SMCHD1 has separable roles in chromatin architecture and gene silencing that could be targeted in disease

Andres Tapia del Fierro[1,2], Bianca den Hamer[3], Natasha Jansz[1,2], Kelan Chen[1,2], Tamara Beck[1], Hannah Vanyai[4], Natalia Benetti[1,2], Alexandra D Gurzau[1,2], Lucia Daxinger[5], Shifeng Xue[6,7], Thanh Thao Nguyen Ly[6,7], Iromi Wanigasuriya[1,2], Kelsey Breslin[1], Harald Oey[5], Yvonne Krom[3], Dinja van der Hoorn[3], Linde F Bouwman[3], Matthew E. Ritchie[1,2], Bruno Reversade[7,8,], Fabrice Prin[4], Timothy Mohun[4], Silvère M van der Maarel[3], Edwina McGlinn[9,10], James M Murphy[1,2], Andrew Keniry[1,2], Jessica C de Greef[3], Marnie E Blewitt[1,2]*

*Corresponding author, blewitt@wehi.edu.au

1. The Walter and Eliza Hall Institute of Medical Research, Melbourne VIC Australia

2. The Department of Medical Biology, University of Melbourne, Melbourne VIC Australia

3. Leiden University Medical Center, Leiden Netherlands

4. Crick Advanced Light Microscopy Facility, The Francis Crick Institute, London UK

5. Queensland Institute of Medical Research, Brisbane QLD Australia

6. Department of Biological Sciences, National University of Singapore, Singapore

7. Institute of Molecular and Cell Biology, A*STAR, Singapore

8. Genome Institute of Singapore, A*STAR, Singapore

9. EMBL Australia, Monash University, Clayton, VIC Australia.

10. Australian Regenerative Medicine Institute, Monash University, Clayton, VIC Australia.

Abstract

The interplay between 3D chromatin architecture and gene silencing is incompletely understood. Here, we report a novel point mutation in the non-canonical SMC protein SMCHD1 that enhances its silencing capacity at endogenous developmental targets and at the facioscapulohumeral muscular dystrophy associated macro-array, D4Z4. Heightened SMCHD1 silencing perturbs developmental *Hox* gene activation, causing a homeotic transformation in mice. Paradoxically, the mutant SMCHD1 appears to enhance insulation against another epigenetic regulator complex, PRC2, while depleting long range chromatin interactions akin to what is observed in the absence of SMCHD1. These data suggest that SMCHD1's role in long range chromatin interactions is not directly linked to gene silencing or insulating the chromatin, refining the model for how the different levels of SMCHD1-mediated chromatin regulation interact to bring about gene silencing in normal development and disease.

Introduction

The importance of correct epigenetic regulation to normal development and differentiation has long been known. Modern genomic techniques have revealed sophisticated mechanisms behind epigenetic control, where DNA methylation, post-translational histone modifications and chromatin conformation come together in a dynamic fashion to regulate gene expression. However, the complex interplay between different modes of epigenetic control, often between those considered to have opposing function (like bivalent chromatin[1]), has precluded straightforward interpretations. One approach to understanding such a complex biological problem to unravel the factors involved in the process using unbiased genetic screens[2]. Indeed, genetic screening approaches have identified novel epigenetic regulators in yeast[3], plants[4], flies[5; 6] and mammals.

We previously reported a sensitised *in vivo* screen in mice that paired N-ethyl-N-nitrosourea (ENU) mutagenesis with a variegating GFP transgene array to identify modifiers of transgene variegation, and therefore epigenetic regulation[7]. This screen led to the discovery of the *Smchd1* gene that encodes an epigenetic repressor, since shown to play a role in X-chromosome inactivation[8; 9], silencing of clustered gene families such as select imprinted clusters[10-12], the clustered protocadherins[10; 11; 13] and *Hox* genes[14]. By understanding more about SMCHD1 we can also learn more about how gene silencing works in each of these cases.

