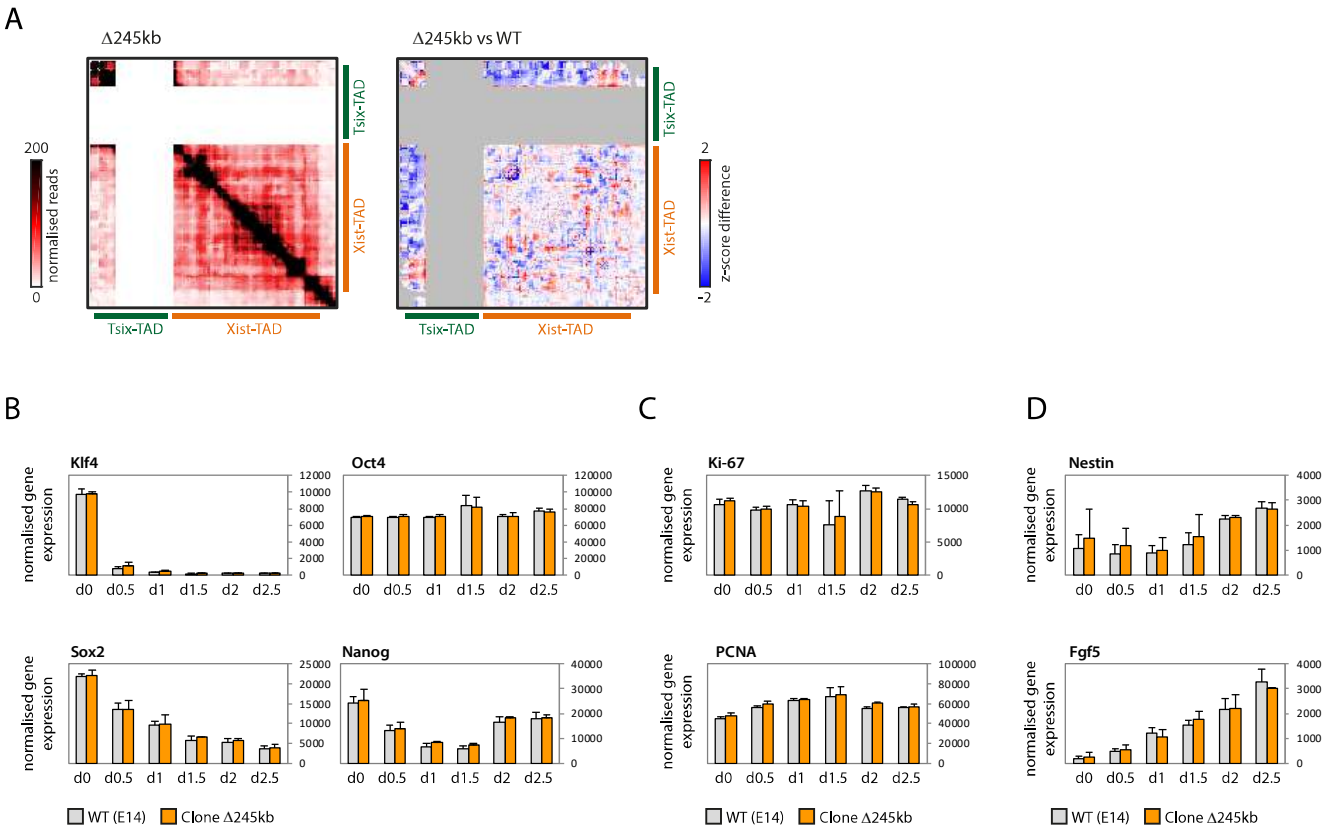
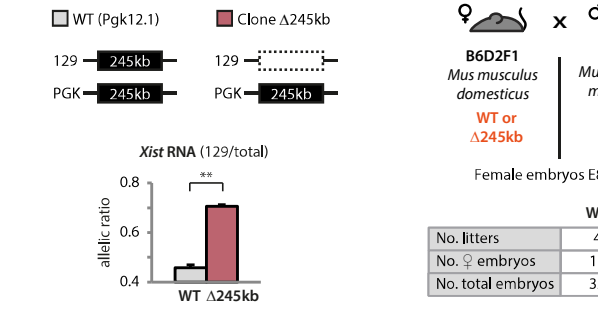
Molecular Cell, Volume *77*

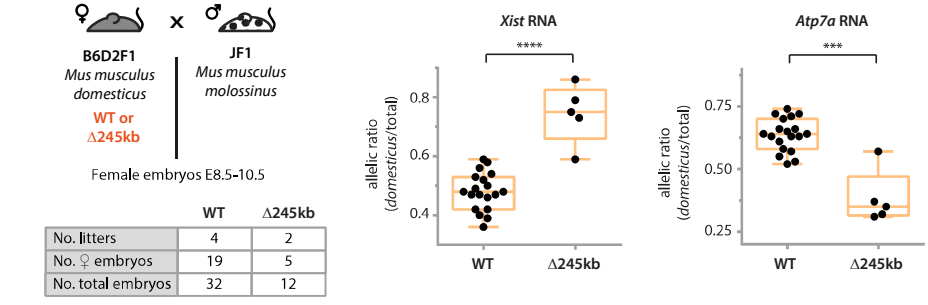
Supplemental Information

A Conserved Noncoding Locus Regulates Random Monoallelic *Xist* Expression across a Topological Boundary

Rafael Galupa, Elphège Pierre Nora, Rebecca Worsley-Hunt, Christel Picard, Chris Gard, Joke Gerarda van Bemmel, Nicolas Servant, Yinxiu Zhan, Fatima El Marjou, Colin Johanneau, Patricia Diabangouaya, Agnès Le Saux, Sonia Lameiras, Juliana Pipoli da Fonseca, Friedemann Loos, Joost Gribnau, Sylvain Baulande, Uwe Ohler, Luca Giorgetti, and Edith Heard



