Allele pairing at Sun1-enriched domains at the nuclear periphery via T_1A_3 tandem DNA repeats

Gayan I. Balasooriya[1,2*] and David L. Spector[1*]

[1]Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, New York 11724. USA.

[2] Current address: School of Cardiovascular and Metabolic Medicine and Sciences, King's College London, 125 Coldharbour Lane, London SE5 9NU, UK.

[*]Co-corresponding authors: spector@cshl.edu, gayan.balasooriya@kcl.ac.uk

Abstract

Spatiotemporal gene regulation is fundamental to the biology of diploid cells. Therefore, effective communication between two alleles and their geometry in the nucleus is important. However, the mechanism that fine-tunes the expression from each of the two alleles of an autosome is enigmatic. By developing an allele-specific gene expression visualization system in living cells, we show that alleles of biallelically expressed *Cth* and *Ttc4* genes are paired prior to acquiring monoallelic expression. We found that active alleles of monoallelic genes are preferentially localized at Sun1-enriched domains at the nuclear periphery. These peripherally localized euchromatic DNA loci harboring monoalleleic genes are enriched with adenine-thymidine-rich tandem repeats, which are bound by Hnrnpd, the "molecular wrench", to localize these loci in a Hi-C-defined A compartment within the B compartment. Our results demonstrate the biological significance of T_1A_3 tandem repeat sequences in genome organization and how the regulation of gene expression, at the level of individual alleles, relates to their spatial arrangement.

Introduction

Numerous studies have shown that chromatin is organised into nuclear territories in the interphase mammalian nucleus, seemingly linking chromosome architecture to gene regulation and transcriptional heredity[1-7]. In a diploid genome, precise regulation of transcription dosage requires effective communication between the two alleles; therefore, the positioning of the two alleles of a gene in the nucleus may play a significant role in "alleleic communication" in the regulation of autosomal gene expression[8-13]. Although in theory, both alleles of a gene could be equally exposed to the identical *trans* environment in the nucleus, the presence of nuclear domains indicates a compartmentalized nucleoplasm, potentially facilitating gene-gathering in "hubs" owing to genome folding and allele pairing potential[14-18], which could significantly impact the regulation of allelic-specific gene expression. Autosomal monoallelic gene expression has been identified in mammalian diploid cells in which only a stochastically or deterministically chosen single allele is transcriptionally active in a given cell[19-21]. Although a recent study showed allele-specific differential regulation of monoallelic genes[22], the mechanism that regulates the allele choice of monoallelic genes, which is distinctive to gene imprinting, and the allelic geometry of those genes remain poorly understood.

Pioneering molecular techniques have been developed to visualize mRNA, DNA, and proteins in live cells, advancing our understanding of the spatial and temporal dynamics of those molecules[23-31]. However, the techniques available to date for exploring genome regulation in living cells can not capture the spatio-temporal dynamics of allelic gene expression, especially referring to the parental origin of transcripts and alleles simultaneously, precluding a deep understanding of the molecular mechanisms of monoallelic gene expression regulation in a context-dependent manner.

61 Therefore, developing molecular tools and techniques to gain mechanistic insights into
62 spatiotemporal allelic gene expression regulation in living cells will immensely deepen our
63 fundamental understanding of the governing principles of gene expression in organismal
64 development, homeostasis and disease[32-34].
65
66 In this study, we developed a molecular tool to distinguish the spatiotemporal expression of RNAs,
67 referring to their maternal and paternal alleles' spatial organization in a living cell nucleus, using
68 polymorphic F1 hybrid mouse embryonic stem cells (mESCs). Utilizing this system to probe two
69 monoallelically expressed autosomal genes in neural progenitor cells (NPCs), we showed that
70 when the genes are biallelically expressed in mESCs, alleles are paired, and the transcriptionally
71 active alleles are localized at Sun1-enriched domains at the nuclear periphery (NP). We further
72 showed, upon mESCs differentiation to NPCs, that those genes gain monoallelic status, and the
73 active alleles remain localized at the Sun1-enriched domains. By immunoprecipitating the
74 chromatin associated with Sun1-enriched domains, we further showed that monoallelic genes are
75 preferentially expressed at the Sun1-enriched domains at the NP compared to biallelic genes.
76 Using publicly available data, we further showed that the monoallelic genes are localized in
77 "micro-TADs" in Hi-C-defined euchromatic A compartments within peripherally located LADs.
78 In addition, we discovered that genes localized at the NP are associated with *de novo* T_1A_3 tandem
79 DNA repeat sequences and that these T_1A_3 repeats are involved in monoallelic gene localization to
80 the Sun1-enriched domains by binding to Hnrnpd.
81
82
83

Results

Visualizing allele-specific gene expression in living cells

Live imaging of subnuclear structures is challenged by molecular noise inherent to biological systems[31,35,36], which significantly impacts image processing and analysis due to the low signal-to-noise ratio. In addition, ectopically expressed fluorescently labelled reporter proteins may cause molecular crowding in cells, potentially altering cellular physiology. Furthermore, cell-penetrating fluorescent reporter molecules[37] may, in some cases, become a limiting factor for long-term time-lapse imaging. To overcome these limitations, we first generated a doxycycline-inducible, tunable EYFP gene fused to the MS2 coat protein (MS2-EYFP) and an mCherry gene fused to the PP7 coat protein (PP7-mCherry) expressing mESC line (Fig. 1a and Extended Data Fig. 1a) using a CRISPR-based approach. Second, we conditionally inserted 24 PP7 stem-loop repeats (PP7-SL) sequence into the 3' UTR of the CAST/EiJ allele (paternal) and 24 MS2 stem-loop repeats (MS2-SL) sequence into the 3' UTR of the C57BL/6J allele (maternal) of two different genes using Cas9-D10A nickase and allele-specific SNP-targeting gRNAs (Fig.1a). We established this allele-specific gene expression visualization system for two random autosomal monoallelic (RaMA) genes, *Cth* and *Ttc4*, which are biallelically expressed in mESCs but become monoallelic upon differentiation[22], using two independent mESC clones (Fig. 1c).

Active *Cth* and *Ttc4* alleles are paired

To confirm that the MS2-SL and PP7-SL are integrated into the targeted *Cth* and *Ttc4* gene alleles, but not off-targeted to different loci, we performed dual-color DNA-FISH using nick-translated probes designed to detect the endogenous alleles (green) and the inserted stem-loop cassettes (red).

107 In the IF images, we observed the red and green fluorescent signals exclusively adjacent to each
108 other, confirming on-target stem-loop knock-in for both the *Cth* and *Ttc4* genes (Extended Data
109 Fig.1b). We observed that both alleles of the *Cth* gene were within ~1µm distance of each other,
110 and the alleles of the *Ttc4* gene were within less than ~3µm distance of each other in mESCs
111 (Extended Data Fig.1b, c). To assess the coat-protein binding specificity toward their cognate
112 stem-loops, we induced the expression of MS2-EYFP and PP7-mCherry coat proteins in mESCs
113 and mESC-derived NPCs to label nascent RNAs from the *Cth* and *Ttc4* alleles. Imaging of fixed
114 mESCs and NPCs, confirmed stem-loop-coat protein specificity in our knock-in clones (Fig. 1b
115 and Extended Data Fig. 1d).
116
117 Homologous allele pairing has been shown to influence gene regulation[8,11,38], yet the broader
118 impact of allele pairing and three-dimensional nuclear arrangement on transcriptional heredity
119 remains poorly understood. By examining nascent RNAs from the maternal and paternal alleles of
120 the *Cth* gene every 20 seconds over 20 minutes by live-cell imaging (Fig.1c and Supplementary
121 Movie 1), we observed that the two actively transcribing alleles frequently colocalize in mESCs
122 (Fig.1d, e; n=22 cells).
123
124 Because genome editing at specific loci can potentially perturb chromatin organization[39,40], we
125 asked whether the observed allele localization is an artifect caused by insertion of stem-loops to
126 both alleles. To test this, we performed nascent RNA-FISH for the *Cth* and *Ttc4* genes in wild-
127 type mESCs and mESC-derived NPCs (Fig.1f and Extended Data Fig.1e). By quantifying the
128 distance between flurosent signals (nasent RNA-FISH), we observed that the *Cth* alleles are paired
129 in 81% of the mESCs (8% are monoallelic), and the *Ttc4* alleles are paired in 59% of the mESCs

130 (Fig.1f - column graph). These results confirm that insertion of MS2 and PP7 stem-loops did not
131 alter the chromatin positioning; therefore, allele colocalizion of the *Cth* and *Ttc4* genes represents
132 a true biological phenomenon.
133
134
135 **Active *Cth* and *Ttc4* alleles localize to the nuclear periphery**
136 In the fluorescent images obtained from the mESC and NPC (Fig. 1b, f) allele-specific gene
137 expression systems, and the nasent RNA-FISH images for the *Cth* and *Ttc4* genes (Fig. 1f), we
138 observed that transcriptionally active *Cth* and *Ttc4* alleles localized to the NP (Fig.2a and Extended
139 Data Fig.2a). Therfore, since *Cth* and *Ttc4* are monoallelic genes, to assess whether the active
140 monoallelic gene alleles localized at the NP is a unique characteristic of this gene class, we
141 quantitatively assessed the nuclear positioning of the transcriptionally active alleles of a panel of
142 13 RaMA and 14 deterministic monoallelic (DeMA) genes by performing a nasent RNA-FISH
143 assay. We found that the DeMA and RaMA genes we assessed were preferentially localized
144 (DeMA genes 95% on average; RaMA genes 93% on average; n=100 cells per gene) at the NP of
145 NPCs (Fig.2b and Extended Data Fig.2b,c). In line with our observation, previous studies have
146 reported that transcriptionally active olfactory receptor genes, -an autosomal monoallelic gene
147 family-, are also transcribed at the NP[41-43], suggesting the existance of transcriptionally active
148 monoallelic gene hubs at the NP.
149
150
151
152

153 **Monoallelic genes localize at Sun1 territories**

154 Following our observation of the localization of transcriptionally active *Cth* and *Ttc4* alleles at the
155 NP, we asked whether these active alleles associate with specific proteins or protein territories at
156 the inner nuclear membrane that might provide a mechanism for harboring alleles. Sun1, an inner-
157 nuclear-membrane protein (Extended Data Fig.2d), is involved in positioning the nucleus in the
158 cytoplasm by interacting with actin filaments via nesprin[44,45]. In addition, Sun1 is associated with
159 the nuclear pore complex, which plays a role in mRNA transport[46,47] and in harboring active
160 genes[14,48]. Since Sun1-enriched nuclear domains are associated with stabilizing chromatin
161 architecture[49], their specific molecular function in gene regulation is unknown. Although overall,
162 the lamina-associated domains at the NP harbor heterochromatin[50], the transcriptional status of
163 genes specifically associated with Sun1-enriched domains is unknown. Using an
164 immunofluorescent assay for Sun1 protein in *Cth* and *Ttc4* mESC generated clones (Fig.1a), we
165 identified that nascent RNAs from both alleles of the *Cth* and *Ttc4* genes colocalized with Sun1
166 protein at the NP (Fig.2c and Extended Data Fig.2e). Based on this observation, we hypothesized
167 that monoallelic genes might preferentially localize to Sun1-enriched territories at the NP. To test
168 our hypothesis, we performed Sun1 Chromatin ImmunoPrecipitation (ChIP) in mESCs and *in*
169 *vitro*-derived NPCs. To validate the specificity of the Sun1 antibody, we performed mass
170 spectrometric (MS) analysis of proteins isolated from the Sun1 ChIP assay. In both mESC and
171 NPC MS datasets, Sun1 was the most significantly enriched protein in the pull-down (Fig. 2d and
172 Extended Data Table 1), confirming the antibody specificity for Sun1, which was subsequently
173 confirmed by immunoblot analysis (Fig.2e).

174

175 By performing qPCR analysis for the DNA isolated from chromatin pulled down by the Sun1 ChIP
176 assay followed by Sanger sequencing, we confirmed that both *Ttc4* alleles are localized to Sun1-
177 enriched territories in mESCs, whereas only a single allele was associated with Sun1 in NPCs
178 (Fig.2f,g and Extended Data Fig.2f, g). Although only a single *Cth* allele was detected in the Sun1
179 pull-down in NPCs (Fig.2g), the *Cth* gene could not be amplified from the Sun1 pull-down DNA
180 in mESCs (Fig.2f), likely due to low enrichment of *Cth* in the Sun1 pull-down chromatin
181 fraction. Therefore, to assess *Cth* and other genes associated with Sun1-enriched chromatin, we
182 deep-sequenced the DNA fragments isolated from the Sun1-ChIP chromatin. The allele-specific
183 ChIP-seq data showed that the *Cth* peaks were enriched on both alleles in mESCs and on a single
184 allele in NPCs, located at ~4.5 kb downstream of the *Cth* gene (Fig.2h – upper panel). In NPCs, a
185 single peak was homologously enriched within the intronic region of both *Cth* alleles. For *Ttc4,*
186 allele-specific ChIP-seq peaks were enriched at distinct genomic regions in mESC versus NPCs:
187 in mESCs, peaks localized to the gene body of *Ttc4 or* neighboring gene *Mroh7*, whereas in NPCs,
188 peaks occurred ~4 kb downstream of the *Fam151a* and *Acot11* gene bodies (Fig.2h – lower panel).
189 In addition, we found that the monoallelic genes we assessed in the nascent RNA-FISH assay were
190 also enriched in the Sun1-associated chromatin (Extended Data Fig. 2h,i). Moreover, we also
191 found that 49% of transcriptionally active DeMA genes and 61.5% of RaMA genes in mESCs
192 (from our previous study[22]) overlapped with the genes enriched in Sun1-associated chromatin
193 (Extended Data Table 2), whereas only 27.8% of transcriptionally active biallelic genes were
194 Sun1-associated (Fig.2i). Furthermore, when we assessed how common these monoallelic gene
195 associations were in Sun1 territories in NPCs, we found that 74.1% of monoallelic genes in mESCs
196 are enriched in the Sun1 chromatin pull-down (Extended Data Fig.2j and Extended Data Table 3).
197

198 **“micro-TADs” within LADs harbor active genes**

199 Although lamina-associated domains (LADs) have been shown to harbor heterochromatin[51-54],
200 emerging evidence suggests that some of the genes within LADs are transcriptionally active[55,56].
201 Since we found that the active alleles of monoallelic genes were preferentially localized at the NP,
202 we examined whether the Cth, Ttc4, and the monoallelic gene panels we used in the nascent RNA-
203 FISH assay were associated with Hi-C-defined compartments in ESCs and NPCs[57,58] using the
204 published data in the 4DN data portal[59,60]. We found that, compared to low-resolution
205 compartments, in high-resolution compartments defined by an insulation score-diamond, the Cth
206 and Ttc4 genes (Fig.3a, Extended Data Fig.3a,b) and the monoallelic gene panel we used
207 (Extended Data Fig.3c,d) were housed in “micro A” compartments within B compartments. This
208 finding confirmed our observation of active monoallelic genes at the NP and demonstrated the
209 presence of euchromatic compartments within LADs, which we named “micro-TADs”.

210

211

212 **TA-tandem repeats are enriched in monoallelic genes**

213 Chromatin loci are positioned within specific domains in the three-dimensional nuclear space.
214 Although it has been shown that the nuclear lamina acts as a scaffold to anchor chromatin at the
215 NP via nuclear envelope transmembrane proteins (e.g., HP1, Lap2β, cKrox, Lamin_B receptor)[61-
216 63], the DNA motifs that facilitate chromatin positioning at the nuclear envelope are poorly
217 documented. Since monoallelic genes are preferentially located in Sun1-enriched territories, we
218 then asked whether specific motif sequences are present in the chromatin associated with Sun1
219 protein territories. By performing motif analysis, we discovered that T_1A_3 and A_1T_3 tandem repeats
220 (which we refer to as TA-bridges) are significantly (p=1e-511 and p=1e-242, respectively)

221 enriched in Sun1 pulled-down chromatin fractions (Fig.3c, Extended Data Fig.3e,f, Extended Data
222 Table 4). We also found that those TA-bridges were present approximately 5 kb downstream of
223 the 3' end and on the fourth intron of *the Cth* gene, and ~10 kb downstream of the 3' end and ~100
224 kb upstream of the TSS (in the first intron of *the Acot11 gene*) of the *Ttc4* gene (Fig.2h).
225 Additionally, some genes known to be expressed monoallelically were located near these TA-
226 bridges (*Cth*- downstream: *Zranb2*; upstream: Ankrd13c, Srsf11, Lrrc40; and Ttc4-
227 downstream: Pars2, Dhcr24; upstream: Ssbp3, Mrpl37, Cdcp2). The monoallelic gene panels
228 examined in the RNA-FISH assay are also linked to TA-bridges, suggesting a potential mechanism
229 by which TA-bridges may facilitate the establishment and maintenance of monoallelic gene
230 expression within Sun1 territories.
231
232 Through genome-wide scanning, we identified 4,572 TA-bridge motifs across the entire mouse
233 genome. Of these, 887 (19.4%) were located on the Y chromosome (Extended Data Fig.3g).
234 Considering the comparatively smaller size of the Y chromosome and its higher enrichment of
235 TA-bridges compared to other chromosomes (Fig.3d), as well as its predominant localisation to
236 the A compartment (Fig.3e), we examined the position of the Y-chromosome territory within the
237 nucleus using published LaminB1-DamID data in ESCs from the 4DN data portal[59,60]. Although
238 the LaminB1-DamID data (4DNFI32K5L77 – from the Bas van Steensel lab, NKI, NL) indicated
239 that the Y chromosome is localized at the NP (Fig.3e - middle panel of enlarged area), Hi-C-
240 defined compartments[57] showed that the Y chromosome mainly resides in the A compartment, and
241 its RNA expression profile[64] aligned with the active compartment profiles (Fig.3e - lower panel of
242 enlarged area), further supporting the presence of "micro-TADs" within LADs.
243

244 **TA-bridges localize to Sun1 territories**

245 To further investigate whether the TA-bridge motif could direct their associated chromatin to the
246 NP, we developed an assay to visualize TA-bridge positioning within the nucleus in live cells,
247 using immunofluorescence microscopy, and to identify the *in vivo* TA-bridge-interacting proteins
248 (Fig.3f). We transiently transfected biotin-labeled single-stranded $(T_1A_3)_8$ and $(A_1T_3)_8$ DNA
249 oligonucleotides, as well as double-stranded $(T_1A_3)_8$ DNA into mESCs and NPCs, and observed
250 that biotin-labeled single-stranded $(T_1A_3)_8$ and $(A_1T_3)_8$ DNA, as well as double-stranded $(T_1A_3)_8$
251 DNA, were located at the NP among the nuclear-localized signals in ESCs and NPCs (Fig.3g,h
252 and Extended Data Fig.3h).

253

254 Next, we aimed to identify the protein candidates interacting with TA-bridges by performing mass
255 spectrometry (MS) analysis of the protein pull-down by AT-bridges. Although previously known
256 Sun1-binding proteins (Lmnb1, Lmnb2, Emerin) were pulled down by TA-bridges, Sun1 was not
257 significantly enriched in the $(T_1A_3)_8$ or double-stranded TA-bridge DNA motif pull-down protein
258 fractions (Fig. 4a, b, and Extended Data Table 5), and was also not pulled down by the $(A_1T_3)_8$
259 motif (Extended Data Fig. 5a). Therefore, we could not conclude that Sun1 is the direct binding
260 protein of the TA-bridges. To further investigate the "molecular hook" that binds to the double-
261 stranded TA-bridges, we conducted an electrophoretic mobility shift assay (EMSA) using nuclear
262 protein extracted from mouse ESCs. In the EMSA, we observed multiple shifted biotin-positive
263 bands for $(T_1A_3)_8$ and $(A_1T_3)_8$, as these DNA probes most likely interacted with multiple protein
264 complexes (Fig. 4c). Interestingly, although we noted that the double-stranded TA-bridge DNA
265 motifs were localised at the Sun1 territory in the NP of ESCs and NPCs in the immunofluorescence
266 assay, these motifs did not show any binding with nuclear proteins.

