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

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

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

[*]Corresponding author: spector@cshl.edu

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. Establishing 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 active DNA loci are enriched with adenine-thymidine-rich tandem repeats that interact with Hnrnpd and reside 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 localized in chromosome 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 autosomal gene regulation[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 result in differential allelic gene regulation. Autosomal monoallelic gene expression has been identified in a sub-set of mammals in which only a stochastically or deterministically chosen single allele is transcriptionally active in a given cell[19-21]. Although a recent study suggested allele-specific differential regulation[22], the mechanism that regulates the allele choice of a monoallelic gene and the allelic geometry of those genes is not fully understood.

Pioneering molecular techniques have been developed to visualize mRNA, DNA, and proteins in live cells, allowing the understanding of the spatial and temporal dynamics of those molecules[23-31]. However, the techniques thus far established to explore genome regulation in living cells could not capture the temporal dynamics of allelic gene expression. Thus, insights into temporal allelic gene regulation will immensely advance our fundamental understanding of governing aspects of biology, including homeostasis, development, and disease[32-34].

In this study, we developed a molecular approach to distinguish the temporal expression of a gene's maternal and paternal alleles in living or fixed cells using polymorphic F1 hybrid mouse

42 embryonic stem cells (mESCs). We showed that this system could be used to detect the specific
43 spatial arrangement of each of the two alleles of a gene. Utilizing this system to probe two
44 monoallelically expressed autosomal genes, we showed that when the genes are biallelically
45 expressed in mESCs, alleles are paired, and the transcriptionally active alleles are localized at
46 Sun1-enriched domains at the nuclear periphery (NP). Upon differentiation to neural progenitor
47 cells (NPCs) those genes gain monoallelic status, and the active alleles are localized at the Sun1-
48 enriched domain. By pulling down the chromatin associated with Sun1-enriched domains, we
49 further showed that monoallelic genes are preferentially expressed at the NP compared to biallelic
50 genes. Using publicly available data, we further showed that the monoallelic genes are localized
51 in "micro-TADs" in Hi-C-defined euchromatic A compartments within peripherally located
52 LADs. In addition, we discovered that *de novo* T_1A_3 tandem DNA repeat sequences are associated
53 with genes localized at the NP and these T_1A_3 repeats are involved in the localization of the alleles
54 to the Sun1-enriched domains by Hnrnpd.
55
56 **Results**
57 **Visualizing allele-specific gene expression in living cells**
58 One of the main hurdles in live imaging of subnuclear structures is the molecular noise
59 common to any biological system[31,35,36] that significantly impacts image processing and analysis
60 because of the low signal-to-noise ratio. Further, ectopically expressed fluorescent-labeled reporter
61 proteins may cause molecular-crowding in cells, affecting the cells' physiology. In addition, cell-
62 penetrating fluorescent reporter molecules[37] may in some cases become a limiting factor during
63 long-term imaging. To overcome these issues, we generated a doxycycline-inducible and tunable
64 EYFP gene fused MS2 coat protein (MS2-EYFP) and mCherry gene fused PP7 coat protein (PP7-

65 mCherry) expressing mESC line (Fig.1a and Extended Data Fig.1a). Next, we conditionally
66 knocked-in 24-repeats of the PP7 stem-loop (PP7-SL) sequence into the 3' UTR of the CAST/EiJ
67 allele and 24-repeats of an MS2 stem-loop (MS2-SL) sequence into the 3' UTR of the C57BL/6J
68 allele using Cas9-D10A nickase and allele-specific SNP targeting gRNAs (Fig.1a). We established
69 this system for two autosomal monoallelic genes, *Cth* and *Ttc4*, biallelically expressed in mESCs
70 but monoallelically expressed upon differentiation[22], using two independent mESC clones (Fig.
71 1c).
72
73 **Active *Cth* and *Ttc4* alleles are paired**
74 To confirm that the MS2-SL and PP7-SL are integrated into the targeted *Cth* and *Ttc4* gene
75 alleles but not off-targeted to different loci, we performed dual-color DNA-FISH using GFP-
76 coupled nucleotide incorporated probes designed to detect the endogenous alleles of the genes and
77 the RFP-couplled nucleotide incomporated probes designed to detect the inserted stem-loop
78 cassettes. We detected the RFP and GFP fluorescent signals to be exclusively adjacent, confirming
79 the stem-loop knock-in into the *Cth* and *Ttc4* genes (Extended Data Fig.1b). To our surprise we
80 observed that both alleles of the *Cth* and *Ttc4* genes were paired (*Cth* alleles are within ~1μm
81 distance of each other and *Ttc4* alleles are within ~1.5 μm distance of each other) in mESCs
82 (Extended Data Fig.1b - green dots & c). To assess the binding specificity of the coat proteins to
83 their unique stem-loops, we induced the MS2-EYFP and PP7-mCherry coat proteins expression in
84 mESCs and mESC-derived NPCs to target the *Cth* and *Ttc4* genes. By imaging fixed mESCs and
85 NPCs, we confirmed that the coat proteins interact specifically with the appropriate stem-loops
86 (MS2 or PP7) and they do not cross-react in our clones (Fig.1c and Extended Data Fig.1d).

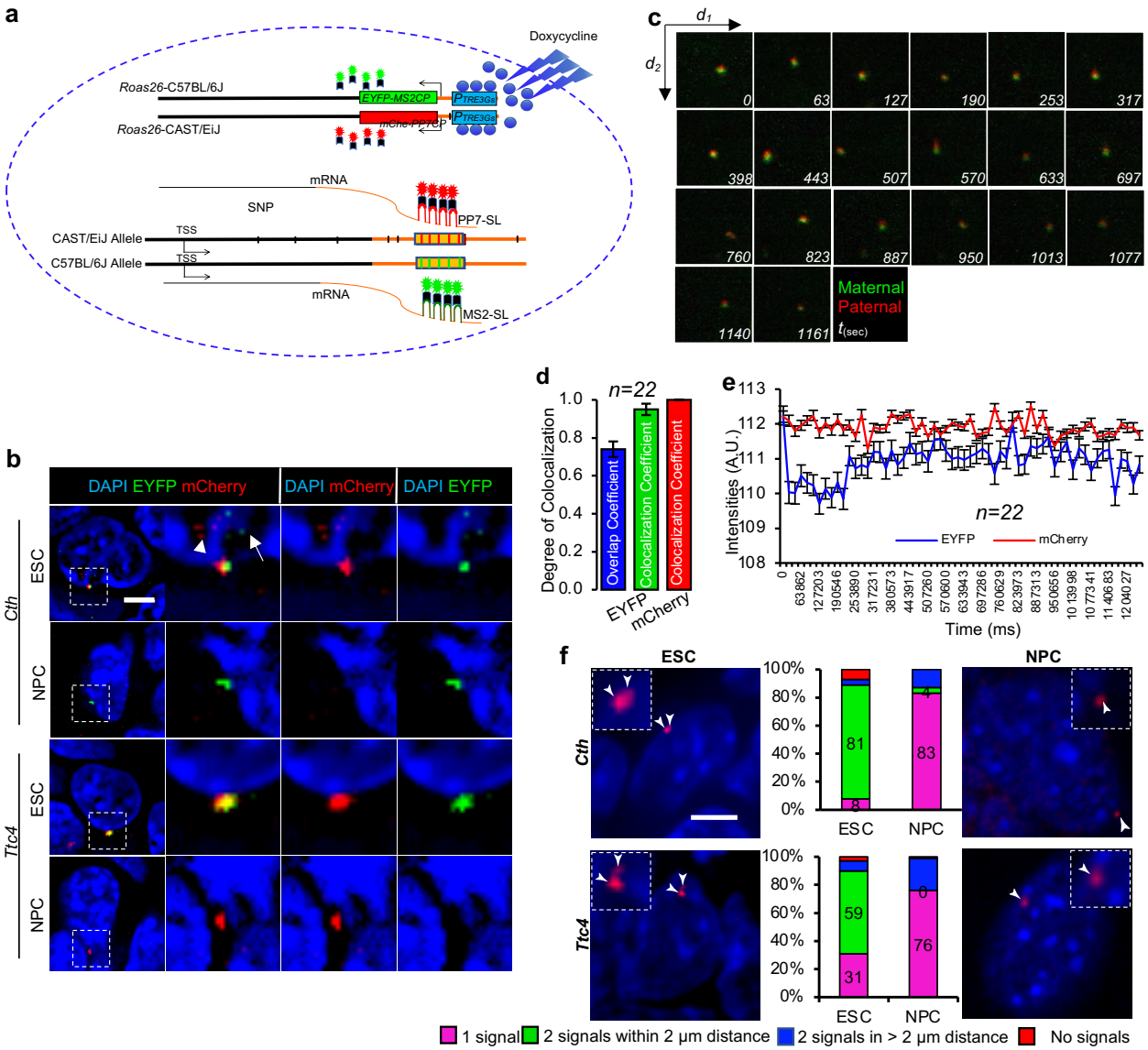


Figure 1 | Development of an allele-specific gene expression system to assess the localization of *Cth* and *Ttc4* alleles.

a, Schematic illustration of allele-specific gene expression visualization system. Inducible expression of EYFP-MS2CP and mChe-PP7CP is sensitive to the dosage of doxycycline. These coat proteins interact with unique stem-loops, MS2-SL with EYFP-MS2CP and PP7-SL with mChe-PP7CP at the 3' UTR of nascent mRNAs. Each RNA molecule contains 24 stem-loops. The maternal (C57BL/6J) mRNAs are marked by EYFP, and the paternal (CAST/EiJ) mRNAs are marked by mCherry. Low-background fluorescent signals are an advantage in our system as they will allow for easier allele detection and significant background noise reduction will not be necessary for image analysis. In addition, this introduces minimal disturbance to the cell's physiology as the fluorescent molecules crowding in the cells is minimal and controllable. **b,** EYFP-MS2CP and mChe-PP7CP signals in doxycycline indused *Cth* and *Ttc4* ESC and NPC cells. Cells were fixed at 36 hrs post doxycycline induction. Dotted areas in the left image panels are enlarged to demonstrate the fluorescent signal colocalization at the nuclear periphery. Fluorescent signals, possibly from single RNA molecules (maternal: arrow, paternal: arrowhead) are visible in enlarged images in the first row. Scale bar - 6 µm. **c,** Snapshots of a tracked single pair of alleles at different time points from a 20 min video using the 3i lattice lightsheet microscope. Green and red dots represent the maternal and paternal *Cth* alleles, respectively in ESCs. **d,** Degree of colocalization of the EYFP and mCherry signals calculated from tracked 22 EYFP and mCherry signal pairs in 22 cells from a 20 min video. Columns represent the overlap coefficient of green and red

dots (blue column) and colocalization coefficients of EYFP (green column) and mCherry (red column). Error bars: SEM. n=22 cells. **e,** Line plot showing the fluctuation in intensity of the colocalized EYFP and mCherry dots. Live images were captured every minute using the 3i lattice lightsheet microscope. Error bars: SEM. n=22. **f,** Representative RNA-FISH images show that *Cth* and *Ttc4* allele's nascent mRNAs are localized at the nuclear periphery in mESCs and NPCs. Arrowheads indicate the nascent RNA-FISH signals. Magnified images (in the dotted lined box) show the active *Cth* and *Ttc4* alleles pairing. Scale bar - 6 µm. The column bar graphs show the percentages of the number of observations made for the distance between the two alleles. Two alleles within 2 µm proximity of each other are considered as paired alleles.

87 Homologous allele pairing has been shown to impact a gene's regulation[8,11,38], but the
88 broader impact of allele pairing in terms of transcriptional heredity and the spatial arrangement of
89 the allele in the 3D nucleus of mammalian cells in particular is not well understood. By examining
90 the *Cth* genes' maternal and paternal alleles' nascent RNAs in every 20 sec for 20 minutes by live
91 cell imaging (Fig.1d and Supplementary Movie 1); we observed that both alleles are colocalized
92 (Fig.1e) and actively being transcribed (Fig.1f) in mESCs (n=22 cells). However, DNA
93 manipulation of specific loci could create altered genome organization in 3D nuclear space due to
94 potential changes in chromatin structure[39,40]. Therefore, to test whether the homologous allele
95 pairing we observed in regard to the *Cth* and *Ttc4* genes in our stem-loop targeted mESC clones
96 was due to genetic manipulation, we performed nascent RNA-FISH for these two genes in wild-
97 type mESCs and mESC derived NPCs (Fig.1g and Extended Data Fig.1e). We observed that
98 the *Cth* alleles are paired in 81% of the mESCs (8% are monoallelic), and *Ttc4* alleles are paired
99 in 59% of the mESCs (Fig.1g-column graph), confirming that the allele-specific gene labeling
100 technique did not alter the chromatin loci; thus, the allele pairing of the *Cth* and *Ttc4* genes are
101 true biological events.
102

103 **Active *Cth* and *Ttc4* alleles localize to the nuclear periphery**

104 In mESCs and NPCs having MS2-SLs and PP7-SLs knocked into the *Cth* and *Ttc4* gene
105 alleles (Fig.1c), and in addition by a nascent RNA-FISH assay (Fig.1g), we observed that the

active *Cth* and *Ttc4* alleles were localized to the NP (Fig.2a and Extended Data Fig.2a). Because the *Cth* and *Ttc4* are monoallelic genes, we asked whether this allele localization is common to monoallelic genes using a panel of known 13 random autosomal monoallelic genes (RaMA) and 14 deterministic monoallelic (DeMA) genes to assess the nuclear positioning of their transcriptionally active alleles in NPCs by nascent RNA-FISH. We found that the DeMA and RaMA genes we assessed were preferentially (DeMA genes 95% on average, RaMA genes 93% on average, n=100 per gene) localized at the NP of NPCs (Fig.2b and Extended Data Fig.2b,c). Interestingly, it has been reported in previous studies that the transcriptionally active some olfactory receptor genes, a family of autosomal monoallelic genes, are also localized at the NP[41-43], further confirming our observation in relation to RaMA and DeMA genes' expression at the NP.

Monoallelic genes localize at Sun1 territories

Following our observation of the localization of *Cth* and *Ttc4* transcriptionally active alleles at the NP, we asked whether these transcriptionally active alleles exhibit any association with specific protein or protein territories at the inner nuclear membrane. Sun1, a inner-nuclear-membrane protein (Extended Data Fig.2d), is involved in positioning the nucleus in the cytoplasm by interacting with actin filaments via nesprin[44,45]. In addition, Sun1 is associated with the nuclear pore complex, which has a role in mRNA transport[46,47] and harboring active genes[14,48]. While the Sun1-enriched nuclear domains are associated with stabilizing chromatin architecture[49], their specific molecular function link to gene regulation is unknown. Although overall, the lamina-associated domains at the NP harbor heterochromatic chromatin[50], the transcriptional status of genes associated with Sun1-enriched nuclear domains is unknown. Immunofluorescent labeling of

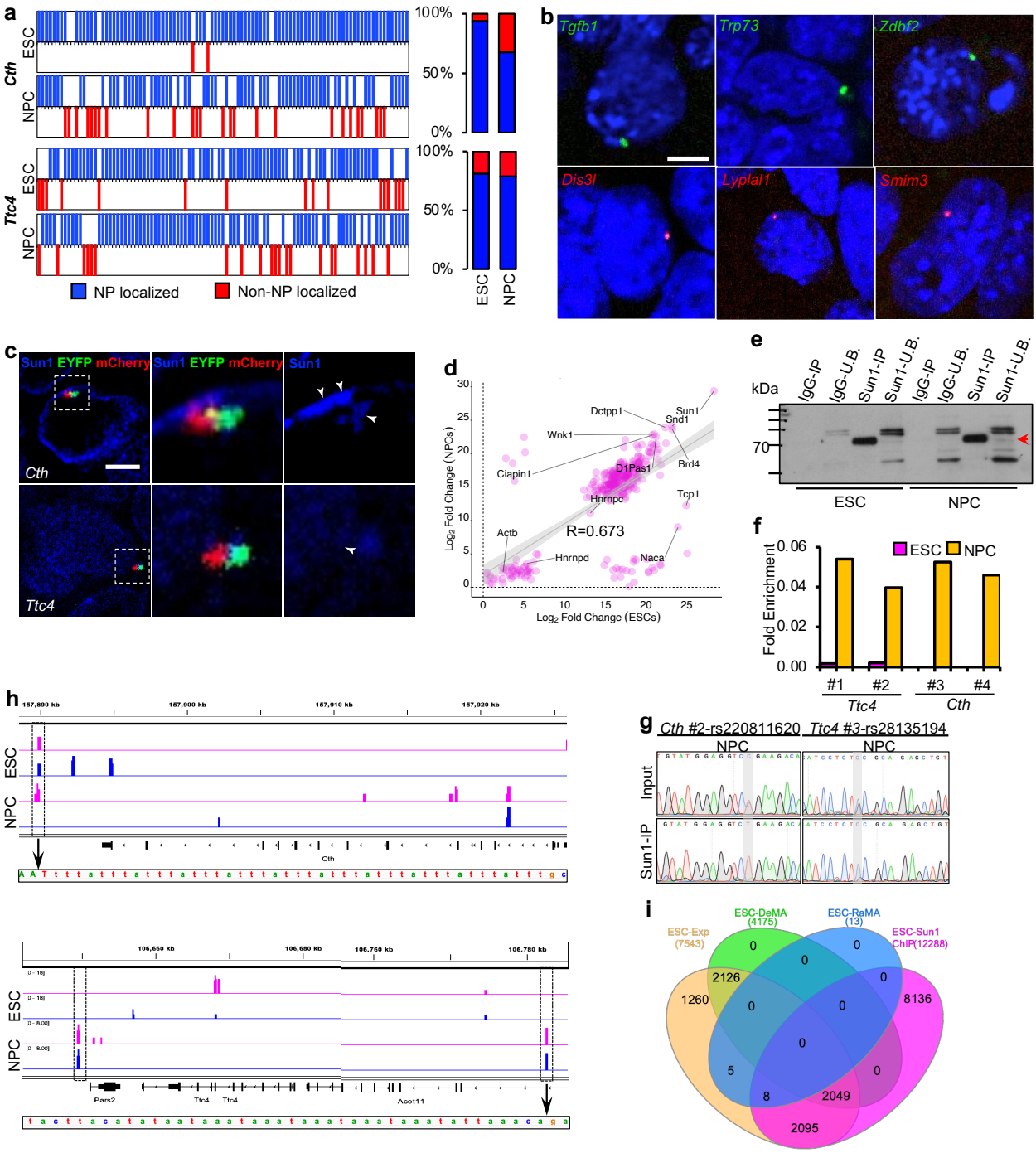


Figure 2 | Active alleles are localized at Sun1 protein territories.
a, Quantitation of *Cth* and *Ttc4* nascent RNA-FISH signals localization in the nucleus. Each line represents a counted cell and empty points on the X-axis presents the cells with one signal at the nuclear periphery and the other signals more internally in the nucleoplasm. Blue and red lines indicate nuclear periphery or nucleoplasm localized nascent RNA-FISH signals, respectively (n=100 cells in each group). Column graphs showing the percentages of the RNA-FISH signals at the nuclear periphery versus nucleoplasm in ESCs and NPCs. **b,** Representative nascent RNA-FISH images of several transcriptionally active random autosomal and deterministic monoallelic genes in NPCs. Imprinted genes are labeled in green font, and stochastic monoallelic genes are labeled in red font. Scale bars - 6 µm. **c,** *Cth,* and *Ttc4* genes' nascent RNA localization at Sun1 protein territories. Scale bars - 6 µm. **d,** Correlation of proteins pulled

down by Sun1 antibody in mESCs and NPCs. The highest enrichment was shown for the Sun1 protein confirming the specificity of the Sun1 antibody. The overall Pearson's correlation R is 0.673 with 95% confidence interval from 0.606 to 0.73 and p-value of 3.9e-41. **e,** Immunoblot of the Sun1 pull-down. Sun1 antibody recognize five Sun1 protein isoforms (lane Sun1-U.B). **f,** Enrichment of *Cth* and *Ttc4* genes in Sun1 pulled-down chromatin. Enrichment was assayed by qRT-PCR using primer sets (*Cth*: #1 and #2, *Ttc4*: #3 and #4) covering known SNPs. Fold enrichment was calculated against the input. **g,** Histogram of the Sanger sequenced qRT-PCR products from Sun1 pull-down chromatin from NPCs. Parental allele status in the pull-down was confirmed by SNPs (*Cth*: rs220811620-primer set #2 and *Ttc4*: rs28135194-primer set #3) unique to each allele. **h,** Allele-specific genome browser tracks illustrating Sun1 ChIP-seq peaks. TA-bridge motif-containing peaks associated with *Cth* and *Ttc4* genes are shown by the dotted boxes. Pink lines: Maternal alleles, blue lines: Paternal alleles. **i,** Venn diagram showing the intersects of Sun1 associated genes (including X and Y chromosome genes) identified in ChIP-seq assay and DeMA, RaMA, and the total expressed genes in mESCs. RNA expression data is from Balasooriya and Spector, Nat. Commns, 2022[22].