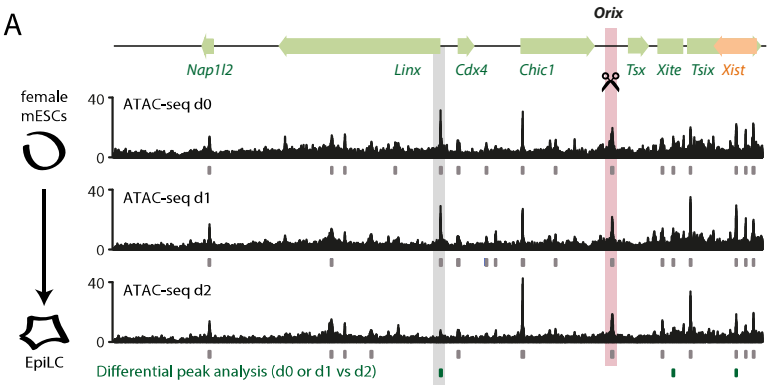


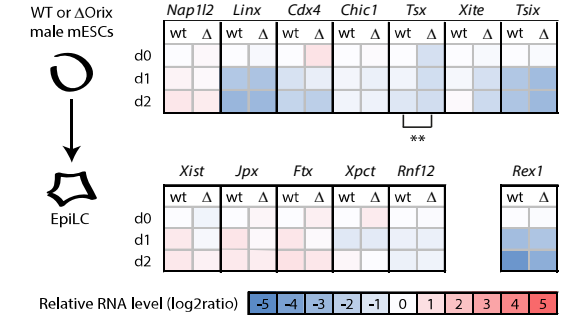


	WT	Δ245kb
No. litters	4	2
No. ♀ embryos	19	5
No. total embryos	32	12

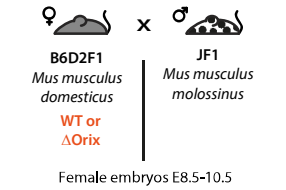
Figure S1. Characterisation of Δ245kb mutants, Related to Figure 1

(**A**) 5C profiles of Δ254kb male mESCs; pooled data from two biological replicates. Differential maps represent the subtraction of Z-scores calculated for wildtype data from Z-scores calculated for mutant data (see Methods). Grey pixels correspond to either the deleted region or to contacts that were filtered because they did not meet the quality control threshold (see Methods). (**B**, **C**, **D**) Gene expression analysis using nCounter (see Methods) of wild type (grey) and Δ245kb (orange) male mESCs during differentiation. Data is normalised to six reference genes (see Methods), and represents the average of RNA counts from two biological replicates for each genotype. (**E**) Analysis of Xist RNA allelic ratios in wildtype and heterozygous Δ245kb female mESCs at day 4 of differentiation. In the mutant female mESCs, the 129 allele harbours the deletion. Average of three replicates is shown with error bars representing SEM. Statistical analysis: paired two-tailed t-test (** $p<0.01$). (**F**) Reciprocal cross of analysis shown in Fig. 1E-G. On the left, schematic illustration of the crosses used for analysis of RNA allelic ratios in wildtype and heterozygous E8.5-E10.5 female hybrid embryos (*molossinus*/*domesticus*). Table summarises number of embryos collected. On the right, analysis of allelic ratios for Xist and Atp7a RNA. Each black dot represents the ratio for a single female embryo. Statistical analysis: two-tailed t-test (*** $p<0.001$, **** $p<0.0001$). **Note 1:** Given that Δ245kb heterozygous female ESCs also showed skewed *Xist* expression during early differentiation (**Fig. S1E**), our results indicate that the Δ245kb allele affects primary XCI choice (*Xist* upregulation). We cannot rule out that the effects we see *in vivo* are further intensified by secondary choice mechanisms (such as counter-selection of cells inactivating the wildtype allele); we note, however, that the Δ245kb allele is not deleterious for cell viability, as male mice with a single Δ245kb X-chromosome are viable. **Note 2:** Male and female Δ245kb mutants (hemizygous or homozygous) are viable, survive to adulthood and generate live descendants, despite lacking several coding and noncoding loci (*Tsx*, *Chic1, Cdx4, Linx, Ppnx, Nap1L2*). However, homozygous crosses are subfertile. Given that either paternal or maternal transmission of the Δ245kb allele result in viable male and female pups, imprinted XCI seems to be unaffected. This indicates that the Δ245kb region is not involved in regulating *Xist* expression during imprinted XCI, which is consistent with previous observations from transgene studies (Okamoto et al., 2005).

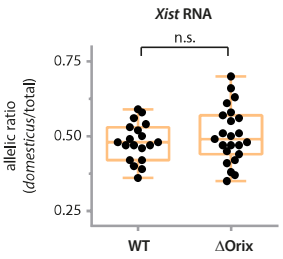




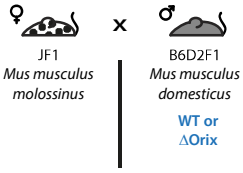
C



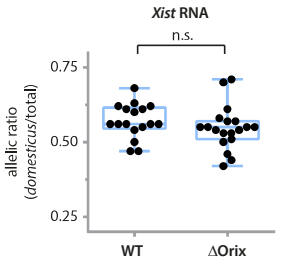
	WT	ΔOrix
No. litters	4	2
No. ♀ embryos	19	5
No. total embryos	32	12



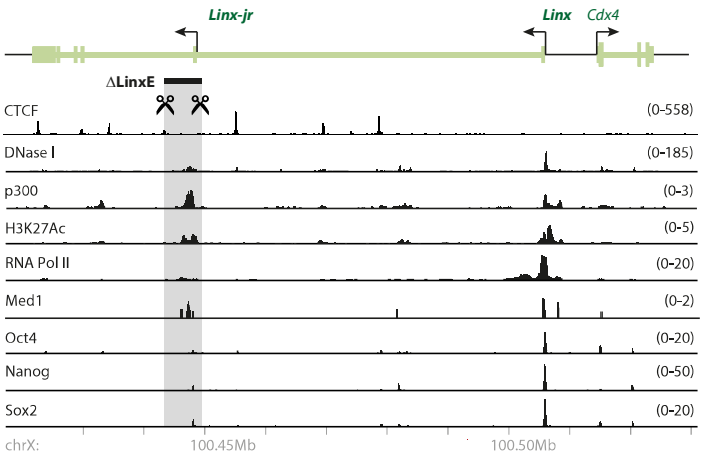
D



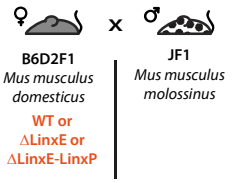
	WT	ΔOrix
No. litters	6	4
No. ♀ embryos	17	18
No. total embryos	32	27



