HAL
open science

The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome

Michal R Gdula, Tatyana B Nesterova, Greta Pintacuda, Jonathan Godwin, Ye Zhan, Hakan Ozadam, Michael Mcclellan, Daniella Moralli, Felix Krueger, Catherine M Green, et al.

To cite this version:

Michal R Gdula, Tatyana B Nesterova, Greta Pintacuda, Jonathan Godwin, Ye Zhan, et al.. The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome. Nature Communications, 2019, 10 (1), pp.30. 10.1038/s41467-018-07907-2 . hal-01973783

HAL Id: hal-01973783

https://hal.sorbonne-universite.fr/hal-01973783v1

Submitted on 8 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-018-07907-2 OPEN

The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome

Michal R. Gdula [1], Tatyana B. Nesterova [1], Greta Pintacuda [1], Jonathan Godwin[1], Ye Zhan [2], Hakan Ozadam[2], Michael McClellan[3], Daniella Moralli [4], Felix Krueger [5], Catherine M. Green[4], Wolf Reik[6], Skirmantas Kriaucionis[3], Edith Heard [7], Job Dekker [2] & Neil Brockdorff[1]

The inactive X chromosome (Xi) in female mammals adopts an atypical higher-order chromatin structure, manifested as a global loss of local topologically associated domains (TADs), A/B compartments and formation of two mega-domains. Here we demonstrate that the non-canonical SMC family protein, SmcHD1, which is important for gene silencing on Xi, contributes to this unique chromosome architecture. Specifically, allelic mapping of the transcriptome and epigenome in SmcHD1 mutant cells reveals the appearance of sub-megabase domains defined by gene activation, CpG hypermethylation and depletion of Polycomb-mediated H3K27me3. These domains, which correlate with sites of SmcHD1 enrichment on Xi in wild-type cells, additionally adopt features of active X chromosome higher-order chromosome architecture, including A/B compartments and partial restoration of TAD boundaries. Xi chromosome architecture changes also occurred following SmcHD1 knockout in a somatic cell model, but in this case, independent of Xi gene derepression. We conclude that SmcHD1 is a key factor in defining the unique chromosome architecture of Xi.

[1] Developmental Epigenetics, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. [2] Program in Systems Biology, Department of Medicine, Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Howard Hughes Medical Institute, Worcester, MA 01605, USA. [3] Ludwig Cancer Research, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7DQ, UK. [4] The Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. [5] Bioinformatics Group, The Babraham Institute, Cambridge CB22 AT, UK. [6] Epigenetics Program, The Babraham Institute, Cambridge CB22 AT, UK. [7] Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, UPMC Paris-Sorbonne, 26 Rue d'Ulm, 75005 Paris, France. These authors contributed equally: Michal R. Gdula, Tatyana B. Nesterova. Correspondence and requests for materials should be addressed to N.B. (email: neil.brockdorff@bioch.ox.ac.uk)

X chromosome inactivation is the mechanism that evolved in mammals to equalise levels of X-linked gene expression in XX females relative to XY males. Cells of early female embryos selectively inactivate a single X chromosome, usually at random, resulting in the formation of a stable heterochromatic structure, the Barr body. The inactive X chromosome (Xi), once established, is highly stable, and is maintained in somatic cells throughout the lifetime of the animal[1,2]. The X inactivation process is triggered by the non-coding RNA Xist, which localises to the Xi territory to induce chromosome-wide gene silencing[3–6].

Chromatin features that distinguish Xi and the active X chromosome (Xa) include specific histone post-translational modifications, variant histones and CpG DNA methylation (reviewed in ref. [2]). Additionally, Xi acquires a characteristic higher-order chromosome structure. Specifically, A-type chromatin compartments, corresponding to gene-rich regions which normally replicate in early S-phase, switch to replication in mid- or late-S-phase (reviewed in ref. [7]). Additionally, topologically associated domains (TADs), sub-megabase scale domains which are formed by the activity of cohesin, restricted at boundaries by oppositely oriented binding sites for the insulator protein CTCF[8–13], are in large part absent on Xi, being replaced instead by two large mega-domains that are separated by a hinge that encompasses the DXZ4 repeat sequence[14–18]. The basis for this unique TAD structure is not well understood, but is thought to depend, at least in part, on ongoing expression of Xist RNA[17].

Barr body formation is a multistep process. Thus, Xist RNA recruits specific chromatin modifiers, including the SPEN-NCoR-HDAC3 complex[19–22], required for histone deacetylation[22], and the PRC1 and PRC2 Polycomb complexes, required for deposition of H2A lysine 119 ubiquitylation (H2AK119u1) and H3 lysine 27 methylation (H3K27me3), respectively[23–27]. The lamin B receptor[22,28] and m6A RNA modification complex[19,29] have also been implicated in establishment of chromosome-wide gene silencing. Other factors are recruited to Xi at later stages. Examples include the variant histone macroH2A[30], and the non-canonical SMC protein SmcHD1[31]. The role of these factors remains to be defined, although is likely to be linked to the long-term stability of the inactive state.

SmcHD1 is classified as an SMC protein by virtue of an SMC hinge domain at the C-terminal end, but differs from canonical SMC complexes in having a functional GHKL-ATPase domain rather that a Walker A/B type ATPase domain[32]. Biochemical and biophysical studies indicate that SmcHD1 homodimerises via the hinge and GHKL domains to form a complex that is reminiscent of bacterial SMC proteins, both in form and scale[33], albeit forming a functional homodimer rather than a trimeric complex. SmcHD1 performs an important role in silencing on Xi, and at selected mono-allelically expressed autosomal loci[31,32,34,35]. Whilst it is known that a proportion of Xi genes are activated in SmcHD1 mutant embryos[34,35], the molecular mechanism is not well understood. Notably, although SmcHD1 is required for DNA methylation at CpG island (CGI) promoters of many Xi genes, loss of CGI methylation does not appear to account for the observed gene activation[34]. An alternative hypothesis is that SmcHD1-mediated compaction of Xi, inferred by microscopy based analyses in human cell lines[36], imposes gene repression. Given the important role of SMC family proteins in genome topology, we set out to investigate the role of SmcHD1 in the higher-order architecture of Xi. Thus, we performed high-resolution analysis of Xi transcription, epigenetic features, and higher-order chromatin features in SmcHD1 mutant cell lines.

Here we find that SmcHD1 loss of function results in the appearance of sub-megabase domains defined by gene activation, CpG hypermethylation and depletion of Polycomb-mediated H3K27me3. These domains, which correlate with sites of SmcHD1 enrichment on Xi in wild-type cells, additionally adopt features of active X chromosome higher-order chromosome architecture, including A/B compartments and partial restoration of TAD boundaries. Xi chromosome architecture changes also occurred following SmcHD1 knockout in a somatic cell model, but in this case, independent of Xi gene derepression.

Results

Widespread Xi gene activation in SmcHD1 mutant MEFs. In order to gain insight into the role of SmcHD1 in X inactivation we set out to analyse epigenomic and long-range chromatin features of Xi at high resolution. Thus, we derived XX MEF lines from *Mus musculus domesticus* (domesticus) × *Mus musculus castaneus* (castaneus) female embryos, that were either wild-type (wt) or SmcHD1 *null* (SmcHD1 mut). Xi was of castaneus origin in both cases (Supplementary Fig. 1a, b). The high frequency of SNPs between domesticus and castaneus genomes allows assignment of high-throughput sequencing reads to either maternal or paternal genomes. The breeding strategy enabled us to obtain cell lines from F2 embryos in which the entire X chromosome was either of domesticus or castaneus origin. Autosomes on the other hand were mosaic as a result of recombination in the F1 generation. X inactivation in the inter-specific embryos is random, so stable MEF lines were sub-cloned to obtain wt and SmcHD1 mutant lines. Xist RNA FISH and karyotype analysis confirmed the presence of Xi and Xa chromosome(s) (Supplementary Fig. 1c–e).

Initially, we performed allelic ChIP-seq analysis of SmcHD1. As shown in Fig. 1a, b, SmcHD1 is highly enriched over Xi, notably over domains that correlate with gene-dense regions across the length of the chromosome. The distribution of Smchd1 enrichment sites determined by peak-calling analysis is summarised in Fig. 1c. More than 25% of observed peaks are on the X chromosome, where they locate more frequently in intergenic regions compared with autosomes (65 vs 30%). We also observed SmcHD1 enrichment over specific regions on autosomes, including known SmcHD1 target loci such as the protocadherin locus on chromosome 18 and the PWS/AS imprinted gene cluster on chromosome 7 (Supplementary Fig. 2a, b). SmcHD1 peaks in the latter region accord with a previously published dataset analysing SmcHD1 occupancy in XY neuronal stem cells[37].

We went on to analyse Xi transcription in SmcHD1 mutant compared to wt cells using allelic chromatin RNA-seq (ChrRNA-seq), which enriches for nascent unprocessed mRNAs, thereby maximising the number of informative SNPs due to inclusion of intron sequences. The results are summarised in Fig. 1d–h. In wt cells, Xi expression was detected for a small number of genes (28/403), largely corresponding to genes previously reported to escape X inactivation (Fig. 1g, Supplementary Fig. 2c). However, in SmcHD1 mutant cells, Xi expression was seen for the majority of Xi genes (254/403) (Fig. 1d–f). For 67 SmcHD1-dependent genes, Xi expression was in a similar range to that found on Xa, with a further 163 genes showing partial or low-level activation (Fig. 1d–g, Supplementary Fig. 2c). We observed extensive overlap with 64 SmcHD1-dependent genes identified in a prior study in which we performed non-allelic microarray analysis of SmcHD1 mutant embryos[34] (Supplementary Fig. 2d). The larger number of Xi expressed genes observed in this study can be attributed to the increased sensitivity afforded by allelic ChrRNA-seq. We did not observe a clear link between SmcHD1 occupancy over gene promoters and derepression (Fig. 1h).

Analysis of the association of SmcHD1-dependent genes with genomic sequence features in the immediate chromosomal environment identified a correlation with gene-dense regions, and the location of SINE repeats, as reported previously[34]. Because a