267 **Hnrnpd binds to TA-bridges**

268 Although we identified several protein candidates known to interact with single-stranded DNA in
269 TA-bridge MS data, to identify proteins that directly bind to TA-bridges, we performed MS
270 analysis of proteins extracted from the multiple shifted bands identified in the EMSA assay. MS
271 data from the five excised bands showed no enrichment of Sun1 protein (Extended Data Fig.4b
272 and Extended Data Table 6), confirming that Sun1 does not directly interact with TA-bridges.
273 However, by intersecting the most significant protein (Extended Data Fig.4b – circled clusters)
274 candidates across multiple bands from the EMSA assay, we identified Hnrnpd as the only common
275 protein candidate, which was also enriched in proteins pulled down by Sun1 (Fig.2d). To further
276 investigate the direct interaction between Hnrnpd and TA-bridges, we examined known functions
277 attributed to Hnrnpd. Analogous to AT-rich interacting domain (ARID)-containing proteins,
278 Hnrnpd has been reported to bind to the adenine-thymine-rich minor grooves of B-form DNA[65]
279 (GO:0003680) (Extended Data Fig.4c), supporting our identification of Hnrnpd as a "molecular
280 hook" that binds TA-bridge tandem repeats. Furthermore, our findings suggest that Hnrnpd could
281 bind to both the noncanonical 5'-$(T_1A_3)_n$-3' and the 5'-$(A_1T_3)_n$-3' tandem repeat sequences. To
282 investigate whether Hnrnpd binds to TA-bridges before or after their localisation to Sun1-enriched
283 territories, we performed co-immunoprecipitation of Sun1-associated proteins using Sun1
284 antibody on nucleic acid-depleted ESC and NPC nuclear extracts, followed by MS analysis.
285 Evaluation of the proteins immunoprecipitated with Sun1 revealed that Hnrnpd was less abundant
286 in the Sun1 pull-down from ESC nuclear extracts (Fig.4d) and could not be detected in the Sun1
287 pull-down from NPC nuclear extracts (Fig.4e and Extended Data Table 7), suggesting Hnrnpd
288 most likely binds to the AT-bridge motifs prior to the allele's localisation to Sun1-enriched
289 territories.

290 **Discussion**

291 Although allele pairing has been observed for decades, the presence of this phenomenon in
292 mammals has not been extensively studied[38]. Furthermore, it has not been previously demonstrated
293 that allele pairing is necessary for some genes to be expressed from both alleles. Using
294 the *Cth* and *Ttc4* genes, which are biallelically expressed in mESCs but become monoallelic upon
295 differentiation, we showed that allele pairing for certain genes coincides with transcription from
296 both alleles. In this study, we also demonstrated that for the monoallelic *Cth* and *Ttc4* genes, when
297 the alleles are positioned further apart, only the allele associated with the Sun1-enriched domain
298 is transcriptionally active. The establishment of *monoallelic expression, in terms of gene*
299 *imprinting,* is described as a mechanism to regulate the transcriptional dosage of dosage-sensitive
300 genes during early development, but detailed studies evaluating whether imprinted genes can be
301 expressed biallelically in differentiated cells or in rare diseases are lacking. Although
302 the *Cth* and *Ttc4* genes attain monoallelic status in differentiated cells, their previous biallelic
303 expression, distinct from imprinted genes, suggests a degree of plasticity in the transcriptional
304 status of the inactive allele. Consequently, reactivation of the silent allele—particularly in the
305 context of "disease causative monoallelic genes"—may have clinical significance. Additionally,
306 our finding that the active alleles of monoallelic genes are localised to Sun1-enriched territories at
307 the NP suggests that their gene expression may be facilitated within these specialised nuclear
308 compartments.

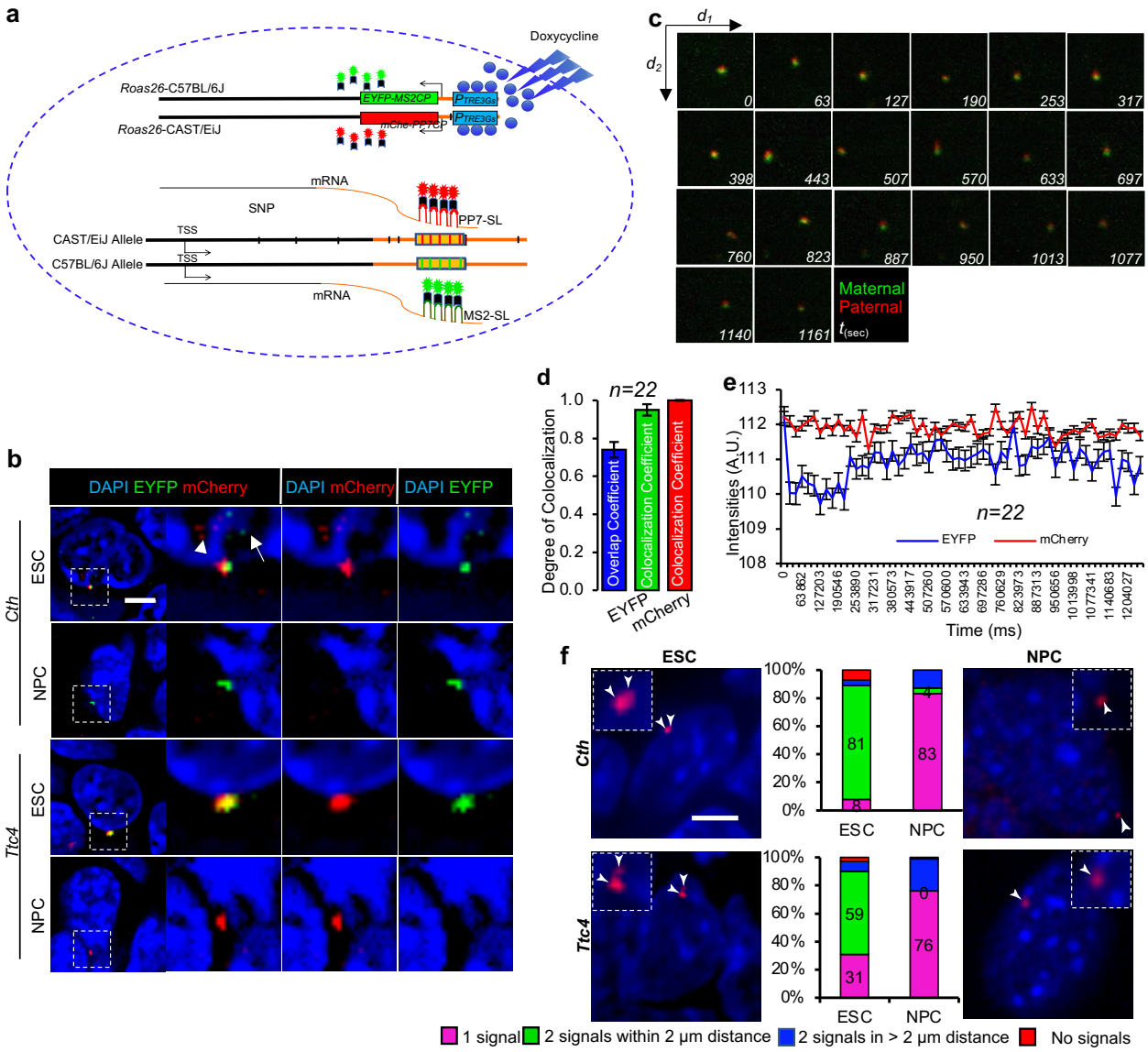
309

310 Our observation of monoallelic genes' preferential expression at the NP supports the idea that the
311 chromatin located at the NP is not solely heterochromatic[66-70]. Evaluating published data, we also
312 demonstrated that the active alleles of monoallelic genes are cluster-embedded within the B

313 compartment at the NP. However, why many monoallelic genes adopt this unique arrangement at
314 the NP remains enigmatic. In our previous studies, we showed that most monoallelic genes encode
315 only one RNA isoform, which is translated into a functional protein[22]. . Therefore, it is reasonable
316 to argue that NP positioning may assist RNAs in being exported through nuclear pore complexes
317 to avoid the high rate of RNA turnover in the nucleus— acting as a "defensive mechanism," as
318 proposed in the gene gating hypothesis, an idea worth exploring further. Moreover, since the
319 transcriptionally active alleles are localized at the NP, these genes (monoallelic: Rac1, S100a10,
320 Junb; biallelic: Sun1, Scn5a) may be at risk of misregulation in diseases caused by laminopathy
321 and nuclear deformation[71], and thus could have clinical relevance.
322
323 Although in association with SPDYA, SHELTERIN, and TTM complex proteins, Sun1 stabilizes
324 the telomeric rings at the inner nuclear membrane[72], Sun1 does not serve as a direct binding partner
325 to the DNA. Therefore, in line with our findings, it is likely part of a scaffold that harbors
326 chromatin. The discovery of novel repetitive elements in genomes has been achieved using long-
327 read DNA sequencing methods[73,74]. However, so far, only a few repetitive elements have been
328 linked to genome regulation[75-78], and the functional significance of many remains elusive.
329 Although some transcriptionally active genes have been reported to be localized at the NP[66-70], the
330 mechanism by which alleles are physically positioned and maintained at the NP remains unknown.
331 Here, we discovered that Hnrnpd physically interacts with TA-bridges, highlighting the potential
332 importance of other yet-to-be-identified DNA tandem repeat elements. While Hnrnpd binds to TA-
333 bridges, it is not an enriched component of Sun1 territories, suggesting that Hnrnpd's binding to
334 the TA-bridges occurs before the localisation of alleles to the Sun1 territories. Furthermore, we
335 observed that known helicases (Mcm5[79] and Top2a[80]) are significantly enriched among TA-

336 bridge interacting proteins (Fig.4d,e), indicating that strand separation of the TA-bridge likely
337 initiates prior to the motifs being localised at the Sun1 nuclear territories.
338
339 During stem cell differentiation into more specialized cells, the gene expression program changes
340 alongside chromatin reorganisation, leading to different architectures depending on the
341 progenies[81,82]. Furthermore, it has been shown that asynchronous allele replication occurs more
342 frequently during cell differentiation, potentially promoting the early-replicating allele to become
343 the active allele of a monoallelic gene[83,84]. Because Hnrnpd binds to TA-bridges independently of
344 Sun1, we propose a mechanistic model where Hnrnpd attaches to single-stranded TA-bridges of
345 the early replicating allele in prophase, resulting in the targeting of the TA-bridge-bound Hnrnpd
346 to the newly forming nuclear membrane Sun1-enriched domains in telophase via a molecular
347 targeting mechanism to be elucidated in future studies.
348
349
350
351
352
353
354
355
356
357

358 **Figure 1**



359

360

361 **Figure 1 | Development of an allele-specific gene expression system to assess the localization**

362 **of *Cth* and *Ttc4* alleles.**

363 **a,** Schematic illustration of the allele-specific gene expression visualization system. Inducible

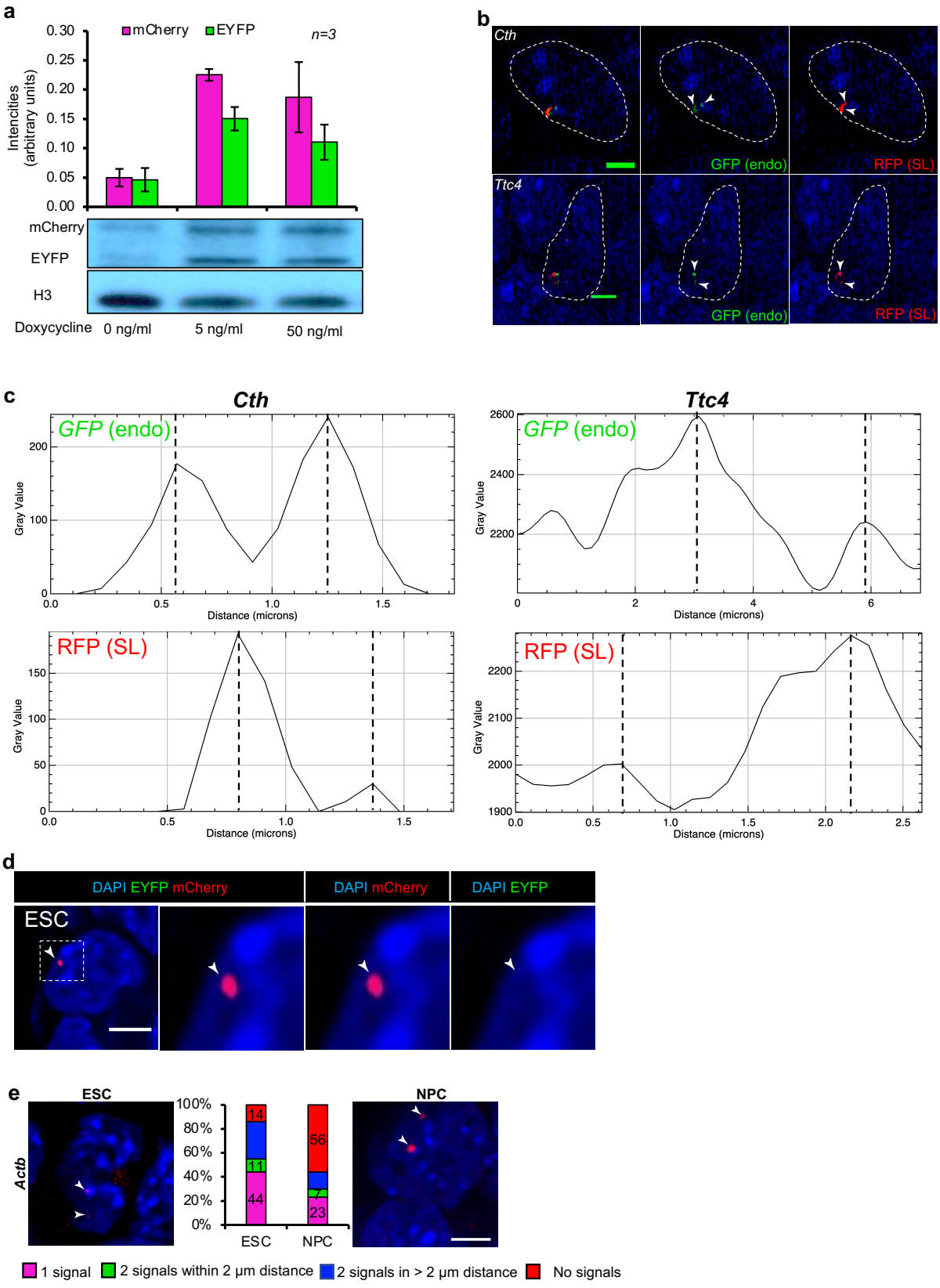
364 expression of EYFP-MS2CP and mCherry-PP7CP is sensitive to doxycycline dosage. These coat

365 proteins interact with unique stem-loops, MS2-SL with EYFP-MS2CP and PP7-SL with mCherry-

366 PP7CP at the 3' UTR of nascent mRNAs. Each RNA molecule contains 24 stem-loops. The
367 maternal (C57BL/6J) mRNAs are marked by EYFP, and the paternal (CAST/EiJ) mRNAs are
368 marked by mCherry. Low-background fluorescent signals are an advantage in this system, as they
369 enable easier allele detection and do not require significant background noise reduction during
370 image analysis. Additionally, this introduces minimal disturbance to the cell's physiology because
371 the crowding of fluorescent molecules within the cells is minimal and controllable. **b,** EYFP-
372 MS2CP and mChe-PP7CP signals in doxycycline- induced *Cth* and *Ttc4* ESC and NPC cells. Cells
373 were fixed 36 hours post- doxycycline induction. Dotted areas in the left image panels are enlarged
374 to demonstrate the fluorescent signal colocalisation at the nuclear periphery. Fluorescent signals,
375 possibly from single RNA molecules (maternal: arrow, paternal: arrowhead), are visible in the
376 enlarged images in the first row. Scale bar - 6 µm. **c,** Snapshots of a tracked single pair of alleles
377 at different time points from a 20-minute video using the 3i lattice lightsheet microscope. Green
378 and red dots represent the maternal and paternal *Cth* alleles, respectively. **d,** Degree of
379 colocalisation of the EYFP and mCherry signals calculated from tracked 22 EYFP and mCherry
380 signal pairs in 22 cells over a 20-minute video. Columns represent the overlap coefficient of green
381 and red dots (blue column) and colocalisation coefficients of EYFP (green column) and mCherry
382 (red column). Error bars: SEM. n = 22 cells. **e,** Line graph showing fluctuations in the intensity of
383 the colocalised EYFP and mCherry dots. Live images were captured every minute using the 3i
384 lattice lightsheet microscope. Error bars: SEM. n = 22. **f,** Representative RNA-FISH images
385 indicate localisation of nascent mRNAs of *Cth* and *Ttc 4* alleles at the nuclear periphery in mESCs
386 and NPCs. Arrowheads indicate the nascent RNA-FISH signals. Magnified images (within the
387 dotted-lined box) show active pairing of the *Cth* and *Ttc4* alleles. Scale bar - 6 µm. The column

388 graphs show the percentage of cells quantified for distances between red and green dots in nascent
389 RNA-FISH images. Two alleles within a 2 μm proximity are considered as paired alleles.
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410

411 **Extended Data Figure 1**

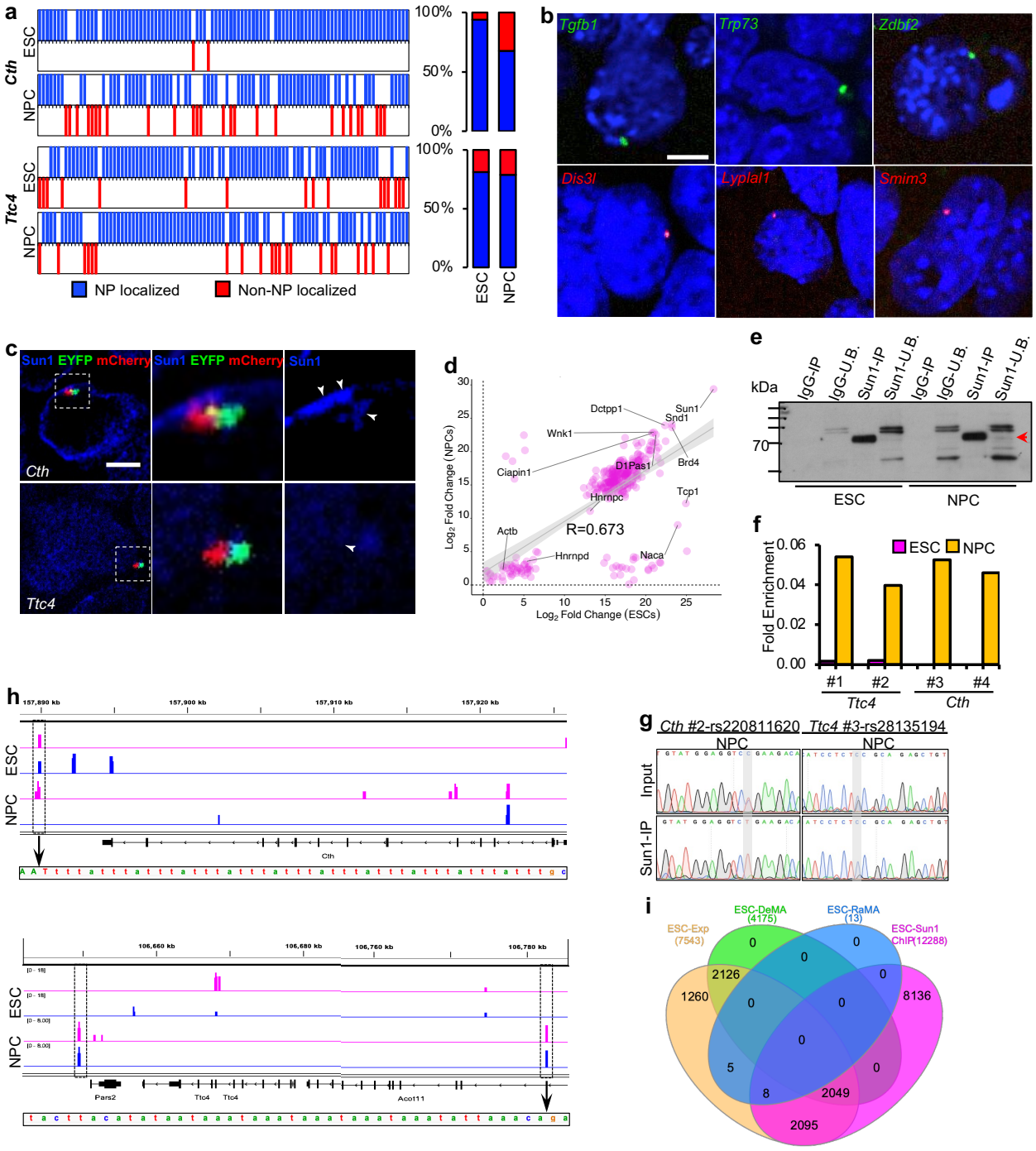


412

413 **Extended Data Figure 1| Development of an allele-specific gene expression system to assess**
414 **the localization of *Cth* and *Ttc4* alleles.**

415 **a**, Immunoblot (lower panel) showing EYFP and mCherry responsiveness to doxycycline dosage.
416 Bar graph presenting quantification of the EYFP (using GFP antibody) and mCherry bands in
417 immunoblots from three independent experiments (n=3). **b,** DNA-FISH of *Cth* and *Ttc4* genes
418 targeted with MS2 and PP7 stem-loops in NPCs. Endogenous alleles are labelled with the nick-
419 translated green fluorescent probes, and the stem-loops are labelled with the nick-translated red
420 fluorescent probes. Total DNA is labeled with DAPI. Two signals were detected for green and red
421 in these positively labeled (71%) nuclei. No nuclei exhibited only green or only red signals. Scale
422 bar – 6 µm. **c,** The distance between the two *Cth* alleles and the two *Ttc4* alleles in NPC nuclei.
423 The average distances between the endogenous *Cth* alleles are approximately 0.6 µm, and the two
424 inserted stem-loops are around 0.55 µm, whereas the average distance between the endogenous
425 Ttc4 alleles is about 2.5 µm and the stem-loops approximately 1.5 µm. **d,** Monoallelic expression
426 of the *Ttc4* gene in mESCs. Although *Ttc4* was expected to be expressed biallelically in mESCs,
427 we observed that *Ttc4* is monoallelically expressed in 31% of the mESCs. **e,** Nascent mRNA FISH
428 assay for *Actb* in wild-type mESCs and NPCs. Arrowheads indicate nascent RNA-FISH signals.
429 The accompanying column graphs show the percentages of observations for the distance between
430 the two alleles. Alleles within 2 µm of each other are considered paired. Scale bar - 6 µm. Active
431 *Actb* alleles were paired in only 11% of mESCs, while the remaining cells positive for RNA-FISH
432 signals showed single allelic (44%) and biallelic (11%) expression. Compared to *Cth* and *Ttc4*
433 (Fig.1f), the percentage of cells lacking signals (ESCs: 14% and NPCs: 56%) and the previous
434 observation[22] of the imbalance of mature versus nascent RNAs of Actb, along with published
435 data[85], suggest that single allelic expression (44%) of Actb may be due to gene pulsing.

