be contained within ΔE2 and ΔE3

112 (Extended Data Fig. 2e). Next, we narrowed down the search by creating additional deletions
113 within regions E2 and E3, confirmed by WGS and RNA-seq (Fig. 1e and Extended Data Fig. 1e)
114 and discovered that ΔE2B and ΔE3B displayed a *Xist* phenotype at day 4 and day 8 of
115 differentiation, but not ΔE2A and ΔE3A (Fig. 1f). RNA-seq analysis revealed that loss of an 80kb
116 region encompassing *Hcfc1* and *Irak1* severely affects XCI at both analysed timepoints, whereas
117 downregulation of pluripotency factors and upregulation of factors associated with ESC
118 differentiation were broadly uncompromised compared to the parental cell line (Fig. 1e and
119 Extended Data Fig. 2f). This analysis confirmed downregulation of *Xist* in ESC lines harboring
120 ΔE2B and ΔE3B encompassing *Hcfc1* and *Irak1* (Fig. 1g,h) and as expected, impaired inactivation
121 of X-linked genes (Extended Data Fig. 2g). Analysis of previously published datasets show that,
122 similar to *Rnf12*, *Hcfc1* is upregulated between days 1 and 3 of ESC differentiation when XCI is
123 initiated (Fig. 1i and Extended Data Fig. 2g) [5,19], whereas *Irak1* is not upregulated. HCFC1 has
124 also been shown to interact with the crucial *Xist*-regulator YY1, pointing at *Hcfc1* as a candidate
125 XCI activator [7,20].

126

127 ## *Hcfc1* is a dose-dependent activator of XCI

128 To test whether *Hcfc1* possesses XCI-activator activity, we restored *Hcfc1* expression in ΔE2B
129 and ΔE3B ESC lines that showed an XCI phenotype by introducing a wild-type *Hcfc1* (129 origin)
130 or a control transgene (Fig. 2a, Extended Data Fig. 3a). RT-qPCR analysis showed that *Xist* levels
131 were rescued upon restoration of *Hcfc1* expression levels at day 4 of ESC differentiation in a
132 transfected pool of cells, and in individual clones with a stably integrated transgene, confirming a
133 dose-dependent regulatory role for *Hcfc1* in XCI (Fig. 2b,c, Extended Data Fig. 3b,c). To test
134 whether HCFC1 is required for activation of XCI, we studied the effect of dTAG-13-mediated
135 depletion of HCFC1 (Fig. 2d). Correct 3' integration of a 2xFLAG-V5-tag combined with
136 FKBP12^{F36V} in both *Hcfc1* (HCFC1-FKBP) alleles was confirmed by PCR (Extended Data Fig.
137 3d). HCFC1 is glycosylated and cleaved by OGT to form dimers that stay associated[21,22]. We
138 confirmed loss of expression of both the N- and C-terminal domains after 48 hours of dTAG13
139 treatment (Fig. 2e, Extended Data Fig. 3e and 4e). Loss of HCFC1 occurs very rapidly, as
140 examination of the C-terminal peptide showed a complete loss within 2 hours after addition of
141 dTAG-13 (Fig. 2e). Depletion of HCFC1 during different time intervals of ESC differentiation

142 resulted in a significant reduction of *Xist* expression (Fig. 2d,f). *Xist* RNA-FISH analysis
143 confirmed loss of *Xist* expression and a severely reduced percentage of *Xist*-coated X
144 chromosomes at day 3 of differentiation (Fig. 2g,h). Allele-specific RNA-seq analysis also showed
145 that positive *Xist cis*-regulator *Jpx* was significantly downregulated, and *Ftx* showing a
146 downregulation trend, whereas expression of pluripotency factors *Nanog*, *Pou5f1*, *Sox2* and *Rex1*
147 remained unaffected, indicating that ESC differentiation proceeded normally in the absence of
148 HCFC1 within the analysed time windows (Fig. 2i, Extended Data Fig. 3f). On the contrary,
149 expression of several *trans*-regulators of *Xist*, including *Rnf12, Yy1*, *Kdm5c* and *Rif1* was
150 significantly increased upon loss of HCFC1, suggesting a compensatory mechanism in *Xist*
151 regulation (Extended Data Fig. 3g). dTAG-mediated loss of HCFC1 did not affect negative cis-
152 regulators of *Xist*, such as *Tsx* and *Tsix* (Fig. 2i). Consistent with the observed substantial *Xist*
153 downregulation, allele-specific expression analysis of X-linked genes indicated loss of XCI in
154 HCFC1-depleted ESCs at day 3 of differentiation (Fig. 2j). To discern if this loss of silencing is
155 due to *Xist* loss or HCFC1 loss, we removed HCFC1 and forced *Xist* expression in undifferentiated
156 ESCs by means of a doxycyclin inducible promoter at the endogenous *Xist* locus. We observed
157 that XCI still happens properly since X-linked *Rnf12* was silenced while a gene escaping XCI,
158 *Kdm6a,* was not (Extended Data Fig. 3h-k). These findings identify HCFC1 as a dose-dependent
159 activator of XCI and indicate that *Hcfc1* is required for upregulation of *Xist* and positive *Xist*
160 regulatory genes.

161

162 ## Overlapping genome-wide enrichment of HCFC1, YY1 and Integrator

163 HCFC1 is a member of the Set1A/B and MLL1/2 COMPASS complexes, mediating H3K4me3
164 deposition at gene regulatory elements which promotes transcription activation through Integrator
165 complex recruitment and promoter-proximal eviction of paused RNA Pol II[23–26]. To
166 mechanistically understand the role of HCFC1 in XCI and gene regulation in general, we re-
167 analysed published ChIP-seq datasets examining the genome-wide distribution of HCFC1 and the
168 COMPASS core member DPY30, as well as YY1, H3K4me3 and Ser-5 phosphorylated Pol II in
169 undifferentiated mouse ESCs[25,27,28]. This analysis confirmed genome-wide co-enrichment of
170 DPY30 with HCFC1 and their catalytic product H3K4me3 (Fig. 3a). Of 31,247 HCFC1 peaks in
171 ESCs, 70.5% were found in genes, of which 48.7% were at promoters (Extended Data Fig. 4a).

172 Within promoters, 81.5% of YY1 peaks overlapped with HCFC1 (Fig. 3b,c), confirming the
173 previously described overlap in HCFC1 and YY1 binding to chromatin in human cells[20,29]. This
174 analysis also shows that many HCFC1 peaks are not co-bound by YY1, indicating that other
175 transcription factors might be recruiting HCFC1 to those sites.

176 To understand HCFC1-mediated regulation of *Xist*, we examined HCFC1's distribution by ChIP-
177 seq on day 3 of ESC differentiation when *Xist* expression is upregulated. For this, we generated
178 2xFLAG-V5-tagged HCFC1 (HCFC1-FLAGV5) female ESC lines through CRISPR/Cas9-
179 mediated integration (Extended Data Fig. 4b). ChIP-seq analysis revealed 20,660 HCFC1 binding
180 peaks, the majority of which (77.1%) were located within genes, of which 44.6% were at promoter
181 regions (Extended Data Fig. 4a). Comparison of HCFC1 binding between day 0 and day 3 revealed
182 only minor changes (Fig. 3c,d), and similarly, the overlap with YY1 binding sites was evident,
183 with 84.5% of YY1 promoter peaks overlapping with HCFC1 peaks at day 3 of differentiation
184 (Fig. 3c,e). As described for human cells, HCFC1 peaks were highly enriched at CpG islands and
185 promoters (Extended Data Fig. 4c)(Michaud 2013). Moreover, HCFC1 peaks showed strong
186 enrichment for the YY1 binding motif and motifs of other transcription factors implicated in
187 HCFC1 recruitment, such as RONIN and ZNF143/STAF (Extended Data Fig. 4d)[29]. Similar to
188 undifferentiated ESCs, H3K4me3 ChIP-seq analysis at day 3 of differentiation revealed a high
189 level of overlap with HCFC1 peaks, consistent with HCFC1's role as a member of the COMPASS
190 complexes responsible for H3K4me3 catalysis (Fig. 3a).

191 Recent studies have indicated an important role for COMPASS-mediated H3K4me3 in recruitment
192 of Integrator to target genes[25]. To investigate the role of HCFC1-COMPASS-mediated catalysis
193 of H3K4me3 and subsequent recruitment of Integrator, we studied the presence and distribution
194 of H3K4me3 and INTS11, a member of the endonuclease module of the Integrator complex, before
195 and after dTAG-13-mediated depletion of HCFC1. Western blot analysis revealed a global
196 decrease of H3K4me3 presence at day 3 of differentiation upon HCFC1 removal by dTAG-13
197 treatment (Extended Data Fig. 4e,f). ChIP-seq analysis identified a genome-wide reduction of
198 H3K4me3 enrichment at 4,991 sites (9.8% of all 51,021 H3K4me3 peaks) upon HCFC1 removal,
199 confirming an important role for HCFC1 in H3K4me3 deposition, although the loss was less
200 prominent than reported for the depletion of DPY30 or RBPP5 (Fig. 3f, Extended Data Fig. 4g)[25].
201 As expected, correlation analyses between RNA-seq and ChIP-seq data showed that gene

202 expression changes following HCFC1 depletion are positively associated with changes in
203 H3K4me3 levels, (Fig. 3g, Extended Data Fig. 4h,i). Sites that displayed decreased H3K4me3
204 enrichment upon HCFC1 depletion did not exhibit or only exhibited a mild loss of INTS11
205 binding, indicating that INTS11 recruitment to those sites is predominantly H3K4me3 independent
206 (Fig. 3h, Extended Data Fig. 4g,i). Interestingly, HCFC1 sites that showed reduced H3K4me3
207 deposition upon HCFC1 removal also displayed very strong YY1 binding (Fig. 3h). Our results
208 indicate that HCFC1 plays a crucial role in H3K4me3 catalysis at HCFC1-YY1 co-bound sites,
209 but that the presence of this histone modification is unrelated to the recruitment of Integrator.

210

211 ## YY1-mediated recruitment of the COMPASS and Integrator complexes

212 Our results suggest a putative interaction between YY1 and HCFC1 in regulating H3K4me3
213 deposition. To test whether YY1 is indeed involved in recruitment of COMPASS to target
214 sequences, we generated two female 2xFLAG-V5-tagged YY1 ES cell lines and performed FLAG
215 pull-downs followed by mass spectrometry analysis (Extended Data Fig. 5a,b). Our data confirmed
216 strong interactions with the INO80 chromatin remodeling complex, as previously described[30] and
217 with HCFC1 and COMPASS member RBBP5, and revealed weak interactions with other
218 COMPASS complex subunits, likely reflecting indirect protein–protein interactions (Fig. 4a).
219 Interestingly, we also identified most Integrator complex subunits (INTS1, INTS2, INTS4, INTS7,
220 INTS8, INTS9, INTS11 and INTS12) as strong YY1 interactors, suggesting a role for YY1 in
221 recruitment of Integrator to gene regulatory elements. This includes all subunits of the RNA
222 endonuclease module of Integrator, responsible for cleavage of nascent RNA associated with
223 paused RNA pol II[31].

224 Interrogation of published ChIP-seq datasets of INTS11 distribution in undifferentiated ESCs
225 indicates a high level of overlap between genome-wide distribution of YY1 and the Integrator
226 complex (Fig. 4b, Extended Data Fig. 5c), with the immense majority (95.8%) of YY1 peaks at
227 promoters overlapping with INTS11 (Fig. 4b,c). Indeed, overlapping YY1-INTS11 peaks are
228 enriched for the YY1 motif while INTS11-only peaks are not (Extended Data Fig. 5d). These
229 results suggest a dual role for YY1 in recruitment of both HCFC1 and Integrator complexes.

230 To functionally test whether HCFC1 and Integrator recruitment is YY1-dependent, we generated
231 compound knock-in 2xFLAG-V5-Hcfc1:FKBP12^{F36V}– 2xHA-Yy1 female (XX) ESC lines

232 (HCFC1-FLAGV5/YY1-FKBP ESCs) and analysed the effect of dTAG-13-mediated depletion of
233 YY1 on HCFC1 and Integrator recruitment to regulatory elements (Fig. 4d). Western blot analysis
234 confirmed complete loss of YY1 within 2h of dTAG-13 treatment (Extended Data Fig. 5e). We
235 then proceeded to differentiate these cells for 3 days and treated them with dTAG-13 for the last 2
236 days of differentiation (Extended Data Fig. 5e). HCFC1 ChIP-seq analysis at day 3 of
237 differentiation revealed that HCFC1 enrichment at the great majority of its peaks does not change
238 upon YY1 removal, indicating that HCFC1 recruitment is predominantly mediated by other
239 transcription factor (TFs) (Fig. 4e). Nevertheless, 7.9% of HCFC1 peaks are YY1-dependent, and
240 analysis of DNA binding motifs at HCFC1 peaks with decreased enrichment in the YY1 dTAG-
241 13 condition compared to peaks with increased enrichment confirmed strong presence of the YY1
242 consensus binding motif (Fig. 4e,f). At 70 promoter-associated loci within this subset, loss of both
243 YY1 and HCFC1 is associated with reduced gene expression (Fig 4h), indicating that YY1 recruits
244 HCFC1 to a subset of promoters to activate their expression.

245 Since eight subunits of the Integrator complex were significantly pulled down by YY1[32], we
246 analysed INTS11 recruitment dependence on YY1. Loss of YY1 leads to reduction of INTS11
247 recruitment at 23.6% of all INTS11 binding sites genome-wide (Fig. 4e). INTS11 peaks with
248 decreased enrichment in the YY1 dTAG-13 condition are strongly enriched for the YY1 binding
249 motif (Extended Data Fig. 5f). Moreover, concomitant loss of YY1 and INTS11 at certain loci is
250 associated with reduction in expression of 93 proximal genes such as *Cnnm2* and *Zfp37* at day 0[33]
251 and day 3 of differentiation (Fig. 4i, Extended Data Fig. 5g-k). Although western blot analysis
252 showed a general decrease in H3K4me3 (Extended Data Fig. 5i,m), H3K4me3 ChIP-seq analysis
253 indicated that 2,057 sites (3.5% of all 58,455 H3K4me3 peaks) showed a reduction of H3K4me3,
254 a lower percentage compared to HCFC1 (Fig. 3f and Fig. 4e,i). Interestingly, most sites displaying
255 a reduction of Integrator recruitment after YY1 depletion show persistent H3K4me3 enrichment
256 (Fig. 4j). Importantly, at these sites INTS11 recruitment is HCFC1 independent (Fig. 4j), in line
257 with our finding that dTAG-13-mediated depletion of HCFC1 revealed limited effects on INTS11
258 recruitment (Fig. 3h). These results highlight an important dual regulatory role for YY1 in genome-
259 wide recruitment of COMPASS or Integrator to target genes to activate their expression.

260

COMPASS- and Integrator-mediated regulation of XCI

To understand the XCI phenotype in *Hcfc1* mutant cells, we studied the role of HCFC1/COMPASS, YY1 and Integrator in XCI regulation. ChIP-seq analysis of HCFC1 and YY1 binding in day 0 ESCs showed clear HCFC1 and YY1 overlapping peaks around the *Xist, Jpx* and *Ftx* promoter and regulatory elements (Fig. 5a)[27,28]. Analysis of day 3 differentiated female ESCs revealed HCFC1 binding at the *Xist* regulatory element, as well as at regulatory elements of negative (*Tsix,*) and positive *(Jpx*, *Ftx)* regulators of *Xist* (Fig. 5a). Importantly, dTAG-13-mediated loss of HCFC1 or YY1 both leads to reduction of H3K4me3 at the *Xist* regulatory element accompanied by a reduction in *Xist* expression at day 3 of ESC differentiation (Fig. 2f and Fig. 5a,b). Integrator recruitment at the *Xist* regulatory element appears largely unaffected by HCFC1 and YY1 depletion and their concomitant loss of H3K4me3. *Jpx* expression is also significantly reduced upon HCFC1 or YY1 removal (Fig. 2i and 5b). In contrast to *Xist*, H3K4me3 deposition at the *Jpx* promoter remains unchanged in both conditions despite loss of HCFC1 recruitment at the *Jpx* promoter following YY1 depletion (Fig. 5a). Notably, INTS11 recruitment to the *Jpx* promoter is completely abrogated upon YY1 loss indicating a crucial role for YY1 in Integrator recruitment to the *Jpx* promoter. Although loss of HCFC1 results in a mild, non-significant reduction in *Ftx* expression, YY1 depletion causes a significant decrease in *Ftx* expression. In both cases, only INTS11 recruitment, and not H3K4me3 deposition, is affected (Fig. 2i and 5a,b), suggesting an indirect effect of HCFC1 on *Ftx* expression, possibly *via* loss of promoter co-regulation and/or mass action. While *Tsix* expression is affected by YY1 removal it is not affected by HCFC1 removal. Also, INTS11 recruitment to Tsix' promoter is YY1-dependent and accompanied by a slight reduction in H3K4me3 deposition (Fig. 2i and 5a,b). Interestingly, the more severe effect of YY1 depletion on *Xist*, *Jpx* and *Ftx* compared to HCFC1 could be explained by the positive effect YY1 has on *Hcfc1* expression (Fig. 5b).

These findings indicate distinct mechanisms in the regulation of XCI. HCFC1- and YY1-mediated H3K4 trimethylation is essential for *Xist* activation, with HCFC1 activity confined to *Xist*, while YY1 plays a dual role by recruiting the Integrator complex to both *Xist*-activating (*Jpx*, *Ftx*) and -repressing (*Tsix*) cis-regulatory elements. Interestingly, both HCFC1 and YY1 are required for H3K4me3 enrichment at the *Xist* locus, despite continued HCFC1 recruitment in the absence of YY1, suggesting an additional role for YY1 in catalyzing H3K4me3 at the *Xist* regulatory element.