In addition to the developmental roles of SMCHD1 elegantly characterised in mice, SMCHD1 is also relevant to human disease. Heterozygous pathogenic variants have been found in the developmental disorders Bosma-arrhinia and microphthalmia syndrome[15; 16] (BAMS) and facioscapulohumeral muscular dystrophy[17; 18] (FSHD). BAMS is a craniofacial malformation syndrome and FSHD is a typically adult-onset debilitating progressive muscular dystrophy. FSHD is caused by the death of skeletal muscle cells due to aberrant expression of the germline and cleavage stage transcription factor DUX4. DUX4 is expressed when epigenetic silencing is relaxed by one of two distinct molecular mechanisms that both result in the same clinical outcome[19]. The *DUX4* gene is located within the *D4Z4* macrosatellite repeat on chromosome 4q35, which is normally comprised of 8-100 D4Z4 tandem repeat units. FSHD type 1 is caused by repeat contraction to 1-10 units[20], wherein the number of repeat units is insufficient to trigger efficient silencing of the whole repeat array. FSHD type 2 is most commonly caused by loss-of-function mutations in SMCHD1, which results in de-repression of the *D4Z4* repeat array[17]. Patients who carry both SMCHD1 mutations and repeat contraction present with more severe disease, suggesting that SMCHD1 silencing activity is relevant to both subtypes of the disease[18], and that FSHD patients reside on a disease spectrum[21]. With SMCHD1 now known to contribute to at least two human diseases, it is of high interest to determine how SMCHD1 works and thus how its activity may be manipulated to treat disease.

SMCHD1 is a member of the structural maintenance of chromosomes (SMC) family. It interacts with chromatin through its hinge domain, which also enables homodimerisation[13; 22; 23]. Unlike the canonical SMC proteins, SMCHD1 is non-canonical in part because it possesses a GHKL ATPase domain[24; 25]. The pathogenic variants identified in each of these two domains of the protein suggest that they are critical for SMCHD1 function. In BAMS, heterozygous missense variants restricted to the extended ATPase domain underlie disease[15; 16; 26]. By contrast, FSHD loss-of-function mutations occur across the length of the protein[27; 28]. The differing mutation types and locations suggest that while loss of SMCHD1 function causes FSHD, in BAMS it may be due to altered protein function, in some cases mediated by gain of ATPase activity[15; 26]. However, a comprehensive explanation as to why variants in SMCHD1 can cause such disparate phenotypes is currently unclear as one pathogenic variant has been identified in both BAMS and FSHD[29].

Recent work has begun to reveal both how SMCHD1 functions at, and is recruited to, the chromatin. For the inactive X chromosome, SMCHD1 recruitment is dependent on the polycomb repressive complex 1 (PRC1) catalysed mark, histone 2A lysine 119 ubiquitination[30; 31] (H2AK119ub). Somewhat like the canonical SMC protein complex, Cohesin, upon recruitment to target loci, SMCHD1 is involved in mediating long-range chromatin interactions[14; 32; 33]. SMCHD1 also appears to insulate against the occupancy and effect of other epigenetic regulators, such as CTCF and polycomb repressive complex 2 (PRC2)[13; 14; 32; 33]. To date, our understanding of SMCHD1 function has largely been informed by the study of systems upon complete loss of SMCHD1 protein. While these studies have been informative, they are limited in that they reflect a complete loss-of-function scenario of the SMCHD1 protein and therefore cannot appropriately model all aspects of disease that arise due to *SMCHD1* perturbation.

In this study, we report a new *Smchd1* missense mutant that was identified in the same ENU mutagenesis screen that led to the original discovery of *Smchd1*. In this screen, mutations were named Modifiers of Murine Metastable Epialleles (Mommes)[7; 34], with the original report detailing the *Smchd1* loss-of-function *MommeD1* mutation[7]. Here we describe the effects of the *MommeD43 Smchd1* mutation on gene expression, development and chromatin architecture. These studies reveal that the *MommeD43* variant produces a gain-of-function effect on the expression of critical SMCHD1 targets, including in a mouse model of FSHD where the *MommeD43* variant can partially rescue *DUX4* silencing. Interestingly, our data on chromatin architecture and insulation suggest that SMCHD1's role in regulating long range chromatin interactions is not required for silencing and is divorced from its role in chromatin insulation. Therefore, this study suggests an attractive starting point for SMCHD1 modulation in FSHD treatment and expands our understanding of how the multiple layers of SMCHD1-dependent chromatin regulation interact to elicit epigenetic silencing.

Results

MommeD43 is a *Smchd1* mutant with increased transgene array silencing activity

In our ENU mutagenesis screen looking for epigenetic modifiers of transgene silencing, we identified a mutant line of mice, called MommeD43. The system used a transgene array of 11 units containing a GFP gene in a tandem repeat, directed to express in erythrocytes[35] (Fig. 1a). The *MommeD43* mutation caused a significant decrease in the proportion of erythrocytes expressing the GFP transgene as measured by flow cytometry, indicating enhanced silencing (Fig. 1a). The mutation was mapped using positional cloning in conjunction with linkage analysis to mouse chromosome 17. Further refinement by whole exome sequencing revealed a cytosine to adenosine mutation in exon 15 of the *Smchd1* gene, translating as an Alanine 667 to Glutamic acid conversion (A667E) in the extended ATPase domain of SMCHD1[25] (Fig. 1b). Interestingly, *MommeD43* has the opposite effect on transgene silencing to the nonsense mutation in *Smchd1* identified in the MommeD1 strain[7; 8]. These data suggest that *MommeD43* could be a gain-of-function mutation in *Smchd1*.