Sun1 in doxycycline-treated *Cth* and *Ttc4* mESC clones identified nascent RNA at both alleles of the *Cth* and *Ttc4* genes that are localize in association with Sun1 at the NP (Fig.2c and Extended Data Fig.2e). Therefore, we hypothesized that monoallelic genes might preferentially localize at the Sun1 territories at the NP. To investigate this further we performed Chromatin ImmunoPrecipitation (ChIP) using a Sun1 antibody in mESCs and *in vitro* derived NPCs. First, to validate the Sun1 antibodies' specificity to Sun1 proteins, we performed MS analysis for the protein isolated from the Sun1 ChIP assay. In MS data from mESCs and NPCs, Sun1 protein appeared as the highest significant protein in the pull-down (Fig.2d and Extended Data Table 1), confirming the antibody specificity to Sun1, which was subsequently confirmed by immunoblot analysis (Fig.2e).

By performing qPCR analysis from the chromatin isolated from the Sun1 ChIP assay followed by Sanger sequencing, we confirmed that both *Ttc4* alleles are localized at the Sun1 territories in mESCs, but only a single allele was associated with Sun1 in NPCs (Fig.2f,g and Extended Data Fig.2f,g). Although we observed only a single *Cth* allele pulled-down in NPCs (Fig.2g), we could not amplify the *Cth* gene from the DNA pull-down by Sun1 in mESCs (Fig.2f), perhaps due to the low yield of *Cth* in the Sun1 pull-down. Therefore, to assess the genes associated with Sun1-enriched chromatin in a high throughput manner, we deep-sequenced the DNA fragments isolated from the Sun1-ChIP assay. In the allele-specific ChIP-seq analysis, we

147 observed that the ChIP-seq peaks were enriched for both alleles in mESCs and a single allele in
148 NPCs at ~4.5 kb downstream of the *Cth* gene (Fig.2h – upper panel). However, we observed the
149 ChIP-seq peak enrichment for the *Ttc4* alleles at different regions in mESCs vs NPCs (in mESC:
150 on the gene body of *Ttc4* or neighboring gene *Mroh7* and in NPCs: ~4 kb
151 downstream *Fam151a* and *Acot11* gene bodies) (Fig.2h – lower panel). In addition, the
152 monoallelic genes we used in the nascent RNA-FISH assay were also found to be overlapping with
153 Sun1 protein territory associated genes (Extended Data Fig.2h,i). Moreover, we also found that
154 49% of transcriptionally active DeMA genes and 61.5% of RaMA genes in mESCs from our
155 previous study[22] overlapped with the genes enriched in Sun1 associated chromatin in mESCs
156 (Extended Data Table 2); however, the enrichment of transcriptionally active biallelic genes is
157 27.8% (Fig.2i). Furthermore, when we assessed how common these monoallelic gene associations
158 were in Sun1 territories in NPCs, we found that 74.1% of monoallelic genes in mESCs are enriched
159 in the Sun1 chromatin pull-down (Extended Data Fig.3j and Extended Data Table 3).
160

161 **"micro-TADs" within LADs harbor active genes**

162 Although lamina-associated domains (LADs) are shown to harbor heterochromatin[51-54],
163 emerging evidences suggest that genes within LADs are not all transcriptionally inactive[55,56]. Since
164 we observed the active alleles of monoallelic genes to be preferentially localized at the NP, we
165 examined the *Cth*, *Ttc4* and the monoallelic gene panels we used in the nascent RNA-FISH assay
166 for their association in published Hi-C-defined compartments in ESCs and NPCs[57,58] using the
167 4DN data portal[59,60]. We found that compared to low-resolution compartments, in high-resolution
168 compartments defined by an insulation score-diamond, the *Cth*, *Ttc4* genes (Fig.3a, Extended Data
169 Fig. a,b) and the monoallelic gene panel we used (Extended Data Fig.3c,d) are localized in "micro"

170 A compartments within B compartments. This finding confirms our observation of active
171 monoallelic genes at the NP and also suggests the possibility of accommodating active genes
172 within LADs by positioning them in "micro-TADs".
173
174 **TA-bridges are enriched in monoallelic genes**
175 Chromatin loci are found to be positioned at specific domains in 3D nuclear space.
176 Although it has been shown that the nuclear lamina functions as a scaffold to anchor chromatin at
177 the NP via nuclear envelop transmembrane proteins (e.g. HP1, Lap2β, cKrox, Lamin-B receptor)[61-
178 63], the DNA motifs that are involved in chromatin positioning at the nuclear envelope have been
179 poorly documented. Since monoallelic genes are preferentially localized at Sun1-enriched
180 territories, we asked whether specific motif sequences are present in the Sun1 protein territory
181 associated chromatin. We discovered that T_1A_3 (in ESCs) and A_1T_3 (in NPCs) tandem repeats (we
182 refer to them as TA-bridges) are significantly (ESCs: p=1e-511 and NPCs: p=1e-242) enriched in
183 Sun1 pulled-down chromatin fractions (Fig.3c, Extended Data Fig.3e,f, Extended Data Table 4).
184 We also found the presence of TA-bridges at ~5 kb upstream to the TSS and associated with the
185 fourth intron of *the Cth* gene, and ~5 kb downstream and ~105 kb upstream (on the first intron
186 of *Acot11*) of the *Ttc4* gene (Fig.2h). In addition, we further noticed that genes surrounding these
187 motifs (For *Cth*: *-Zranb2, +Ankrd13c, +Srsf11, +Lrrc40* for *Ttc4*: *-Pars2, -Dhcr24, +Ssbp3,*
188 *+Mrpl37, +Cdcp2*) are also known monoallelically expressed genes. Furthermore, we also found
189 that TA-bridges are associated with the monoallelic gene panel we used in our nascent RNA-FISH
190 assay suggesting an association between TA-bridges and the establishment and/or maintrnance of
191 monoallelic expression (Extended Data Fig.2h,i – lower panels of each track).

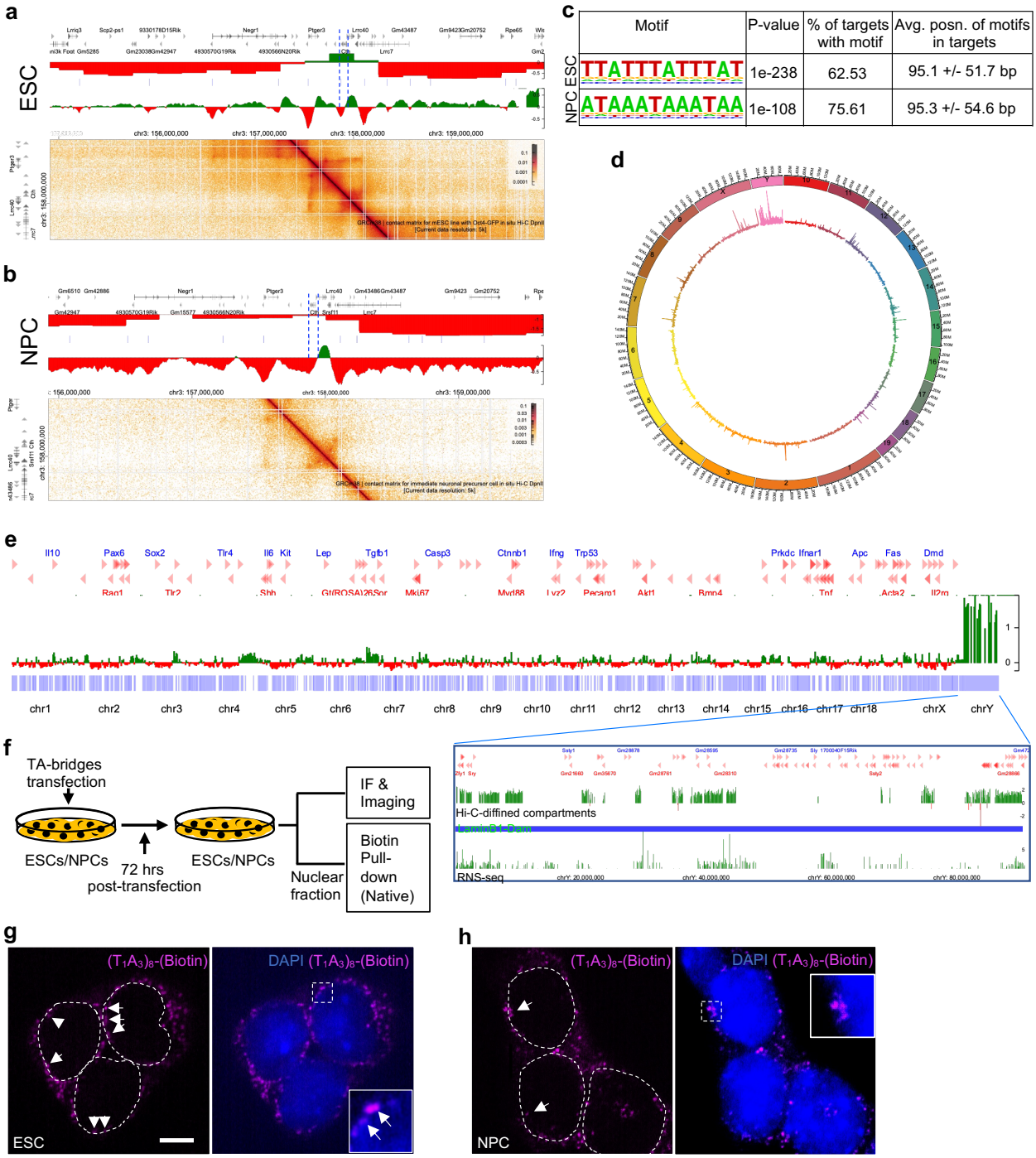


Figure 3 | TA-bridges position the *Cth* gene in micro-TADs within the chromatin B compartment.
a & b, *Cth* positioning in Hi-C-defined "micro-TADS". The *Cth* gene is located in the A compartment (green) within the predominantly represented B compartment (red). Contact Matrix (lower panels) for each gene showing the broader TADs and the "micro-TADs" within the TADs. Vertical lines show the boundaries of the TADs. Upon ESCs differentiation to NPCs, the monoallelically expressed *Cth* allele A compartment represents an isolated island within the B compartment. Data is from references 59, 60, 61 and 62. **c,** *De novo* TA-bridge DNA motif. **d,** genome-wide association of TA-bridges shown in circular genome representation. The Y chromosome is enriched with TA-bridges. Hight of the lines represent the score of the motif at the specific genomic region. We used Motif Cluster Alignment and Search tool (MCAST) for genome wide TA-bridge motif analysis using below parameters. Sequence length: 11

bases, Motif hit threshold: Hit p-value < 5.0E-4, Adjacent motif spacing: less than or equal to 50, Match threshold: E-value < 10.0 **e,** Genome-wide Hi-C-defined insulation score-diamond A (green) and B (red) compartment distribution in ESCs. Marked lower panel showing the magnified Y chromosome for its insulation score-diamond (upper track), LaminB1-DamID (middle track) and RNA-seq read counts (lower track). **f,** Schematic diagram showing the assay development to visualize the TA-bridges *in vivo* localization and TA-bridge *in vivo* interacting protein discovery. **g** & **h,** transiently transfected $(T_1A_3)_8$ single stranded DNA motif localization in the nuclei. mESCs and NPCs were transiently transfected with biotin tagged $(T_1A_3)_8$. At 72 hrs post-transfection, cells were fixed and immunolabeled for biotin and DNA was labeled with DAPI.

192 By genome-wide scanning, using our motif search criteria, we scored 4572 TA-bridge
193 motifs spanning through the whole mouse genome. From the total, we found 887 (19.4%) of these
194 motifs were located on the Y chromosome (Extended Data Fig.3g). Given the the comparatively
195 smaller size of the Y chromosome and the higher enrichment of TA-bridges than on other
196 chromosomes (Fig.3d), and its predominant localization to the A compartment (Fig.3e), we
197 assessed the location of the Y-chromosome territory in the nucleus using published LaminB1-
198 DamID data in ESCs in 4DN data portal[59,60]. Although the LaminB1-DamID data
199 (4DNFI32K5L77 – from Bas van Steensel lab, NKI, NL) showed the Y-chromosome is localized
200 at the NP (Fig.3e-middle panel of enraged area), Hi-C-define compartments[57] found that the Y
201 chromosome is predominantly in the A compartment and that the RNA expression profile[64] agreed
202 with the active compartment profiles (Fig.3e-lower panel of enraged area) further suggesting the
203 presence of "micro-TADs" within LADs.
204

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

206 To further confirm that the TA-bridge motif associates with allele localization to Sun1
207 territories, we developed an assay to visualize TA-bridge localization using immunofluorescence
208 microscopy, and identified the *in vivo* TA-bridge interacting proteins upon pull-down followed by
209 MS (Fig.3f). We transiently transfected biotin-labeled single stranded $(T_1A_3)_8$ and $(A_1T_3)_8$ DNA
210 oligonucleotides, and double-stranded $(T_1A_3)_8$ DNA into mESCs and NPCs. In our

211 immunofluorescence assay, we observed that biotin-labeled single-stranded $(T_1A_3)_8$ and $(A_1T_3)_8$
212 DNA, as well as double-stranded $(T_1A_3)_8$ DNA, were located at the NP among the nuclear-
213 localized signals in ESCs and NPCs (Fig.3g,h and Extended Data Fig.3h).

214 Next, we attempted to identify the protein candidates interacting with TA-bridges by
215 performing MS analysis for the pritein pull down by AT-bridges. Although the previously known
216 Sun1 binding proteins (Lmnb1, Lmnb2, Emerin) were pulled-down by TA-bridges, Sun1 was not
217 significantly enriched in the $(T_1A_3)_8$ and the double-stranded TA-bridges DNA motif pulled-down
218 protein fractions (Fig.4a,b and Extended Data Table 5), and also was not pulled-down by the
219 $(A_1T_3)_8$ motif (Extended Data Fig.5a), therefore, we could not conclude that Sun1 is the direct
220 binding protein of the TA-bridges. To further investigate the "molecular hook" that binds to the
221 double-stranded TA-bridges, we performed an electrophoretic mobility shift assay (EMSA) using
222 nuclear protein extracted from mouse ESCs. In the EMSA assay, we observed multiple shifted
223 biotin-positive bands for $(T_1A_3)_8$ and $(A_1T_3)_8$ since those DNA probes most likely interacted with
224 multiple protein complexes (Fig.4c). Interstingly, even though we observed that the double-
225 stranded TA-bridge DNA motifs were localized at the Sun1 territory at the NP in ESCs and NPCs
226 in the immunofluorescence assay, double-stranded TA-bridge DNA motifs did not show any
227 binding with nuclear proteins in our EMSA assay.

228

229 **Hnrnpd binds to TA-bridges**

230 Although we found several protein candidates that are known to interact with single-
231 stranded DNA in TA-bridge MS data, to identify the proteins that directly bind to TA-bridges, we
232 performed MS analysis of the proteins extracted from the multiple shifted bands identified in the
233 EMSA assay. MS data from the five excised bands, did not identify Sun1 protein enrichment

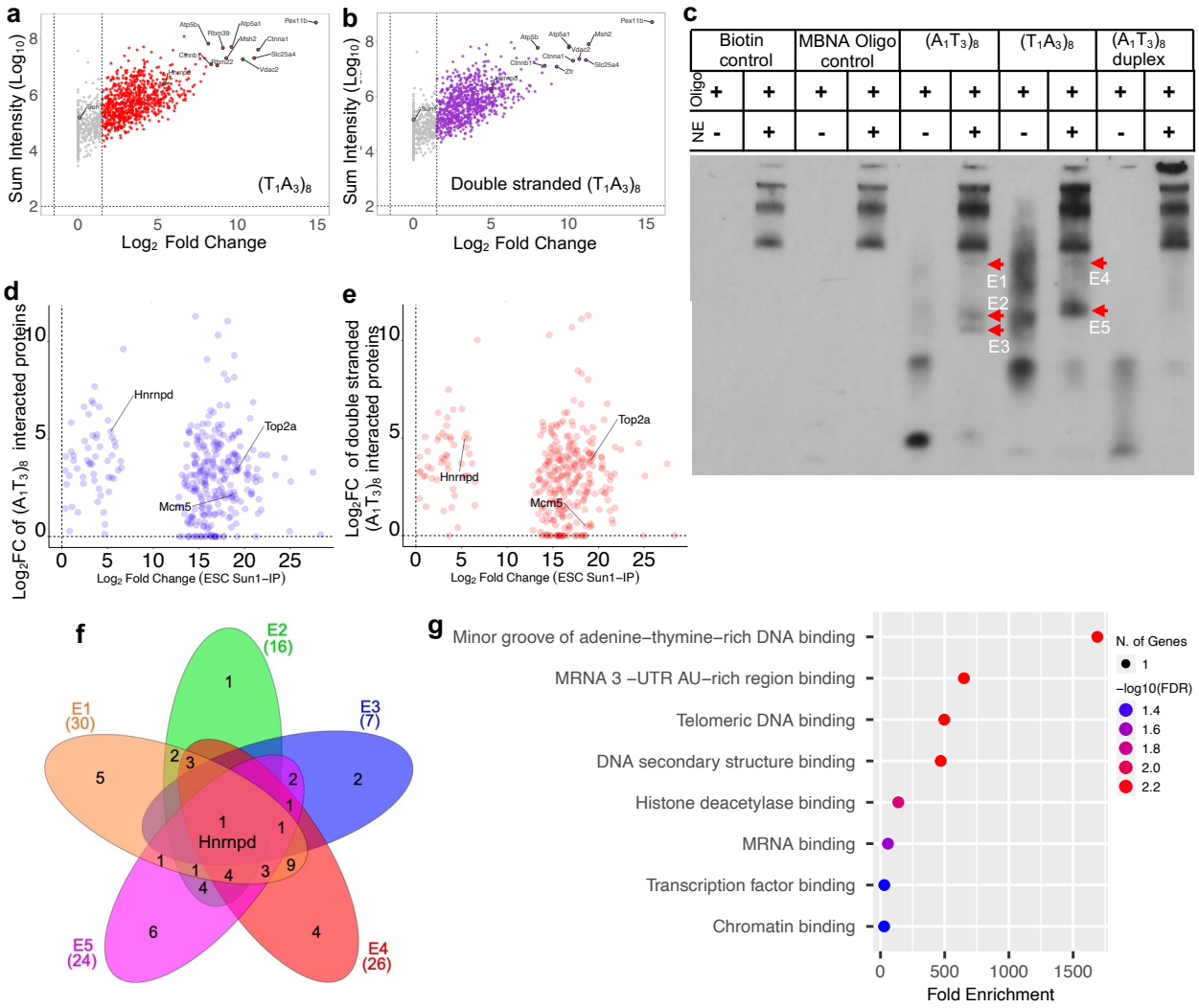


Figure 4 | Hnrnpd interacting TA-bridges localize to Sun1 territories.
a & **b,** Volcano plots showing the pull-downed single-stranded $(T_1A_3)_8$ (**a**) and double-stranded $(T_1A_3)_8$ (**b**) bridges *in vivo* interacting proteins. Sun1 proteins were pulled down by both $(T_1A_3)_8$ and double-stranded $(T_1A_3)_8$ but were not significantly enriched ($Log_{10}2$ of sum intensity and $Log_{10}1.5$ of fold change was used as cut-off values). **c.** EMSA assay for $(T_1A_3)_8$, $(A_1T_3)_8$, and double-stranded $(T_1A_3)_8$ bridges. ESC nuclear extract was used. MBNA oligos provided with the assay kit were used as the negative control. Proteins were extracted from three excised shifted bands in the $(A_1T_3)_8$ bridge lane and two shifted bands in double-stranded $(T_1A_3)_8$ bridge lanes. Higher molecular weight biotin-positive bands in all the samples are unspecific interactions. **d** & **e,** Comparisons of significantly enriched proteins pulled-down with Sun1, $(T_1A_3)_8$ single-stranded DNA and $(T_1A_3)_8$ double-stranded DNA. Each dot represents a protein. Hnrnpd shows similar enrichment for $(T_1A_3)_8$ single-stranded DNA (**d**) and $(T_1A_3)_8$ double-stranded DNA (**e**). **f,** Venn diagram showing the intersects of the highly significant protein clusters from five bands excised from the EMSA assay blot. Hnrnpd is the common protein extracted from all the bands. **g,** Gene Ontology (molecular function) for Hnrnpd. FDR 0.05 was used as cut-off threshold.