291

292 ## Discussion

293 Here, we provide evidence that HCFC1 plays a critical role in regulating XCI. Knockout and
294 rescue experiments indicate that HCFC1 is a dosage-sensitive X-encoded activator of XCI
295 involved in the regulation of *Xist* and *Xist* regulatory genes. Previous studies involving HCFC1
296 heterozygous knockout mice described an XCI phenotype with complete skewing of XCI towards
297 inactivation of the mutant allele[34]. As described for *Rnf12*, this skewing phenotype could be
298 explained by a continued requirement for *Hcfc1* during the XCI process but could also be explained
299 by the essential function of HCFC1 in gene regulation in general and cell cycle control[34].
300 Unfortunately, early lethality of *Hcfc1*$^{-/-}$ embryos due to imprinted XCI problems prohibited
301 studying the role of HCFC1 in random XCI in the developing embryo.

302 Although the effect of the combined heterozygous deletions of large regions containing *Hcfc1* and
303 *Rnf12* on *Xist* expression and XCI is severe, a low percentage of cells is still capable of initiating
304 XCI upon ESC differentiation pointing to the presence of additional XCI-activators located in
305 other regions and not detected in this study. Interestingly, our study did involve deletions
306 encompassing genes previously implicated in regulating XCI including *Kdm5c* and *Gata1*, yet
307 these deletions had no detectable effect on XCI in our system[14,15]. This result may be due to limited
308 expression (*Gata1*) or sensitivity of our model system and readout (*Kdm5c*), which might not have
309 been able to capture subtle effects on XCI.

310 HCFC1 is a member of the SET1A/B-COMPASS and MLL1/2-COMPASS complexes in
311 mammals (reviewed in[35]). Our study confirms the importance of HCFC1-mediated recruitment of
312 COMPASS to target sequences, H3K4me3 deposition, and proximal gene expression. However,
313 in line with its non-core function within the complex, the effect of loss of HCFC1 on H3K4me3
314 deposition is less severe than depletion of core COMPASS subunits such as DPY30 or RBBP5[25,35].
315 Moreover, we show that Integrator recruitment to HCFC1-dependent H3K4me3 promoters is
316 largely invariant, indicating other mechanisms at play to recruit Integrator to those sites.
317 Interestingly, we noticed strong enrichment of most Integrator subunits in our YY1 pulldown
318 studies, suggesting a direct role for YY1 in recruiting Integrator to target sites. Indeed, many YY1
319 sites showed reduced INTS11 recruitment upon YY1 removal and as with HCFC1, this seemed to
320 be independent of H3K4me3, while the great majority of proximal promoters were downregulated.

321 Also, although most HCFC1 binding seemed to be YY1 independent and thus mediated by other
322 TFs, we observed a subset of YY1-dependent promoter-proximal HCFC1 sites that correlate with
323 gene expression. Our study highlights the versatile role of YY1 in mediating recruitment of
324 Integrator and HCFC1-COMPASS genome-wide. Although previous studies indicated a role for
325 histone modification H3K4me3 in recruitment of the Integrator complex, our studies reveal a role
326 for YY1 in recruitment of Integrator that seems independent of H3K4me3[25].

327 By dissecting the contribution of YY1 and HCFC1 to XCI regulation, we show that YY1 and
328 HCFC1 direct dose-dependent catalysis of H3K4 tri-methylation at the *Xist* regulatory element
329 and activation of *Xist* transcription (Fig. 5c). Interestingly, although HCFC1 recruitment to this
330 element appears unaffected in the absence of YY1, HCFC1 acute depletion still leads to loss of
331 H3K4me3 suggesting a role for YY1 in recruiting COMPASS complexes independent of HCFC1,
332 or indirectly by recruitment of other factors that facilitate H3K4me3 deposition. HCFC1 depletion
333 also leads to expression reduction of the positive regulators of *Xist, Jpx*, *Ftx*, all located in the *Xist*
334 topological associated domain (TAD). Since HCFC1 depletion does not lead to a loss of H3K4me3
335 at these co-activators, downregulation of *Jpx* and *Ftx* might be an indirect effect driven through
336 loss of *Xist* co-activation. As HCFC1 has also been described as a member of the NSL complex,
337 its loss may affect histone acetylation and thereby decrease gene expression of genes involved in
338 XCI, a hypothesis that was not tested in the present study but warrants further investigation[36,37]. In
339 contrast to the effects of HCFC1 loss, YY1 reduction decreases expression of genes in both the
340 Xist-TAD and the neighbouring Tsix-TAD, which harbours negative regulators of *Xist*. Beyond
341 its role in catalysing H3K4me3 at the *Xist* regulatory element, YY1 recruits both HCFC1 and
342 INTS11 to the *Jpx* and *Tsix* promoters, and INTS11 alone to the *Ftx* regulatory element.
343 Interestingly, YY1 depletion affects Integrator recruitment to these elements but loss of HCFC1
344 recruitment seems to have limited effect on H3K4me3 deposition pointing to an additional
345 mechanism to recruit the core COMPASS complex independently of YY1 and HCFC1. Our results
346 therefore indicate a role for YY1 in regulating *Xist* expression through H3K4me3 catalysis at its
347 regulatory element, and regulating *Jpx, Ftx* and *Tsix* expression through recruitment of Integrator.

348 Our findings highlight the cooperative activities of HCFC1 and RNF12 in female exclusive
349 initiation of XCI (Fig. 5d); through dose-dependent breakdown of REX1, RNF12 facilitates
350 increased recruitment of YY1 to different regulatory elements at the XIC in female cells[7]. The
351 effect on *Xist* expression is more pronounced in female cells likely through the dual action of YY1

352 on *Xist* and its positive regulators *Jpx* and *Ftx*, and *Hcfc1* itself. HCFC1 on the contrary, only
353 appears to act through dose-dependent recruitment of COMPASS to the *Xist* regulatory region,
354 where the combined action of YY1 and HCFC1 is required for catalysis of H3K4me3. In female
355 cells, the increased activity of HCFC1, RNF12 and other XCI-activators amplifies the activation
356 of *Xist* to sufficient levels to spread locally in *cis* and silence its negative regulator *Tsix.* Negative
357 feedback will prevent inactivation of all X chromosomes except one. These findings highlight the
358 complex and concerted action of RNF12 and HCFC1 in female exclusive initiation of XCI by
359 dose-dependent action converging on the same XCI regulatory elements.

360

361

362

363 # Methods

364 ## Cells and Cell Culture

365 ### *Cell culture*

366 F1 2-1 mESCs were grown on irradiated male mouse embryonic fibroblasts (MEFs) in medium
367 containing DMEM (Gibco, 11995065) supplemented with 15% fetal calf serum, 100 U ml^{-1}
368 penicillin, 100 µg mL^{-1} streptomycin (Sigma-Aldrich, P0781), 0.1 mM non-essential amino acids
369 (Lonza, BE13-114E), 0.1 mM 2-mercaptoethanol (Gibco, 31350010) and 5.000 U mL^{-1} home-
370 made LIF. Cells were passaged at least two times prior to differentiation. mESC were depleted of
371 MEFs by pre-plating dissociated cells for 45 minutes at 37°C in non-gelatinized plates. The
372 supernatant containing mESCs was counted and subsequently spun down for 5 min at 300 × g.
373 Specific cell numbers were plated for different time points for monolayer differentiation in
374 medium consisting of IMDM GlutaMAX™ (Gibco, 31980030) supplemented with 15% FCS,
375 0.1 mM NEAA, 100 U mL−1 penicillin, 100 µg mL−1 streptomycin, 23,75 µL L−1
376 monothioglycerol (Sigma-Aldrich, M6145) and 42 µg mL−1 ascorbic acid (Sigma-Aldrich,
377 A92902). Degradation of HCFC1 or YY1 in mESCs or differentiation was achieved by refreshing
378 the media with addition of 500 nM dTAG-13 (Tocris, 6605) or equivalent amounts of DMSO at
379 the indicated time points.

380 ### *Generating Rnf12 lines*

381 $Rnf12^{+/-}$ mESC lines were generated in F1 2–1 hybrid (129/Sv-Cast/Ei) mESC by CRISPR/Cas9
382 mediated targeting using different sets of gRNAs. The $Rnf12^{+/-}$ ESC line Δ1 was generated by
383 targeting mESCs with PX459 vectors with gRNAs targeting intron 2 and the 3'UTR using Gene
384 Pulser 0.2 cm-gap cuvettes (Biorad, 1652086) at 118 kV, 1200 µF and ∞Ω in a Gene Pulser xCell
385 electroporation system (Biorad). Twenty-four hours post-transfection, medium was refreshed and
386 supplemented with 1 µg mL−1 puromycin (Sigma-Aldrich, P8833) for 36h. Drug-resistant clones
387 were picked, expanded and characterized by PCR. Due to chromosomal instability upon inducing
388 ΔJ, a second $Rnf12^{+/-}$ line was generated ($Rnf12^{+/-}$ Δ2). $Rnf12^{+/-}$ Δ2 was generated by transfecting
389 2×1 µg PX459 alternative gRNA expressing vectors as well as 2 µg homology arm vector with
390 lipofectamine 2000 (Invitrogen, 11668019). Twenty-four hours post-transfection, the medium was
391 refreshed and supplemented with 1 µg mL−1 puromycin for 48h. Drug-resistant clones were
392 picked, expanded and characterized by PCR.

393 *Generating deletion lines*

394 Deletion regions were designed so that genes within the same evolutionary strata and syntenic
395 blocks were removed together[17,18]. As described previously, deletion lines were generated by
396 transfecting 2×1 µg PX459 gRNA expressing vectors as well as 2 µg homology arm vector
397 (containing a neomycin resistance cassette) with lipofectamine. 24h post-transfection, the medium
398 was refreshed and supplemented with 1 µg mL−1 puromycin for 48h followed by 285 µg mL−1
399 Geneticin (G418) (Gibco, 11811-031) for 4-6 days for ΔA through ΔK. Combinations of 5′ and 3′
400 gRNAs were used to generate the incremental deletions ΔE1 to Δ E3 and ΔE2A to ΔE3B. No
401 Geneticin selection was performed for the incremental deletions. Drug-resistant clones were
402 picked, expanded, and characterized by microsatellite PCR. ΔA through ΔI and ΔK as well as all
403 incremental deletions were generated in Rnf12$^{+/-}$ Δ1. ΔJ was generated in Rnf12$^{+/-}$ Δ2. No deletions
404 were generated between ΔK and the telomere due to genomic instability.

405 *Generating rescue lines*

406 The coding sequence of *Hcfc1* was amplified from 129/Sv mESC cDNA and cloned into the
407 pCAG-2×FLAG-V5-IRES-Puro expression vector. Rnf12$^{+/}$, ΔE2B and ΔE3B mESCs were
408 transfected with 2 µg of either the *Hcfc1* expression vector or an empty vector control, as described
409 previously. 24h post-transfection, the culture medium was refreshed and supplemented with 1 µg
410 mL^{-1} puromycin for 5 days to select for stable integration. After two passages without selection,
411 cells were split: one fraction was used to isolate individual clones, and the remaining pool was
412 subjected to differentiation. Cells were harvested at day 4 of differentiation and analyzed by RT-
413 qPCR.

414 *Generation of knock-in ESC lines*

415 Female mESCs harboring a heterozygous doxycycline responsive endogenous *Xist* promoter[38]
416 were used as a parental cell line for generation of endogenously tagged *Hcfc1* lines. In short,
417 mESCs were transfected with gRNAs targeting the 3' end and HCFC1-2×FLAG-V5-P2A-
418 mTagBFP2 or HCFC1-2×FLAG-V5-FKBP^{F36V}-P2A-mTagBFP2 constructs for homology-
419 directed repair. 24 hours post-transfection cells were selected with medium supplemented with
420 puromycin for 48 hours. Drug-resistant clones were picked on EVOS M5000 Imaging System
421 (Thermo Scientific, AMF5000SV) for highest fluorescent intensity. To generate endogenous C-
422 terminally tagged YY1 ESCs for dTAG-inducible degradation, HCFC1-2×FLAG-V5-P2A-
423 mTagBFP2 cells were transfected with gRNAs targeting the 3'UTR and using a

424 pAW62.YY1.FKBP.knock-in.mCherry vector[33] (Addgene #10437) for homology-directed repair.
425 Due to a SNP in the 3'UTR of *Yy1*, two separate gRNAs targeting either 129/Sv or Cast/Ei were
426 selected. Drug-resistant clones were picked on EVOS M5000 Imaging System. All positive clones
427 were expanded and screened for homozygous insertions flanking the integration site by PCR and
428 western blotting.

429 Female ESCs expressing 2xFLAG-V5-YY1 were generated in a *Tsix-STOP* background by
430 random integration of a pCAG 2xFLAG V5 YY1 construct. Cells were electroporated and
431 subjected to puromycin selection 24 h later. After one week, individual clones were isolated and
432 expanded. Transgene expression was confirmed by western blotting, and karyotypes were
433 assessed. Two independent clones with normal karyotypes and YY1 expression comparable to
434 endogenous levels were selected for FLAG pull down experiments.

435 Western Blot

436 Nuclear protein extracts were performed as described previously[39] with modifications. All buffers
437 were supplemented with 1× protease inhibitors (Roche, 11836170001). Cells were harvested by
438 scraping in cold PBS on ice and subsequently centrifuged for 5 min, 300 x g at 4°C. The pellet
439 was resuspended in 5× pellet volume Buffer A (10 mM Hepes pH 7.6, 1.5 mM MgCl2 and 10 mM
440 KCl, 0.5 mM DTT), incubated for 10 min at 4°C, vortexed for 30 sec and spun down at 900 x g,
441 for 5 min, 4°C. The supernatant was transferred and the nuclear pellet was resuspended in 2× pellet
442 volume Buffer C (20 mM Hepes pH 7.6, 25% glycerol, 420 mM NaCl, 1.5 mM MgCl2, and
443 0.2 mM EDTA, 0.5 mM DTT, 0.05% NP-40). After rotating the solution for 30 min, 4°C, it was
444 subsequently spun down for 10 min at 18.000 x g, 4°C. The supernatant containing the nuclear
445 fraction was collected and the concentration was determined using Nanodrop or The Pierce BCA
446 Protein Assay Kit (Thermo Scientific, 23227). For western blot analysis, lysates were diluted 1:1
447 with 2× Laemmli Sample Buffer (0.125M Tris-HCl pH 6.8, 4% SDS, 20% (w/v) glycerol, 0.01%
448 bromophenol blue, 150 mM DTT) and boiled for 5 min at 95°C. Equal concentration of protein
449 lysates were loaded in the 4–15% Criterion™ TGX™ Precast Midi Protein Gel, 18-well (Biorad,
450 #5671084) with Precision Plus Protein Dual Color Standards (Biorad, #1610374) and run in 1x
451 Tris/Glycine/SDS buffer (Biorad, #1610772). Samples were transferred to a Trans-Blot Turbo
452 PVDF membrane (Biorad, #1704157) using semi-dry transfer in the Trans-Blot Turbo Transfer
453 System for 7 min at 2.5 A constant up to 25V. After transfer, membranes were blocked with 5%
454 milk powder in 1×TBS (0.02M Tris-HCl, 0.15M NaCl, pH 7.6). The membrane was washed 3×

455 with 1×TBST (0.02M Tris-HCl, 0.15M NaCl, 0.1% Tween-20, pH 7.6), incubated with the
456 appropriate primary antibodies, HCFC1 Carboxy-terminal Antigen antibody (Cell signaling,
457 50708S, 1:1000), V5 antibody (Invitrogen, R96025, 1:5000), YY1 antibody (Santa Cruz, sc-7341,
458 1:200) and/or H3K4me3 antibody (Sigma-Aldrich, 05-1339, 1:1000). Membranes were washed
459 3× with 1× TBST and incubated with the appropriate secondary antibodies, IRDye 800CW Goat
460 anti-Mouse IgG (LI-COR Biosciences, 926-32210), IRDye 800CW Donkey anti-Mouse IgG (LI-
461 COR Biosciences, 926-32212) and/or IRDye® 680RD Donkey anti-Rabbit IgG (LI-COR
462 Biosciences, 926-68073) and visualised on the Odyssey CLx Imager. Last, the membranes were
463 incubated with β-Actin−Peroxidase (Sigma-Aldrich, A3854), stained with ECL Western blotting
464 detection reagents (Cytyva, RPN2209) and visualised on the AI600 Imager.

465 Expression analysis and RNA-seq

466 mESCs were depleted of MEFs as described above, supernatants were spun down and cell pellets
467 were lysed. Differentiated cells were washed 1× in PBS and lysed directly in the culture dish prior
468 to extraction. Total RNA was extracted using the ReliaPrep™ RNA Miniprep kit (Promega,
469 Z6012) with increased DNase treatment (25 min at room temperature). Reverse transcription was
470 performed on 500 ng RNA using Superscript III (Invitrogen, 18080093), random hexamers
471 (Invitrogen, N8080127) and RNaseOUT (Invitrogen, 10777019). The cDNA was diluted 1:1 in
472 water prior to amplification. RT-qPCRs, allele-specific RT-qPCRs and qPCRs were performed in
473 triplicate in the GoTaq qPCR Master Mix (Promega, A6002) in a CFX384 Touch Real-Time PCR
474 Detection System (BioRad, LJB22YE8Z). Allele-specific primer concentrations were determined
475 and optimized using pure gDNA of 129/Sv and Cast/Ei mice. Rplp0 was used as reference gene
476 for relative quantification of HCFC1-FKBP cells. Hist2h2aa1 was used as reference gene for all
477 other qPCRs. All primers are included in Extended Data Table 1.

478 Sequencing libraries were prepared using the Illumina TruSeq stranded mRNA library preparation
479 kit. Libraries from the deletion cell lines were sequenced on a HiSeq 2500 sequencer, generating
480 single-end 50 bp reads. Libraries from the HCFC1 dTAG-13 and DMSO-treated samples were
481 sequenced on an Illumina NextSeq 2000 sequencer, resulting in paired-end read clusters of 50
482 bases in length.

483 ## RNA-seq analysis

484 RNA-seq reads were processed allele-specifically. First, we created an N-masked reference
485 genome from mm10 using SNPsplit v0.3.4[40], masking all SNPs between the 129/Sv and Cast/Ei
486 strains based on the dbSNP v142 database[41]. Reads were trimmed using TrimGalore v0.6.7[42] and
487 aligned to the N-masked reference genome using HISAT2 v2.2.1[43]. Mapped reads were then
488 assigned to either the 129/Sv, Cast/Ei, or unassigned BAM file based on the best alignment using
489 SNPsplit (default settings with --paired for the paired-end data). All allele-specific and unassigned
490 reads were merged into a composite BAM file using Samtools v1.10[44].