Homozygous *Smchd1* null animals show complete female embryonic lethality due to failure of X chromosome inactivation and reduced viability in males, dependent on strain background[8; 11]. By contrast, $Smchd1^{MommeD43/+}$ and $Smchd1^{MommeD43/MommeD43}$ mice of both sexes are viable and fertile, with no observed transmission ratio distortion of each genotype from heterozygous intercrosses (Supplementary Fig. 1).

To explore how the *MommeD43* mutation in *Smchd1* may influence SMCHD1 function, we first considered whether the mutation results in increased levels of SMCHD1 protein. We found no differences in SMCHD1 protein levels when measured by western blot or flow cytometry (Fig. 1c, Supplementary Fig. 2). We then tested whether the *MommeD43* mutation resulted in altered chromatin occupancy by chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq). Lacking antibodies against SMCHD1 that were of high enough quality for ChIP-seq, we developed a mouse line carrying the *MommeD43* mutation by CRISPR/Cas9 editing of our previously described *Smchd1-GFP* fusion mouse model[14], allowing immunoprecipitation of SMCHD1 with an antibody to GFP. Both are used interchangeably for the remainder of this study. We performed anti-GFP ChIP-seq in *Smchd1-GFP*$^{+/+}$ and *Smchd1-GFP*$^{MommeD43/MommeD43}$ primary female neural stem cells (NSCs, a cell type in which we had extensive data on *Smchd1*[14]). We found very high correlation of SMCHD1 levels around (±5kb) known SMCHD1 binding sites (2840 peaks[14]) between the wild-type and *MommeD43* mutant cells (Fig. 1e, Pearson R>0.99), such as the *Hoxb* cluster (Fig. 1d). These results suggest that *MommeD43* does not alter chromatin localization of SMCHD1. In sum, these data indicate that the *MommeD43* mutation does not alter the levels of SMCHD1 protein, nor its binding to target loci, suggesting that the mutation may instead alter protein function.