436 **Figure 2**



437

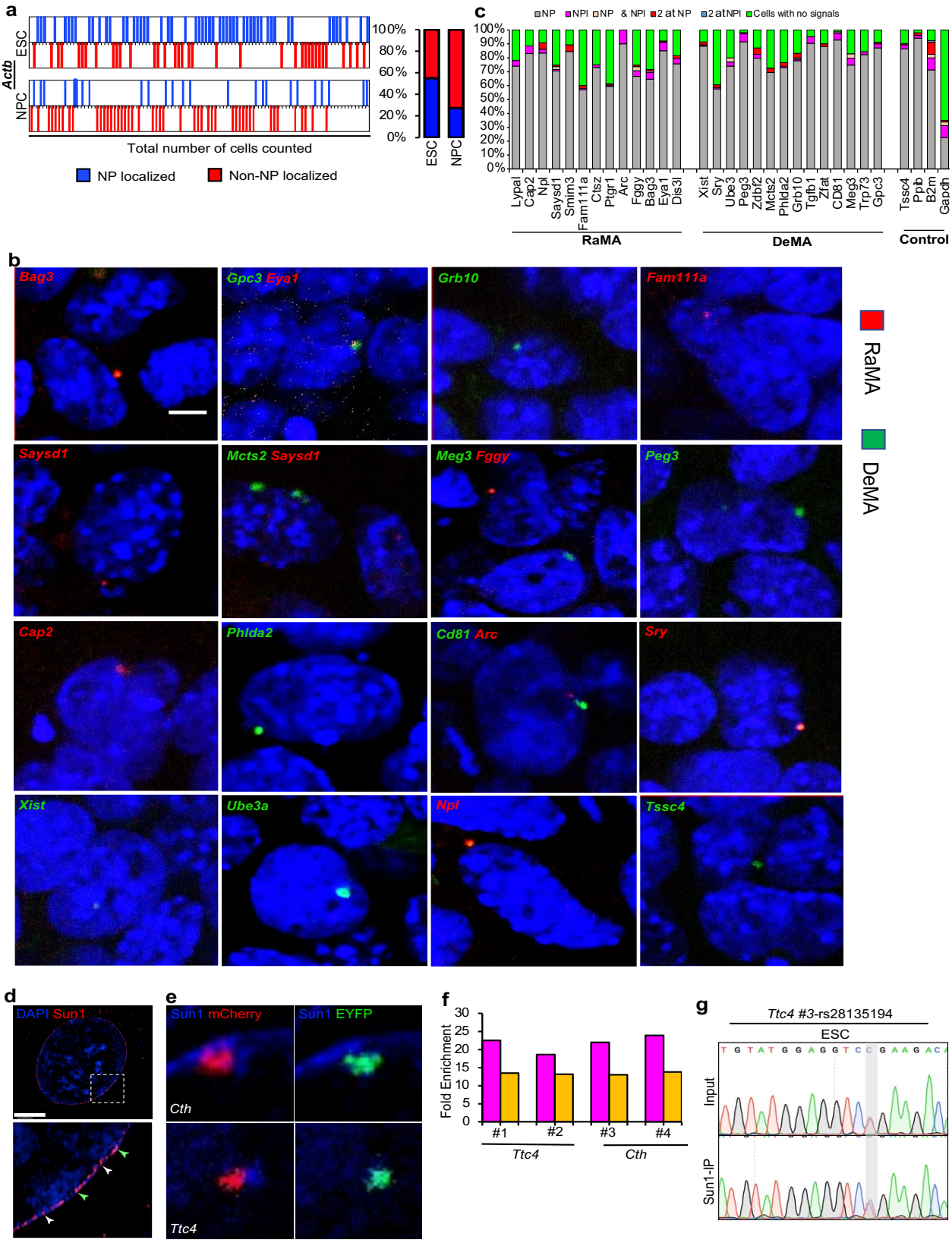
438 **Figure 2 | Active alleles are localized at Sun1 protein territories.**

439 Quantification of *Cth* and *Ttc4* nascent RNA- FISH signal localization in the nucleus. Each line

440 represents a counted cell, and empty points on the X- axis indicate cells with one signal at the

441 nuclear periphery and the other more internally located in the nucleoplasm. Blue and red lines
442 denote cells with nuclear periphery or nucleoplasm- localized nascent RNA- FISH signals,
443 respectively (n = 100 cells per group). Column graphs display the percentages of RNA-FISH
444 signals at the nuclear periphery versus nucleoplasm in ESCs and NPCs. **b,** Representative nascent
445 RNA-FISH images of transcriptionally active random autosomal and deterministic monoallelic
446 genes we tested in NPCs. Green fonts - imprinted genes, red fonts - stochastic monoallelic genes.
447 Scale bars - 6 µm. **c,** *Cth,* and *Ttc4* genes' nascent RNA localization at Sun1 protein territories at
448 the NP. Scale bars - 6 µm. **d,** Correlation of proteins pulled down by Sun1 antibody in mESCs and
449 NPCs. The highest enrichment was observed for the Sun1 protein, confirming the antibody's
450 specificity. The overall Pearson's correlation coefficient is 0.673, with a 95% confidence interval
451 from 0.606 to 0.73 and a p-value of 3.9e-41. **e,** Immunoblot of the Sun1 pull-down. The Sun1
452 antibody recognises five Sun1 protein isoforms (lane: Sun1-U.B). **f,** Enrichment of *Cth* and *Ttc4*
453 genes in Sun 1 chromatin pull-downs. Enrichment was assessed via qRT-PCR using primer sets
454 (*Cth*: #1 and #2, *Ttc 4*: #3 and #4 – sequences are in the method section) covering known SNPs.
455 Fold enrichment was calculated relative to input. **g,** Histogram plot of Sanger-sequenced qRT-
456 PCR products from Sun1 chromatin pull-down DNA in NPCs. Parental backgrounds of the alleles
457 were confirmed by SNPs covering primers (*Cth*: rs 220811620- primer set #2 and *Ttc4*: rs
458 28135194- primer set #3) unique to each allele. **h,** Allele-specific genome browser tracks
459 illustrating Sun1 ChIP-seq peaks. TA-bridge motif-containing peaks associated with the *Cth* and
460 *Ttc4* genes are shown in dotted boxes. Pink lines: Maternal alleles; blue lines: Paternal alleles. **i,**
461 Venn diagram showing the intersections of Sun1-associated genes (including X and Y
462 chromosome genes) identified in ChIP-seq assays and DeMA, RaMA, and all expressed genes in
463 mESCs. RNA expression data are from Balasooriya and Spector, Nat. Commun., 2022[22].

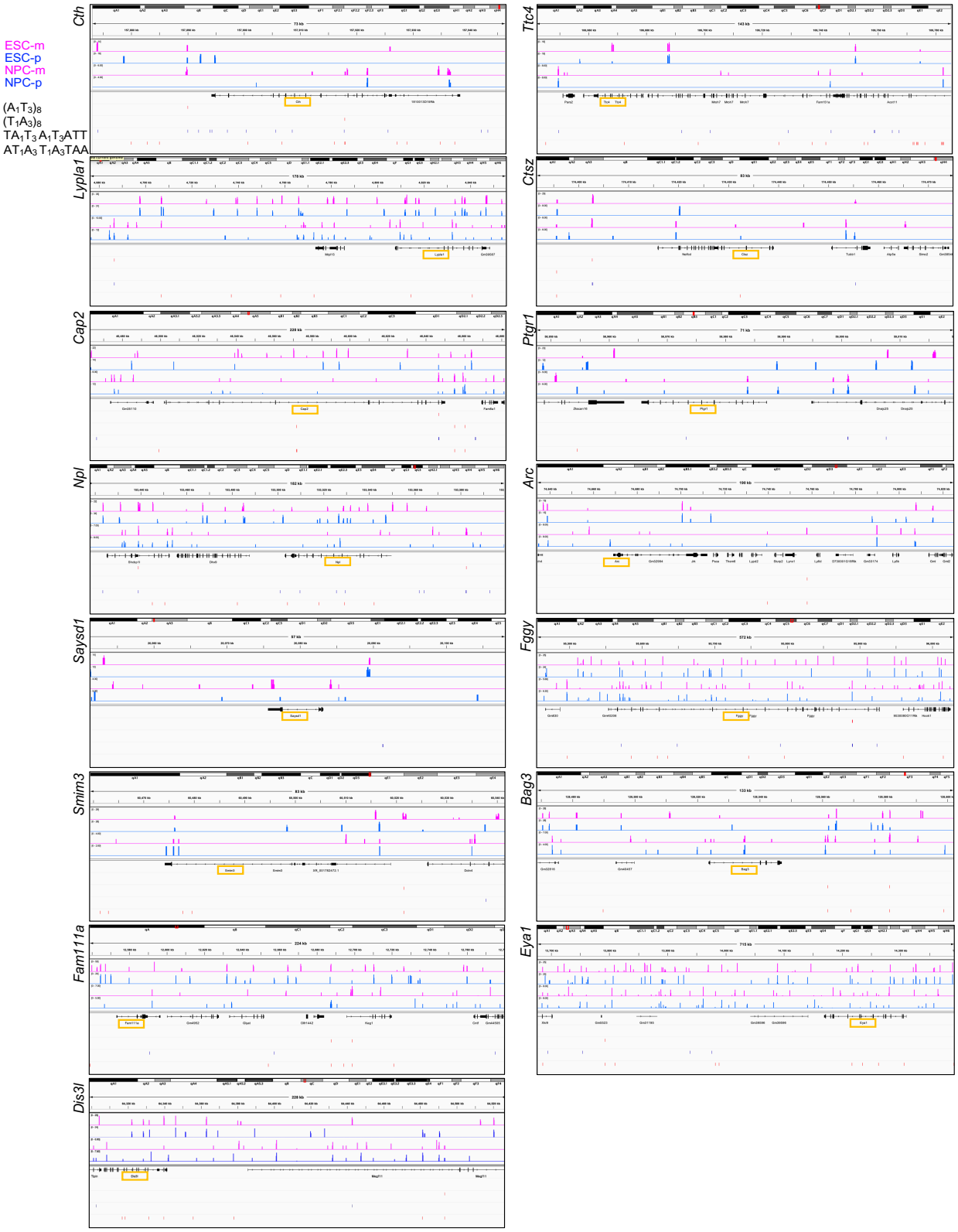
464 **Extended Data Figure 2**



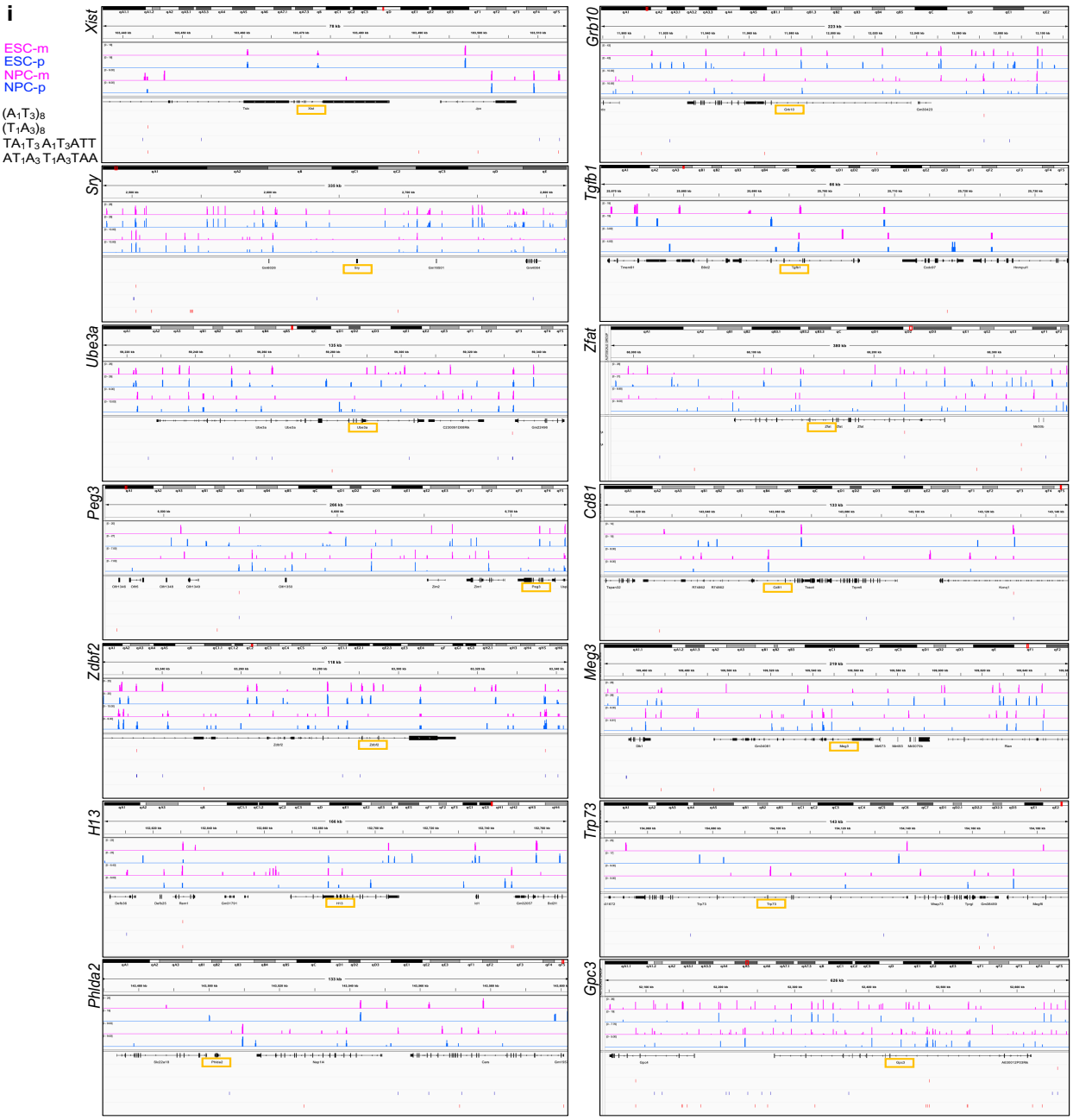
465

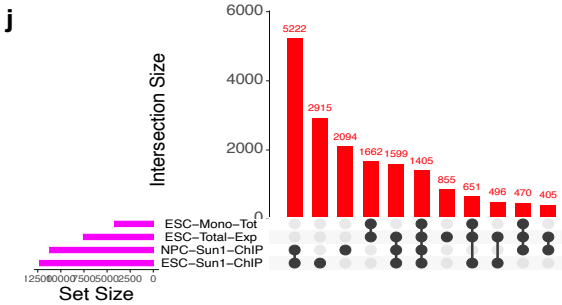
Extended Data Figure 2 Continued

h



468 **Extended Data Figure 2 Continued**



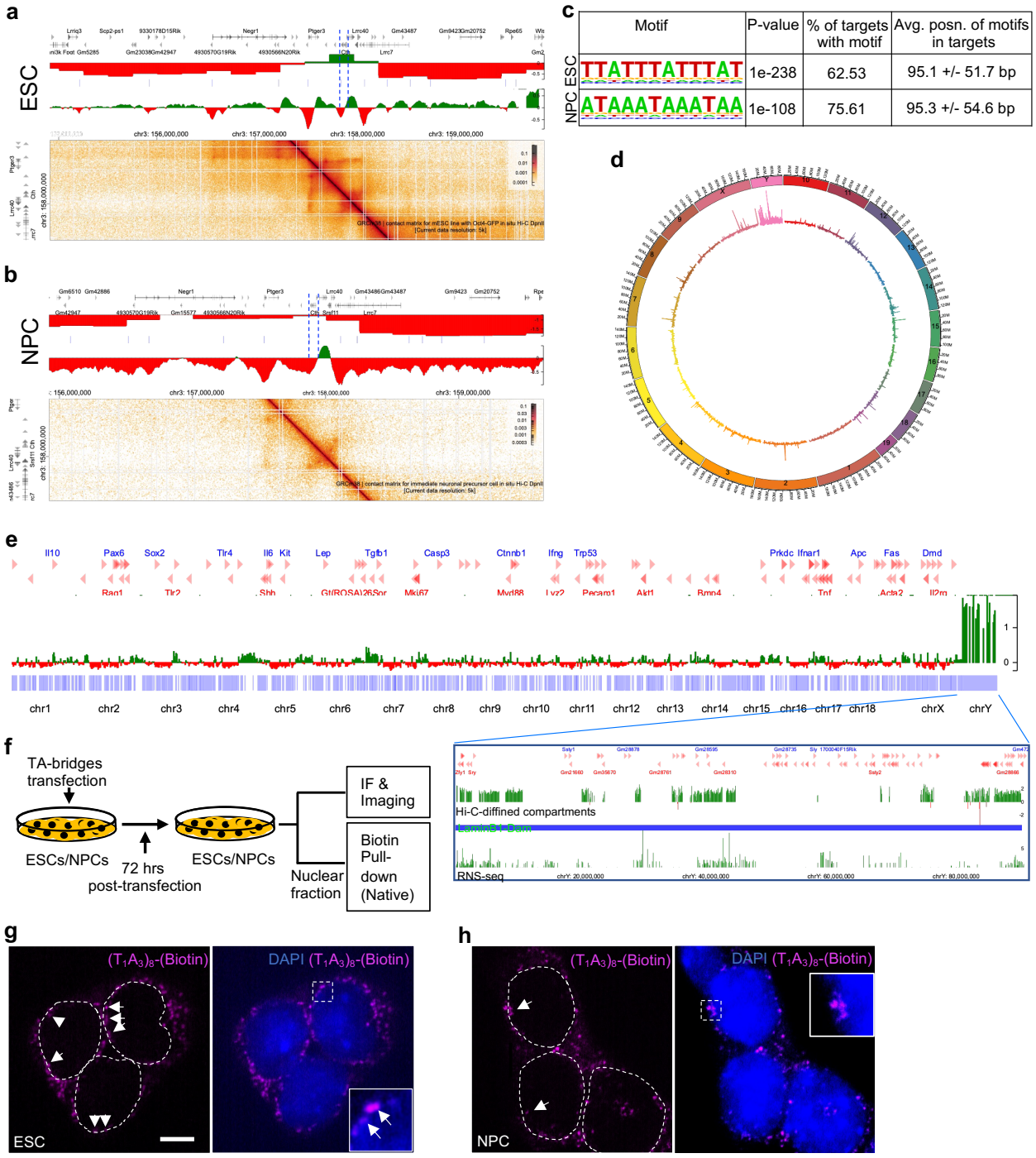


469

470 **Extended Data Figure 2 | Active alleles of monoallelic and bi-allelic genes are localized at**
471 **Sun1 protein territories.**
472 **a,** Quantification of *Actb* nascent RNA- FISH signals in the nucleus. Each line represents a counted
473 cell, and empty points on the X-axis indicate cells with one signal at the nuclear periphery and the
474 other further inside the nucleoplasm. Blue and red lines indicate nascent RNA-FISH signals
475 localized to the nuclear periphery or nucleoplasm (n=100 cells in each group). **b,** Representative
476 images from the nascent RNA-FISH assay for a panel of monoallelic genes. DeMA genes are
477 labeled in green font, and RaMA genes are labelled in red font. Scale bars - 6 µm. **c,** Quantification
478 of the nascent RNA-FISH signals for RaMA and DeMA genes in NPCs. The Y-axis shows the
479 percentage of cells with RNA-FISH signals located in the nucleus (NP: Nuclear periphery, NPl:
480 Nucleoplasm). **d,** Immunofluorescent image showing Sun1 protein localization at the nuclear
481 periphery. The enlarged dotted area is shown in the lower panel. Green arrowheads indicate Sun 1
482 and DNA overlapped loci. White arrowheads indicate Sun 1 with no DNA overlapping. Scale bars
483 - 6 µm. **e,** *Cth* nascent RNAs and Sun1 protein, and, *Ttc4* nascent RNA and Sun1 protein
484 localization at the nuclear periphery. Scale bars - 200 nm. **f,** Fold enrichment of Sun1 pull-downed
485 *Ttc4* and *Cth* genes. The same primer sets were used as in Fig. 2f. Fold enrichment was calculated
486 relative to each gene's DNA fragment enrichment in the input. **g,** Histogram of the Sanger
487 sequenced qRT-PCR products from Sun1 pull-down chromatin from mESCs using *Ttc4* primers
488 (rs28135194). Grey areas indicate the SNPs. Both maternal and paternal alleles were pulled down
489 by Sun1. **h** & **i,** Allele-specific genome browser tracks showing the ChIP-seq peak association of
490 RaMA (**h**) and DeMA (**i**) genes. Pink tracks: maternal alleles; blue tracks: Paternal alleles. The
491 positions of the shortest TA-bridge motifs (n=2) and the longest (n=8) TA-bridge motifs are shown
492 in the panels below the allele tracks. (n: number of motifs). **j,** UpSet plot showing the gene

