Long-range chromatin interactions on the inactive X and at *Hox* clusters are regulated by the SMC protein Smchd1.

Natasha Jansz[1,2], Andrew Keniry[1,2], Marie Trussart[1,2], Heidi Bildsoe[3,+], Tamara Beck[1], Ian D. Tonks[4], Arne W. Mould[4,^], Peter Hickey[5], Kelsey Breslin[1], Megan Iminitoff[1], Matthew E. Ritchie[1,2,6], Edwina McGlinn[3], Graham F. Kay[4], James M. Murphy[1,2], Marnie E. Blewitt[1,2,7,*]

1. The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia
2. The Department of Medical Biology, The University of Melbourne, Parkville, Victoria 3052, Australia
3. EMBL Australia, Australian Regenerative Medicine Institute, Monash University, Clayton, Victoria 3800, Australia
4. The Queensland Institute of Medical Research Berghofer, Herston, Queensland 4006 Australia
5. Department of Biostatistics, Johns Hopkins University, Baltimore Maryland, USA
6. School of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria 3010, Australia
7. School of Biosciences, The University of Melbourne, Parkville, Victoria 3052

^ Current address, Mammalian Development, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK.

+ Current address, Department of Molecular and Translational Science, Centre for Genetic Diseases, Hudson Institute of Medical Research, Monash University, Clayton, Victoria 3168 Australia

* To whom correspondence should be addressed, blewitt@wehi.edu.au

Abstract

The regulation of higher order chromatin structure is complex and dynamic, however we do not yet understand the full suite of mechanisms governing architecture. Here we reveal the non-canonical SMC protein Smchd1 as a novel regulator of long-range chromatin interactions in mice, and add it to the canon of epigenetic proteins required for *Hox* gene regulation. The effect of losing Smchd1-dependent chromatin interactions has varying outcomes dependent on chromatin context. At autosomal targets transcriptionally sensitive to Smchd1 deletion, we find increased short-range interactions and ectopic enhancer activation. By contrast, the inactive X chromosome is transcriptionally refractive to Smchd1 ablation, despite chromosome-wide increases in short-range interactions. There we observe spreading of H3K27me3 domains into regions not normally decorated by this mark. Together these data suggest Smchd1 has the capacity to insulate the chromatin, thereby limiting access to other chromatin modifying proteins.

In recent years chromatin conformation capture and sequencing (Hi-C) techniques have dramatically increased our understanding of how DNA forms long-range interactions[1-6]. Hi-C first revealed the presence of topologically associated domains (TADs): Mb-sized self-interacting regions[7]. With increasing resolution, Hi-C also enabled study of chromosome looping. However, our knowledge of these genomic structures has far preceded an understanding of mechanism and function. Models have been developed to explain the mechanisms of structure formation and maintaineance. The loop extrusion model has greatest traction, where DNA extrudes through a ring formed by the SMC complex, Cohesin, until the complex is stabilised by convergent Ctcf binding at TAD boundaries. As such, Ctcf and Cohesin have been the focus of much research in the field. However, genome-wide loss of Ctcf and Cohesin mediated structures often have a surprisingly limited effect on transcription[8,9]. A host of epigenetic regulators also contribute to higher order chromatin structures that modulate transcription. For example, polycomb repressive complex 1 (PRC1) mediateslong-range repressive chromatin interactions distinct from TADs, resulting in compaction of the locus[10,11]. Continued identification of epigenetic regulators mediating long-range chromatin interactions will help to resolve the function of DNA structures defined using Hi-C and refine the principles of genome organisation.

Structural maintenance of chromosomes hinge domain containing 1 (Smchd1) is a non-canonical SMC protein[12], that similar to the canonical SMC proteins[13-15], dimerises and interacts with nucleic acids through its hinge domain[16,17]. Smchd1 differs to the canonical SMC proteins as its N-terminal ATPase domain belongs to the GHKL family rather than the ABC-type[18,19], and has no identified obligate binding partners. The role of Smchd1 in regulating chromatin organisation has not been studied, though work from our group has suggested Smchd1 may functionally oppose Ctcf[17].

Smchd1 plays a role in epigenetic repression throughout the genome. We identified Smchd1 via its role in repeat-induced silencing in mice[12,20]. It is also important for silencing the *D4Z4* macrosatellite repeat in humans[21]. Smchd1 has a critical role in silencing the inactive X (Xi) chromosome in females[12,22], and in transcriptional repression of specific autosomal loci including select imprinted genes and the clustered protocadherins[17,23-25]. Each of these sites are direct targets of Smchd1[12,17,26], but interestingly so are other autosomal loci including the *Hox* genes[17], where Smchd1's role has not been explored.

Several epigenetic alterations are seen at Smchd1 targets in *Smchd1* null embryos. *Smchd1* null females die around embryonic day (E) 10.5, with failed X chromosome inactivation (XCI)[12,22,23,25]: the early stages of XCI proceed normally, but by mid-gestation failed XCI is accompanied by DNA hypomethylation of a subset of CpG islands (CGI) on the Xi[12,22]. Depleted SMCHD1 is also associated with hypomethylation at its autosomal targets[17,21,25,27]. These findings place Smchd1 upstream of DNA methylation but they have not revealed how Smchd1 enables transcriptional repression.

At all of Smchd1's targets, chromosome conformation at different scales plays a role in regulating gene expression. For example, on the Xi, two mega-domains much larger than TADs exist, which hinge on the clustered macrosatellite repeat *DXZ4*. In contrast, the active X (Xa) has a TAD structure similar to autosomes[5,28-30]. At the clustered protocadherins, short range enhancer-promoter chromatin looping occurs to regulate gene activation[31]. We previously reported that Smchd1 represses the protocadherins by binding to promoters and the enhancer at the cluster. Based on this we proposed a model where Smchd1 mediates chromosome conformation[17], however we did not analyse conformation, or the Xi because we used male cells. Nozawa *et al*. also put forward a

model, specifically for SMCHD1 regulating the chromosome structure of the human Xi. They found SMCHD1 knockdown caused decompaction of the Xi in 293 cells, implicating SMCHD1 in the maintenance of a condensed Xi in human cells[26].

Here we tested whether Smchd1 has a role in maintaining chromosome structure using *in situ* chromosome conformation capture followed by sequencing (*in situ* Hi-C). We report genome-wide changes to long-range chromatin interactions upon deletion of *Smchd1* in female cells, predominantly at Smchd1 targets. Disruption of long-range interactions at the *HoxB* locus prompted us to investigate the role of Smchd1 in regulating the *Hox* clusters, thereby implicating Smchd1 in the developmental regulation of *Hox* gene silencing. Remarkably, the widespread increase in X-linked chromatin interactions in the absence of Smchd1 did not result in the reactivation of genes from the Xi. Rather, we see increased H3K27me3 on the Xi that is not associated with a loss of DNA methylation, suggesting H3K27me3 enrichment is not due to enhanced recruitment of PRC2 to unmethylated CpG-rich regions[32-34]. These data suggest Smchd1 limits the access of chromatin modifying proteins. Furthermore, genome-wide profiling of chromatin modifications and DNA accessibility suggests that Smchd1 plays a role in limiting long-range chromatin interactions that are permissive for transcription.

Results:

Smchd1 is required to mainain higher-order chromatin organisation at target loci

To test whether loss of Smchd1 alters long-range chromatin interactions at its targets, we performed *in situ* Hi-C in primary neural stem cells (NSCs) derived from the forebrains of E14.5

female *Smchd1*$^{fl/fl}$ embryos, and in paired cells one week post Cre-mediated deletion of *Smchd1* (Fig. 1a). Conditional deletion of *Smchd1*[35] circumvented female-specific embryonic lethality [12] and enabled us to address whether Smchd1 is involved in maintaining X-linked chromatin interactions in female cells. The three biological replicates for each genotype were highly correlated and clustered by genotype, suggesting higher order chromatin conformation is altered upon deletion of *Smchd1* (Supplementary Fig. 1).

We used the bioinformatics package diffHic to analyse differential chromatin interactions between *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ female NSCs. DiffHic uses edgeR statistics to perform a differential analysis of chimeric read counts that map to different genomic bins between experimental conditions[36]. We performed analyses at two resolutions. Analysis at 1Mb afforded us greater power to assess regions with fewer read counts; namely very long-range interactions, as Hi-C interactions are known to decay as genomic distance increases[37,38] (Supplementary Fig. 1). Analysis at 100kb resolution allowed us to assess short range interactions between regions less than 1Mb apart (Supplementary Fig. 2).

At 1Mb resolution we detected 2382 significant differential interactions (DI) genome-wide (Fig. 1b). We also detected these differential interactions at 500kb resolution (87% overlap, Supplementary Data Set 2). At 100kb resolution, we detected 643 significant differential interactions between Smchd1 replete and deleted cells (Fig. 1b). Many previously described Smchd1 targets were found in the differential interactions from analyses at both resolutions, including the *Snrpn* imprinted cluster and three of the four *Hox* clusters (*HoxA*, *B* and *C*, Supplementary Data Set 2). As expected, X-linked differential interactions were over-represented (Fig. 1c, d).

We next tested whether the chromatin conformation alterations post-*Smchd1* deletion represent a disruption to the partitioning of the genome into A and B compartments, and TADs. We found 90% of A and B compartments are maintained upon Smchd1 deletion (Supplementary Fig. 2). Our Hi-C heatmaps and genome-wide analysis indicate that TAD boundaries are unchanged in the absence of Smchd1 (Fig. 1g), though the strength of some TAD boundaries changed (Supplementary Data Set 3). These data suggest Smchd1 is not required for maintenance of compartments or TAD boundaries, as expected given Smchd1 occupancy is limited to a small set of autosomal loci.

We analysed differential interactions in relation to Smchd1 occupancy in NSCs as determined by chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq). We performed Smchd1 ChIP under two distinct conditions using our new *Smchd1*^{GFP} knock-in mice, which express a functional Smchd1-GFP fusion protein: the first condition enabled profiling of Smchd1 occupancy in accessible DNA by enzymatic fragmentation using MNase; the second profiled more inaccessible chromatin, including the Xi, using long sonication in non-ionic detergent (Supplementary Fig. 3). Both methods resolved binding sites concordant with previous studies (Supplementary Data Set 4)[17,26]. We compared differential interactions with Smchd1 occupancy, and found that the two different resolutions revealed something of biological interest. At 1Mb resolution 88.1% DIs are anchored in a region that contains a Smchd1 binding site, though this is only true for 31.4% DIs at 100kb resolution (Fig. 1f). Intriguingly, while we detected long-range interactions that were both strengthened (43%) and weakened (57%) upon Smchd1 deletion at 1Mb resolution, at 100 kb resolution the vast majority of differential interactions were strengthened (Fig. 1g, Supplementary Data Set 2). Similarly, analysis using the bioinformatics tool HiCCUPS[39], which uses a local background model to identify short-range looping interactions at

10 kb resolution, also revealed increased short-range interactions upon *Smchd1* deletion (Supplementary Data Set 5). To better understand the nature of these differential interactions, we compared changes in higher order chromatin structure to transcription.

Smchd1 is required to maintain gene silencing at a subset of autosomal targets

Smchd1 has an important role in gene silencing during development, but it is unknown whether it is required to maintain gene silencing. To test whether changes in genome organisation upon acute loss of Smchd1 were associated with changes in gene expression we performed RNA-seq on samples taken at the same stage as Hi-C. This revealed 463 significantly differentially-expressed genes (DEG) between *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ female NSCs (Fig. 1h, Supplementary Data Set 6), including known Smchd1 targets such as clustered protocadherin genes and *Snrpn* cluster imprinted gene *Ndn*. Gene set enrichment analysis confirmed concordant gene expression between *Smchd1*$^{del/del}$ and male *Smchd1* null (*Smchd1*$^{MommeD1/MommeD1}$) NSCs (p=0.0015, Fig. 1i). Interestingly 69% of DEG fell within a differentially interacting region detected with diffHic (p=0.00000164, hypergeometric probability), including the clustered protocadherin genes, which display both increased expression and strengthened interactions post-*Smchd1* deletion (Supplementary Data Set 2), suggesting Smchd1 is required for the maintenance of chromosome architecture and gene silencing at some of its autosomal targets. By contrast, X-linked and *Hox* genes were not over-represented in the DEG (Supplementary Data Set 6), indicating Smchd1 may not be required to maintain gene silencing in these cases. Together our analyses have identified two classes of Smchd1 targets: those dependent on Smchd1 to maintain both chromatin architecture and gene silencing; and those where gene silencing is refractory to loss of Smchd1 in a differentiated cell, despite changes to higher-order chromatin organisation.

Smchd1 deletion results in activation of autosomal enhancers

Our previous study found Smchd1 bound at distal *cis*-regulatory elements in male NSCs, but little is known about Smchd1 function in these regions. Therefore, we profiled DNA accessibility using Assay for Transposase Accessible Chromatin using sequencing (ATAC-seq), revealing 4205 differentially accessible peaks between *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ in male and female NSCs (FDR<0.05, Supplementary Data Set 7). Here, *Smchd1* deletion was associated with increased accessibility (Fig. 2a, b). We focused on differentially accessible peaks with the greatest change in accessibility upon loss of Smchd1 (913 peaks, >two-fold change).

Whereas ATAC-seq reads tend to be highly enriched at the transcriptional start site (TSS) of genes, the majority of differentially accessible peaks sit 50-500 kb from the TSS (Fig. 2c). Consistent with this, differentially accessible peaks were enriched for E14.5 embryonic brain enhancers compared with the population (132/913, Hypergeometric probability = 0.0047), but not promoters (33/913, Hypergeometric probability = 1, Fig. 2d)[40]. Looking at the distribution of differential reads over E14.5 embryonic brain enhancers, we observed a subtle increase in accessibility centred around these regions upon *Smchd1* deletion (Fig. 2e).

Enhancer usage is highly tissue and cell-type specific[40]. To determine whether the remaining differentially accessible peaks contained ectopically activated non-neural enhancers, we performed ChIP-seq in *Smchd1*$^{del/del}$ and control NSCs for H3K27ac, a mark found at active enhancers, and compared to H3K27me3, which is synonymous with inactive or poised enhancers. The regions under the 913 differentially accessible peaks showed increased H3K27ac but depleted H3K27me3 coverage upon loss of Smchd1 (Fig. 2f, g, Supplementary Data Set 8). On occasion,

the short-range interactions brought differentially expressed genes into proximity with H3K27ac marked regions. These data suggest there could be ectopic activation of enhancers upon deletion of Smchd1.