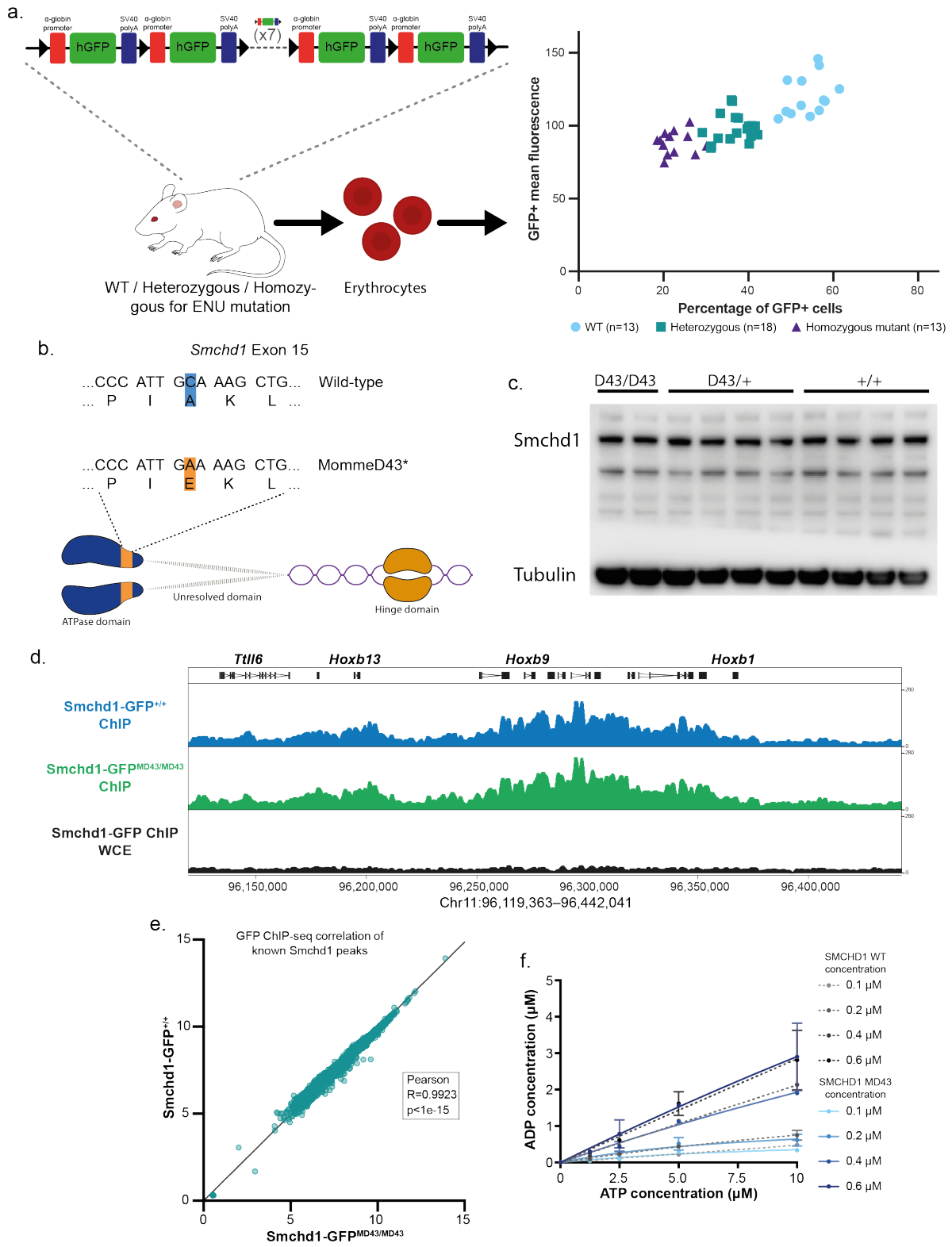


Figure 1. *MommeD43* is a *Smchd1* mutant with increased transgene array silencing activity

a. Diagram of ENU screen experiment (left) and results (right). Erythrocytes from mice with an 11-unit GFP transgene repeat array carrying heterozygous or homozygous *MommeD43* mutations and wild-type littermate controls were analyzed by FACS to measure GFP transgene expression levels. b. Schematic representation of murine Smchd1 (resolved ATPase and hinge domains linked by flexible still unresolved domain) and the location of the MommeD43 mutation in its structure (orange). c. Western blot of SMCHD1 in *Smchd1*$^{MommeD43/MommeD43}$, *Smchd1*$^{MommeD4/+}$ and *Smchd1*$^{+/+}$ cells showing no noticeable change in SMCHD1 levels. d. ChIP-seq genome browser tracks of the *Hoxb* cluster locus showing GFP ChIP in primary NSCs with endogenous SMCHD1-GFP fusion protein. This region is heavily marked by both Smchd1 and *MommeD43* does not alter localization of either. On top are the name of a few genes in the locus for reference. e. Scatter plot of log2-transformed normalized GFP ChIP-seq counts in *Smchd1*$^{GFP/GFP}$ and *Smchd1*$^{GFP\text{-}MommeD43/MommeD43}$ NSCs around (±5kb) previously published peaks[14]. The Pearson coefficient indicates very high positive correlation showing no noticeable changes in SMCHD1 DNA binding sites. f. ATPase assay using recombinant purified wild-type murine SMCHD1 extended ATPase domain (grayscale) compared with the *MommeD43* mutant equivalent (blue). *MommeD43* abbreviated to *MD43.*

The *MommeD43* mutation does not alter the conformation or activity of the extended ATPase domain