491 Transcript-level counts were obtained from each BAM file using featureCounts v2.0.6[45] with the
492 Ensembl v98 gene annotations (-t exon -s 2). We normalized the raw counts to transcripts per
493 million (TPM) using a custom Python script. Gene expression levels were visualized in bar plots,
494 showing TPM values normalized to the average TPM of DMSO-treated or untreated control
495 samples. To validate the deleted regions, $\log_2$(TPM+1) values were calculated per allele and
496 visualized in a heatmap across all samples.

497 Differential expression analyses were performed using DESeq2 v1.30.1[46] with parametric
498 dispersion fitting. Fold changes were shrunk using the ashr v2.5-54 method[47]. Differentially
499 expressed genes (DEGs) were defined as those with an adjusted p-value < 0.05 and an absolute
500 $\log_2$(fold change) > 1. Results were visualized in volcano plots, with autosomal and X-linked
501 DEGs coloured separately.

502 To assess X-chromosome silencing, we calculated allelic ratios for *Rnf12*$^{+/-}$, ΔE, ΔE2B and ΔE3B
503 as well as HCFC1 dTAG-13 and DMSO-treated samples. For each condition, genes with more
504 than 20 allele-specific reads across all replicates were included. The allelic ratio was defined as:

505 $$\frac{Xi}{Xi + Xa}$$

506 where Xi represents 129/Sv and Xa represents Cast/Ei. Allelic ratios of X-linked genes were
507 visualized using violin plots combined with boxplots. Statistical significance was assessed using a
508 two-sided Mann-Whitney U test ($\alpha < 0.05$).

509 We reanalysed YY1-FKBP and wild-type RNA-seq data from GSE99521[33] using the same allele-
510 specific pipeline as described above, except that the N-masked reference genome was generated

511 using SNPs between C57BL/6 and 129/Sv. Additionally, to assess expression dynamics of
512 candidate genes during early differentiation, we analysed a bulk RNA-seq dataset from
513 GSE151009[19]. Allele-specific count tables were downloaded and normalized to TPM values. For
514 both ΔXic^{B6} and ΔXic^{Cast} cell lines, allelic expression of *Hcfc1*, *Irak1, Rnf12* and *Xist* was plotted
515 over time in line plots, displaying the mean and 95% confidence interval. Statistical significance
516 between day 0 and day 1 was assessed using one-sided independent t-tests, followed by multiple
517 testing correction with the Benjamini-Hochberg method.

518 RNA-FISH

519 Xist probes were made from a 19 Kb Xist/Tsix plasmid (p510) using the Nick Translation Kit
520 (Abbott, 07J00-001) with red dUTPs (Abbott 02N34-050) according to the manufacturer's
521 instructions. Per coverslip, 3 µl of Xist probe was mixed with 1 µl salmon sperm DNA, 0.5 µl
522 NaOAc 3M and 100 µl 100% EtOH and centrifuged 30 min, 12.800 rpm at 4°C. Pellet was washed
523 with 500 µl cold 70% EtOH and let dry. 7 µl of formamide was added to the probe, incubated 10
524 min at 75°C and subsequently quenched on ice. 7 µl of hybridization buffer (20% Dextran sulfate,
525 2 mg/ml BSA, 20 mM VRC, 4×SSC), mixed and added to the coverslip.

526 ESCs were grown on gelatin coated coverslips at different time points for monolayer
527 differentiation. Cells were fixed in 4% PFA in PBS for 10 min while rocking and rinsed 3× with
528 PBS. Cells were subsequently permeabilized in 0.5% Triton, 20 mM VRC (NEB, S6639) in PBS
529 for 4 min on ice. Cells were sequentially dehydrated by addition of 70, 90 and 100% ethanol,
530 followed by airdrying. The probe, dissolved in hybridization buffer was added to the coverslip and
531 incubated overnight at 37°C in humid chamber with 50% formamide in 2× SSC. After
532 hybridization, coverslips were washed 3 × with 50% formamide in 2×SSC at 42°C, 2 × with 2×SSC
533 at 42°C and once with 2 × SSC supplemented with 4 µM Hoechst (Thermo Scientific, 62249)
534 before mounting with ProLong™ Gold Antifade (Invitrogen, P36930).

535 WGS

536 ESCs were depleted of MEFs by pre-plating dissociated cells for 45 minutes at 37°C in non-
537 gelatinized plates and subsequently grown for several days in a 6-well dish without MEFs. Cells
538 were collected and lysed overnight in low-SDS lysis buffer (50 mM Tris-HCl pH 8.0, 100 mM
539 EDTA, 0.2% SDS, 100 mM NaCl) with Proteinase K for Phenol-Chloroform DNA extraction.
540 Following DNA extraction, 10 µg DNA was treated with 4 µg/ml RNase A treated for 1 hour at

541 37°C followed by clean up with the QIAamp DNA Micro Kit (QIAGEN, 56304). Sequencing
542 libraries were prepared using the Nextera DNA Flex protocol from Illumina using 50 ng of
543 genomic DNA and 5 cycles of PCR. Libraries were sequenced on an Illumina HiSeq 2500
544 sequencer, generating 50 bp single reads. For ΔE, over 50 million reads were obtained, while all
545 other samples yielded at least 20 million reads.

546 WGS analysis

547 Reads were trimmed using TrimGalore v0.6.7[42] and aligned to the N-masked reference genome
548 (see RNA-seq analysis) using Bowtie2 v2.5.1[48]. Mapped reads were assigned to either the 129/Sv
549 or Cast/Ei BAM file using SNPsplit v0.3.4[40], based on the best alignment. Allele-specific read
550 counts were calculated for genomic windows of 100 kb using deepTools v3.5.5[49] bamCoverage (-
551 bs 100000 -of bedgraph --normalizeUsing None). For each window with more than 10 allele-
552 specific reads, the $\log_2$(fold change) between the 129/Sv and Cast/Ei allele was computed.
553 Statistical significance between both alleles was assessed using two-sided binomial tests with
554 p=0.5, assuming equal read counts from both alleles in non-deleted regions. Resulting p-values
555 were corrected for multiple testing using the Benjamini-Hochberg method. Data were visualized
556 in scatter plots showing the $\log_2$(fold change) across the entire X-chromosome and within the 67-
557 76.5 Mb region. For visualization, extreme $\log_2$(fold change) values were capped at ±2.
558 Statistically significant windows ($\alpha < 0.05$) were highlighted, as well as the window containing
559 the *Hcfc1* locus.

560 DNA FISH

561 BACs/Fosmids were selected inside (MeCP2; WI1-1364A8) and outside the deleted region
562 (HPRT; RP23-412J16 and Xist; WI1-2363H9). Probes were labeled using the Nick Translation
563 Kit (Abbott, 07J00-001) with red (Abbott 02N34-050) or green (Abbott 02N32-050) dUTPs
564 according to the manufacturer's instructions.

565 DNA FISH was performed as described previously[50]. Cells were blocked in metaphase by addition
566 of 12 µl/ml KaryoMAX Colcemid (Gibco, 15210040) to the culture medium for 2 hours. Cells
567 were trypsinized and resuspended in hypotonic buffer (0.075 M KCl) followed by fixation by
568 addition of 3:1 methanol-acetic acid solution. Metaphase spreads were made by dropping fixed
569 cells on coverslips and aged overnight at 37°C. Coverslips were equilibrated in 2×SSC (1× SSC:
570 0.15 M NaCl, 0.015 M sodium citrate), treated with RNAse A (1 µg/ml in 2×SSC) for 5 min at

571 37°C in humid chamber and washed 3× with 2×SSC at room temperature. Cells were
572 permeabilized in pre-warmed 1% pepsin in 0.01M HCl solution for 7.5 min at 37°C. After
573 permeabilization, cells were washed 2× with PBS and prefixed in 1% formaldehyde for 5 min at
574 room temperature. Coverslips were washed 2× with PBS and sequentially dehydrated by addition
575 of 70, 90 and 100% ethanol and airdrying.

576 5 µl of Nick-translated DNA probes was combined with 3 µg Mouse Cot-1 DNA (Invitrogen,
577 18440016) and 10 µg UltraPure™ Salmon Sperm DNA (Roche, 11467140001) and precipitated
578 using NaAc and ethanol. The pellets were washed with 70% ethanol, spun down and air dried. The
579 dried pellet was dissolved in formamide and denatured by incubating for 10 min at 75°C while
580 shaking (1000 Rpm) and subsequently mixed 1:1 with cold 2×hybridization buffer (50%
581 formamide, 2× SSC, 50 mM phosphate buffer (pH 7.0), 10% dextran sulfate and 1 ng/µl mouse
582 Cot DNA). Following preparation of the probe mix, it was added to the coverslips and incubated
583 on heatplate for 2 min at 75°C. Slides were transferred to a humid chamber with 50% formamide
584 in 2× SSC and incubated overnight at 37°C. The next day, slides were washed 2× in 50%
585 formamide in 2× SSC at 42°C, washed 2× in 2× SSC at 42°C, air dried and mounted with ProLong
586 Gold Antifade with DAPI (Invitrogen P36931).

587 ## ChIP-seq

588 ChIP-seq was performed as previously described[39] with several modifications. Up to $1x10^8$
589 attached cells were fixed with 20mM disuccinimidyl glutarate (ThermoFisher, 20593-50MG) in
590 PBS for 45min RT, washed twice with PBS followed by fixing in 1% PFA (Sigma-Aldrich, F1635-
591 25ML) in PBS for 10min RT for the V5-HCFC1 and Integrator ChIPs, or only 1% PFA in medium
592 for 10min RT for the H3K4me3 ChIPs (V5-HCFC1 ChIP was performed in FLAGV5-HCFC1
593 YY1-FKBP ESCs). After crosslinking, cells were quenched by addition of 125 mM glycine for 5
594 min. Cells were washed 1× with PBS before scraping in PBS with 1x protease inhibitors (Roche,
595 11836170001). From this point onwards, all buffers contained protease inhibitors (Roche,
596 4693132001) unless stated. Fixed cells were pelleted and subsequently resuspended in LB1 buffer
597 (50 mM HEPES pH 8.0, 140 mM NaCL, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton
598 X-100) and incubated for 10 min at 4 °C while rotating, then in LB2 buffer (10 mM Tris-HCl pH
599 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) and last in LB3 buffer (10 mM Tris-HCl pH
600 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Sodium Deoxycholate, 0.5% N-

601 laurolylsarcosine). DNA was sheared using a Bioruptor Pico (Diagenode, B01060001) for 10
602 cycles of 30 sec on and 30 sec off (single crosslinked samples) or 14 cycles of 30 sec on and
603 30 sec off (for double crosslinked samples) to an average fragment length of 300 bp. 1% triton
604 X-100 was added to the sheared chromatin and incubated for 10 minutes at 4°C and spun down
605 for 10 min at 4500 x g. Chromatin was aliquoted, flash frozen in liquid nitrogen and stored at -
606 80°C until further use. Different amounts of chromatin were used for the H3K4me3 ChIP (25 ug),
607 HCFC1-V5 and INTS11 ChIPs (150 ug). Each chromatin sample was diluted to 500 ul using ChIP
608 dilution buffer (1.1% Triton X-100, 0.01% SDS, 167 mM NaCl, 16.7 mM Tris-HCl and 1.2 mM
609 EDTA) and 1% of it was substracted for INPUT and frozen at -20°C. A specific amount of
610 antibody was added (Ab; 2 ug H3K4me3 Ab, Cell Signaling Technologies, 9751S [C42D8]; 10
611 ug INTS11 Ab, Bethyl Laboratories, A301-274A; 7ul V5 Ab, Invitrogen R96025) to each
612 chromatin sample and incubated with rotation overnight at 4°C. The next day, 30 ul Dynabeads
613 protein G beads were washed (Invitrogen, 10004D) 2x with 500 ul ChIP dilution buffer. Chromatin
614 with antibody was added to the beads and incubated 2h at 4°C. Afterwards, by magnetic separation,
615 beads were washed 5 min 4°C 3x with Low salt wash buffer (20 mM Tris-HCl, 2 mM EDTA, 1%
616 Triton X-100, 150 mM NaCl, 0.1% SDS), 1x with High salt wash buffer (20 mM Tris-HCl, 2 mM
617 EDTA, 1% Triton X-100, 500 mM NaCl, 0.1% SDS) and 1x with LiCl wash buffer (10 mM Tris-
618 HCl, 1 mM EDTA, 0.25 M LiCl, 1% NP-40, 1% deoxycholate). Beads were finally washed with
619 TE without protease inhibitors. 200 ul of ChIP elution buffer (0.1 M $NaHCO_3$, 10 mM EDTA and
620 1% SDS) were added to the beads with 22 ml proteinase K (10 mg/mL) and 5 ml RNase (10
621 mg/mL), and shaken at 1,000 rpm on an orbital thermo-mixer for 2 h at 37°C and finally at 65°C
622 overnight. DNA was then cleaned using Zymo ChIP DNA clean concentrator kit (Zymo
623 Research, D5202) and eluted in 12 ul. 0.5 ul was used to confirm ChIP efficiency. The
624 concentration of the ChIP samples was measured, and sequencing libraries were prepared using a
625 ThruPLEX DNA- V H T . L W 7 D N D U D % L R 5 I R O O R Z L b r a r i e s P D Q X I D F
626 were sequenced on an Illumina NextSeq 2000 sequencer, generating 50 bp paired-end reads.

627 ## ChIP-seq analysis

628 Reads were trimmed using TrimGalore v0.6.7[42] and aligned to the N-masked reference genome
629 (see RNA-seq analysis) using Bowtie2 v2.5.1[43]. For datasets containing Drosophila melanogaster
630 spike-ins, reads were also aligned to the dm6 genome, and those with higher alignment scores to
631 dm6 than to mm10 were separated into a dedicated FASTQ file. All remaining reads were assigned

632 to either the 129/Sv, Cast/Ei, or unassigned BAM file based on the best alignment using SNPsplit
633 (--paired). Allele-specific and unassigned reads were merged into a composite BAM file using
634 Samtools v1.10[44].

635 Peak calling was performed on the composite BAM files using MACS3 v3.0.1[51] with the callpeak
636 function (-f BAMPE), comparing ChIP samples to their corresponding input controls. Bigwigs
637 with signal tracks were generated using MACS3 bdgcmp with the ppois method. Peaks were
638 annotated using the annotatePeaks.pl script from Homer v4.11[52], and motif enrichment analysis
639 was conducted with findMotifsGenome.pl (-size 200). Small annotation categories were merged
640 for clarity: 3' UTR and TTS were combined into "3' UTR - TTS", and exon and non-coding into
641 "exon".

642 Several ChIP-seq datasets were reanalysed using the same approach as described above. Reads
643 from a YY1 ChIP-seq dataset in day 0 ESCs were obtained from GSE240684[27] and analysed allele-
644 specifically. In addition, we downloaded ChIP-seq data for H3K4me3, DPY30, PolIISer5, and
645 INTS11 in day 0 ESCs from GSE181714[25], as well as a HCFC1 ChIP-seq dataset in day 0 ESCs
646 from GSE49847[28]. As these datasets were generated from non-hybrid cells, reads were mapped to
647 the mm10 genome, and the resulting BAM files were used as composite BAM files for all
648 subsequent analyses.

649 ### Overlap between peak locations

650 To assess overlap between HCFC1 binding and other chromatin features, HCFC1 day 0 peaks
651 were used as a reference to evaluate enrichment of HCFC1 (day 3), H3K4me3 (day 0 and 3), YY1,
652 DPY30, PolIISer5, and INTS11 (all day 0). Heatmaps were generated using deepTools v3.5.5 with
653 computeMatrix (reference-point mode, averageTypeBins mean, 10bp bins, ±3kb) followed by
654 plotHeatmap.

655 Overlap between HCFC1 (day 0 and 3) and YY1 (day 0) peaks was quantified using intervene
656 v0.6.5[53] venn (--type genomic --save-overlaps). Resulting BED files for each overlap category
657 were annotated with Homer annotatePeaks.pl. The proportion of peaks per annotation category
658 was calculated for each overlap group and visualized in stacked bar plots. To specifically evaluate
659 promoter-associated binding, peaks were filtered for those located in promoter regions, and their
660 overlap was visualized in a Venn diagram using matplotlib-venn v1.1.1. Overlap between YY1
661 and INTS11 binding on day 0 and 3 was assessed following the same approach.

662 Correlation between HCFC1 binding at day 0 and day 3 was further assessed using deepTools
663 multiBigwigSummary in BED-file mode, with a BED file containing all TSSs ±500 bp in mm10.
664 Counts (obtained with outRawCounts setting) were compared in a scatter plot using log1p-
665 transformed values. Pearson correlation was calculated using the plotCorrelation function with the
666 removeOutliers option.

667 Differential peak calling in H3K4me3 enrichment

668 To detect global changes in H3K4me3 levels, *Drosophila melanogaster* spike-ins were
669 incorporated into all H3K4me3 samples. The number of reads mapping better to dm6 than mm10
670 (see ChIP-seq analysis) were counted and used to scale CPM-normalized tracks. Initial
671 visualization of these tracks revealed large variability between replicates (data not shown).
672 Consequently, the spike-in normalization was excluded from the final analysis. Instead, all
673 H3K4me3 tracks were normalized using the ChIPseqSpikeInFree v1.2.4 method[54]. Composite
674 BAM files were processed to compute read coverage in sliding windows, which were normalized
675 to CPM and used to calculate scaling factors for each sample.

676 Differential peak calling between dTAG-13 and DMSO conditions was performed using DESeq2
677 v1.30.1[46]. Peaks identified by MACS3 in each condition were merged using BEDTools v2.31.1
678 merge[55], and the resulting BED file was converted to SAF format. Read counts per merged peak
679 were quantified with featureCounts v2.0.6[45] and provided to DESeq2. The estimated size factors
680 were adjusted using the scaling factors computed by ChIPseqSpikeInFree. A Wald test was
681 performed to compare dTAG-13 and DMSO conditions, and fold changes were shrunk using the
682 ashr v2.5-54 method[47]. Differential peaks were defined as those with an adjusted p-value < 0.05
683 and an absolute $\log_2$(fold change) > 1.