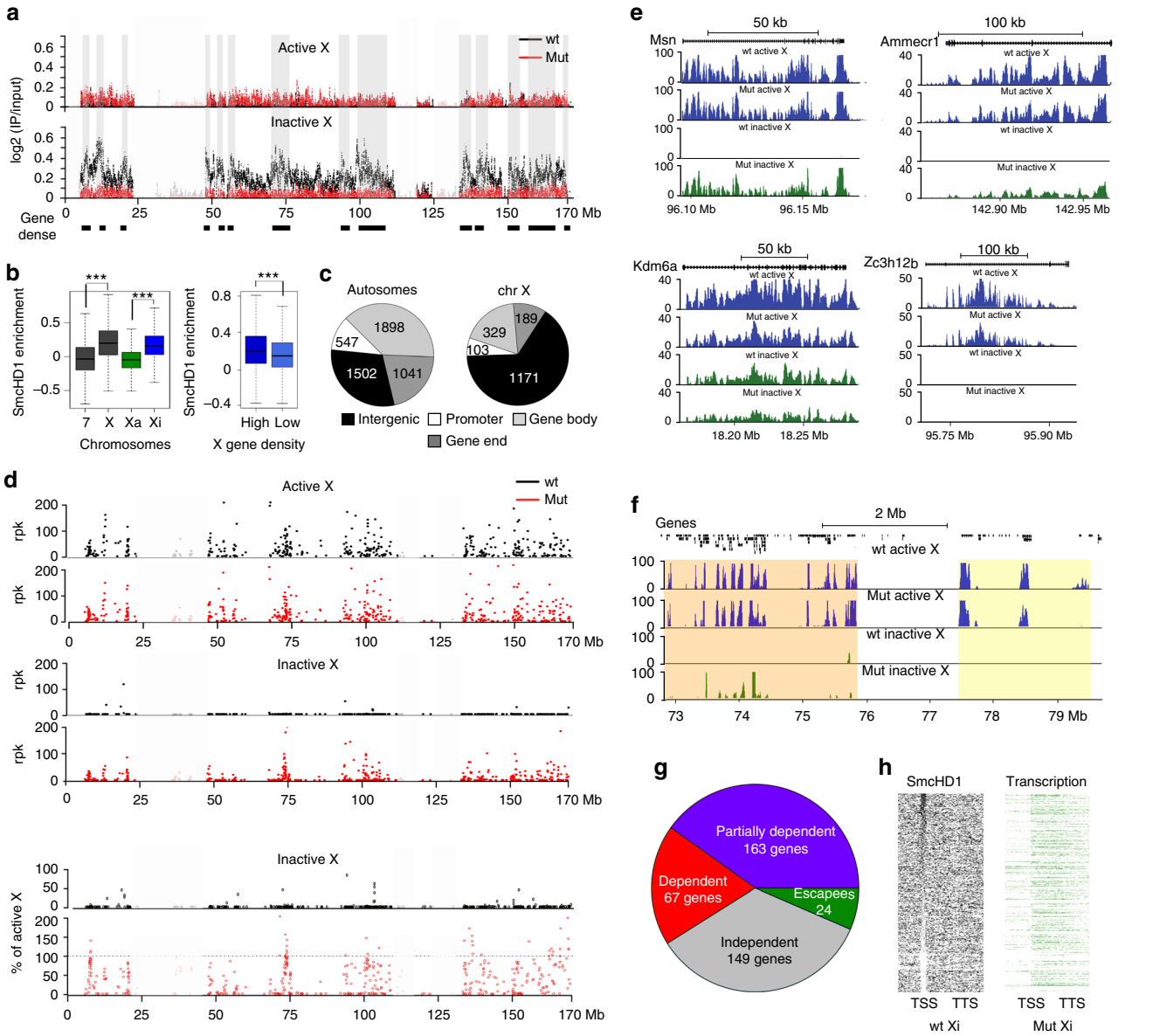


Fig. 1 Widespread derepression of Xi genes in SmcHD1 mutant MEFs. **a** Chromosome-wide profiles of SmcHD1 occupancy on Xa and Xi depicted as allele-specific ChIP-seq enrichment (log2 ratios of IP to input per 500 bp) binned into 10 kb intervals. Profiles from wt and germline SmcHD1 mutant (mut) MEFs are presented. Shaded areas and black bars below the profiles depict gene-dense areas. Throughout the figure regions of low mappability are indicated with reduced colour intensity. **b** SmcHD1 enrichment within 10 kb bins on chr7, X (both alleles), Xa and Xi. Right panel shows differences in SmcHD1 enrichment over chrX regions with high and low gene density. Boxplots present quartiles, median and outliers of ChIP enrichment. Significance estimated with Mann–Whitney test (p-value < 10^{-4}). **c** Genic and intergenic distribution of SmcHD1 peaks (MACS2, FDR < 0.05) on autosomes and chrX. **d** Quantification of gene expression on Xa and Xi with allele-specific chromatin RNA-seq in wt and mut MEFs. Normalized read count per 1 kb of gene body is shown. Lower panel depicts gene expression on wt and germline SmcHD1 mutant Xi presented as % of Xa expression. **e** Representative examples of individual X-linked genes (UCSC screen shots). **f** Example of a ~7 Mb region with a domain of strongly activated Xi genes, followed by a domain of Xi genes activated to a lesser degree (both with orange shading), and by a domain with genes expressed on Xa and silenced on Xi in mut cells (yellow shading). **g** Summary of gene expression analysis. X-linked genes expressed on Xa were divided based on their transcriptional status on Xi into SmcHD1-dependent (expression not significantly different from Xa), partially SmcHD1-dependent (de-repressed but expressed significantly lower level than Xa), SmcHD1-independent (silenced on Xi) and wt escapees. **h** Occupancy of SmcHD1 on genes (expressed on wt Xa, > 5 kb). Heatmap presents log2(IP/input) values from wt cells reduced by values for the same regions in mut cells. Right panel: de-repressed genes (transcription) on mut Xi. Genes on both heatmaps are ordered in the same way according to decrease of SmcHD1 at the TSS

relatively large proportion of Xi genes show SmcHD1 dependence, the latter correlation likely reflects the preferential location of SINE repeats within gene-dense regions. No other genomic features were found to correlate with SmcHD1 dependence.

SmcHD1 recruitment is a late step in Xist-mediated chromosome silencing[31], suggesting it has a role in the continuation or maintenance of gene silencing. To further investigate this issue we generated a somatic SmcHD1 knockout MEF line (SmcHD1 somKO) by CRISPR/Cas9-mediated mutagenesis in wt MEFs (Supplementary Fig. 3a–c). We then performed ChrRNA-seq as described above. Interestingly, and in contrast to germline SmcHD1 loss of function, no derepression of Xi genes was

observed (Supplementary Fig. 3d). This result suggests that SmcHD1 is required to reinforce gene silencing during a specific window in development, and is then dispensable, presumably reflecting compensation through other maintenance pathways.

Unique features of the Xi epigenome in SmcHD1 mutant MEFs. Previous work established that SmcHD1 is important for DNA methylation of Xi CGI promoters[31], but its role in chromosome-wide intergenic and intragenic DNA methylation on Xi has not been investigated. To address this we used whole genome bisulfite analysis (WGBS) to determine the methylome of Xa and Xi at single nucleotide resolution in wt and SmcHD1 mutant MEFs. A summary of data illustrating overall methylation density in 100 kb windows across the entire genome is shown in Fig. 2a. CpG methylation across autosomes was generally in the

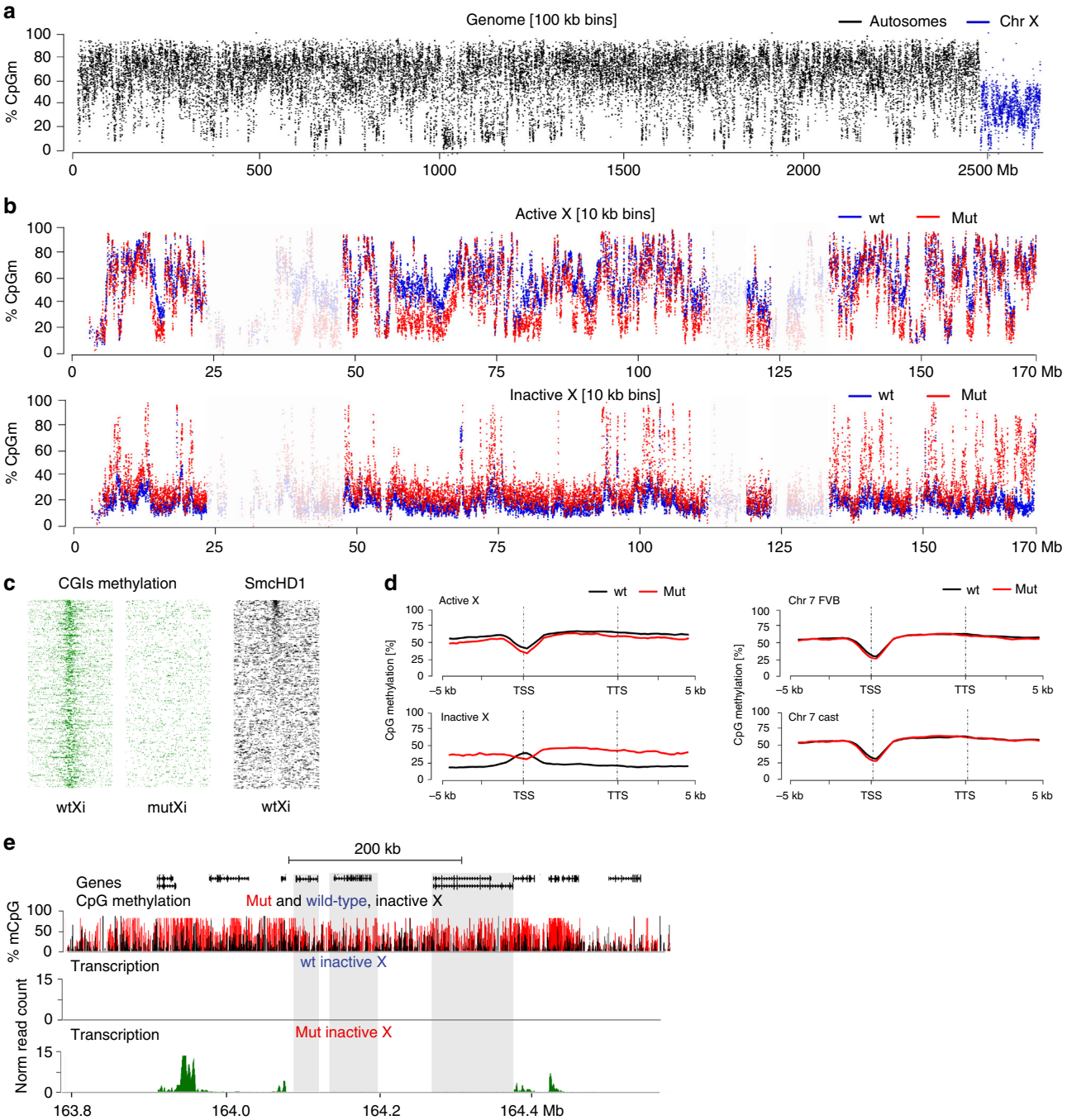


Fig. 2 The Xi methylome in wt and SmcHD1 mutant MEFs. **a** WGBS reads binned into 100 kb intervals illustrating reduced level of CpG methylation on the X chromosome relative to autosomes. **b** Allele-specific DNA methylation profile of Xa (top) and Xi (bottom) in wt and germline SmcHD1 mutant (mut) MEFs plotted as % of CpGm averaged within 10 kb bins. Regions of low mappability are indicated with reduced colour intensity. **c** DNA methylation of CGIs on wt and mut Xi (left) and SmcHD1 occupancy (right). CGIs in all heatmaps sorted according to the SmcHD1 enrichment. **d** Metagene plots of DNA methylation in wt and mut MEFs generated for Xa and Xi as well as for maternal and paternal Chr7. **e** Example of a hypermethylated domain on Xi in mut cells, showing that the region of high DNA methylation extends beyond expressed genes to encompass genes that are silenced. Non-expressed genes with DNA methylation above the background level are highlighted with shadowed boxes

range of 60–80%, but was significantly lower on the X chromosome. We noted that gene-poor regions are CpG hypomethylated across all chromosomes. An example, chromosome 7 is illustrated in Supplementary Fig. 4a, b. A similar pattern of hypomethylation is apparent in available MEF WGBS[38] (Supplementary Fig. 4c). Accordingly, we find that total CpG methylation levels in MEFs, as determined by HPLC analysis, are moderately reduced relative to ES cells and adult tissue (Supplementary Fig. 4d). Interestingly, hypomethylation of gene-poor regions was accentuated on Xa (Fig. 2b), and autosomes (Supplementary Fig. 4a) in SmcHD1 mutant cells. The molecular basis for this pattern of CpG hypomethylation is currently unknown.

We went on to compare allelic methylation levels on Xa and Xi chromosomes (Fig. 2b). In wt MEFs, Xa methylation levels were similar to autosomes (56% compared with 66% on autosomes), but Xi was extensively CpG hypomethylated across the entire chromosome (19% CpG methylation). CpG hypomethylation of gene-poor regions, as described above, likely contributes to the observed pattern on Xi. However, CpG hypomethylation was also evident in gene-rich regions, presumably linked to X inactivation status. This observation is consistent with prior studies showing reduced levels of CpG methylation on Xi relative to Xa[39,40]. Xi CGIs in wt MEFs were, as expected, highly methylated (Fig. 2c, left), and at most CGIs, methylation does not correlate with SmcHD1 occupancy (Fig. 2c, right). Additionally, we observed high levels of CpG methylation over bodies of genes that escape X inactivation (Supplementary Fig. 4e).