234 (Extended Data Fig.4b and Extended Data Table 6), confirming that Sun1 does not directly interact
235 with TA-bridges. However, intersecting the highest significant protein (Extended Data Fig.4b –
236 circled clusters) candidates from multiple bands from the EMSA assay, we identified Hnrnpd as
237 the only common protein candidate, and it was also enriched in proteins pulled-down by Sun1
238 (Fig.2d). To further investigate the Hnrnpd and TA-bridge direct interaction, we evaluated known
239 functions attributed to Hnrnpd. With a similar function to AT-rich interacting domain (ARID)
240 containing proteins, it has been reported to bind to the adenine-thymine-rich miner groves of B-
241 form of DNA[65] (GO:0003680) (Extended Data Fig.4c), supporting our identity of Hnrnpd as a
242 "molecular hook" that binds TA-bridge tandem repeats. Moreover, our finding suggests that,
243 Hnrnpd could bind to the non-canonical 5'-$(T_1A_3)_n$-3' as well as the 5'-$(A_1T_3)_n$-3' tandem repeat
244 sequences. Further, to investigate whether Hnrnpd binding to the TA-bridges occurs prior to or
245 after TA-bridges localize to Sun1-enriched territories, we co-immunoprecipitated the Sun1
246 associated proteins in nucleic acid-depleted ESC and NPC nuclear extracts using Sun1 antibody
247 and then performed MS analysis. Upon evaluation of the proteins immunoprecipitated with Sun1,
248 we found that Hnrnpd was less abundant in the Sun1 pulled-down protein fraction from the ESC
249 nuclear extract (Fig.4d) and could not be detected in the Sun1 pulled down protein fraction from
250 NPC nuclear extract (Fig.4e and Extended Data Table 7). These findings suggest that Hnrnpd most
251 likely binds to the AT-bridge motifs prior to the allele's localizing to the Sun1-enriched territories.
252

253 **Discussion**

254 Although allele pairing has been observed for decades, the presence of this phenomenon
255 in mammals has not been widely studied[38]. Moreover, it has not been previously shown that allele-
256 pairing is critical for some genes to be expressed from both alleles. Using the *Cth* and *Ttc4* genes,

257 which are biallelically expressed in mESCs, but become monoallelic upon differentiation, here we
258 showed that allele-pairing for some genes is concomitant with transcriptional activity from both
259 alleles. When the alleles are separated, only the allele that is associated with the Sun1-enriched
260 domain is active. Establishment of monoallelic expression, in terms of gene imprinting, is
261 described as a mechanism to control the transcriptional dosage of dosage-sensitive genes in early
262 development, but no in depth studies have been carried out to evaluate whether the imprinted genes
263 could be expressed biallelically in differentiated cells in adults or in some reare diseases. Even
264 though the *Cth* and *Ttc4* genes acquire monoallelic status in differentiated cells, their prior
265 biallelic expression, distinctive from imprinted genes, suggests the plasticity of the inactive allele's
266 transcriptional status. Consequently, the spontaneous silent allele's reactivation of "disease
267 causative monoallelic genes" presumably may bear clinical relevance. Further, our finding of the
268 active alleles of monoallelic genes to be localized to Sun1-enriched territories at the NP suggests
269 that the expression of those genes may be facilitated in specialized nuclear compartments.

270 Our observation of monaolleic genes' preferential expression at the NP supports the idea
271 that the NP localized chromatin is not solely heterochromatic[66-70]. Evaluating published data, we
272 also showed that the active allele of monoallelic genes are cluster-embedded within the B
273 compartment at the NP. However, why a large number of monoallelic genes acquire this unique
274 arrangement at the NP is enigmatic. In our previous studies, we showed that most monoallelic
275 genes contain only one RNA isoform known to be translated into a protein[22]. Therefore, it is
276 reasonable to argue that NP positioning may help RNAs to be exported through nuclear pore
277 complexes to avoid the high rate of RNA turnover in the nucleus, as a "defensive mechanism," as
278 proposed in the gene gating hypothesis, an idea worth pursuing in future investigations. Moreover,
279 given that the transcriptionally active alleles are localized at the NP, these genes (monoallelic:

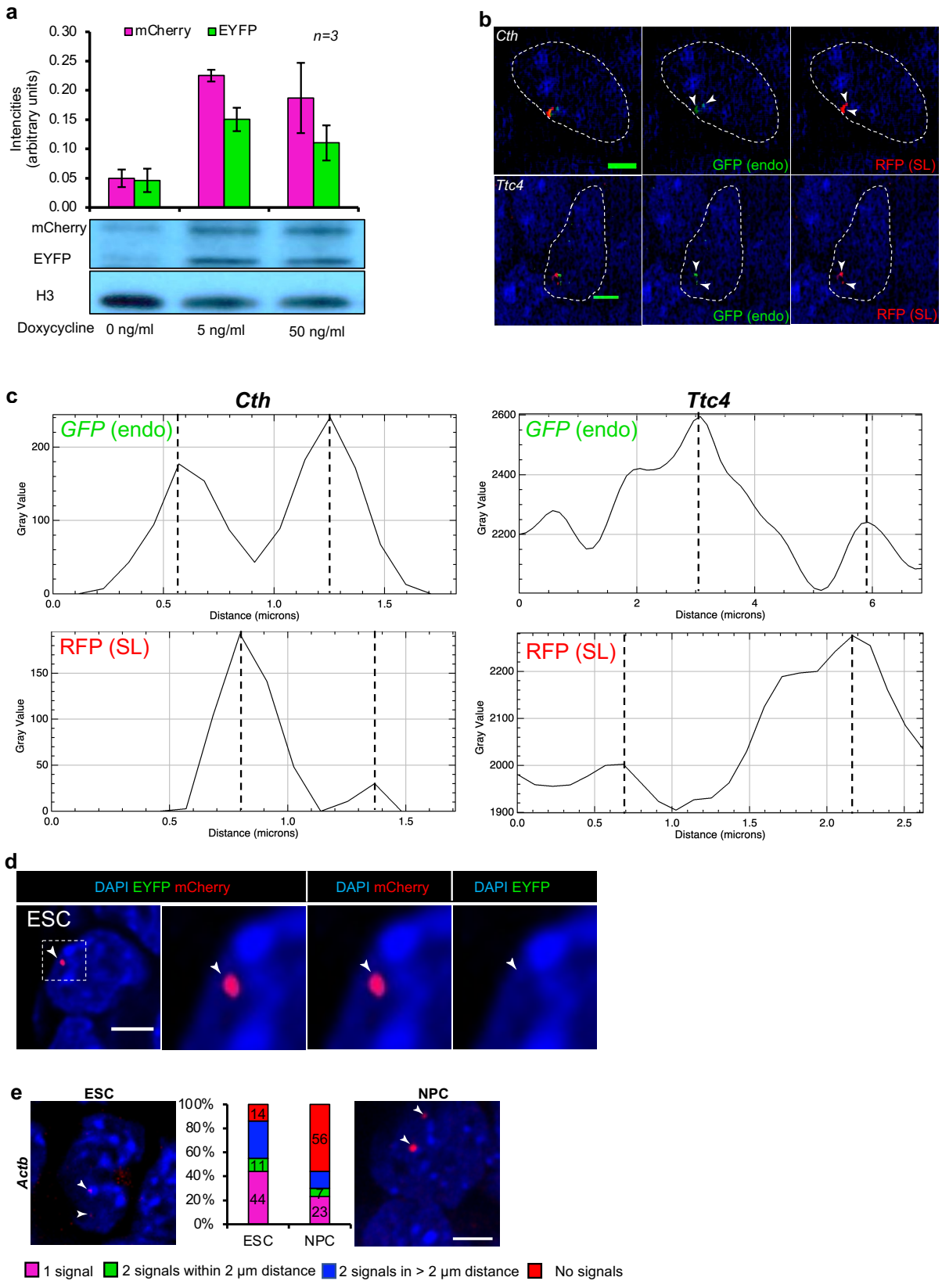
280 Rac1, S100a10, Junb. Biallelic: Sun1, Scn5a) are at risk of miss-regulation in diseases caused by
281 laminopathy and nuclear deformation[71] and therefore may bear clinical relevance.

282 Although in association with SPDYA, SHELTERIN, and TTM complex proteins, Sun1
283 stabilizes the telomeric rings at the inner nuclear membrane[72], Sun1 does not act as a direct binding
284 partner to the DNA, therefore in line with our finding, it is likely part of a scaffold to harbor
285 chromatin. Discoveries of novel repetitive elements in genomes have been made using long-read
286 DNA sequencing methods[73,74]. However, thus far, only a few repetitive elements have been
287 identified in relation to genome regulation[75-78], the functional relevance of many remain elusive.
288 Although some transcriptionally active genes have been reported to be localized at the NP[66-70], the
289 mechanism by which alleles are physically localized and maintained at the NP is not known. Here,
290 we discovered that Hnrnpd physically interacts with TA-bridges, thus raising the potential
291 importance of other DNA tandem repeat elements yet to be uncovered. Although Hnrnpd binds to
292 TA-bridges, it is not an enriched component of Sun1 territories, suggesting that Hnrnpd binding
293 to the TA-bridges occurs prior to allele positioning to the Sun1 teritories. Further, we found that
294 known helicases (Mcm5[79] and Top2a[80]) are significantly enriched among TA-bridge interacting
295 proteins (Fig.4d,e) suggesting that TA-bridge strand separation likely initiates prior to the motifs
296 localization at the Sun1 nuclear territories.

297 During stem cell differentiation into more specialized cells, the gene expression program
298 changes concomitant with chromatin reorganization, resulting in different architectures depending
299 on the resulting cell type[81,82]. Furthermore, it has been shown that asynchronous allele replication
300 occurs more frequently during cell differentiation perhaps promoting the early replicating allele to
301 be established as the active allele of a monoallelic gene[83,84]. Because Hnrnpd binds to TA-bridges
302 independently of Sun1, we propose a mechanistic model where Hnrnpd binds to single-stranded

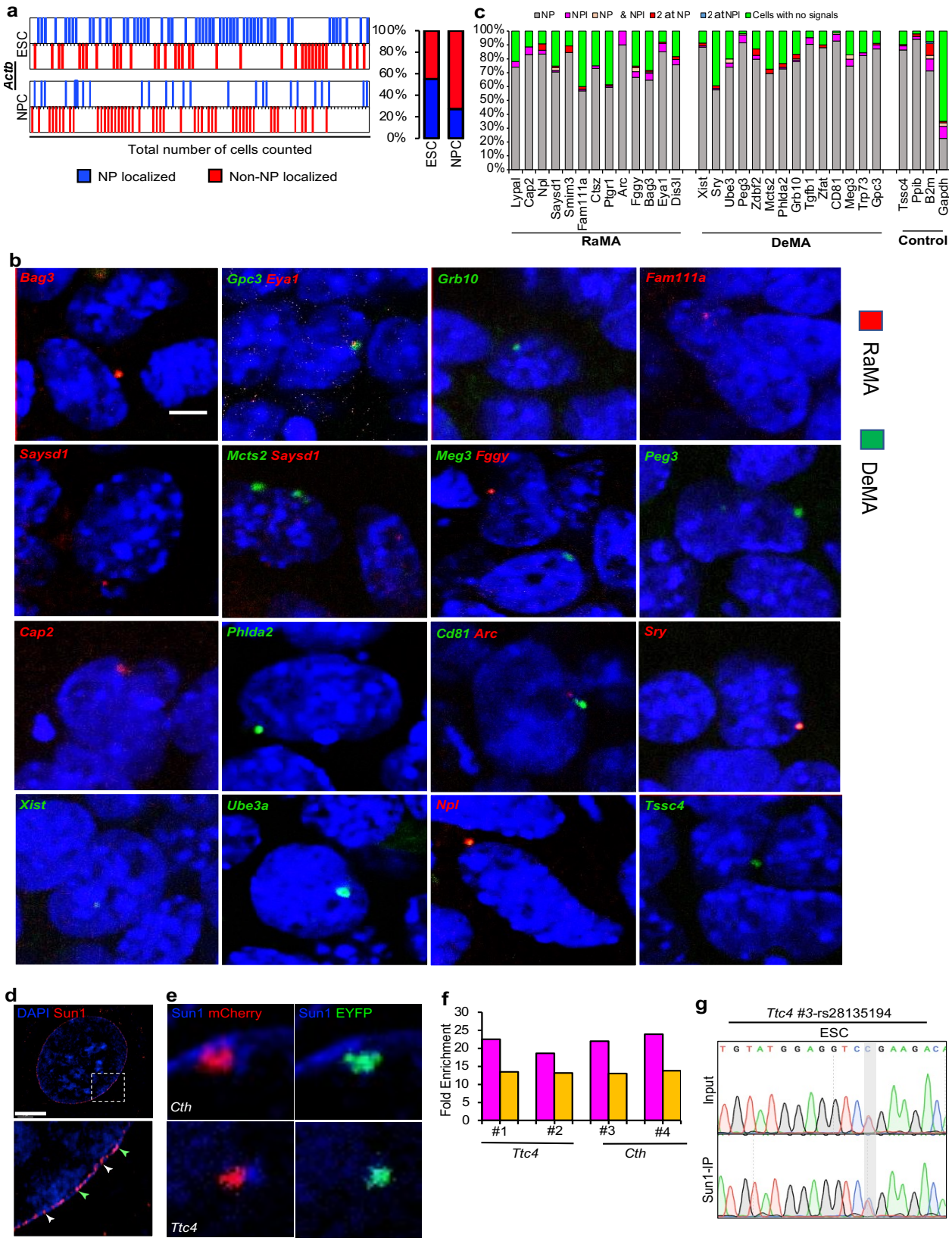
303 TA-bridges of the early replicating allele in prophase resulting in targeting of the TA-bridge bound
304 Hnrnpd to the newly forming nuclear membrane Sun1-enriched domains in telophase by a
305 molecular targeting mechanism to be elucidated in future sudies.

Extended Data Figures



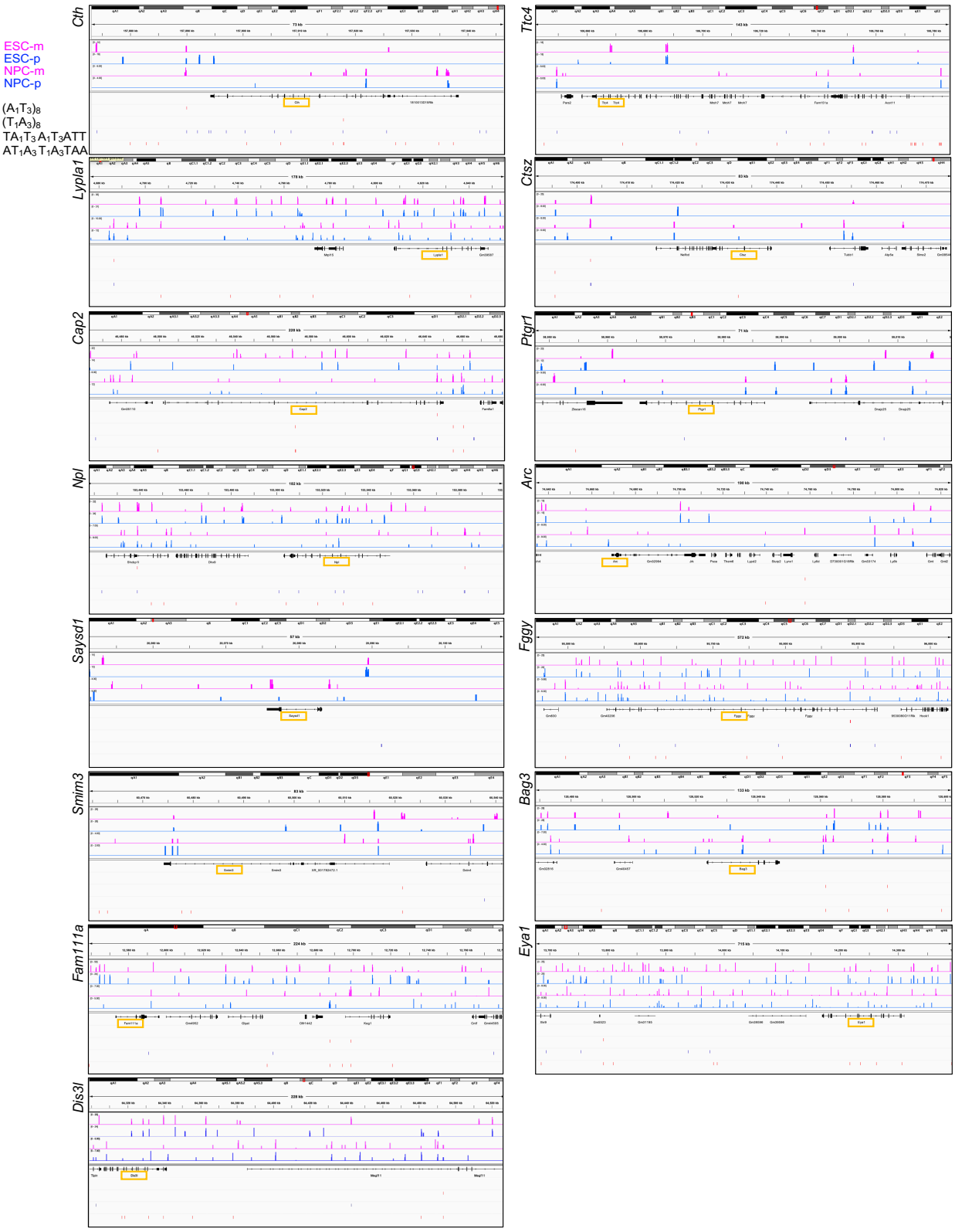
Extended Data Figure 1| Development of an allele-specific gene expression system to assess the localization of *Cth* and *Ttc4* alleles.

a, Immunoblot (lower panel) showing the EYFP and mCherry responsiveness to doxycycline dosage. Bar graph showing the results of quantitation of the EYFP (GFP antibody was used) and mCherry bands in immunoblots from three independent experiments (n=3). **b,** DNA-FISH of *Cth* and *Ttc4* genes targeted with MS2 and PP7 stem-loops in NPCs. Endogenous alleles are labeled with a GFP probe, and knocked in stem-loops are labeled with an RFP probe. Total DNA is labeled with DAPI. Two signals were detected for GFP and RFP in these positively labeled (71%) nuclei. We did not observe only the GFP or only the RFP signals in any nucleus. Scale bar – 6 µm. **c,** Distance between the two *Cth* alleles and the two *Ttc4* alleles in NPC nuclei. The average distances between the endogenous *Cth* alleles is ~0.6 mm and the two inserted stem-loops is ~0.55 mm, whereas the average distance between endogenous *Ttc4* alleles is ~2.5 mm and the stem-loops is ~1.5 mm. **d,** Monoallelic expression of *Ttc4* gene in mESCs. Although *Ttc4* was expected to be expressed biallelically in mESCs, we observed that *Ttc4* is monoallelically expressed in 31% of the mESCs. **e,** Nascent mRNA FISH assay for *Actb* in wild-type mESCs and NPCs. Arrowheads indicate the nascent RNA-FISH signals. The column graphs show the percentages of the number of observations made for the distance between the two alleles. Two alleles within 2 µm proximity to each other are considered as paired alleles. Scale bar - 6 µm. *Actb* gene's active alleles were paired in only 11% of mESCs, and the remaining cells positive for RNA-FISH signals showed single allelic (44% of cells) and biallelic (11% of cells) expression. Given that, compared to *Cth* and *Ttc4* (Fig.1f), the percentage of cells lacking signals (ESCs: 14% and NPCs: 56%) and also agreeing with our previous observation[22] of the imbalance of mature versus nascent RNAs of *Actb* and the published data[85], single allelic expression (44%) of *Actb* is perhaps due to gene pulsing.