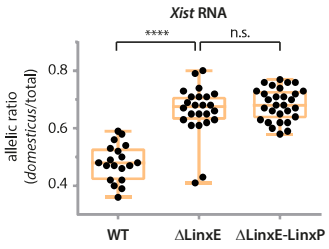
E

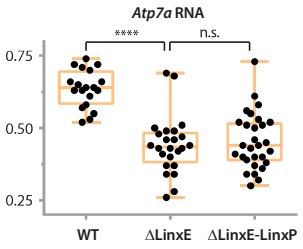


F

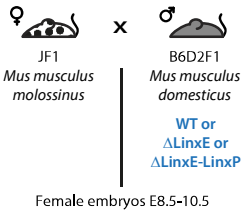


	WT	ΔLinxE	ΔLinxE-LinxP
No. litters	4	6	6
No. ♀ embryos	19	25	29
No. total embryos	32	42	46

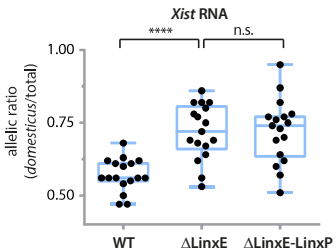




G



	WT	ΔLinxE	ΔLinxE-LinxP
No. litters	6	8	5
No. ♀ embryos	17	17	17
No. total embryos	32	40	32



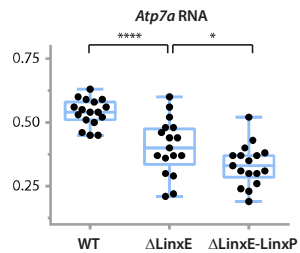
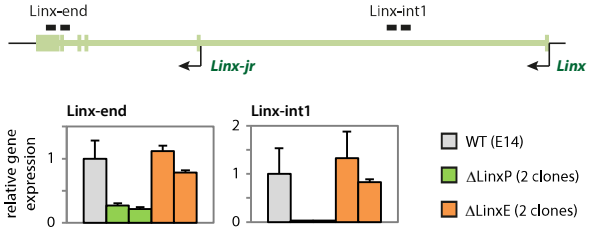


Figure S2. Characterisation of ΔOrix, ΔLinxE and ΔLinxE-LinxP, Related to Figure 2

(**A**) ATAC-seq data for the Tsix-TAD region in differentiating XX mESC – second replicate shown. See legend of Fig. 2A for more details. Pale red box highlights the *Orix* element. (**B**) Heatmap representation of nCounter analysis (see Methods) of wildtype (wt) and ΔOrix (Δ) male mESCs during differentiation. Data is normalised to wt-d0 for each gene, and represents the average of two biological replicates (wt) or the average of two biological replicates from two independent mutant clones (Δ). Statistical analysis: two-way ANOVA (** $p<0.01$). (**C**, **D**) On the left, schematic illustrations of the crosses used for analysis of RNA allelic ratios in wildtype and heterozygous E8.5-E10.5 female hybrid embryos (*molossinus*/ *domesticus*). Tables summarise number of embryos collected. On the right, analysis of allelic ratios for Xist RNA; each black dot represents the ratio for a single female embryo. Statistical analysis: two-tailed t-test. (**E**) Schematic representation of the *Linx* locus and its chromatin features (see Methods for sources of datasets represented). Position of introns and exons is based on Nora et al, 2012 (Nora et al., 2012) and mESC RNA SCRIPTURE (Guttman et al., 2010). Targeted region *LinxE* (~6kb) is indicated. Coordinates (mm9) – chrX: 100416637-100531447. (**F, G**) Schematic illustration of the crosses used for analysis of RNA allelic ratios in wildtype and heterozygous E8.5-E10.5 female hybrid embryos (*molossinus*/*domesticus*). Tables summarise number of embryos collected. Graphs show analysis of RNA allelic ratios for Xist and Atp7a, an X-linked gene. Each black dot represents the ratio for a single female embryo. Statistical analysis: Tukey's multiple comparisons test (* $p<0.05$; **** $p<0.0001$). **Note:** We could still detect some transcripts at the 3' of the locus in ~10% of cells (**Fig. S3A, S3C**), likely corresponding to a reported smaller isoform of *Linx* with an alternative first exon (Nora et al., 2012), here referred to as *Linx-jr*. We also generated mice knockout for the promoter region of *Linx-jr* (ΔLinxE, ~6kb) (**Fig. S2E**), either alone or in combination with *LinxP*. Similar to ΔLinxP, ΔLinxE led to increased *Xist* expression in cis and preferential inactivation of *Atp7a* (0.68 vs 0.48, $p<0.0001$, **Fig. S2F-G**). Double *cis*-knockout of *LinxE* and *LinxP* did not have a stronger effect than *LinxE* knockout alone (**Fig. S2F-G**). The *Linx* locus therefore harbours two different negative cis-regulators of *Xist* with an impact on XCI choice. The repetitive nature of the *LinxE* DNA sequences did not allow us to study this element in more detail, and we therefore focused on *LinxP*.