684 Tracks for each sample were scaled using the size factors computed by ChIPseqSpikeInFree.
685 Average tracks per condition were generated with deepTools bigwigAverage. Moreover, tracks
686 with $\log_2$(fold change) between dTAG-13 and DMSO were computed using deepTools
687 bigwigCompare. To show HCFC1- and YY1-associated changes in H3K4me3 enrichment, we
688 generated heatmaps with deepTools computeMatrix (reference-point mode, averageTypeBins
689 mean, 100bp bins, ±3kb) and plotHeatmap. Differential peaks with significantly reduced
690 H3K4me3 signal (as identified by DESeq2) were compared to unaffected peaks, which were

691 randomly downsampled to match the number of reduced peaks. Both $\log_2$(fold change) tracks and
692 DMSO tracks were plotted for comparison.

693 ### Differential peak calling in HCFC1 and INTS11 binding

694 Changes in HCFC1 and INTS11 binding in YY1-FKBP cells after dTAG-13 treatment were
695 evaluated using MACS3 bdgdiff. The genome-wide number of unique and overlapping peaks were
696 visualized in Venn diagrams using matplotlib-venn. Peaks from each category were annotated
697 using the annotatePeaks.pl script from Homer v4.11 (Heinz et al. 2010), after which small
698 annotation categories were merged for clarity: 3' UTR and TTS were combined into "3' UTR -
699 TTS", and exon and non-coding into "exon". Motif enrichment analyses were conducted on each
700 peak category separately with Homer findMotifsGenome.pl (-size 200). Enrichment results for
701 known motifs were compared using scatter plots, where p-values for each motif were plotted
702 between two categories, and motifs were coloured by TF family.

703 ### Correlation between RNA-seq and ChIP-seq in HCFC1-FKBP cells

704 To correlate RNA-seq changes between HCFC1-FKBP dTAG-13 and DMSO treatments with
705 changes to H3K4me3 and INTS11 ChIP-seq signal under the same conditions, we first selected
706 DEGs from the RNA-seq analysis (adjusted p-value < 0.05, absolute $\log_2$(fold change) > 1). To
707 allow comparison with unaffected expressed genes, we matched each gene in the downregulated
708 group to a gene in the non-significant group with a similar expression level. This was done using
709 the NearestNeighbors algorithm (n_neighbors=1) from scikit-learn v1.5.1 (Pedregosa et al., 2011),
710 applying the fit and kneighbors functions. To avoid duplicate matches, each selected nearest
711 neighbor was removed from the pool after selection. This resulted in a non-significant gene subset
712 of equal size and similar expression distribution as the downregulated group. Finally, we visualized
713 fold changes in H3K4me3 and INTS11 enrichment following dTAG-13 treatment, as well as
714 baseline enrichment of HCFC1, YY1, H3K4me3, and INTS11 at TSSs of upregulated,
715 downregulated, and matched unaffected genes.

716 We quantified H3K4me3 changes in the promoter region using deepTools multiBigwigSummary
717 in BED-file mode, with a bed file containing regions spanning TSS + 1kb, based on the enrichment
718 pattern observed in the ChIP-seq heatmap. The resulting mean signal change per promoter region
719 was visualized in violin plots for each gene category separately. Statistical significance between

720 groups was assessed using a two-sided Mann-Whitney U test, with p-values corrected for multiple
721 testing using the Benjamini-Hochberg method ($\alpha < 0.05$).

722 Correlation between YY1-FKBP dTAG-13 vs DMSO RNA-seq and INTS11 ChIP-seq

723 To assess the relationship between transcriptional changes upon YY1 degradation and alterations
724 in INTS11 binding, we compared RNA-seq data from YY1-FKBP dTAG-13 and DMSO-treated
725 samples with changes in the INTS11 ChIP-seq signal under the same conditions. INTS11 peaks
726 that were significantly reduced upon YY1 dTAG-13 treatment (see 'Changes in HCFC1 and
727 INTS11 binding') were intersected with YY1 ChIP-seq peaks using intervene venn (type genomic,
728 save-overlaps). Overlapping peaks were annotated with Homer annotatePeaks.pl, and genes with
729 shared peaks located within 1 kb of a transcription start site (TSS) were selected. The resulting
730 gene list was compared to the set of DEGs from the RNA-seq analysis. The number of up- and
731 downregulated genes among these shared targets was visualized in a bar plot. Statistical
732 significance was assessed using a Pearson's chi-squared (χ^2) goodness-of-fit test, comparing the
733 observed number of up- and downregulated genes to the expected distribution based on the overall
734 proportions of up- and downregulated genes.

735 Genome browser views

736 Genome browser views were created using IGV v2.18.2[56]. ChIP-seq signal tracks were generated
737 using the bdgcmp function in MACS3 (see ChIP-seq analysis). Log_2(fold change) tracks
738 comparing dTAG-13 and DMSO conditions were computed based on CPM-normalized coverage
739 for HCFC1 and INTS11 binding, or ChIPseqSpikeInFree-scaled CPM-normalized tracks for
740 H3K4me3 (see "Differential peak calling in H3K4me3 enrichment") using deepTools
741 bigwigCompare.

742 Mass Spectometry

743 *Nuclear extracts preparation*

744 All subsequent steps for nuclear extract preparation and FLAG-purification were performed at 4°C
745 using buffers supplemented with 1× protease inhibitors (Roche, 4693132001) and 15 µM MG-132
746 proteasome inhibitor (Sigma, C2211). Pre-chilled buffers and equipment were used throughout,
747 and low protein-binding Falcon and Eppendorf tubes were employed to minimize sample loss. One
748 day prior to harvest, two independently generated FLAG-V5-YY1 and 4 control mESC lines were
749 trypsinized, MEFs were removed by pre-plating for 45 min in non-gelatinized culture dishes. The

750 resulting mESCs were cultured without MEFs for 16h before processing. Cells were washed with
751 cold PBS, collected by scraping on ice in PBS+ (1×PBS, 0.5 mM DTT), transferred to a Falcon
752 tube and pelleted by centrifugation at 1500 rpm for 5 min. Pellets were resuspended in 2 pellet
753 volumes of PBS+ and centrifuged at 1500 rpm for 5 min. The pellets were resuspended in 5 pellet
754 volumes of buffer A (10 mM Hepes pH 7.6, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT), incubated
755 on ice for 10 min and centrifuged at 3000 rpm for 5 min. The supernatant was removed, and the
756 pellets were resuspended in 2 pellet volumes of buffer A. Cells were lysed with 10 strokes of a
757 Dounce homogenizer (type A pestle) on ice. The homogenate was transferred to Falcon tubes and
758 centrifuged at 3000 rpm for 5 min. The supernatant, containing the cytoplasmic fraction, was
759 discarded. The nuclear pellet was resuspended in 1.5 pellet volumes of Buffer C (20 mM Hepes
760 pH 7.6, 20% glycerol, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT) and
761 homogenized with 10 strokes of a Dounce homogenizer (type B pestle) on ice. The nuclear
762 suspension was transferred to a Falcon tube, incubated while rotating for 30 min and centrifuged
763 at 14000 rpm for 15min. The supernatant, containing the nuclear fraction, was collected and
764 protein concentration was determined using a Nanodrop.

765 *FLAG pull down*

766 Protein concentrations of nuclear extracts were adjusted to the lowest concentration using buffer
767 C. For each sample, 1 mL nuclear extract was diluted with 1 mL buffer C-0 (20 mM Hepes pH
768 7.6, 20% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA), to reduce the salt concentration to
769 approximately 150 mM NaCl. Samples were incubated on ice for 15 min, centrifuged at 4000 rpm
770 for 15 min, and the supernatant was transferred to a new tube. Benzonase nuclease (200 U;
771 Novagen, 70664) was added to prevent RNA- or DNA-mediated interactions. Anti-Flag M2
772 Affinity Gel (Sigma-Aldrich, A2220) beads slurry (120 µl per sample) was washed four times with
773 1 ml Buffer C-150 (20 mM Hepes pH 7.6, 20% glycerol, 150 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM
774 EDTA, 0.02% NP40, 0.5 mM DTT) by centrifugation at 1000 rpm for 1 min. Nuclear extracts
775 were added to the washed beads, resuspended, and incubated with gentle rotation for 3 h. Anti-
776 FLAG-beads were pelleted by centrifugation at 2000 rpm for 2 min and washed five times with 1
777 mL buffer C-100 (20 mM Hepes pH 7.6, 10% glycerol, 100 mM KCl, 1.5 mM MgCl2, 0.2 mM
778 EDTA, 0.02% NP40, 0.5 mM DTT). Each wash was performed with 5 min rotation (10 min for
779 the final wash), followed by centrifugation at 2000 rpm for 2 min. The bound fraction was eluted
780 by resuspending the Anti-FLAG-beads with 70 µL 0.2 mg/mL FLAG tripeptide peptide (Sigma,

F4799) in buffer C-150 (without NP40), incubating for 15 min on ice, and centrifuging at 2000 rpm for 2 min. The supernatant was collected as the eluted fraction, and the elution step was repeated four times. An aliquot of the eluted fraction was analyzed by Western blot to confirm the presence of FLAG-V5-YY1 protein. Eluted fractions were combined and centrifuged at 2000 rpm for 2 min to remove residual Anti-FLAG beads. Protein eluates were concentrated by precipitation with trichloroacetic acid (w/v) / 0.1% sodium deoxycholate (w/v). Pellets were resuspended in 2× Laemmli sample buffer, boiled for 5 min at 95℃, separated by SDS-PAGE on 10% NuPAGE Bis-Tris Gel (Invitrogen, NP0315BOX) and stained using a colloidal blue staining kit (Invitrogen, LC6025).

Sample preparation

Complete lanes of SDS-PAGE gels were excised and divided into smaller slices. Proteins were reduced with 10 mM dithiothreitol (DTT) for 1 h at room temperature, alkylated with 55 mM chloroacetamide (CAM) for 1 h in the dark, and digested overnight at room temperature with sequencing-grade trypsin (1:100, w/w; Roche). Peptides were desalted using Sep-Pak tC18 Vac cartridges (Waters) and eluted with 80% acetonitrile (AcN).

MS analysis

Peptides were analyzed by nanoflow LC-MS/MS using an EASY-nLC system (Thermo) coupled to a Fusion Tribrid Orbitrap mass spectrometer (Thermo) operating in positive ion mode and equipped with a nanospray source. Peptide mixtures were first trapped on a ReproSil C18-reversed S K D V H F R O X P Q ’ U 0 D L V F K * P E + - K R F X P V H D W P D S O D F H U H G W R then separated on a ReproSil C18 reversed S K D V H F R O X P Q ’ U 0 D L V F K * P E + packed in-house) using a linear gradient of 5–90% buffer B (0.1% FA, 80% v/v AcN) over 70 or 120 min at a constant flow rate of ~200 nL/min. The column effluent was directly sprayed into the ESI source of the mass spectrometer. Spectra were acquired in profile mode with an MS1 resolution of 70,000 (AGC target: 3E6) across an m/z range of 350–1750. Data-dependent acquisition (Top15) was performed with HCD fragmentation, using an isolation window of 3.0 m/z and a normalized collision energy of 26–28%. MS2 spectra were recorded in the ion trap. Singly charged precursors were excluded, dynamic exclusion was set to 20 s, and the intensity threshold was set to 8.0E3.

Data analysis

YY1

2. Luchsinger-Morcelle, S. J., Gribnau, J. & Mira-Bontenbal, H. Orchestrating Asymmetric Expression: Mechanisms behind Xist Regulation. *Epigenomes* 8, 6 (2024).

3. Schwämmle, T. & Schulz, E. G. Regulatory principles and mechanisms governing the onset of random X-chromosome inactivation. *Curr. Opin. Genet. Dev.* 81, 102063 (2023).

4. Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F. & Gribnau, J. X inactivation counting and choice is a stochastic process: evidence for involvement of an X-linked activator. *Cell* 132, 410–421 (2008).

5. Jonkers, I. et al. RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. *Cell* 139, 999–1011 (2009).

6. Gontan, C. et al. RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature* 485, 386–390 (2012).

7. Makhlouf, M. et al. A prominent and conserved role for YY1 in Xist transcriptional activation. *Nat. Commun.* 5, 4878 (2014).

8. Barakat, T. S. et al. The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. *Mol. Cell* 53, 965–978 (2014).

9. Shin, J. et al. Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. *Nature* 467, 977–981 (2010).

10. Gontan, C. et al. REX1 is the critical target of RNF12 in imprinted X chromosome inactivation in mice. *Nat. Commun.* 9, 4752 (2018).

11. Shin, J. et al. Rnf12/RLIM is dispensable for X-chromosome inactivation in the mouse embryonic epiblast. *Nature* 511, 86–89 (2014).

12. Tian, D., Sun, S. & Lee, J. T. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* 143, 390–403 (2010).

13. Chureau, C. et al. Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X inactivation center region. *Hum. Mol. Genet.* 20, 705–718 (2011).

14. Samanta, M. K. et al. Activation of Xist by an evolutionarily conserved function of KDM5C demethylase. *Nat. Commun.* 13, 2602 (2022).

15. Ravid Lustig, L. et al. GATA transcription factors drive initial Xist upregulation after fertilization through direct activation of long-range enhancers. *Nat. Cell Biol.* 25, 1704–1715 (2023).

16. Gribnau, J. & Grootegoed, J. A. Origin and evolution of X chromosome inactivation. *Curr. Opin. Cell Biol.* 24, 397–404 (2012).

17. Disteche, C. M. Escapees on the X chromosome. *Proc. Natl. Acad. Sci. U. S. A.* 96, 14180–14182 (1999).

875 18. Prada, C. F. & Laissue, P. A high resolution map of mammalian X chromosome fragile
876 regions assessed by large-scale comparative genomics. *Mamm. Genome Off. J. Int. Mamm.*
877 *Genome Soc.* **25**, 618–635 (2014).

878 19. Pacini, G. *et al.* Integrated analysis of Xist upregulation and X-chromosome inactivation
879 with single-cell and single-allele resolution. *Nat. Commun.* **12**, 3638 (2021).

880 20. Yu, H. *et al.* The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1
881 and HCF-1 and is a critical regulator of gene expression. *Mol. Cell. Biol.* **30**, 5071–5085 (2010).

882 21. Capotosti, F. *et al.* O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1.
883 *Cell* **144**, 376–388 (2011).

884 22. Wilson, A. C., LaMarco, K., Peterson, M. G. & Herr, W. The VP16 accessory protein
885 HCF is a family of polypeptides processed from a large precursor protein. *Cell* **74**, 115–125
886 (1993).

887 23. Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N. & Herr, W. Human Sin3
888 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered
889 together selectively by the cell-proliferation factor HCF-1. *Genes Dev.* **17**, 896–911 (2003).

890 24. Shilatifard, A. The COMPASS family of histone H3K4 methylases: mechanisms of
891 regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* **81**, 65–95 (2012).

892 25. Wang, H. *et al.* H3K4me3 regulates RNA polymerase II promoter-proximal pause-
893 release. *Nature* **615**, 339–348 (2023).

894 26. van Nuland, R. *et al.* Quantitative Dissection and Stoichiometry Determination of the
895 Human SET1/MLL Histone Methyltransferase Complexes. *Mol. Cell. Biol.* **33**, 2067–2077
896 (2013).

897 27. Bowness, J. S., Almeida, M., Nesterova, T. B. & Brockdorff, N. YY1 binding is a gene-
898 intrinsic barrier to Xist-mediated gene silencing. *EMBO Rep.* **25**, 2258–2277 (2024).

899 28. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome.
900 *Nature* **515**, 355–364 (2014).

901 29. Michaud, J. *et al.* HCFC1 is a common component of active human CpG-island
902 promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor
903 occupancy. *Genome Res.* **23**, 907–916 (2013).

904 30. Cai, Y. *et al.* YY1 functions with INO80 to activate transcription. *Nat. Struct. Mol. Biol.*
905 **14**, 872–874 (2007).

906 31. Beckedorff, F. *et al.* The Human Integrator Complex Facilitates Transcriptional
907 Elongation by Endonucleolytic Cleavage of Nascent Transcripts. *Cell Rep.* **32**, 107917 (2020).

908 32. Mendoza-Figueroa, M. S., Tatomer, D. C. & Wilusz, J. E. The Integrator Complex in
909 Transcription and Development. *Trends Biochem. Sci.* **45**, 923–934 (2020).

910 33. Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*
911 **171**, 1573-1588.e28 (2017).

912 34. Minocha, S. *et al.* Epiblast-specific loss of HCF-1 leads to failure in anterior-posterior
913 axis specification. *Dev. Biol.* **418**, 75–88 (2016).

914 35. Cenik, B. K. & Shilatifard, A. COMPASS and SWI/SNF complexes in development and
915 disease. *Nat. Rev. Genet.* **22**, 38–58 (2021).

916 36. Mendjan, S. *et al.* Nuclear pore components are involved in the transcriptional regulation
917 of dosage compensation in Drosophila. *Mol. Cell* **21**, 811–823 (2006).

918 37. Cai, Y. *et al.* Subunit composition and substrate specificity of a MOF-containing histone
919 acetyltransferase distinct from the male-specific lethal (MSL) complex. *J. Biol. Chem.* **285**,
920 4268–4272 (2010).

921 38. Loda, A. *et al.* Genetic and epigenetic features direct differential efficiency of Xist-
922 mediated silencing at X-chromosomal and autosomal locations. *Nat. Commun.* **8**, 690 (2017).

923 39. Robert-Finestra, T. *et al.* SPEN is required for Xist upregulation during initiation of X
924 chromosome inactivation. *Nat. Commun.* **12**, 7000 (2021).

925 40. Krueger, F. & Andrews, S. R. SNPsplit: Allele-specific splitting of alignments between
926 genomes with known SNP genotypes. *F1000Research* **5**, 1479 (2016).

927 41. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**,
928 308–311 (2001).