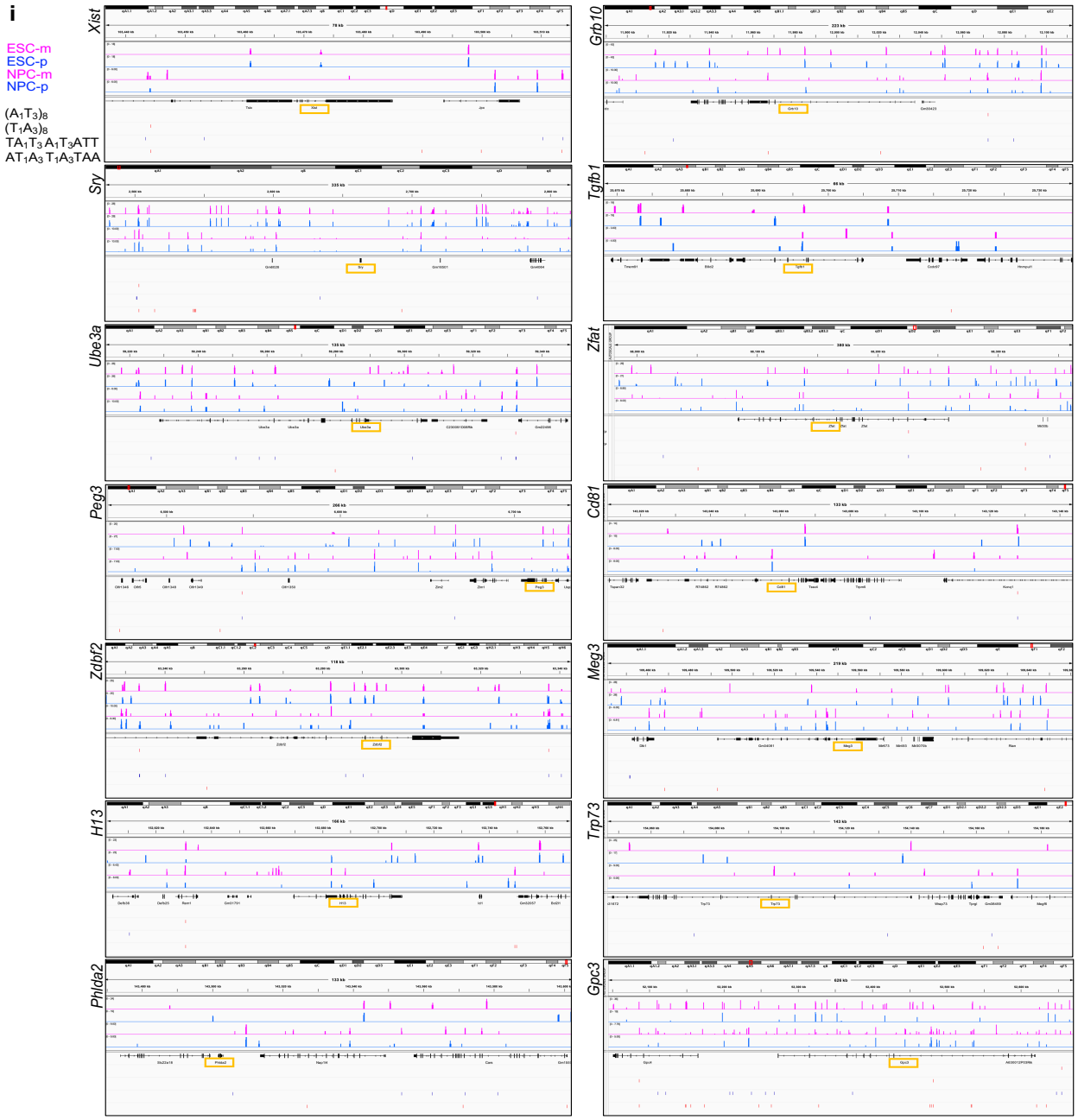

a
Actb
ESC
NPC
Total number of cells counted
NP localized
Non-NP localized
100%
80%
60%
40%
20%
0%
ESC
NPC
c
NP
NPl
NP & NPl
2 at NP
2 at NPl
Cells with no signals
Lypal
Cap2
Npl
Saysd1
Smim3
Fam111a
Ctsz
Ptgr1
Arc
Fggy
Bag3
Eya1
Dis3l
Xist
Sry
Ube3
Peg3
Zdbf2
Mcts2
Phlda2
Grb10
Tgfb1
Zfat
CD81
Meg3
Trp73
Gpc3
Tssc4
Ppib
B2m
Gapdh
RaMA
DeMA
Control
b
Bag3
Gpc3 Eya1
Grb10
Fam111a
Saysd1
Mcts2 Saysd1
Meg3 Fggy
Peg3
Cap2
Phlda2
Cd81 Arc
Sry
Xist
Ube3a
Npl
Tssc4
RaMA
DeMA
d
DAPI Sun1
e
Sun1 mCherry
Sun1 EYFP
Cth
Ttc4
f
Fold Enrichment
30
25
20
15
10
5
0
#1
#2
#3
#4
Ttc4
Cth
g
Ttc4 #3-rs28135194
ESC
Input
Sun1-IP

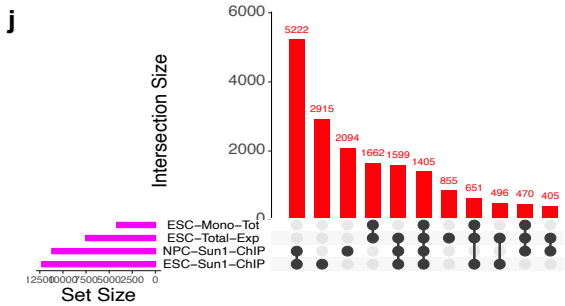
Extended Data Figure 2 Continued

h



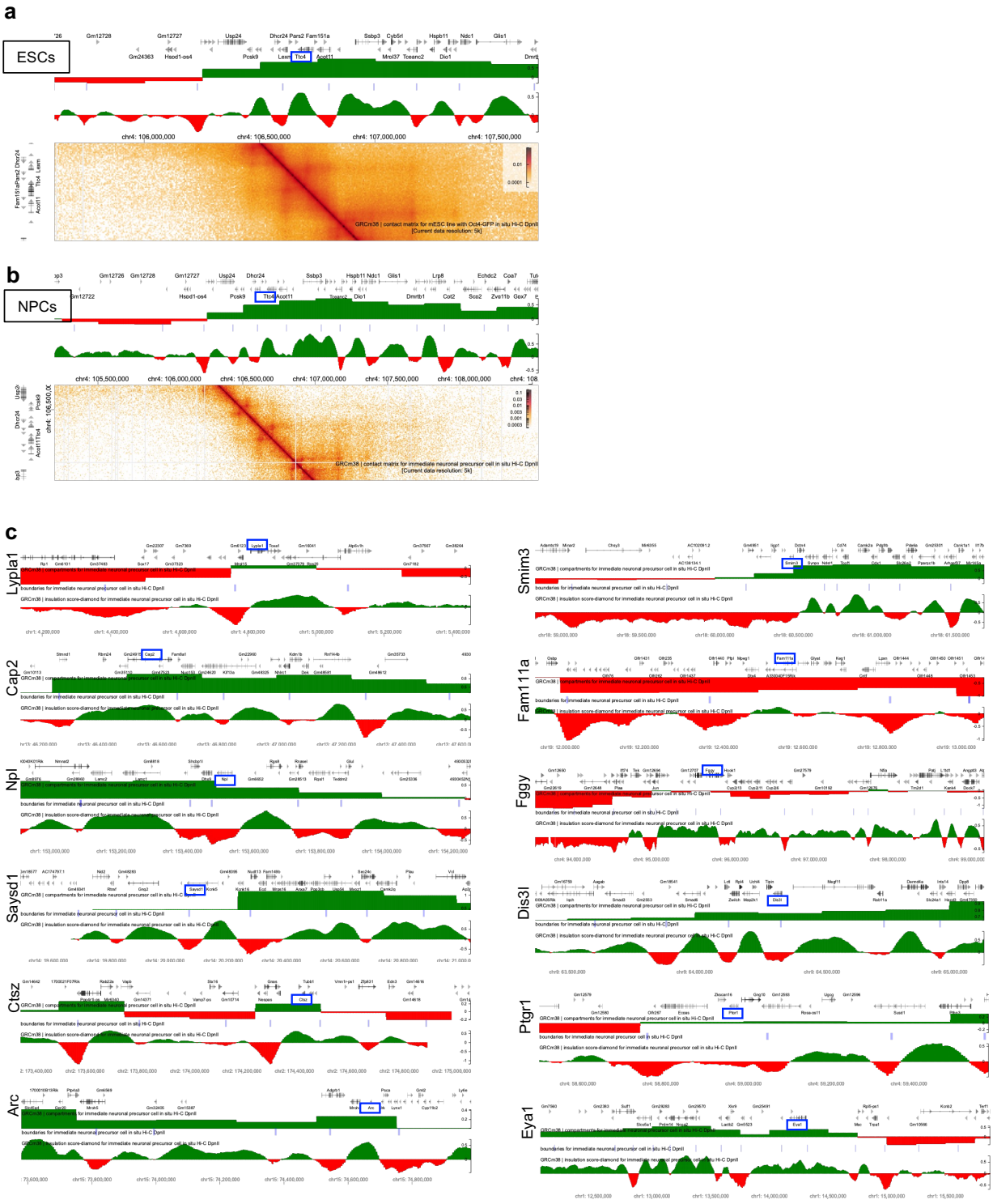
Extended Data Figure 2 Continued





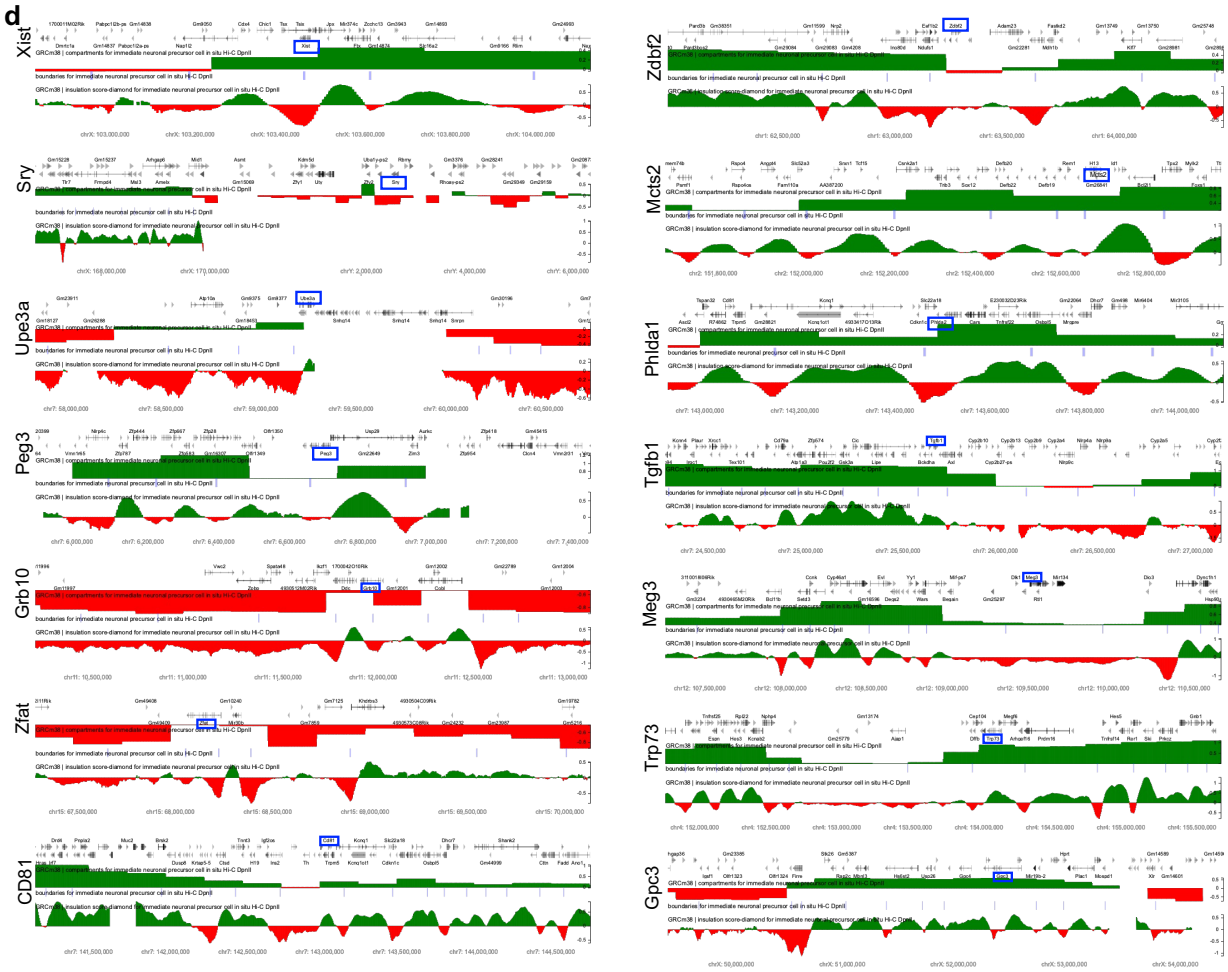
Extended Data Figure 2 | Active alleles of monoallelic and bi-allelic genes are localized at Sun1 protein territories.

a, Quantitation of *Actb* nascent RNA-FISH signals in the nucleus. Each line represents a counted cell and empty points on the X-axis presents the cells with one signal at the nuclear periphery and the other signal further inside the nucleoplasm. Blue and red lines indicate nascent RNA-FISH signals localized to the nuclear periphery or nucleoplasm (n=100 cells in each group). **b,** Representative images from nascent RNA-FISH assay for a panel of monoallelic genes. DeMA genes are labeled in green font, and RaMA genes are labelled in red font. Scale bars - 6 µm. **c,** Quantitation of the nascent RNA-FISH signals for RaMA and DeMA genes in NPCs. Y axis shows the percentage of cells with RNA-FISH signals located in the nucleus (NP: Nuclear periphery, NPl: Nucleoplasm). **d,** Immunofluorescent image showing Sun1 protein localization at the nuclear periphery. The enlarged dotted area is shown in the lower panel. Green arrowheads show Sun1 and DNA overlapped loci. White arrowheads show Sun1 with no DNA overlapping. Scale bars - 6 µm. **e,** *Cth* nascent RNAs and Sun1 protein and, *Ttc4* nascent RNA and Sun1 protein localization at the nuclear periphery. Scale bars 200 nm. **f,** Fold enrichment of Sun1 pull-down of the *Ttc4* and *Cth* genes. The same primer sets were used as in Fig.2f. Fold enrichment was calculated with reference to each gene's DNA fragment enrichment in input. **g,** Histogram of the Sanger sequenced qRT-PCR products from Sun1 pull-down chromatin from mESCs using *Ttc4* primers (rs28135194). Grey areas are showing the SNPs. Both maternal and paternal alleles were pulled down by Sun1. **h** & **i,** Allele-specific genome browser tracks showing the ChIP-seq peak association of RaMA (**h**) and DeMA (**i**) genes. Pink tracks: maternal alleles, blue tracks: Paternal alleles. The positions of the shortest TA-bridge motifs (n=2) and the longest (n=8) TA-bridge motifs are shown in the panels below the allele tracks. (n: number of motifs). **j,** UpSet plot showing the gene intersects between Sun1-ChIPed genes in ESCs and NPCs with total genes expressed and monoallelic genes (DeMA and RaMA) expressed in ESCs.


a
ESCs
chr4: 106,000,000
chr4: 106,500,000
chr4: 107,000,000
chr4: 107,500,000
GRCm38 | contact matrix for mESC line with Oct4-GFP in situ Hi-C DpnII
[Current data resolution: 5k]
b
NPCs
chr4: 105,500,000
chr4: 106,000,000
chr4: 106,500,000
chr4: 107,000,000
chr4: 107,500,000
chr4: 108,000,000
GRCm38 | contact matrix for immediate neuronal precursor cell in situ Hi-C DpnII
[Current data resolution: 5k]
c
Lypla1
Cap2
Npl
Saysd1
Ctsz
Arc
Smim3
Fam111a
Fggy
Dis3l
Ptgr1
Eya1
boundaries for immediate neuronal precursor cell in situ Hi-C DpnII
GRCm38 | insulation score-diamond for immediate neuronal precursor cell in situ Hi-C DpnII