In SmcHD1 mutant MEFs, Xi CGIs were in most cases hypomethylated (Fig. 2c), as previously reported[31]. Conversely, we observed several domains with relatively high CpG methylation (Fig. 2b, bottom). These domains correspond in most cases with Xi genes that are activated in SmcHD1 mutant MEFs, although CpG hypermethylation is not restricted to transcribed sequences, extending both 5′ and 3′ (Fig. 2d, e). Individual genes that normally escape X inactivation were also CpG hypermethylated on Xi, similar to wt MEFs (Supplementary Fig. 4e).

The histone modification H3K27me3 catalysed by the major Polycomb complex PRC2, is highly enriched on Xi as determined by immunostaining[23,24], and high-resolution ChIP-seq analysis[41]. Previously we noted that this feature is not grossly affected by germline SmcHD1 loss of function, as determined by immunostaining of interphase nuclei in XX embryos[32]. Similarly, H3K27me3 enrichment on Xi was readily detected by immunostaining in the wt and SmcHD1 mutant MEF lines described herein (Supplementary Fig. 3b). However, allelic ChIP-seq analysis revealed domains in which Xi H3K27me3 is markedly depleted in SmcHD1 mutant cells (Fig. 3a). Similar data were obtained in an independent SmcHD1 mutant cell line in which the Xi is of FVB origin (Fig. 3b, Supplementary Fig. 5a). These domains (estimated 77 with mean length of 140 kb), comprise ~10% of the Xi regions over which H3K27me3 is normally enriched. The location of the H3K27me3 depleted domains correlates closely with CpG hypermethylation and Xi gene activation (Fig. 3c–f). However, as noted for CpG hypermethylation domains, the H3K27me3 depleted regions extend beyond the boundaries of activated genes (Supplementary Fig. 5b), and in some cases do not include known genes (Supplementary Fig. 5c). Together these results suggest that modified epigenomic features in SmcHD1 mutant cells correlate with domains in which Xi genes are activated. We note that in SmcHD1 mutant compared to wt cells global levels of H3K27me3 are higher on Xi, Xa and autosomes, shown in Supplementary Fig. 5d.

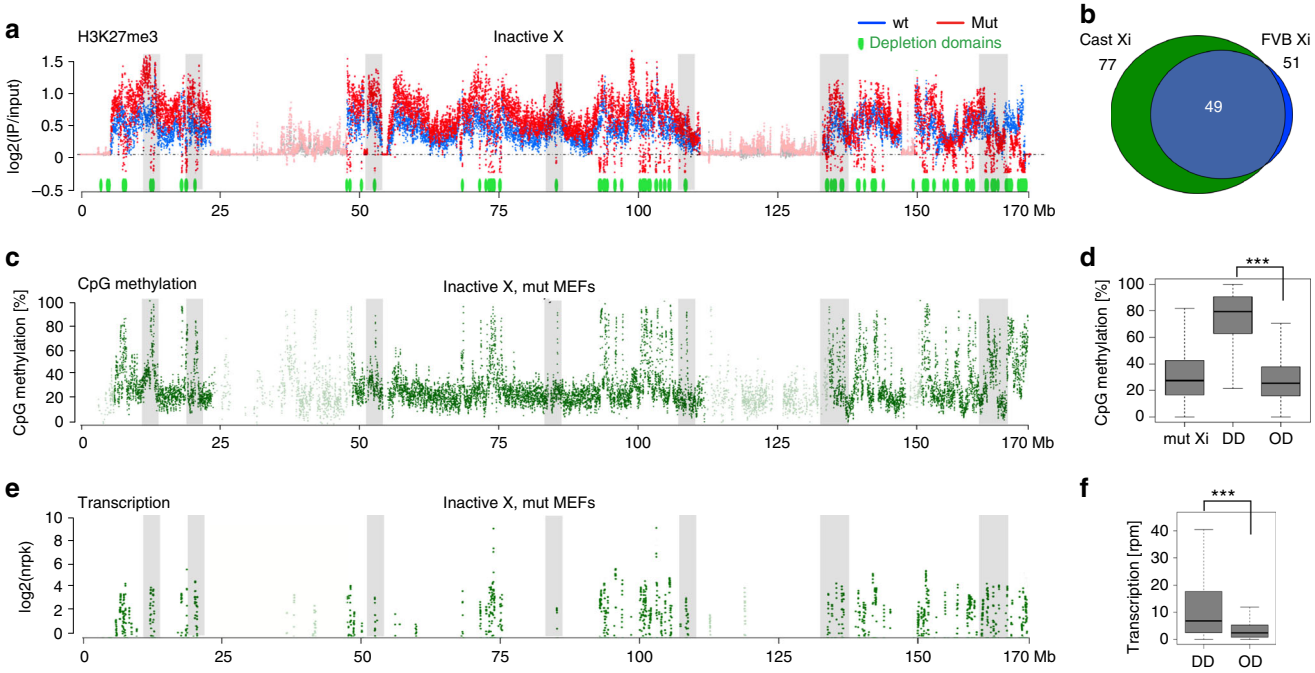


Fig. 3 H3K27me3 depletion highlights domains of Xi gene activation in SmcHD1 mutant MEFs. **a** Allele-specific H3K27me3 ChIP-seq illustrating multiple H3K27me3 depleted domains not present on wt Xi. Scatterplot depicts ChIP log2(IP/input) values averaged within 10 kb bins. 77 H3K27me3 depleted domains are plotted in green below the H3K27me3 profile. Regions of low probe mappability are indicated with reduced transparency. **b** Overlap between H3K27me3 depletion domains detected in the germline SmcHD1 mutant (mut) mut5 A6 clone used throughout the study with Xi derived from M.m. castaneus (Xa from FVB strain) with different MEF cell line mut5 E1 coming from different parents and with Xi from FVB (Xa from M.m.castaneus). Location of H3K27me3 depleted domains on Xi in mut MEFs correlates with, CpG hypermethylation domains (**c**, **d**) and genes activated on Xi (**e**, **f**). Significance of differences in DNA methylation level (**d**) and transcription (**f**) in the domains depleted of H3K27me3 and the rest of the chromosome was estimated with Mann–Whitney ***p-value < 10^{-4}. Grey shading in **a**, **c**, **e** highlights the correlation between H3K27me3, CpG methylation and transcription at the representative examples of loci depleted of H3K27me3. Boxplots in **d**, **f** present quartiles, median and outliers

Modified Xi architecture in SmcHD1 mutant MEFs. In light of evidence that SmcHD1 influences Xi chromatin at the level of sub-megabase domains, we went on to directly analyse parameters of long-range chromatin architecture. Mammalian chromosomes are comprised of distinct gene-rich and gene-poor compartments with sizes ranging from sub-megabase to several megabases in length. These regions are replicated co-ordinately, either during early- or mid/late-S phase, respectively. Replication-timing domains are broadly synonymous with A- and B- type chromatin, which in studies on long-range chromosome topology, have been shown to self-associate[42]. Additionally, B-type domains overlap extensively with Lamin associated domains (LADs) which localise to the nuclear periphery[43]. Xi is unusual in that both gene-rich and gene-poor compartments replicate synchronously in mid/late S-phase[7].

A previous analysis of a human XX somatic cell line in which SMCHD1 was depleted using siRNA, revealed an aberrant replication timing pattern for Xi, with the appearance of early replicating domains as determined using a cytogenetic assay[36]. With this observation in mind we set out to determine allelic temporal replication patterns in our wt and SmcHD1 mutant MEFs using RepliSeq, a high-resolution sequencing based approach[44] (Supplementary Fig. 6a, b). In wt MEFs we observed mid/late S-phase replication across Xi, contrasting with Xa where gene-rich and gene-poor chromosome domains replicate in early- and mid/late-S phase, respectively (Fig. 4a). However, in SmcHD1 mutant MEFs, replication patterns on Xi are more similar to the Xa pattern in wt cells (Fig. 4b). Thus, we observed regions of the chromosome in which Xi replication timing overlaps with that seen on Xa (orange shading, Fig. 4b), and other regions where replication timing is either partially advanced (yellow shading, Fig. 4b), or largely unaffected. The location of Xi regions showing a more pronounced shift in replication timing correlates closely with domains of H3K27me3 depletion (Fig. 4c), a proxy for SmcHD1 localisation, Xi gene activation and CpG hypermethylation, as above (Supplementary Fig. 6c–e).

A further level of chromosome architecture is TADs, to which replication-timing domains are thought to be linked[45]. To determine if TAD organisation on Xi is affected by SmcHD1 loss of function we performed Hi–C in the germline SmcHD1 mutant, somatic SmcHD1 knockout, and wt MEF cell lines. Analysis of data for autosomes and for Xa indicated that SmcHD1 does not significantly impact on global TAD structure. An example, chromosome 7, is illustrated in Supplementary Fig. 7a–d. Along the Xi in wt cells we detect the presence of two mega-domains separated by the DXZ4-containing boundary and the general absence of TADs, as previously reported[14–18] (Fig. 5a). However, we observed several distinct features both in germline SmcHD1 mutant and SmcHD1 somatic knockout cells. The mega-domain structure and DXZ4 hinge characteristic of Xi are discernible (Fig. 5b, c), but there are increased short-range interactions and increased compartmentalisation (Fig. 5a–f). Notably, we observed restoration of Xa TAD structure across the length of the chromosome (Fig. 5e–g, Supplementary Fig. 7e). This is most clear for the distal 20 Mb of Xi (Fig. 5e), where there are also correlated changes in transcription, epigenetic modifications, and replication timing (Fig. 5h, Supplementary Fig. 7f). Analysis of interaction frequencies along the Xi further illustrates the increase in short-range interactions in germline SmcHD1 mutant MEFs (Fig. 5i), with variation across the chromosome