493 intersections between Sun1-tritories associated genes in ESCs and NPCs, along with the total
494 genes expressed and monoallelic genes (DeMA and RaMA) expressed in ESCs.
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515

516 **Figure 3**



517

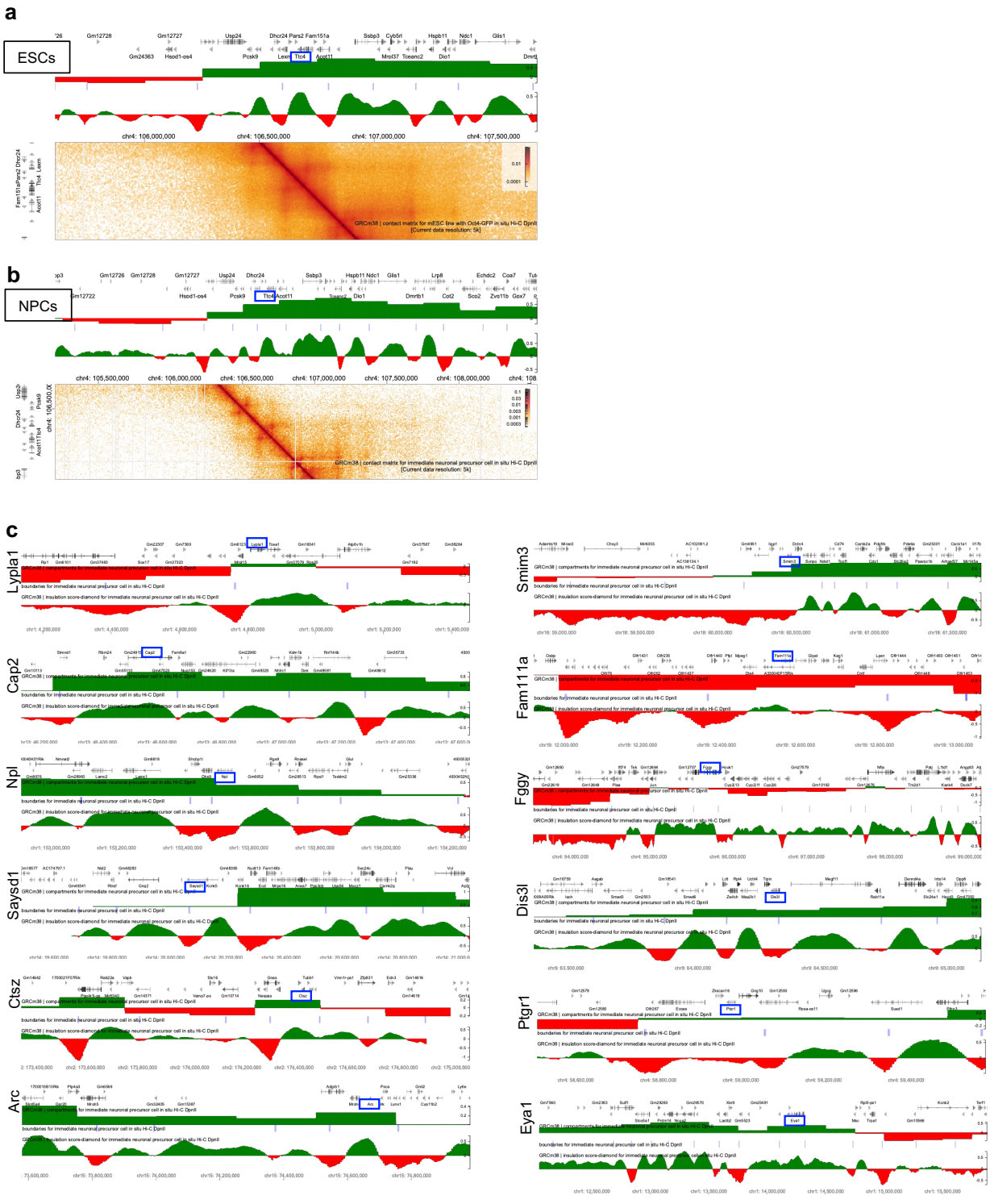
518

519 **Figure 3 | TA-bridges position the *Cth* gene in micro-TADs within the chromatin B**

520 **compartment.**

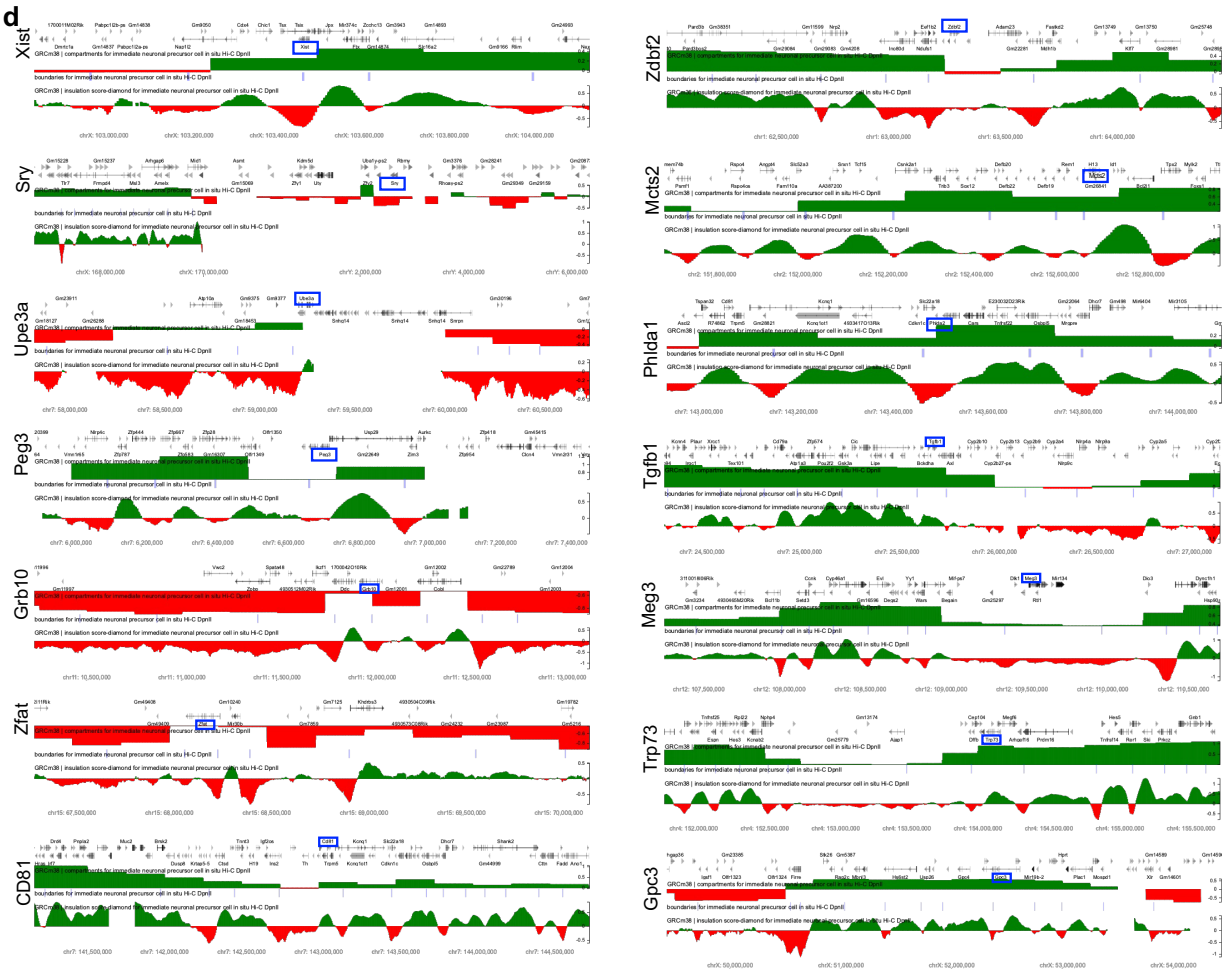
521 **a & b,** *Cth* positioning in Hi-C-defined "micro-TADS". The *Cth* gene is located in the A
522 compartment (green) within the predominantly represented B compartment (red). The contact
523 Matrix (lower panels) for each gene shows the broader TADs and the "micro-TADs" within the
524 TADs. Vertical lines indicate the boundaries of the TADs. Upon ESCs differentiation into NPCs,
525 the monoallelically expressed *Cth* allele forms an isolated A compartment within the B
526 compartment. Data is from references 59, 60, 61 and 62. **c,** *De novo* TA-bridge DNA motif. **d,**
527 genome-wide association of TA-bridges shown in circular genome representation. The Y
528 chromosome is highly enriched with TA-bridges. The height of the lines represents the motif score
529 at a specific genomic region. Motif analyzis was performed using the Motif Cluster Alignment and
530 Search tool (MCAST) with the following parameters: Sequence length: 11 bases, motif hit
531 threshold: p-value < 5.0E-4, adjacent motif spacing: less than or equal to 50, and match threshold:
532 E-value < 10.0. **e,** Genome-wide Hi-C-defined insulation score-diamond A (green) and B (red)
533 compartment distribution in ESCs. The marked lower panel shows a magnified view of the Y
534 chromosome, including its insulation score-diamond (upper track), LaminB1-DamID (middle
535 track), and RNA-seq read count density (lower track). **f,** Schematic diagram showing the
536 development of the assay used to visualize the TA-bridges spatial localisation and TA-bridge
537 interacting (*in vivo*) proteins. **g** & **h,** Transiently transfected $(T_1A_3)_8$ single-stranded DNA motif
538 localization in nuclei. mESCs and NPCs were transiently transfected with biotin-tagged $(T_1A_3)_8$
539 oligonucleotides. At 72 hrs post-transfection, cells were fixed and immunolabeled for biotin, and
540 DNA was counterstained with DAPI.
541
542
543

Extended Data Figure 3



Extended Data Figure 3 Continued

d



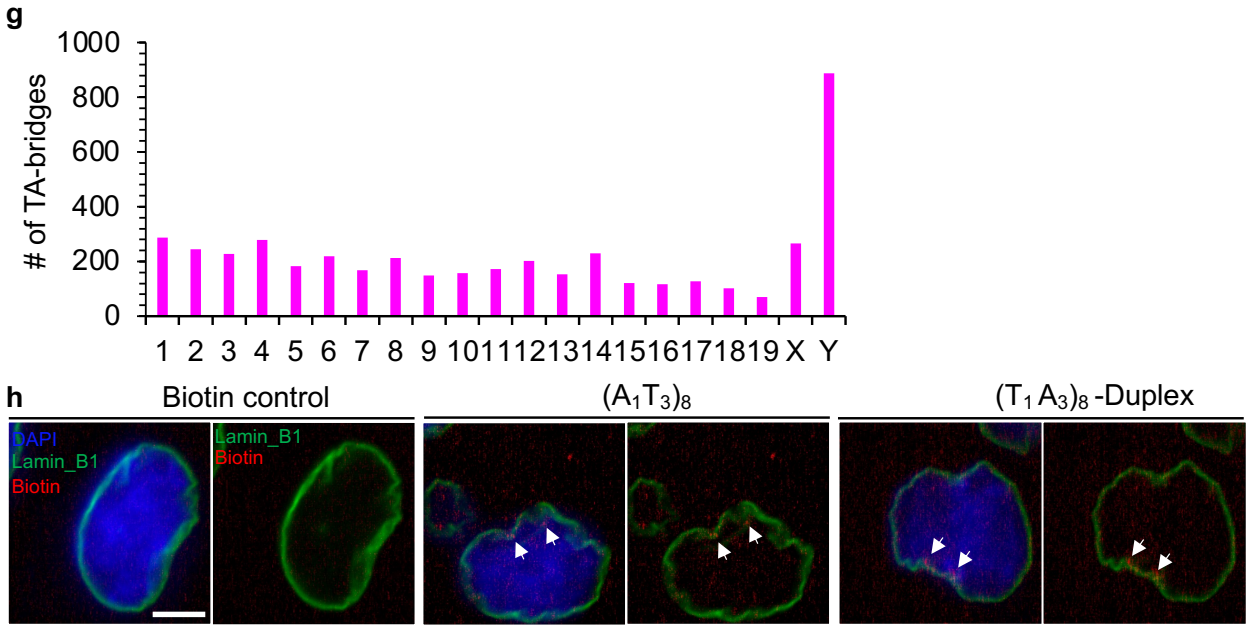
e ESC

Best Match	Motif	P-value	% of Targets	% of Background
MA0042.1_FOXI1/Jaspar	TATTTATTTATT	1e-242	63.09%	3.17%
grh/dmmpmm(Bigfoot)/fly	TCACAACCAGCT	1e-32	14.33%	1.54%
YML081W(MacIsaac)/Yeast	GAGTTTGAGGAC	1e-32	25.62%	6.07%
MA0091.1_TAL1::TCF3/Jaspar	ACCAGATGGTGG	1e-30	15.43%	2.11%
dsx/dmmpmm(Bergman)/fly	TACACTGTAGTT	1e-28	14.60%	1.98%
ZNF143|STAF(Zf)/CUTLL-ZNF143-ChIP-Seq(GSE29600)/Homer	TGCTGGGAATTG	1e-24	12.67%	1.73%
MA0430.1_YLR278C/Jaspar	TGAACTCCGG	1e-22	16.25%	3.39%
MA0403.1_TBF1/Jaspar	AGACAGGGCT	1e-21	21.49%	6.12%
MA0308.1_GSM1/Jaspar	AACTCCTGACCT	1e-20	12.95%	2.25%

f NPC

Best Match	Motif	P-value	% of Targets	% of Background
MA0042.1_FOXI1/Jaspar	ATAAATAAATAA	1e-511	84.29%	3.77%
z/dmmpmm(Bigfoot)/fly	TGGAACTCAG	1e-57	25.65%	4.52%
BORIS(Zf)/K562-CTCFL-ChIP-Seq/Homer	CCAGAAGAAGGC	1e-46	20.68%	3.62%
ASH1/Literature(Harbison)/Yeast	AGTCAGGGCTT	1e-40	26.04%	6.78%
Smad4(MAD)/ESC-SMAD4-ChIP-Seq(GSE29422)/Homer	GTCTTCAGACAC	1e-39	14.31%	1.86%
SeqBias:	GATGTTTGTGAG	1e-39	14.71%	1.99%
MAC1/Literature(Harbison)/Yeast	GAGCAACC	1e-39	42.54%	17.29%
PB0091.1_Zbtb3_1/Jaspar	CACTGTAG	1e-37	22.66%	5.51%
MA0091.1_TAL1::TCF3/Jaspar	ACCAGATGGTGG	1e-35	13.72%	1.93%

549 **Extended Data Figure 3 Continued**



550

551

552 **Extended Data Figure 3 | TA-bridge associated monoallelic gene's positioning in micro-**

553 **TADs within the B chromatin compartment.**

554 **a,** & **b,** *Ttc4* positioning in Hi-C-defined "micro-TADS". The *Ttc4* gene is located in the A

555 compartment (green). Monoallelic genes are enriched within the *Ttc4*-associated A compartment

556 that lies within the surrounding B compartment. The contact Matrix (lower panels) for each gene

557 shows the broader TADs and the "micro-TADs" within TADs. Vertical lines show the boundaries

558 of the TADs. Data is from references 59, 60, 61 and 62. **c** & **d**, Association of DeMA and RaMA

559 genes with "micro-TADs". Panel **c** shows RaMA genes; panel **d** shows DeMA genes. **e** & **f,** De

560 novo motif discovery. Narrow-peak search criteria were used to identify the shortest possible

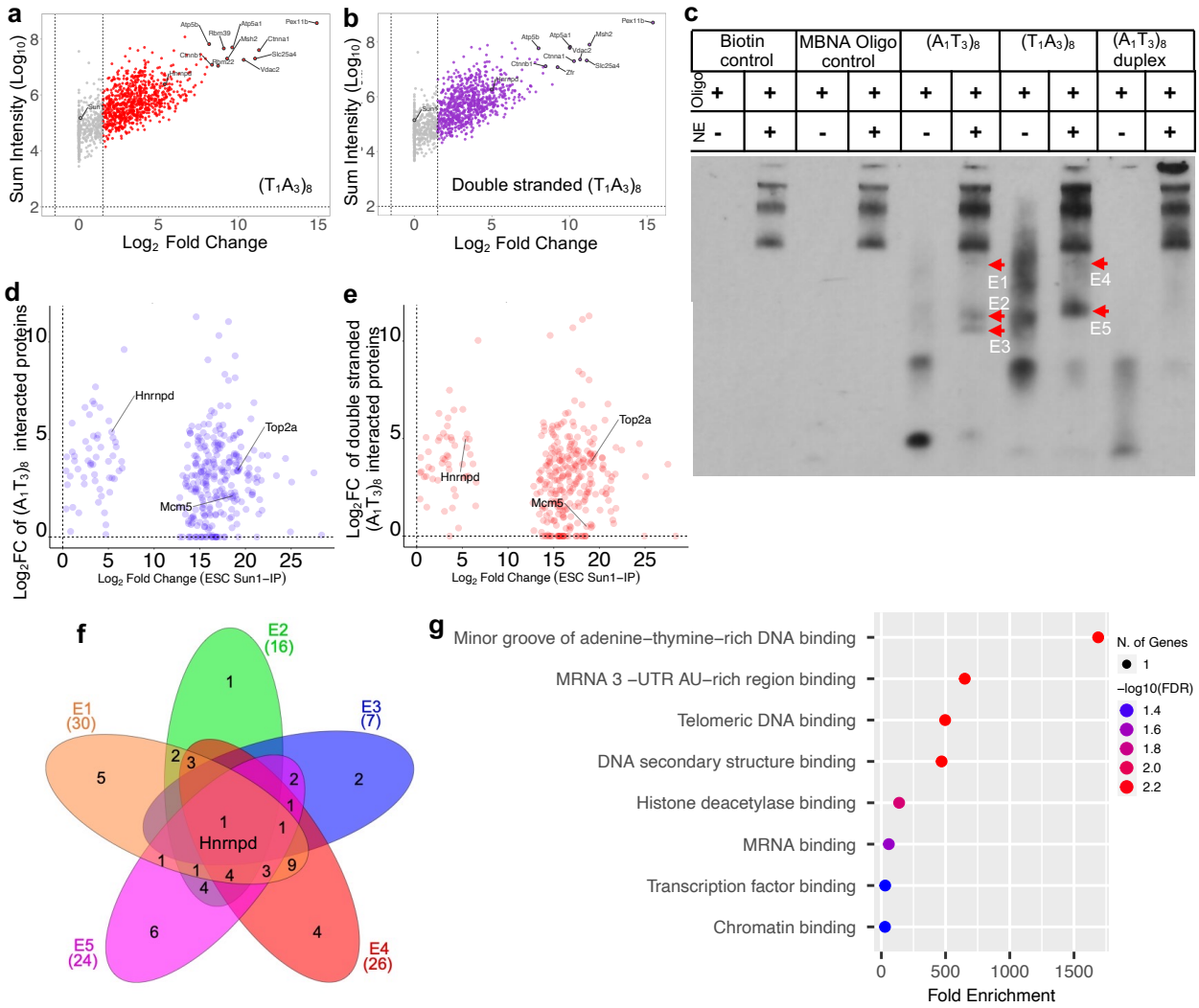
561 motifs. Only the nine most significant motifs identified in Sun1-ChIP chromatin in ESCs (**e**) and