Extended Data Figure 3 Continued

d



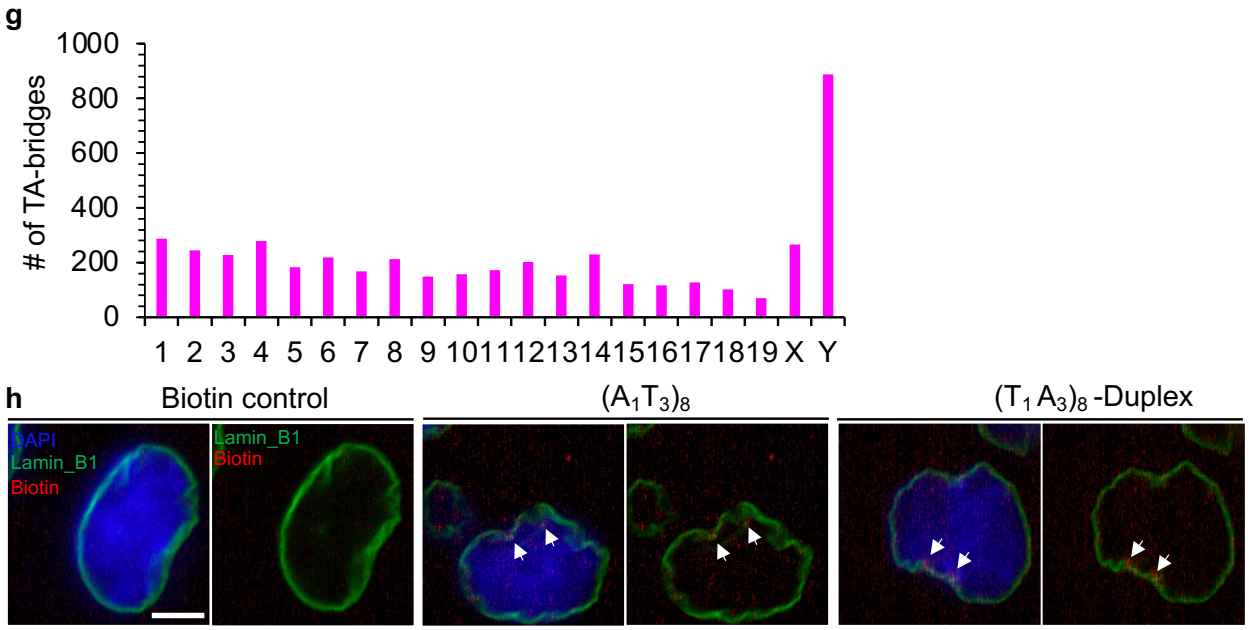
e ESC

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

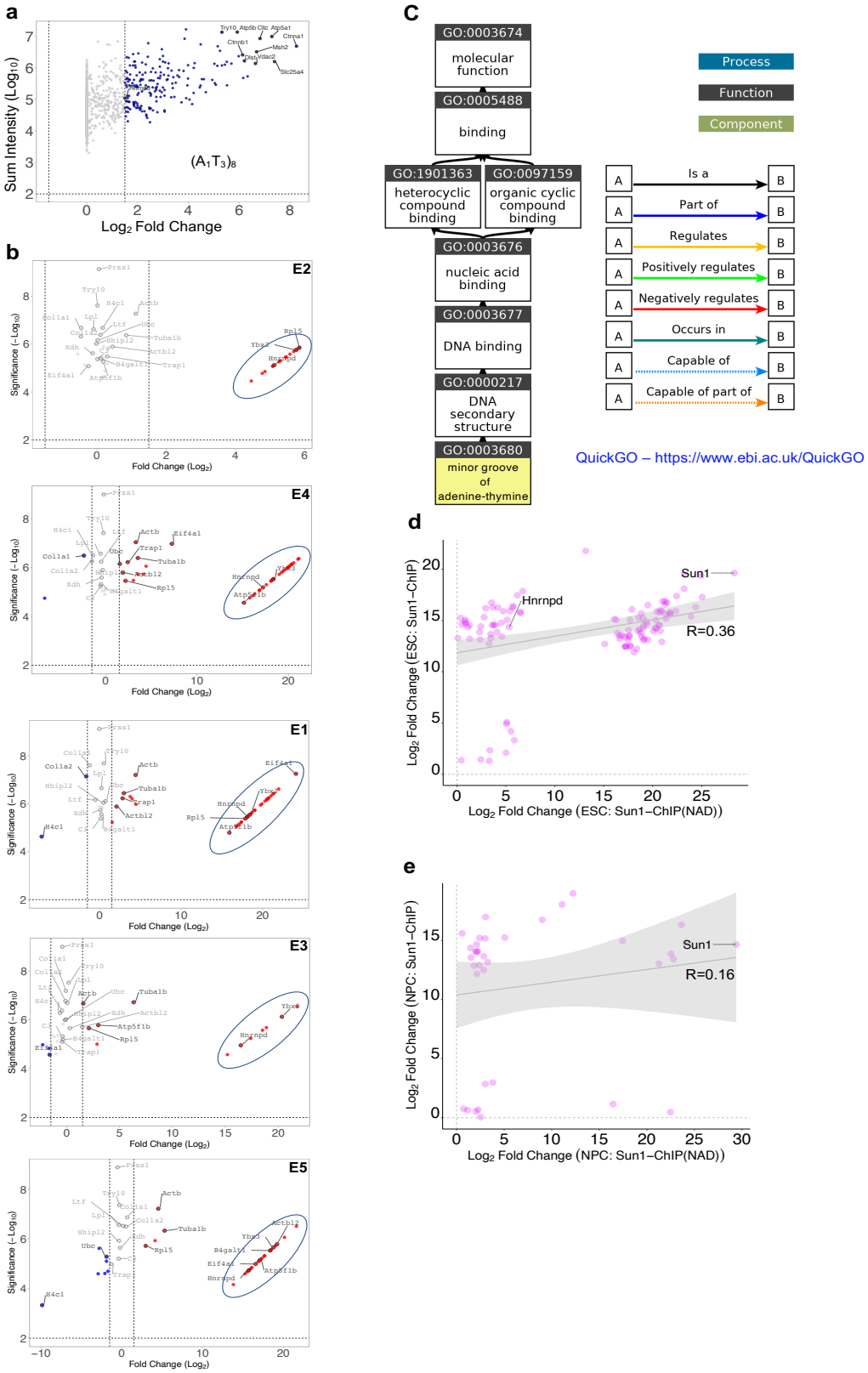
f NPC

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

Extended Data Figure 3 Continued



Extended Data Figure 3 | TA-bridge associated monoallelic gene's positioning in micro-TADs within the B chromatin compartment.
a, & **b,** *Ttc4* positioning in Hi-C-defined "micro-TADS". The *Ttc4* gene is located in the A compartment (green). Monoallelic genes are dense in the *Ttc4* associated A compartment within the B compartment. Contact Matrix (lower panels) for each gene showing the broader TADs and the "micro-TADs" within the TADs. Vertical lines showing the boundaries of the TADs. Data is from references 59, 60, 61 and 62. **c** & **d**, Association of DeMA and RaMA genes in "micro-TADs". c: RaMA genes and d: DeMA genes. **e** & **f,** De Novo motif discovery. Narrow peak search criteria were used to mine the shortest possible motifs. Only the nine highest significant motifs found in Sun1-ChIP chromatin in ESCs (**a**) and NPCs (**f**) are listed. **g,** Number of TA-bridges in each chromosome. Y chromosome contains the highest number of TA-bridges. **h,** Single-stranded $(A_1T_3)_8$ DNA and double-stranded $(T_1A_3)_8$ DNA localization in NPC nuclei. At 72 hrs post-transfection, cells were fixed and immunolabeled for biotin, Lamin B1 and DNA (DAPI). Lamin B1 was used to mark the nuclear periphery boundary.


a
Sum Intensity (Log10)
Log2 Fold Change
(A1T3)8
b
E2
E4
E1
E3
E5
Significance (−Log10)
Fold Change (Log2)
c
GO:0003674
molecular function
GO:0005488
binding
GO:1901363
heterocyclic compound binding
GO:0097159
organic cyclic compound binding
GO:0003676
nucleic acid binding
GO:0003677
DNA binding
GO:0000217
DNA secondary structure
GO:0003680
minor groove of adenine-thymine
Process
Function
Component
Is a
Part of
Regulates
Positively regulates
Negatively regulates
Occurs in
Capable of
Capable of part of
QuickGO – https://www.ebi.ac.uk/QuickGO
d
Sun1
Hnrnpd
R=0.36
Log2 Fold Change (ESC: Sun1−ChIP)
Log2 Fold Change (ESC: Sun1−ChIP(NAD))
e
Sun1
R=0.16
Log2 Fold Change (NPC: Sun1−ChIP)
Log2 Fold Change (NPC: Sun1−ChIP(NAD))

Extended Data Figure 4 | Hnrnpd interacting TA-bridges localize to Sun1 territories.
a, Volcano plots showing the significantly enriched $(A_1T_3)_8$ single-stranded motif *in vivo* interacting proteins. Sun1 protein is not pulled-down by $(A_1T_3)_8$ single stranded motifs. $Log_{10}2$ of sum intensity and $Log_{10}1.5$ of fold change were used as cut-off values. **b,** Volcano plots showing the significantly enriched proteins in five biotin bands (E1-E5) from the EMSA assay. Hnrnpd is present in all the highest significant protein clusters (circled). **c,** Diagram showing the functional gene ontology annotation of Hnrnpd protein mined using QuickGO. GO term 0003680 indicates it binds to the adenine-thymine-rich minor grove of DNA. **d** & **e,** Correlations of significantly enriched proteins. Sun1 pull-down proteins from non-nucleic acid depleted (d) and nucleic acid depleted (e) nuclear extracts in ESCs and NPCs. Hnrnpd is not enriched in the nucleic acid depleted Sun1 pulled-down proteins (ESC - The overall Pearson's correlation R is 0.359 with a 95% Confidence Interval from 0.186 to 0.511 and a p-value of 1e-4. NPC - The overall Pearson's correlation R is 0.156 with a 95% Confidence Interval from -0.182 to 0.461 and a p-value of 0.364).

306 **Methods**

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

308 Polymorphic F1 mESCs were generated and maintained as previously described[22]. NPC in vitro
309 differentiation was caried out as follows. Male F1 mESCs were seeded on 0.1% gelatin coated 8-
310 well chamber ibidi slides at $0.5x10^4$ cells per well in 300 ul of 2i medium. At 24 hrs post seeding,
311 cells were washed three times, five min each, with 1x PBS. After washing, 1 ml of differentiation
312 medium (10% FBS (Millipore, Cat.No. ES-009-B, 1X L-Glutamine (Thermo Fisher,
313 Cat.No.25030081, 1X Non-Essential AA, Millipore, Cat.No. TMS-001-C), 0.15 mM 2-
314 Mercaptoethanol (Millipore, Cat.No. ES-007-E), 100 U/ml Penicilline-Streptomycin (Gibco,
315 Cat.No. 15140-112) in Knockout DMEM (Gibco, Cat.no.10829-018)) was added and incubated at
316 37^0C, 5% CO_2. Around 12 hrs post-differentiation initiation, mESCs changed their morphology to
317 neural-like cells. At 24 hrs post differentiation induction, cells were washed 3 times with 1xPBS
318 and 2.5 ml of basic NPC differentiation medium was added supplemented with 2.5 ug/ul murine
319 EGF (PeproTech Inc, Cat.No. 315-09) and 2.5 ug/ul murine Fgf2 (PeproTech Inc, Cat.No.450-33)
320 and cultures were incubated for another 36 hrs. Subsequently, cells were trypsinised and seeded at
321 $1x10^5$ cells per well density with the medium supplemented with EGF and Fgf2 to maintain the
322 long-term NPC cultures. Cells were harvested by trypsinization and the total RNA was extracted
323 using Trizol (Thermo Fisher, Cat.No. 1559618). Using 1 ug of total RNA, cDNA was synthesized
324 with SuperScript lll (Thermo Fisher, Cat.No. 18080-093) reverse transcriptase using Oligo(dt)20
325 primers (Thermo Fisher, Cat.No.18418-020).

326

327

328

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

330 Generating a parental mESC line with mCherry-PP7CP and EYFP-MS2CP Tet-On inducible
331 expression system. First, Tet-On Inducible Expression System® plasmid (Clontech, Cat.No.
332 634301) was linearized by digestion with EcoRI and AgeI. Then, mCherry-PP7CP DNA fragments
333 were excised out (double-digested with BaeI from PP7-mCherry plasmid -from Spector lab) and
334 adaptors compatible with the digested Tet-On plasmid were ligated to the 5' end (annealed Oligo#1
335 + Oligo#2. 3' end: annealed Oligo#3 + Oligo#4) to generate Tet-On-PP7CP-mCherry plasmid.
336 EYFP-MS2CP fragment was excised from the pMS2-cp-YFP (Spector lab plasmid) with NheI and
337 NgoMIV and the adaptors (3'-end: annealed Oligo#5 + Oligo#6) compatible to the Tet-On
338 linearized plasmid were ligated to generate the Tet-On-MS2CP-EYFP plasmid. Then, both Tet-
339 On-PP7CP-mCherry and Tet-On-MS2CP-EYFP plasmids were double digested (with BglII and
340 PclI) to excise the expression cassettes. Subsequently, annealed Adpt A (Oligo#7 + Oligo#8) and
341 annealed Adpt B (Oligo#9 + Oligo#10) were ligated into the mCherry end of the excised
342 expression cassette and Tet-On-3G end of the Tet-On-PP7CP-mCherry plasmid, respectively.
343 Next, the annealed Adpt C (Oligo#11 + Oligo#12) and the annealed Adpt D (Oligo#13 +
344 Oligo#14) were ligated into the EYFP end of the excised expression cassette and Tet-On-3G of
345 the Tet-On-MS2CP-EYFP plasmid, respectively. These expression cassettes with overhanging
346 sequences complimentary to the insert site sequences of the Rosa26 alleles were used to transfect
347 into mESCs.

348 Oligos:

349 Oligo#1-CCGAGTGAGTCGTATTA

350 Oligo#2-TCGATAATACGACTCACT

351 Oligo#3-AAGCCAAAAAAGAAGAGAAAGGTATAA

352 Oligo#4-AATTTTATACCTTTCTCTTCTTTTTTGGCTTGGACA

353 Oligo#5-CTAGTAATACGACTCACTATAGG

354 Oligo#6-AATTCCTATAGTGAGTCGTATTA

355 Oligo#7-

356 GATCGGATGGATCCATCGGAAGGGGAGCCTTTCTCTCTGGGCAAGAGCGGTGCAAT

357 GGTGTGTAAAGGTAGCCGCTC

358 Oligo#8-CGATGGATCCATCC

359 Oligo#9- GGATGGATCCACTG

360 Oligo#10-

361 CATGCAGTGGATCCATCCTCTCGGCTACCTTTACACACCATTGCACCGCTCTTGCCCA

362 GAGAGAAAGGCTCCCCTTC

363 Oligo#11-GATCGGATGGATCCATCGTCTGAAGGAGAGCCTTTC

364 Oligo#12- CGATGGATCCATCC

365 Oligo#13-CGATGGATCCATCC

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

367

368 <u>Designing allele-specific gRNAs for expression cassette knock-in to the Rosa26 alleles.</u> gRNAs

369 were designed to include SNPs in the R26 alleles at the PAM site cloned into pSpCas9n(BB)

370 (PX460) (Addgene, Cat.No. 48873). The following gRNA sequences were used in combination.

371 gR26#1 (Annealed R26-5'Cast1.1 and R26-5'Cast1.2), gR26#2 (Annealed R26-3'Cast-Comp1.1

372 and R26-3'Cast-Comp 1.2), gR26#3 (Annealed R26-5'B6 1.1 and R26-5'B61.2), gR26#4

373 (Annealed R26-3'B61.1 and R26-3'B61.2).

374 Next, the following combinations of gRNAs and complimentary adaptor ligated expression
375 cassettes were transfected into the F1 mESCs using Xfect transfection kit (Takara,
376 Cat.No.631317).
377 To target the Tet-On-PP7CP-mCherry expression cassette to the R26 CAST (paternal) allele,
378 gR26#1, gR26#2 and Tet-On-PP7CP-mCherry combination was used. To targt the Tet-On-
379 MS2CP-EYFP to the R26 C57 (maternal) allele, gRNA#3, gRNA#4 and Tet-On-MS2CP-EYFP
380 combination was used. We used 2 ug of gRNA-Cas9 plasmids and 1.0 ug of adaptors in each
381 transfection. Prior to the transfection, mESCs were seeded at $0.5x10^5$ cell/ well in 6-well Corning
382 tissue culture plates (Corning, Cat.No. 353046) and incubated for 12 hrs. At 24 hrs post-
383 transfection, cells were rinsed with 10% FBS containing KnockOut™ DMEM(1X) (Gibco,
384 Cat.No.10829-018) and replaced with new 2i medium. At 48 post-transfection, single clones were
385 picked under the microscope, expanded and genotyped using R26-PP7/MS2-R26-F and R26-
386 PP7/MS2-R26-R.
387 Oligos:
388 R26-5'Cast1.1-CACCGAGTGGGGTCGACTATCTGAAGGG
389 R26-5'Cast1.2-AAACCCCTTCAGATAGTCGACCCCACTC
390 R26-3'Cast-Comp1.1-CACCGGCTTGCCCTTTTCGTCTTCTCGG
391 R26-3'Cast-Comp1.2-AAACCCGAGAAGACGAAAAGGGCAAGCC
392 R26-5'B61.1-CACCGTAGTGGGGTCGACTATCTGAAGG
393 R26-5'B61.2-AAACCCTTCAGATAGTCGACCCCACTAC
394 R26-3'B61.1-CACCGCCGCTCTTGCCCAGAGAGAAAGG
395 R26-3'B61.2-AAACCCTTTCTCTCTGGGCAAGAGCGGC
396

397 MS2 and PP7 stem loops construction. First pCR4-24xMS2SL (Addgene, Cat.No.31865) and
398 pCR-24xPP7SL (Addgene, Cat.No. 31864) plasmids were double digested using SpeI and NotI
399 restriction enzymes and the 24xMS2SL (1352bp) and 24xPP7SL (1492bp) DNA fragments were
400 gel purified. The annealed adapters were ligated into the excised stem-loop DNA fragments as
401 follows:
402 For Ttc4 gene's C57a allele (Lig#1): anneled Ttc4-B6-5.1 + Ttc4-B6-5.2 (and annealed Ttc4-B6-
403 3.1 + Ttc4-B6-3.2.
404 For the Ttc4 gene CAST allele (Lig#2): annealed Ttc4-CAST-5.1 + Ttc4-CAST -5.2 and Anealed
405 Ttc4-CAST -3.1 + Ttc4-CAST -3.2.
406 For the Cth gene's C57 alle (Lig#3): Anneneld Cth-B6-5.1 + Cth-B6-5.2 and annealed Cth-B6-3.1
407 + Cth-B6-3.2.
408 For the Cth gene's CAST allele (Lig#5): Cth-CAST-5.3 + Cth-CAST-5.4 and annealed Cth-
409 CAST-3.3 + Cth-CAST-3.4.
410 Oligos:
411 Ttc4-B6-5.1 TAATACGACTCACTATAGG
412 Ttc4-B6-5.2
413 GGCCCCTATAGTGAGTCGTATTAAAAGGGAAAACAACACAAGCCTACTTGCAAACC
414 ACTGCCCATTGACTCCTCAAGCTTCTGTCAGATGTGGCACAAGCCCTTGCATGGGCA
415 TTGAGAAGTAGAGAATCCAGTAGGGCTCTGAGTAGTCCCAGGATCCAAGT
416 Ttc4-B6-3.1
417 CTAGTAATACGACTCACTATAGGACTTGGATCCTGGGACTACTCAGAGCCCTACTGG
418 ATTCTCTACTTCTCAATGCCCATGCAAGGGCTTGTGCCACATCTGAC