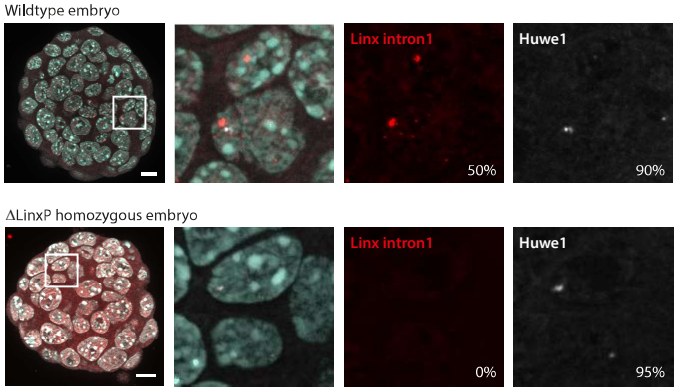
A

qRT-PCR primers



B

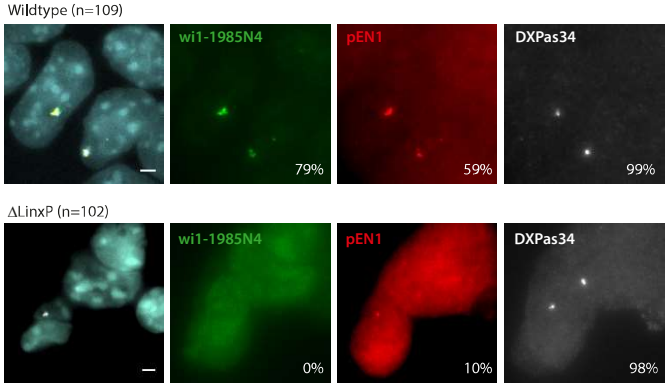
RNA FISH on E4.5 ICM



C

RNA FISH probes

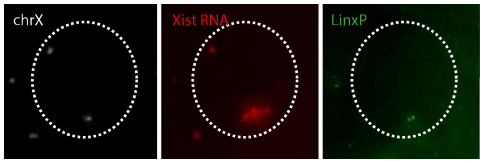
RNA FISH on male mESC



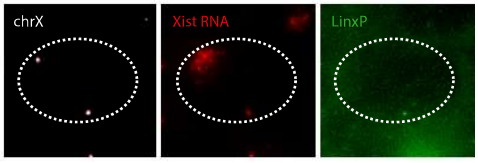
D

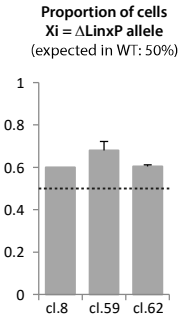
RNA/DNA FISH on female differentiating mESC (d4)

Xi = wild type X chromosome

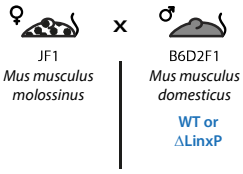


Xi = ΔLinxP X chromosome



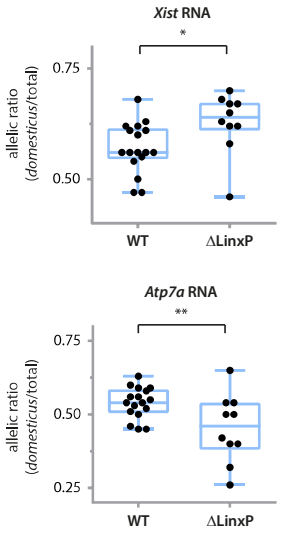


E

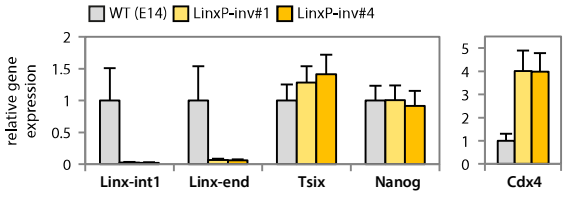


Female embryos E8.5-10.5

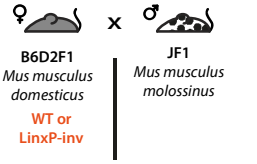
	WT	ΔLinxP
No. litters	6	5
No. ♀ embryos	17	10
No. total embryos	32	17