929 42. Krueger, F., James, F., Ewels, P., Afyounian, E. & Schuster-Boeckler, B.
930 FelixKrueger/TrimGalore: v0.6.7 - DOI via Zenodo. Zenodo
931 https://doi.org/10.5281/zenodo.5127899 (2021).

932 43. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome
933 alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915
934 (2019).

935 44. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.*
936 **25**, 2078–2079 (2009).

937 45. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
938 assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* **30**, 923–930 (2014).

939 46. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
940 for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

941 47. Stephens, M. False discovery rates: a new deal. *Biostat. Oxf. Engl.* **18**, 275–294 (2017).

942 48. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
943 **9**, 357–359 (2012).

944 49. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible
945 platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187-191 (2014).

946 50. Jonkers, I. *et al.* Xist RNA is confined to the nuclear territory of the silenced X
947 chromosome throughout the cell cycle. *Mol. Cell. Biol.* **28**, 5583–5594 (2008).

948 51. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137
949 (2008).

950 52. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime
951 cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589
952 (2010).

953 53. Khan, A. & Mathelier, A. Intervene: a tool for intersection and visualization of multiple
954 gene or genomic region sets. *BMC Bioinformatics* **18**, 287 (2017).

955 54. Jin, H. *et al.* ChIPseqSpikeInFree: a ChIP-seq normalization approach to reveal global
956 changes in histone modifications without spike-in. *Bioinforma. Oxf. Engl.* **36**, 1270–1272 (2020).

957 55. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
958 genomic features. *Bioinforma. Oxf. Engl.* **26**, 841–842 (2010).

959 56. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer
960 (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**,
961 178–192 (2013).

962 57. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass
963 spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).

964

965 Acknowledgements

966 We thank Mariana Pelicano de Almeida and Debbie van den Berg for the INTS11 ChIP protocol
967 and feedback. We also thank Stanley van Herk for access to the Bioruptor Pico. Finally, we thank
968 Danilo Remmers and Koert van den Bergh for technical assistance.

969

970

971 Funding

972 This study was supported by a ZonMW-TOP subsidy (nr 91215046) to JG.

973

Author Information

Author contribution

J.B., B.T., S.M., C.G., H.M.-B. and J.G. conceptualised and designed experiments. J.B, S.M., E.R., R.v.H., S.J.L.-M., C.M., R.R., B.G., C.G. and H.M.-B. performed experiments and analysed data. B.T. performed all bioinformatic analyses. J.D. and W.v.I performed the mass spectrometry and sequencing, respectively. J.B., B.T., C.G., H.M.-B. and J.G wrote the manuscript and all authors edited it. J.B., S.M., C.M. and B.G. generated all cell lines. C.G. performed the pulldowns. J.B., S.M. and R.v.H. genotyped cell lines. J.B., S.J.L.-M. and H.M.-B. performed ChIP-seq experiments. J.B., E.R. and R.v.H. performed WB and RNA-FISH experiments. C.G., H.M.-B. and J.G. jointly supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

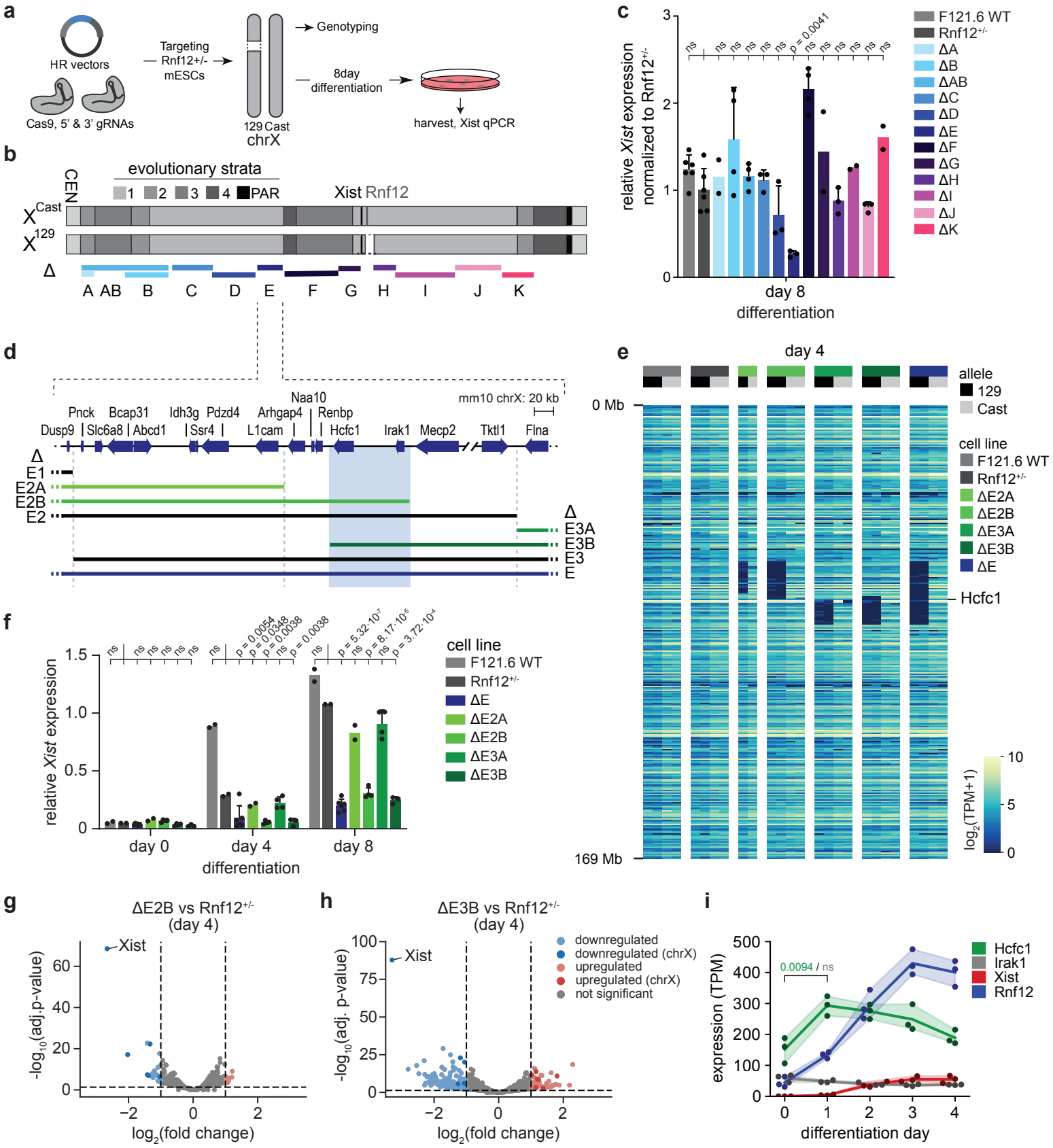
Supplementary information

The online version contains supplementary material available at XXX.

Materials and correspondence

All requests for materials and related questions should be addressed to Hegias Mira-Bontenbal and Joost Gribnau.

Figure 1

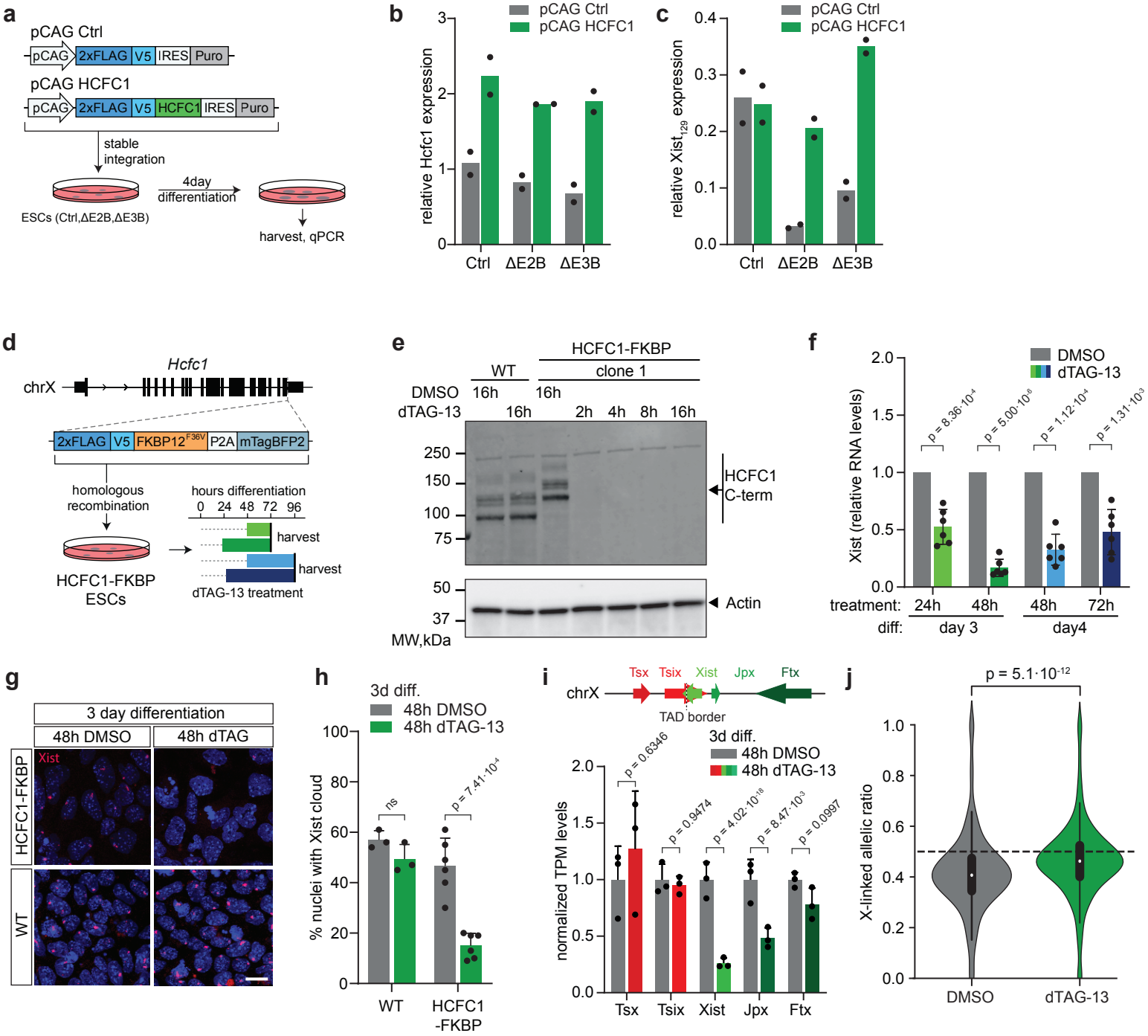

a
HR vectors
Cas9, 5' & 3' gRNAs
Targeting Rnf12+/- mESCs
129 Cast
chrX
Genotyping
8day differentiation
harvest, Xist qPCR
b
evolutionary strata
1 2 3 4 PAR
CEN
Xist Rnf12
Cast
129
Δ
A AB B C D E F G H I J K
c
relative Xist expression normalized to Rnf12+/-
day 8 differentiation
F121.6 WT
Rnf12+/-
ΔA ΔB ΔAB ΔC ΔD ΔE ΔF ΔG ΔH ΔI ΔJ ΔK
d
mm10 chrX: 20 kb
Dusp9 Pnck Slc6a8 Bcap31 Abcd1 Idh3g Ssr4 Pdzd4 Naa10 Arhgap4 L1cam Renbp Hcfc1 Irak1 Mecp2 Tktl1 Flna
Δ E1 E2A E2B E2
Δ E3A E3B E3 E
e
day 4
allele
129
Cast
cell line
F121.6 WT
Rnf12+/-
ΔE2A
ΔE2B
ΔE3A
ΔE3B
ΔE
Hcfc1
0 Mb
169 Mb
log2(TPM+1)
f
relative Xist expression
cell line
F121.6 WT
Rnf12+/-
ΔE
ΔE2A
ΔE2B
ΔE3A
ΔE3B
day 0
day 4
day 8
differentiation
g
ΔE2B vs Rnf12+/- (day 4)
Xist
-log10(adj.p-value)
log2(fold change)
h
ΔE3B vs Rnf12+/- (day 4)
Xist
downregulated
downregulated (chrX)
upregulated
upregulated (chrX)
not significant
-log10(adj. p-value)
log2(fold change)
i
expression (TPM)
0.0094 / ns
Hcfc1
Irak1
Xist
Rnf12
differentiation day

993 ## Legends

994 Figure 1: Large mega-domain deletions on the X chromosome identify HCFC1 as a potential XCI
995 activator.

996 **a**, Targeting and differentiation strategies. *Rnf12*$^{+/-}$ cells were transiently transfected with
997 CRISRPR-Cas9 and gRNAs targeting different locations on the X chromosomes. Correctly
998 targeted clones were differentiated for several days. **b**, X chromosome scheme depicting
999 evolutionary strata 1, 2, 3, 4 and the pseudo-autosomal region (PAR)[17,18]. The locations of *Xist* and
1000 *Rnf12* are depicted with a black and grey line, respectively, while the deleted *Rnf12* allele on the
1001 129 chromosome is also shown. The different deletions (ΔA to ΔK) are shown with different colors
1002 underneath. **c**, Relative *Xist* expression at day 8 of differentiation of WT, *Rnf12*$^{+/-}$ and the different
1003 *Rnf12*$^{+/-}$ cell lines with X chromosomal deletions, normalized to *Rnf12*$^{+/-}$. Average ± SD, n=2 to 5
1004 biological replicates. A one-sided Welch's t-test was used to test significant differences between
1005 *Rnf12*$^{+/-}$ and WT or deletion cell lines. P-values were corrected for multiple testing with the
1006 Benjamini-Hochberg method. **d**, Zoom-in on a specific region of ΔE. The locations of different
1007 overlapping deletions (ΔE1, ΔE2, ΔE3) are shown with black lines; the smaller incremental
1008 deletions (ΔE2A, ΔE2B, ΔE3A, ΔE3B) are shown in green. The blue rectangle highlights the
1009 location of *Hcfc1* and *Irak1*. **e**, Expression heatmap of X-linked genes, sorted by genomic location
1010 and split by allele (129 or Cast) for 2 WT, 2 *Rnf12*$^{+/-}$, 1 ΔE2A, 2 ΔE2B, 2 ΔE3A, 2 ΔE3B and 2
1011 ΔE clones. Notice the deleted genes per clone on the 129 alleles. **f**, Relative *Xist* expression at day
1012 0 and day 8 of WT, *Rnf12*$^{+/-}$, ΔE and smaller ΔE deletion cell lines. Average expression ± SD; n=2
1013 (WT, *Rnf12*$^{+/-}$ and ΔE2A lines), n=4 (ΔE2B, ΔE3A and ΔE3B lines) and n=5 (ΔE lines) biological
1014 replicates. A one-sided Welch's t-test was used to test significant differences per day of
1015 differentiation between *Rnf12*$^{+/-}$ and WT or deletions cell lines. P-values were corrected for
1016 multiple testing with the Benjamini-Hochberg method. **g**, Volcano plot portraying gene expression
1017 changes between ΔE2B and *Rnf12*$^{+/-}$ cells. Downregulated and upregulated genes in ΔE2B are
1018 shown in blue and red, respectively. Notice *Xist* is depicted on the left, as the most significant
1019 downregulated gene. **h**, Same as **g** but for ΔE3B vs *Rnf12*$^{+/-}$ cells. **i**, Expression (TPM) of *Hcfc1*
1020 (green), *Irak1* (grey), *Rnf12* (blue) and *Xist* (red) in ΔXic_{Cast} ESCs from Pacini et al., 2021[19]. P-
1021 values were calculated between day 0 and 1 for *Hcfc1* (green) and *Irak1* (grey) using one-sided
1022 independent t-tests and corrected for multiple testing with the Benjamini-Hochberg method.