562 NPCs (**f**) are listed. **g,** Number of TA-bridges per chromosome. The Y chromosome contains the

563 highest number of TA-bridges. **h,** Localization of single-stranded $(A_1T_3)_8$ DNA and double-

564 stranded $(T_1A_3)_8$ DNA in NPC nuclei. At 72 hrs post-transfection, cells were fixed and
565 immunolabeled for biotin, Lamin_B1 and DNA (DAPI). Lamin_B1 marks the nuclear periphery.
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586

587 **Figure 4**



588

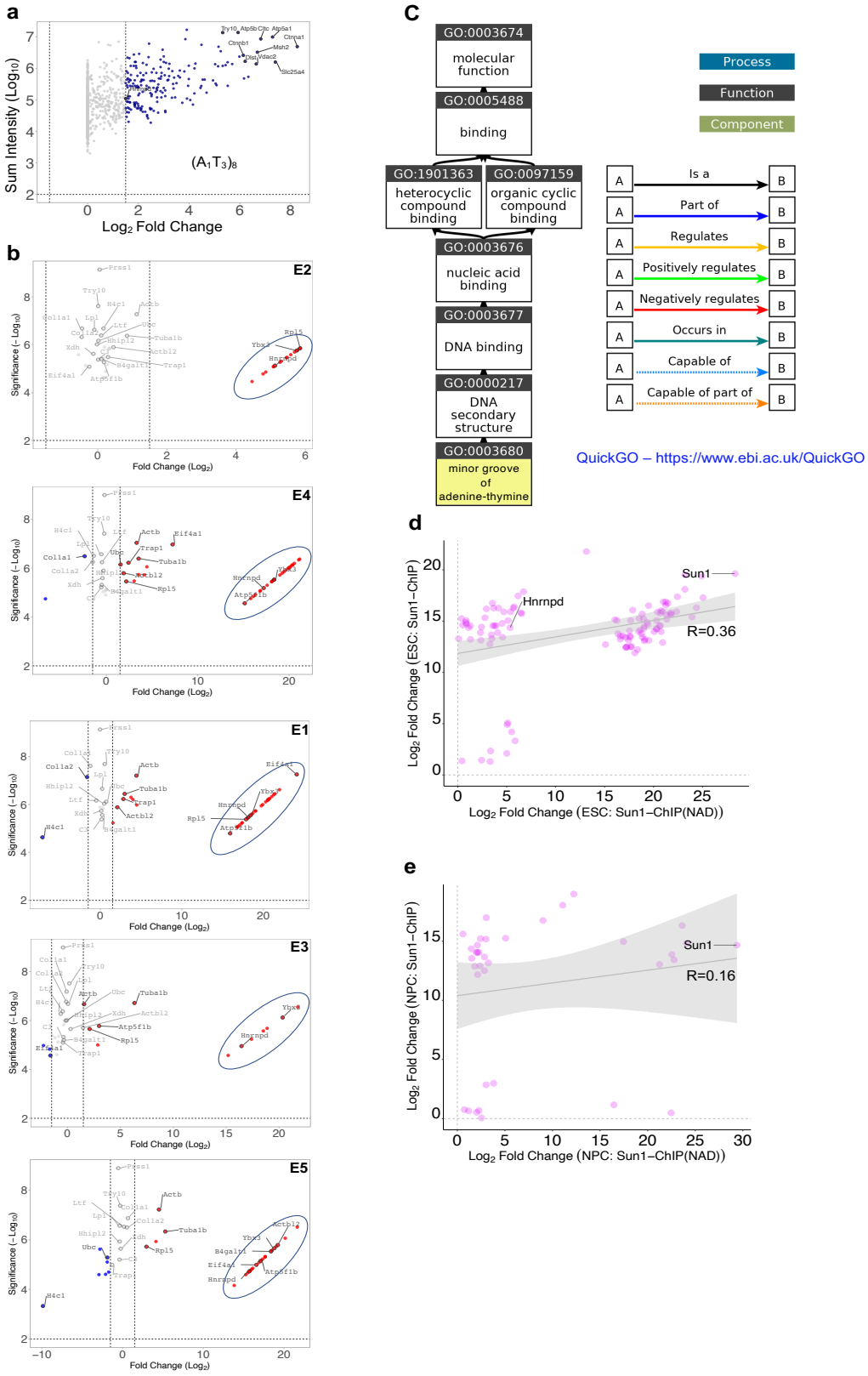
589

590 **Figure 4 | Hnrnpd interacting TA-bridges localize to Sun1 territories.**

591 **a** & **b,** Volcano plots showing in vivo interacting protein pulled down (Fig.3 f) with the single-
592 stranded $(T_1A_3)_8$ (**a**) and double-stranded $(T_1A_3)_8$ (**b**) DNA motifs. Sun1 protein was detected in
593 pull-down with both single-stranded and double-stranded $(T_1A_3)_8$ DNA motifs; however, Sun1
594 was not significantly enriched. $Log_{10}2$ of summed intensity and $Log_{10}1.5$ of fold change were used
595 as cutoff values for significant enrichment. **c.** EMSA assay for $(T_1A_3)_8$, $(A_1T_3)_8$, and double-
596 stranded $(T_1A_3)_8$ bridges. ESC nuclear extract was used. MBNA oligonucleotide supplied with the

597 EMSA kit was used as the negative control. Proteins were extracted from three excised shifted
598 bands in the $(A_1T_3)_8$ lane and from two shifted bands in the double-stranded $(T_1A_3)_8$ lane. Higher-
599 molecular-weight biotin-positive bands in all samples represent non-specific interactions. **d** & **e,**
600 Comparisons of significantly enriched proteins pulled down with Sun1, single-stranded $(T_1A_3)_8$
601 DNA, and double-stranded $(T_1A_3)_8$ DNA. Each dot represents a protein. Hnrnpd shows similar
602 enrichment for single-stranded $(T_1A_3)_8$ DNA. Each dot represents a protein. Hnrnpd shows similar
603 enrichment for single-stranded $(T_1A_3)_8$ DNA (**d**) and double-stranded $(T_1A_3)_8$ DNA (**e**). **f,** Venn
604 diagram showing the intersection of the most significantly enriched protein clusters obtained from
605 five bands excised from the EMSA assay blot. Hnrnpd is the only protein common to all excised
606 bands. **g,** Gene Ontology (molecular function) annotation for Hnrnpd. A cutoff value of FDR 0.05
607 was used as the significance threshold.
608
609
610
611
612
613
614
615
616
617
618
619

Extended Data Figure 4



622 **Extended Data Figure 4 | Hnrnpd interacting TA-bridges localize to Sun1 territories.**

623 **a,** Volcano plot showing significantly enriched in vivo-interacting proteins pulled down with the
624 $(A_1T_3)_8$ single-stranded DNA motif. Sun1 protein was not pulled down by $(A_1T_3)_8$ single-stranded
625 DNA motifs. $Log_{10}2$ of summed intensity and $Log_{10}1.5$ of fold change were used as cutoff values
626 for significant enrichment. **b,** Volcano plots showing significantly enriched proteins identified
627 from five biotin-positive shifted bands (E1-E5) excised from the EMSA blot. Hnrnpd is present in
628 all highly significant protein clusters (circled). **c,** Diagram showing functional Gene Ontology
629 annotation of Hnrnpd using QuickGO. GO term 0003680 indicates that Hnrnpd binds to the
630 adenine-thymine-rich minor groove of DNA. **d** & **e,** Correlations analysis of significantly enriched
631 proteins pulled down with Sun1 from non-nucleic-acid-depleted (**d**) and nucleic-acid-depleted (**e**)
632 nuclear extracts in ESCs and NPCs. Hnrnpd is not enriched in nucleic-acid-depleted Sun1 pull-
633 protein fractions. For ESC, the overall Pearson's correlation coefficient (R) is 0.359 with a 95%
634 confidence interval from -0.186 to 0.511 and a p-value of 1e-4. For NPC, the overall Pearson's
635 correlation coefficient (R) is 0.156 with a 95% confidence interval from -0.182 to 0.461 and a p-
636 value of 0.364.

637

638

639

640

641

642

643

644

645 **Methods**

646 **Mouse ESC (mESC) culture and *in vitro* neural progenitor cell (NPC) differentiation**

647 Polymorphic F1 mESCs were generated and maintained as previously described[22]. NPC in vitro
648 differentiation was carried out as follows. Male F1 mESCs were seeded on 0.1% gelatin-coated 6-
649 well plates at a density of 1.0 x10^5 cells per well in 2.5 ml of 2i/LIF medium. At 24 hrs post-
650 seeding, cells were washed three times with 1x PBS and switched to NDiff® 227 neural induction
651 medium (TaKaRa, Cat.No. Y40002). Cells were incubated at 37^0C, 5% CO_2. Approximately 48
652 hrs after neural induction, mESCs changed their morphology to neural-like cells. At this point,
653 cells were washed three times with 1x PBS, and 2.5 ml of NPC maintenance medium supplemented
654 with 20 ng/ml murine EGF (PeproTech Inc, Cat.No. 315-09) and 20 ng/ml murine Fgf2
655 (PeproTech Inc, Cat.No.450-33) was added to each well and cultures were incubated for an
656 additional 36 hrs. Subsequently, cells were trypsinised and reseeded at a density of 1x10^5 cells per
657 well in NPC maintenance medium to establish long-term NPC cultures.

658

659

660 **Establishment of an allele-specific gene expression visualization system**

661 Generating a parental mESC line with mCherry-PP7CP and EYFP-MS2CP Tet-On inducible
662 expression system: The Tet-On Inducible Expression System® plasmid (Clontech, Cat. No.
663 634301) was linearized by double digestion with EcoRI and AgeI. The mCherry-PP7CP DNA
664 fragment was excised from the PP7-mCherry plasmid (Spector lab) by double digestion with BaeI,
665 and adaptors compatible with the linearized Tet-On vector were ligated to the 5' end (annealed
666 Oligo#1 + Oligo#2) and to the 3' end (annealed Oligo#3 + Oligo#4), generating the Tet-On-
667 PP7CP-mCherry plasmid. The EYFP-MS2CP fragment was excised from the pMS2-cp-YFP

668 plasmid (Spector lab) using NheI and NgoMIV restriction enzymes, and adaptors compatible with
669 the linearized Tet-On vector (annealed Oligo#5 + Oligo#6 at the 3' end) were ligated to generate
670 the Tet-On-MS2CP-EYFP plasmid. Both Tet-On-PP7CP-mCherry and Tet-On-MS2CP-EYFP
671 plasmids were double-digested with BglII and PclI to excise the expression cassettes.
672 Subsequently, annealed Adpt A (Oligo#7 + Oligo#8) and annealed Adpt B (Oligo#9 + Oligo#10)
673 were ligated into the mCherry end of the excised expression cassette and the Tet-On-3G end of the
674 Tet-On-PP7CP-mCherry plasmid, respectively. Likewise, Adpt C (Oligo#11 + Oligo#12) and
675 annealed Adpt D (Oligo#13 + Oligo#14) were ligated into the EYFP end of the excised MS2CP
676 expression cassette and the Tet-On-3G end of the Tet-On-MS2CP-EYFP plasmid, respectively.
677 The resulting expression cassettes, constructed with overhanging sequences complementary to the
678 insertion site at the Rosa26 locus, were used for transfection into mESCs.

679 Oligos:

680 Oligo#1-CCGAGTGAGTCGTATTA

681 Oligo#2-TCGATAATACGACTCACT

682 Oligo#3-AAGCCAAAAAAGAAGAGAAAGGTATAA

683 Oligo#4-AATTTTATACCTTTCTCTTCTTTTTTGGCTTGGACA

684 Oligo#5-CTAGTAATACGACTCACTATAGG

685 Oligo#6-AATTCCTATAGTGAGTCGTATTA

686 Oligo#7-GATCGGATGGATCCATCGGAAGGGGAGCCTTTCTCTCTGGGCAAGAGCGGT
687 GCAATGGTGTGTAAAGGTAGCCGCTC

688 Oligo#8-CGATGGATCCATCC

689 Oligo#9- GGATGGATCCACTG

690 Oligo#10-CATGCAGTGGATCCATCCTCTCGGCTACCTTTACACACCATTGCACCGCTC

691 TTGCCCAGAGAGAAAGGCTCCCCTTC

692 Oligo#11-GATCGGATGGATCCATCGTCTGAAGGAGAGCCTTTC

693 Oligo#12- CGATGGATCCATCC

694 Oligo#13-CGATGGATCCATCC

695 Oligo #14-CATGCAGTGGATCCATCCGAAAGGCTC**T**CCTTCATC

696

697 Designing allele-specific gRNAs for expression cassette knock-in at the Rosa26 locus: Allele-
698 specific gRNAs were designed to incorporate SNPs in the Rosa26 (R26) alleles at the PAM site
699 and were cloned into pSpCas9n(BB) (PX460) (Addgene, Cat.No. 48873) plasmid. The following
700 gRNA sequences were used in combination. gR26#1 (annealed R26-5'Cast1.1 and R26-
701 5'Cast1.2), gR26#2 (annealed R26-3'Cast-Comp1.1 and R26-3'Cast-Comp 1.2), gR26#3
702 (annealed R26-5'B6 1.1 and R26-5'B61.2), gR26#4 (annealed R26-3'B61.1 and R26-3'B61.2).
703 Next, the indicated combinations of gRNAs and adaptor-ligated expression cassettes were
704 transfected into F1 mESCs using the Xfect transfection kit (Takara, Cat.No.631317). To target the
705 Tet-On-PP7CP-mCherry expression cassette to the R26 CAST (paternal) allele, gR26#1, gR26#2
706 and Tet-On-PP7CP-mCherry cassettes were co-transfected. To targt the Tet-On-MS2CP-EYFP to
707 the R26 C57 (maternal) allele, gRNA#3, gRNA#4, and the Tet-On-MS2CP-EYFP were co-
708 transfected. For each transfection, 2 μg of gRNA-Cas9 plasmids and 1.0 μg of adaptor-ligated
709 cassette DNA were used. Prior to transfection, mESCs were seeded at $0.5x10^5$ cell/ well in 6-well
710 Corning tissue culture plates (Corning, Cat. No. 353046) and cultured for 12 hrs. At 24 hours post-
711 transfection, cells were rinsed with KnockOut™ DMEM (Gibco, Cat. No. 10829-018) containing
712 10% FBS, then switched to fresh 2i medium. At 48 hours post-transfection, single clones were
713 picked under a microscope and expanded.

714 Oligos:

715 R26-5'Cast1.1-CACCGAGTGGGGTCGACTATCTGAAGGG

716 R26-5'Cast1.2-AAACCCCTTCAGATAGTCGACCCCACTC

717 R26-3'Cast-Comp1.1-CACCGGCTTGCCCTTTTCGTCTTCTCGG

718 R26-3'Cast-Comp1.2-AAACCCGAGAAGACGAAAAGGGCAAGCC

719 R26-5'B61.1-CACCGTAGTGGGGTCGACTATCTGAAGG

720 R26-5'B61.2-AAACCCTTCAGATAGTCGACCCCACTAC

721 R26-3'B61.1-CACCGCCGCTCTTGCCCAGAGAGAAAGG

722 R26-3'B61.2-AAACCCTTTCTCTCTGGGCAAGAGCGGC

723

724 MS2 and PP7 stem-loops construction: The pCR4-24xMS2SL (Addgene, Cat. No.31865) and

725 pCR-24xPP7SL (Addgene, Cat. No. 31864) plasmids were double-digested using SpeI and NotI,

726 and the 24xMS2SL (1352bp) and 24xPP7SL (1492bp) DNA fragments were gel-purified.

727 Annealed adapters were ligated into each of the excised stem-loop DNA fragments as follows:

728 Ttc4 gene C57a allele (Lig#1): annealed Ttc4-B6-5.1 + Ttc4-B6-5.2 and annealed Ttc4-B6-3.1 +

729 Ttc4-B6-3.2.

730 Ttc4 gene CAST allele (Lig#2): annealed Ttc4-CAST-5.1 + Ttc4-CAST -5.2 and annealed Ttc4-

731 CAST -3.1 + Ttc4-CAST -3.2.

732 Cth gene C57 allele (Lig#3): annealed Cth-B6-5.1 + Cth-B6-5.2 and annealed Cth-B6-3.1 + Cth-

733 B6-3.2.

734 Cth gene CAST allele (Lig#5): annealed Cth-CAST-5.3 + Cth-CAST-5.4 and annealed Cth-

735 CAST-3.3 + Cth-CAST-3.4.

736 Oligos:

737 Ttc4-B6-5.1 TAATACGACTCACTATAGG

738 Ttc4-B6-5.2

739 GGCCCCTATAGTGAGTCGTATTAAAAGGGAAAACAACACAAGCCTACTTGCAAACC

740 ACTGCCCATTGACTCCTCAAGCTTCTGTCAGATGTGGCACAAGCCCTTGCATGGGCA

741 TTGAGAAGTAGAGAATCCAGTAGGGCTCTGAGTAGTCCCAGGATCCAAGT

742 Ttc4-B6-3.1

743 CTAGTAATACGACTCACTATAGGACTTGGATCCTGGGACTACTCAGAGCCCTACTGG

744 ATTCTCTACTTCTCAATGCCCATGCAAGGGCTTGTGCCACATCTGACAGAAGCTTGA

745 GGAGTCAATGGGCAGTGGTTTGCAAGTAGGCTTGTGTTGTTTTCCCTTT

746 Ttc4-B6-3.2 CCTATAGTGAGTCGTATTA

747 Ttc4-CAST-5.1 AATTAACCCTCACTAAAGGG

748 Ttc4-CAST-5.2 CTAGCCCTTTAGTGAGGGTTAATTACAGGATGTGGCTTCATGGCTCC

749 AGTC

750 Ttc4-CAST-3.1 GGCCAATTAACCCTCACTAAAGGGACTTGGATCCTGGGACTACTCA

751 GAGCCCTACTGGATTCTCTACTTCTCAATGCCCATGCAAGGGCTTGTGCCACATCTG

752 ACAGAAGCTTGAGGAGTCAATGGGCAGTGGTTTGCAAGTAGGCTTGTGTTGTTTTCC

753 CTTTATCGTGAGGCCTCAT

754 Ttc4-CAST-3.2 CCCTTTAGTGAGGGTTAATT

755 Cth-B6-5.1 AGAACTGTCTTCTGTTTATCTTCTAACTAACAGGTTGTTCTGTTAGTAT

756 CATTTCGGTAATTTTGCTATATTTGTGTCCAAGGAAGTAAGAGTTGTTCTGTAATACG

757 ACTCACTATAGG

758 Cth-B6-5.2 GGCCCCTATAGTGAGTCGTATTA

759 Cth-B6-3.1 CTATAGTGAGTCGTATTAGATC

760 Cth-B6-3.2 TAATACGACTCACTATAGGTTCATTCTCAACAAGACAAAACCCTAGTAC
761 AACAGAAAAAAGAAAGAAGTCATCGAATACTATATAAAATTAGTACAAATGTTACA
762 ATGTTAAATCATAACTACAAAATACTTCAATTTAATAATTTACTTACCAGAATTTAG
763 GTGAC
764 Cth-CAST-5.3 AATTAACCCTCACTAAAGGG
765 Cth-CAST-5.4 CTAGCCCTTTAGTGAGGGTTAATTCCTTGGACACAAATATAGCAAAA
766 TTACCGAAATGATACTAACAGAACAACCTGTTAGTTAGAAGATAAACAGAAGACAG
767 TCCTTAA
768 Cth-CAST-3.3 GGCCAATTAACCCTCACTAAAGGGTAAAGGTTGTACTTGATACTTGTT
769 ATTTTTCTTAAATAAATCCAGTT
770 Cth-CAST-3.4 CCCTTTAGTGAGGGTTAATT
771
772 gRNA cloning for allele-specific insertion of stem-loops: To clone gRNAs into pSpCas9n(BB)
773 (PX460) plasmids, we generate plasmid libraries by cloning allele-specific gRNA pairs as
774 described below.
775 T8: Ttc4-C57BL-gRNA-1.1: + Ttc4-C57BL-gRNA-1.2:
776 T9: Ttc4-C57BL-gRNA-2.1: + Ttc4-C57BL-gRNA-2.2:
777 T10: Ttc4-CAST-gRNA-3.1: + Ttc4-CAST-gRNA-3.2:
778 **T11:** Ttc4-CAST-gRNA-3.1a: + Ttc4-CAST-gRNA-3.2a:
779 T12: Ttc4-CAST-gRNA-4.1a: + Ttc4-CAST-gRNA-4.2a:
780 T13: Ttc4-CAST-gRNA-4.1: + Ttc4-CAST-gRNA-4.2:
781 C1: Cth-C57BL-gRNA-1.1: + Cth-C57BL-gRNA-1.2:
782 C2: Cth-C57BL-gRNA-2.1: + Cth-C57BL-gRNA-2.2:

783 C3: Cth-CAST-gRNA-3.1: + Cth-CAST-gRNA-3.2:

784 C4: Cth-CAST-gRNA-3.1a: + Cth-CAST-gRNA-3.2a:

785 C5: Cth-CAST-gRNA-4.1a: + Cth-CAST-gRNA-4.2a:

786 C6: Cth-CAST-gRNA-4.1b: + Cth-CAST-gRNA-4.2b:

787 C7: Cth-CAST-gRNA-4.1: + Cth-CAST-gRNA-4.2:

788 Ttc4-C57BL-gRNA-1.1: CACCGTTCTGTTGACTGTATAAACTTGG

789 Ttc4-C57BL-gRNA-1.2: CAAGACAACTGACATATTTGAACCCAAA

790 Ttc4-C57BL-gRNA-2.1: CACCGAAATGAGACCTCACAATAAAGGG

791 Ttc4-C57BL-gRNA-2.2: CTTTACTCTGGAGTGTTATTTCCCCAAA

792 Ttc4-CAST-gRNA-3.1: CACCGTGTGGATGCCATGAAGTGACTGG

793 Ttc4-CAST-gRNA-3.2: CACACCTACGGTACTTCACTGACCCAAA

794 Ttc4-CAST-gRNA-3.1a: CACACCTACGGTACTTCACTGACCCAAA

795 Ttc4-CAST-gRNA-3.2a: CAAGACAACTGATATATTTGAACCCAAA

796 Ttc4-CAST-gRNA-4.1a: CACCGGTTTATATAGTCAACAGAACAGG

797 Ttc4-CAST-gRNA-4.2a: CCAAATATATCAGTTGTCTTGTCCCAAA

798 Ttc4-CAST-gRNA-4.1: CACCGCTACTCCTATTTCAGAAATGAGG

799 Ttc4-CAST-gRNA-4.2: CGATGAGGATAAAGTCTTTACTCCCAAA

800 Cth-C57BL-gRNA-1.1: CACCGAAACAGAAGACAGTTCTTAAGGG

801 Cth-C57BL-gRNA-1.2: AAACCCCTTAAGAACTGTCTTCTGTTTC

802 Cth-C57BL-gRNA-2.1: CACCGTGGTCTTAAATCCACTGTGGTGG

803 Cth-C57BL-gRNA-2.2: AAACCCACCACAGTGGATTTAAGACCAC

804 Cth-CAST-gRNA-3.1: CACCGCTTCCATGGATCTTCCCTTAAGG

805 Cth-CAST-gRNA-3.2: AAACCCTTAAGGGAAGATCCATGGAAGC

806 Cth-CAST-gRNA-3.1a: CACCGCAGAACAACTCTTACTTCCTTGG

807 Cth-CAST-gRNA-3.2a: AAACCCAAGGAAGTAAGAGTTGTTCTGC

808 Cth-CAST-gRNA-4.1a: CACCGAGTAAGAGTTGTTCTGTTTTGGG

809 Cth-CAST-gRNA-4.2a: AAACCCCAAAACAGAACAACTCTTACTC

810 Cth-CAST-gRNA-4.1b: CACCGAAACCACCACAGTGGATTTAAGG

811 Cth-CAST-gRNA-4.2b: AAACCCTTAAATCCACTGTGGTGGTTTC

812 Cth-CAST-gRNA-4.1: CACCGAAGAAAAATAACTTGATAACTGG

813 Cth-CAST-gRNA-4.2: AAACCCAGTTATCAAGTTATTTTTCTTC

814

815 Allele-specific MS2 and PP7 stem-loop knock-in to the *Cth* and *Ttc4* genes: A single-cell-derived
816 mESC clone expressing mCherry-PP7CP and EYFP-MS2CP was transfected with a mixture of
817 DNA constructs and gRNA-Cas9 plasmids using the Xfect transfection reagent to knock in stem
818 loops to individual alleles. Prior to transfection, $0.5x10^5$ mESCs per well were plated in 6-well
819 plates and cultured for 12 hours. To target the *Cth* gene alleles, Lig#3 and Lig#5 (3.45 μg of DNA
820 in each) were co-transfected with gRNAs C1, C2, C3, C4, C5, C6 and C7 (0.3 μg of plasmid each).
821 To target the *Ttc4* gene alleles, Lig#1 and Lig#2 (3.6 μg of DNA each) were co-transfected with
822 gRNAs T8, T9, T10, T11, T12 and T13 (0.3 μg of plasmid each). Twelve hours after transfection,
823 the medium was replaced with fresh 2i medium. At 24 hrs post-transfection, single colonies were
824 picked under a light microscope, expanded and genotyped using Cth-Wt gene F
825 (AAAGTTCGAGTVAAAGCTGTCAT and Cth-Wt gene R
826 (TCAAGTACAACCTTTAATCTCTTGAA) for Cth clones, and Ttc4-Wt gene F
827 (GGCAGTGGTTTGCAAGTAGG) and Ttc4-Wt gene R

828 (CCAGCTGGTTTCAGTTTAATAATGT) for *Ttc4* clones. The mESC clones for both *Cth* and
829 *Ttc4* were further validated for on-target integration by DNA-FISH assay.
830
831
832 **SDS-polyacrylamide gel electrophoresis and immunoblotting**
833 Protein electrophoresis was performed using 1.0 mm-thick and 8x10 cm Tris-glycine SDS-
834 polyacrylamide gels cast with 5% Stacking gel (30% acrylamide mix, 1.0 M Tris (pH6.8), 10%
835 SDS, 10% ammonium persulfate, 0.1% TEMED in dH_2O) and 12 % resolving gel (30%
836 acrylamide mix, 1.5 M Tris (pH8.8), 10% SDS, 10% ammonium persulfate, 0.1% TEMED in
837 dH_2O). Gels were run in 1x Tris-glycine SDS running buffer, and proteins were transferred in ice-
838 cold 1x transfer buffer onto Immobilon®-P Transfer membranes (Merck Millipore, Cat.No.
839 IPVH00005). The transfer membranes were air-dried to dehydrate and stored until use. For
840 immunoblotting, membranes were rehydrated in 0.1% PBST (0.1% Tween-20 in 1xPBS),
841 changing the buffer several times until the membranes were fully wetted. Rehydrated membranes
842 were blocked for 1 ½ hours in 5% blocking buffer containing skim milk (Sigma, Cat.No.
843 1153630500) prepared in 0.1% PBST and then incubated overnight with primary antibodies
844 diluted in the same blocking buffer. The next day, membranes were washed four times in 0.1%
845 PBST for 10 minutes each and incubated for 2 hours at room temperature with HRP-conjugated
846 secondary antibody buffer (3% milk in 0.1% PBSTX). After the incubation, membranes were
847 washed four times for 10 minutes each with 0.1% PBST, followed by a final wash in 1x PBS. ECL
848 reagent (Thermo Fisher, Cat. No. PI32209) was added to fully cover the membrane, and
849 chemiluminescent signals were captured on X-ray film (Amersham Hyperfilm™ ECL, GE
850 Healthcare Limited, Cat.No. 28906836). Primary antibodies used: Goat anti-YFP polyclonal

851 antibody (1:500) (ORIGENE, Cat.No. AB1166-100); mouse anti-mCherry antibody (1:500)
852 (Thermo Fisher, Cat.No. MA5-32977); Histone H3 antibody (1:1000) (Novus Biological, Cat. No.
853 NB500-171-0.025 ml). Secondary antibodies: Donkey anti-Goat IgG (H+L), HRP-conjugated
854 (1:500) (Thermo Fisher, Cat.No. A15999); Donkey anti-Mouse IgG (H+L), HRP-conjugated
855 (1:500) (Abcam, Cat. No. AB205719).
856
857
858 **Immunofluorescence assay**
859 Cells were cultured on ibidi u-Slide 8-well glass-bottom chambers (ibidi GmbH, Cat. No. 80827).
860 Cells were washed once with 1x PBS and then fixed in 4% PFA for 12 minuites with gently
861 rocking. PFA was removed by rinsing three times with 400 µl of 1x PBS. Cells were
862 permeabilized with 0.3% PBSTx (0.3% Triton X-100 in 1xPBS) for 10 minuites at RT with gental
863 rocking, followed by three washes with 400 µl of 0.1% PBSTx. Cells were blocked for one houre
864 at RT in blocking buffer (3% BSA, 10% NDS, 0.1% PBSTx). Then 200 µl of blocking buffer
865 containing primary antibodies was added to each well, and incubated overnight at 4ºC on a shaker.
866 The next day, primary antibodies were removed by washing four times with 400 µl of 0.1% PBSTx
867 for 10 minuits each. Secondry antibodies diluted in buffer (5% NDS in 0.1% PBSTx) were added
868 at 200 µl per well and incubated for 1 ½ hours at RT on a slow rocker. After incubation, secondary
869 antibodies were removed by washing with 0.1% PBSTx three times, fllowed by a final wash with
870 1X PBS. DAPI diluted 1:10000 in 1x PBS (200 ul) was added to each well and incubated for 5
871 minuites, followed by three washes with 1x PBS. A total of 100 µl of ibidi mounting medium (ibidi
872 GmbH, Cat.No. 50001) was added to each well, and chamber slides were wrapped in aluminium
873 foil and stored at 4ºC until imaging. Primary antibodies: Sun1 (1:700, gift from Min lab), anti-

874 Biotin (1: 1000) (Santa Cruz, Cat. No. 101339), Lamin_B1 (1:500) (Abcam, Cat. No. ab16048).
875 Secondary antibodies: Donkey anti-Rabbit AF568 (Thermo Fisher, Cat.No. A10042), Donkey
876 anti-Mouse, AF647 (Thermo Fisher, Cat.No. A-21235), Goat anti-Rabbit AF488 (Thermo Fisher,
877 Cat. No. A-11008). Secondary antibodies were used at 1:1000 dilution.
878
879
880 **Live imaging using lattice light-sheet microscopy and quantitation**
881 mESCs were cultured overnighrt on small round NO. 1 cover glass, 5 mm diameter (Thomas
882 Scientific, Cat.No.1217H19). Four hourse before image acquisition, medium was replaced with
883 medium containing 20 ng/ml doxycycline (Sigma, Cat.No. 324385-1GM), and the same medium
884 condition was used during image acquisition. Live images was performed using a 3i lattice light-
885 sheet microscope (LLSM) with the following parameters: 62.5x LLS objective, magnification
886 changer 1.0x, binning 1x1, average time-lapse interval 20 seconds, pixel size (XY): 0.104 μm, Z-
887 step 0.553 μm. Channel 1: 488nm excitation at 300 ms; Channel 2: 560nm excitation at 300 ms.
888 The system uses a 25x/ 1.1 NA detection objective, with additional optical manginification to
889 achive a final magnification of 62.5x. Each laser was set to 9% AOTF output, with a laser head
890 power of 50 mW each. Images were deskewed and deconvolved using the algorithms embedded
891 in SlideBook6 software.
892
893
894 **Nick translated probe preparation**
895 One microgram of PCR-generated DNA probe was used in each reaction. The reaction mixture
896 (total 45 μl) contained DNA fragments, 3.8 μl labeled dUTP (0.2 mM), 2.5 μl of 0.1 mM dTTP,

897 10 µl (0.1 mM) dNTP mix, 5.0 µl of 10x Nick Translation buffer, and nuclease-free ddH_2O
898 (volume up to 45 µl). The mixture was vortexed and briefly spun down, followed by addition of
899 5.0 µl of Nick Translation enzyme. Sampels were gently vortexed and centrifuged, then incubated
900 at 15ºC for 10 hours, heat-inactivated at 70ºC for 10 minutes, and cooled to 4ºC. Samples were
901 transferred to 1.5 ml RNase-free microcentrifuge tubes, and DNA was precipitated by adding 1 µl
902 of 0.5 M EDTA, 1 µl of linear acrylamide (Ambion), 5 ul of 3M NaOAc (pH5.2), and 125 µl of
903 cold (-20ºC) 100% ethanol, followed by mixing and incubation at -80ºC for 2 hours. Samples
904 were centrifuged at 20000 rcf for 60 minuites at 4ºC to pallet DNA. The pellet was washed twice
905 with 1 ml of 75% ethanol, centrifuged at 14000 rpm for 5 minuites each time. Pellet were dried at
906 37ºC for 15 minuites, resuspended in 50 µl of DEPC-treated DNase/RNase free H_2O, and
907 incubated at 37ºC on a heat block with shaking for 1 hour (at 200 rpm). Probes were stored at -
908 20ºC until use.
909
910
911 **Nascent RNA-FISH assay**
912 NPCs were cultured in 24-well plates (MatTek, Cat.No. P24G013F) and used at ~80% confluency.
913 All buffers were prepared using DEPC-treated 1x PBS. Freshly prepared 4% PFA (in DEPC-
914 treated PBS) was chilled on ice, then warmed at 37ºC for 30 minutes before use. Cells were rinsed
915 once with 1x PBS, fixed with 200 µl of 4% PFA for 15 minutes at RT with gentle shaking, and
916 washed three times with 1x PBS (10 minutes each). A 200 mM VRC (stock) was warmed at 60ºC
917 before used. A 0.5% Triton X-100/ 2 mM VRC permiabalization buffer was in 1x PBS, and 300
918 µl was added per well wth incubation on ice for 10 minutes. Cells were washed three times with
919 1x PBS, then saturated with 400 µl of 2x SSC for 2 hours at 37ºC in a humidified chamber. Probe

920 preparation: A mixture of 100 ng DNA probe and 2 µl competitor yeast tRNA (10 µg/ul) was dried
921 in SpeedVac at maximum heat for 25 minutes. Dried probes were dissolved in 10 ul deionized
922 formamide (Ambion), incubated at 37ºC for 20 minutes, denatured at 95ºC for 10 minutes, and
923 chilled on ice for 10 min. Hybridization buffer was prepared by mixing 4 µl warmed 50% dextran
924 sulfate (42ºC), 3 µl DEPC-treated ddH_2O, 1 µl RNasin (Promega), and 2 µl 20x SSC. Ten
925 microliters of hybridization buffer was added to a each probe tube and mixed gently. A total of
926 40 µl of probe mixture was added to each well, sealed with Parafilm, wrapped in foil, and incubated
927 for 16 hours at 37ºC in a humidified incubator. After hybridization, cells were washed three times
928 (5 minutes each) with pre-warmed (42ºC) 2x SSC/ 50% deionized formamide, then sequentially
929 washed with: pre-warmed (42ºC) 2x SSC three times (5 minutes each), pre-warmed (37ºC) 1x SSC
930 three times (5 minuits each), 0.5% SSC three times (5 minutes each). Cells were then incubated
931 with 200 µl of DAPI in 4x SSC for 5 minutes and washed once with 4x SSC. Finally, 100 µl ibidi
932 mounting medium (ibidi GmbH, Cat.No.50001) was added per well, and plates were wrapped in
933 aluminum foil and stored at 4ºC until imaging.
934 Primers used for PCR amplification to generate nick translated probes:
935 *Acyp2*: F- GGGCCGACTTTTCTAACGA, R- TGAATGACTCCATAAAACAACATT
936 *Arc*: F- GTGAAGACAAGCCAGCATGA, R- CTCCAGGGTCTCCCTAGTCC
937 *Fggy*: F-CCTGGCATTGCAAAATATGA, R- CCTGGATCCAATGAGTGAGC
938 *Bag3*: F-CGAGAGCCTCCACCTGTTAC, R- CCAGACAACCAGAGGGGTTA
939 *Eya1*: F- TGTCAGTGCTGGACTTCAGG, R- CTGCTGCCCCCAGTATTAAA
940 *Dis3l*: F- CGCTGCCACTCTGATACAAT, R- TCTGAAGACAGAGGCAAGAGG
941 *Lypla1*: F- AGTCACCTCCTCCTGCTGAA, R- CCAGACAACCAGAGGGGTTA
942 *Cap2*: F- GCCTTCCCTGAGTCTCCTCT, R- GCGCTATGCTTCGTTACACA