F

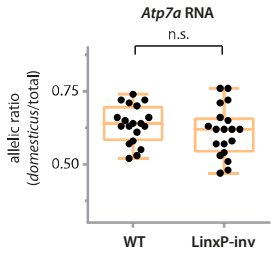


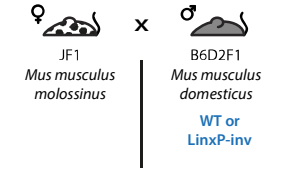
G



Female embryos E8.5-10.5

	WT	LinxP-inv
No. litters	4	3
No. ♀ embryos	19	19
No. total embryos	32	31





Female embryos E8.5-10.5

	WT	LinxP-inv
No. litters	6	5
No. ♀ embryos	17	17
No. total embryos	32	35

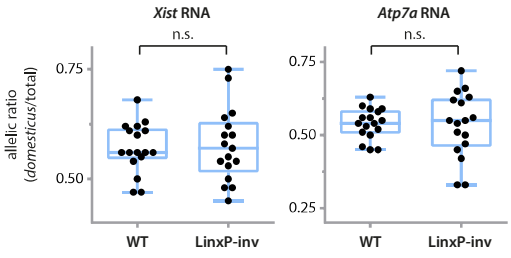


Figure S3. Characterisation of ΔLinxP mutants, Related to Figure 2

(**A**) Schematic representation of the *Linx* locus and position of the primers used for qRT-PCR analysis of wildtype, ΔLinxP and ΔLinxE mESC. Gene expression levels relative to wildtype and normalised to three control genes (geNorm; see Methods). Bars represent averages of three biological replicates for each genotype/clone. (**B**) RNA FISH for *Linx* and *Huwe1* (X-linked) on immuno-dissected inner cell masses (ICM) from E4.5 wildtype and ΔLinxP embryos. Max projections of six z-planes (~2.5μm). Exposure acquisitions and image processing were the same for wildtype and mutant. Percentages of cells in the ICM positive for *Linx* or *Huwe1* are represented. Position of *Linx* probe (wi1-1985N4) is represented in (J). Equivalent results found in two additional embryos for each genotype (data not shown). Scale bar: 10μm. (**C**) RNA FISH for different regions of *Linx* and *DXPas34* (control) on wildtype and ΔLinxP mutant male mESC. Position of the *Linx* probes used is illustrated in the scheme above. Percentages of cells positive for each probe are indicated. Equivalent results found in an independent experiment (data not shown). Scale bar: 2μm. (**D**) Determining which allele is more frequently coated by Xist RNA in isogenic female ESCs, wildtype or heterozygous for ΔLinxP, using RNA/DNA FISH. The two alleles are distinguished using a probe for the deleted region (*LinxP*). X chromosomes are identified by using a probe for the *Tsix/Xist* region. Data are presented as means and error bars represent standard deviation (two biological replicates, more than 80 cells per genotype counted for each). In wildtype cells, the proportion of cells with either one or the other X chromosome inactivated is expected to be 50:50 (dotted line) because the X chromosomes are genetically identical. (**E**) Reciprocal cross of analysis shown in Fig. 2E-G. Left, schematic illustration of the crosses used for analysis of RNA allelic ratios in wildtype and heterozygous E8.5-E10.5 female hybrid embryos (*molossinus*/*domesticus*). Table summarises number of embryos collected. Right, analysis of allelic ratios for Xist and Atp7a RNA. Each black dot represents the ratio for a single female embryo. Statistical analysis: two-tailed t-test (* $p<0.05$, ** $p<0.01$). (**F**) Gene expression analysis by qRT-PCR of wildtype and LinxP-inv mESC. Gene expression levels relative to wildtype and normalised to three control genes (geNorm; see Methods). Bars represent averages of three biological replicates for each genotype/clone. (**G**) Complementary analysis (Atp7a RNA allelic ratios) and reciprocal cross of analysis shown in Fig. 2L. Schematic illustrations represent the crosses used for analysis of RNA allelic ratios in wildtype and heterozygous E8.5-E10.5 female hybrid embryos (*molossinus*/*domesticus*) and tables summarise number of embryos collected. Graphs show analysis of allelic ratios for Xist and Atp7a RNA; each black dot represents the ratio for a single female embryo. Statistical analysis: two-tailed t-test. **Note:** The *Linx-jr* RNA does not seem involved in regulating *Xist*. Upon inversion of the *LinxP* element, which does not have an impact on *Xist* expression nor XCI choice in mouse (**Fig. 2H-J; S3G**), *Linx-jr* transcripts cannot be detected (**Fig. S3F**). The absence of *Linx-jr* transcripts is therefore associated with an absence of an effect on *Xist* expression or XCI choice.

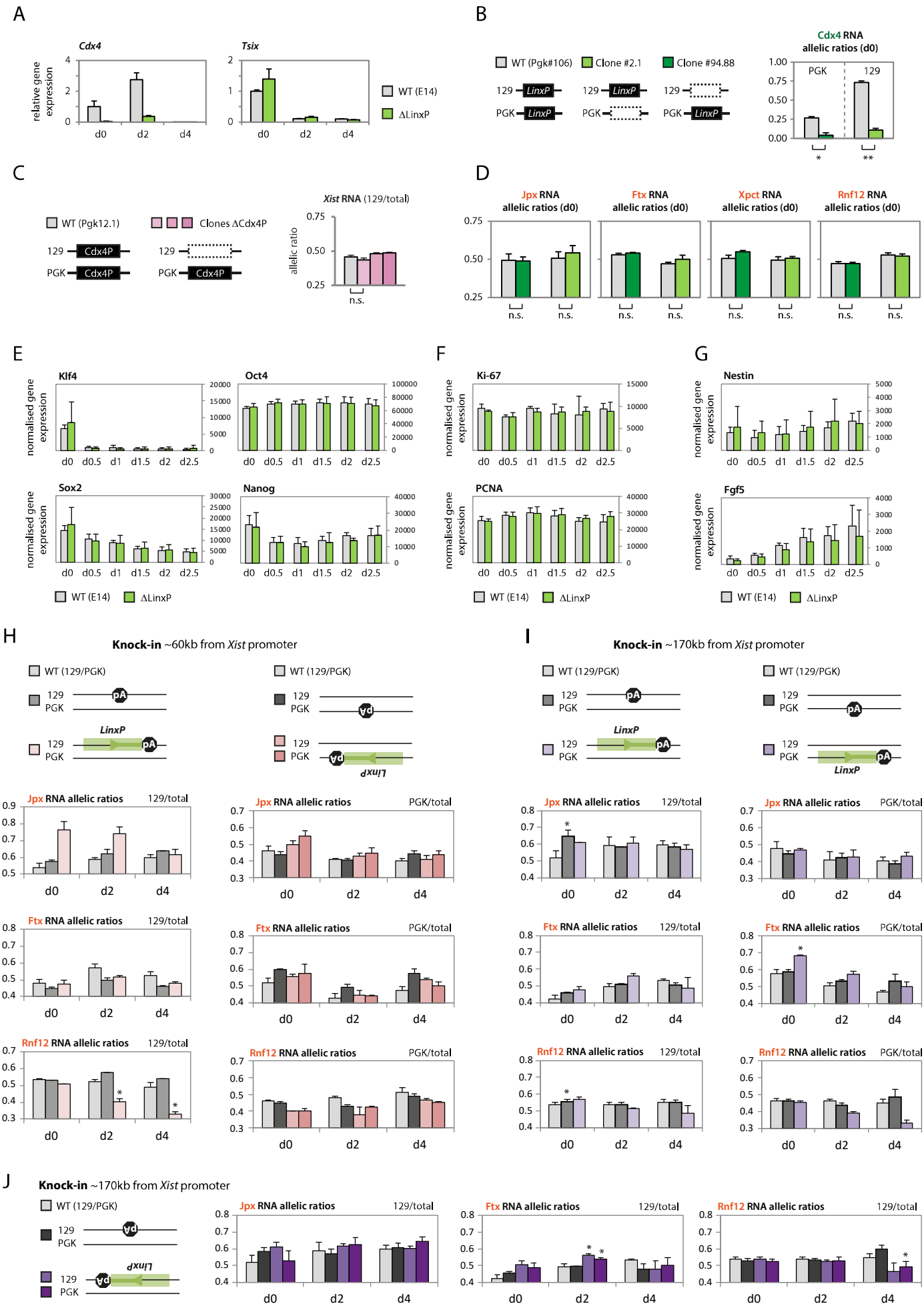
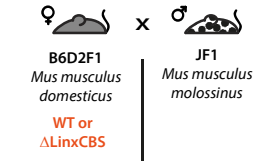