Figure 2

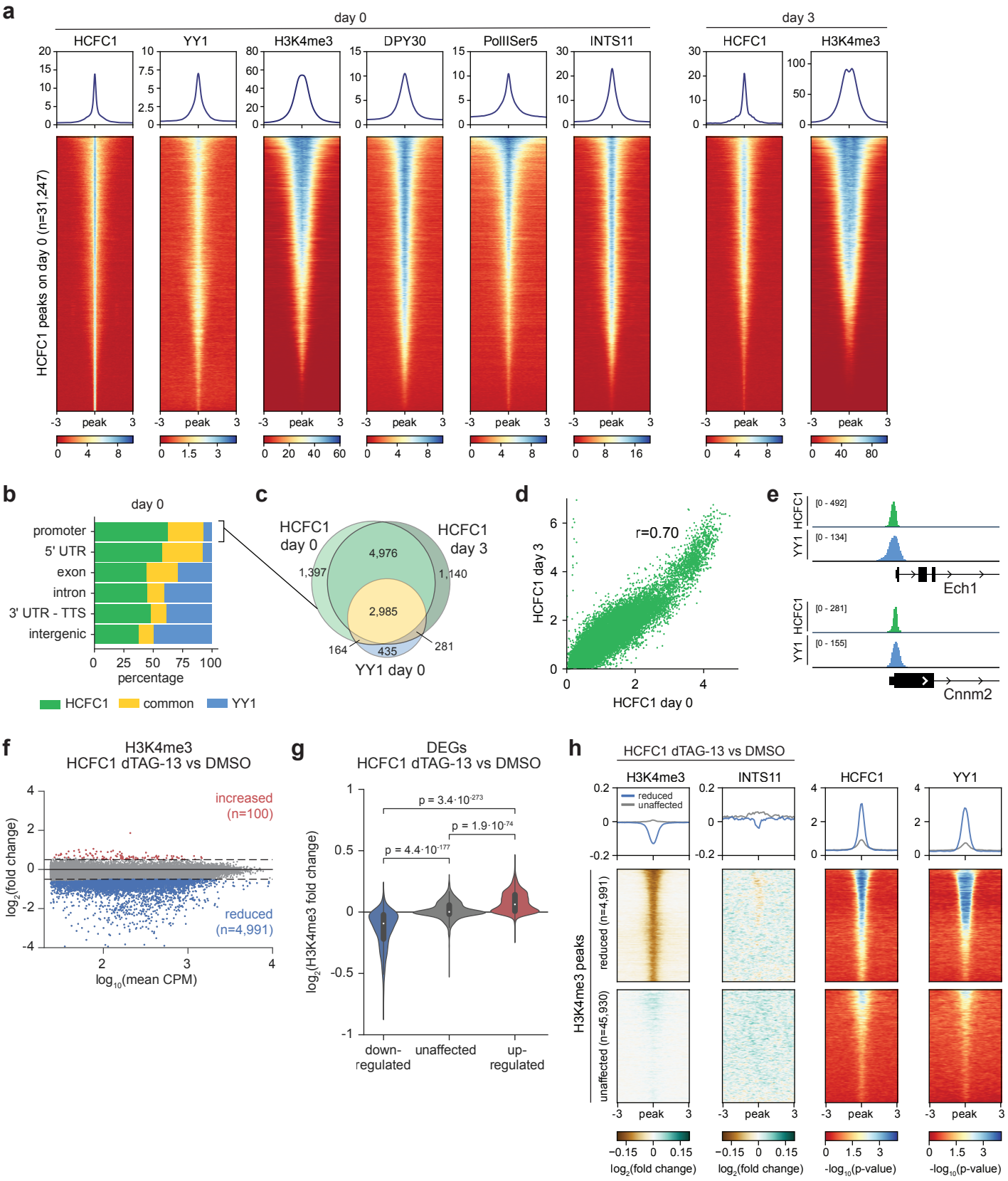

a
pCAG Ctrl
pCAG
2xFLAG
V5
IRES
Puro
pCAG HCFC1
HCFC1
stable integration
4day differentiation
ESCs (Ctrl,ΔE2B,ΔE3B)
harvest, qPCR
b
pCAG Ctrl
pCAG HCFC1
relative Hcfc1 expression
Ctrl
ΔE2B
ΔE3B
c
relative Xist129 expression
d
Hcfc1
chrX
FKBP12F36V
P2A
mTagBFP2
homologous recombination
hours differentiation
0 24 48 72 96
harvest
HCFC1-FKBP ESCs
dTAG-13 treatment
e
HCFC1-FKBP
WT
clone 1
DMSO
dTAG-13
16h
2h 4h 8h 16h
250
150
100
75
50
37
MW,kDa
HCFC1 C-term
Actin
f
DMSO
dTAG-13
Xist (relative RNA levels)
p = 8.36·10-4
p = 5.00·10-6
p = 1.12·10-4
p = 1.31·10-3
treatment: 24h 48h 48h 72h
diff: day 3 day4
g
3 day differentiation
48h DMSO
48h dTAG
HCFC1-FKBP
WT
Xist
h
3d diff.
48h DMSO
48h dTAG-13
% nuclei with Xist cloud
ns
p = 7.41·10-4
WT
HCFC1 -FKBP
i
Tsx
Tsix
Xist
Jpx
Ftx
chrX
TAD border
normalized TPM levels
p = 0.6346
p = 0.9474
p = 4.02·10-18
p = 8.47·10-3
p = 0.0997
j
p = 5.1·10-12
X-linked allelic ratio
DMSO
dTAG-13

1023 Figure 2: HCFC1 is an XCI activator.

1024 **a**, HCFC1 rescue and differentiation strategy of control (Ctrl; parental Rnf12$^{+/-}$), ΔE2B and ΔE3B
1025 ESCs. **b**, RT-qPCR analysis of *Hcfc1* expression upon *Hcfc1* rescue in pools of transfected cells
1026 after 4 days of differentiation of in Ctrl (parental *Rnf12*$^{+/-}$), ΔE2B and ΔE3B. The average of n=2
1027 biological replicates is shown. **c**, RT-qPCR analysis of *Xist* expression in pools of transfected cells
1028 after 4 days of differentiation. The average of n=2 biological replicates is shown. **d**, 2xFLAG-V5-
1029 FKBP12^{F36V} tagging strategy of HCFC1 and differentiation schemes with different dTAG-13
1030 treatment windows. **e**, Western blot of the C-terminal fragments of HCFC1 upon dTAG-13
1031 treatment in undifferentiated ESCs. **f**, RT-qPCR analysis of *Xist* expression at different timepoints
1032 after start of dTAG-13 treatment and differentiation normalised to DMSO control; see Fig. 2a. A
1033 two-sided Welch's t-test was used to assess significant differences between DMSO- and dTAG-13
1034 treated cells. P-values were corrected for multiple testing using the Benjamini-Hochberg method.
1035 **g**, *Xist* RNA-FISH analysis in HCFC1-FKBP ESCs differentiated for 3 days and treated with
1036 dTAG-13 for the last 48h vs DMSO and parental ESCs. **h**, Quantification of **g**. Average ± SD,
1037 n=3-6 biological replicates, n=157-236. A two-sided Welch's t-test was used to assess significant
1038 differences between DMSO- and dTAG-13 treated cells. P-values were corrected for multiple
1039 testing using the Benjamini-Hochberg method. **i**, Gene expression levels of several genes located
1040 in the XIC determined by RNA-seq analysis of HCFC1-FKBP ESCs differentiated for 3 days and
1041 treated with dTAG-13 for the last 48h. Shown in green are genes residing in the *Xist* TAD and
1042 involved in its activation. In red, genes located in the *Tsix* TAD and involved in *Xist* repression.
1043 Average TPMs normalized to 48h DMSO control ± SD, n= 3 biological replicates. P-values were
1044 calculated using DESeq2, which corrects for multiple testing using the Benjamini-Hochberg
1045 method. **j**, Violin plot of allelic ratios of X-linked gene expression in the DMSO and dTAG-13
1046 conditions. The box plots display the median (black line), the interquartile range (box limits) and
1047 1.5x of the interquartile range (whiskers). Dashed line indicates ratio of 0.5 as expected ratio
1048 without silencing. P-value was calculated using a two-sided Mann-Whitney U test.

1049

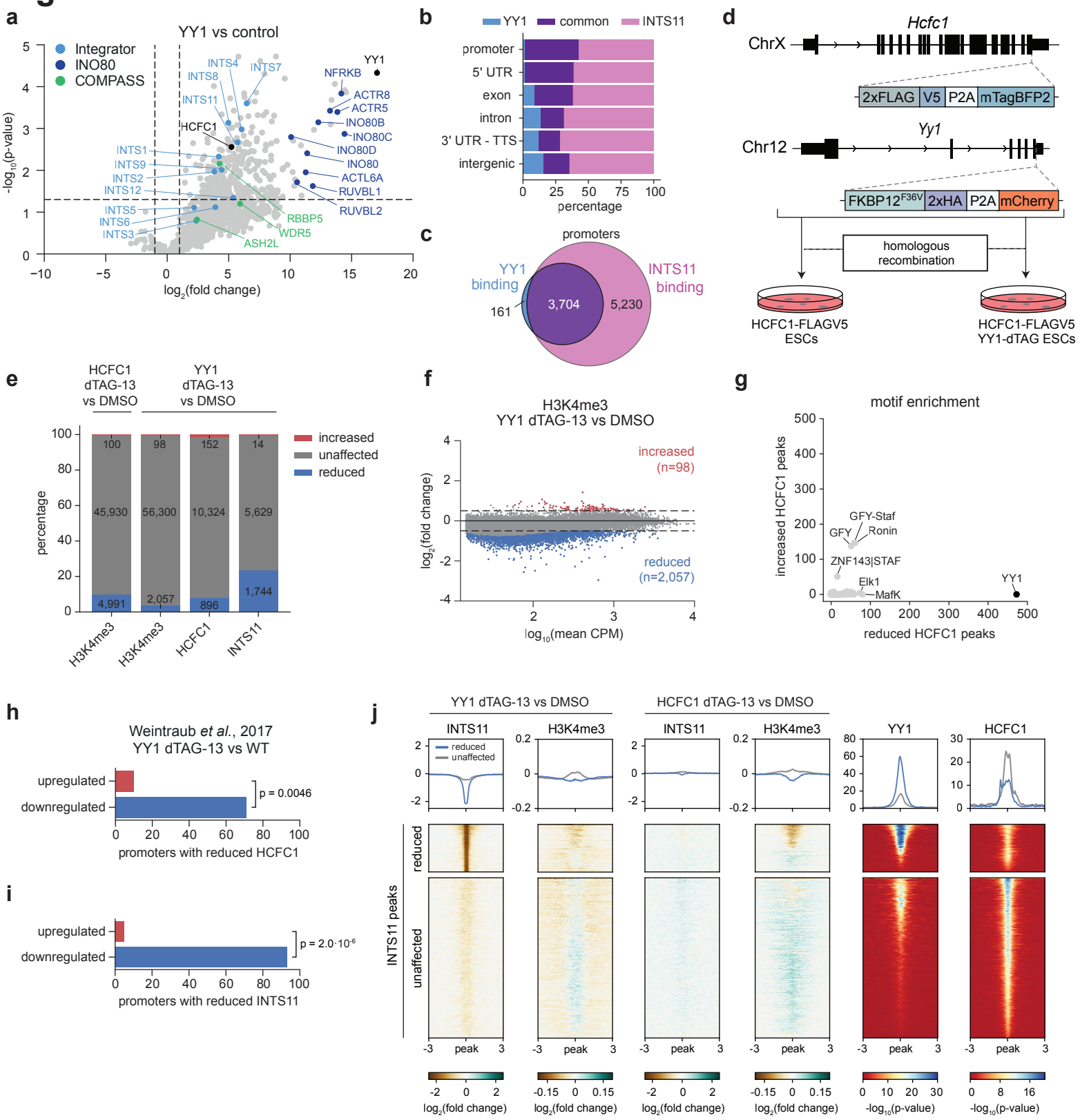
Figure 3


a
day 0
day 3
HCFC1
YY1
H3K4me3
DPY30
PolIISer5
INTS11
HCFC1 peaks on day 0 (n=31,247)
-3 peak 3
b
day 0
promoter
5' UTR
exon
intron
3' UTR - TTS
intergenic
percentage
HCFC1
common
YY1
c
HCFC1 day 0
HCFC1 day 3
4,976
1,397
1,140
2,985
164
435
281
YY1 day 0
d
r=0.70
HCFC1 day 3
HCFC1 day 0
e
HCFC1
YY1
[0 - 492]
[0 - 134]
Ech1
[0 - 281]
[0 - 155]
Cnnm2
f
H3K4me3
HCFC1 dTAG-13 vs DMSO
increased (n=100)
reduced (n=4,991)
log2(fold change)
log10(mean CPM)
g
DEGs
HCFC1 dTAG-13 vs DMSO
p = 3.4·10-273
p = 1.9·10-74
p = 4.4·10-177
log2(H3K4me3 fold change)
down-regulated
unaffected
up-regulated
h
HCFC1 dTAG-13 vs DMSO
H3K4me3
INTS11
HCFC1
YY1
reduced
unaffected
H3K4me3 peaks
reduced (n=4,991)
unaffected (n=45,930)
log2(fold change)
-log10(p-value)

1050 Figure 3: HCFC1 binding correlates with promoters, H3K4me3 deposition and YY1 binding.
1051 **a**, Heatmap of HCFC1, YY1, H3K4me3, DPY30, PolIISer5 and INTS11 enrichment at HCFC1
1052 peaks in undifferentiated ESCs (left). The right panels show HCFC1 and H3K4me3 enrichment at
1053 day 3 of differentiation at HCFC1 peaks identified at day 0. **b**, HCFC1-only and YY1-only and
1054 shared HCFC1-YY1 peak distribution at different genomic locations at day 0. **c**, Venn diagram
1055 with the overlap of HCFC1 and YY1 peaks in promoter regions at day 0 and 3 of differentiation.
1056 **d**, Correlation between HCFC1 binding at day 0 and day 3 in promoter regions, with r representing
1057 the Pearson correlation coefficient. **e**, Genome browser views showing examples of HCFC1-YY1
1058 overlapping peaks in promoter regions. **f**, MA plot depicting H3K4me3 enrichment changes upon
1059 dTAG-13-mediated removal of HCFC1. In red, peaks with increased H3K4me3 enrichment; in
1060 grey, unaffected peaks and in blue, peaks with reduced H3K4me3 enrichment. **g**, Violin plots
1061 showing H3K4me3 changes in promoter regions of differentially expressed genes from the RNA-
1062 seq comparison between HCFC1-FKBP dTAG-13 and DMSO. P-values were calculated using
1063 two-sided Mann-Whitney U tests, corrected for multiple testing using the Benjamini-Hochberg
1064 approach. **h**, Heatmaps showing differences in H3K4me3 and INTS11 enrichment at H3K4me3
1065 peaks that are lost upon dTAG-13-mediated removal of HCFC1 and unaffected H3K4me3 peaks.
1066 For both H3K4me3 and INTS11, the fold change between dTAG-13 and DMSO treatment is
1067 shown. HCFC1 binding (day 3) and YY1 binding (day 0) are displayed on the right. For the
1068 unaffected peaks, a random subset of 4,991 peaks is shown.

1069

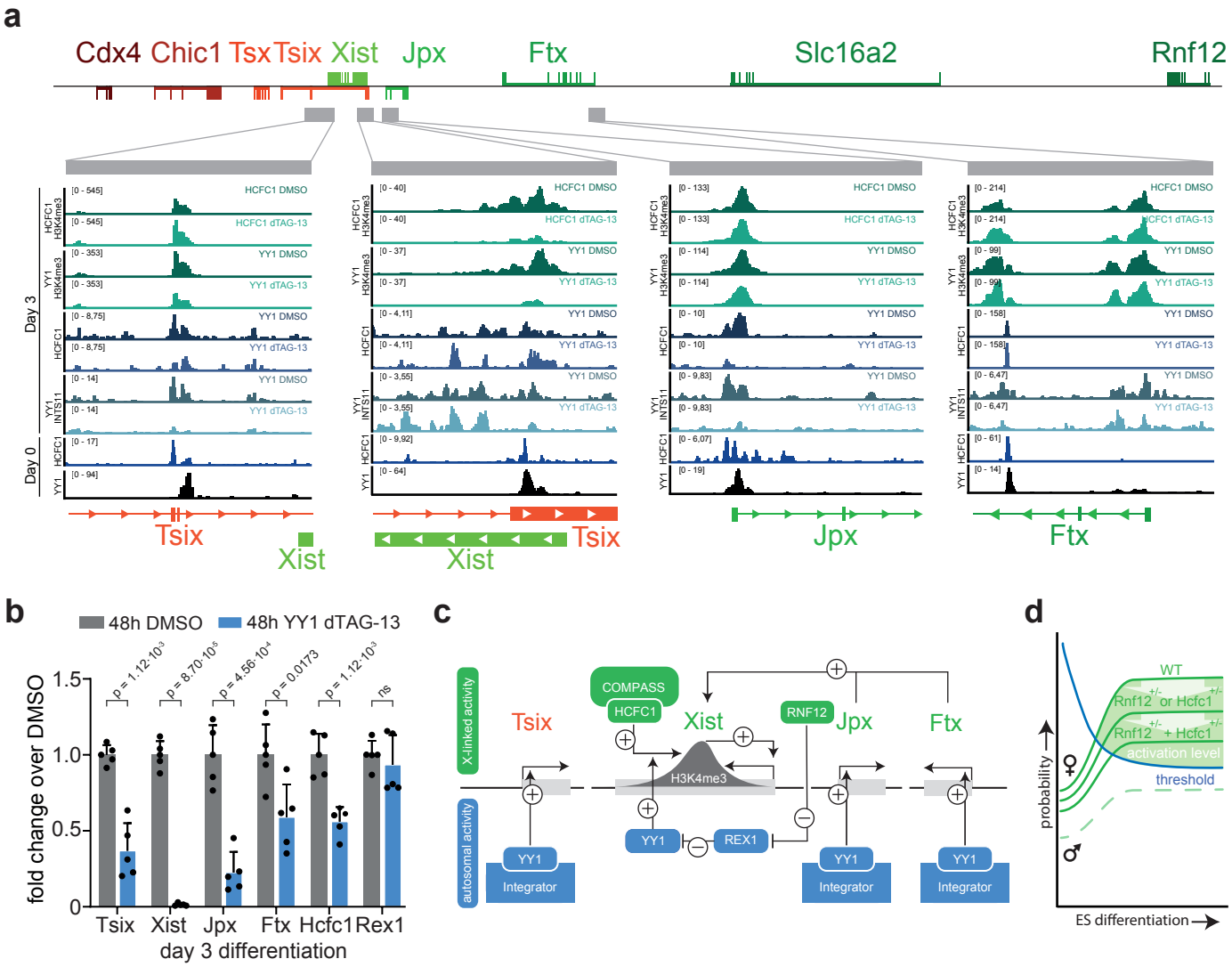
Figure 4


a
YY1 vs control
Integrator
INO80
COMPASS
-log10(p-value)
log2(fold change)
YY1
NFRKB
ACTR8
ACTR5
INO80B
INO80C
INO80D
INO80
ACTL6A
RUVBL1
RUVBL2
RBBP5
WDR5
ASH2L
INTS4
INTS7
INTS8
INTS11
HCFC1
INTS1
INTS9
INTS2
INTS12
INTS5
INTS6
INTS3
b
YY1
common
INTS11
promoter
5' UTR
exon
intron
3' UTR - TTS
intergenic
percentage
c
promoters
YY1 binding
INTS11 binding
161
3,704
5,230
d
Hcfc1
ChrX
2xFLAG
V5
P2A
mTagBFP2
Yy1
Chr12
FKBP12F36V
2xHA
P2A
mCherry
homologous recombination
HCFC1-FLAGV5 ESCs
HCFC1-FLAGV5 YY1-dTAG ESCs
e
HCFC1 dTAG-13 vs DMSO
YY1 dTAG-13 vs DMSO
increased
unaffected
reduced
percentage
100
98
152
14
45,930
56,300
10,324
5,629
4,991
2,057
896
1,744
H3K4me3
H3K4me3
HCFC1
INTS11
f
H3K4me3
YY1 dTAG-13 vs DMSO
increased (n=98)
reduced (n=2,057)
log2(fold change)
log10(mean CPM)
g
motif enrichment
increased HCFC1 peaks
reduced HCFC1 peaks
GFY-Staf
GFY
Ronin
ZNF143|STAF
Elk1
MafK
YY1
h
Weintraub et al., 2017
YY1 dTAG-13 vs WT
upregulated
downregulated
p = 0.0046
promoters with reduced HCFC1
i
upregulated
downregulated
p = 2.0·10-6
promoters with reduced INTS11
j
YY1 dTAG-13 vs DMSO
HCFC1 dTAG-13 vs DMSO
INTS11
H3K4me3
INTS11
H3K4me3
YY1
HCFC1
reduced
unaffected
INTS11 peaks
reduced
unaffected
peak
log2(fold change)
-log10(p-value)