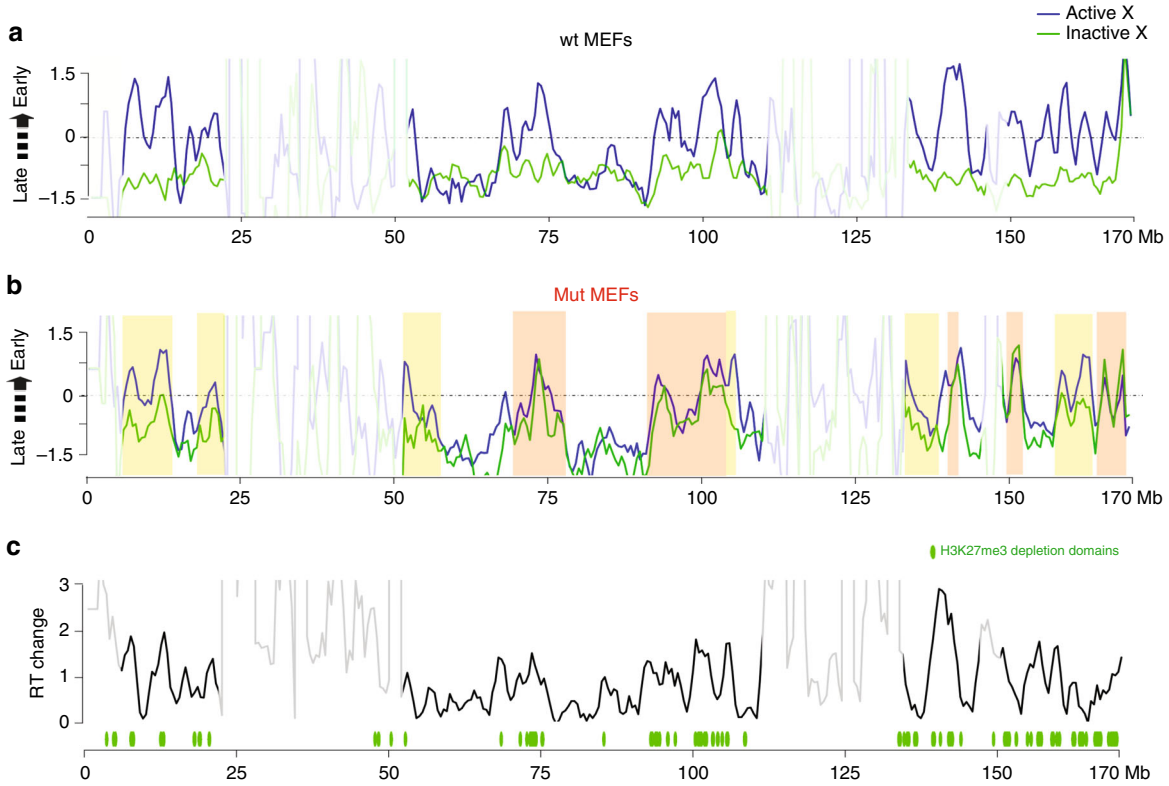
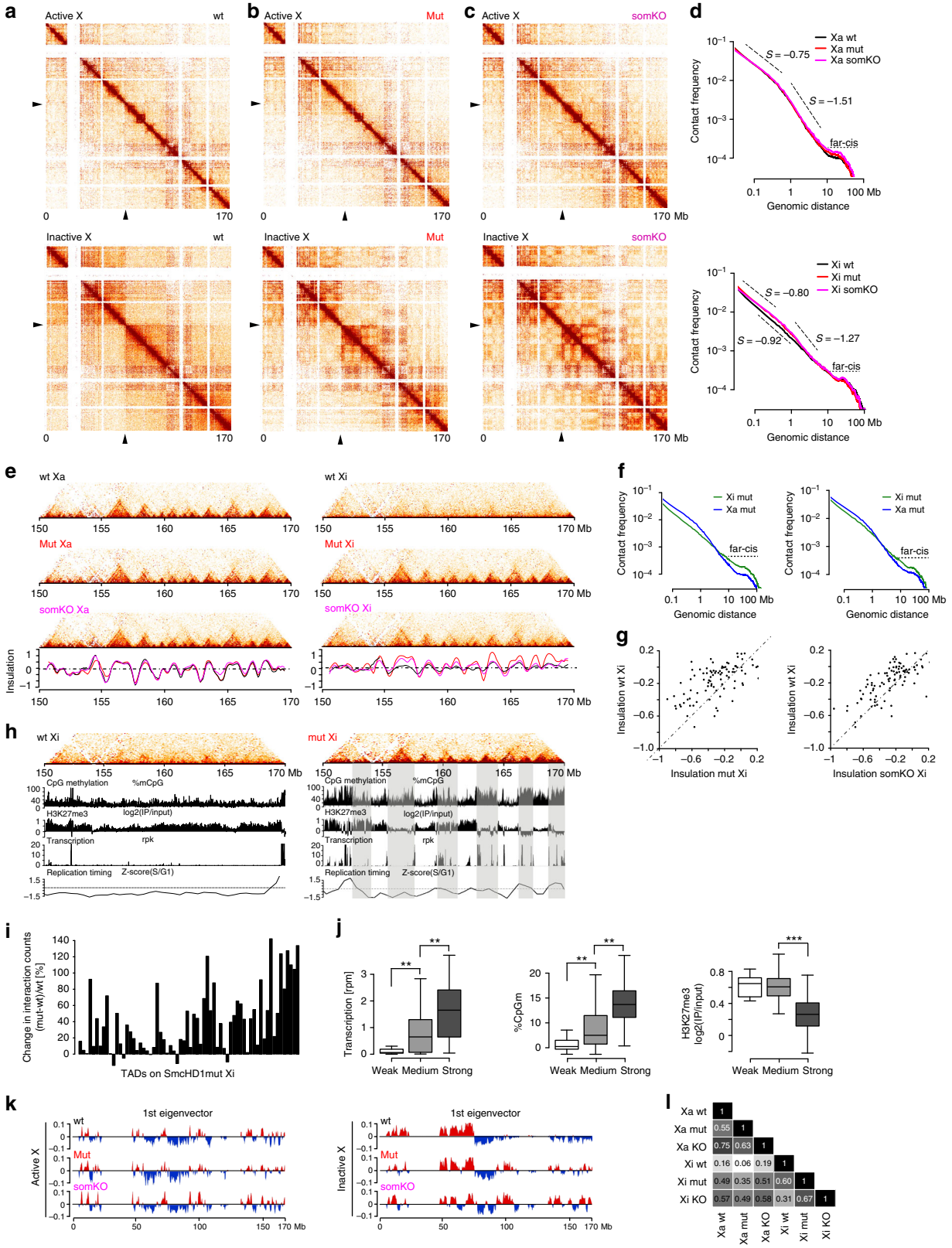


Fig. 4 Advanced Xi replication timing in SmcHD1 mutant MEFs. Allele-specific Repli-seq profiles of wt (**a**) and germline SmcHD1 mutant (mut) (**b**) MEFs. Z-score of ratio of read densities from asynchronously dividing cells in S-phase to the read densities of G1 cells, 500 kb bins. Higher Z-score value corresponds to earlier replication. Boxes represent examples of regions in which replication timing on Xi and Xa in germline SmcHD1 mutant cells are equivalent (orange), or moderately advanced/equivalent to wt Xi (yellow). Regions of low probe mappability are indicated with reduced transparency. **c** Difference in the replication timing (Z-score of G1/S read densities) shift between Xa and Xi in wt and mut MEFs within and outside H3K27me3 depletion domains



correlating with changes in Xi transcription and epigenetic landscape (Fig. 5j). In addition to TAD restoration, we observed compartmentalisation similar to Xa/autosomes, again both in germline SmcHD1 mutant and somatic knockout cell lines (Fig. 5b, c, k, l, Supplementary Fig. S7a, g). SmcHD1-dependent changes in TAD structure were reproducible in independently derived MEF lines (Supplementary Fig. 7h, i). Taken together these results demonstrate that SmcHD1 is a key factor in defining the unique higher-order chromatin domain organisation on Xi.

SmcHD1 depletes CTCF and cohesin at TAD boundaries on Xi. Recent studies have established that TAD boundaries result

Fig. 5 Altered TAD structure on Xi in SmcHD1 null MEFs. **a** Heatmap depicting allele-specific Hi-C interactions for Xa and Xi for wt MEFs. Interaction matrices were normalised for the number of available valid interactions per chromosome and KR-balanced within 250 kb bins. Black triangles indicate DXZ4 locus ("hinge"). **b**, **c** Hi-C interactions, as in **a**, for Xa and Xi in germline SmcHD1 mutant (mut) MEFs (**b**) and SmcHD1 somatic knockout (somKO) (**c**). **d** Distribution of linear genomic distances separating interacting loci for distinct cell lines for Xa (top) and Xi (bottom). Slope values are given for linear parts of the curves. Horizontal dashed line separates the far-cis interactions (>10 Mb). **e** Chromatin conformation within the distal 20 Mb of Xa and Xi for wt, mut, and somKO MEFs. Heatmaps, as above, but interactions aggregated within 50 kb bins. Insulation score plotted in black for wt, red for mut and magenta for somKO MEFs (bottom). **f** As in **d**, distribution of interaction distances for Xi and Xa in wt and mut cells. **g** Insulation at wt Xa TAD borders calculated for wt vs mut Xi (left) and wt vs somKO MEFs (right). **h** Hi-C interactions of Xi in wt (left) and mut (right) MEFs, aligned with DNA methylation, H3K27me3, transcription and replication timing profiles for the distal 20 Mb of ChrX. Shaded bars highlight TADs for mut cells. **i**. Difference in the interaction counts within TADs along the Xi between mut and wt MEFs related to the counts in wt. Interactions >100 kb were scored. **j** Boxplots summarising mean expression, CpG methylation and H3K27me3 enrichment on mut Xi for each TAD (Mann–Whitney test with Bonferroni correction). TADs were grouped based on the degree of re-establishment according to the quartiles of values in (**i**) into weak < Q1, medium > Q1 and < Q3 and strong > Q3. **k** First eigenVector values for Xi and Xa in all cell lines (500 kb bin). A/B compartments coloured in red/blue, respectively. **l**. Correlation between the first eigenVector values for Xa and Xi in all cell lines analysed (Pearson's coefficients). Boxplots in **j** presents quartiles, median and outliers

from the insulator protein CTCF restraining processive activity of the cohesin complex[11–13]. Accordingly, CTCF and cohesin occupancy is reduced at many sites on Xi compared with Xa[17,46,47]. With this in mind we performed ChIP-seq to determine the occupancy of CTCF and the cohesin subunit, Rad21, on Xa and Xi in wt compared with germline SmcHD1 mutant and SmcHD1 somatic knockout MEFs. Consistent with these prior studies, we observed that both CTCF and Rad21 occupancy is reduced on Xi compared with Xa (Fig. 6a, b). In contrast, in germline SmcHD1 mutant and somatic knockout MEFs we observed restoration of both CTCF and Rad21 occupancy at many sites (Fig. 6a–d, Supplementary Fig. 8a). Restoration of Rad21 peaks was more pronounced in germline SmcHD1 mutant MEFs. The observed effects were most apparent over chromosomal regions at which chromosome architecture changes occur in germline SmcHD1 mutant and SmcHD1 somatic knockout cells (Fig. 6e, f). At selected regions for which allelic assignment of CTCF/Rad21 enrichment sites was possible, we were able to correlate restoration of CTCF/Rad21 occupancy on Xi with the reappearance of specific TADs (Fig. 6c, d, Supplementary Fig. 8b). The majority of Xa CTCF/Rad21 binding sites that were absent on Xi in wt MEFs acquired significant levels of CTCF/cohesin occupancy in germline SmcHD1 mutant and SmcHD1 somatic knockout (Fig. 6g, h). Given that CpG methylation of CTCF sites can block binding[48], we investigated whether restoration of Xi CTCF binding could be linked to loss of DNA methylation in germline SmcHD1 mutant and SmcHD1 somatic knockout cells. As shown in Fig. 6i, we observed that loss of DNA methylation could account for restoration of CTCF binding only at a small subset of sites.