A
Cdx4
Tsix
relative gene expression
WT (E14)
ΔLinxP
d0
d2
d4
B
WT (Pgk#106)
Clone #2.1
Clone #94.88
129
PGK
LinxP
Cdx4 RNA allelic ratios (d0)
C
WT (Pgk12.1)
Clones ΔCdx4P
Cdx4P
Xist RNA (129/total)
allelic ratio
n.s.
D
Jpx RNA allelic ratios (d0)
Ftx RNA allelic ratios (d0)
Xpct RNA allelic ratios (d0)
Rnf12 RNA allelic ratios (d0)
E
Klf4
Oct4
Sox2
Nanog
normalised gene expression
d0
d0.5
d1
d1.5
d2
d2.5
F
Ki-67
PCNA
G
Nestin
Fgf5
H
Knock-in ~60kb from Xist promoter
pA
LinxP
Jpx RNA allelic ratios 129/total
Jpx RNA allelic ratios PGK/total
Ftx RNA allelic ratios 129/total
Ftx RNA allelic ratios PGK/total
Rnf12 RNA allelic ratios 129/total
Rnf12 RNA allelic ratios PGK/total
I
Knock-in ~170kb from Xist promoter
J
Knock-in ~170kb from Xist promoter

Figure S4. Characterisation of ΔLinxP and LinxP-knockin mutants, Related to Figure 3 and 5

(**A**) Gene expression analysis by qRT-PCR of wildtype and ΔLinxP male mESC during early differentiation. Gene expression levels relative to wildtype (d0) and normalised to three control genes (geNorm; see Methods). Bars represent averages of three biological replicates for each genotype. (**B**) Allelic quantification of Cdx4 RNA by pyrosequencing in hybrid (129/PGK) female ESCs, wildtype or heterozygous for ΔLinxP. Note that each clone harbours the deletion in a different allele and Cdx4 RNA allelic ratios are shown from one or the other allele (PGK or 129), depending on the mutant clone that is being compared. Data are presented as means and error bars represent SEM (three biological replicates). Statistical analysis: two-tailed paired t-test (** $p<0.01$). (**C**) Analysis of Xist RNA allelic ratios in wildtype female mESC and heterozygous ΔCdx4P clones at day 4 of differentiation. In each mutant female clone, the 129 allele harbours the deletion. Average of three replicates is shown for each genotype/clone with error bars representing SEM. Statistical analysis: paired two-tailed t-test. (**D**) Allelic quantification of Jpx, Ftx, Xpct and Rnf12 RNA by pyrosequencing in hybrid (129/PGK) female ESCs, wildtype or heterozygous for ΔLinxP. Note that each clone harbours the deletion in a different allele and Cdx4 RNA allelic ratios are shown from one or the other allele (PGK or 129), depending on the mutant clone that is being compared. Data are presented as means and error bars represent SEM (three biological replicates). Statistical analysis: two-tailed paired t-test. (**E**, **F**, **G**) Gene expression analysis using nCounter (see Methods) of wild type (grey) and ΔLinxP (green) male mESCs during differentiation. Data is normalised to six reference genes (see Methods), and represents the average of RNA counts from four biological replicates for each genotype. (**H**, **I**, **J**) Allelic quantification of Jpx, Ftx or Rnf12 RNA by pyrosequencing in hybrid (129/PGK) female ESCs, wildtype or harbouring a knock-in cassette, at differentiation time points d0, d2 and d4. Note that for each clone, the cassette was knocked-in either on the 129-X chromosome or the PGK-X chromosome, and the RNA allelic ratios are shown for each clone relative to the knock-in allele. Data are presented as means and error bars represent SEM (three biological replicates each). Statistical analysis: two-tailed paired t-test (* $p<0.05$). Clones harbouring the polyA cassette alone (shades of grey) were compared to wild type (WT), while clones harbouring the *LinxP* element (shades of salmon and purple) were compared to the clones harbouring the polyA cassette alone. **Note 1:** We characterised the transcription status of all *Xic* genes in ΔLinxP and also ΔLinxE male mESC using nCounter technology, qPCR and/or RNA-seq (see Methods). The ΔLinxE allele is not associated with any changes in gene expression across the *Xic* (data not shown). In ΔLinxP male mESC, we observed that *Cdx4,* located ~10kb upstream of *Linx*, was dramatically downregulated (**Fig. S4A**). *Cdx4* expression was also affected in *cis* in ΔLinxP heterozygous female mESC (**Fig. S4B**) and in mutants harbouring a polyA cassette downstream of *LinxP* (**Fig. S6C**). To address whether *Cdx4* expression could be involved in regulating *Xist* in *cis*, we generated heterozygous mutants of the *Cdx4* promoter (ΔCdx4P) in female ESCs and compared *Xist* allelic ratios upon differentiation. No difference was found between heterozygous ΔCdx4P mutants and control ESCs (**Fig. S4C**),

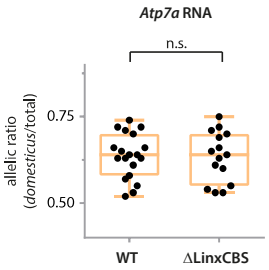
excluding the hypothesis that *LinxP* could be affecting *Xist* expression in *cis* via *Cdx4*. We also assessed whether ΔLinxP could be affecting other genes within the Xist-TAD, but allelic ratios for *Jpx*, *Ftx*, *Xpct* or *Rnf12* were not significantly different between ΔLinxP heterozygous and control female ESCs (**Fig. S4D;** unlike *Xist,* **Fig. 3B**). No other gene within the *Xic* or genome-wide (as revealed by RNA-seq; data not shown but available with this paper) was consistently affected by ΔLinxP, including markers for pluripotency, differentiation and proliferation (**Fig. S4E, S4F, S4G**). **Note 2:** We observed preferential expression of either *Jpx* or *Ftx* for the *LinxP* knock-ins between *Jpx* and *Ftx*, or *Ftx* and *Xpct*, respectively; however, this effect was not consistent across clones nor across differentiation (**Fig. S4H-J**), in contrast to the effect on *Xist*. *Xist* activation in *cis* by the *LinxP* knock-ins was accompanied by skewed silencing of *Rnf12* during differentiation in some clones (**Fig. S4H-J**).