943 *Cth*: F- CCTCTGTGCCTGAGAAGGAC, R- TGAATTTCACAGTAGGTATTCCAAA

944 *Npl*: F- CATCATGACGCTGGTCTCTG, R- CCCAAAAGGCAGGTGAAAT

945 *Saysd1*: F- GTCCTGCTGGGACTGTTTGT, R- TCCTCAGGCACATGGTACAC

946 *Smim3*: F- CTATCACCCAATCCCCTGTG, R- TGGGTTGTAGCTCAGTGGTG

947 *Fam111a*:F- GGGACCAAAGACCACTTTCA, R- GATCCAGAAGACCCACCAAA

948 *Ctsz*: F- TCGTGACCAGCACCTACAAG, R- GGATTCCAGGGCTCCAGAT

949 *Ptgr1*: F- TCCCTGGAGGGTTAGCTTTT, R- AATGGGCTTAGGGAATGGAC

950 *Cd81*: F- ATTCTGAGCATGGTGCTGTG, R- CTTGAAATGGAGGCAGGAAA

951 *Meg3*: F- TCAAAGGGCTGAGGAGAAAA, R- AAGTGGTCAGATTGCGAAGC

952 *Xist*: F- TGTGTATTGTGGGTGTGTCTATTT, R- AGTCACGAAAGGCAGGAGTG

953 *Gpc3*: F- CAAGCAGCATGGAAATCAGA, R- TGTCAAAGAAATCCATGCAGA

954 *Tssc4*: F- ATGACACCCTGCCTTCTGAC, R- CACACCTTCCCACTTGGACT

955 *Ube3a*: F- CCACCTAGTCCTCTCGTCCA, R- CGCCTCACTGGTCCTTGT

956 *Sry*: F- CTGCAGTTGCCTCAACAAAA, R- CATGAGACTGCCAACCACAG

957 *Peg3*: F- TTCATCCTGAGCGAGAAGGT, R- GATGGGTTGATTTGGGTCAC

958 *Zdbf2*: F- AAGCGACCGACATCATCAG, R- TGGGTATGTGTGCAAATGAGA

959 *Mcts2*: F- GAACCAGTGGGGTAATCACG, R- CACACACAGAGATACAACGCAAT

960 *Phlda2*:F-TCGAAATGGCTTCGAAAATC, R- AAGAAAAGTGACATTGTTTATTTGGA

961 *Grb10*: F- TATGACCTCCTTGCCCACTC, R- AGAGGCCATCAAGCTTTCAA

962 *Tgfb1*: F- GCCCCAGAGTCTGAGACCA, R- ACGCGGGTGACCTCTTTAG

963 *Zfat*: F- CCCAGTTCCAACCACACAGT, R- CAGTGTCCTTGCACTGAGGA

964 *Trp73*: F- GCCCATCAAAGAGGAGTTCA, R- GCTCCTCTGTGCTTCTCCAC

965 *Actb*: F- GCGCAAGTACTCTGTGTGGA, R- CTGGCTGCCTCAACACCT

966 *Gapdh*: F- GTCACCAGGGCTGCCATT, R- GGCCTTCTCCATGGTGGT

967 *Ppib*: F- ATGTGGTACGGAAGGTGGAG, R- CCACTTTATTGATCAGCATTGG

968 *B2m*: F- TGAAGATTCATTTGAACCTGCTTA, R- AAAAGCAGAAGTAGCCACAGG

969

970

971 **DNA-FISH assay**

972 NPCs were cultured and fixed as described for the nascent RNA-FISH assay. Cells were rinsed in
973 1x PBS three times (5 min each wash) and permeabilized by adding ice-cold 0.5% Triton X-100
974 in 1x PBS and incubating on ice for 10 minutes. Cells were washed again with 1x PBS three times
975 (5 minutes each wash) and equilibrated in 2x SSC, changing the solution twice for 5 minutes each.
976 Cells were then incubated in 2x SSC containing 0.1 mg/ml RNase A for 1 hour at 37ºC.
977 Hybridization probe preparation: Probe mix was prepared by combining 10 µg of mouse Cot-1
978 DNA, 5 µl salmon sperm DNA, 5 µl of yeast tRNA, and 40 ng of nick-translated probe in a
979 microcentrifuge tube and adjusting the final volume to 20 µl with ddH_2O. The mixture was gently
980 tapped to mix and dehydrated in a SpeedVac at 65 ºC for 30 minutes. Ten microliters of deionized
981 formamide was added to the dried pellet, resuspended gently, and incubated at 37ºC for 60 minutes
982 with gentle shaking until fully dissolved. After the RNase treatment, cells were washed twice (5
983 minutes each) with 2x SSC at 42ºC. A 70% deionized formamide in 2x SSC (pre-warmed to 80ºC)
984 was added to each well, and 24-well MatTek plates were floated on 80ºC water bath and incubated
985 for 3 minutes. The solution was removed as quickly as possible, and ice-cold (-20ºC) 70% ethanol
986 was added to each well and incubated for 5 minutes on ice. Cells were dehydrated sequentially in
987 95% ethanol (5 minutes), followed by 100% ethanol (5 minutes), and kept in 100% ethanol until
988 proceeding to hybridization. Immediately before the hybridization, the probe mixture was

989 denatured at 95ºC for 10 minutes and cooled on ice for 2 minutes. Ten microliters of pre-warmed
990 2X hybridization buffer (20% Dextran Sulfate, 4x SSC in dH_2O at 37ºC) was added to 10 µl of
991 probe and 20 µl of the final probe mixture was added to each well. Plates were sealed with Parafilm
992 and incubated overnight at 37ºC in a humidified chamber. After probe hybridization, the probe
993 mixture was removed, and 50% deionized formamide in 2X SSC (pH 7.0) was added to each well
994 and incubated at 37ºC for 30 minutes. Cells were then washed sequentially with 2x SSC (15
995 minutes at 37ºC), 1x SSC (15 minutes at 37ºC), and equilibrated in 4x SSC for 5 minutes at RT.
996 DAPI diluted 1:10,000 in 4x SSC was added to each well and incubated for 3 minutes at RT,
997 followed by two washes with 4x SSC (5 minutes each). A total 100 µl of ibidi mounting medium
998 (ibidi GmbH, Cat.No.50001) was added to each well, and plates were wrapped in aluminium foil
999 and stored at 4ºC until imaging.
1000 Primers used for PCR amplification (probes for nick translation):
1001 Cth Nik F: GGTCAAGTGCTTGCAACAGA
1002 Cth Nik R: GAGCTTCCGGAGACACAGTC
1003 Ttc4 Nik F: CTCCATGGGAGGATGAGAAA
1004 Ttc4 Nik R: CCAAACCAAACCAAACCAAC
1005
1006
1007 **Chromatin Immunoprecipitation by Sun1 and Sequencing**
1008 ESCs and NPCs at ~80% confluence were harvested by trypsinization, rinsed and resuspended in
1009 Knockout DMEM containing 10% FBS (Gibco, Cat.no.10829-018). Crosslinking was performed
1010 by incubating $5X10^7$ cells 20 ml of Knockout DMEM containing 1% Formaldehyde and 10% FBS
1011 for 10 minutes at RT while gently shaking. Crosslinking was quenched by adding 2 ml of 1.5M

1012 glycine and incubating for an additional 10 minutes at RT with gentle shaking. Cells were
1013 centrifuged at 2000 rpm for 6 min at 4ºC, medium was removed, and cell plates were washed twice
1014 with 20 ml of ice-cold 1x PBS. Cell pellets were then resuspended in 500 µl of ChIP buffer (0.1%
1015 SDS, 10 mM EDTA, 50mM Tris-HCl, pH 7.5, 1 mM PMSF in dH_2O) and kept on ice. Chromatin
1016 was sheared using a Bioruptor® Pico sonication device (diagenode, Cat.No. B01060010) with 30
1017 seconds ON/ 30 seconds OFF for 15 cycles. After sonication, samples were centrifuged at 10,000
1018 g for 10 minutes at 4ºC. Supernatants were transferred to 1.5 ml Eppendorf microcentrifuge tubes
1019 and stored at -80ºC until use. To generate Sun1-ChIP nucleic-acid-depleted extracts, nuclear
1020 extracts were treated with Benzonase® Nuclease (Millipore Sigma, Cat. No. E1014-5KU)
1021 following the manufacturer's instructions. Sheared chromatin DNA and fragments ranged from
1022 300-500 bp, and concentrations were measured using a Nanodrop spectrophotometer. For each
1023 ChIP reaction, 150 µl of Dynabeads™ Protein G (Invitrogen, Cat. No. 10004D) were used. Beads
1024 were washed with ice-cold 0.02% Tween-20 in 1x PBS and resuspended in 150 µl of the same
1025 buffer. Sun1 antibody (0.8 µg; Proteintech, Cat. No. 24568-I-AP) or Rabbit IgG Isotype Control
1026 (Thermo Fisher, Cat. No. 10500c) was added to washed beads, and the volume was adjusted to
1027 200 µl with 0.02% Tween-20 in 1x PBS. Bead-antibody mixtures were incubated for 2 hours at
1028 4ºC on a tube rotator. After incubation, excess liquid was removed, and beads were gently rinsed
1029 once with RIPA buffer (50 mM Tris-HCl (pH8.0), 150 mM NaCl, 2 mM EDTA (pH8.0), 1% NP-
1030 40, 0.5% sodium deoxicholate, 0.05% SDS, and 1 Protease inhibitor tablet (cOmplete Mini,
1031 EDTA-free, Cat. No. 39802300) in 10 ml of dH_2O). For each IP, 150 µg of DNA-containing lysates
1032 was brought to 800 µl with 1x RIPA buffer, added to the bead-antibody complexes, and incubated
1033 overnight at 4ºC on a tube rotator. The next day, beads and bound proteins-DNA complexes were
1034 separated from unbound lysate using a magnetic rack. Beads were washed three times with 1000

1035 µl of 1x TE buffer (10 mM Tris-HCl (pH8.0), 2 mM EDTA in dH_2O). For elution, 200 µl of
1036 elution buffer (50 mM NaCl, 50 mM Tris-HCl (pH7.5), 0.1 mM PMSF, 5 mM EDTA, and 0.1%
1037 SDS in ddH_2O) was added and incubated at 31ºC for 15 minutes in a heat block with shaking at
1038 800 rpm. For paired-end sequencing, DNAs was extracted from 100 µl of elute using the Genomic
1039 DNA Clean & Concentrator™ -25 Kit (Zymo Research, Cat.No. D4064). Libraries were prepared
1040 and sequenced on a NextSeq500 platform (PE-150, mid output) at the Next Generation Sequencing
1041 Core Facility at Cold Spring Harbor Laboratory. Nucleic acids remaining in protein fractions for
1042 for Sun1 interacting pulldown assay were removed by incubating samples with Benzonase
1043 (Millipore Sigma, Cat. No.E1014) for 20 minutes on ice prior to use.
1044
1045
1046 **qPCR and Sanger sequencing**
1047 One µg of genomic DNA from Sun1-ChIP assay and from the input sample was used as template.
1048 Quantitative PCR was performed using PowerUp™ SYBR® Green Master Mix (Thermo Fisher,
1049 Cat. No. A25743) and the primers covering SNPs in the *Cth* gene (F-
1050 GAGCCTGGGAGGATATGAGA, R- AAGCTCGATCCAGGTCTTCA) and *Ttc4* gene (F –
1051 GACAGGGCGGAACTATACCA, R-genes. qPCR products were Sanger sequenced using the
1052 GenScript Biotec Sanger sequencing service.
1053
1054
1055 **Solution Digestion Analysis by LC-MS/MS**
1056 Fifty microliters of elute from each sample was sent to the Taplin Biological Mass Spectrometry
1057 Facility at Harvard Medical School. In brief, 10ul (20ng/µl) of modified sequencing-grade trypsin

1058 (Promega, Madison, WI) was spiked into 300 µl PBS, and samples were incubated overnight at
1059 37ºC. Samples were acidified by adding 20 µl of 20% formic acid and desalted using STAGE tip.
1060 On the day of analysis, the samples were reconstituted in 10 µl of HPLC solvent A. A nano-scale
1061 reverse-phase HPLC capillary column was prepared by packing 2.6 µm C18 spherical silica beads
1062 into a fused silica capillary (100 µm inner diameter x ~30 cm length) with a flame-drawn tip. After
1063 column equilibration, samples were loaded using a Famos autosampler (LC Packings, San
1064 Francisco, CA) onto the column. A gradient was formed, and peptides were eluted with increasing
1065 concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid). As peptides eluted, they were
1066 subjected to electrospray ionization and then entered into an LTQ Orbitrap Velos Elite ion-trap
1067 mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Peptides were detected, isolated,
1068 and fragmented to generate tandem mass spectra of specific fragment ions for each peptide.
1069 Peptide sequences (and hence protein identity) were determined by matching acquired MS/MS
1070 fragmentation patterns to protein databases using Sequest software program (Thermo Fisher
1071 Scientific, Waltham, MA). All databases include a reversed version of all the sequences, and the
1072 data was filtered to between a one and two percent peptide false discovery rate.
1073
1074
1075 **Allele-specific ChIP-seq analysis and motif discovery**
1076 Allele-specific ChIP-seq analysis was performed using the MEA pipeline[4.] An in silico diploid
1077 C57BL/6J / CAST/EiJ genome was reconstructed using the SHAPEIT2 pipeline by combining the
1078 mm10 reference genome (mgp.v5.merged.snps_all.dbSNP142.vcf.gz) with the corresponding
1079 genetic variants (mgp.v5.merged.indels.dbSNP142.normed.vcf.gz). Sequencing reads (.fastq)
1080 were aligned to the reconstructed in silico genome using Bowtie2 aligner (version 2.3.4.3). The

1081 resulting BAM files were processed through the ChIP-seq pipeline embedded in SeqMonk
1082 bioinformatic web portal, and peaks were called using the MACS algorithm with a p-value
1083 threshold of 0.05 and a fragment size of 500 bp. Motif analysis was performed using HOMER
1084 (v4.11) within the Basepair web portal (Version 3.2.7).
1085
1086
1087 **TA-bridge *in vivo* visualization, interacting protein pull-down, and mass spectrometry**
1088 mESCs were seeded at $0.4x10^4$ cells per well in 8-well ibidi chamber slides and incubated for 24
1089 hours prior to transfection with biotinylated DNA oligos. For immunofluorescence assays, 0.03
1090 µg of biotinylated oligos per well were used. For *in vivo* TA-bridge interacting protein pull-down,
1091 $2.5x10^5$ cells per well were seeded in three 6-well plates 24 hrs before transfection. A total of 0.5
1092 µg of biotinylated oligos per well was transfected using the Xfect™ transfection reagent kit
1093 (Clonetech, Cat.No. 631317) according to the manufacturer's instructions. Cells were incubated
1094 overnight and then fixed with 4% PFA and stored at 4ºC until use for imaging. For TA-bridges
1095 interacting protein pull-down, cells were rinsed three times with 1x PBS and nuclear proteins were
1096 extracted using the nuclear extraction kit (ActiveMotif®, Cat.No. 40010). For each pull-down
1097 reaction, 200 µl of Dynabeads™ MyOne Streptavidin T1 and 200 µg of nuclear extract were used.
1098 Beads were washed three times with 500 µl of 2x binding buffer and resuspended in 150 µl of 2x
1099 B&W buffer. The nuclear extract volume was adjusted to 150 µl by adding Complete Lysis Buffer
1100 (from the nuclear extract kit) and combined with the beads. Binding was carried out at RT for 1
1101 hour on a tube rotator. After incubation, tubes were briefly centrifuged and supernatants were
1102 removed using a magnatic rack. Beads were washed four times with 1 ml of 1x B&W buffer and
1103 then resuspended in 20 µl elution buffer (50mM NaCl, 50mM Tris-HCl (pH7.5), 0.1 mM PMFS,

1104 5 mM EDTA, 0.1% SDS in Milli-Q water) and boiled at 95ºC for 7 min while gently vortexing in
1105 a heat rack. Tubes were briefly centrifuged and elutes were collected using a magnetic rack. Elutes
1106 were stored at -80ºC until analysis. For MS analysis, 60 µl (20 µl combined from 3 preps) was
1107 used.
1108
1109
1110 **Identification of proteins bound to TA-bridges by MS**
1111 MS was performed at the Cold Spring Harbor laboratory Mass Spectrometry Core Facility. On-
1112 membrane proteolysis was carried out as follows. Excised nylon membranes were hydrated and
1113 washed three times with 500 ul of 100 mM ammonium bicarbonate (ABC) for 10 minutes at 55ºC
1114 under agitation. Next, 500 µl of fresh 3 mM TCEP (Tris(2-carboxyethyl)phosphine hydrochloride
1115)/50mM ABC was added and incubated at 55ºC for 20 minutes while vortexing at 500 rpm. The
1116 TCEP solution was removed, and 500 µl of fresh 10 mM CEMTS (2-carboxymethyl
1117 methanethiosulfonate) in 10% ethanol was added. Membranes were incubated at 55ºC for 5
1118 minutes while vortexing at 500 rpm. The CEMTS solution was removed, and membranes were
1119 washed three times with 500 µl of 50 mM ABC for 5 minutes at 55ºC with agitation. Trypsin
1120 solution (100 ng/µl; Sequencing-grade modified porcine trypsin, Promega) was prepared in 50 mM
1121 ABC. Five microliters of trypsin solution was added directly onto each membrane piece, followed
1122 by 200 microliters of trypsin solution was added into the middle of the membrane piece, followed
1123 by 200 µl of 50 mM ABC containing 0.1% ProteaseMax surfactant. Membranes were digested
1124 overnight at 37ºC while vortexing at 500 rpm. After digestion, 200 µl of 80% acetonitrile
1125 (CAN)/1% TFA was added and incubated for 15 minutes at 55ºC with agitation. Elutes were
1126 collected into clean tubes, lyophilized by vacuum centrifugation, and the peptide pellets were

1127 resuspended in 20 µ of loading buffer (55% DMSO, 0.1% TFA in water). Peptides were loaded on
1128 a 30 cm x 75 µM inner-diameter column packed with Reprosil 1.9 µm C18 silica particles and
1129 resolved using a 5-35% acetonitrile gradient in water (0.1% formate) at a flow rate of 250
1130 nl/minute. Eluting peptides were ionized by electrospray (2200V) and analyzed using an Exploris
1131 480 mass spectrometer (Thermo). The instrument was operated to acquire 120K-resolution MS1
1132 scans (m/z 380-2000 Th) and 60K HCD MS2 spectra at stepped normalized collision energies of
1133 28,33 and 38%, with the first mass locked to 100 Th. Raw files were searched using the Mascot
1134 scoring engine within ProteomeDiscoverer, with mass tolerances set at 5 ppm for MS1 and 0.01
1135 Da for MS2. Spectra were matched against the UniProt mouse consensus database supplemented
1136 with the common contaminants atabase (cRAP). Methionine oxidation and N/Q deamidation were
1137 set as variable modifications. Peptide spectral matches were filtered to maintain FDR < 1%.
1138
1139
1140 **Electrophoretic mobility shift assay (EMSA)**
1141 EMSA was performed using LightShift® Chemiluminescent EMSA Kit (Thermo Scientific,
1142 Cat.No. 20148). Biotinylated oligos were synthesized by bioSYNTHESIS®. The single-strandard
1143 biotinylated oligos usd were: BM-1: TAAATAAATAAATAAATAAATAAATAAATAAA, BM-
1144 2: ATTTATTTATTTATTTATTTATTTATTTATTT, Control: GCAT. Oligos were tagged at the
1145 3' end with Desthiobiotin-TEG, separated from the oligo sequence by six -Sp18- spacers. Double-
1146 stranded TA-bridges were generated by anneling BM-1 with a non-biotinilated complimentary
1147 oligo, and purified using the Genomic DNA Clean & Concentrator™ (ZYMO RESEARCH, Cat.
1148 No. D4065). A total of 30 fmol of each biotinylated oligo was used per reaction. Biotinylated
1149 Epstein-Barr nuclear antigen (EBNA) DNA was used as a negative control. 6% native

1150 polyacrylamide gels (1.0 mm, 8x10 cm) were used for gel-shift assay. Nuclear extracts were
1151 prepared from 70% confluent ESC cultures using the nuclear extraction kit (ActiveMotif®,
1152 Cat.No. 40010). Binding reactions were assembled by combining 10x binding buffer, 1 µl
1153 poly(dI.dC), Nuclear extract protein, and biotinylated DNA oligos to finl volume of 20 µL. Binding
1154 mixture were incubated for 20 minutes at RT. Then, 5 µl of 5x loading buffer was added to each
1155 reaction and mixed by pipetting. Prior to sample loading, native gels were pre-run for 60 minutes
1156 at 100 V. Samples were electrophoresed at 100 V until the bromophenol blue dye migrated to
1157 approximately three-quarters of the gel length. For protein transfer, Biodyne® A pre-Cut
1158 membranes (Thermo Scientific, Cat.No. 77015) were soaked in 0.5X TBE for 10 minutes, and
1159 gels, nylon membranes, and blotting paper were assembled as specified by the manufacturer.
1160 Transfers were performed in ice-cold 0.5X TBE, with the transfer apparatus fully submerged on
1161 ice, for 45 minutes at 380 mA. Nylon membranes were semi-dried on a paper towel with the
1162 staining side facing upward. Crosslinking was performed using a UV StrataLinker™ 1800
1163 (Stratagene) at 120,000 µJ for 60 seconds. Membranes were transferred directly to blocking buffer
1164 pre-warmed to 37ºC and incubated for 15 minutes in the blocking buffer on a slow rocker.
1165 Membranes were then rinsed once in 20 ml 1x wash buffer and washed four times for 5 minutes
1166 each 20 ml 1x wash buffer with gentle shaking. Membranes were transferred to a clean container
1167 and incubated in substrate equilibration buffer for 5 minutes. Excess buffer was removed by
1168 blotting, and 12 ml of substrate working solution was added and incubated for 5 minutes without
1169 shaking. After removing excess substrate (without over-drying), membranes were wrapped in
1170 plastic film and exposed to X-ray film (Amersham Hyperfilm™ ECL, GE Healthcare Limited,
1171 Cat.No. 28906836) for 7 minuits.
1172

1173 **Data availability**

1174 The Sun1 ChIP-seq data (raw and processed) supporting the finding of this study have been
1175 deposited in the Gene Expression Omnibus (GEO) under accession code GSE228344. All other
1176 data supporting the finding of this study has been submitted with the manuscript.

1177

1178

1179 **References**

1180 1 Bickmore, W. A. The spatial organization of the human genome. *Annu Rev Genomics Hum*
1181 *Genet* **14**, 67-84, doi:10.1146/annurev-genom-091212-153515 (2013).

1182 2 Binder, H. *et al.* Transcriptional regulation by histone modifications: towards a theory of
1183 chromatin re-organization during stem cell differentiation. *Phys Biol* **10**, 026006,
1184 doi:10.1088/1478-3975/10/2/026006 (2013).