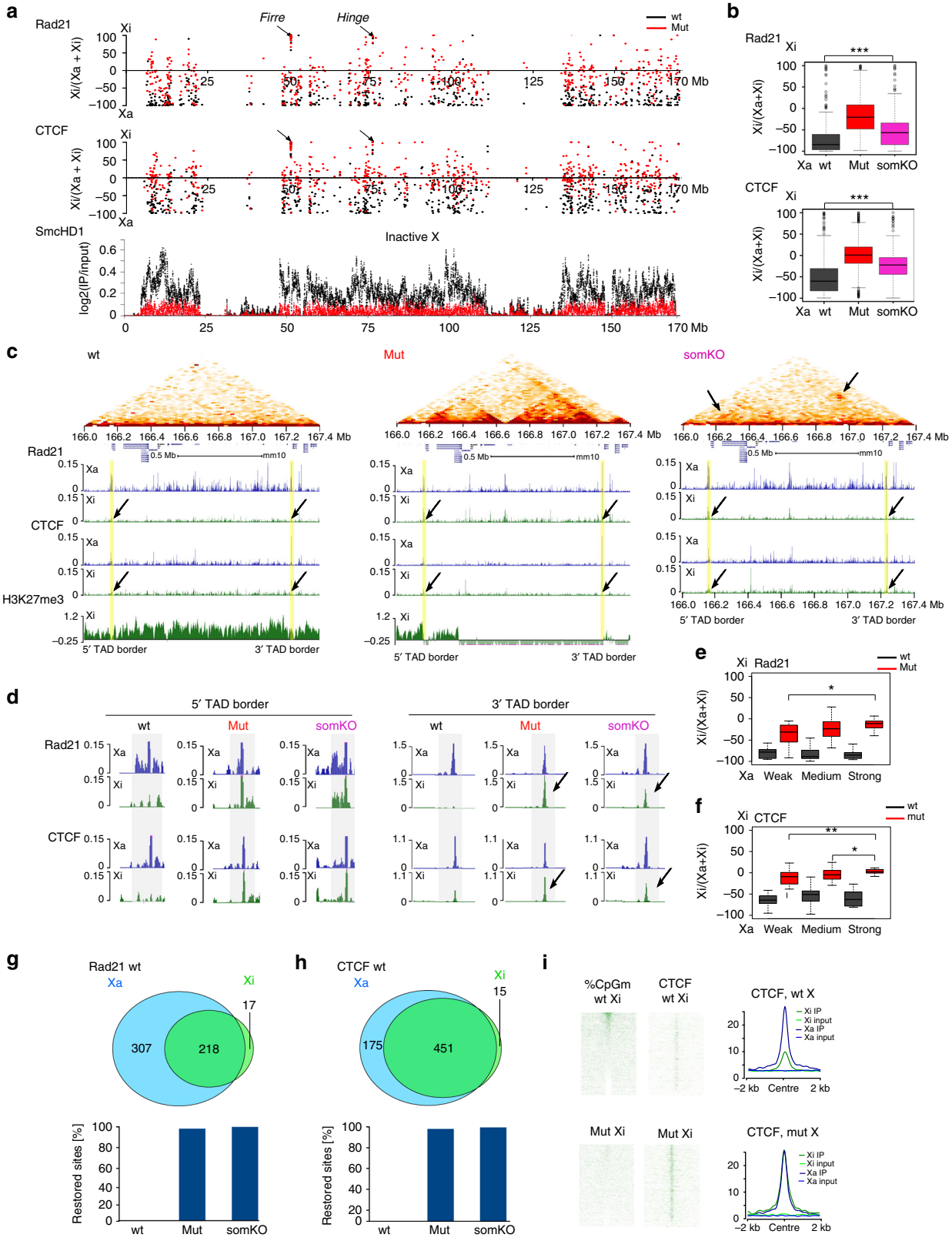
Discussion

Allelic ChIP-seq analysis of SmcHD1 on Xi shows strong enrichment over gene-dense domains. This pattern mirrors the localisation of Xist RNA and Xist-dependent histone modifications[49,50], and is therefore consistent with a previous study using a human cell line that reported a requirement for ongoing XIST RNA expression to maintain SMCHD1 enrichment on Xi[36]. Using allelic RNA-seq we found that a large proportion of Xi genes within SmcHD1 enriched regions are de-repressed in germline SmcHD1 mutant MEFs, albeit, in most cases expression does not reach the level seen on Xa. At present we cannot discriminate between the possibility that Xi gene activation in SmcHD1 mutant cells occurs in a graded fashion within individual cells, or that low-level activation reflects probabilistic expression in a subset of cells that varies for different loci.

In prior work we found that SmcHD1 enrichment on Xi is a late step in the X inactivation cascade in differentiating XX ESCs[31], suggesting a role in maintenance rather than establishment of X inactivation. Paradoxically, in this study we found that there is no activation of Xi genes following knockout of SmcHD1 in MEFs. In light of this we speculate that SmcHD1 is required for Xi to transition from the initial repressed state to long-term silencing, but that downstream silencing pathways, for example CGI DNA methylation, maintain the inactive state even in the absence of SmcHD1. Although we cannot at present define how SmcHD1 mediates transitional Xi gene silencing, its role in establishing higher-order chromosome folding, discussed below, is one possible mechanism.

Analysis of the Xi methylome at single nucleotide resolution confirms prior studies demonstrating Xi hypomethylation[39,40]. We observed widespread and extensive hypomethylation, encompassing both gene-dense and gene-poor compartments. As noted, we cannot rule out that hypomethylation of gene-poor regions genome-wide contributes to the overall hypomethylation of Xi. However, hypomethylation of gene-dense regions on Xi is presumably linked to the X inactivation process. Mechanistically, this could be attributable to reduced binding of the de novo DNA methyltransferases Dnmt3a/b, which recognise the histone modification H3K36me3 in gene bodies of active genes[51,52]. Consistent with this idea, hypermethylated regions on Xi correspond to genes that escape X inactivation and in germline SmcHD1 mutant MEFs, hypermethylation occurs within regions encompassing activated Xi genes. However, Xi CpG hypermethylation in germline SmcHD1 mutant MEFs is not restricted to gene bodies, where H3K36me3 is normally enriched, but rather occurs across domains that include upstream and downstream sequences, and in some cases more than one gene. Reciprocal depletion of PRC2-mediated H3K27me3 further reinforces the hypothesis that deregulation of the Xi epigenome in SmcHD1 mutant cells occurs at the level of sub-megabase domains that encompass one or more genes, together with intergenic/flanking sequences. These observations suggest that alternative mechanisms, for example increased accessibility to nucleosome remodelling complexes, are involved in establishing the distinct epigenomic features observed on Xi in germline SmcHD1 mutant MEFs.

Further support for SmcHD1 functioning on Xi at the level of domain organisation comes from direct analysis of replication-timing domains and TADs/compartments. In both cases we observe a shift towards Xa-specific organisation across Xi, most obviously associated with sites where Xi gene activation and modified epigenomic features occur. The mechanism for replication-timing changes is not known. One possibility is that SmcHD1 affects recruitment of Rif1, required to direct PP1 to



reverse Cdc7-mediated phosphorylation of the MCM complex[53]. Thus, at chromosomal regions that are normally early replicating, SmcHD1 may facilitate Rif1 recruitment, for example through affecting chromosome architecture. SmcHD1 has been reported to disrupt Xi replication-timing independent of gene activation following SMCHD1 knockdown in a human XX cell line[36], implying that transcriptional activation is not the cause of shifted replication-timing patterns.

The unique organisation of TADs on Xi has been linked to reduced binding of CTCF at TAD boundaries and/or reduced recruitment of cohesin complexes[17,18]. Accordingly, we observe restoration of both CTCF binding and cohesin (Rad21)

Fig. 6 CTCF and cohesin occupancy on Xi in SmcHD1 null cells. **a** Chromosome-wide occupancy of Rad21 (top) and CTCF (middle) for Xi or Xa, estimated for distinct peaks in wt, germline SmcHD1 mutant (mut) and somatic SmcHD1 knockout (somKO) MEFs, as a ratio indicating relative levels on Xa and Xi. 100 = fully Xi-, −100 = fully Xa-specific. Black arrows depict Firre and Hinge loci highly enriched on wt Xi. Lower part, shows SmcHD1 profile on the Xi log2 (IP/input) in 10 kb bins. **b** Summary of preferential occupancy of Rad21 (top) and CTCF (bottom) for Xi or Xa in all cell lines (wt, mut and somKO). Boxplots present quartiles, median and outliers of values estimated as in (**a**). **c** Example illustrating Hi-C interactions, CTCF and Rad21 occupancy within a region harbouring a TAD that is re-established in mut and to lesser extent in somKO cells. Top part shows Hi-C heatmap of chromatin interaction counts in 50 kb bins. Black arrows depict occupancy changes within the borders of the re-established TAD. Yellow shading highlights regions which differ in CTCF/Rad21 binding between mut/somKO and wt cells correlating with external borders of the re-established TAD. **d** Enlarged views of 5′ and 3′ TAD borders from **c**. **e**, **f** Boxplots summarising preferential occupancy of Rad21 (**e**) and CTCF (**f**) for Xi or Xa (values as in **a**) for mut and wild-type Xi and each TAD classified as in Fig. 5i based on their degree of re-establishment in mutant cells. **g**, **h** Venn diagram quantifying Rad21 (**g**) and CTCF (**h**) occupancy on Xi and Xa within peaks in wt, mut and somKO MEFs. Peaks were considered Xi- or Xa-specific if >90% of reads mapped to Xi or Xa, respectively. Remaining peaks were counted as present on both chromosomes. Barplots show the proportion of the Xa-specific peaks which gained Rad21 (**d**) and CTCF (**e**) in mut and somKO MEFs. **i** DNA methylation at Xi-linked CTCF sites alongside CTCF enrichment for wt and mut MEFs both ordered according to the DNA methylation level. Right part shows averaged profile of CTCF enrichment for wt and mut MEFs. **b**, **e**, **f** *0.05 < *p*-value > 10^{-2}, **10^{-2} < *p*-value > 10^{-4}, ***p-value < 10^{-4}(Mann-Whitney test)

accumulation in germline SmcHD1 mutant and SmcHD1 somatic knockout MEFs, and sites of restoration correlate well with the reappearance of Xa TAD structure. These findings accord with prior analysis of SmcHD1 at an autosomal target, the protocadherin gene cluster in neural stem cells, which suggested that SmcHD1 and CTCF have opposing roles in transcriptional regulation[37]. The antagonistic effects of SmcHD1 on TAD boundaries may be analogous to the role of condensin in TAD dissolution during mitosis[54].

We observed that whilst restoration of CTCF binding also occurs in MEFs following depletion of SmcHD1 in somatic cells (where there is no Xi gene derepression), restoration of Rad21 on Xi was of a lower magnitude. Prior work has suggested a link between cohesin loading and active transcription[55], and this may account for the reduced level of restoration relative to the germline SmcHD1 mutant model. The mechanism for restoration of CTCF binding is not known, albeit, our analysis indicates that changes in DNA methylation could account for a small proportion of restored sites. Regardless, this result together with Hi–C analysis demonstrates that changes in the long-range architecture of Xi in SmcHD1 mutant cells is not a consequence of gene derepression. We note that prior studies have reported that depletion of Xist RNA in somatic cells leads to partial restoration of Xa chromosome architecture/TAD structure, also in the absence of changes in Xi gene repression[17,56]. Given that SmcHD1 recruitment to Xi is dependent on ongoing Xist expression[36], we suggest that these effects may be attributable in part, or entirely, to loss of SmcHD1 on Xi.

During revision of this paper three related studies were published[57–59]. In close agreement with our findings these studies report that SmcHD1 is important for silencing many Xi genes[57,58], for chromosome-wide changes in the Xi epigenome[58], and for establishment of Xi-specific higher-order chromosome architecture[57,59].

In summary, our findings illustrate that SmcHD1 functions in X inactivation at the level of chromatin domain organisation. In future studies it will be important to determine the molecular interactions that underpin SmcHD1 function in antagonising TAD/compartment formation both on Xi and potentially at other SmcHD1 target loci.

Methods

Derivation and culture of mouse embryonic fibroblasts (MEFs). Animal studies were carried out under the United Kingdom Home Office ASPA project, licence numbers 30/2800 (until 2015) and 30/3326 (from 2015 to present).

Interspecific crosses between SmcHD1 mutation carrier strain (FVB) and *castaneus* strain were designed to prevent meiotic recombination between X^{cast} and X^{FVB} chromosomes (Fig.1). FVB MommeD1[32] heterozygous male was crossed with *castaneus* WT female. F1 SmcHD1 heterozygote male was selected for F2 cross with SmcHD1 heterozygous FVB female. Embryos from this cross were dissected at E9.5, and MEF lines were derived by culturing trypsinised wt and $SmcHD1^{mut/mut}$ littermate embryos. Established MEF lines were sub-cloned, and individual clones were selected for further analysis. SmcHD1 genotype was determined by sequencing PCR fragments overlapping the SmcHD1 point mutation in exon 23[32]. FVB or *castaneus* origin of the X chromosome was determined by sequencing PCR fragments across Xist SNP region. XX/XY genetic content was established by PCR analysis using primers that give distinct bands for the Uba1 and Uba1y genes[60]. All primers used for genotyping are listed below.