A

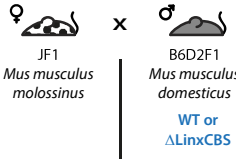


Female embryos E8.5-10.5

	WT	ΔLinxCBS
No. litters	4	5
No. ♀ embryos	19	15
No. total embryos	32	36

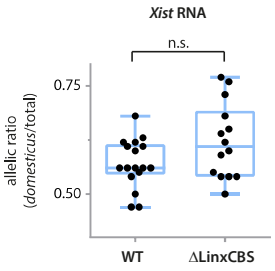


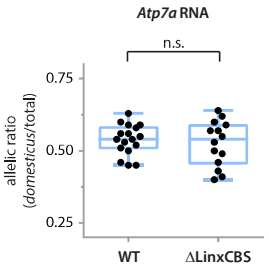
B



Female embryos E8.5-10.5

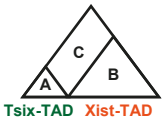
	WT	ΔLinxCBS
No. litters	6	5
No. ♀ embryos	17	10
No. total embryos	32	17





E

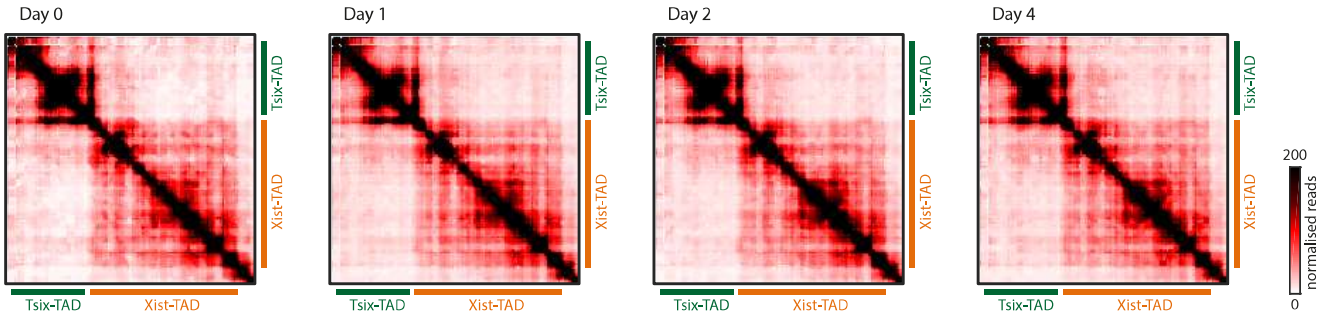
Proportion of inter-TAD interactions (Fig. 4J)
Calculated as C / (A+B+C)



A = sum of interactions within Tsix-TAD
B = sum of interactions within Xist-TAD
C = sum of inter-TAD interactions

C

5C profiles of differentiating female mESCs (Pgk12.1)



D

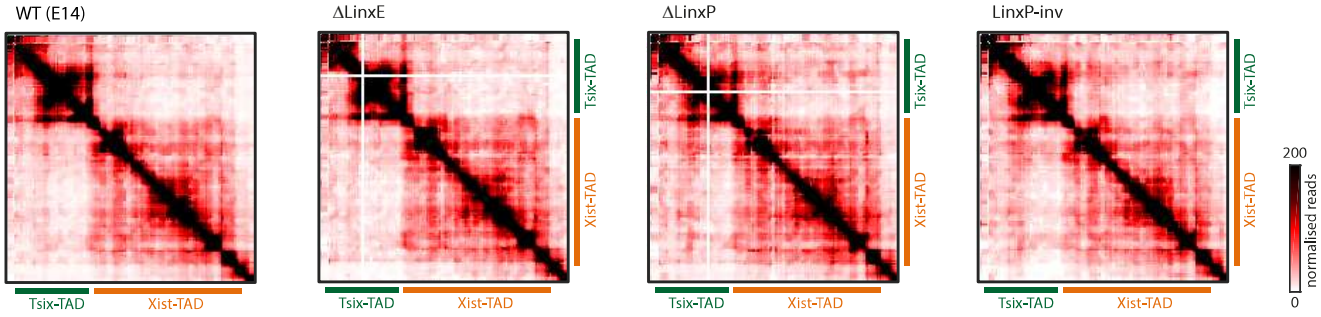


Figure S5. Characterisation of ΔLinxCBS mutants and chromosome conformation analysis of female mESCs, Related to Figure 4

(**A**) Complementary analysis (Atp7a RNA allelic ratios) of Fig. 4D. Left, schematic illustration of the crosses used for analysis of RNA allelic ratios in wildtype and heterozygous E8.5-E10.5 female hybrid embryos (*molossinus*/*domesticus*) and table summarising number of embryos collected. Right, analysis of allelic ratio for Atp7a RNA; each black dot represents the ratio for a single female embryo. Statistical analysis: two-tailed t-test. (**B**) Reciprocal cross of analysis shown in Fig. 4D. Left, schematic illustration of the crosses used for analysis of RNA allelic ratios in wildtype and heterozygous E8.5-E10.5 female hybrid embryos (*molossinus*/*domesticus*) and table summarising number of embryos collected. Right, analysis of allelic ratios for Xist and Atp7a RNA; each black dot represents the ratio for a single female embryo. Statistical analysis: two-tailed t-test. (**C**) 5C profiles of female mESC (Pgk12.1) during early differentiation; pooled data from two or three biological replicates for each time point. See Methods for more details. (**D**) 5C profiles of wildtype, ΔLinxE, ΔLinxP and LinxP-inv male mESC; pooled data from two biological replicates for each genotype. Differential maps to wildtype shown in Fig.4G-I. See Methods for more details. (**E**) Supporting figure for Fig. 4J, depicting the calculations of the proportion of inter-TAD contacts.