Smchd1 is involved in regulating *Hox* cluster conformation, and *Hox* gene silencing in development and differentiation

The most significant differential interaction from *in situ* Hi-C analysis at 1 Mb resolution was a lost interaction anchored at Chr11:96-97 Mb, which contains the *HoxB* cluster where Smchd1 is highly enriched by ChIP-seq (Fig. 3a). We detected 23 weakened long-range interactions from this region upon *Smchd1* deletion, 14 of which were anchored in other Smchd1-bound clustered gene families, including the olfactory receptors and *Slc36* gene family (Fig. 3b, Supplementary Fig. 4). Long-range interactions were lost from the *HoxB* family in a range of 2–50 Mb, suggesting that many are extra-TAD interactions, although the TAD boundaries remained unchanged (Fig. 3c, Supplementary Data Set 3). The differential interaction with the lowest FDR was weakening of an interaction between the *HoxB* and *Keratin* gene clusters, also occupied by Smchd1 (Fig. 3d, Supplementary Fig. 4). We validated this finding by DNA FISH, observing a decreased frequency of interaction between the *HoxB* and *Keratin* gene clusters post-*Smchd1* deletion in NSCs (Fig. 3e, f, Chi-squared test $p = 0.000361883$). These data suggest Smchd1 mediates these long-range, extra-TAD interactions. Intriguingly, we also detect enhanced short-range interactions with diffHic and the formation of a loop between the distal end of the *HoxB* cluster and the neighbouring region (Supplementary Data Set 5); however these short-range interactions do not bring the *HoxB* cluster into contact with H3K27ac-marked regions.

Although all four *Hox* clusters are Smchd1 targets, and we observed altered chromatin interactions at the *Hox* clusters in NSCs, we do not observe upregulation of *Hox* genes upon *Smchd1* deletion. NSCs are derived from the embryonic forebrain, where *Hox* genes are not activated during development. Furthermore, epigenetic control of *Hox* loci is regulated by multiple, redundant mechanisms in committed cell lineages, meaning NSCs may be an intractable system for studying perturbations to the maintenance of *Hox* gene silencing. To test if Smchd1 plays a role in establishment or maintenance of developmental *Hox* gene silencing, we analysed the effect of *Smchd1* deletion in murine embryonic stem cells (mESC) differentiating into neuromesodermal progenitors. mESC do not express *Hox* genes, but as they are directed to differentiate into neuromesodermal progenitors they turn on and off *Hox* gene expression in a way that mimics the collinear expression pattern of the tailbud in the developing embryo[41,42].

One week post-*Smchd1* deletion, paired male *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ mESCs were differentiated using neuromesodermal progenitor protocols to activate the full complement of *Hox* genes[42,43]. Cells were sampled for RNA-seq at days 1, 5 and 8 of differentiation to capture different stages of *Hox* gene activation and silencing (Fig. 4a). At days 1 and 5 there were minimal changes to gene expression (Supplementary Fig. 5). At day 8 there were 1425 DEG (Supplementary Fig. 5, Supplementary Data Set 9), including Smchd1 targets such as the clustered protocadherin and *Snrpn* imprinted cluster genes. Control samples behaved as expected, with anterior *Hox* genes activated by day 5 and posterior *Hox* genes at day 8. Interestingly, at day 8, 19 *Hox* genes along the length of all four paralogous clusters were upregulated in *Smchd1*$^{del/del}$ cells compared to controls (FDR<0.05, Fig. 4b).

As *in vitro* differentiation assays may not be subject to the same selective pressures and strict regulatory pathways employed during development, we tested whether Smchd1 is involved in *Hox* gene silencing *in vivo*. We and others previously identified *Hoxb7* and *Hoxb1* as differentially-expressed genes in microarray data for whole embryos lacking Smchd1[23,25]; however, this approach was limited based on the whole embryo samples. The presomitic mesoderm is a region in the embryo that continually expands and segments during development, giving rise to vertebral precursors known as embryonic somites. Patterning within the presomitic mesoderm is critically regulated by the collinear activation of an anterior-to-posterior *Hox* code as development proceeds. This has proved a sensitive tissue to investigate other *Hox* regulators[46], so we performed RNA-seq on the presomitic mesoderm of somite matched E9.5 *Smchd1*$^{+/+}$ and *Smchd1*$^{MommeD1/MommeD1}$ embryos (Supplementary Fig. 6). We found premature activation of posterior *Hox* genes, but no differences in anterior *Hox* genes (Fig. 4c, Supplementary Data Set 10). The discrepancy in anterior *Hox* regulation may stem from inherent difference between *in vitro* and *in vivo* conditions, or differences in the developmental state of day 8 *in vitro* versus E9.5 *in vivo*. Nonetheless, both studies support a functional role for Smchd1-mediated chromosome conformation in *Hox* gene regulation.

To test the functional consequence of the *Hox* gene changes without Smchd1, we performed whole-mount skeletal preparations from E17.5 *Smchd1*$^{MommeD1/MommeD1}$ and *Smchd1*$^{+/+}$ male embryos, as the effects of *Hox* gene dysregulation in mouse are most pronounced in the embryonic skeleton[45]. As expected, all wild-type embryos displayed 26 presacral vertebral elements, with no deviation from the normal configuration of C7-T13-L6 (10/10). In contrast, all *Smchd1* null embryos exhibited a reduction in the number of presacral elements to 25, with varying changes in vertebral morphology centred around the thoraco-lumbar junction (10/10). In 5/10 cases we observed a complete loss of

the thirteenth thoracic element, with no concomitant gain of lumbar or sacral elements (C7-T12-L6). A variation was seen in 3/10 embryos, with a dramatically reduced rib process on T13 and loss of one lumbar element (C7-T13[reduced]-L5, Fig. 4d-g). The final two *Smchd1*$^{MommeD1/MommeD1}$ embryos exhibited a similar partial rib reduction or complete rib loss of at T13, but with additional asymmetric changes at the lumbo-sacral junction. Three of the *Smchd1* null embryos additionally showed presence of an ectopic rib on one side of C7 (Supplementary Fig. 6). These data demonstrate an essential role for Smchd1 in formation of the vertebral column, with changes at the lumbo-sacral junction being viewed as either a loss of one element, or alternatively, serial homeotic transformation from the 20th vertebral element (T13) onwards. Such posteriorizing transformations are consistent with precocious or enhanced posterior *Hox* activation[46,47]. Moreover, they are commonly seen in mice deficient for polycomb group proteins[48-51], demonstrating a role for Smchd1 in *Hox* gene silencing *in vivo*.

Our data provide the first evidence that *in vivo* Smchd1 is involved in silencing posterior *Hox* gene expression prior to activation. Given we see perturbations to long-range chromatin interactions at the *Hox* loci in *Smchd1*$^{del/del}$ NSCs, we propose that developmental Smchd1-mediated *Hox* gene silencing occurs through regulation of the higher order chromatin structure.

Acute loss of Smchd1 alters X chromosome conformation but does not result in reactivation of the Xi.

Much like *Hox* gene regulation, XCI is under strict developmental control, and once in the maintenance stage of silencing, reactivation requires disruption of multiple redundant epigenetic pathways. Our RNA-seq data suggested acute loss of Smchd1 did not result in gross reactivation

of the Xi in female NSCs, despite major changes in X-linked chromatin interactions. To ensure that we were not missing stochastic X reactivation, we also performed RNA-seq in NSCs derived from female $X_{129}{}^{XistAA}X_{Castaneus}$; *Smchd1*$^{del/fl}$ embryos, in which the *Xist*^{AA} allele ensures that the Castaneus X chromosome is the obligate Xi[52]. One week post-*Smchd1* deletion, we found no DEG on the Xi or indeed the Xa (Fig. 5a, b, Supplementary Data Set 11). These data demonstrate Smchd1-mediated higher-order chromatin organisation is dispensable for the short-term maintenance of XCI in NSCs.

Previous work suggested that the maintenance of the mega-domain structure of the Xi may not be required for gene silencing, because deletion of the *DXZ4* macro-repeat array dramatically altered the Xi architecture, but did not result in failed XCI during differentiation or X chromosome reactivation in differentiated cells[28,30]. We therefore used our Hi-C data to generate interaction maps for the X chromosome in *Smchd1* deleted and replete samples[53]. While we do not have the polymorphic samples required to detect genetic differences between the Xa and Xi in the Hi-C Data sets, Smchd1 is not found on the Xa in males[17], and our Smchd1-GFP ChIP-seq reveals Smchd1 covering the majority of the presumptive Xi (Fig. 5c), similar to the report of SMCHD1 ChIP-seq in human cells[26]. Smchd1-GFP is notably absent from the gene-poor region that shows highest enrichment for H3K9 methylation[54]: a region where we observe no differential interactions. Therefore, we assume changes in X-linked chromosome architecture post-*Smchd1* deletion relate to the Xi. We also performed *in situ* HiC followed by analysis with diffHic in paired male samples and found the overrepresentation of X-linked DIs was specific to female samples, further supporting that X-linked changes reflect changes to the architecture of the Xi (Supplementary Data Set 2). In control samples, the interaction map is consistent with the known mega-domain structure of the Xi, hinged at *DXZ4* (Fig. 5d)[5,28-30]. In the absence of Smchd1, the mega-domain structure is

less apparent (Fig. 5e), which is also observed in heatmaps from subsampled data (Supplementary Fig. 7). However, analysis with diffHic showed *Smchd1* deletion results in strengthening of short-range interactions concomitant with loss of the long-range mega-domain structure (Fig. 5f, g). These data are consistent with a role for Smchd1 contributing to the unique structure of the Xi by limiting short range interactions.

Smchd1 is not required to maintain Xi chromosome compaction or accessibility in mouse

Given the role of SMCHD1 in the compaction of the human Xi[26], we tested whether the large-scale changes in X-linked chromatin interactions upon *Smchd1* deletion indicated a loss of Xi compaction in mouse. We performed DNA FISH using an X chromosome paint in female *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ NSCs. The paint does not discern between Xa and Xi, but we observed approximately 30% difference in volume of the two X chromosomes within individual nuclei, consistent with reported 20-30% difference between the volume of the Xi and the Xa[55]. Therefore we assumed the smaller X within each nucleus was the Xi. There was no difference in the mean volume of the presumptive Xi or Xa in female *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ NSCs, nor was there a change in the difference of the volumes of the two X chromosomes measured within individual nuclei (Fig. 6a-c, Supplementary Fig. 7). Consistent with this, we did not see female-specific changes in X-linked chromatin accessibility by ATAC-seq in female *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ NSCs, when compared with males (Supplementary Figure 7, Supplementary Data Set 7). Together these data suggest that Smchd1 is not involved in maintaining chromatin compaction of the murine Xi.

Loss of Smchd1 results in widespread H3K27me3 enrichment on the Xi

It was intriguing that extensive loss of long-range chromatin interactions on the Xi was not accompanied by concurrent alterations in gene expression, compaction, or chromatin accessibility. We therefore looked for changes to an alternative silencing pathway on the Xi in the absence of Smchd1. Surprisingly, upon Smchd1 deletion, there was a 2.25-fold increase in intensity of the PRC2 mark H3K27me3 on the Xi and an increase in volume detected by immunofluorescence (Fig. 6d-f, Paired t-test, $p<0.00001$). We also detected an enrichment of H3K27me3 on the Xi by ChIP-seq in female $X_{129}^{XistΔA}X_{Castaneus}$; *Smchd1*$^{del/del}$ NSCs compared with controls (Fig. 6g, Supplementary Data Set 12). *Smchd1* deletion additionally resulted in detection of H3K27me3 MACS2 peaks in regions where they would not normally be found on the Xi (Fig. 6g, h, Supplementary Fig. 7), indicating spreading of H3K27me3. These data prompted us to re-examine H3K27me3 on the Xi in *Smchd1*$^{MommeD1/MommeD1}$ female nuclei, and there too we saw increased intensity of H3K27me3 on the Xi (Supplementary Fig. 7)[12], suggesting this change is not exclusive to acute loss of Smchd1. Interestingly, we did not observe enrichment of H3K27me3 at Smchd1's autosomal targets. Spreading of H3K27me3 domains on the Xi in the absence of Smchd1 has not been reported before.

One possible explanation for H3K27me3 enrichment in the absence of Smchd1, is that DNA hypomethylation could facilitate PRC2 recruitment[32-34]. Indeed, female embryos heterozygous or homozygous for a null allele of *Smchd1* display a dose-dependent reduction in DNA methylation at CGIs on the X chromosome[12,22]. Therefore, we assessed DNA methylation in NSCs derived from $X_{129}^{XistΔA}X_{Castaneus}$; *Smchd1*$^{+/fl}$ and $X_{129}^{XistΔA}X_{Castaneus}$; *Smchd1*$^{del/fl}$ female embryos, one week post-*Smchd1* deletion, using reduced representation bisulphite sequencing (RRBS). As expected[12], female cells that underwent XCI with a heterozygous deletion of *Smchd1* displayed reduced DNA methylation on the Xi (Fig. 6i, Supplementary Data Set 13). However, we did not detect hypomethylation on the

Xi upon acute heterozygous or homozygous deletion of *Smchd1*, at all CpGs or CGIs (Fig. 6i, j, Supplementary Data Set 13). Therefore an increase in H3K27me3 occurs independently of DNA hypomethylation following acute removal of Smchd1. Our data raise the possibility that Smchd1 contributes to the maintenance of the local chromatin environment at target loci, potentially by blocking the access of other epigenetic regulators.

Discussion

Smchd1 has well known roles in developmental gene silencing in mouse and human[12,21-23,25,56,57]. In this study we have implicated Smchd1 in the developmental silencing of another fundamentally important set of genes, the *Hox* genes. We present a genome-wide study of gene expression and chromatin architecture upon acute loss of Smchd1. Our data implicate Smchd1 in the maintenance of gene silencing and higher order chromatin organisation of target loci. Intriguingly loss of Smchd1 had a much more dramatic effect on the maintenance of chromatin conformation and composition than it did on accessibility and transcription. While *Smchd1* deletion resulted in altered conformation and expression at the clustered protocadherins and imprinted targets, the same was not true for the Xi. During development, Smchd1 is critical for gene silencing later in the ontogeny of XCI, implicating Smchd1 in the maintenance of XCI; however we have shown that dramatic remodelling of the X chromosome upon acute loss of Smchd1 is not associated with reactivation of genes from the Xi. We made similar observations at three paralogous *Hox* clusters. These findings show that select genes have a strictly developmental requirement for Smchd1 in regulating transcriptional silencing. Additionally we identify a subset of autosomal genes that require Smchd1 for the maintenance of gene silencing.

While we and others have previously suggested that Smchd1 may have a role in regulating higher order chromatin architecture, Smchd1's role in chromatin structure had not been explored[17,26,58]. Here we have determined that loss of Smchd1 results in genome-wide changes to long-range chromatin interactions. We found long-range interactions that reflect Smchd1 occupancy were weakened, whereas shorter-range interactions that are not enriched for Smchd1 occupancy were strengthened (Fig. 7a, b). Further examination of specific loci, and investigation of the changing chromatin landscape, along with gene expression, led us to propose that Smchd1 may function to maintain both chromatin architecture and gene expression, by limiting access of other epigenetic complexes that alter the chromatin. This is illustrated by our previous observation that the absence of Smchd1 at the clustered protocadherin locus results in an increase in Ctcf occupancy[17]. We now report that this too is true for H3K27me3 on the Xi, although at a much larger scale. Taken together, these data suggest that Smchd1 has the capacity to function at the chromatin as an insulating protein (Fig. 7b, c). It would be interesting to determine whether Smchd1 possesses insulating potential in other contexts, and towards other chromatin modifying proteins. Furthermore, given the tissue-specific manner in which enhancers are regulated, it will be informative to study Smchd1 occupancy and its effect on chromatin architecture in cells derived from different tissues, such as those where SMCHD1 plays an important role in disease[56-60].