1070 Figure 4: YY1 interacts with the Integrator complex.

1071 **a**, YY1 interactome identified by FLAG pull-down and mass spectrometry. Volcano plot showing
1072 proteins enriched in YY1 pull-downs compared to control. Known members of the Integrator (light
1073 blue), INO80 (dark blue), and COMPASS (green) complexes are highlighted. Selected proteins,
1074 including YY1 itself and HCFC1, are labeled in black. Dashed lines denote thresholds for
1075 significance ($p < 0.05$, $\log_2$ fold change > 1). Data are based on two biological replicates for FLAG-
1076 YY1 and four biological replicates for the control. **b**, Percentages of shared YY1-INTS11, YY1-
1077 only or INTS11-only peaks at different genomic locations at day 0. **c**, Venn diagram showing the
1078 overlap of YY1 and INTS11 binding in promoter regions. **d**, Generation strategy of the
1079 FKBP12^{F36V}-HA-Yy1:2xFLAG-V5-Hcfc1 compound knockin ESC lines by homologous
1080 recombination. **e**, Stacked barplot showing percentages of H3K4me3, HCFC1 and INTS11 peaks
1081 that significantly lost or gained enrichment or remained unchanged, upon HCFC1 or YY1 loss. **f,**
1082 MA plot depicting H3K4me3 enrichment changes upon dTAG-13-mediated removal of YY1. In
1083 red, peaks with increased H3K4me3 enrichment; in grey, unaffected peaks and in blue, peaks with
1084 reduced H3K4me3 enrichment. **g**, Scatter plot comparing motif enrichment between gained and
1085 lost HCFC1 peaks shown in **e**. Each point represents a motif, plotted by its $-\log_{10}$(p-value) in both
1086 analyses, and motifs with the lowest p-values in either condition are labeled. **h**, Barplot showing
1087 the number of up- and downregulated genes with reduced HCFC1 binding in the promoter region
1088 upon YY1 depletion. P-value was calculated using a Pearson's chi-squared (χ^2) goodness-of-fit
1089 test, based on the overall proportion of up- and downregulated genes. **i**, Same as **h**, but for promoter
1090 regions with reduced INTS11 binding. **j**, Heatmap showing changes in INTS11 and H3K4me3
1091 enrichment upon dTAG-13-mediated depletion of YY1 or HCFC1, alongside YY1 and HCFC1
1092 binding at INTS11 peaks. Peaks are grouped based on whether INTS11 enrichment is reduced
1093 (top) or unaffected (bottom) by YY1 depletion.

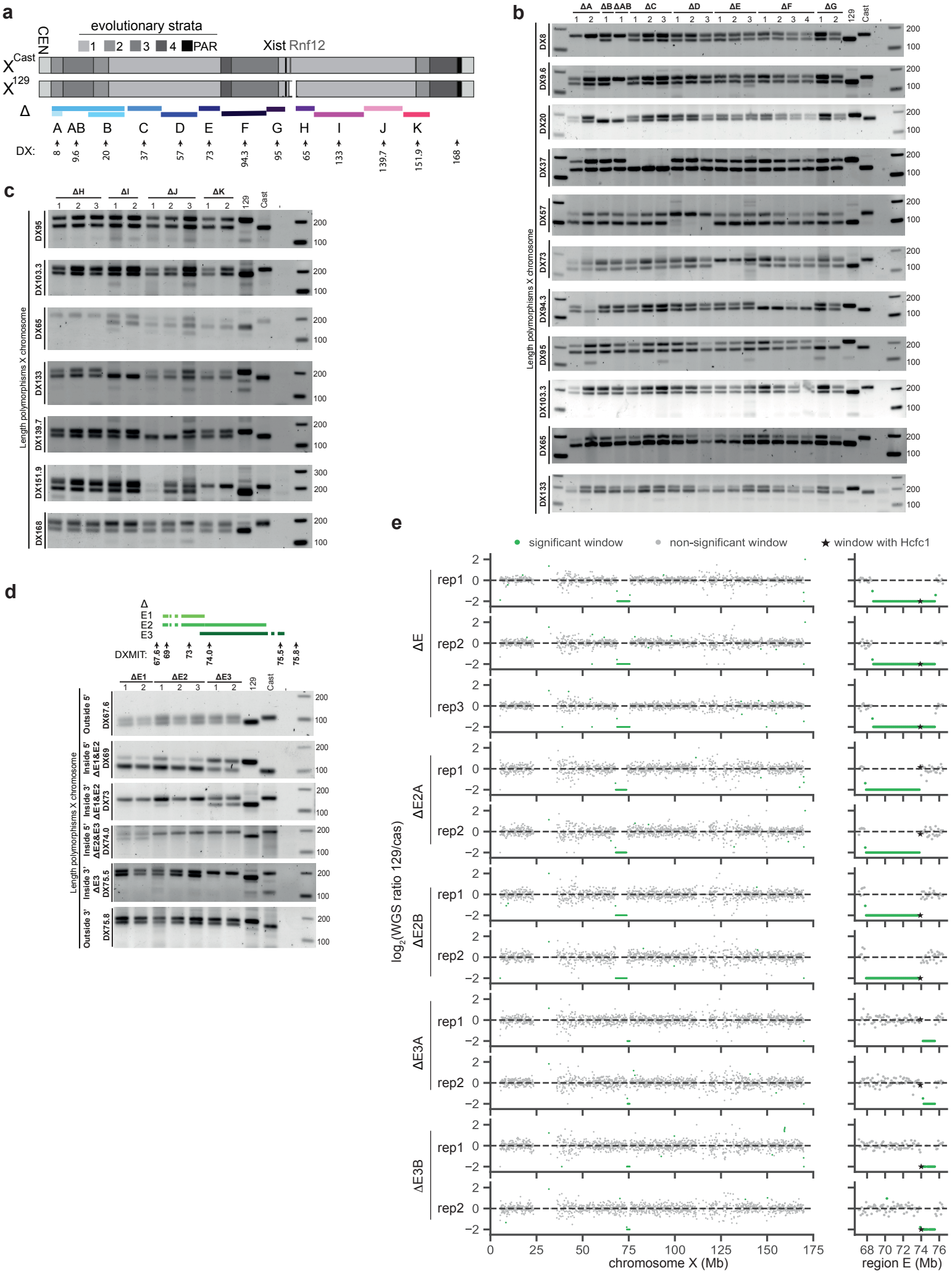
Figure 5


a
Cdx4 Chic1 Tsx Tsix Xist Jpx Ftx Slc16a2 Rnf12
Day 3
Day 0
HCFC1 DMSO
HCFC1 dTAG-13
YY1 DMSO
YY1 dTAG-13
Tsix
Xist
Jpx
Ftx
b
48h DMSO
48h YY1 dTAG-13
fold change over DMSO
p = 1.12·10^-3
p = 8.70·10^-6
p = 4.56·10^-4
p = 0.0173
p = 1.12·10^-3
ns
Tsix Xist Jpx Ftx Hcfc1 Rex1
day 3 differentiation
c
X-linked activity
autosomal activity
COMPASS
HCFC1
Tsix
Xist
RNF12
Jpx
Ftx
H3K4me3
YY1
REX1
Integrator
d
WT
Rnf12 +/- or Hcfc1 +/-
Rnf12 +/- + Hcfc1 +/-
activation level
threshold
probability
ES differentiation

1094 Figure 5: HCFC1 binds to and regulates the expression of several non-coding genes in the XIC.
1095 **a**, Genome browser views showing ChIP-seq tracks of H3K4me3 in HCFC1-depleted or YY1-
1096 depleted versus DMSO controls at day 3 of differentiation, HCFC1 ChIP-seq in YY1-depleted
1097 versus its DMSO control at day 3 of differentiation, INTS11 ChIP-seq in YY1-depleted versus its
1098 DMSO control at day 3, and YY1 and HCFC1 ChIP-seq in female ESCs at day 0 around the *Tsix*,
1099 *Xist*, *Jpx* and *Ftx* promoter regions. **b,** Relative expression of *Tsix, Xist*, *Jpx*, *Ftx, Hcfc1* and
1100 pluripotency factor *Rex1* at day 3 of differentiation of YY1-FKBP ESC cells normalised to DMSO
1101 control. A two-sided Welch's t-test was used to test significant differences between DMSO- and
1102 dTAG-13 treated cells. P-values were corrected for multiple testing using the Benjamini-Hochberg
1103 method. **c,** Gene-regulatory network controlling *Xist* expression. RNF12 promotes YY1 binding
1104 by targeting the *Xist* repressor REX1 for proteasomal degradation. HCFC1/COMPASS and YY1
1105 are both required for H3K4me3 catalysis at the *Xist* regulatory element. YY1 also recruits
1106 Integrator to activate transcription of *Xist* cis-regulatory genes within the XIC. **d**, Dose-dependent
1107 activation of *Xist* expression. Elevated levels of X-linked activators in XX cells raise the
1108 probability for *Xist* transcription initiation above threshold; this probability is reduced in *Rnf12*$^{+/-}$
1109 or *Hcfc*1$^{+/-}$ cells and further diminished in double-heterozygous *Rnf12*$^{+/-}$;*Hcfc*1$^{+/-}$ cells.

1110

Extended Data Figure 1



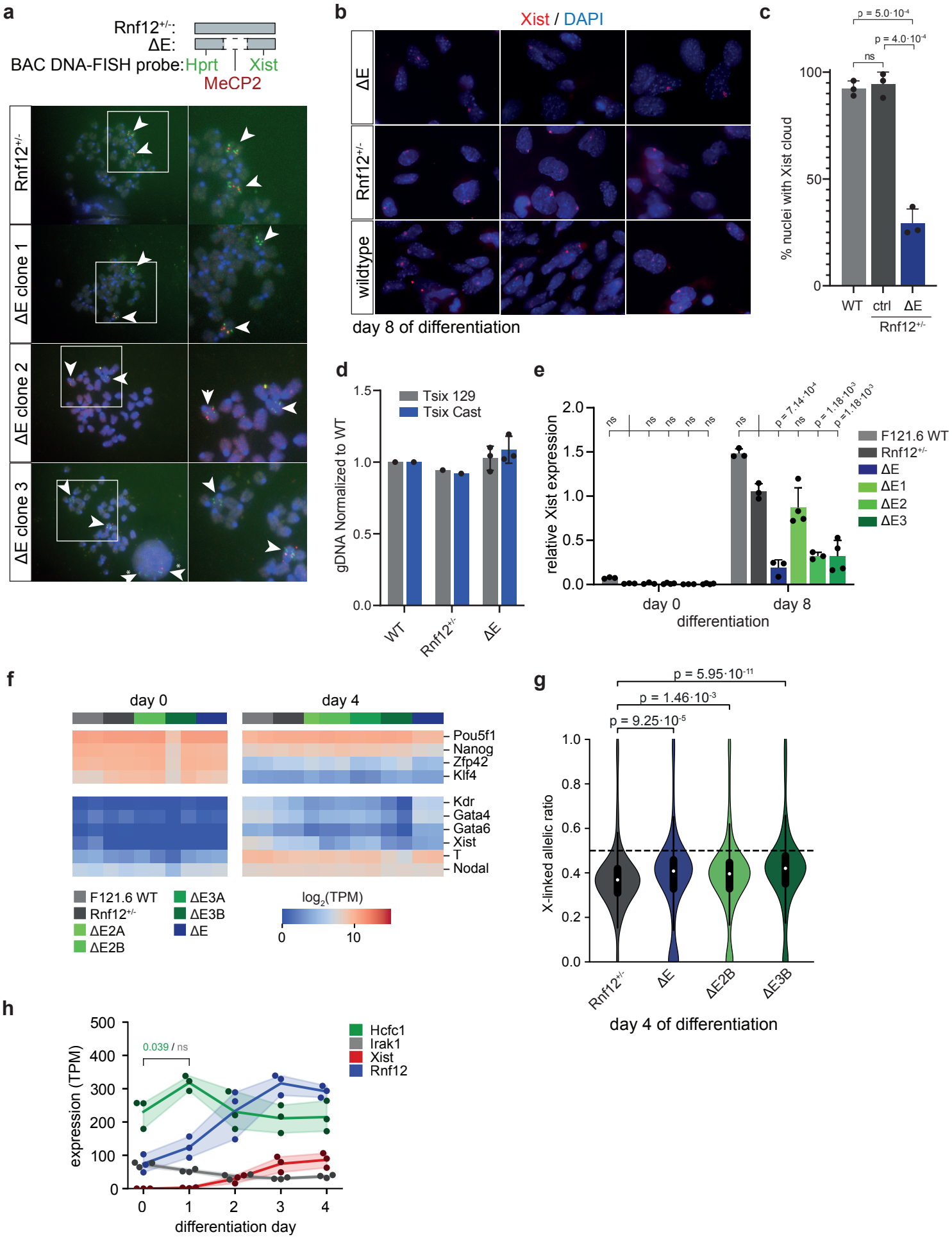
1111 Extended Data Figure 1: Megadomain deletion confirmation by PCR and WGS.

1112 **a**, Overview of the locations of the different deletions generated in this study. Different X-linked
1113 length polymorphisms between 129 and Cast are shown (DX amplicons). **b-c,** PCR analysis of
1114 different DX amplicons for the different deletions. **d**, Location of the different DX amplicons with
1115 respect to the deleted regions, and length polymorphism analysis inside and around several
1116 incremental deletions within region E. **e**, Scatter plots of allelic ratios along the X-chromosome
1117 from shallow whole-genome sequencing for different ΔE, ΔE2A, ΔE2B, ΔE3A and ΔE3B clones
1118 (left). Y-axis represents the $\log_2$(fold change) between the 129 and Cast alleles for each window.
1119 Windows with significant signal differences between alleles are highlighted in green. The zoomed-
1120 in plots on the right show the region around 67-77 Mb, with the *Hcfc1*-containing window
1121 indicated by a star.

1122

1123

Extended Data Figure 2



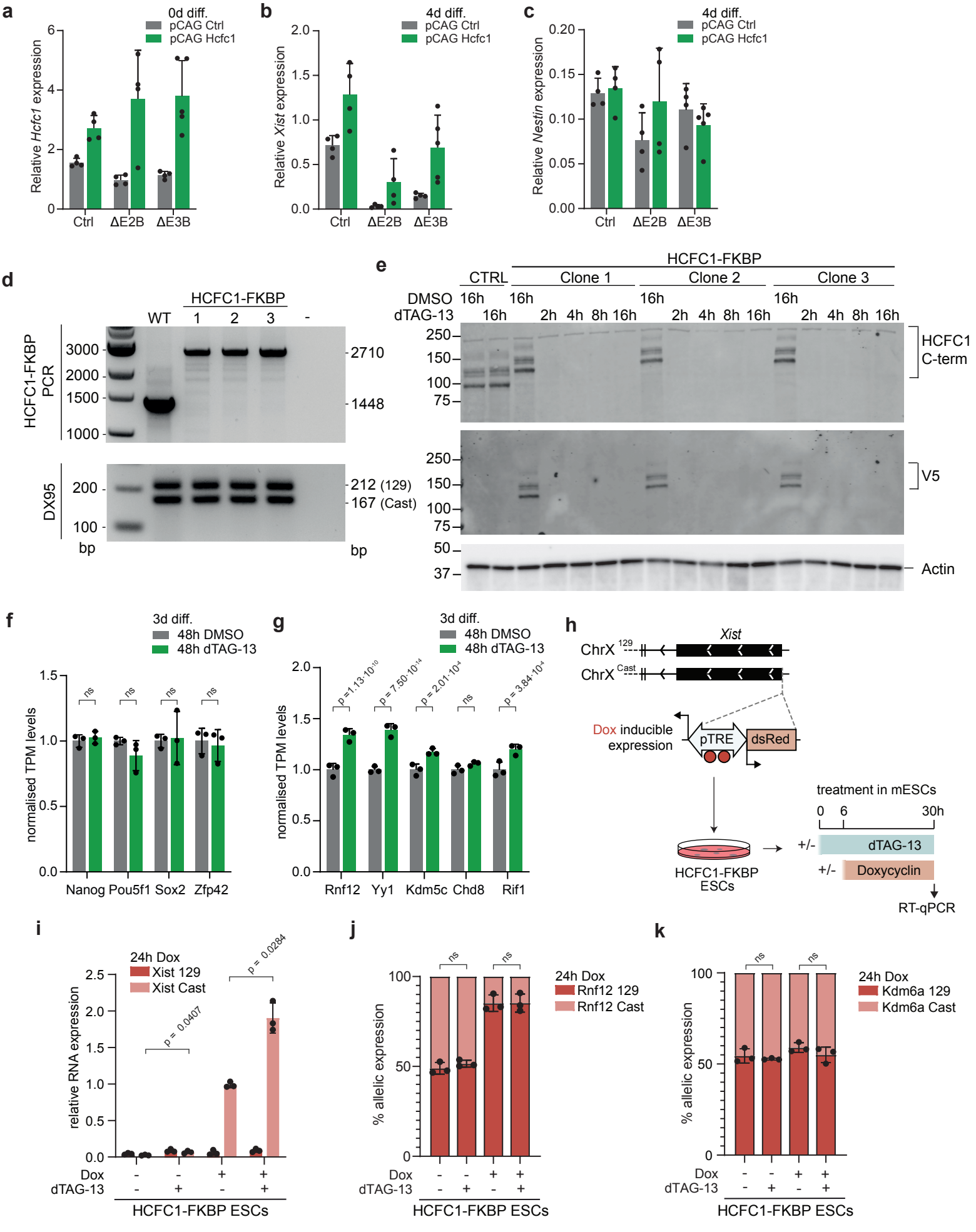
1124 Extended Data Figure 2: Deletions containing *Hcfc1* show reduced Xist upregulation but can
1125 differentiate.