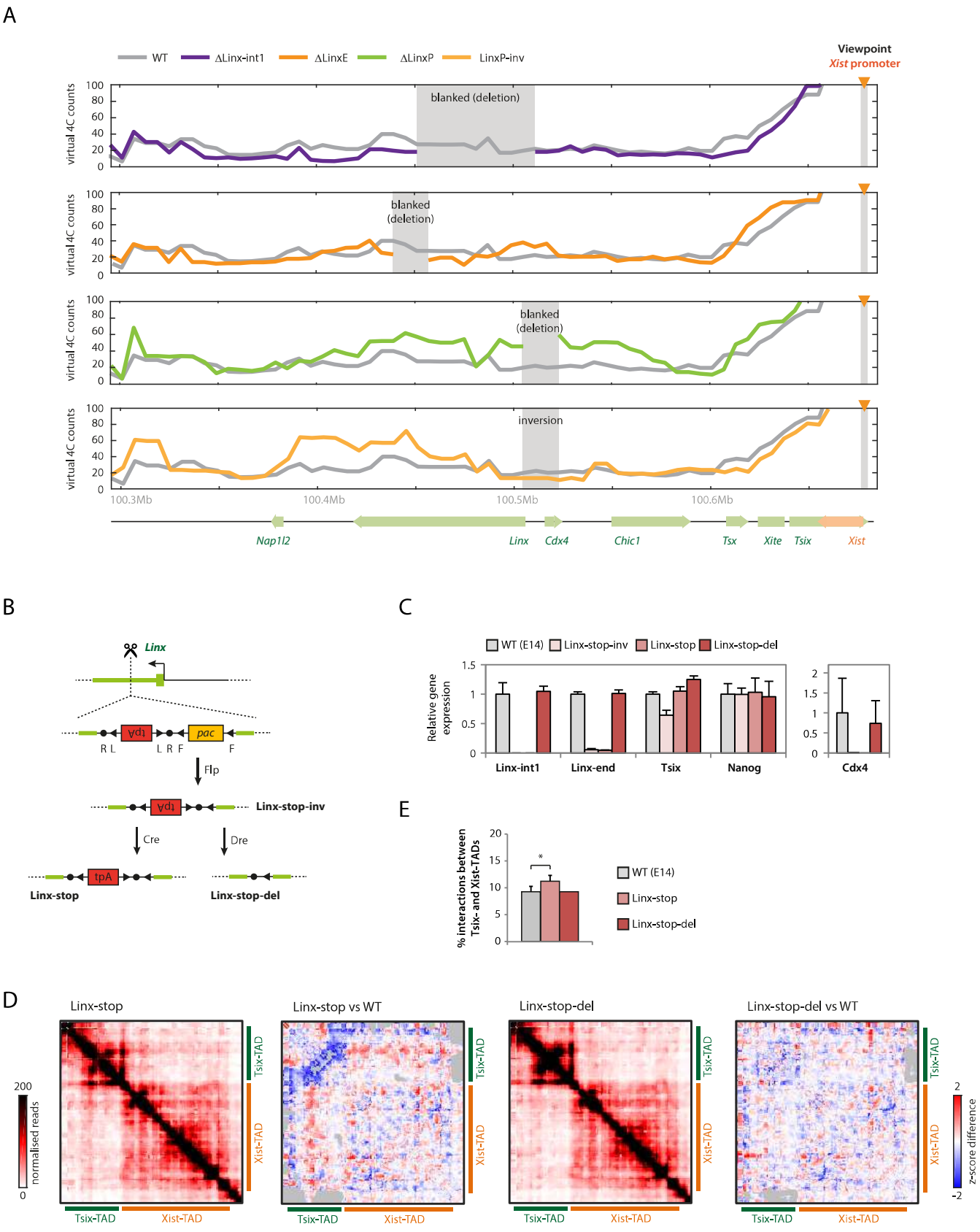

A
WT
ΔLinx-int1
ΔLinxE
ΔLinxP
LinxP-inv
Viewpoint
Xist promoter
virtual 4C counts
blanked (deletion)
inversion
100.3Mb
100.4Mb
100.5Mb
100.6Mb
Nap1l2
Linx
Cdx4
Chic1
Tsx
Xite
Tsix
Xist
B
Linx
pac
R L
L R F
F
Flp
Linx-stop-inv
Cre
Dre
tpA
Linx-stop
Linx-stop-del
C
WT (E14)
Linx-stop-inv
Linx-stop
Linx-stop-del
Relative gene expression
Linx-int1
Linx-end
Tsix
Nanog
Cdx4
E
% interactions between Tsix- and Xist-TADs
WT (E14)
Linx-stop
Linx-stop-del
D
Linx-stop
Linx-stop vs WT
Linx-stop-del
Linx-stop-del vs WT
normalised reads
z-score difference
Tsix-TAD
Xist-TAD

Figure S6. Characterisation of Linx-stop mutants and chromosome conformation analysis of ΔLinxP, ΔLinxE and LinxP-inv mutants, Related to Figure 4

(**A**) Virtual 4C plot generated from 5C data, using the bin containing the *Xist* promoter as view point. (**B**) Schematic representation of the knock-in strategy for inserting a stop-cassette ~1kb downstream of the *LinxP* element. Selection marker was removed (flipase, Flp) and polyA signal inverted to correct orientation (Cre). As a control, the cassette was removed (Dre). (**C**) Gene expression analysis by qRT-PCR of wildtype and Linx-stop mESC. Gene expression levels relative to wildtype and normalised to three control genes (geNorm; see Methods). Bars represent averages of three biological replicates for each genotype/clone. (**D**) 5C profiles of Linx-stop male mESCs; pooled data from two biological replicates. Differential maps represent the subtraction of Z-scores calculated for wildtype data from Z-scores calculated for mutant data (see Methods). Grey pixels correspond to to contacts that were filtered because they did not meet the quality control threshold (see Methods). (**E**) Quantification of 5C inter-TAD contacts in wildtype, Linx-stop and Linx-stop-del mESC (see Fig.S5E for details on calculations). Bars represent the average of the calculated proportions of four (E14, Linx-stop) or two (Linx-stop-del) independent replicates. Statistical analysis: two-tailed t-test (* $p<0.05$).