Many of the lost long-range autosomal chromatin interactions upon *Smchd1* deletion occur across topologically associated domain (TAD) boundaries at a scale of tens of Mb, without concommitant changes to TAD boundaries. These data provide evidence that like some other epigenetic regulators, Smchd1 may be involved in mediating long-range chromatin interactions at target loci, potentially forming Smchd1 clusters like those observed between H3K27me3 targets[61]. By focusing on the *HoxB* cluster, we found that in tissue restricted cells, there was a decoupling of a loss of

Smchd1-mediated chromatin interactions, and transcriptional regulation at the locus; however by studying gene expression more dynamically *in vitro* and *in vivo*, we identified a role for Smchd1 in *Hox* gene silencing. We thus propose Smchd1-mediated long-range interactions are repressive and function in developmentally regulated silencing pathways.

Given the prominent role of Smchd1 in XCI, it was no surprise to find X-linked regions over-represented in the differentially interacting regions. While we could not discern between the Xi and Xa in our *in situ* Hi-C data, two complementary studies released while our paper was under revision have the appropriate data. Entirely consistent with our work, they report that Smchd1 is required for higher-order conformation of the Xi[62,63]. All three of our studies suggest that in *Smchd1* null cells the Xi adopts a structure that more closely resembles the structure of the Xa. Loss of Smchd1 and associated long-range chromatin interactions did not result in female-specific changes in chromatin accessibility, or the compaction of the murine X chromosomes as measured by DNA FISH. This is in contrast to observations by Nozawa *et al*., who suggested that SMCHD1 is involved in the compaction of the human Xi by bringing together H3K27me3 and H3K9me3 heterochromatin domains through an HBiX-HP1 pathway[26]. Our data suggest that in mouse, Smchd1 functions through a divergent pathway to regulate chromatin structure on the Xi. This is supported by Brideau *et al*., who found that the interaction between Smchd1 and the murine HBiX homologue Lrif1 is not conserved on the Xi in mouse[18]. Instead, all three of our studies suggest Smchd1 plays a role in maintaining the structure of the Xi, by insulating structures that are associated with the Xa, often mediated by Ctcf[62,63].

In conclusion, we have identified a role for the non-canonical SMC protein Smchd1 in the maintenance of higher order chromatin structures. We propose that Smchd1-mediated long-range

interactions have insulating capabilities, that limit access to other factors including chromatin modifying complexes and chromatin proteins that enable structures permissive for transcription. Given the important developmental role of SMCHD1 in mouse and human, it remains of interest to determine how Smchd1 is targeted to genes on autosomes and the inactive X chromosome to perform its functions as a structural protein.

Accession Codes

All genomics data has been deposited in the Gene Expression Omnibus, under accession number GSE111726.

Acknowledgements

We thank H. Coughlin, R. Allan and T. Johansson for useful discussions. This work was funded by the Australian National Health and Medical Research Council grant to MEB, JMM and MER (GNT1098290), and fellowships to JMM (GNT1105754) and MER (GNT1104924). NJ was supported by an Australian Research Training Program Fellowship. MEB was supported by a Bellberry-Viertel Senior Medical Research Fellowship. The Australian Regenerative Medicine Institute is supported by grants from the State Government of Victoria and the Australian Government. This work was made possible through Victorian State Government Operational Infrastructure Support and Australian National Health and Medical Research Council Research Institute Infrastructure Support Scheme.

Author Contributions

NJ designed experiments, performed experiments, interpreted and analysed data and wrote the paper. AK designed experiments, performed experiments, interpreted and analysed data. MT, PH and MER contributed to bioinformatic analyses of the data. TB, KB and MI performed experiments. HB designed and performed experiments. EM designed experiments, performed experiments and interpreted and analysed data. GFK, IDT and AWM designed and performed experiments, JMM designed experiments, interpreted data and edited the paper. MEB designed experiments, interpreted data, wrote and edited the paper.

Competing financial interests

All authors declare no competing financial interests.

References

1. Bouwman, B.A. & de Laat, W. Getting the genome in shape: the formation of loops, domains and compartments. *Genome Biol* **16**, 154 (2015).
2. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110-1121 (2016).
3. Kakui, Y. & Uhlmann, F. SMC complexes orchestrate the mitotic chromatin interaction landscape. *Curr Genet* **64**, 335-339 (2018).
4. Lieberman-Aiden, E. et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289-293 (2009).
5. Rao, S.S.P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
6. Schmitt, A.D., Hu, M. & Ren, B. Genome-wide mapping and analysis of chromosome architecture. *Nat Rev Mol Cell Biol* **17**, 743-755 (2016).
7. Dixon, J.R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).
8. Nora, E.P. et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22 (2017).
9. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56 (2017).

10. Eskeland, R. et al. Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. *Molecular Cell* **38**, 452-464 (2010).
11. Kundu, S. et al. Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Molecular Cell* **65**, 432-446.e5 (2017).
12. Blewitt, M.E. et al. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nature Genetics* **40**, 663-669 (2008).
13. Chiu, A., Revenkova, E. & Jessberger, R. DNA interaction and dimerization of eukaryotic SMC hinge domains. *J Biol Chem* **279**, 26233-42 (2004).
14. Haering, C.H., Lowe, J., Hochwagen, A. & Nasmyth, K. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell* **9**, 773-88 (2002).
15. Hirano, M. & Hirano, T. Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. *EMBO J* **21**, 5733-44 (2002).
16. Chen, K., Czabotar, P.E., Blewitt, M.E. & Murphy, J.M. The hinge domain of the epigenetic repressor Smchd1 adopts an unconventional homodimeric configuration. *The Biochemical journal* **473**, 733-742 (2016).
17. Chen, K. et al. Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E3535-E3544 (2015).
18. Brideau, N.J. et al. Independent Mechanisms Target SMCHD1 to Trimethylated Histone H3 Lysine 9-Modified Chromatin and the Inactive X Chromosome. *Molecular and Cellular Biology* **35**, 4053-4068 (2015).
19. Chen, K. et al. The epigenetic regulator Smchd1 contains a functional GHKL-type ATPase domain. *The Biochemical journal* **473**, 1733-1744 (2016).
20. Blewitt, M.E. et al. An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7629-7634 (2005).
21. Lemmers, R.J.L.F. et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nature Genetics* **44**, 1370-1374 (2012).
22. 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**, 265-79 (2012).
23. Gendrel, A.-V. et al. Epigenetic Functions of Smchd1 Repress Gene Clusters on the Inactive X Chromosome and on Autosomes. *Molecular and Cellular Biology* **33**, 3150-3165 (2013).
24. Leong, H.S. et al. Epigenetic regulator Smchd1 functions as a tumor suppressor. *Cancer Research* **73**, 1591-1599 (2013).
25. 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**, 1-1 (2013).
26. Nozawa, R.-S. et al. Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nature Structural & Molecular Biology* **20**, 566-573 (2013).
27. Mason, A.G. et al. SMCHD1 regulates a limited set of gene clusters on autosomal chromosomes. *Skeletal Muscle* **7**, 12 (2017).

28. Darrow, E.M. et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E4504-12 (2016).
29. Deng, X. et al. Bipartite structure of the inactive mouse X chromosome. *Genome Biology* **16**, 152 (2015).
30. Giorgetti, L. et al. Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575-579 (2016).
31. Guo, Y. et al. CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 21081-21086 (2012).
32. Brinkman, A.B. et al. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Research* **22**, 1128-1138 (2012).
33. Saksouk, N. et al. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. *Molecular Cell* **56**, 580-594 (2014).
34. Walter, M., Teissandier, A., Perez-Palacios, R. & Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* **5**(2016).
35. de Greef, J.C. et al. Smchd1 haploinsufficiency exacerbates the phenotype of a transgenic FSHD1 mouse model. *Hum Mol Genet* **27**, 716-731 (2018).
36. Lun, A.T.L. & Smyth, G.K. diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. *BMC Bioinformatics*, 1-12 (2015).
37. Serra, F. et al. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput Biol* **13**, e1005665 (2017).
38. Lajoie, B.R., Dekker, J. & Kaplan, N. The Hitchhiker's Guide to Hi-C Analysis: Practical guidelines. *Methods (San Diego, Calif.)* **72**, 65-75 (2015).
39. Durand, N.C. et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-8 (2016).
40. Shen, Y. et al. A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116-120 (2012).
41. Gouti, M. et al. In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. *PLoS Biol* **12**, e1001937 (2014).
42. Gouti, M. et al. A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Dev Cell* **41**, 243-261 e7 (2017).
43. Lippmann, E.S. et al. Deterministic HOX patterning in human pluripotent stem cell-derived neuroectoderm. *Stem Cell Reports* **4**, 632-44 (2015).
44. Wong, S.F. et al. Independent regulation of vertebral number and vertebral identity by microRNA-196 paralogs. *Proc Natl Acad Sci U S A* **112**, E4884-93 (2015).
45. Wellik, D.M. Hox genes and vertebrate axial pattern. *Curr Top Dev Biol* **88**, 257-78 (2009).
46. Boulet, A.M. & Capecchi, M.R. Duplication of the Hoxd11 Gene Causes Alterations in the Axial and Appendicular Skeleton of the Mouse. *Developmental Biology* **249**, 96-107 (2002).
47. Carapuco, M., Novoa, A., Bobola, N. & Mallo, M. Hox genes specify vertebral types in the presomitic mesoderm. *Genes Dev* **19**, 2116-21 (2005).

48. Akasaka, T. et al. A role for mel-18, a Polycomb group-related vertebrate gene, during theanteroposterior specification of the axial skeleton. *Development* **122**, 1513-22 (1996).
49. Akasaka, T. et al. Mice doubly deficient for the Polycomb Group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. *Development* **128**, 1587-97 (2001).
50. Suzuki, M. et al. Involvement of the Polycomb-group gene Ring1B in the specification of the anterior-posterior axis in mice. *Development* **129**, 4171-83 (2002).
51. van der Lugt, N.M. et al. Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes Dev* **8**, 757-69 (1994).
52. Royce-Tolland, M.E. et al. The A-repeat links ASF/SF2-dependent Xist RNA processing with random choice during X inactivation. *Nature Structural & Molecular Biology* **17**, 948-954 (2010).
53. Trussart, M. et al. Defined chromosome structure in the genome-reduced bacterium Mycoplasma pneumoniae. *Nat Commun* **8**, 14665 (2017).
54. Keniry, A. et al. Setdb1-mediated H3K9 methylation is enriched on the inactive X and plays a role in its epigenetic silencing. *Epigenetics & Chromatin*, 1-20 (2016).
55. Naughton, C., Sproul, D., Hamilton, C. & Gilbert, N. Analysis of active and inactive X chromosome architecture reveals the independent organization of 30 nm and large-scale chromatin structures. *Molecular Cell* **40**, 397-409 (2010).
56. Gordon, C.T. et al. De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. *Nature Genetics* **49**, 249-255 (2017).
57. Shaw, N.D. et al. SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. *Nature Genetics* **49**, 238-248 (2017).
58. Jansz, N., Chen, K., Murphy, J.M. & Blewitt, M.E. The Epigenetic Regulator SMCHD1 in Development and Disease. *Trends Genet* **33**, 233-243 (2017).
59. Lemmers, R.J.L.F. et al. Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nature Genetics* **32**, 235-236 (2002).
60. Gurzau, A.D. et al. FSHD2- and BAMS-associated mutations confer opposing effects on SMCHD1 function. *J Biol Chem* (2018).
61. Vieux-Rochas, M., Fabre, P.J., Leleu, M., Duboule, D. & Noordermeer, D. Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. *Proc Natl Acad Sci U S A* **112**, 4672-7 (2015).
62. Gdula, M.R. et al. The non-canonical SMC protein SmcHD1 antagonises TAD formation on the inactive X chromosome. *bioRxiv* (2018).
63. 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* (2018).

Figure legends

Figure 1. Smchd1 is required to maintain higher-order chromatin organisation at target loci, and to maintain gene silencing at a subset of its autosomal targets. a. Western blot for Smchd1 one week post Cre-mediated deletion in NSCs, with Actin as a loading control (Supplementary Data Set 1). **b.** Summarised data (mean ± standard deviation) from diffHic analyses between *Smchd1*$^{del/del}$ and *Smchd1*$^{fl/fl}$ NSCs at 1Mb and 100kb resolutions. n =3, FDR<0.05. **c.** Bar graph showing the proportion of autosomal (grey) and X-linked (black) differential interactions. **d.** Individual differential interactions along the X chromosome (bottom) with chromosome 2 (top) as a representative autosome for comparison. Genes are represented as black lines, and each 1Mb DI anchor point is represented as a blue box. **e.** The proportion of DIs that overlap a Smchd1 bound region of the genome (black) and those that do not (grey). **f.** The proportion of DIs that are weakened (black) vs strengthened (grey) in the absence of Smchd1. Strengthened interactions are those detected more frequently in *Smchd1*$^{del/del}$ samples than *Smchd1*$^{fl/fl}$ samples. Weakened interactions are the reciprocal. **g.** Normalised Hi-C interaction frequencies at 100kb resolution displayed as a heat map and rotated 45° for the 42-107 Mb region on chromosome 11. Left: *Smchd1*$^{fl/fl}$ heatmap, right: *Smchd1*$^{del/del}$ heatmap. The colour of the contact map (blue to red) indicates the $\log_2$(contact frequency). Black triangles bound TADs detected by TADbit. **h.** Gene expression data from female *Smchd1*$^{fl/fl}$ (n=3) (x axis) against *Smchd1*$^{del/del}$(n=3) samples (y axis). DEG between *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ samples (FDR<0.05) are shown in blue, clustered protocadherins in red and all other genes in grey. Axes are mean $\log_2$ counts per million (x) and $\log_2$ fold-change (y). **i.** Barcode plot of genes significantly upregulated (red bars, top of plot) and down-regulated (blue bars, bottom of plot) in *Smchd1*$^{MommeD1/MommeD1}$ NSCs compared with wild-type, in the *Smchd1*$^{del/del}$ versus

Smchd1$^{fl/fl}$ contrast. An enrichment line for each set shows the relative enrichment of the vertical bars in each part of the plot.