Fibroblasts were grown in Dulbecco's modified Eagle medium (DMEM, Life Technologies) supplemented with 10% foetal calf serum (FCS; Seralab), 2 mM L-glutamine, 1x nonessential amino acids, 50 µM 2-mercaptoethanol, and 100 U/ml penicillin/100 µg/ml streptomycin (Life Technologies) in a humidified 37 °C incubator under 5% CO_2.

CRISPR/Cas9-mediated SmcHD1 knockout in MEFs. sgRNAs targeting *SmcHD1* in the vicinity of the original MommeD1 mutation (exon 23) and neighbouring exons were cloned into plasmid pX459 as described[61]. SmcHD1 wt1A2 cell line that carries *wt SmcHD1* (two FVB alleles and one *castaneus* allele) was used for mutagenesis. Fibroblasts were plated on 90 mm Petri dishes a day before transfection. A few hours before transfection growth medium was replaced with antibiotic-free medium. Cells were transfected with a pool of two sgRNAs, 3 µg each, using Lipofectamine 3000 (Life Technologies) according to the manufacturer's instructions. DNA:Lipofectamine ratio was 1:3. Cells were trypsinised 18 h after lipofection and plated on 145 mm petri dishes at densities 1/10; 1/3 and the rest. Puromycin selection at final concentration 4 µg/ml was applied next day and maintained for 72 h, after which cells were grown in EC10 medium for 10 days until colonies were ready to be picked. Individual well-spaced colonies were scraped from the dishes and transferred into 48-well plate without trypsinisation. Initial screening of knockout clones was by genomic PCR across intronic region (*SmcHD1_TNK208* + *SmcHD1_TNK291*). Selected clones were subsequently analysed by sequencing of cloned PCR products from regions around sgRNAs and across introns when appropriate. Candidate KO clones D4 and A3.3 were subsequently further characterised by immunofluorescence with SmcHD1, H3K27me3 and H2Aub1 antibodies, and by Western blot of nuclear extracts with SmcHD1 antibody. Used oligonucleotides are listed in Supplementary Table 1.

Metaphase spreads. Cells were plated on 140 mm Petri dishes 2 days before metaphase collection. Semi-confluent cultures with actively dividing cells were fed with fresh medium supplemented with 1.5 µg/mL ethidium bromide (Life Technologies) and incubated for 1 h 20 min. Mitotic cells were arrested by addition of KaryoMAX Colcemid (Life Technologies) at a final concentration of 0.1 µg/ml for further 40 min. Cells were carefully rinsed once with PBS and trypsinised briefly at room temperature until the top layer of cells enriched with mitotic cells started to move. Trypsin was inactivated with fresh medium and top layer of cells was collected and pelleted at 200 × *g* for 3 min, RT. All supernatant was aspirated and pellets were gently resuspended in 1 ml of hypotonic solution (75 mM KCl). After incubation for 5 min at RT, swollen cells were placed on ice, and 200 µL of freshly prepared methanol/acetic acid fixative (3:1, 4 °C) was added drop-wise to the solution to pre-fix cells. Cells were pelleted at 200 × *g* for 3 min, and most of the supernatant was removed, leaving behind about 100 µL to re-suspend the cells in by gentle flicking. 1 ml of ice-cold fixative was added to the resuspended cells and incubated overnight at 4 °C without agitation. The following day the cells were carefully resuspended in the same fixative and pelleted as before. The fixative was replaced 3–5 times in total until good metaphase chromosome spreading was observed after the cell suspension was dropped onto microscope slides and air dried.

M-FISH. Metaphase spreads were prepared as described above, omitting the addition of ethidium bromide. The cells were hybridised with the 21XMouse MFISH kit (Zeiss Metasystem), following the manufacturer instructions. The slides were mounted in DAPI/Vectashield mounting medium (Oncor), under a glass coverslip, and analysed with an Olympus BX60 microscope for epifluorescence equipped with a Sensys CCD camera (Photometrics, USA). Images were collected using Genus Cytovision software (Leica). A minimum of twenty-five cells were analysed for each cell line.

Chromosome paint (DNA FISH) on metaphase spreads. Slides with freshly fixed metaphase spreads were dehydrated through ethanol series (2 × 70%, 2 × 90% 2 min each followed by 100% for 5 min), and incubated in an oven at 65 °C for 1 h. Slides were cooled down on a bench for a few minutes and denatured in 70% (v/v) Formamide/2xSSC at 65 °C for 90 s. Slides were quenched in ice-cold 70% ethanol for 4 min and then dehydraded again through the ethanol series (the same as above), and dried in the vacuum dessicator for 5 min at RT. A 1:1 mix of directly labelled chromosome 8 (Cy-3, Cambio Ltd) and X (FITC) paints was denatured at 65 °C for 10 min, spun down and incubated at 37 °C for 40–60 min. Probe was incubated with the denatured metaphases overnight at 37 °C. Next day the slides were washed twice with a solution of 1xSSC/50% formamide followed by two washes with 1xSSC and three washes with 4xSSC/0.05% Tween 20 in a water bath at 45 °C, 5 min each. Slides were mounted in Vectashield containing 4,6-diamidino-2-phenylindole (DAPI) (Vector Laboratories) and sealed with nail varnish.

RNA-FISH. Cells were plated on Superfrost Plus gelatinised slides (VWR) and grown at least overnight. After washing twice with PBS, cells were permeabilised for 5 min in CSK buffer (100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 10 mM PIPES) with 0.5% Triton X-100 (Sigma) on ice. Slides were rinsed briefly in PBS and fixed in 4% formaldehyde/PBS for 10 min on ice, followed by two washes in 70% ethanol. Slides were either stored in 70% ethanol at 4 °C until use or dehydrated (80, 95, 100% ethanol, 3 min each, RT) and air dried immediately before hybridisation with Xist probe. Xist probe was generated from an 18 kb cloned cDNA spanning the whole Xist transcript using a nick translation kit (Abbott Molecular) as previously described[19]. Directly labelled probe (1.5 µL) was co-precipitated with 10 µg salmon sperm DNA, 1/10 volume 3 M sodium acetate (pH 5.2) and 3 volumes of 100% ethanol. After washing in 75% ethanol, the pellet was dried, resuspended in 6 µL deionised formamide and denatured at 75 °C for 7 min before quenching on ice. Probe was diluted in 6 µL 2x hybridisation buffer (5xSSC, 12.5% dextran sulphate, 2.5 mg/mL BSA (NEB)), added to the denatured slides and incubated overnight at 37 °C in a humid chamber. After incubation, slides were washed three times with a solution of 2xSSC/50% formamide followed by three washes with 2xSSC in a water bath at 42 °C. Slides were mounted with Vectashield with DAPI and sealed with nail varnish.

Immunofluorescence. Cells were plated on Superfrost Plus gelatinised slides (VWR) at least a day before the experiment. On the day of the experiment, cells were washed with PBS and then fixed with 2% formaldehyde in PBS for 15 min at RT, followed by 5 min of permeabilisation in 0.4% Triton X-100. Cells were rinsed with PBS three times, 2 min each and pre-blocked with a 0.2% w/v PBS-based solution of fish gelatine (Sigma) three times, 5 min each. Primary antibody dilutions were prepared in fish gelatin solution with 5% normal goat serum. Cells were incubated with primary antibodies for 2 h in a humid chamber at room temperature, then washed three times in fish gelatin solution to remove non-bound and non-specifically bound antibodies. Secondary antibodies were diluted in fish gelatin solution and incubated with cells for 1 h at RT in a humid chamber. Primary and secondary antibody dilutions are listed in Supplementary Table 2. After incubation, slides were washed twice with fish gelatin and once with PBS before mounting with Vectashield mounting medium with DAPI. Excess mounting medium was removed and the coverslips were sealed using nail varnish.

Microscopy. Z stack images were acquired on a Zeiss AX10 microscope equipped with AxioCam MRm charge-coupled device camera using AxioVision software (Carl Zeiss International, UK). Best exposure time for each field and channel was manually determined and kept fixed among experiments. Further image editing and refinement was achieved through Fiji/ImageJ.

RNA extraction. Cells grown on T25 flasks (Nunc) were washed twice in PBS and lysed directly with 1 ml of Trizol reagent (Life Technologies). Samples were incubated for 5 min at 25 °C, cellular lysates were collected and transferred into 1.5 ml RNase-free Eppendorf tubes. 0.2 volume of chloroform was added to each sample and mixed by vigorous shaking for 15 s. Samples were incubated for 2 min at 25 °C and centrifuged at 12,000 × *g* for 5 min at 4 °C. The upper aqueous phase was transferred into a clean tube and an equal volume of isopropanol was added. Samples were incubated for 10 min, and RNA was pelleted at 12,000 × *g* for 10 min at 4 °C. The pellets were washed once in 1 ml of 75% ethanol and then air dried for 5–10 min and resuspended in 50–100 µL RNase-free water. Contaminating DNA was removed using the Ambion DNA-free DNase Treatment kit (Life Technologies) according to the manufacturer's instructions. cDNA was generated with $pd(N)_6$ random hexamers (GE Healthcare) using SuperScript III reverse transcriptase (Life Sciences) according to the manufacturer's instructions.

Genomic DNA extraction. Cells from a confluent 90–140 mm petri dish were harvested and resuspended in 5–10 ml lysis buffer (10 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM EDTA-NaOH pH 8.0, 0.5% sodium lauroyl sarcosinate) with proteinase K added to a final concentration of 100 µg/ml. Samples were incubated overnight at 55 °C, and then genomic DNA was phenol/chloroform extracted and purified using 15 ml MaXtract High Density Tubes (Qiagen). Genomic DNA was precipitated with 1/25 volume of 5 M NaCl and 2.5 volume of ice-cold 100% ethanol. High molecular weight genomic DNA was spooled and transferred to a new Eppendorf tube containing 1 ml of 70% ethanol. DNA was pelleted, air dried and resuspended in 300–400 µl 10 mM Tris pH 8.5. DNA concentration was measured by Nanodrop.

Nuclear extraction. Nuclear extracts for Western blot analysis of CRISPR/Cas9-mediated SmcHD1 mutant cell lines were prepared essentially according to a method described previously[62]. Briefly, cells were trypsinised and washed in PBS and then resuspended in 10 packed cell volumes of buffer A (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl with 0.5 mM DTT, 0.5 mM PMSF, complete protease inhibitors (Sigma) added fresh, and incubated on ice for 10 min. Cells were recovered by centrifugation at 1500 × *g* for 5 min at 4 °C. Cells were then lysed in three volumes of buffer A + 0.1% NP40 and incubated on ice for another 10 min. Nuclei were collected by centrifugation at 400 × *g*, 5 min at 4 °C and washed once in five volumes of PBS with protease inhibitors. Recovered nuclei were resuspended in one volume of buffer C (5 mM HEPES pH 7.9, 26% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 250 mM NaCl with complete protease inhibitors and 0.5 mM DTT added fresh). Salt concentration was increased to 350 mM NaCl and extraction was performed on ice for 1 h with occasional agitation. Nuclei were pelleted at 16,000 × *g* for 20 min at 4 °C and supernatants were collected as nuclear extracts. Concentration of the extracts was measured using Bradford assay (Bio-Rad) according to manufacturers' instructions. Samples were stored at −80 °C until use.