419 AGAAGCTTGAGGAGTCAATGGGCAGTGGTTTGCAAGTAGGCTTGTGTTGTTTTCCCT

420 TT

421 Ttc4-B6-3.2 CCTATAGTGAGTCGTATTA

422 Ttc4-CAST-5.1 AATTAACCCTCACTAAAGGG

423 Ttc4-CAST-5.2

424 CTAGCCCTTTAGTGAGGGTTAATTACAGGATGTGGCTTCATGGCTCCAGTC

425 Ttc4-CAST-3.1

426 GGCCAATTAACCCTCACTAAAGGGACTTGGATCCTGGGACTACTCAGAGCCCTACTG

427 GATTCTCTACTTCTCAATGCCCATGCAAGGGCTTGTGCCACATCTGACAGAAGCTTG

428 AGGAGTCAATGGGCAGTGGTTTGCAAGTAGGCTTGTGTTGTTTTCCCTTTATCGTGA

429 GGCCTCAT

430 Ttc4-CAST-3.2 CCCTTTAGTGAGGGTTAATT

431 Cth-B6-5.1

432 AGAACTGTCTTCTGTTTATCTTCTAACTAACAGGTTGTTCTGTTAGTATCATTTCGGT

433 AATTTTGCTATATTTGTGTCCAAGGAAGTAAGAGTTGTTCTGTAATACGACTCACTAT

434 AGG

435 Cth-B6-5.2 GGCCCCTATAGTGAGTCGTATTA

436 Cth-B6-3.1 CTATAGTGAGTCGTATTAGATC

437 Cth-B6-3.2

438 TAATACGACTCACTATAGGTTCATTCTCAACAAGACAAAACCCTAGTACAACAGAA

439 AAAAGAAAGAAGTCATCGAATACTATATAAAATTAGTACAAATGTTACAATGTTAA

440 ATCATAAC TACAAAATACTTCAATTTAATAATTTACTTACCAGAATTTAGGTGAC

441 Cth-CAST-5.3 AATTAACCCTCACTAAAGGG

442 Cth-CAST-5.4

443 CTAGCCCTTTAGTGAGGGTTAATTCCTTGGACACAAATATAGCAAAATTACCGAAAT

444 GATACTAACAGAACAACCTGTTAGTTAGAAGATAAACAGAAGACAGTCCTTAA

445 Cth-CAST-3.3

446 GGCCAATTAACCCTCACTAAAGGGTAAAGGTTGTACTTGATACTTGTTATTTTTCTTA

447 AATAAATCCAGTT

448 Cth-CAST-3.4 CCCTTTAGTGAGGGTTAATT

449

450 <u>gRNA cloning for allele-specific insertion of stem-loops.</u> For cloning gRNAs into pSpCas9n(BB)

451 (PX460) plasmids we generated a gRNA targeting library to target maternal and paternal alleles

452 separately. gRNA libraries are as follows:

453 T8: Ttc4-C57BL-gRNA-1.1: + Ttc4-C57BL-gRNA-1.2:

454 T9: Ttc4-C57BL-gRNA-2.1: + Ttc4-C57BL-gRNA-2.2:

455 T10: Ttc4-CAST-gRNA-3.1: + Ttc4-CAST-gRNA-3.2:

456 **T11:** Ttc4-CAST-gRNA-3.1a: + Ttc4-CAST-gRNA-3.2a:

457 T12: Ttc4-CAST-gRNA-4.1a: + Ttc4-CAST-gRNA-4.2a:

458 T13: Ttc4-CAST-gRNA-4.1: + Ttc4-CAST-gRNA-4.2:

459 C1: Cth-C57BL-gRNA-1.1: + Cth-C57BL-gRNA-1.2:

460 C2: Cth-C57BL-gRNA-2.1: + Cth-C57BL-gRNA-2.2:

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

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

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

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

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

466 Ttc4-C57BL-gRNA-1.1: CACCGTTCTGTTGACTGTATAAACTTGG

467 Ttc4-C57BL-gRNA-1.2: CAAGACAACTGACATATTTGAACCCAAA

468 Ttc4-C57BL-gRNA-2.1: CACCGAAATGAGACCTCACAATAAAGGG

469 Ttc4-C57BL-gRNA-2.2: CTTTACTCTGGAGTGTTATTTCCCCAAA

470 Ttc4-CAST-gRNA-3.1: CACCGTGTGGATGCCATGAAGTGACTGG

471 Ttc4-CAST-gRNA-3.2: CACACCTACGGTACTTCACTGACCCAAA

472 Ttc4-CAST-gRNA-3.1a: CACCGTTCTGTTGACTATATAAACTTGG

473 Ttc4-CAST-gRNA-3.2a: CAAGACAACTGATATATTTGAACCCAAA

474 Ttc4-CAST-gRNA-4.1a: CACCGGTTTATATAGTCAACAGAACAGG

475 Ttc4-CAST-gRNA-4.2a: CCAAATATATCAGTTGTCTTGTCCCAAA

476 Ttc4-CAST-gRNA-4.1: CACCGCTACTCCTATTTCAGAAATGAGG

477 Ttc4-CAST-gRNA-4.2: CGATGAGGATAAAGTCTTTACTCCCAAA

478 Cth-C57BL-gRNA-1.1: CACCGAAACAGAAGACAGTTCTTAAGGG

479 Cth-C57BL-gRNA-1.2: AAACCCCTTAAGAACTGTCTTCTGTTTC

480 Cth-C57BL-gRNA-2.1: CACCGTGGTCTTAAATCCACTGTGGTGG

481 Cth-C57BL-gRNA-2.2: AAACCCACCACAGTGGATTTAAGACCAC

482 Cth-CAST-gRNA-3.1: CACCGCTTCCATGGATCTTCCCTTAAGG

483 Cth-CAST-gRNA-3.2: AAACCCTTAAGGGAAGATCCATGGAAGC

484 Cth-CAST-gRNA-3.1a: CACCGCAGAACAACTCTTACTTCCTTGG

485 Cth-CAST-gRNA-3.2a: AAACCCAAGGAAGTAAGAGTTGTTCTGC

486 Cth-CAST-gRNA-4.1a: CACCGAGTAAGAGTTGTTCTGTTTTGGG

487 Cth-CAST-gRNA-4.2a: AAACCCCAAAACAGAACAACTCTTACTC

488 Cth-CAST-gRNA-4.1b: CACCGAAACCACCACAGTGGATTTAAGG

489 Cth-CAST-gRNA-4.2b: AAACCCTTAAATCCACTGTGGTGGTTTC

490 Cth-CAST-gRNA-4.1: CACCGAAGAAAAATAACTTGATAACTGG

491 Cth-CAST-gRNA-4.2: AAACCCAGTTATCAAGTTATTTTTCTTC

492

493 Allele-specific MS2 and PP7 stem loops knock-in to the *Cth* and *Ttc4* genes. An mCherry-PP7CP
494 and EYFP-MS2CP expressing single clone derived mESC line was transfected with a cocktail of
495 DNAs and gRNA-Cas9 plasmids using Xfect transfection reagents to knock-in stem-loops to
496 individual-alleles as follows. $0.5x10^5$ mESCs per well of 6-plates were plated and cultured for 12
497 hrs before the transfection. To target the *Cth* gene alleles, Lig#3 and Lig#5 (3.45 ug of DNA in
498 each) combined with gRNAs C1, C2, C3, C4, C5, C6 and C7 (0.3 ug of plasmid from each) were
499 used. To target the *Ttc4* gene alleles, Lig#1 and Lig#2 (3.6 ug of DNA in each) combined with
500 gRNAs T8, T9, T10, T11, T12 and T13 (0.3 g of plasmid from each) were used. Medium was
501 removed 12 hrs after transfection and refreshed with 2i medium. At 24 hrs post transfection, single
502 colonies were picked under a light microscope and expanded for genotyping using Cth-Wt gene F
503 (AAAGTTCGAGTVAAAGCTGTCAT and Cth-Wt gene R
504 (TCAAGTACAACCTTTAATCTCTTGAA) (for Cth clones) and Ttc4 Wt gene F
505 (GGCAGTGGTTTGCAAGTAGG) and Ttc4 Wt gene R
506 (CCAGCTGGTTTCAGTTTAATAATGT) (for *Ttc4* gene). The mESC clones for both *Cth* and
507 *Ttc4* were further validated for on-targeted integration by a DNA-FISH assay (Extended Data Fig.1
508 b).

509

510

511 **Immunoblotting**

512 1.0 mm thick and 8x10 cm Tris-glycine SDS-polyacrylamide gels cast with 5% Stacking gel (30%
513 acrylamide mix, 1.0 M Tris (pH6.8), 10% SDS, 10% ammonium persulfate, 0.1% TEMED in
514 dH2O) and 12 % resolving gel (30% acrylamide mix, 1.5 M Tris (pH8.8), 10% SDS, 10%
515 ammonium persulfate, 0.1% TEMED in dH2O) were used for protein electrophoresis. Gels were
516 run in 1x running buffer and then the proteins were transferred from the gel *in* ice cold 1x Buffer
517 onto Immobilon®-P Transfer membranes (Merck Millipore, Cat.No. IPVH00005). Transfer
518 membranes were air dried to dehydrate and stored until use. To blot with antibodies, membranes
519 were rehydrated in 0.1% PBST (0.1% Tween20 in 1xPBS) changing the buffer several times until
520 the membrane became wet. Rehydrated membranes were blocked in 5% blocking buffer
521 containing skim milk (Sigma, Cat.No. 1153630500) in 0.1% PBST for 1½ hrs and then incubated
522 with primary antibodies diluted in 5% blocking buffer, overnight. Next day, the membranes were
523 washed in 0.1% PBST for 4 times changing the buffer every 10 mins and then incubated with
524 secondary antibodies conjugated with horseradish peroxidase diluted in secondary antibody buffer
525 (3% milk in 0.1% PBSTX) for 2 hrs at RT on a shaker. After the incubation, membranes were
526 washed 4 times 10 moins each with 0.1% PBST buffer. As a final wash 1x PBS was used. The
527 ECL (Thermo Fisher, Cat. No. PI32209) was added onto the membrane to cover the whole
528 membrane and exposed to X-Ray film (Amersham Hyperfilm™ ECL, GE Healthcare Limited,
529 Cat.No. 28906836). Primary antibodies used: YFP Goat polyclonal antibody (1:500) (ORIGENE,
530 Cat.No. AB1166-100), mouse mCherry antibody (Thermo Fisher, Cat.No. MA5-32977).
531 Secondary antibodies: Donkey anti-Goat IgG, (H+L) Secondary Antibody, HRP (Thermo Fisher,
532 Cat.No. A15999). Donkey anti-Mouse IgG H&L, HRP (1:500) (abcam, Cat.No. AB205719).
533 Histone H3 antibody (1:1000) (Novus Biological, Cat.No. NB500-171-0.025 ml).

534 **Immunofluorescence assay**

535 Cells were cultured on ibidi u-Slide 8 well Glass Bottom chambers (ibidi GmbH, Cat.No. 80827).
536 Cells were washed once with 1x PBS and then fixed in 4% PFA for 12 min while gently rocking
537 and the PFA was washed off by rinsing the cells with 400 ul of 1x PBS, three times. Then the cells
538 were permeabilized with 0.3% PBSTx (Triton X-100 in 1xPBS) for 10 min at RT while gently
539 rocking and washed with 400 ul of 0.1% PBSTx three times. Cells were then blocked with blocking
540 buffer (3% BSA, 10% NDS, 0.1% PBSTx) for 1 hr at RT and 200 ul of blocking buffer containing
541 primary antibodies was added to each well and then incubated overnight at 4ºC on a shaker. Next
542 day, primary antibodies were washed off by rinsing 4 times with 400 ul of 0.1% PBSTx, 10 min
543 for each wash. 200 ul of the secondary antibody diluted in buffer (5% NDS in 0.1% PBSTx) was
544 added to each well and incubated for 1½ hrs at RT on a slow rocker. After the incubation period,
545 secondary antibodies were washed off with 0.1% PBSTx three times and washed once with 1X
546 PBS as the last wash. DAPI diluted at 1:10000 in 1xPBS was (200 ul) added to each well and
547 incubated for another 5 min then washed 3 times with 1X PBS. 100 ul of ibidi mounting medium
548 (ibidi GmbH, Cat.No. 50001) was added to each well and the chamber slides were kept wrapped
549 in aluminum foil at 4ºC until imaged. Primary antibodies used: Sun1 (1:700) a gift from Min lab,
550 Biotin antibody (33) 1: 1000 (Santa Cruz, Cat.No. 101339), LaminB1 (1:500) (Abcam, Cat.No.
551 ab16048). Secondary antibodies used: Donkey anti-Rabbit, AF 568 (Thermo Fisher, Cat.No.
552 A10042), Donkey anti-Mouse, AF 647 (Thermo Fisher, Cat.No. A-21235, Goat anti-Rabbit, AF
553 488 (Thermo Fisher, Cat.No. A-11008). Secondary antibodies were in 1:1000 dilution.

554

555

556

557 **Live imagine using lattice light-sheet microscopy and quantitation**

558 mESCs were cultured on small diameter round cover glass, #1 thickness, 5 mm (Thomas Scientific,
559 Cat.No.1217H19) overnight. Four hrs prior to image acquisition, medium was replaced with
560 medium containing 20 ng/ml doxycycline (Sigma, Cat.No. 324385-1GM) and during image
561 acquisition, the chamber was filled with the same medium.

562 The live images were captured using the 3i lattice light sheet microscope (LLSM) with the
563 following parameters. Objective: 62.5x LLS (magnification: 62.50x), Mag. Changer: 1.0x)
564 magnification: 1.00x), Binning: 1x1, Average timelaps interval: 20 sec, Microns per pixel (XY):
565 0.104 um, step size in microns (Z): 0.553, Channel 1: 488 at 300 ms and Channel 2: 560 at 300
566 ms. The system uses a 25x 1.1 NA detection objective but increased the magnification using optics
567 after the objective to reach 62.5x as the final magnification on the system. Each laser was set to a
568 total power of 9% out of the AOTF with the laser power at the head being 50 mW each. The images
569 were deskewed and deconvolved using the algorithms embedded in SlideBook6 software.

570

571 **Nick translated probe preparation**

572 One microgram of DNA probe (from PCR) was used in each reaction. Reaction was prepared
573 mixing DNA fragments, 3.8 ul of labelled dUTP (from 0.2 mM), 2.5 ul of 0.1 mM dTTP, 10 ul of
574 0.1 mM dNTP mix, 5.0 ul of 10x Nick translation buffer adding nuclease free ddH_2O to volume
575 up to 45 ul. Next, the sample mixture was vortexed and briefly spin-down and 5.0 ul of Nick
576 translation enzyme was added. Reaction mixtures were gently vortexed and centrifuged and
577 proceeded to amplification by incubation at 15ºC for 10 hrs, and then heat inactivated at 70ºC for
578 10 min and cooled down to 4ºC. Samples were briefly spin-down and transferred into 1.5 ml RNase
579 free micro centrifuge tubes. Then 1 ul of 0.5 M EDTA, 1 ul of linear acrylamide (Ambion), 5 ul

580 of 3M NaOAc (pH5.2) and 125 ul of 100% ethanol (from -20ºC cold) was added, briefly mixed
581 by tipping and then incubated at -80ºC for 2 hrs. After the incubation, the samples were centrifuged
582 at 20000 rcf for 60 min at 4 ºC. A pellet should be visible at this point. Next, 1 ml of 75% ethanol
583 was added and centrifuged at 14000 rpm for 5 min and repeated for one time. The pellet was then
584 dried at 37ºC in an incubator for 15 min. Next, 50 ul of DEPC treated DNase and RNase free H_2O
585 was added to each tube containing a sample pellet and placed in a heat block at 37ºC and vortexed
586 for 1 hr 200 rmp speed. Probes were kept at -20ºC until used.
587
588 **Nascent RNA-FISH assay**
589 NPCs were cultured in 24-well plates (MatTek, Cat.No. P24G013F) and used when the cultures
590 were 80% confluent. In all the steps, DEPC treated 1x PBS was used. Freshly made 4% PFA (made
591 in DEPC treated 1xPBS) was chilled on ice and then warmed in a 37ºC water bath for 30 min.
592 Cells were rinsed once with 1xPBS and 200 ul of 4% PFA was added to each well and incubated
593 for 15 min at room temperature with gentle shaking to fix the cells. 200 mM VRC was warmed on
594 60ºC heat block until thawed. Cells were then washed with 1xPBS three times (10 min per wash)
595 with gently shaking. 0.5% Triton X-100/2 mM VCR buffer was made in 1xPBS and 300 ul of the
596 mix was added to each well and incubated on ice for 10 min. Cells were then washed three times
597 with 1xPBS on a shaker and then saturated in 400 ul of 2x SSC for 2 hrs in a 37ºC humidified
598 chamber. While saturating, probes were prepared as follows. 100 ng of probe and 2 ul of
599 competitor yeast tRNA (10ug/ul) were mixed in a microcentrifuge tube and dried down in a Speed
600 Vac (at maximum heat) for 25 min. Once dried ,10 ul of deionized formamide (Ambion) was added
601 and pipetted up-down to dissolve the probes. Tubes were then placed on a 37ºC heat-block for 20
602 min until probes were completely dissolved and then probes were denatured by heating at 95ºC for

603 10 min and chilled on ice for 10 min. Hybridization buffer was prepared combining 4 ul of 50%
604 dextran sulfate (warmed to 42ºC), 3 ul of DEPC treated ddH_2O, 1 ul of RNasin (Promega), and 2
605 ul of 20x SSC per probe mix and 10 ul of the hybridization buffer was added to a tube with probe
606 pipetting five times up-down to mix. 40 ul of probe mixture was added into each well containing
607 cells, sealed with parafilm, wrapped with aluminum foil and incubated for 16 hrs in 37ºC I a
608 humidified incubator. After incubation, cells were washed with pre-warmed (42ºC) 2x SSC/ 50%
609 deionized formamide solution three times (5 min per each wash). Next, the cells were washed with
610 pre-warmed (42ºC) 2x SSC for three times every 5 min, pre-warmed (37ºC) 1x SSC three times (5
611 min per each wash), 0.5% SSC three times (5 min per each wash). Then the cells were incubated
612 with 200 ul of DAPI containing 4x SSC for 5 min and washed with 4x SSC for 5 min. 100 ul of
613 ibidi mounting medium (ibidi GmbH, Cat.No.50001) was added to each well and plates were kept
614 wrapped in aluminum foil at 4ºC until imaged.
615 Primers used for PCR amplification and use as a template:
616 *Acyp2*: F- GGGCCGACTTTTCTAACGA, R- TGAATGACTCCATAAAACAACATT
617 *Arc*: F- GTGAAGACAAGCCAGCATGA, R- CTCCAGGGTCTCCCTAGTCC
618 *Fggy*: F-CCTGGCATTGCAAAATATGA, R- CCTGGATCCAATGAGTGAGC
619 *Bag3*: F-CGAGAGCCTCCACCTGTTAC, R- CCAGACAACCAGAGGGGTTA
620 *Eya1*: F- TGTCAGTGCTGGACTTCAGG, R- CTGCTGCCCCCAGTATTAAA
621 *Dis3l*: F- CGCTGCCACTCTGATACAAT, R- TCTGAAGACAGAGGCAAGAGG
622 *Lypla1*: F- AGTCACCTCCTCCTGCTGAA, R- CCAGACAACCAGAGGGGTTA
623 *Cap2*: F- GCCTTCCCTGAGTCTCCTCT, R- GCGCTATGCTTCGTTACACA
624 *Cth*: F- CCTCTGTGCCTGAGAAGGAC, R- TGAATTTCACAGTAGGTATTCCAAA
625 *Npl*: F- CATCATGACGCTGGTCTCTG, R- CCCAAAAGGCAGGTGAAAT