Figure 2. Smchd1 deletion results in activation of autosomal enhancers and increased promoter-enhancer interactions. a-e. Data from n=5, three female samples, and two males. **a/b.** Box plots showing coverage at all ATAC-seq peaks (**a**) and 4205 differential ATAC-seq (**b**) in log(cpm) in *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ NSCs. The box indicates the first to third quartile, horizontal line indicates the median, the whiskers indicate the median plus/minus the interquartile (25-75%) range multiplied by 2, and individual points that fall outside of this range are shown as circles. **c.** Distribution of the 913 differential ATAC-seq peaks that showed a greater than two-fold change in accessibility, relative to transcriptional start sites (TSS) calculated using the GREAT algorithm. The percentage of associations is plotted on the y-axis, and the distance from the TSS in windows as indicated in kb are plotted on the x-axis. **d.** Fold enrichment of chromatin features in 913 differential ATAC-seq peaks over the population. The colour scale indicates the fold enrichment. Smchd1 binding sites were defined in this study (Supplementary Data set 3) and all other genomic features were taken from Shen *et al.*[40]**. e.** Coverage in log(cpm) of the 4205 differential ATAC-seq peaks centred over neural enhancers ± 500 bp. Reads from *Smchd1*$^{fl/fl}$ in red and reads from *Smchd1*$^{del/del}$ NSCs in blue. Note the truncated axis. **f.** Heat maps depicting H3K27ac signals and WCE control centred around the 913 differential ATAC-seq peaks ± 10,000 kb (x axis). y-axis presents the 913 differentially accessible peaks sorted by H3K27ac enrichment. n=1. **g**. Average H3K27me3 and H3K27ac over the 913 DAP shown in **e**, for *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ NSCs.

Figure 3. Smchd1 is required to maintain long-range interactions between the *HoxB* cluster and linked clustered gene families. a. ChIP-seq for Smchd1-GFP over the *HoxB* cluster. Genes

are shown, followed by MACS2-like Smchd1-GFP ChIP-seq peaks as black bars. Smoothed Smchd1-GFP ChIP-seq track, normalised to reads from the whole cell extract, displaying positive enrichment of GFP ChIP-seq in *Smchd1*$^{GFP/GFP}$ NSCs, alongside wild-type cells as an isotype control. Scale is normalised ChIP-seq reads. **b.** Representation of *cis* differential interactions detected between female *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ NSCs at 1Mb resolution at the *HoxB* cluster on chromosome 11. *HoxB* cluster is indicated in purple, differentially interacting regions in blue. All interactions depicted are weakened without Smchd1. Half circles represent the interaction. **c.** Normalised Hi-C interaction frequencies at 100kb resolution displayed as a heat map and rotated 45° for the region surrounding the differential interaction with the lowest FDR from diffHic analysis at 1Mb resolution, anchored at the *HoxB* and *Keratin* gene clusters. Above the horizontal is the *Smchd1*$^{fl/fl}$ heatmap, below the *Smchd1*$^{del/del}$. The colour of the contact map (blue to red) indicates the $\log_2$(contact frequency). Black triangles bound TADs called by TADbit, dotted white region indicates the portion that is shown in higher magnification in **d.** Region of differential interaction in (**d.**) is depicted in a white dotted square. Between the heatmaps, genes are shown as bars (black, green for *Keratin* cluster genes, light blue for *HoxB* genes, red are differentially expressed in NSCs), Smchd1 ChIP-seq peaks are shown as black bars. **e.** DNA FISH using BAC probes against the *HoxB* cluster (magenta) and the *Keratin* gene cluster (yellow). DNA is stained with DAPI (cyan). The scale bar represents 4 μm. Images show loci scored as not interacting (orange arrows, left and middle panels) or interacting (white arrows, middle and right panels). **f.** Quantification of number of *HoxB-Keratin* interactions from n=146 alleles n=148 alleles from *Smchd1*$^{fl/+}$ and *Smchd1*$^{del/del}$ cells, respectively.

Figure 4. Smchd1 is required for developmentally regulated *Hox* gene silencing *in vitro* and *in vivo*. a. Schematic of *in vitro* differentiation of mESC to neuromesodermal progenitors. mESC

are derived in 2i + LIF medium. One week prior to differentiation (D-7) cells were infected with MSCV-Cre-puro or MSCV-puro retrovirus (green arrow), and transduced cells selected with puromycin (blue arrow). Cells were weaned into 15% Serum + LIF medium. For differentiation cells were changed into N2B27 medium containing Fgf-2. At day 2 (D2) Gsk3i, and at D6 Gdf11 were added. Cells were harvested for RNA-seq at day (D) 1, 5 and 8 (red arrows). **b.** Heatmaps showing the $\log_2$(fold change) expression differences in *Hox* genes between *Smchd1* deleted and control samples at day 1, 5 and 8 of differentiation from mESC to neuromesodermal progenitors. Red to blue colour indicates fold change. Genes written in red have statistically significant differences at day 8, n=2. **c.** As for **b**, except samples are presomitic mesoderm from somite matched *Smchd1*$^{+/+}$ and *Smchd1*$^{MommeD1/MommeD1}$ E9.5 embryos. Somite numbers are indicated at the left. Log_2fold change is shown for n=4 female *Smchd1*$^{MommeD1/MommeD1}$ and n=3 female *Smchd1*$^{+/+}$ embryos. **d/f.** Representative sagittal views of whole-mount E17.5 male *Smchd1*$^{+/+}$ (**d**) and *Smchd1*$^{MommeD1/MommeD1}$ (**f**) embryos with forelimbs and hindlimbs removed, stained with Alizarin red and Alcian blue for bone and cartilage, respectively. **e/g.** As per **d/f** showing dorsal view. Putative T13 is indicated with an asterix, note absence of rib on T13 in the *Smchd1*$^{MommeD1/MommeD1}$ embryo, representing homeotic transformation.

Figure 5. Smchd1 deletion appears to alter the conformation of the Xi, but is not required to maintain gene silencing on the Xi a-b. X-linked gene expression data from female *Smchd1*$^{del/fl}$ (n=2) (x axis) versus *Smchd1*$^{del/del}$ (n=2) samples (y-axis). X-linked genes are represented in blue; no differentially expressed genes were detected on the Xi or Xa between *Smchd1*$^{del/fl}$ and *Smchd1*$^{del/del}$ samples (FDR<0.05) **a.** Plot for the 129-derived *Xist*$^{\Delta A}$ allele containing Xa **b.** Plot for the Castaneus-derived obligate Xi. **c**. Enrichment of Smchd1-GFP over whole cell extract on the X Chromosome. Height and colour of peaks indicates strength of read counts normalised to whole

cell extract; red is most enriched, blue is most depleted. Position of MACS2-like peaks called using Seqmonk are indicated in blue. Enriched domain detector (EDD) domains are shown in green. n=2. Histones track depicts H3K27me3 (salmon) and H3K9me3 (blue) derived from Keniry *et al*.[54]. DiffHic track depicts differential interactions from our own analyses, n=3. **d-e.** Normalised Hi-C interaction frequencies at 100 kb resolution displayed as a heat map for the X chromosome in *Smchd1*$^{fl/fl}$ (**c**) and *Smchd1*$^{del/del}$ (**d**) NSCs. The colour of the contact map, from blue to red, indicates the $\log_2$(contact frequency). The arrow indicates *DXZ4*. **e.** A 45° rotated transformed interaction matrix displayed as a heatmap for the X chromosome in **c-d** that shows the distribution of absolute distances against the mean distances of all interactions in the original matrix. The colour from yellow to red indicates the density estimated from the absolute and mean distances. **f.** Differential plot of the transformed interaction matrices for the X chromosome in *Smchd1*$^{del/del}$ minus *Smchd1*$^{fl/fl}$ in part **e**. The colour from blue to red indicates the difference between the two densities estimated in part **e**.

Figure 6. Smchd1 limits H3K27me3 spreading on the Xi, though it is not required to maintain Xi compaction a. Confocal imaging with Airyscan processing of *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ NSC nuclei following DNA FISH with X chromosome paint. n=23 observations per genotype, n=2 cell lines. DAPI stained DNA (blue) and X chromosome paint (magenta), shown with 3D rendering to measure volume, below confocal images. **b-c.** The volumes of the largest and smallest X chromosome (**b**) or percent difference between them per nucleus in each genotype (**c**). Two-tailed student's t-test, mean ± SEM. **d.** Confocal images with Airyscan processing of H3K27me3 immunofluorescence in female *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ NSCs (green). **e.** Volume and **f.** mean intensity of focal H3K27me3 enrichment in 3D using Imaris. Two-tailed student's t-test, mean ±

SEM. 38 observations, n=3 cell lines, p <0.00001. **g.** Genome browser tracks for H3K27me3 ChIP-seq on the Xi in *Smchd1*$^{del/fl}$ and *Smchd1*$^{del/del}$ NSCs (n=2). Scale is normalised ChIP-seq read counts, normalised for read depth using total autosomal read count. Genes in black, MACS2-called peaks in *Smchd1*$^{del/fl}$ (blue) and *Smchd1*$^{del/del}$ NSCs (red). **h.** Coverage of MACS2-called H3K27me3 peaks in *Smchd1*$^{del/del}$ and *Smchd1*$^{del/fl}$ samples, centred around MACS2-called H3K27me3 peaks detected on the Xi of *Smchd1*$^{del/fl}$ samples ± 5000 kb. **i.** Bean plots showing DNA methylation (%mC) on the Xa (blue) and Xi (red) as determined by RRBS in *Smchd1*$^{del/del}$ (n=1), *Smchd1*$^{del/fl}$ (n=1), *Smchd1*$^{del/+}$ (n=1) and *Smchd1*$^{fl/+}$ (n=1) NSCs. CpG analysed were those with a minimum coverage of 10 reads in each condition. The former 2 genotypes were derived from embryos haploinsufficient for Smchd1 during embryogenesis, the latter 2 were Smchd1 replete. **j.** DNA methylation of CpG Islands (%5mC) that map to the Castaneus genome in *Smchd1*$^{del/fl}$ (x-axis) and *Smchd1*$^{del/del}$ NSCs. CGIs on the Xi are indicated in blue. CGIs hypomethylated during development in *Smchd1* null female embryos are in red. Pearson's correlation value (R) is shown.

Figure 7. An insulating model for Smchd1 function: Smchd1 is involved in the maintenance of long-range repressive chromatin structures, which limit promoter-enhancer interactions that are permissive for transcription, and access to chromatin modifiers. a. In *Smchd1*$^{+/+}$ cells, Smchd1 is targeted to regions of the genome that are repetitive in nature (red and blue repeating blocks), such as clustered gene families, including the *Hox* clusters, where it enables long-range chromatin looping. **b.** In the absence of Smchd1, long-range repressive interactions are lost, which based on our previous data allows for the formation of Ctcf-mediated (blue ovals) promoter-enhancer interactions that are permissive for transcription (depicted in purple-green and red-orange). **c.** On the inactive X, loss of Smchd1 results in enhanced H3K27me3. We propose this is

due to loss of Smchd1's capacity to insulate the chromatin from other chromatin modifiers (orange circles).

Online methods

Mouse strains

All mice were bred and maintained using standard animal husbandry procedures, approved by the WEHI Animal Ethics Committee under animal ethics numbers AEC 2014.026 and AEC 2018.004.

Smchd1$^{fl/fl}$ mice were generated on a C57BL/6 background, as described previously[35]. Mice carrying the *Smchd1*^{MommeD1} mutation as previously described[20] were maintained on the FVB/N inbred background. Both strains were genotyped as previously described (primers in Supplementary Table 1). *Xist*$^{\Delta A}$ mice were generated on a 129 background as previously described[52]. Castaneus strain (Cast/EiJ) animals were obtain from Jackson laboratories.

Construction of the *Smchd1*^{GFP} targeted allele is described in Supplementary Note 1.

Embryo genotyping

Embryo tails or yolk sacs were used to prepare DNA using standard methods. Genotypes were determined by PCR using GoTaq Green (Promega) or allelic discrimination for MommeD1.

Primers for genotyping were as for standard genotyping of the strain, except that sex was additionally determined by PCR for *Otc* (X chromosome) and *Zfy* (Y chromosome) (Supplementary Table 1). Sex was often also determined at E14.5 by dissection of embryonic gonads.

Derivation of Mouse Embryonic Stem Cells (mESCs)

Female mice were super-ovulated by injecting 5 IU folligon (MSD Animal Health Australia) two days prior to, and 5 IU chorulon (MSD Animal Health Australia) on the day of mating. At E3.5, dams were sacrificed, uteri removed and blastocysts flushed from the uterine horns with M2 medium (Sigma-Aldrich). Blastocysts were washed in M2 medium twice, and 2i+LIF medium [KnockOut DMEM (Life Technologies), 1X Glutamax (Life Technologies), 1 X MEM Non-Essential Amino Acids (Life Technologies), 1 X N2 Supplement (Life Technologies), 1 X B27 Supplement (Life Technologies), 1 X 2-mercaptoethanol (Life Technologies), 100 U/mL Penicillin/100 μg/mL Streptomycin (Life Technologies), 10 μg/mL Piperacillin (Sigma-Aldrich), 10 μg/mL Ciprofloxacin (Sigma-Aldrich), 25 μg/mL Fluconazol (Selleckchem), 1000 U/mL ESGRO Leukemia Inhibitory Factor (Merck), 1 μM StemMACS PD0325901 (Miltenyi Biotech), 3 μM StemMACS CHIR99021 (Mitenyi Biotech)] twice. Blastocysts were plated in non-tissue culture treated 24-well plates in 2i+LIF medium. Following 7 days in cultured at 37°C in a humidified atmosphere with 5 % (v/v) carbon dioxide and 5 % (v/v) oxygen, outgrowths were moved by mouth-pipetting through trypsin-EDTA for 2 minutes, mESC wash media [KnockOut DMEM (Life Technologies), 10 % KnockOut Serum Replacement (Life Technologies), 100 IU/ml penicillin/100 μg/ml streptomycin (Life Technologies)], and finally 2i+LIF. Outgrowths were disrupted by pipetting, and transferred into a 24-well plate to be cultured as mESC lines.

mESCs were maintained in suspension culture in serum-free 2i+LIF medium on non-tissue culture treated plates at 37°C in a humidified atmosphere with 5 % (v/v) carbon dioxide and 5 % (v/v) oxygen. mESC were passaged daily. mESC colonies were collected and allowed to settle in a tube for < 5 minutes. Supernatant containing cellular debris was removed, and mESC colonies were resuspended in Accutase (Sigma-Aldrich) and incubated at 37°C for 5 minutes to achieve a single-cell suspension. At least 2X volume of mESC wash media was added to the suspension, and cells were pelleted by centrifugation at 600 x g for 5 minutes.