1185 3 Cremer, T. & Cremer, M. Chromosome Territories. *Csh Perspect Biol* **2**, doi:ARTN
1186 a003889 10.1101/cshperspect.a003889 (2010).

1187 4 Fraser, P. & Bickmore, W. Nuclear organization of the genome and the potential for gene
1188 regulation. *Nature* **447**, 413-417, doi:10.1038/nature05916 (2007).

1189 5 Hubner, M. R., Eckersley-Maslin, M. A. & Spector, D. L. Chromatin organization and
1190 transcriptional regulation. *Curr Opin Genet Dev* **23**, 89-95, doi:10.1016/j.gde.2012.11.006
1191 (2013).

1192 6 Meaburn, K. J. & Misteli, T. Cell biology: chromosome territories. *Nature* **445**, 379-781,
1193 doi:10.1038/445379a (2007).

1194 7 Takizawa, T., Meaburn, K. J. & Misteli, T. The meaning of gene positioning. *Cell* **135**, 9-
1195 13, doi:10.1016/j.cell.2008.09.026 (2008).

1196 8 Abed, J. A. *et al.* Highly structured homolog pairing reflects functional organization of the
1197 Drosophila genome. *Nat Commun* **10**, doi:ARTN 4485 10.1038/s41467-019-12208-3
1198 (2019).

1199 9 Duncan, I. W. Transvection effects in Drosophila. *Annu Rev Genet* **36**, 521-556,
1200 doi:10.1146/annurev.genet.36.060402.100441 (2002).

1201 10 Galouzis, C. C. & Prud'homme, B. Transvection regulates the sex-biased expression of a
1202 fly X-linked gene. *Science* **371**, 396-+, doi:10.1126/science.abc2745 (2021).
1203 11 Hogan, M. S., Parfitt, D. E., Zepeda-Mendoza, C. J., Shen, M. M. & Spector, D. L.
1204 Transient Pairing of Homologous Oct4 Alleles Accompanies the Onset of Embryonic Stem
1205 Cell Differentiation. *Cell Stem Cell* **16**, 275-288, doi:10.1016/j.stem.2015.02.001 (2015).
1206 12 Jack, J. W. & Judd, B. H. Allelic pairing and gene regulation: A model for the zeste-white
1207 interaction in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **76**, 1368-1372,
1208 doi:10.1073/pnas.76.3.1368 (1979).
1209 13 Viets, K. *et al.* Characterization of Button Loci that Promote Homologous Chromosome
1210 Pairing and Cell-Type-Specific Interchromosomal Gene Regulation. *Dev Cell* **51**, 341-+,
1211 doi:10.1016/j.devcel.2019.09.007 (2019).
1212 14 D'Angelo, M. A. Nuclear pore complexes as hubs for gene regulation. *Nucleus-Phila* **9**,
1213 142-148, doi:10.1080/19491034.2017.1395542 (2018).
1214 15 Mao, Y. T. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies.
1215 *Trends Genet* **27**, 295-306, doi:10.1016/j.tig.2011.05.006 (2011).
1216 16 Meldi, L. & Brickner, J. H. Compartmentalization of the nucleus. *Trends Cell Biol* **21**, 701-
1217 708, doi:10.1016/j.tcb.2011.08.001 (2011).
1218 17 Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome
1219 Organization in the Nucleus. *Cell* **174**, 744-+, doi:10.1016/j.cell.2018.05.024 (2018).
1220 18 Zimber, A., Nguyen, Q. D. & Gespach, C. Nuclear bodies and compartments: Functional
1221 roles and cellular signalling in health and disease. *Cell Signal* **16**, 1085-1104,
1222 doi:10.1016/j.cellsig.2004.03.020 (2004).

1223 19 Eckersiey-Maslin, M. A. & Spector, D. L. Random monoallelic expression: regulating gene
1224 expression one allele at a time. *Trends Genet* **30**, 237-244, doi:10.1016/j.tig.2014.03.003
1225 (2014).
1226 20 Ferguson-Smith, A. C. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat*
1227 *Rev Genet* **12**, 565-575, doi:10.1038/nrg3032 (2011).
1228 21 Reik, W. & Walter, J. Imprinting mechanisms in mammals. *Curr Opin Genet Dev* **8**, 154-
1229 164, doi:10.1016/s0959-437x(98)80136-6 (1998).
1230 22 Balasooriya, G. I. & Spector, D. L. Allele-specific differential regulation of monoallelically
1231 expressed autosomal genes in the cardiac lineage. *Nat Commun* **13**, 5984,
1232 doi:10.1038/s41467-022-33722-x (2022).
1233 23 Chen, M. *et al.* Live imaging of RNA and RNA splicing in mammalian cells via the
1234 dcas13a-SunTag-BiFC system. *Biosens Bioelectron* **204**, 114074,
1235 doi:10.1016/j.bios.2022.114074 (2022).
1236 24 Daigle, N. & Ellenberg, J. LambdaN-GFP: an RNA reporter system for live-cell imaging.
1237 *Nat Methods* **4**, 633-636, doi:10.1038/nmeth1065 (2007).
1238 25 Haim, L., Zipor, G., Aronov, S. & Gerst, J. E. A genomic integration method to visualize
1239 localization of endogenous mRNAs in living yeast. *Nat Methods* **4**, 409-412,
1240 doi:10.1038/nmeth1040 (2007).
1241 26 Hocine, S., Raymond, P., Zenklusen, D., Chao, J. A. & Singer, R. H. Single-molecule
1242 analysis of gene expression using two-color RNA labeling in live yeast. *Nat Methods* **10**,
1243 119-121, doi:10.1038/nmeth.2305 (2013).
1244 27 Janicki, S. M. *et al.* From silencing to gene expression: real-time analysis in single cells.
1245 *Cell* **116**, 683-698, doi:10.1016/s0092-8674(04)00171-0 (2004).

1246 28 Ma, H. H. *et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs
1247 using CRISPRainbow. *Nature Biotechnology* **34**, 528-530, doi:10.1038/nbt.3526 (2016).
1248 29 Maass, P. G. *et al.* Spatiotemporal allele organization by allele-specific CRISPR live-cell
1249 imaging (SNP-CLING). *Nat Struct Mol Biol* **25**, 176-+, doi:10.1038/s41594-017-0015-3
1250 (2018).
1251 30 Tsukamoto, T. *et al.* Visualization of gene activity in living cells. *Nat Cell Biol* **2**, 871-878,
1252 doi:10.1038/35046510 (2000).
1253 31 Vera, M., Biswas, J., Senecal, A., Singer, R. H. & Park, H. Y. Single-Cell and Single-
1254 Molecule Analysis of Gene Expression Regulation. *Annual Review of Genetics, Vol 50* **50**,
1255 267-291, doi:10.1146/annurev-genet-120215-034854 (2016).
1256 32 Chen, H. M. & Larson, D. R. What have single-molecule studies taught us about gene
1257 expression? *Gene Dev* **30**, 1796-1810, doi:10.1101/gad.281725.116 (2016).
1258 33 Shao, S. P., Xue, B. X. & Sun, Y. J. Intranucleus Single-Molecule Imaging in Living Cells.
1259 *Biophys J* **115**, 181-189, doi:10.1016/j.bpj.2018.05.017 (2018).
1260 34 Shav-Tal, Y., Singer, R. H. & Darzacq, X. Imaging gene expression in single living cells.
1261 *Nat Rev Mol Cell Bio* **5**, 855-862, doi:10.1038/nrm1494 (2004).
1262 35 Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression.
1263 *Nature* **422**, 633-637, doi:10.1038/nature01546 (2003).
1264 36 Larson, D. R., Singer, R. H. & Zenklusen, D. A single molecule view of gene expression.
1265 *Trends Cell Biol* **19**, 630-637, doi:10.1016/j.tcb.2009.08.008 (2009).
1266 37 Wang, L., Frei, M. S., Salim, A. & Johnsson, K. Small-Molecule Fluorescent Probes for
1267 Live-Cell Super-Resolution Microscopy. *J Am Chem Soc* **141**, 2770-2781,
1268 doi:10.1021/jacs.8b11134 (2019).

1269 38 Joyce, E. F., Erceg, J. & Wu, C. T. Pairing and anti-pairing: a balancing act in the diploid
1270 genome. *Curr Opin Genet Dev* **37**, 119-128, doi:10.1016/j.gde.2016.03.002 (2016).
1271 39 Blattner, G., Cavazza, A., Thrasher, A. J. & Turchiano, G. Gene Editing and Genotoxicity:
1272 Targeting the Off-Targets. *Front Genome Ed* **2**, 613252, doi:10.3389/fgeed.2020.613252
1273 (2020).
1274 40 Bohne, J. & Cathomen, T. Genotoxicity in gene therapy: An account of vector integration
1275 and designer nucleases. *Curr Opin Mol Ther* **10**, 214-223 (2008).
1276 41 Yoon, K. H. *et al.* Olfactory receptor genes expressed in distinct lineages are sequestered
1277 in different nuclear compartments. *P Natl Acad Sci USA* **112**, E2403-E2409,
1278 doi:10.1073/pnas.1506058112 (2015).
1279 42 Armelin-Correa, L. M., Gutiyama, L. M., Brandt, D. Y. C. & Malnic, B. Nuclear
1280 compartmentalization of odorant receptor genes. *P Natl Acad Sci USA* **111**, 2782-2787,
1281 doi:10.1073/pnas.1317036111 (2014).
1282 43 Clowney, E. J. *et al.* Nuclear Aggregation of Olfactory Receptor Genes Governs Their
1283 Monogenic Expression. *Cell* **151**, 724-737, doi:10.1016/j.cell.2012.09.043 (2012).
1284 44 Gundersen, G. G. & Worman, H. J. Nuclear Positioning. *Cell* **152**, 1376-1389,
1285 doi:10.1016/j.cell.2013.02.031 (2013).
1286 45 Haque, F. *et al.* SUN1 interacts with nuclear lamin A and cytoplasmic nesprins to provide
1287 a physical connection between the nuclear lamina and the cytoskeleton. *Mol Cell Biol* **26**,
1288 3738-3751, doi:10.1128/Mcb.26.10.3738-3751.2006 (2006).
1289 46 Liu, Q. *et al.* Functional association of Sun1 with nuclear pore complexes. *J Cell Biol* **178**,
1290 785-798, doi:10.1083/jcb.200704108 (2007).

1291 47 Li, P. *et al.* The function of the inner nuclear envelope protein SUN1 in mRNA export is
1292 regulated by phosphorylation. *Sci Rep-Uk* **7**, doi:ARTN 9157 10.1038/s41598-017-08837-
1293 7 (2017).
1294 48 Summer, M. C. & Brickner, J. The Nuclear Pore Complex as a Transcription Regulator.
1295 *Csh Perspect Biol* **14**, doi:ARTN a039438 10.1101/cshperspect.a039438 (2022).
1296 49 Matsumoto, A. *et al.* Loss of the integral nuclear envelope protein SUN1 induces alteration
1297 of nucleoli. *Nucleus-Phila* **7**, 68-83, doi:10.1080/19491034.2016.1149664 (2016).
1298 50 van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome
1299 Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780-791,
1300 doi:10.1016/j.cell.2017.04.022 (2017).
1301 51 Gonzalez-Sandoval, A. & Gasser, S. M. On TADs and LADs: Spatial Control Over Gene
1302 Expression. *Trends Genet* **32**, 485-495, doi:10.1016/j.tig.2016.05.004 (2016).
1303 52 Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of
1304 nuclear lamina interactions. *Nature* **453**, 948-951, doi:10.1038/nature06947 (2008).
1305 53 Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina
1306 interactions during differentiation. *Mol Cell* **38**, 603-613,
1307 doi:10.1016/j.molcel.2010.03.016 (2010).
1308 54 Leemans, C. *et al.* Promoter-Intrinsic and Local Chromatin Features Determine Gene
1309 Repression in LADs. *Cell* **177**, 852-864 e814, doi:10.1016/j.cell.2019.03.009 (2019).
1310 55 Luo, L. *et al.* The nuclear periphery of embryonic stem cells is a transcriptionally
1311 permissive and repressive compartment. *J Cell Sci* **122**, 3729-3737,
1312 doi:10.1242/jcs.052555 (2009).

1313 56 Wu, F. & Yao, J. Identifying Novel Transcriptional and Epigenetic Features of Nuclear
1314 Lamina-associated Genes. *Sci Rep* **7**, 100, doi:10.1038/s41598-017-00176-x (2017).
1315 57 Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*
1316 **171**, 557-+, doi:10.1016/j.cell.2017.09.043 (2017).
1317 58 Monahan, K., Horta, A. & Lomvardas, S. LHX2-and LDB1-mediated trans interactions
1318 regulate olfactory receptor choice. *Nature* **565**, 448-+, doi:10.1038/s41586-018-0845-0
1319 (2019).
1320 59 Dekker, J. *et al.* The 4D nucleome project. *Nature* **549**, 219-226, doi:10.1038/nature23884
1321 (2017).
1322 60 Reiff, S. B. *et al.* The 4D Nucleome Data Portal as a resource for searching and visualizing
1323 curated nucleomics data. *Nat Commun* **13**, doi:ARTN 2365 10.1038/s41467-022-29697-4
1324 (2022).
1325 61 Shevelyov, Y. Y. & Ulianov, S. V. The Nuclear Lamina as an Organizer of Chromosome
1326 Architecture. *Cells* **8**, doi:ARTN 136 10.3390/cells8020136 (2019).
1327 62 Czapiewski, R., Robson, M. I. & Schirmer, E. C. Anchoring a Leviathan: How the Nuclear
1328 Membrane Tethers the Genome. *Front Genet* **7**, doi:ARTN 82 10.3389/fgene.2016.00082
1329 (2016).
1330 63 Snyder, M. J. *et al.* Anchoring of Heterochromatin to the Nuclear Lamina Reinforces
1331 Dosage Compensation-Mediated Gene Repression. *Plos Genet* **12**, doi:ARTN e1006341
1332 10.1371/journal.pgen.1006341 (2016).
1333 64 Rivera-Mulia, J. C. *et al.* Allele-specific control of replication timing and genome
1334 organization during development. *Genome Research* **28**, 800-811,
1335 doi:10.1101/gr.232561.117 (2018).

1336 65 Arao, Y. *et al.* A+U-rich-element RNA-binding factor 1/heterogeneous nuclear
1337 ribonucleoprotein D gene expression is regulated by oestrogen in the rat uterus. *Biochem J*
1338 **361**, 125-132, doi:Doi 10.1042/0264-6021:3610125 (2002).
1339 66 Wu, F. N. & Yao, J. Identifying Novel Transcriptional and Epigenetic Features of Nuclear
1340 Lamina-associated Genes. *Sci Rep-Uk* **7**, doi:ARTN 100 10.1038/s41598-017-00176-x
1341 (2017).
1342 67 Leemans, C. *et al.* Promoter-Intrinsic and Local Chromatin Features Determine Gene
1343 Repression in LADs. *Cell* **177**, 852-+, doi:10.1016/j.cell.2019.03.009 (2019).
1344 68 Wu, C. H. *et al.* NELF and DSIF cause promoter proximal pausing on the hsp70 promoter
1345 in Drosophila. *Gene Dev* **17**, 1402-1414, doi:10.1101/gad.1091403 (2003).
1346 69 Gressel, S. *et al.* CDK9-dependent RNA polymerase II pausing controls transcription
1347 initiation. *Elife* **6**, doi:ARTN e29736 10.7554/eLife.29736 (2017).
1348 70 Nazer, E., Dale, R. K., Chinen, M., Radmanesh, B. & Lei, E. P. Argonaute2 and LaminB
1349 modulate gene expression by controlling chromatin topology. *Plos Genet* **14**, doi:ARTN
1350 e1007276 10.1371/journal.pgen.1007276 (2018).
1351 71 Hah, J. & Kim, D. H. Deciphering Nuclear Mechanobiology in Laminopathy. *Cells* **8**,
1352 doi:10.3390/cells8030231 (2019).
1353 72 Chen, Y. *et al.* The SUN1-SPDYA interaction plays an essential role in meiosis prophase
1354 I. *Nat Commun* **12**, 3176, doi:10.1038/s41467-021-23550-w (2021).
1355 73 Hoyt, S. J. *et al.* From telomere to telomere: The transcriptional and epigenetic state of
1356 human repeat elements. *Science* **376**, eabk3112, doi:10.1126/science.abk3112 (2022).

1357 74 Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology,
1358 bioinformatics and applications. *Nat Biotechnol* **39**, 1348-1365, doi:10.1038/s41587-021-
1359 01108-x (2021).
1360 75 Bzymek, M. & Lovett, S. T. Instability of repetitive DNA sequences: the role of replication
1361 in multiple mechanisms. *Proc Natl Acad Sci U S A* **98**, 8319-8325,
1362 doi:10.1073/pnas.111008398 (2001).
1363 76 Altemose, N. *et al.* DiMeLo-seq: a long-read, single-molecule method for mapping protein-
1364 DNA interactions genome wide. *Nat Methods* **19**, 711-723, doi:10.1038/s41592-022-
1365 01475-6 (2022).
1366 77 Nguyen, G. H. *et al.* Regulation of gene expression by the BLM helicase correlates with
1367 the presence of G-quadruplex DNA motifs. *Proc Natl Acad Sci U S A* **111**, 9905-9910,
1368 doi:10.1073/pnas.1404807111 (2014).
1369 78 Mao, S. Q. *et al.* DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol*
1370 *Biol* **25**, 951-957, doi:10.1038/s41594-018-0131-8 (2018).
1371 79 Tye, B. K. MCM proteins in DNA replication. *Annu Rev Biochem* **68**, 649-686,
1372 doi:10.1146/annurev.biochem.68.1.649 (1999).
1373 80 Roca, J. Topoisomerase II: a fitted mechanism for the chromatin landscape. *Nucleic Acids*
1374 *Res* **37**, 721-730, doi:10.1093/nar/gkn994 (2009).
1375 81 Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation.
1376 *Nature* **518**, 331-336, doi:10.1038/nature14222 (2015).
1377 82 Winick-Ng, W. *et al.* Cell-type specialization is encoded by specific chromatin topologies.
1378 *Nature* **599**, 684-+, doi:10.1038/s41586-021-04081-2 (2021).

1379 83 Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor
1380 gene expression. *Cell* **78**, 823-834, doi:10.1016/s0092-8674(94)90562-2 (1994).
1381 84 Singh, N. *et al.* Coordination of the random asynchronous replication of autosomal loci.
1382 *Nat Genet* **33**, 339-341, doi:10.1038/ng1102 (2003).
1383 85 Lionnet, T. *et al.* A transgenic mouse for in vivo detection of endogenous labeled mRNA.
1384 *Nat Methods* **8**, 165-170, doi:10.1038/nmeth.1551 (2011).
1385
1386
1387 **Acknowledgements**
1388 We wold like to thank the members of the Spector laboratory for their critical discussions and
1389 advice throughout this study. We thank Prof. Min Han for generously providing the Sun1 antibody.
1390 We thank Osama El Demerdash for assistance in establishing the computational pipeline, Tse-
1391 Luen Wee and Nour El-Amine for assistance with microscopy, and Andrew Barazia (Intelligent
1392 Imaging Innovation- 3i) for enabling use of the 3i lattice light-sheet microscope for live imaging
1393 experiments. We would also thank Paolo Cifani (CSHL) and Ross Tomaino (Taplin Mass
1394 Spectrometry Facility, Harvard Medical School) for assisting with mass spectrometry experiments,
1395 and Sara Goodwin (CSHL) for assisting with DNA sequencing. We acknowledge the CSHL
1396 Cancer Center Shared Resources (Microscopy, Mass Spectrometry, Next-Gen Sequencing) for
1397 services and technical expertise (NCI 2P3OCA45508). Sequencing was performed using
1398 equipment purchased through NIH grant S10OD028632-01. This research was supported by
1399 R35GM131833 (D.L.S.) and grant 18-26 from the Charles H. Revson Foundation (G.I.B.).
1400
1401

1402 **Author Contribution**

1403 G.I.B. and D.L.S. conceived and designed the overall study. G.I.B. designed and performed all the

1404 experiments and analysis. G.I.B. and D.L.S. wrote and edited the manuscript.

1405

1406

1407 **Competing Interest**

1408 The authors declare no competing interests.

1409

1410

1411 **Co-corresponding authors**

1412 David L. Spector (spector@cshl.edu), Gayan Balasooriya (gayan.balasooriya@kcl.ac.uk)

1413