Western blotting. Samples were diluted in 6xSMASH buffer (375 mM Tris.HCl pH 6.8, 35% Glycerol, 12% SDS, 0.1% bromophenol blue, 4.3 M β-mercaptoethanol), boiled for 10 min at 95 °C, separated on 6% polyacrylamide gel and transferred onto a nitrocellulose membrane by wet transfer in Tris/glycine buffer (100 V for 70 min at 4 °C). Membranes were blocked in TBST buffer (100 mM Tris-HCl pH 7.5, 0.9% NaCl, 0.1% Tween 20, 5% w/v Marvel milk powder) for 1 h at room temperature and then incubated with primary antibodies overnight at 4 °C with gentle rocking. Membranes were washed three times for 10 min with TBST and incubated for 1 h with secondary antibody conjugated to horseradish peroxidase or IRDye 800CW anti-rabbit IgG (Li-COR). After washing three times for 10 min with TBST and one 10 min wash with PBS, bands were visualised either using ECL (GE Healthcare) or on Odyssey Fc Imaging System (Li-COR). Primary and secondary antibody dilutions are listed in Supplementary Table 2.

Whole-genome bisulfite sequencing (WGBS). Genomic DNA and RNA were extracted from MEF cultures using AllPrep DNA/RNA/Protein Mini Kit (Qiagen) according to the manufacturer's instructions. Concentration of DNA was measured on Nanodrop and adjusted to 20 ng/µl after which the DNA was sheared to 100–500 bp fragments on Covaris sonicator with the following settings: 10% duty cycle, 200 bursts/s, intensity 4.0, 80 s duration; mode–frequency sweeping). Fragment ends were repaired, A-tailed and ligated with PE Illumina methylated adapter oligos using NEBNext DNA Library Prep Master Mix Set for Illumina (NEB). Adaptor-ligated gDNA fragments were subjected to bisulfite conversion using two-step modification procedure with Imprint™ DNA Modification Kit (Sigma). Converted and purified DNA was amplified for 9–12 cycles with KAPA HiFi HS Uracil enzyme mix (Anachem). Precise concentration and size of the libraries were determined by qPCR with universal Illumina primers and also on Agilent 2100 Bioanalyzer (Agilent Technologies) using High Sensitivity DNA assay. Libraries were sequenced on HiSeq2000 system (Illumina) using 100 bp Paired End protocol.

Chromatin RNA sequencing. wt, SmcHD1 mutant and SmcHD1 somatic knockout MEFs were grown on 2 × 145 mm petri dishes in EC10 medium to semi-confluency. RNA-Seq enriching for chromatin-bound nuclear RNA was performed according to a modified chromatin RNA-Seq protocol[63]. The cells were trypsinised, collected in EC10 medium and cell number was counted on LUNAII cell counter (Logos Biosystems). 1×10^7 cells for each line were spun down, resuspended in 12 ml of ice-cold PBS and supplemented with 2.5×10^6 SG4 drosophila cells for calibration. Cells were spun down at 500 × *g*, 5 min at 4 °C and cell pellets were resuspended in 800 µl of HLBN hypotonic buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 2.5 mM $MgCl_2$, 0.05% NP40). 480 µl of buffer HLBNS (HLBN, 25% sucrose) was carefully under-layered to create sucrose cushion, and nuclei were isolated by centrifugation for 5 min at 1000 × *g* at 4 °C. Supernatant containing cytoplasmic debris was discarded and the nuclear pellet was resuspended in 100 µl of ice-cold buffer NUN1 (20 mM Tris-HCl pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 50%

glycerol; 1 mM DTT and cOmplete EDTA free protease inhibitors (Sigma) added fresh). Nuclei were lysed in 1200 µl of ice-cold lysis buffer NUN2 (20 mM HEPES pH 7.6, 300 mM NaCl, 7.5 mM $MgCl_2$, 0.2 mM EDTA, 1 M urea, 1% NP40; 1 mM DTT) during 15 min incubation on ice and RNA-bound chromatin was pelleted at 16,000 × *g* for 10 min at 4 °C. Chromatin-RNA pellet was resuspended in 200 µl of high salt buffer HSB (10 mM Tris-HCl pH 7.5, 500 mM NaCl, 10 mM $MgCl_2$). DNA and proteins were digested with Turbo DNAse (Life Sciences) and proteinase K (10 mg/ml, ThermoFisher, nuclease free), incubating on ThermoMixer at 37 °C for 10 min and 30 min, respectively. RNA was extracted with 1 ml of TRIzol (Life Sciences) according to the manufacturer guidelines. RNA was dissolved in 1xTURBO DNAse buffer, digested with TURBO DNAse for 30 min at 37 °C on a ThermoMixer and extracted with TRIzol. RNA was washed three times with 75% ethanol, dissolved in water and quantified using a NanoDrop (ND-1000).

RNA quality was checked with RNA 6000 Pico Chip (Agilent Technologies) on Agilent 2100 Bioanalyser (Agilent Technologies). Samples were depleted of ribosomal RNA with Ribo-Zero Gold Kit (MRZG12324, Illumina) according to manufacturer's guidelines. RNA was isolated from 3 biological replicates. RNA-Seq libraries were prepared with NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina (E7420S) using NEBNext® Multiplex Oligos for Illumina for multiplexing (E7335S and E7500S). Libraries were sequenced on HiSeq2000 and NextSeq500 using NextSeq 500 High-Output Kit: 1 lane, 150 cycles, 75 bp paired-end sequencing (Illumina).

Chromatin immunoprecipitation sequencing. wt, SmcHD1 mutant and SmcHD1 somatic knockout MEFs were grown on 5–8 × 145 mm petri dishes in EC10 medium to semi-confluency. Cells were trypsinised, washed in PBS and counted. For SmcHD1 and Rad21, 5×10^7 cells were then cross-linked in 2 µM Ethylene glycol-bis(succinic acid N-hydroxysuccinimide ester) (EGS, Sigma) in PBS at RT for 1 h followed by 15 min cross-linking in 1% formaldehyde. For H3K27me3 and CTCF 1×10^7 cells were cross-linked in 1% formaldehyde alone for 15 min at RT. Formaldehyde was quenched by addition of glycine to a final concentration of 125 mM, and incubation at RT for 3 min. Cells were spun down at 700 × *g* for 4 min at 4 °C and lysed in 50 mM HEPES pH 7.9, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.25% Triton X-100 and 2% NP40. Cellular lysis was assisted by 20 strokes of a large clearance pestle in Dounce grinder, with subsequent incubation for 10 min at 4 °C with constant rotation. Released nuclei were spun down and washed once in 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA. Nuclei were lysed by addition of 0.1% sodium deoxycholate and 0.5% N-lauroylsarcosine. All solutions contained cOmplete EDTA-free protease inhibitors (Sigma) added fresh. Chromatin was sonicated on BioRuptor sonicator (Diagenode) to produce fragments of ~500 bp. Triton X-100 was added to a final concentration of 1%, and insoluble pellet was removed by centrifugation at maximum speed for 10 min at 4 °C. Chromatin was aliquoted and stored at −80°C until use.

Immunoprecipitation was performed overnight at 4 °C with 3–5 µg of specific antibody and 100 µl of chromatin corresponding to 1×10^5 (H3K27me3, CTCF) or 5×10^5 cells (Rad21, SmcHD1) in 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 150 mM NaCl and 1% Triton X-100, with proteinase inhibitors. Antibody-bound chromatin was isolated on rProtein A Sepharose Fast Flow beads (GE Healthcare) that had been blocked for 1 h at 4 °C with 1 mg/ml bovine serum albumin (NEB) and 1 mg/ml yeast tRNA (Sigma). Agarose beads with immunoprecipitated material were washed thoroughly in low salt buffer (LSB, 20 mM Tri-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% Triton X-100), high salt buffer (HSB, the same as LSB but with 500 mM NaCl), LiCl buffer (10MM Tris-HCl pH 8.0, 1 mM EDTA, 0.25 M LiCl, 1% NP40, 1% deoxycholate) and twice in TE buffer, all with proteinase inhibitors. Immunoprecipitated material was eluted from beads in elution buffer (0.1 M $NaHCO_3$, 1% SDS) with shaking on Thermomixer at RT for 30 min. Beads were removed by centrifugation, eluted chromatin was reverse cross-linked and treated with RNAse and proteinase K in the presence of 200 mM of NaCl at 37 °C for 2 h followed by 65 °C overnight with shaking at 800 rpm for 1 min every 2 min. DNA was purified using ChIP DNA Clean and Concentrator kit (Zymo Research) according to the manufacturer's instructions.

DNA size was assessed on Bioanalyzer (Agilent) and DNA was post-sonicated on Bioruptor Pico sonicator for 18–20 cycles of 30 s on/30 s off if necessary. DNA Concentration was quantified using PicoGreen® dsDNA Quantitation Kit (Molecular Probes), Bioanalyzer High sensitivity DNA assay and/or Qubit dsDNA HS assay kit.

NEBNext (H3K27me3) or NEBNext Ultra II (CTCF, Rad21, SmcHD1) DNA Library Prep Kits for Illumina were used to prepare libraries for sequencing, following the manufacturer's instructions. End-repaired, A-tailed and adapter-ligated libraries were amplified for 7–11 cycles depending on the initial DNA amount. Indexed libraries were quantified, normalised and pooled for sequencing either on Illumina HiSeq2000 50 bp paired end run (H3K27me3, eight samples on three lanes) or on Illumina NextSeq 550 System (CTCF, Rad21, SmcHD1, 8–12 libraries per flowcell).

Repli-seq. Repli-seq was performed as described[64]. Briefly, asynchronously proliferating cells were flow-sorted based on their DNA content (Hoechst 33342) into G1- and S-phase fractions (Fig. S6) into ice-cold PBS. Genomic DNA was isolated immediately after flow-sorting with Quick-gDNA™ MiniPrep (Zymo Research). Sequencing libraries were prepared with Illumina TruSeq DNA PCR-Free whole genome sequencing kit and sequenced on NextSeq 550 with 150 cycles High Output kit using paired end protocol.

Hi–C. Hi–C was performed as described[65]. For each Hi–C library 25×10^6 cells were cultured in EC10 media. The day of harvesting, cells were incubated in 22.5 mL of fresh media without serum, and cross-linked by adding 625 µL of 37% formaldehyde (1% final concentration). Plates were immediately mixed thoroughly after formaldehyde addition, and subsequently rocked every 2 min for exactly 10 min at RT. The cross-linking was quenched by addition of 1.25 mL of 2.5 M glycine. After 5 min at RT, plates were placed on ice for an additional 15 min. Cells were finally harvested by a cell lifter, transferred to a 15 mL falcon tubes, and centrifuged at 800 × *g* for 10 min (4 °C). The cell pellet was snap-frozen in liquid nitrogen, and stored at −80 °C until use. Hi–C libraries were then sequenced on a HiSeq4000 (paired-end reads of 100 bases each). Two biological replicates of wt and SmcHD1 mutant MEFs were sequenced in separate lanes, each yielding ~350 million reads per replicate.

Allele-specific alignment. All NGS data were obtained from interspecific FVB-CAST/EiJ cells, which enabled allele-specific analysis. Assigning of reads into one of the parental genomes was performed in two stages.

First reads were mapped to mm10 reference genome with the aligner optimal for the assay. To minimise mapping biases due to differences in the similarity of FVB/Cast genomes to the reference genome all known SNP loci were N-masked before generating appropriate genome index files. SNP coordinates were obtained from The Sanger Institute (ftp://ftp-mouse.sanger.ac.uk/REL-1505-SNPs_Indels/).

Second using SNPsplit programme[66] and confident SNPs extracted from the file mentioned above reads were sorted into paternal, maternal and unassigned subsets. The procedure in our experimental set-up allowed us to assign ~30% reads to parental genomes.

Chromatin RNA-seq analysis. Reads were mapped with STAR 2.5b aligner[67] with index generated from FVB-Cast SNP N-masked genome. Counts per gene were obtained with Htseq-count[68].

Initial analysis showed that there is high correlation between chromosome copy number and gene expression values on distinct parental chromosomes therefore HTSeq counts were normalised for chromosomal copy number differences estimated from comparison of ChIP-seq input values and repli-seq alignments.