Differentiation of mESC into Neuromesodermal Progenitors (NMPs)

This protocol was adapted from three previous reports[41–43,45]. Four days prior to differentiation, mESCs were put onto tissue-culture ware that had been coated with 0.1% porcine gelatin (Sigma-Aldrich). mESC were passaged into 75 % 2i+LIF medium 25 % ES cell DMEM+LIF [High glucose DMEM, 0.085 mM MEM Non-Essential Amino Acids (Life Technologies), 34 mM NaHCO3, 0.085 mM 2-mercaptoethanol (Life Technologies), 100 μg/ml streptomycin, 100 IU/ml penicillin, 15% foetal bovine serum (FBS) (Life Technologies), 1000 U/mL ESGRO Leukemia Inhibitory Factor (Merck), 10 μg/mL Piperacillin (Sigma-Aldrich), 10 μg/mL Ciprofloxacin (Sigma-Aldrich), 25 μg/mL Fluconazol (Selleckchem)] . After 24 h, the medium was changed to 50 % 2i+LIF medium 50 % ES cell DMEM+LIF. After 48 h, the medium was changed to 25 % 2i+LIF medium 75 % ES cell DMEM+LIF. After 72 h, the cells were split and seeded at 2.5 x 10^5 cells per cm^2 in 100 % ES cell DMEM+LIF. The next morning (D0 differentiation), ES cell DMEM+LIF was removed, cells were washed with PBS to remove any traces of FBS, and medium was replaced with N2B27 medium [1:1 mix of Advanced DMEM/F12 (Life Technologies) and Neurobasal media (Life Technologies), 0.5 X N2-supplement (Life Technologies), 0.5 X B27-

supplement, 1 X Glutamax (Life Technologies), 40 μg/mL BSA Fraction V (Life Technologies), 1X Beta-mercaptoethanol (Life Technologies), 100 U/mL Penicillin/100μg/mL Streptomycin (Life Technologies), 10 Piperacillin μg/mL (Sigma-Aldrich), 10 μg/mL Ciprofloxacin (Sigma-Aldrich), 25 μg/mL Fluconazol (Selleckchem)] supplemented with 10 ng/mL recombinant human basic FGF (Peprotech). After 48 hours (D2), medium was replaced with N2B27 medium supplemented with 10 ng/mL recombinant human basic FGF (Peprotech) and 5 μM StemMACS CHIR99021 (Mitenyi Biotech). After a further 48 hours (D4), medium was replaced with N2B27 plus bFGF and StemMACS CHIR99021. After 48 hours (D6), medium was replaced with N2B27 medium supplemented bFGF, StemMACS CHIR99021, and 50 ng/mL recombinant human GDF11 (Mitenyi Biotech). NMPs were maintained in N2B27 medium supplemented bFGF, StemMACS CHIR99021 and GDF11.

Generation of Neural Stem Cells (NSCs)

NSCs were derived as exactly previously described[17] and cultured with minimal changes. For culture, cells were seeded onto either plates coated with 15 ng/mL polyornithine (Sigma-Aldrich) and 10 ng/mL laminin (Sigma-Aldrich), or directly into tissue culture treated plates with 10 ng/mL laminin (Sigma-Aldrich) added to the medium, at a density of 40,000 cells per cm^2. Primary cells were maintained for a maximum of 20 passages.

Generation of *Smchd1*$^{MommeD1/MommeD1}$ MEFs

MEFs were generated as described from E10.5 *Smchd1*$^{MommeD1/MommeD1}$ embryos[12], and cultured at 37°C in 5% CO_2 and 5% O_2 for one week. Cells were passaged using 0.5% trypsin-EDTA (Life

Technologies) and seeded at approximately 40,000 cells/cm^2 on gelatin coated cover slips for immunofluorescence.

Generation of *Smchd1*$^{del/del}$ NSCs

Smchd1$^{fl/fl}$ NSCs were transduced as described below with concentrated retroviral supernatant containing the MSCV-Cre-Puromycin construct. The next day, cells were selected with 1 μg/mL Puromycin (Sigma-Aldrich) for at least 7 days. *Smchd1* deletion was confirmed by PCR, RT-qPCR, and Western Blot.

Generation of *Smchd1*$^{del/del}$ mESCs

Smchd1$^{fl/fl}$ NSCs were transduced as described below with concentrated retroviral supernatant containing either the MSCV-Cre-Puromycin or MSCV-Puromycin construct. The next day, cells were selected with 1 μg/mL puromycin (Sigma-Aldrich) for at least 7 days. Smchd1 deletion was confirmed by PCR and RNA-seq.

$X_{129}{}^{XistΔA}X_{Castaneus}$ *Smchd1*$^{del/del}$ NSCs were generated by Cre-mediated deletion as above from $X_{129}{}^{XistΔA}X_{Castaneus}$ *Smchd1*$^{del/fl}$ NSCs. $X_{129}{}^{XistΔA}X_{Castaneus}$ *Smchd1*$^{del/+}$ controls were generated by Cre-mediated deletion from $X_{129}{}^{XistΔA}X_{Castaneus}$ *Smchd1*$^{fl/+}$ NSCs.

Retrovirus Production and Transduction

VSVg pseudotyped retroviral supernatants were produced using calcium phosphate mediated transient transfection of 293T cells, as previously described[64]. The media was collected at 48 and

72 hours post-transfection, centrifuged to remove residual 293T cells, and either snap frozen or concentrated using Poly(ethylene glycol).

For transduction of NSCs, NSCs were seeded at a density of 40,000 cells per cm^2 8-16 hours before transduction. PEG concentrated viral supernatant was added to the culture together with polybrene at a final concentration of 4 μg/mL. On the next day, cells were selected with 1 μg/mL puromycin.

For transduction of mESCs, ESCs were seeded at 10^5 cells per cm^2 on plates that had been coated with 0.1% gelatin or in suspension at 10^5 cells per mL in 2i+LIF medium containing PEG concentrated viral supernatant and 8 μg/mL polybrene. The next day medium was changed and cells were selected with 1 μg/mL puromycin.

Dissection of the presomitic mesoderm from E9.5 embryos

The PSM was dissected from *Smchd1*$^{MommeD1/MommeD1}$ *Smchd1*$^{MommeD1/+}$ and *Smchd1*$^{+/+}$ E9.5 embryos as previously described[44]. In brief, under a light microscope the tail bud was removed approximately 1.5 somite lengths after final formed somite. The PSM was snap frozen for later RNA harvest using a Qiagen RNeasy micro kit. The yolk sac was taken for genotyping. The embryo body was fixed in 4% paraformaldehyde for 2 hours, then washed through a methanol series as previously described[44]. To image, the embryo was dipped in ethidium bromide solution then photographed on a fluorescence dissection microscope to count somites. RNA-seq comparisons were made between somite-matched, sex-matched *Smchd1*$^{MommeD1/MommeD1}$ and *Smchd1*$^{+/+}$ PSM samples.

Immunofluorescence

Immunofluorescence was performed on *Smchd1*$^{GFP/GFP}$, *Smchd1*$^{fl/fl}$ and Smchd1$^{del/del}$ NSCs and *Smchd1*$^{+/+}$, *Smchd1*$^{MommeD1/MommeD1}$ MEFs as described in Chaumeil *et al*.[65] Following immunofluoresce, cells were mounted in Vectashield HardSet mounting medium with DAPI (Vector Laboratories). When *Smchd1*$^{GFP/GFP}$ cells were used, detection of Smchd1 protein was enabled via detection of native Smchd1-GFP fusion protein. Cells were visualised on an Elite Widefield (DeltaVision), LSM 880 (Zeiss), or Live-Cell AxioObserver (Zeiss) microscope. Images were analysed using the open source ImageJ distribution package, FIJI[66]. For each experiment 100–250 nuclei were analysed.

Imaris Software (Bitplane) was used to measure the volume occupied by markers of the Xi (H3K27me3 or X chromosome paint) within the nucleus. A threshold was manually set to measure the signal only above nucleoplasmic staining or background. A region of interest was then defined based on DNA paint or focal H3K27me3 enrichment. Volume was then calculated for the region of interest above the threshold.

DNA FISH

DNA FISH was performed on $X_{129}^{XistΔA}X_{Castaneus}$ *Smchd1*$^{del/del}$ and *Smchd1*$^{+/fl}$ NSCs. 1 μg RP23-196F5 (*HoxB* locus) or RP2390M1 (*Keratin* locus) BAC DNA (CHORI) were used in a nick translation reaction (Vysis) to generate DNA probes labelled with SpectrumRed dUTP or SpectrumGreen dUTP (Vysis), respectively. ~100 ng probe per sample was precipitated in Ethanol with 10% NaOAc, and 1 μg Salmon Sperm (Life Technologies), and 1μg Mouse Cot-1 DNA before being resuspended in formamide (Sigma-Aldrich), denatured at 75°C for 10 minutes, and allowed to compete with Cot-1 DNA for 1 hour at 37°C. Cells were prepared for DNA FISH exactly as described in Chaumeil *et al*.[65]. Cells were mounted in Antifade Mounting Medium (Vectashield)

and visualised on an LSM 880 (Zeiss) microscope, with Airyscan processing. Images were analysed using the open source ImageJ distribution package, FIJI. Images were processed with the Median 3D filter to reduce background signal, and each image was manually thresholded with the Yen thresholding method. Distances between FISH signals were measured using the ImageJ plugin, DiAna[67]. Overlapping or touching FISH signals for *HoxB* and *Keratin* probes were scored as interacting, whereas those where paired probes fell within 4μm but were not touching were scored as not interacting. 166 alleles were scored for the *Smchd1*$^{fl/+}$ NSCs, 152 alleles were scored for the *Smchd1*$^{del/del}$ NSCs.

X Chromosome Paint

Female *Smchd1*$^{fl/fl}$ and Smchd1$^{del/del}$ NSCs were harvested with Accutase (Sigma-Aldrich), fixed for 5 minutes in Carnoys' fixative (3:1 methanol:acetic acid), followed by centrifugation for 10 minutes at 4°C. Fixation and centrifugation were repeated, and nuclei were spotted onto a Superfrost Plus slide (Thermo Scientific) and allowed to dry. FITC-conjugated X-chromosome paint (AMP 0XG) was used as per the manufacturer's instructions (Cytocell Ltd.). Cells were visualised on an LSM 880 (Zeiss) microscope, with Airyscan processing. Images were analysed using Imaris Software (Bitplane) as described above. For each experiment 10-20 nuclei were analysed.

***In situ* Hi-C**

In situ HiC was performed as described with the following modifications[5]. 10^7 *Smchd1*$^{fl/fl}$ and Smchd1$^{del/del}$ NSCs were cross-linked in 1% formaldehyde in NSC basal medium for 10 minutes at

room temperature with rotation. Formaldehyde quenched with 200 mM Glycine, cells were washed with PBS, and then pellet was snap-frozen, and stored at -80°C. Nuclei were lysed in cold lysis buffer [10mM Tris-HCl pH8.0, 10mM NaCl, 0.2% Igepal CA630, 1 x Protease Inhibitor (Sigma)] and centrifuged at 600 x g for 5 minutes. Nuclei were resuspended in 0.5% (w/v) SDS and incubated at 62°C for 5 minutes. SDS was quenched with 10% Triton X-100 for 15 minutes at 37°C. Chromatin was digested by adding 200 U MboI with 200 mg/mL BSA for 1 hour, and a further 100 U every hour for 4 hours, and then incubated overnight at 37°C with rotation. Subsequent steps were performed exatly as described[5]. Libraries were sequenced to a depth of at least 50 million reads (*Smchd1*$^{del/del}$: 124860050; 100099110; 57467204. *Smchd1*$^{fl/fl}$: 94218954; 200738168; 119847878).

Primary data processing was performed using HiCUP [68] v0.5.8. Differential Interactions were identified using diffHiC v1.12.0, which utilises edgeR statistics[36]. In brief, reads mapped and filtered using HiCUP were counted into 100 kb and 1 Mb bin pairs. Noise was removed by filtering out low abundance reads based on a negative binomial distribution and using inter-chromosomal counts to determine non-specific ligation events. Libraries were then normalised using TMM normalisation, and trended biases removed by fitting libraries to a generalised linear model. EdgeR was then used to test for differential interactions between genotypes, at either 100 kb or 1 Mb resolution, using a quasi-likelihood F-test, then we adjusted for multiple testing using FDR. Overlap with Smchd1 ChIP-seq peaks was performed using genomic coordinates for autosomal and X-linked MACS2 called ChIP peaks from GFP ChIP performed using MNase and sonication in *Smchd1*$^{GFP/GFP}$ NSCs, respectively[69].

Generation of Hi-C contact matrices

To construct the Hi-C interaction matrices of the Smchd1 replete and deleted genomes, all reads were mapped to the *mm10* reference genome using TADbit[37] v0.2.0.58 with the original iterative mapping strategy ICE[70]. TADbit uses the GEM mapper[71] to map Fastq files. The minimal size for mapping was set to 25bp and the iterative mapping procedure increased in steps of 5bp until a maximal read length of 75bp was reached. A filtering strategy was then applied to correct for experimental biases as described in Imakaev *et al*.[37]. Once filtered, the read-pairs were binned at a 100 kb, 500 kb or 1 Mb resolution and columns containing few interactions counts were removed following the two-step strategy described in Serra *et al*.[37]. The remaining bins were further normalised using ICE[37] as implemented in TADbit. As the resulting interaction matrices of the three separate Hi-C experiments for both conditions were highly correlated (Spearman correlation > 0.96), we merged the input reads into a single dataset, one for each condition. The new datasets were also filtered and normalised by TADbit and the resulting merged interaction maps were used to describe the 3D organisation of the *Smchd1* deleted and Smchd1 wild-type genome.

The subsampled contact matrices were constructed from a random selection of 200 million read-pairs for both conditions respectively, then binned at a 100 kb resolution and then filtered and normalised following the same steps described previously.

TAD detection on Hi-C contact maps

The merged interaction maps were used for domain detection at a 100kb resolution. The TADbit program[37,53] employs a breakpoint detection algorithm that returns the optimal segmentation of the chromosome using a BIC-penalized likelihood[72].

Alignment of TAD boundaries

To assess whether TAD borders are conserved throughout different experiments, TADbit performed a multiple-experiment border alignment algorithm, starting from different border definitions of the same genomic region and aligns each TAD to a consensus TAD list, using the classic Needleman-Wunsch algorithm[73].

Compartment Detection on Hi-C Contact Maps

The merged interaction maps were used for compartments detection at a 1 Mb resolution. The TADbit program[37,53] was used to search for compartments in each chromosome, by computing a correlation matrix on the normalised matrix and by using the first eigenvector to identify compartments.

Annotation of chromatin looping with HiCCUPS

Annotation of chromatin looping was performed using the HiCCUPS peak calling algorithm[39]. Replicate HiC fastq files were merged for each condition and processed using the Juicer pipeline v1.8.9 to map HiC reads to the mm10 reference genome, remove duplicate and near duplicate reads, and reads that map to the same restriction fragment. HiC contact maps were produced from filtered reads with a MAPQ >30. HiCCUPS was run using default parameters, and identified 3192 loops in *Smchd1*$^{del/del}$ HiC heatmaps, and 2410 loops in *Smchd1*$^{fl/fl}$ HiC heatmaps.

Transformed contact matrices

The merged interaction maps were used to compute the absolute distance and the mean distance between interacting bins and to construct the transformed interaction matrices.

For each interaction $m_{i,j}$ defined as an interaction between the bin *i* and bin *j*, the absolute distance *AD* was computed as the absolute difference between the genomic location of the bin *i* and the bin *j*. The mean distance *MD* was computed as the mean between the genomic location of the bin *i* and bin *j*.

For an interaction $m_{17,26}$ of the merged interaction map at a 100 kb resolution for the X chromosome in *Smchd1*$^{fl/fl}$

$$AD = |17 * 100000 - 26 * 100000| = 900000$$

$$MD = (17 * 100000 + 26 * 100000/2) = 2150000.$$

The differential plot was computed as the difference between the two transformed interaction matrices *Smchd1*$^{del/del}$ versus in *Smchd1*$^{fl/fl}$.