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

650

651 **DNA-FISH assay**

652 NPCs were cultured and fixed as described in nascent RNA-FISH assay. Cells were rinsed in
653 1xPBS three times (5 min per wash) and permeabilized adding ice-cold 0.5% Triton X-100/1xPBS
654 and keeping on ice for 10 min. After that cells were washed with 1xPBS three times (5 min per
655 wash) and equilibrated in 2x SSC changing the solution twice, each for 5 min and incubated with
656 2x SSC containing 0.1 mg/ml RNase A for 1 hr at 37ºC. The hybridization probe preparation was
657 carried out as follows. 10 ug of mouse Cot-1DNA, 5 ul of salmon sperm DNA, 5 ul of yeast tRNA
658 and 40 ng of nick translated probe were mixed in 1.5 ul microcentrifuge tube and the volume was
659 adjusted up to 20 ul with ddH_2O. The components were mixed gently by tapping and dehydrated
660 (at 65 ºC) in a Speed-Vac for 30 min. 10 ul of deionized formamide was then added to the pellet,
661 resuspended gently and incubated at 37ºC for 60 min while shaking until the probe was completely
662 dissolved. After the RNase treatment, the cells were washed twice (5 min each) with 2x SSC at
663 42ºC. 70% of deionized formamide in 2x SSC (at 80ºC) was added onto the cells (keep the 24-
664 well MatTek plates floating on 80ºC water bath) an incubated for 3 min, liquid was removed from
665 the wells as quickly as possible and ice-cold (-20ºC) 70% ethanol was added onto the cells and
666 incubated for 5 min on ice. Next the cells were dehydrated using 95% ethanol for 5 min and then
667 in 100% ethanol for 5 min and kept in 100% ethanol until proceeding to hybridization steps. Just
668 before the hybridization was started, probe mixture was warmed up on 95ºC on a heat block for
669 10 min and incubated for 2 min on ice. Then 10 ul of pre-warmed 2x hybridization buffer (20%
670 Dextran Sulfate, 4x SSC in dH_2O at 37ºC) was added to 10 ul of probe and 20 ul of the probe
671 mixture was added onto the cells. The lid of the plate was sealed with Parafilm and incubated at

672 37ºC in a humidified chamber overnight. After the probe hybridization, probe mixture was
673 removed from the well and 50% deionized formamide containing 2X SSC (pH 7.0) was added and
674 incubated at 37ºC for 30 min. Then the cells were washed with 2x SSC - 15 min at 37ºC, 1x SSC
675 - 15 min at 37ºC and then equilibrated in 4x SSC for 5 min at RT. 1:10,000 diluted DAPI in 4x
676 SSC was added to each well, incubated for 3 min at RT and then washed in 4x SSC twice (5 min
677 each) at RT. 100 ul of ibidi mounting medium (ibidi GmbH, Cat.No.50001) was added to each
678 well and the plates were kept wrapped in aluminum foil at 4ºC until imaged.
679 Primers used for PCR amplification to use as a template:
680 Cth Nik F: GGTCAAGTGCTTGCAACAGA
681 Cth Nik R: GAGCTTCCGGAGACACAGTC
682 Ttc4 Nik F: CTCCATGGGAGGATGAGAAA
683 Ttc4 Nik R: CCAAACCAAACCAAACCAAC
684
685 **Chromatin Immunoprecipitation by Sun1 and Sequencing**
686 ESCs and NPCs from 80% confluent cultures were harvested by trypsinization and rinsed and
687 resuspended in 10% FBS containing Knockout DMEM (Gibco, Cat.no.10829-018). Crosslinking
688 was carried out incubating $5X10^7$ cells with 1% Formaldehyde in 20 ml of 10% FBS containing
689 Knockout DMEM for 10 min at RT while gently shaking. Crosslinking was blocked by adding 2
690 ml of 1.5M Glycine and further incubation for 10 min at RT with gentle shaking. Then the cells
691 were centrifuged at 2000 rpm for 6 min at 4ºC, medium was removed and washed twice with 20
692 ml of ice-cold 1xPBS. Pellets were then re-suspended in 500 ul of ChIP buffer (0.1% SDS, 10 mM
693 EDTA, 50mM Tris-HCl pH7.5, Protease inhibitor, 1 mM in dH_2O) and kept on ice. Sonication
694 was carried out using a Bioruptor® Pico sonication device (diagenode, Cat.No. B01060010) (30

695 Sec ON and 30 Sec OFF for 15 cycles). After sonication samples were centrifuged at 10,000 g for
696 10 min at 4ºC. Supernatant was transferred to separate 1.5 ml Eppendorf tubes and kept at -80ºC
697 until used. To deplete the nucleic acids to use in Sun1-ChIP nucleic acid depleted extracts, nuclear
698 extracts were treated with Benzonase® Nuclease (Millipore Sigma, Cat. No. E1014-5KU
699 following the manufacturer's guidelines. Sonicated lysates were used to extract the shredded DNA
700 and the average fragment size was 300-500 bps and the concentration of the shredded DNA was
701 measured using a Nanodrop spectrophotometer. 150 ul of Dynabeads™ Protein G was used for
702 Immunoprecipitation (Invitrogen, Cat.No. 10004D) per ChIP reaction. First the beads were washed
703 with ice cold 0.02% Tween20 in 1xPBS and then resuspended in 150 ul of 0.02% Tween20 in
704 1xPBS. 0.8 ug of Sun1 antibody (Proteintech, Cat.No. 24568-I-AP) or Rabbit IgG Isotype Control
705 (Thermo Fisher, Cat.No. 10500c) was added to the beads and the volumes were brought to 200 ul
706 adding 0.02% Tween20 in 1xPBS. The mixtures were incubated for 2 hrs at 4ºC in a tube rotator.
707 After 2 hrs incubation, excess liquid was removed from the bead-antibody reaction mixture and
708 the beads were gently rinsed again with RIPA buffer (50 mM Tris-HCl (pH8.0), 150 mM NaCl, 2
709 mM EDTA (pH8.0), 1% NP40, 0.5% Sodium Deoxicholate, 0.05% SDS, 1 Protease inhibitor
710 tablet (cOmplete Mini, EDTA-free, Cat.No. 39802300), in 10 ml of dH_2O). 150 ug of DNA
711 containing lysates were volumed up to 800 ul with 1x RIPA buffer and added to the bead-antibody
712 mixture and incubated overnight at 4ºC while gently mixing in a tube rotator. Next day the beads
713 and protein complexes were separated from the excess lysate on a magnetic rack. Beads were then
714 washed three times with 1000 ul of 1xTE buffer (10 mM Tris-HCl (pH8.0), 2 mM EDTA in dH_2O)
715 and last, 200 ul of elution buffer (50 mM NaCl, 50 mM Tris-HCl (pH7.5), 0.1 mM PMSF, 5 mM
716 EDTA, 0.1% SDS in ddH_2O) was added and incubated at 31ºC for 15 min in a heat bock while
717 mixing the samples at 800 rpm. For paired-end sequencing, DNAs were extracted from 100 ul of

718 elute, per sample, using Genomic DNA Clean & Concentrator™ -25 Kit (Zymo Research, Cat.No.
719 D4064) and concentrated and paired-end libraries were prepared and sequenced (NextSeq500, PE-
720 150, mid output) at the Next Generation Sequencing Core Facility at Cold Spring Harbor
721 Laboratory. Nucleic acids from the protein samples were removed by incubating the samples with
722 Benzonase (Millipore Sigma, Cat. No.E1014) for 20 min on ice before use in Sun1 interacting
723 pulldown assay.
724

725 **qPCR and Sanger sequencing**

726 1 ug of genomic DNA fragments from Sun1-ChIP assay and from the input sample were used as
727 the templates. Quantitative PCR was performed using PowerUp™ SYBR® Green Master Mix
728 (Thermo Fisher, Cat.No. A25743) and the primers covering SNPs in the *Cth* (F-
729 GAGCCTGGGAGGATATGAGA, R- AAGCTCGATCCAGGTCTTCA) and *Ttc4* (F –
730 GACAGGGCGGAACTATACCA, genes. qPCR products were Sanger sequenced using the
731 GenScript Biotec Sanger sequencing service.
732

733 **Methods for Solution Digestion Analysis by LC-MS/MS**

734 50 ul of elute from a sample was sent to Taplin Biological Mass Spectrometry Facility at the
735 Harvard Medical School. In brief, 10ul (20ng/ul) of modified sequencing-grade trypsin (Promega,
736 Madison, WI) was spiked into 300ul PBS and the samples were placed in a 37ºC room overnight.
737 Samples were acidified by spiking in 20ul 20% formic acid solution and then desalted by STAGE
738 tip (1). On the day of analysis, the samples were reconstituted in 10 µl of HPLC solvent A. A nano-
739 scale reverse-phase HPLC capillary column was created by packing 2.6 µm C18 spherical silica
740 beads into a fused silica capillary (100 µm inner diameter x ~30 cm length) with a flame-drawn

741 tip (2). After equilibrating the column each sample was loaded via a Famos auto sampler (LC
742 Packings, San Francisco, CA) onto the column. A gradient was formed and peptides were eluted
743 with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid).
744 As peptides eluted, they were subjected to electrospray ionization and then entered into an LTQ
745 Orbitrap Velos Elite ion-trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA).
746 Peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific
747 fragment ions for each peptide. Peptide sequences (and hence protein identity) were determined
748 by matching protein databases with the acquired fragmentation pattern by the software program,
749 Sequest (Thermo Fisher Scientific, Waltham, MA) (3). All databases include a reversed version
750 of all the sequences and the data was filtered to between a one and two percent peptide false
751 discovery rate.
752
753 **Allele-specific ChIP-seq analysis and motif analysis**
754 We used the MEA pipeline[4] for ChIP-seq analysis. In silico diploid C57BL/6J / CAST/EiJ genome
755 was reconstructed using the SHAPEIT2 pipeline by combining the mm10 reference genome
756 (mgp.v5.merged.snps_all.dbSNP142.vcf.gz) and genetic variants
757 (mgp.v5.merged.indels.dbSNP142.normed.vcf.gz). We then aligned sequencing read files (.fastq)
758 to the constructed in silico genome using Bowtie2 aligner (version 2.3.4.3). Generated '.bam' files
759 were then used in the ChIP-seq pipeline embedded in SeqMonk bio-informatic web portal, and the
760 peaks were called using the MACS algorithm setting p=0.05 and 500 base fragment size. Motif
761 analysis was performed using HOMER (v4.11) in Basepair (Version 3.2.7) web portal.
762

763 **TA-bridge *in vivo* visualization and pull-down of *in vivo* TA-bridge interacting proteins and**
764 **MS assay**

765 mESCs were seeded at $0.4x10^4$ cells per well in 8-well ibidi chamber slides and incubated for 24
766 hrs before transfection of biotinylated DNAs. For IF assay 0.03 ug of biotinylated oligos per well
767 were used.

768 For *in vivo* TA-bridge interacting proteins pull-down assay, $2.5x10^5$ cells per well in three 6-well
769 plates were seeded with ESCs 24 hrs prior to transfection. 0.5 ug of biotinylated oligos were used
770 per well for transfection. Using Xfect™ transfection reagent kit (Clonetech, Cat.No. 631317),
771 oligos were transfected according to the manufacture's guidelines. Cells and oligos were incubated
772 overnight and then fixed with 4% PFA and kept at 4ºC for IF.

773 The cells set-up for interacting protein pull-down, were rinsed three times with 1xPBS and the
774 nuclear proteins were extracted using a nuclear extraction kit (ActiveMotif®, Cat.No. 40010). Per
775 pull-down reaction 200 ul of Dynabeads MyOne Streptavidin- T1 and 200 ug of nuclear extract
776 was used. Before using the beads, they were washed three times with 500 ul of 2x binding buffer
777 and then resuspended with 150 ul of 2X B&W buffer. The nuclear extracts' volume was then
778 brought up to 150 ul by adding Complete Lysis Buffer from the nuclear extract kit and added to
779 the beads. Binding was performed at room temperature for 1 hr on a tube rotator. At the end, tubes
780 were briefly centrifuged and supernatant was removed on the magnetic rack. Then the beads were
781 washed four times with 1 ml of 1x B&W buffer and then resuspended in 20 ul of elution buffer
782 (50mM NaCl, 50mM Tris-HCl (pH7.5), 0.1 mM PMFS, 5 mM EDTA, 0.1% SDS in Milli-Q water)
783 and boiled at 95ºC for 7 min while slowly vertexing in a heat rack. Tubes were briefly centrifuged
784 and the supernatant was collected on a magnetic rack. Elutes were kept at -80ºC until used. For
785 MS analysis 60 ul (20 ul combined from 3 prep) was used (see below).

786 **Identification of proteins bound to TA-bridges by MS**

787 MS was performed at the Cold Spring Harbor Mass Spectrometry Core Facility. Procedure for on-
788 membrane proteolysis was performed as followed.

789 Excised nylon membranes were hydrated and washed three times with 500 ul of 100 mM
790 ammonium bicarbonate (ABC) for 10 min at 55ºC under agitation. Then 500 ul of fresh 3 mM
791 TCEP/50mM ABC was added and incubated at 55ºC for 20 min while vortexing at 500 rpm.
792 Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) solution was removed and 500 ul of fresh
793 10 mM CEMTS (2-carboxymethyl methanethiosulfonate)/10% ethanol was added and then the
794 tubes were incubated at 55ºC for 5 min while vortexing at 500 rpm. CEMTS solution was removed
795 and the membranes were washed three times with 500 ul of 50 mM ABC for 5 min at 55ºC under
796 agitation. 100 ng/ul of Trypsin (Sequencing grade modified porcine trypsin, Promega) was
797 prepared in 50 mM ABC and 5 ul of it added into the middle of the membrane piece, followed by
798 200 ul of 50 mM ABC/0.1% ProteaseMax surfactant. Membranes were digested at 37ºC overnight
799 while vortexing at 500 rpm. 200 ul of 80% ACN/1% TFA mix was added to each membrane and
800 incubated for another 15 min at 55ºC under agitation. Elutes were collected into clean tubes,
801 lyophilized by vacuum centrifugation and the peptide pellets were resuspended in 20 ul of loading
802 buffer (55% DMSO, 0.1% TFA in water).

803 Peptides were loaded on a 30 cm x 75 uM ID column packed with Reprosil 1.9 C18 silica particles
804 and resolved on a 5-35% acetonitrile gradient in water (0.1% formate) at 250 nl/min. Eluting
805 peptides were ionized by electrospray (2200V) and transferred into an Exploris 480 mass
806 spectrometer (Thermo). The MS was set to collect 120K resolution precursor scans (m/z 380-2000
807 Th) and 60K HCD fragmentation spectra at stepped 28,33,38% normalized energy, with first mass
808 locked to 100 Th. Files were searched using the Mascot scoring function within

809 ProteomeDiscoverer, with mass tolerance set at 5 ppm for MS1and 0.01 for MS2. Spectra were
810 matched against the Uniprot mouse consensus database plus a database of common contaminants
811 (cRAP). M-oxidation and N/Q-deamidation were set as variable modifications. Peptide-spectral
812 matches were filtered to maintain FDR < 1%.
813
814 **EMSA assay**
815 EMSA assay was performed by using LightShift® Chemiluminescent EMSA Kit (Thermo
816 Scientific, Cat.No. 20148). Biotinylated oligos were synthesized by bioSYNTHESIS®.
817 TAAATAAATAAATAAATAAATAAATAAATAAA(BM-1), ATTTATTTATTTATTTATTT
818 ATTTATTTATTT(BM-2), and GCAT(Cont) oligos (5' to 3') were tagged with DesthioBiotin
819 TEG at the 3' end of the oligos placing six -Sp18- spacers between oligo and DesthioBiotin.
820 Double stranded TA-bridges were synthesized annealing BM-1 with non-biotinylated oligos and
821 DNA was purified using Genomic DNA Clean & Concentrator™ (ZYMO RESEARCH, Cat.No.
822 D4065). 30 fmol from each biotinylated oligos were used per reaction and biotinylated Epstein-
823 Barr nuclear antigen (EBNA) DNA was used as a negative control. 6% natural polyacrylamide
824 gels (1.0 mm 8X10 cm) were used for gel shift assay. Nuclear extracts from mESCs were extracted
825 using nuclear extract kit (ActiveMotif®, Cat.No. 40010) from 70% confluent ESC cultures.
826 Binding reactions were set-up mixing 10x Binding buffer, 1 ul Poly (dI.dC), Nuclear extract
827 protein and biotinylated DNA oligos to achieve 20 ul volume of total binding reaction and
828 incubated for 20 min at room temperature. Then, 5 ul of 5X loading buffer was added to each 20
829 ul binding reaction and mixed pipetting up-down 5 times. Before loading the reaction mixture onto
830 the 6% natural polyacrylamide gels, gels were pre-run for 60 min at 100V. The protein gels were
831 run at 100V until the bromophenol blue dye migrated three quarter of the gel.

832 To transfer protein from gel to nylon membranes, nylon membranes (Biodyne® A Pre-Cut
833 Membranes – Thermo Scientific, Cat.No. 77015) were soaked in 0.5X TBE for 10 min and the
834 gels, nylon membrane and blotting paper were sandwiched according to the manufacturer's
835 instructions. Transfer was conducted in ice-cold 0.5X TBE while the transfer apparatus was kept
836 sub-merged in ice all the time. Transfer was carried out for 45 min at 380 mA and the nylon
837 membrane was semi-dried on a paper towel while bromophenol blue staining side was up-sided.
838 Then the cross-linking was carried out in a UV StrataLinker™ 1800 (STRATAGENE) at 120000
839 uJ power for 60 Sec and then the membrane was directly transferred to blocking buffer pre-warmed
840 to 37ºC for 15 min while gently shaking. Then the blocking buffer was removed and membrane
841 soaked in conjugated/bocking buffer was incubated for 15 min on a slow rocker. The membrane
842 was then transferred to a new container, rinsed once with 20 ml 1x wash buffer and then washed
843 for another four times (5 min for each wash) in 20 ml 1X wash buffer while gently shaking. The
844 membrane was then transferred to a clean container and incubated in Substrate Equilibration buffer
845 for 5 min, with gentle shaking. Membrane was then removed into a clean container after blotting
846 off the excess Equilibration buffer and 12 ml of Substrate working solution was added to the
847 membrane and incubated for another 5 min without shaking. The membrane was then taken out
848 from the solution and blotted to remove the excess solution without letting the membrane over-
849 dry. The membrane was wrapped in plastic film and exposed to X-ray film (Amersham
850 Hyperfilm™ ECL, GE Healthcare Limited, Cat.No. 28906836) for 7 min.
851
852
853
854

855 **Data availability**

856 The Sun1 ChIP-seq data (raw and processed) supporting this study's finding have been deposited
857 into the Gene Expression Omnibus (GEO) under accession code GSE228344. All other data
858 supporting the finding of this study has been submitted with the manuscript.

859

860

861 **References**

862 1 Bickmore, W. A. The spatial organization of the human genome. *Annu Rev Genomics Hum*
863 *Genet* **14**, 67-84, doi:10.1146/annurev-genom-091212-153515 (2013).

864 2 Binder, H. *et al.* Transcriptional regulation by histone modifications: towards a theory of
865 chromatin re-organization during stem cell differentiation. *Phys Biol* **10**, 026006,
866 doi:10.1088/1478-3975/10/2/026006 (2013).

867 3 Cremer, T. & Cremer, M. Chromosome Territories. *Csh Perspect Biol* **2**, doi:ARTN
868 a003889 10.1101/cshperspect.a003889 (2010).

869 4 Fraser, P. & Bickmore, W. Nuclear organization of the genome and the potential for gene
870 regulation. *Nature* **447**, 413-417, doi:10.1038/nature05916 (2007).

871 5 Hubner, M. R., Eckersley-Maslin, M. A. & Spector, D. L. Chromatin organization and
872 transcriptional regulation. *Curr Opin Genet Dev* **23**, 89-95, doi:10.1016/j.gde.2012.11.006
873 (2013).