Differential expression analysis was performed with DESeq2[69], a package from the R Bioconductor project. Results were further processed, analysed and visualized with custom R scripts using commonly used base and Bioconductor packages.

ChIP-seq analysis. ChIP-seq reads were mapped with bowtie2 with SNP N-masked genome index and sorted with SNP-split to separate subsets of reads originating from parental genomes. For SmcHD1, CTCF and Rad21 ChIP-seqs, macs2 was used to detect significant enrichments (narrow peaks, q-value < 0.05). Peak calling was performed on unsorted reads. Allele-specificity of distinct peaks was determined by calculating the ratio of allele-specific reads overlapping the peak in maternal and paternal genome. Peak were considered exclusively specific to either paternal or maternal chromosome if they contained more than 90% of all input corrected reads.

Chromosome-wide SmcHD1 and H3K27me3 ChIP-seq data were obtained by averaging log2(IP/input) values calculated for 500 kb bins within 10 kb intervals.

H3K27me3 depletion domains were defined as regions with intervals of mean log2(IP/input) < 0, adjacent regions with negative values separated by no more than 20 kb were merged.

Hi–C analysis. Hi–C reads were mapped with HiCUP pipeline[70] using bowtie2 indexes based on the FVB-Cast SNP N-masked mm10 genome. Valid sets of Hi–C ditags were obtained after removing uninformative reads (re-ligations, dangling ends, etc), as well as exact duplicates. Hi–C ditags were then sorted based on confident SNP content with SNPsplit into following groups: FVB-FVB, FVB-unassigned, Cast-Cast, Cast-unassigned, FVB-Cast, unassigned-unassigned. For allele-specific analysis subsets "FVB-FVB" and "FVB-unassigned" were merged into one "FVB" and similarly subsets "Cast-Cast" and "Cast-unassigned" were merged into subset "Cast". For comparative analysis of the wt and mutant cells, Hi–C data interaction number were down-sampled to the size of the smallest sample and further equalized per chromosome to normalize for differences in DNA copy number. Interactions were further binned into 1 Mb intervals genome-wide and 100 and 50 kb bins chromosome-wide with juicer tools[71]. Matrices with interactions for distinct chromosomes were balanced separately with the Knight-Ruiz balancing algorithm implemented in juicer tools, that ensures that each row and column of the contact matrix sums to the same value. TAD borders were called with TADtool[72]. Insulation scores were calculated with the matrix2insulation.pl script developed in Dekker Lab based on 50 kb balanced matrices with the following options: -is 500000 -ids 200000 -im imean). Eigenvectors for compartment analysis were obtained with matrix2compartment.pl script based on interaction matrices binned into 250 kb using standard settings apart from changing -ca

parameter to 0.005. Both matrix2insulation.pl and matrix2compartment.pl scripts are available from https://github.com/dekkerlab/cworld-dekker.

Heatmaps and sums of interaction numbers for specific parts of chromosomes were generated based on balanced Hi–C matrices with HiTC Bioconductor package[73].

Whole-genome bisulfite sequencing analysis. Whole-genome bisulfite libraries were mapped with Bismark[74], and deduplicated. Further reads were allele-specifically sorted and processed with bismark_methylation_extractor to retrieve methylated cytosines. Cytosines covered by at least three reads in allele-specific alignment were used for further analysis.

Repliseq analysis. Repli-seq reads were mapped with bowtie2. Obtained alignments were sorted into parental genome specific subsets and subsequent analysis were performed as described[60]. In brief, reads from S and G1 cell fractions were analysed in 100 and 500 kb bins and for each bin read number normalized S/G1 ratio was plotted. Z-scores of S/G1 ratios were further used for detailed analysis and visualization of chromosome-wide replication timing profiles.

Code availability. Code used for the bioinformatic analysis is available from the corresponding author upon request.

Data availability

The accession number for all high-throughput sequencing data reported in this study were deposited in GEO under GSE115984. All other data are available from the corresponding author upon reasonable request. A reporting summary for this Article is available as a Supplementary Information file.

Received: 25 May 2018 Accepted: 14 November 2018
Published online: 03 January 2019

References

1. Lyon, M. F. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**, 372–373 (1961).
2. Gendrel, A. V. & Heard, E. Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. *Annu. Rev. Cell. Dev. Biol.* **30**, 561–580 (2014).
3. Brown, C. J. et al. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527–542 (1992).
4. Brockdorff, N. et al. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515–526 (1992).
5. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137 (1996).
6. Lee, J. T. & Jaenisch, R. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. *Nature* **386**, 275–279 (1997).
7. Aladjem, M. I. & Fu, H. A new light on DNA replication from the inactive X chromosome. *Bioessays* **36**, 591–597 (2014).
8. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
9. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
10. Guo, Y. et al. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
11. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 e22 (2017).
12. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 e24 (2017).
13. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
14. Giorgetti, L. et al. Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575–579 (2016).
15. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
16. Deng, X. et al. Bipartite structure of the inactive mouse X chromosome. *Genome Biol.* **16**, 152 (2015).
17. Minajigi, A. et al. Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276 (2015).
18. Darrow, E. M. et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl Acad. Sci. USA* **113**, E4504–E4512 (2016).
19. Moindrot, B. et al. A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. *Cell Rep.* **12**, 562–572 (2015).
20. Chu, C. et al. Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
21. Monfort, A. et al. Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. *Cell Rep.* **12**, 554–561 (2015).
22. McHugh, C. A. et al. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
23. Silva, J. et al. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev. Cell* **4**, 481–495 (2003).
24. Plath, K. et al. Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131–135 (2003).
25. de Napoles, M. et al. Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell.* **7**, 663–676 (2004).
26. Almeida, M. et al. PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* **356**, 1081–1084 (2017).
27. Pintacuda, G. et al. hnRNPK recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to establish polycomb-mediated chromosomal silencing. *Mol. Cell* **68**, 955–969 e10 (2017).
28. Chen, C. K. et al. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* **354**, 468–472 (2016).
29. Patil, D. P. et al. m^6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).
30. Mermoud, J. E., Costanzi, C., Pehrson, J. R. & Brockdorff, N. Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. *J. Cell. Biol.* **147**, 1399–1408 (1999).
31. Gendrel, A. V. et al. SmcHD1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev. Cell.* **23**, 265–279 (2012).
32. Blewitt, M. E. et al. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* **40**, 663–669 (2008).
33. Brideau, N. J. et al. Independent mechanisms target SMCHD1 to trimethylated histone H3 lysine 9-modified chromatin and the inactive X chromosome. *Mol. Cell. Biol.* **35**, 4053–4068 (2015).
34. Gendrel, A. V. et al. Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. *Mol. Cell. Biol.* **33**, 3150–3165 (2013).
35. Mould, A. W. et al. SmcHD1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. *Epigenetics Chromatin* **6**, 19 (2013).
36. Nozawa, R. S. et al. Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nat. Struct. Mol. Biol.* **20**, 566–573 (2013).
37. Chen, K. et al. Genome-wide binding and mechanistic analyses of SmcHD1-mediated epigenetic regulation. *Proc. Natl Acad. Sci. USA* **112**, E3535–E3544 (2015).
38. Maza, I. et al. Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. *Nat. Biotechnol.* **33**, 769–774 (2015).
39. Hellman, A. & Chess, A. Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141–1143 (2007).
40. Weber, M. et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* **37**, 853–862 (2005).
41. Marks, H. et al. High-resolution analysis of epigenetic changes associated with X inactivation. *Genome Res.* **19**, 1361–1373 (2009).
42. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
43. Guelen, L. et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
44. Koren, A. & McCarroll, S. A. Random replication of the inactive X chromosome. *Genome Res.* **24**, 64–69 (2014).
45. Pope, B. D. et al. Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402–405 (2014).
46. Berletch, J. B. et al. Escape from X inactivation varies in mouse tissues. *PLoS. Genet.* **11**, e1005079 (2015).
47. Calabrese, J. M. et al. Site-specific silencing of regulatory elements as a mechanism of X inactivation. *Cell* **151**, 951–963 (2012).

48. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482–485 (2000).
49. Mak, W. et al. Mitotically stable association of polycomb group proteins eed and enx1 with the inactive x chromosome in trophoblast stem cells. *Curr. Biol.* **12**, 1016–1020 (2002).
50. Duthie, S. M. et al. Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. *Hum. Mol. Genet.* **8**, 195–204 (1999).
51. Dhayalan, A. et al. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* **285**, 26114–26120 (2010).
52. Baubec, T. et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243–247 (2015).
53. Hiraga, S. et al. Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex. *Genes Dev.* **28**, 372–383 (2014).
54. Gibcus, J. H. et al. A pathway for mitotic chromosome formation. *Science* **359**, eaao6135 (2018).
55. Kagey, M. H. et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
56. Splinter, E. et al. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev.* **25**, 1371–1383 (2011).
57. Wang, C. Y., Jegu, T., Chu, H. P., Oh, H. J. & Lee, J. T. SMCHD1 merges chromosome compartments and assists formation of super-structures on the inactive X. *Cell* **174**, 406–421 (2018).
58. Sakakibara, Y. et al. Role of SmcHD1 in establishment of epigenetic states required for the maintenance of the X-inactivated state in mice. *Development* **145**, dev166462 (2018).
59. Jansz, N. et al. Smchd1 regulates long-range chromatin interactions on the inactive X chromosome and at Hox clusters. *Nat. Struct. Mol. Biol.* **25**, 766–777 (2018).
60. Chuma, S. & Nakatsuji, N. Autonomous transition into meiosis of mouse fetal germ cells in vitro and its inhibition by gp130-mediated signaling. *Dev. Biol.* **229**, 468–479 (2001).
61. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
62. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
63. Nojima, T. et al. Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161**, 526–540 (2015).
64. Koren, A. et al. Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am. J. Hum. Genet.* **91**, 1033–1040 (2012).
65. Belton, J. M. et al. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
66. Krueger, F. & Andrews, S. R. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Res.* **5**, 1479 (2016).
67. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
68. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
69. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
70. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. . *F1000Res* **4**, 1310 (2015).
71. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
72. Kruse, K., Hug, C. B., Hernandez-Rodriguez, B. & Vaquerizas, J. M. TADtool: visual parameter identification for TAD-calling algorithms. *Bioinformatics* **32**, 3190–3192 (2016).
73. Servant, N. et al. HiTC: exploration of high-throughput 'C' experiments. *Bioinformatics* **28**, 2843–2844 (2012).
74. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).

Acknowledgements

We thank members of the Brockdorff and Klose labs for helpful discussions and advice. We would like to thank Vladimir Benes and EMBL Gene Core for ChrRNA-seq and WGBS, Amanda Williams from Oxford University Zoology Dept. sequencing, the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z and MRC Hub grant G0900747 91070) for the generation of the H3K27me3 Sequencing data, Miguel Branco for advice on whole genome bisulfite sequencing, and Michal Maj, Dunn School Flow Cytometry Facility for assistance in flow-sorting for Repli-seq. N.B. is funded by Wellcome (103768) and the European Research Council (340081). J.D. is an investigator of the Howard Hughes Medical Institute and was funded in part by a grant from the National Human Genome Research Institute (HG003143). S.K. is funded by Ludwig Cancer Research and BBSRC grant BB/M001873/1.

Author contributions

N.B, J.D., E.H., M.R.G. and T.B.N. conceived this study. T.B.N. and M.R.G. performed most of the experimental work and data analysis. G.P., Y.Z. and H.O. contributed to Hi–C analysis. J.G. contributed to mouse genetics and cell line derivation. S.K., W.R., F.K. and M.M. contributed to DNA methylation analysis, C.M.G. and D.M. performed karyotype analysis. N.B., T.B.N. and M.R.G. prepared the manuscript with input from all other authors.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-07907-2.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019