RNA-seq

RNA was extracted using and RNeasy Mini or Micro kit (Qiagen) or a Quick RNA kit (Zymo). RNA was quantified using Qubit RNA Assay Kit (Life Technologies). Libraries were prepared using TruSeq RNA sample preparation kit from 1 μg total RNA as per manufacturers' instructions. 200-400bp products were size selected and cleaned up using AMPure XP magnetic beads. Final cDNA libraries were quantified using a Qubit dsDNA Assay Kit (Life Technologies) for sequencing on the Illumina NextSeq platform. C57BL/6 inbred NSC RNA-seq, ES cell RNA-seq and PSM RNA-seq libraries were subject to 75 bp single-end reads. 129/Cast NSC samples were subject to 75 bp paired-end reads. Reads were mapped with Tophat v2.0.12 using default options[74].

RNAseq data were analysed using Seqmonk v1.36.1 or v1.38.2. Reads under mRNA were quantified, and EdgeR statistics were used to test for differentially expressed genes in Seqmonk (FDR < 0.05)[75]. Geneset enrichment test for differentially expressed genes in *Smchd1*$^{MommeD1/MommeD1}$ was performed using rotation gene set test (ROAST)[76].

ChIP-seq

To ChIP for Smchd1 on the Xi, 4 x 10^7 female *Smchd1*$^{GFP/GFP}$ and *Smchd1*$^{+/+}$ NSCs were harvested using Accutase (Sigma-Aldrich), washed in PBS, and were cross-linked with 1% formaldehyde (vol/vol) for 10 min at room temperature with rotation, then quenched with 125 mM Glycine. Cells were washed twice in ice-cold PBS, and pelleted by centrifugation at 600 x g at 4°C for 5 min. ChIP was performed in crosslinked cells exactly as described[77], with 10 μg anti-GFP antibody (Life Technologies, A-11122). Immunoprecipitated DNA was quantified using a Qubit dsDNA HS Assay Kit (Life Technologies). Libraries were generated using Illumina TruSeq DNA Sample Preparation Kit. 200–400 bp were size selected using AMPure XP magnetic beads. Libraries were quantified on the NanoDrop 1000 Spectrophotometer. Libraries were pooled and sequenced on the Illumina NextSeq platform, using 75bp, paired-end reads.

ChIP for Smchd1 on autosomes (i.e. this protocol does not solubilise the heterochromatin of the Xi) was performed as above with modifications. Following lysis of 4 x 10^7 female *Smchd1*$^{GFP/GFP}$ and *Smchd1*$^{+/+}$ NSCs, cells were dounced 25 times in a cold, tight dounce homogoniser. Nuclei were centrifuged for 1 minute at 12,000 rpm at 4°C, and washed in IP buffer. Cells were resuspended in 40 μL 1 x MNase buffer (New England Biolabs) with 1 x BSA (New England Biolabs) for every million cells. Nuclei were incubated at 37°C for 5 minutes. 500 U MNase (New England Biolabs)

was added per million cells, and nuclei were incubated at 37°C for 5 minutes. EDTA was added to a final concentration of 10 mM, and nuclei were incubated on ice for 10 minutes. Nuclei were centrifuged for 1 minute at 12,000 rpm at 4°C, resuspended in 130 μL IP buffer, and subjected to sonication using the Covaris S220 sonicator (peak power: 125, duty factor: 10, cycle/burst: 200, duration: 15 s). The lysate was cleared by centrifugation for 10 minutes at 12,000g at 4°C. Smchd1 was then immunoprecipitated, and chromatin purified as above.

ChIP for histone proteins was performed in female $X_{129}^{XistAA}X_{Castaneus}$ *Smchd1*$^{del/del}$ and *Smchd1*$^{del/fl}$ NSCs, crosslinked as above. 1 x 10^6 cells were used per ChIP. Nuclei were extracted by washing three times in nuclear isolation buffer [20mM Tris pH 8.0, 10 mM NaCl, 2 mM EDTA pH 8.0, 0.5% Igepal CA-630, 1X c0mplete protease inhibitor (Roche)] pelleted each time by centrifugation at 1500 rpm for 5 minutes at 4°C. Nuclei were resuspended in sonication buffer [20 mM Tris pH 7.5, 150 mM NaCl, 2 mM EDTA, 1 % Igepal CA-630, 0.3% Sodium dodecyl sulfate, 1X c0mplete protease inhibitor (Roche)] and sonicated using the Covaris S220 sonicator (peak power: 105, duty factor:20, cycle/burst: 200, duration: 750 s). Sample was cleared by centrifugation at 12 000 x g for 20 minutes, and 1 volume of dilution buffer [20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 1X c0mplete protease inhibitor (Roche)] was added to cleared chromatin. Sample was pre-cleared with 20 μL of Dynabeads Protein G (Life Technologies) blocked with 0.1 % Bovine Serum Albumin (BSA) for 2 hours. BSA was added to pre-cleared chromatin to a final concentration of 0.1 %. 1 % chromatin was taken as input. Immunoprecipitaion was performed overnight at 4°C with rotation using 2 μg antibody (anti-H3K27me3 antibody from Millipore 07-499 , anti-H3K27ac antibody from Abcam Ab4729). Immunoprecipitatied samples were captured by incubation with 20 μL Dynabeads Protein G (Life Technologies) blocked with 0.1 % BSA for 2 hours. Beads were then washed twice each in wash buffer 1 [20 mM Tris pH 8.0,

150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.15% SDS], wash buffer 2 [20 mM Tris pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS], wash buffer 3 [20 mM Tris pH 8.0, 250 mM LiCl, 2 mM EDTA, 0.5% Igepal CA-630, 0.5% sodium deoxycholate], and TE buffer [10 mM Tris pH 7.5, 1 mM EDTA]. Crosslink reversal, elution, and library preparation was performed as above.

Mapping was performed using Bowtie2[78] v0.12.9. Sequencing data were analysed with the aid of the Seqmonk v1.36.1 or v1.38.2 software. For Smchd1 on the Xi, and ChIP for the histones, samples were pooled in Seqmonk, and peaks were called using the MACS style caller within the Seqmonk package (settings for 10,000 bp fragments, $p < 1 \times 10–05$) using the whole cell extract as the control. For the Smchd1 ChIP peaks used in the in situ HiC analysis, peaks were called using MACS2 with down sampling and a $p < 0.001$, and the WCE samples used as the control. Enrichment analysis for histone domains was performed by determining the percentage coverage over 1 Mb bins along the female X chromosome. Smchd1 domains were called using Enriched Domain Detector (EDD) software[79] v1.1.18, using the following parameters: max_CI_value = 0.25, required_fraction_of_informative_bins = 0.98, p_hat_CI_method = agresti_coull, log_ratio_bin_size = 10.

Seqmonk browser tracks were produced by quantifying regions over 1000 kb bins, sliding by 500 bp each time for histone marks, or 10,000 kb sliding by 5 kb each time for Smchd1, then normalised for library depth with Match normalisation, and the appropriate H3 or isotype control sample subtracted for each IP. ChIP-seq heatmaps over ATAC-seq peaks were generated using SeqMonk's built in Heatmap function.

ATAC-seq

ATACseq was performed exactly as described in[80]. ATAC-seq libraries were subject to to 75 bp, paired-end sequencing on the Illumina NextSeq platform. Reads were aligned to mm10 using Bowtie2 v2.2.5 with the parameter -X 2000[78]. Mitochondrial reads were removed using Samtools, and duplicates were removed with Picard[81]. MACS2 peak calling was performed using v2.1.0.20140616 the following parameters: --bdg --nomodel --shift -75 --extsize 150 --qvalue 0.05[69]. ATAC-seq data were analysed using Seqmonk v1.36.1. Reads under MACS2 peaks were quantified, normalised for library size. Differential ATAC-seq peak analysis was performed using edgeR statistics in Seqmonk. The GREAT tool was used to assess differential ATAC-seq peaks in respect to TSSs[82]. Differential ATAC-seq peaks were compared with H3K4me3-enriched regions, H3K4me1 enriched regions, Ctcf binding sites, RNA polymerase II binding sites and predicted enhancer sites identified in E14.5 mouse brain in Shen *et al*.[40]. These genomic co-ordinates of these features was transferred from mm9 to mm10 for analysis. Hypergeometric probabilities were calculated using the phyper function in R.

Reduced Representation Bisulphite Sequencing (RRBS)

DNA was isolated from $X_{129}^{XistAA}X_{Castaneus}$ *Smchd1*$^{del/del}$, *Smchd1*$^{del/fl}$, *Smchd1*$^{del/+}$ and *Smchd1*$^{fl/+}$ NSCs using a DNeasy Blood and Tissue Kit (Qiagen). RRBS libraries were made from 100 ng DNA using the Ovation RRBS Methyl-Seq System (Nugen) as per manufacturers' instructions. Libraries were quantified using using a Qubit dsDNA Assay Kit (Life Technologies) for sequencing on the Illumina NextSeq platform[83]. Libraries were subject to 75 bp, paired-end sequencing. Sequencing reads were mapped to a bisulphite-converted version of the n-masked mm10 genome described

above which was created using Bismark v0.13.0. Adapter trimming was performed with TrimGalore v0.4.0, and 7 5' most bases and 2 3' most bases were trimmed based on m-Bias results from Bismark[84]. Reads were then split into either *Mus musculus castaneus* or *Mus musculus domesticus* using SNPsplit. Methylation calls were made using the Bismark Methylation Extractor[54]. Analysis was then performed in Seqmonk v1.38.2. Only CpG sites covered by more than 10 reads were considered for analysis.

Western Blotting

Cells were treated with trypsin-EDTA and washed twice in cold MTPBS. Cells were lysed, western blots performed on lysates, and membranes visualised exactly as described[17].

Whole Mount Skeletal Staining

Whole mount skeletal staining of *Smchd1*$^{MommeD1/MommeD1}$ *Smchd1*$^{MommeD1/+}$ and *Smchd1*$^{+/+}$ E17.5 embryos was performed exactly as described[85]. Embryos were stored in 100% glycerol until imaged on a ZEISS SV11 Stereomicroscope.

Antibodies

For Western blotting for Smchd1, we used an in house produced rat monoclonal antibody raised against the Smchd1 ATPase domain, used at 1:1000 dilution. Anti-Actin antibody for Western blot was from Santa Cruz (H1015), used at 1:5000 dilution. For Smchd1-GFP ChIP, we used Life a Technologies antibody (A111-22), using 10 μg antibody per ChIP. For H3K27me3 ChIP, we used a Millipore antibody (07-449), 2 μg per ChIP. For H3K27ac ChIP, we used an Abcam antibody

(Ab4729), 2 μg per ChIP. For H3K27me3 immunofluorescence, we used the same Millipore antibody (07-449) at 1:100 dilution. For H2AK119ub immunofluorescence, we used an antibody from Cell Signalling Technology (#8240) at a 1:100 dilution. For immunofluorescence, the secondary antibody used was anti-Rabbit-488 (Life Technologies, A21206), at 1:500 dilution.

Data availability

All genomics data can be found in the Gene Expression Omnibus, under GSE111726.

Source Data for for Figures 3 and 6 are available with the paper online.

Supplementary Data Sets 1-13 are available online.

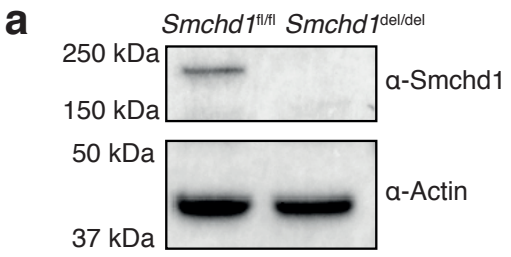
All other data is available upon request.

References

64. Majewski, I.J. et al. Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. *PLoS Biol* **6**, e93 (2008).
65. Chaumeil, J., Augui, S., Chow, J.C. & Heard, E. Combined immunofluorescence, RNA fluorescent in situ hybridization, and DNA fluorescent in situ hybridization to study chromatin changes, transcriptional activity, nuclear organization, and X-chromosome inactivation. *Methods in molecular biology (Clifton, N.J.)* **463**, 297-308 (2008).
66. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676-682 (2012).
67. Gilles, J.F., Dos Santos, M., Boudier, T., Bolte, S. & Heck, N. DiAna, an ImageJ tool for object-based 3D co-localization and distance analysis. *Methods* **115**, 55-64 (2017).
68. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research* **4**, 1310 (2015).
69. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).
70. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* **9**, 999-1003 (2012).
71. Marco-Sola, S., Sammeth, M., Guigo, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat Methods* **9**, 1185-8 (2012).
72. Le Dily, F. et al. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev* **28**, 2151-62 (2014).
73. Needleman, S.B. & Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**, 443-53 (1970).

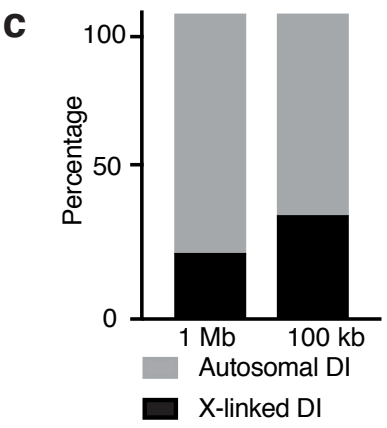
74. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**, R36-R36 (2013).
75. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
76. Wu, H. et al. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* **329**, 444-448 (2010).
77. Nelson, J.D., Denisenko, O. & Bomsztyk, K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nature protocols* **1**, 179-185 (2006).
78. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359 (2012).
79. Lund, E., Oldenburg, A.R. & Collas, P. Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Research* **42**, e92-e92 (2014).
80. Buenrostro, J., Wu, B., Chang, H. & Greenleaf, W. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **109**, 21.29.1-21.29.9 (2015).
81. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
82. McLean, C.Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology* **28**, 495-501 (2010).
83. Yin, D. et al. High concordance between Illumina HiSeq2500 and NextSeq500 for reduced representation bisulfite sequencing (RRBS). *Genom Data* **10**, 97-100 (2016).
84. Krueger, F. & Andrews, S.R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-2 (2011).
85. Rigueur, D. & Lyons, K.M. Whole-mount skeletal staining. *Methods in molecular biology (Clifton, N.J.)* **1130**, 113-121 (2014).