874 6 Meaburn, K. J. & Misteli, T. Cell biology: chromosome territories. *Nature* **445**, 379-781,
875 doi:10.1038/445379a (2007).

876 7 Takizawa, T., Meaburn, K. J. & Misteli, T. The meaning of gene positioning. *Cell* **135**, 9-
877 13, doi:10.1016/j.cell.2008.09.026 (2008).

878 8 Abed, J. A. *et al.* Highly structured homolog pairing reflects functional organization of the
879 Drosophila genome. *Nat Commun* **10**, doi:ARTN 4485 10.1038/s41467-019-12208-3
880 (2019).

881 9 Duncan, I. W. Transvection effects in Drosophila. *Annu Rev Genet* **36**, 521-556,
882 doi:10.1146/annurev.genet.36.060402.100441 (2002).

883 10 Galouzis, C. C. & Prud'homme, B. Transvection regulates the sex-biased expression of a
884 fly X-linked gene. *Science* **371**, 396-+, doi:10.1126/science.abc2745 (2021).
885 11 Hogan, M. S., Parfitt, D. E., Zepeda-Mendoza, C. J., Shen, M. M. & Spector, D. L.
886 Transient Pairing of Homologous Oct4 Alleles Accompanies the Onset of Embryonic Stem
887 Cell Differentiation. *Cell Stem Cell* **16**, 275-288, doi:10.1016/j.stem.2015.02.001 (2015).
888 12 Jack, J. W. & Judd, B. H. Allelic pairing and gene regulation: A model for the zeste-white
889 interaction in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **76**, 1368-1372,
890 doi:10.1073/pnas.76.3.1368 (1979).
891 13 Viets, K. *et al.* Characterization of Button Loci that Promote Homologous Chromosome
892 Pairing and Cell-Type-Specific Interchromosomal Gene Regulation. *Dev Cell* **51**, 341-+,
893 doi:10.1016/j.devcel.2019.09.007 (2019).
894 14 D'Angelo, M. A. Nuclear pore complexes as hubs for gene regulation. *Nucleus-Phila* **9**,
895 142-148, doi:10.1080/19491034.2017.1395542 (2018).
896 15 Mao, Y. T. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies.
897 *Trends Genet* **27**, 295-306, doi:10.1016/j.tig.2011.05.006 (2011).
898 16 Meldi, L. & Brickner, J. H. Compartmentalization of the nucleus. *Trends Cell Biol* **21**, 701-
899 708, doi:10.1016/j.tcb.2011.08.001 (2011).
900 17 Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome
901 Organization in the Nucleus. *Cell* **174**, 744-+, doi:10.1016/j.cell.2018.05.024 (2018).
902 18 Zimber, A., Nguyen, Q. D. & Gespach, C. Nuclear bodies and compartments: Functional
903 roles and cellular signalling in health and disease. *Cell Signal* **16**, 1085-1104,
904 doi:10.1016/j.cellsig.2004.03.020 (2004).

905 19 Eckersiey-Maslin, M. A. & Spector, D. L. Random monoallelic expression: regulating gene
906 expression one allele at a time. *Trends Genet* **30**, 237-244, doi:10.1016/j.tig.2014.03.003
907 (2014).
908 20 Ferguson-Smith, A. C. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat*
909 *Rev Genet* **12**, 565-575, doi:10.1038/nrg3032 (2011).
910 21 Reik, W. & Walter, J. Imprinting mechanisms in mammals. *Curr Opin Genet Dev* **8**, 154-
911 164, doi:10.1016/s0959-437x(98)80136-6 (1998).
912 22 Balasooriya, G. I. & Spector, D. L. Allele-specific differential regulation of monoallelically
913 expressed autosomal genes in the cardiac lineage. *Nat Commun* **13**, 5984,
914 doi:10.1038/s41467-022-33722-x (2022).
915 23 Chen, M. *et al.* Live imaging of RNA and RNA splicing in mammalian cells via the
916 dcas13a-SunTag-BiFC system. *Biosens Bioelectron* **204**, 114074,
917 doi:10.1016/j.bios.2022.114074 (2022).
918 24 Daigle, N. & Ellenberg, J. LambdaN-GFP: an RNA reporter system for live-cell imaging.
919 *Nat Methods* **4**, 633-636, doi:10.1038/nmeth1065 (2007).
920 25 Haim, L., Zipor, G., Aronov, S. & Gerst, J. E. A genomic integration method to visualize
921 localization of endogenous mRNAs in living yeast. *Nat Methods* **4**, 409-412,
922 doi:10.1038/nmeth1040 (2007).
923 26 Hocine, S., Raymond, P., Zenklusen, D., Chao, J. A. & Singer, R. H. Single-molecule
924 analysis of gene expression using two-color RNA labeling in live yeast. *Nat Methods* **10**,
925 119-121, doi:10.1038/nmeth.2305 (2013).
926 27 Janicki, S. M. *et al.* From silencing to gene expression: real-time analysis in single cells.
927 *Cell* **116**, 683-698, doi:10.1016/s0092-8674(04)00171-0 (2004).

928 28 Ma, H. H. *et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs
929 using CRISPRainbow. *Nature Biotechnology* **34**, 528-530, doi:10.1038/nbt.3526 (2016).
930 29 Maass, P. G. *et al.* Spatiotemporal allele organization by allele-specific CRISPR live-cell
931 imaging (SNP-CLING). *Nat Struct Mol Biol* **25**, 176-+, doi:10.1038/s41594-017-0015-3
932 (2018).
933 30 Tsukamoto, T. *et al.* Visualization of gene activity in living cells. *Nat Cell Biol* **2**, 871-878,
934 doi:10.1038/35046510 (2000).
935 31 Vera, M., Biswas, J., Senecal, A., Singer, R. H. & Park, H. Y. Single-Cell and Single-
936 Molecule Analysis of Gene Expression Regulation. *Annual Review of Genetics, Vol 50* **50**,
937 267-291, doi:10.1146/annurev-genet-120215-034854 (2016).
938 32 Chen, H. M. & Larson, D. R. What have single-molecule studies taught us about gene
939 expression? *Gene Dev* **30**, 1796-1810, doi:10.1101/gad.281725.116 (2016).
940 33 Shao, S. P., Xue, B. X. & Sun, Y. J. Intranucleus Single-Molecule Imaging in Living Cells.
941 *Biophys J* **115**, 181-189, doi:10.1016/j.bpj.2018.05.017 (2018).
942 34 Shav-Tal, Y., Singer, R. H. & Darzacq, X. Imaging gene expression in single living cells.
943 *Nat Rev Mol Cell Bio* **5**, 855-862, doi:10.1038/nrm1494 (2004).
944 35 Blake, W. J., Kaern, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression.
945 *Nature* **422**, 633-637, doi:10.1038/nature01546 (2003).
946 36 Larson, D. R., Singer, R. H. & Zenklusen, D. A single molecule view of gene expression.
947 *Trends Cell Biol* **19**, 630-637, doi:10.1016/j.tcb.2009.08.008 (2009).
948 37 Wang, L., Frei, M. S., Salim, A. & Johnsson, K. Small-Molecule Fluorescent Probes for
949 Live-Cell Super-Resolution Microscopy. *J Am Chem Soc* **141**, 2770-2781,
950 doi:10.1021/jacs.8b11134 (2019).

951 38 Joyce, E. F., Erceg, J. & Wu, C. T. Pairing and anti-pairing: a balancing act in the diploid
952 genome. *Curr Opin Genet Dev* **37**, 119-128, doi:10.1016/j.gde.2016.03.002 (2016).
953 39 Blattner, G., Cavazza, A., Thrasher, A. J. & Turchiano, G. Gene Editing and Genotoxicity:
954 Targeting the Off-Targets. *Front Genome Ed* **2**, 613252, doi:10.3389/fgeed.2020.613252
955 (2020).
956 40 Bohne, J. & Cathomen, T. Genotoxicity in gene therapy: An account of vector integration
957 and designer nucleases. *Curr Opin Mol Ther* **10**, 214-223 (2008).
958 41 Yoon, K. H. *et al.* Olfactory receptor genes expressed in distinct lineages are sequestered
959 in different nuclear compartments. *P Natl Acad Sci USA* **112**, E2403-E2409,
960 doi:10.1073/pnas.1506058112 (2015).
961 42 Armelin-Correa, L. M., Gutiyama, L. M., Brandt, D. Y. C. & Malnic, B. Nuclear
962 compartmentalization of odorant receptor genes. *P Natl Acad Sci USA* **111**, 2782-2787,
963 doi:10.1073/pnas.1317036111 (2014).
964 43 Clowney, E. J. *et al.* Nuclear Aggregation of Olfactory Receptor Genes Governs Their
965 Monogenic Expression. *Cell* **151**, 724-737, doi:10.1016/j.cell.2012.09.043 (2012).
966 44 Gundersen, G. G. & Worman, H. J. Nuclear Positioning. *Cell* **152**, 1376-1389,
967 doi:10.1016/j.cell.2013.02.031 (2013).
968 45 Haque, F. *et al.* SUN1 interacts with nuclear lamin A and cytoplasmic nesprins to provide
969 a physical connection between the nuclear lamina and the cytoskeleton. *Mol Cell Biol* **26**,
970 3738-3751, doi:10.1128/Mcb.26.10.3738-3751.2006 (2006).
971 46 Liu, Q. *et al.* Functional association of Sun1 with nuclear pore complexes. *J Cell Biol* **178**,
972 785-798, doi:10.1083/jcb.200704108 (2007).

973 47 Li, P. *et al.* The function of the inner nuclear envelope protein SUN1 in mRNA export is
974 regulated by phosphorylation. *Sci Rep-Uk* **7**, doi:ARTN 9157 10.1038/s41598-017-08837-
975 7 (2017).
976 48 Summer, M. C. & Brickner, J. The Nuclear Pore Complex as a Transcription Regulator.
977 *Csh Perspect Biol* **14**, doi:ARTN a039438 10.1101/cshperspect.a039438 (2022).
978 49 Matsumoto, A. *et al.* Loss of the integral nuclear envelope protein SUN1 induces alteration
979 of nucleoli. *Nucleus-Phila* **7**, 68-83, doi:10.1080/19491034.2016.1149664 (2016).
980 50 van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome
981 Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780-791,
982 doi:10.1016/j.cell.2017.04.022 (2017).
983 51 Gonzalez-Sandoval, A. & Gasser, S. M. On TADs and LADs: Spatial Control Over Gene
984 Expression. *Trends Genet* **32**, 485-495, doi:10.1016/j.tig.2016.05.004 (2016).
985 52 Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of
986 nuclear lamina interactions. *Nature* **453**, 948-951, doi:10.1038/nature06947 (2008).
987 53 Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina
988 interactions during differentiation. *Mol Cell* **38**, 603-613,
989 doi:10.1016/j.molcel.2010.03.016 (2010).
990 54 Leemans, C. *et al.* Promoter-Intrinsic and Local Chromatin Features Determine Gene
991 Repression in LADs. *Cell* **177**, 852-864 e814, doi:10.1016/j.cell.2019.03.009 (2019).
992 55 Luo, L. *et al.* The nuclear periphery of embryonic stem cells is a transcriptionally
993 permissive and repressive compartment. *J Cell Sci* **122**, 3729-3737,
994 doi:10.1242/jcs.052555 (2009).

995 56 Wu, F. & Yao, J. Identifying Novel Transcriptional and Epigenetic Features of Nuclear
996 Lamina-associated Genes. *Sci Rep* **7**, 100, doi:10.1038/s41598-017-00176-x (2017).
997 57 Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*
998 **171**, 557-+, doi:10.1016/j.cell.2017.09.043 (2017).
999 58 Monahan, K., Horta, A. & Lomvardas, S. LHX2-and LDB1-mediated trans interactions
1000 regulate olfactory receptor choice. *Nature* **565**, 448-+, doi:10.1038/s41586-018-0845-0
1001 (2019).
1002 59 Dekker, J. *et al.* The 4D nucleome project. *Nature* **549**, 219-226, doi:10.1038/nature23884
1003 (2017).
1004 60 Reiff, S. B. *et al.* The 4D Nucleome Data Portal as a resource for searching and visualizing
1005 curated nucleomics data. *Nat Commun* **13**, doi:ARTN 2365 10.1038/s41467-022-29697-4
1006 (2022).
1007 61 Shevelyov, Y. Y. & Ulianov, S. V. The Nuclear Lamina as an Organizer of Chromosome
1008 Architecture. *Cells* **8**, doi:ARTN 136 10.3390/cells8020136 (2019).
1009 62 Czapiewski, R., Robson, M. I. & Schirmer, E. C. Anchoring a Leviathan: How the Nuclear
1010 Membrane Tethers the Genome. *Front Genet* **7**, doi:ARTN 82 10.3389/fgene.2016.00082
1011 (2016).
1012 63 Snyder, M. J. *et al.* Anchoring of Heterochromatin to the Nuclear Lamina Reinforces
1013 Dosage Compensation-Mediated Gene Repression. *Plos Genet* **12**, doi:ARTN e1006341
1014 10.1371/journal.pgen.1006341 (2016).
1015 64 Rivera-Mulia, J. C. *et al.* Allele-specific control of replication timing and genome
1016 organization during development. *Genome Research* **28**, 800-811,
1017 doi:10.1101/gr.232561.117 (2018).

1018 65 Arao, Y. *et al.* A+U-rich-element RNA-binding factor 1/heterogeneous nuclear
1019 ribonucleoprotein D gene expression is regulated by oestrogen in the rat uterus. *Biochem J*
1020 **361**, 125-132, doi:Doi 10.1042/0264-6021:3610125 (2002).
1021 66 Wu, F. N. & Yao, J. Identifying Novel Transcriptional and Epigenetic Features of Nuclear
1022 Lamina-associated Genes. *Sci Rep-Uk* **7**, doi:ARTN 100 10.1038/s41598-017-00176-x
1023 (2017).
1024 67 Leemans, C. *et al.* Promoter-Intrinsic and Local Chromatin Features Determine Gene
1025 Repression in LADs. *Cell* **177**, 852-+, doi:10.1016/j.cell.2019.03.009 (2019).
1026 68 Wu, C. H. *et al.* NELF and DSIF cause promoter proximal pausing on the hsp70 promoter
1027 in Drosophila. *Gene Dev* **17**, 1402-1414, doi:10.1101/gad.1091403 (2003).
1028 69 Gressel, S. *et al.* CDK9-dependent RNA polymerase II pausing controls transcription
1029 initiation. *Elife* **6**, doi:ARTN e29736 10.7554/eLife.29736 (2017).
1030 70 Nazer, E., Dale, R. K., Chinen, M., Radmanesh, B. & Lei, E. P. Argonaute2 and LaminB
1031 modulate gene expression by controlling chromatin topology. *Plos Genet* **14**, doi:ARTN
1032 e1007276 10.1371/journal.pgen.1007276 (2018).
1033 71 Hah, J. & Kim, D. H. Deciphering Nuclear Mechanobiology in Laminopathy. *Cells* **8**,
1034 doi:10.3390/cells8030231 (2019).
1035 72 Chen, Y. *et al.* The SUN1-SPDYA interaction plays an essential role in meiosis prophase
1036 I. *Nat Commun* **12**, 3176, doi:10.1038/s41467-021-23550-w (2021).
1037 73 Hoyt, S. J. *et al.* From telomere to telomere: The transcriptional and epigenetic state of
1038 human repeat elements. *Science* **376**, eabk3112, doi:10.1126/science.abk3112 (2022).

1039 74 Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology,
1040 bioinformatics and applications. *Nat Biotechnol* **39**, 1348-1365, doi:10.1038/s41587-021-
1041 01108-x (2021).
1042 75 Bzymek, M. & Lovett, S. T. Instability of repetitive DNA sequences: the role of replication
1043 in multiple mechanisms. *Proc Natl Acad Sci U S A* **98**, 8319-8325,
1044 doi:10.1073/pnas.111008398 (2001).
1045 76 Altemose, N. *et al.* DiMeLo-seq: a long-read, single-molecule method for mapping protein-
1046 DNA interactions genome wide. *Nat Methods* **19**, 711-723, doi:10.1038/s41592-022-
1047 01475-6 (2022).
1048 77 Nguyen, G. H. *et al.* Regulation of gene expression by the BLM helicase correlates with
1049 the presence of G-quadruplex DNA motifs. *Proc Natl Acad Sci U S A* **111**, 9905-9910,
1050 doi:10.1073/pnas.1404807111 (2014).
1051 78 Mao, S. Q. *et al.* DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol*
1052 *Biol* **25**, 951-957, doi:10.1038/s41594-018-0131-8 (2018).
1053 79 Tye, B. K. MCM proteins in DNA replication. *Annu Rev Biochem* **68**, 649-686,
1054 doi:10.1146/annurev.biochem.68.1.649 (1999).
1055 80 Roca, J. Topoisomerase II: a fitted mechanism for the chromatin landscape. *Nucleic Acids*
1056 *Res* **37**, 721-730, doi:10.1093/nar/gkn994 (2009).
1057 81 Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation.
1058 *Nature* **518**, 331-336, doi:10.1038/nature14222 (2015).
1059 82 Winick-Ng, W. *et al.* Cell-type specialization is encoded by specific chromatin topologies.
1060 *Nature* **599**, 684-+, doi:10.1038/s41586-021-04081-2 (2021).

1061 83 Chess, A., Simon, I., Cedar, H. & Axel, R. Allelic inactivation regulates olfactory receptor
1062 gene expression. *Cell* **78**, 823-834, doi:10.1016/s0092-8674(94)90562-2 (1994).
1063 84 Singh, N. *et al.* Coordination of the random asynchronous replication of autosomal loci.
1064 *Nat Genet* **33**, 339-341, doi:10.1038/ng1102 (2003).
1065 85 Lionnet, T. *et al.* A transgenic mouse for in vivo detection of endogenous labeled mRNA.
1066 *Nat Methods* **8**, 165-170, doi:10.1038/nmeth.1551 (2011).
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083

1084 **Acknowledgements**

1085 We would like to thank members of the Spector lab for critical discussions and advice
1086 throughout the course of this study. We thank Prof. Min Han for his generous gift of Sun1 antibody.
1087 We would also like to thank Osama El Demerdash for assistance with setting up a computational
1088 pipeline, Tse-Luen (Erika) Wee and Nour El-amine for assisting with microscopy, and Andrew
1089 Barazia (Intelligent Imaging Innovation- 3i) for allowing us to use the 3i lattice lightsheet
1090 microscope in one of our experiments. We would also like to thank Paolo Cifani (CSHL) and Ross
1091 Tomaino (Taplin Mass Spectrometry Facility, Harvard Medical School) for assisting with Mass
1092 Spectrometry experiments, and Sara Goodwin (CSHL) for assisting with DNA sequencing. We
1093 acknowledge the CSHL Cancer Center Shared Resources (Animal, Microscopy, Mass
1094 Spectrometry, Next-Gen Sequencing) for services and technical expertise (NCI 2P3OCA45508).
1095 Sequencing analysis was performed using equipment purchased through NIH grant
1096 S10OD028632-01. This research was supported by R35GM131833 (D.L.S.) and grant 18-26 from
1097 the Charles H. Revson Foundation (G.I.B.).

1098

1099 **Author Contribution**

1100 G.I.B. and D.L.S. conceived and designed the overall study. G.I.B. designed and performed all the
1101 experiments and analysis. G.I.B. and D.L.S. wrote and edited the manuscript.

1102

1103 **Competing Interest**

1104 The authors declare no competing interests.

1105

1106 **Corresponding author**

1107 David L. Spector (spector@cshl.edu)