1126 **a**, DNA FISH analysis on metaphase spreads of the parental and three ΔE clones detecting *Hprt*
1127 and *Xist* in green, and *Mecp2* in red. Notice red is inside the deleted region. White arrowheads
1128 indicate chromosome X. Arrowheads with an asterisk in ΔE clone 3 indicate the signal within an
1129 interphase nucleus. **b**, Representative images of *Xist* RNA FISH (red) in WT, *Rnf12*$^{+/-}$ and three
1130 ΔE clones at day 8 of differentiation. **c**, Quantification of **b**, average number of cells with a *Xist*
1131 cloud ± SD, n=3 biological replicas, n=201-231, statistical significance was calculated using a
1132 two-sided Welch T-test and corrected for multiple testing using the Benjamini-Hochberg
1133 procedure. P-values are indicated, ns = not significant. **d**, Allelic gDNA PCR of *Tsix* in WT,
1134 *Rnf12*$^{+/-}$ and 3 independent ΔE clones on day 8 of differentiation to confirm the presence of two X
1135 chromosomes. Average ± SD, n=1-3 biological replicates. **e**, Relative *Xist* expression in WT,
1136 *Rnf12*$^{+/-}$, ΔE, ΔE1, ΔE2 and ΔE3 ESCs at day 0 and 8 of differentiation. A one-sided Welch's t-
1137 test was used to test significant differences between *Rnf12*$^{+/-}$ and WT or deletion cell lines. P-
1138 values were corrected for multiple testing with the Benjamini-Hochberg method. **f**, Heatmap of
1139 gene expression (in TPMs) for pluripotency markers (*Pou5f1*, *Nanog*, *Zfp42*/*Rex1*, *Klf4*) and
1140 differentiation markers (*Kdr*/*Flk-1*, *Gata4*, *Gata6*, *T* and *Nodal*) across WT, *Rnf12*$^{+/-}$ and ΔE,
1141 ΔE2A, ΔE2B, ΔE3A and ΔE3B clones. **g**, Violin plot of allelic ratios of X-linked gene expression
1142 in *Rnf12*$^{+/-}$, ΔE, ΔE2B and ΔE3B. The box plots display the median (black line), the interquartile
1143 range (box limits) and 1.5x of the interquartile range (whiskers). Dashed line indicates ratio of 0.5
1144 as expected ratio without silencing. P-values were calculated using a two-sided Mann-Whitney U
1145 test and corrected for multiple testing using the Benjamini-Hochberg approach. **h**, Expression
1146 (TPM) of *Hcfc1* (green), *Irak1* (grey), *Rnf12* (blue) and *Xist* (red) in ΔXic_{B6} ESCs from Pacini et
1147 al., 2021[19]. P-values were calculated between day 0 and 1 for Hcfc1 (green) and Irak1 (grey) using
1148 one-sided independent t-tests and corrected for multiple testing with the Benjamini-Hochberg
1149 method.

1150

1151

Extended Data Figure 3



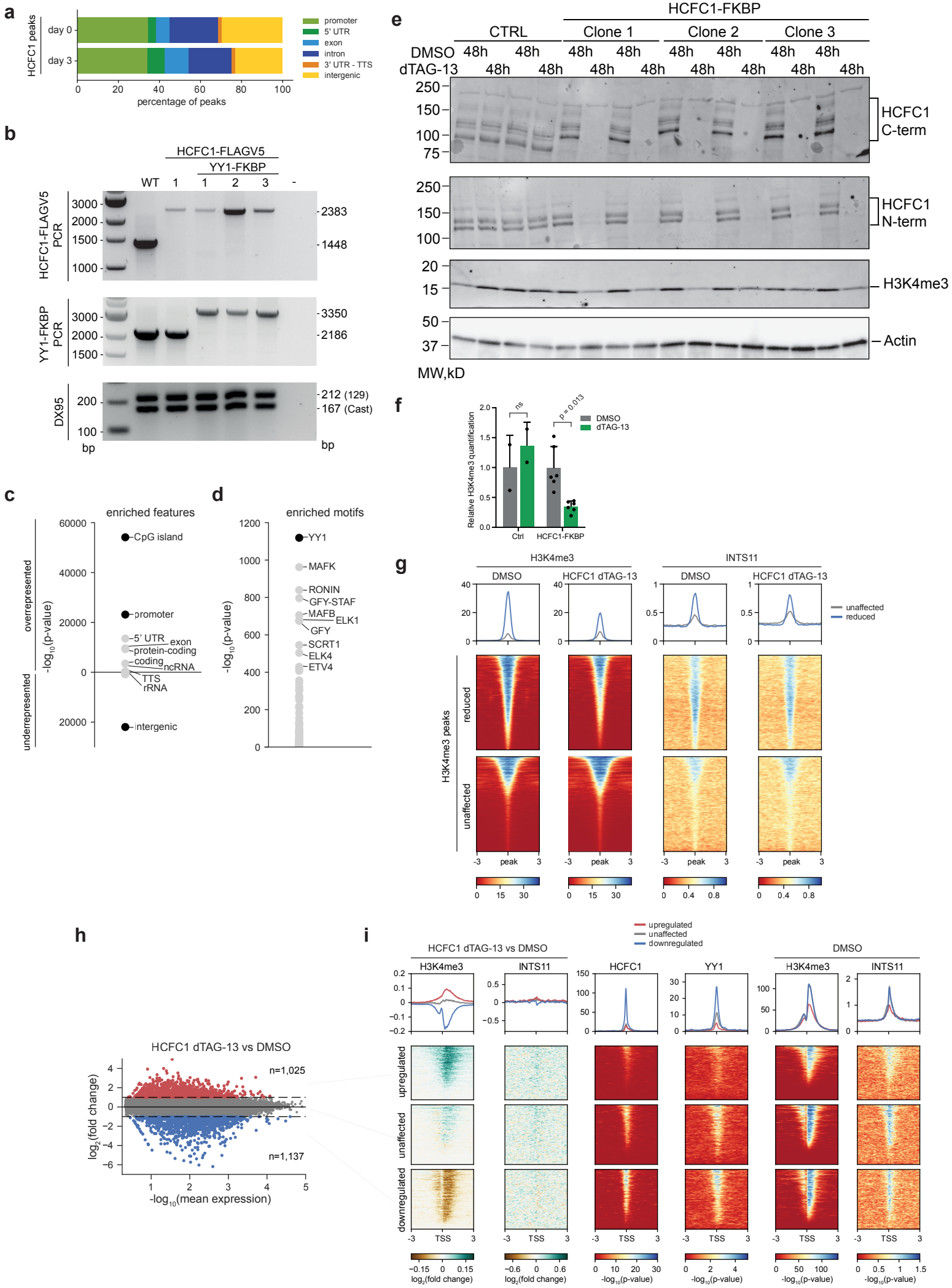
1152 Extended Data Figure 3: HCFC1 rescue and conditional knockout confirmation.

1153 **a**, Relative *Hcfc1* expression in control (*Rnf12*$^{+/-}$), ΔE2B and ΔE3B ESCs upon stable integration,
1154 *Hcfc1* expression vs control plasmid in different individual undifferentiated clones; Average ± SD,
1155 n=4-5 independent clones. **b**, Relative *Xist* expression in Control (*Rnf12*$^{+/-}$), ΔE2B and ΔE3B upon
1156 stable integration, *Hcfc1* expression vs control plasmid at day 4 of differentiation in different
1157 individual clones; Average ± SD, n=4-5 independent clones. **c**, Same as in **b**, but for differentiation
1158 marker *Nestin*. **d**, PCR analysis of HCFC1-FKBP ESCs, top. DX95 PCR to control for the presence
1159 of both X chromosomes, bottom. **e**, HCFC1 C-term and V5 WB in HCFC1-FKBP ESCs after
1160 treatment with dTAG-13 or DMSO control in the ESC state. Actin is used as loading control. **f**,
1161 Expression of several pluripotency markers at day 3 of differentiation after 48h of dTAG-13 or
1162 DMSO treatment. TPM values are normalized to the average expression in the DMSO condition
1163 for each gene. P-values were calculated using DESeq2, which corrects for multiple testing using
1164 the Benjamini-Hochberg method **g**, Same as in **f**, but for genes involved in *Xist* regulation. **h**,
1165 Schematic of HCFC1-FKBP ESCs with a doxycycline-inducible *Xist* promoter and strategy to
1166 induce *Xist* expression in the ESC stage in the presence of dTAG-13. **i**, RT-qPCR analysis of *Xist*
1167 expression in HCFC1-FKBP ESCs upon 24h of dox treatment and/or 30h dTAG-13 treatment.
1168 Average ± SD, n=3 biological replicates. **j**, Allelic expression of X-linked gene *Rnf12* in HCFC1-
1169 FKBP ESCs upon 24h Dox and/or 30h dTAG-13 treatment. Average ± SD, n=3 biological
1170 replicates. **k**, Allelic expression of XCI escapee *Kdm6a* upon forced *Xist* expression (+Dox) for
1171 24h in the presence of dTAG-13 or DMSO. For (i–k), p-values from two-sided Welch's t-test
1172 between DMSO and dTAG-13 treatment, corrected using the Benjamini–Hochberg procedure.

1173

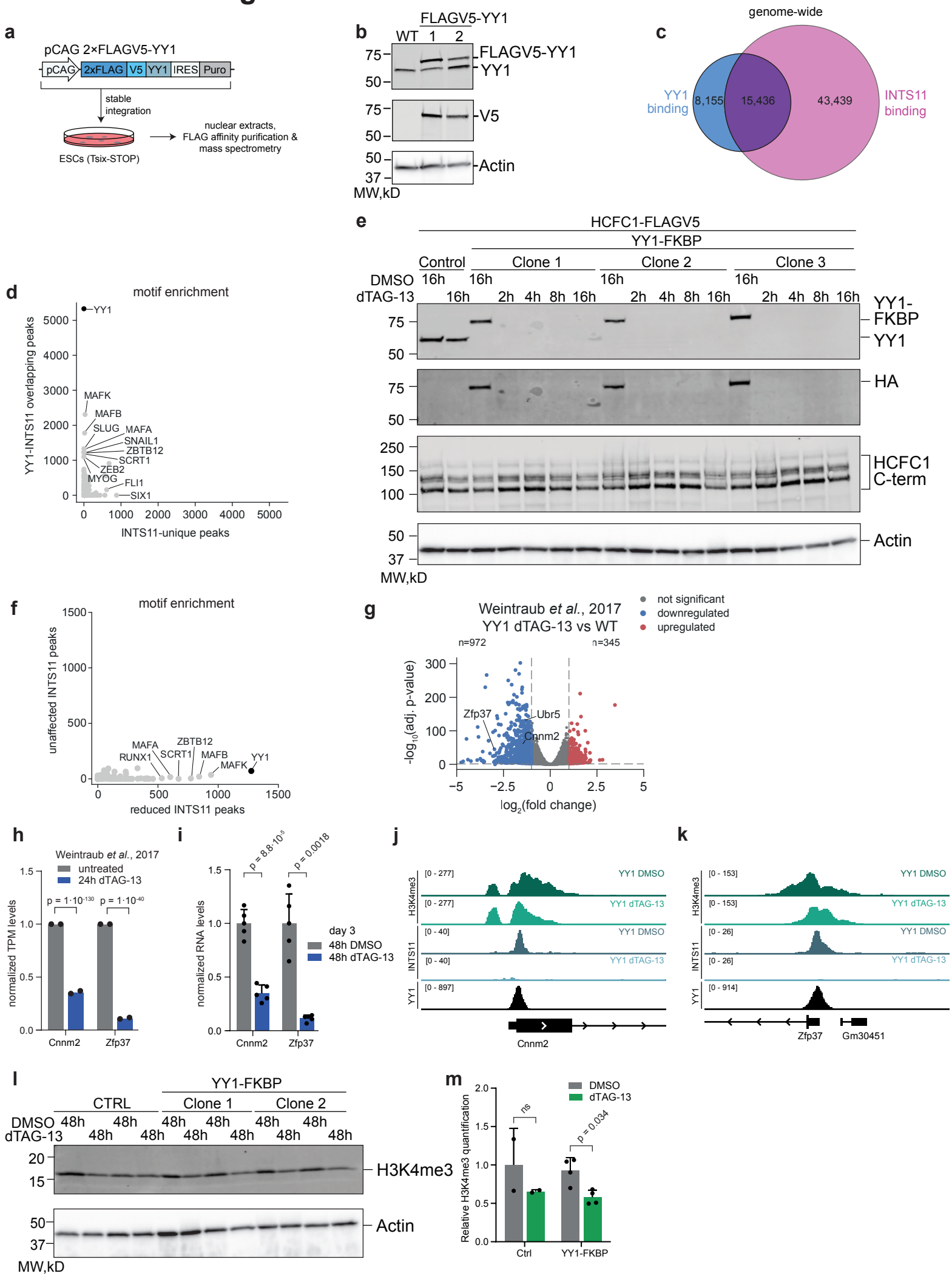
1174

Extended Data Figure 4



1175 Extended Data Figure 4: HCFC1 binding sites analysis and genome-wide H3K4me3 and INTS11
1176 recruitment upon dTAG-13-mediated depletion of HCFC1.
1177 **a**, Genomic annotation of HCFC1 peaks at day 0 and day 3 of ESC differentiation. **b**, PCR analysis
1178 confirming proper integration of HCFC1-V5- and the YY-FKBP-tags and presence of both X
1179 chromosomes (DX95). **c**, Feature enrichment analysis at HCFC1 peaks at day 3 of differentiation.
1180 Features are grouped by over- and underrepresentation, and the top 10 with the lowest p-values
1181 are labeled. **d**, Motif enrichment analysis at HCFC1 peaks at day 3 of differentiation. The 10 most
1182 significantly enriched motifs are labeled. **e**, Western blot analysis of H3K4me3 in HCFC1-FKBP
1183 cells treated with DMSO or dTAG-13 (48h) at day 3 of differentiation. **f**, Quantification of global
1184 H3K4me3 WB signal in Ctrl and HCFC1-FKBP cells treated with DMSO or dTAG-13 (48h) at
1185 day 3 of differentiation. ns, not significant, p-values were calculated using a two-sided Welch's t-
1186 test and corrected *via* Benjamini-Hochberg. **g**, Heatmap of H3K4me3 and INTS11 enrichment at
1187 H3K4me3 peaks lost upon HCFC1 depletion, and at unaffected peaks (random subset of 4,991
1188 peaks). **h**, MA plot showing gene expression changes upon dTAG-13-mediated depletion of
1189 HCFC1. Genes are colored based on significant upregulation (red), downregulation (blue) or no
1190 change (grey). **j**, Heatmap of H3K4me3 and INTS11 changes upon HCFC1 depletion, as well as
1191 HCFC1 and YY1 binding and H3K4me3 and INTS11 enrichment in the DMSO condition, at
1192 promoters of upregulated, unaffected and downregulated genes.

Extended Data Figure 5



1193 Extended Data Figure 5: YY1 and INTS11 binding sites analysis.

1194 **a**, Generation strategy of the 2xFLAG-V5-Yy1 ESC lines by random integration in XX *Tsix*-STOP
1195 ESCs. **b**, Western blotting analysis of YY1 expression in 2xFLAG-V5-YY1 ESCs using an
1196 endogenous YY1 or V5 antibody, and Actin as loading control. **c**, Venn diagram showing the
1197 genome-wide overlap of YY1 and INTS11 peaks at day 0 of differentiation. **d**, Scatter plot
1198 comparing motif enrichment between YY1-INTS11 overlapping peaks and INTS11-unique peaks
1199 identified in **c**. Each point represents a motif, plotted by its $-\log_{10}$(p-value) in both analyses, and
1200 motifs with the lowest p-values in either condition are labeled. **e**, WB analysis of YY1 degradation
1201 upon dTAG-13 treatment vs DMSO for different hours of YY1-FKBP ESCs at day 0 of
1202 differentiation. **f**, Scatter plot comparing motif enrichment between reduced and unaffected
1203 INTS11 peaks in YY1-FKBP ESCs in the dTAG-13 condition at day 3 of differentiation. Each
1204 point represents a motif, plotted by its $-\log_{10}$(p-value) in both analyses, and motifs with the lowest
1205 p-values in either condition are labeled. **g**, Volcano plot showing gene expression changes upon
1206 YY1 depletion. Genes are colored based on significant upregulation (red), downregulation (blue)
1207 or no change (grey). **h**, Expression of *Cnnm2* and *Zfp37* upon 24h YY1 removal in undifferentiated
1208 ESCs[33], average of n=2 biological replicates. TPM values are normalized to the average expression
1209 in the DMSO condition for each gene. P-values were calculated using Wald tests and corrected for
1210 multiple testing using the Benjamini-Hochberg method, as implemented by DESeq2. **i**, *Cnnm2* and
1211 *Zfp37* expression analysis by qRT-PCR in KFBP-YY1 ESCs upon 48h DMSO or dTAG-13
1212 treatment at day 3 of differentiation. A two-sided Welch's t-test was used to assess significant
1213 differences between DMSO- and dTAG-13 treated cells. P-values were corrected for multiple
1214 testing using the Benjamini-Hochberg method. **j**, Views of ChIP-seq tracks for H3K4me3 (green)
1215 and INTS11 (blue) in YY1-depleted ESCs versus its DMSO control at day 3 of differentiation,
1216 and YY1 ChIP-seq track in female ESCs at day 0 (black), around the *Cnnm2* promoter region. **k**,
1217 Same as **j**, for *Zfp37*. **l**, WB analysis of YY1 and H3K4me3 of YY1-FKBP ESCs differentiated
1218 for 3 days and treated with dTAG-13 or DMSO for the last 48h. **m**, Quantification of global
1219 H3K4me3 WB signal in Ctrl and YY1-FKBP cells treated with DMSO or dTAG-13 (48h) at day
1220 3 of differentiation. ns, not significant. P-values were calculated using a two-sided Welch's t-test
1221 and corrected *via* Benjamini-Hochberg.

1222

1223