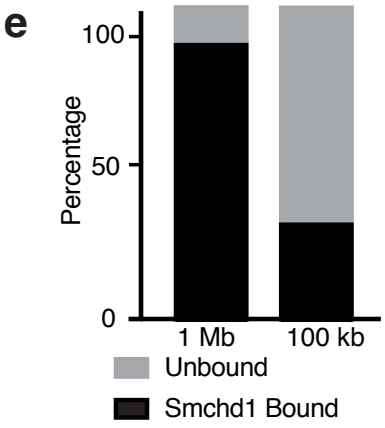
Figure 1

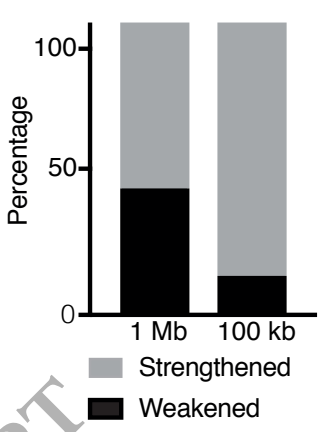


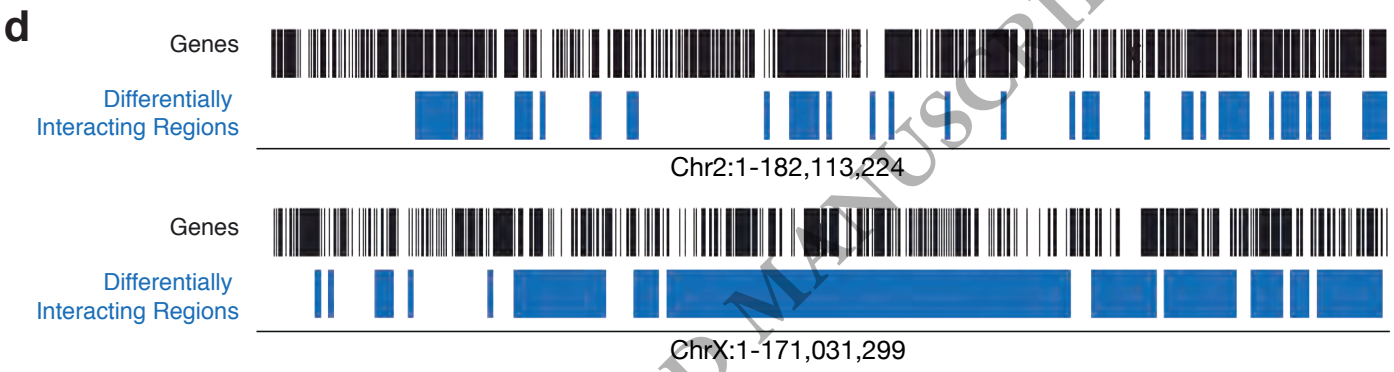
b

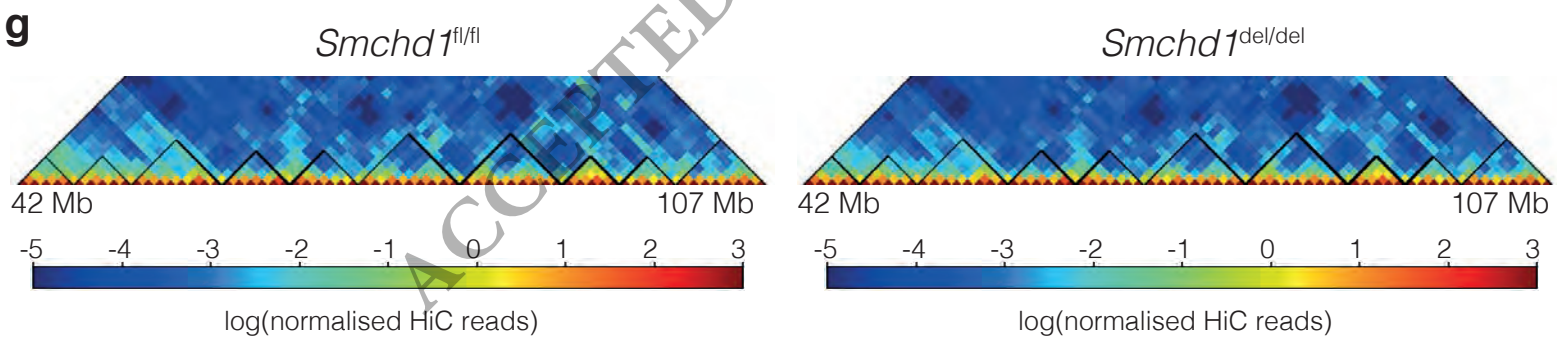
Summarised DiffHiC data	1 Mb resoution	100 kb resolution
DIs (FDR<0.05)	2382	643
Mean distance weakened DIs	21.2 ± 20 Mb	684 ± 774 kb
Mean distance strengthened DIs	3.6 ± 4.9 Mb	270 ± 547 kb

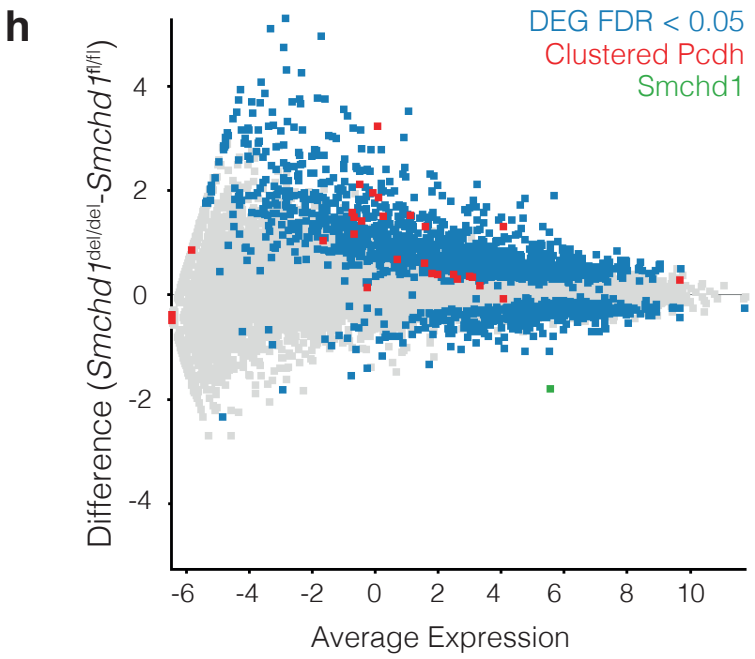












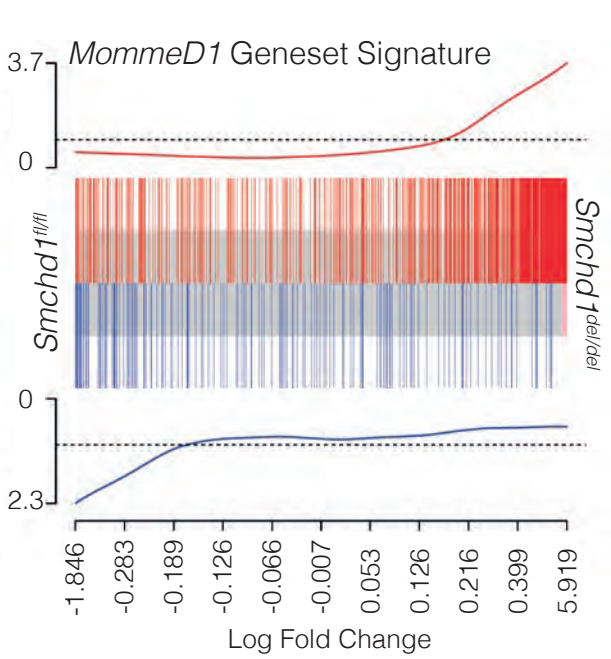


Figure 2

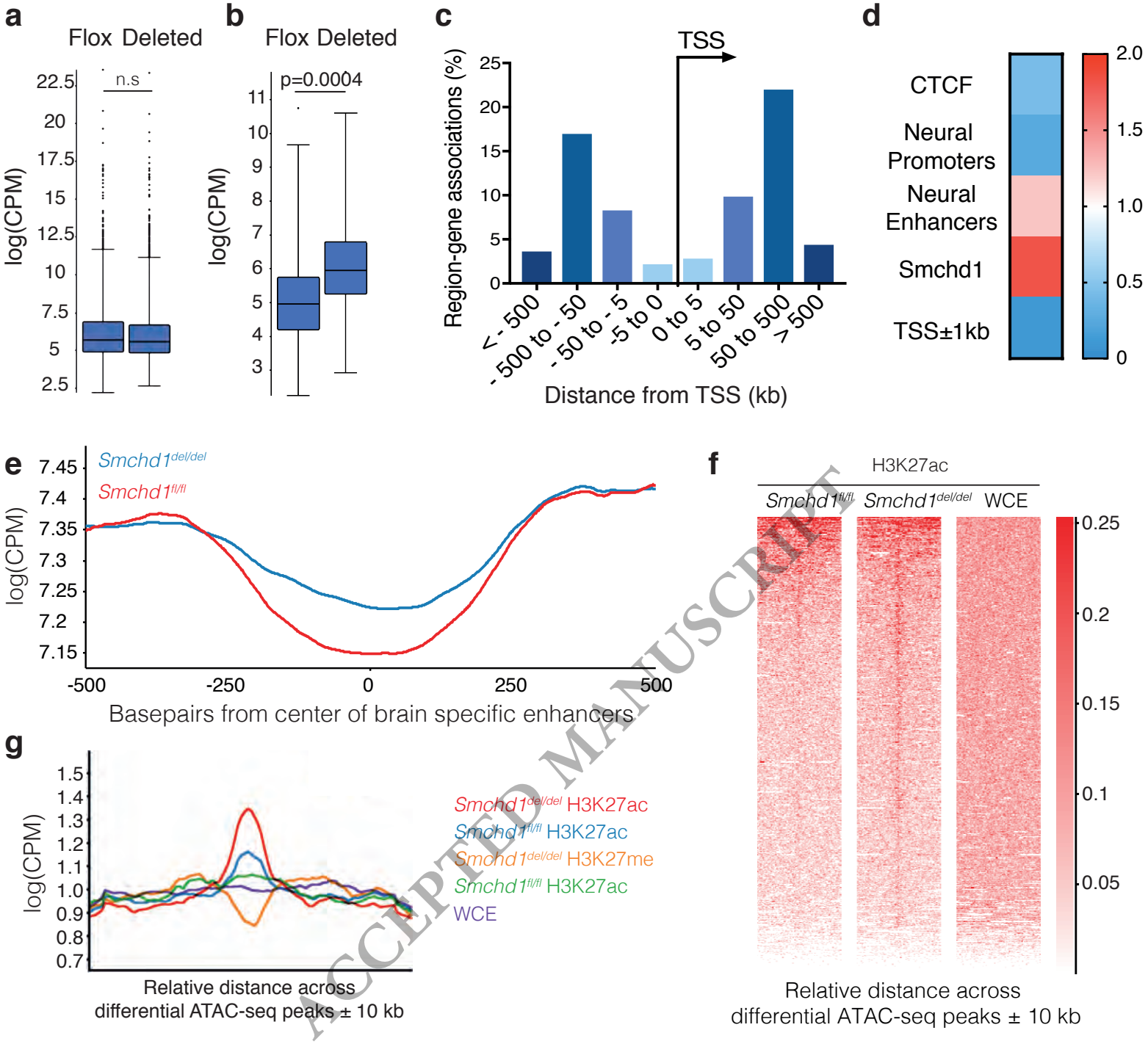
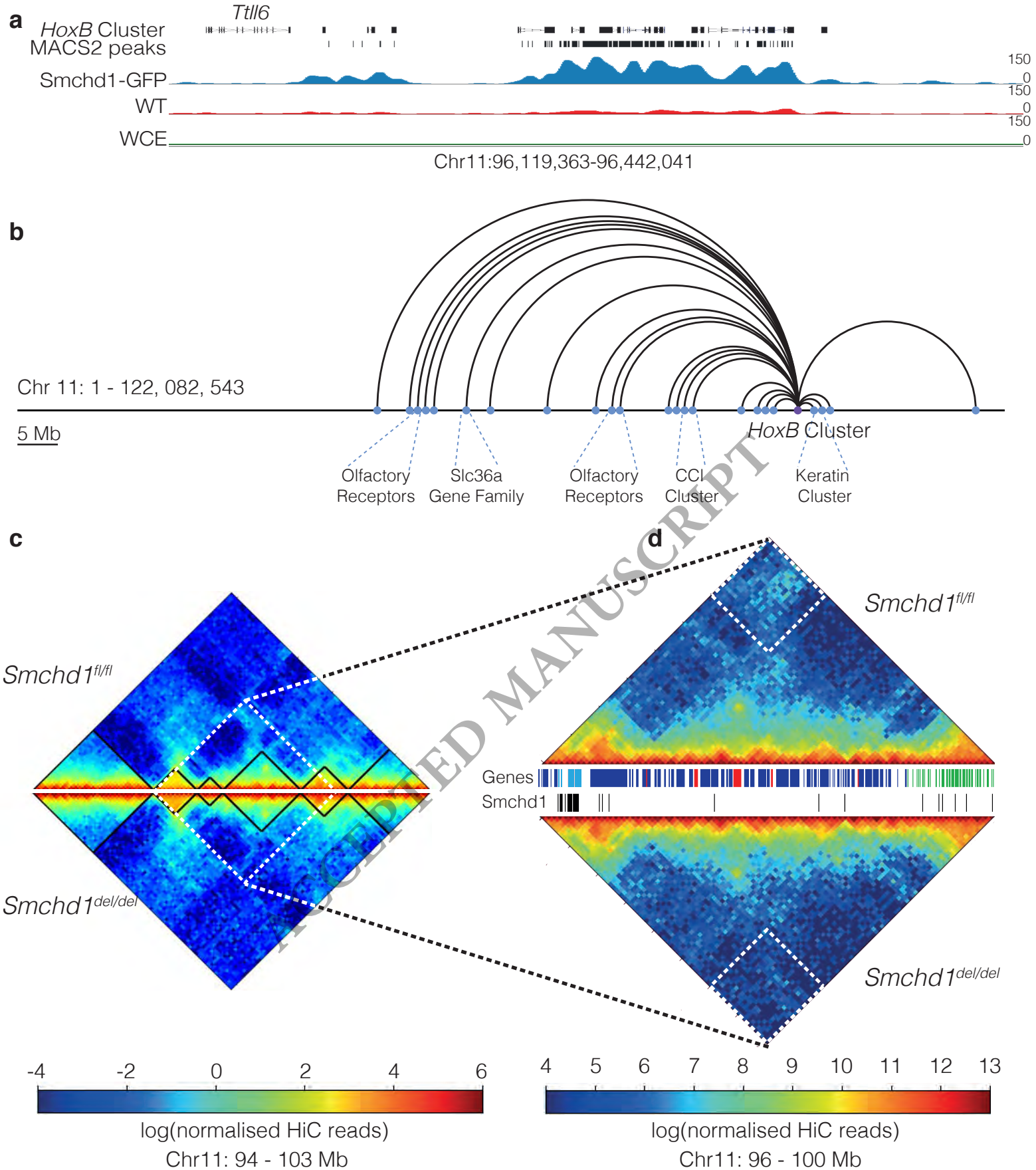
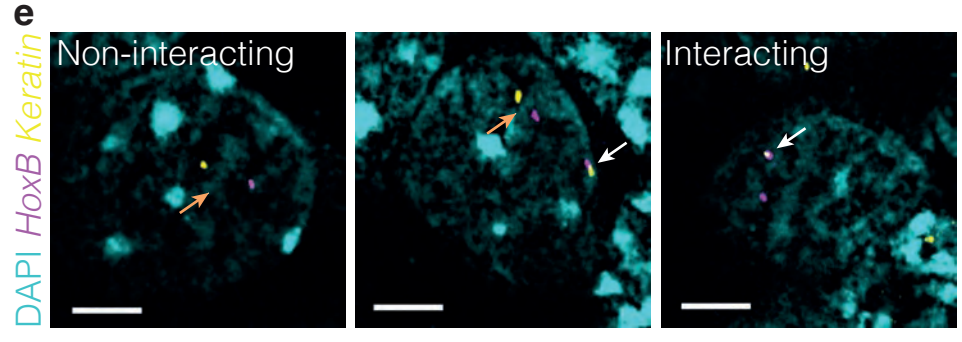


Figure 3

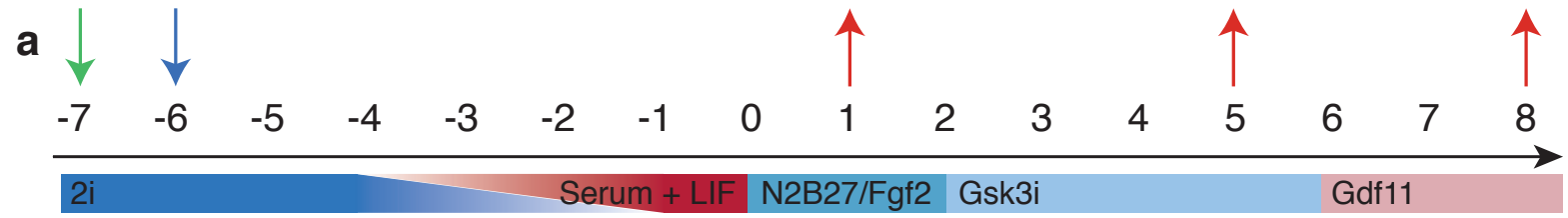


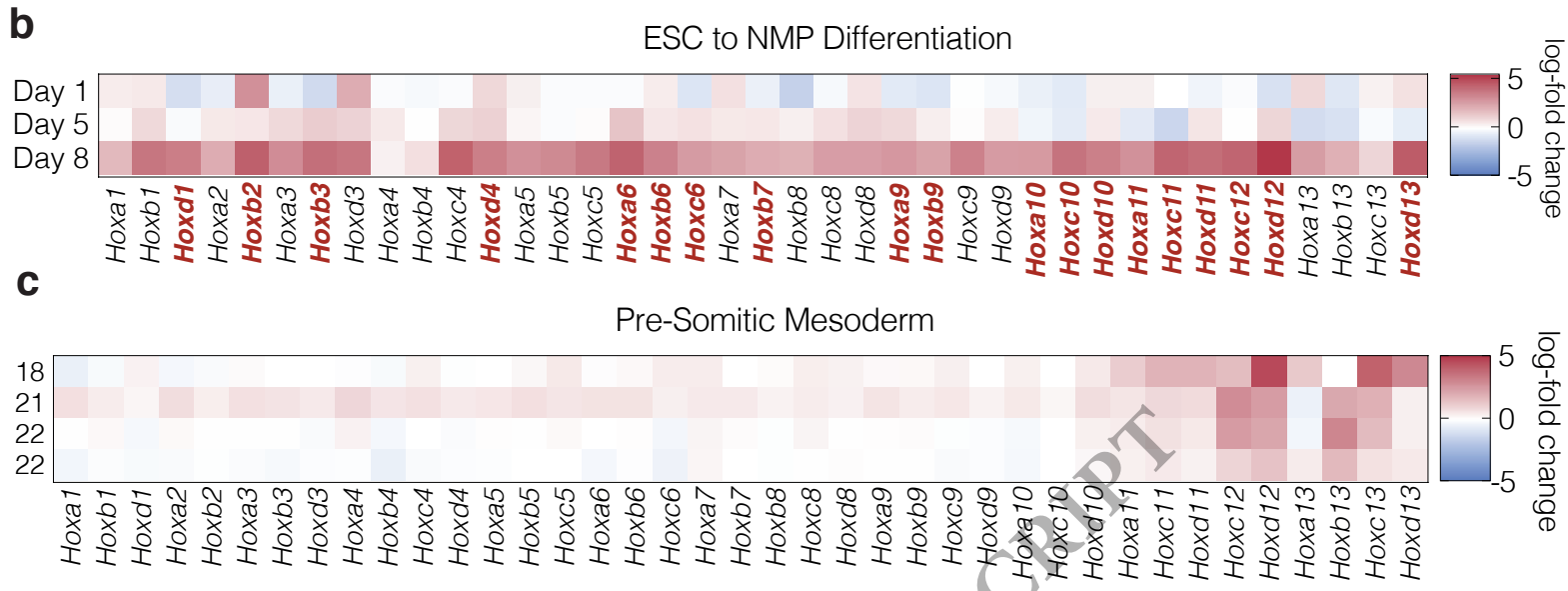


f

	Interactions	Alleles Scored
Smchd1$^{+/fl}$	20	146
Smchd1$^{del/del}$	4	148

Figure 4





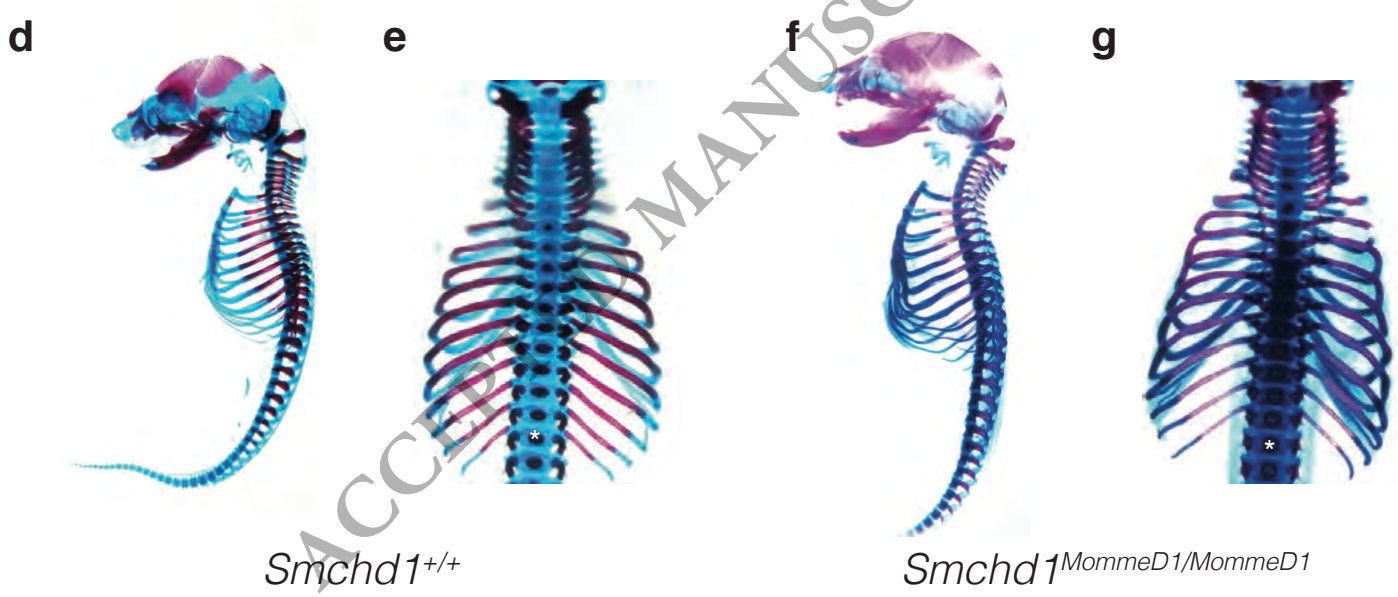


Figure 5

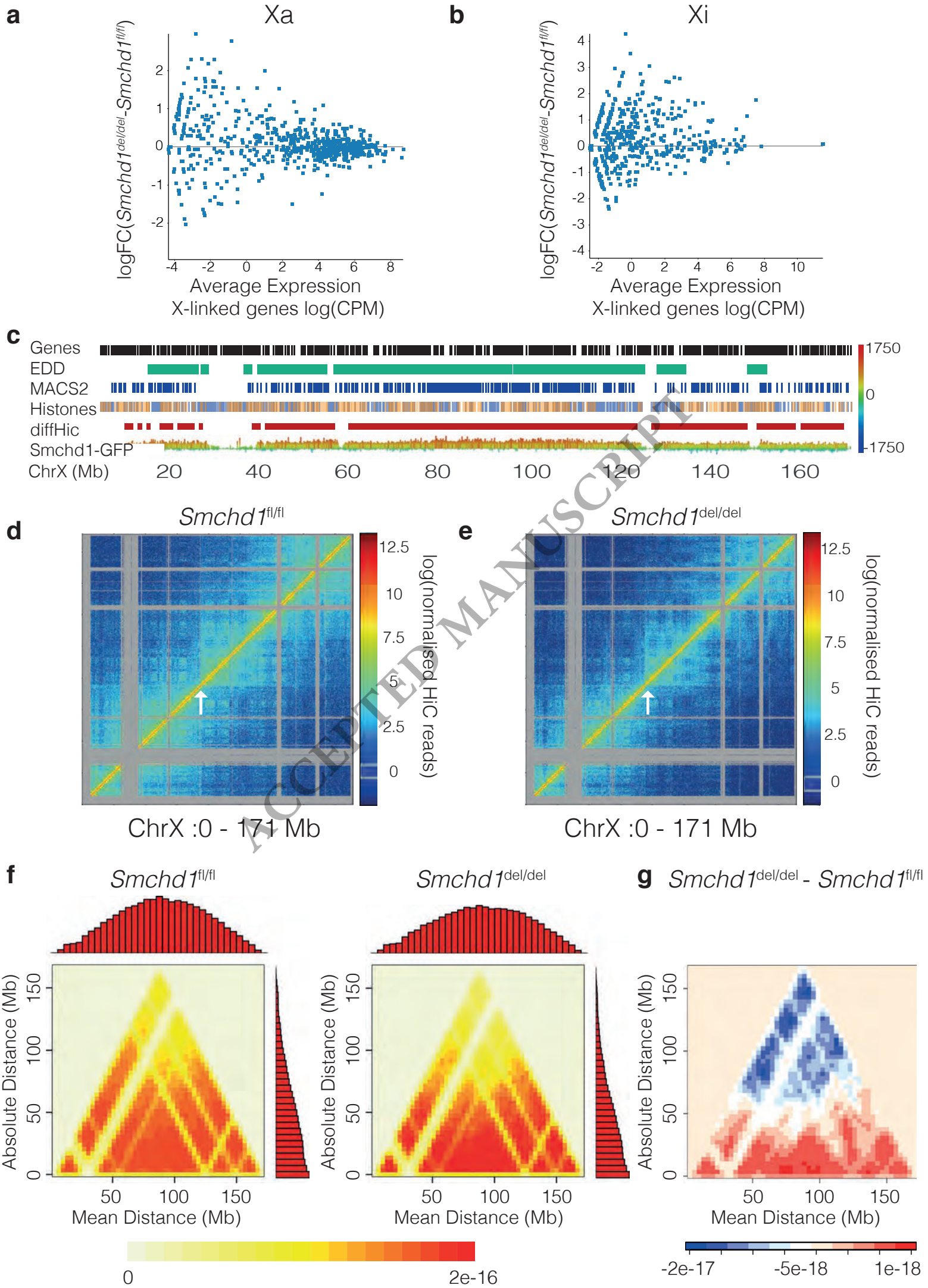


Figure 6

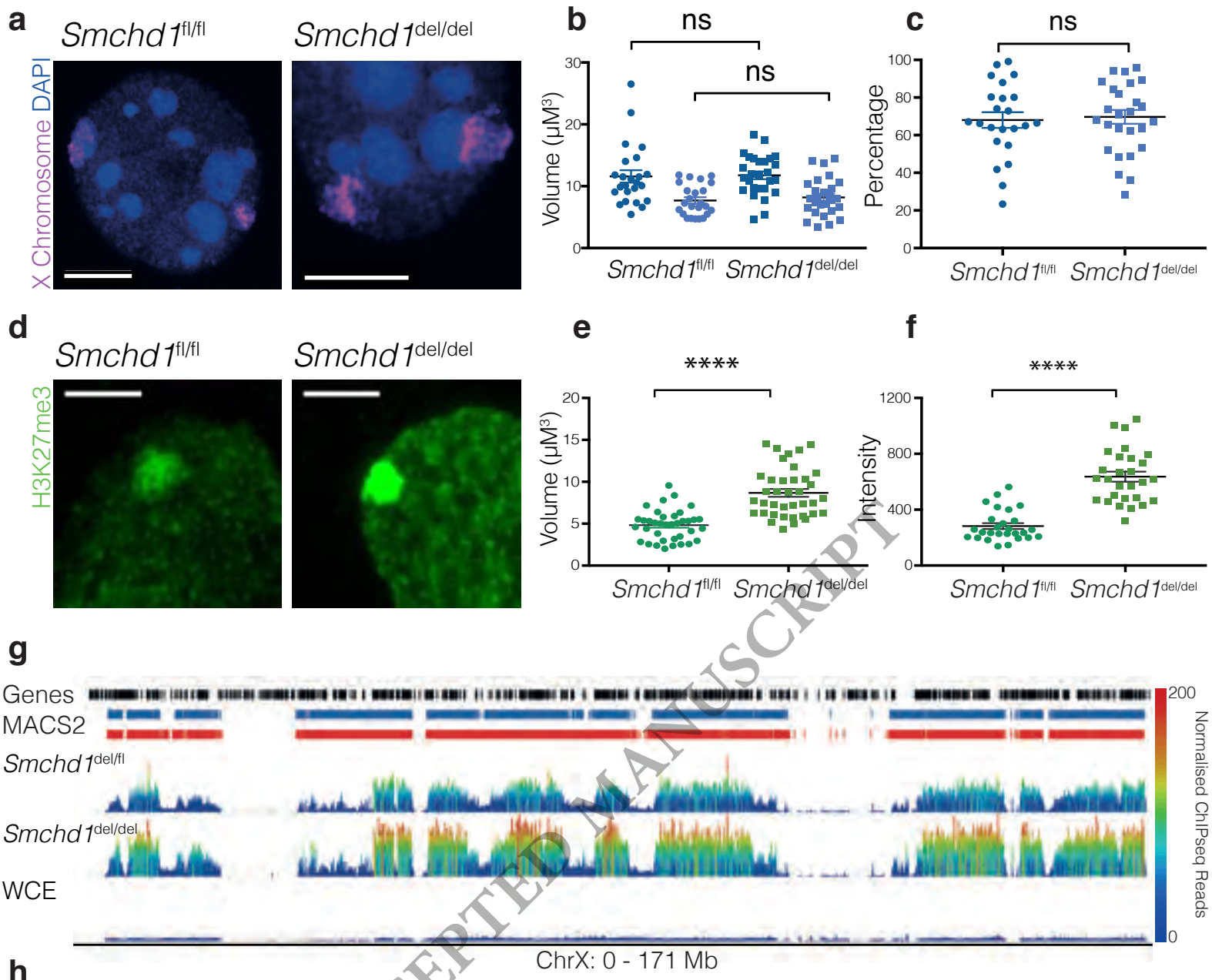


Figure 7