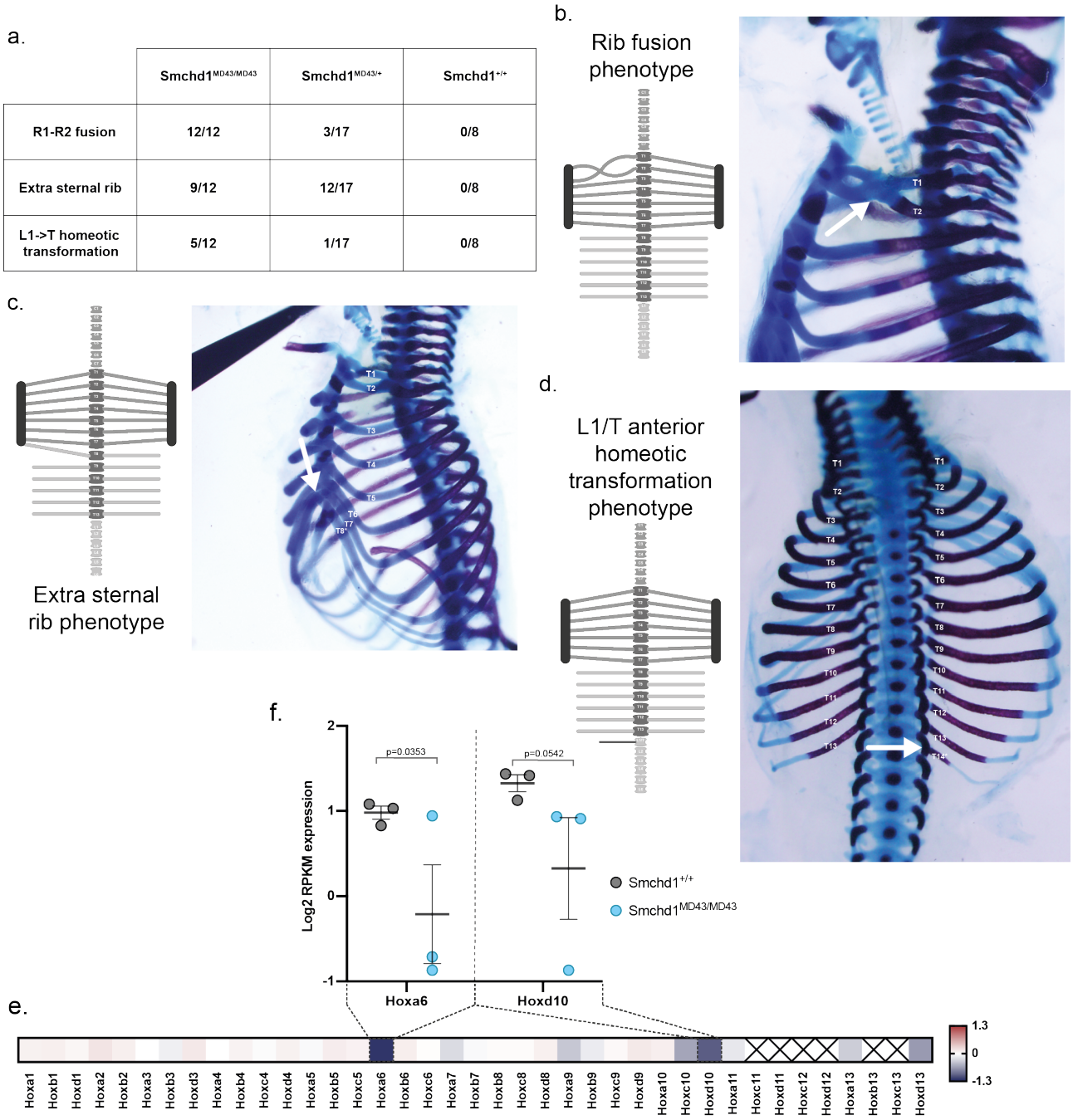
The *MommeD43* mutation is located in the extended ATPase domain of SMCHD1, which contains a catalytically active GHKL ATPase[24; 25]. To test if the *MommeD43* mutation alters the conformation of this region, we analysed the structure of this domain using a recombinant construct to express the extended ATPase domain, as we have previously done[25]. The *MommeD43* mutant protein had no detectable change in ATPase activity (Fig. 1f). Using small-angle X-ray scattering (SAXS), we determined that the *MommeD43* mutant extended ATPase domain has the same conformation as the wild-type protein (Supplementary Fig. 3, Supplementary Table 1). Therefore, we cannot attribute altered SMCHD1 function to a change in enzymatic activity or conformation. However, both of these results were obtained by purifying the recombinant extended ATPase domain and therefore do not exclude the possibility that the mutation may confer changes to ATPase activity or conformation in the context of the full-length protein.

***MommeD43* has a gain of function effect on *Hox* gene silencing and skeletal development**

We next investigated the effect of the *MommeD43* mutation in development. SMCHD1 has a known role in *Hox* gene regulation and subsequent skeletal development[14]. Loss of *Smchd1* causes a posterior homeotic transformation at thoracic vertebra 13 (T13), consistent with the observed failure in posterior *Hox* gene silencing. To determine the effects of MommeD43 on skeletal morphology, we examined whole-mount skeletal preparations from embryonic day (E) 17.5 embryos. We found that the *MommeD43* mutation in *Smchd1* resulted in an anterior homeotic transformation with additional effects on rib formation. We divided these effects into three distinct and independent phenotypes (Fig. 2a, detailed scoring in Supplementary Fig. 4). First, a fusion of the ribs of the first two thoracic elements was observed with complete penetrance in homozygous mutants (*Smchd1*$^{MommeD43/MommeD43}$, 12/12 embryos analysed) and partial penetrance in heterozygous mutants (*Smchd1*$^{MommeD43/+}$, 3/17; Fig. 2b). The second, partially-penetrant phenotype was an extra sternal rib attachment, where the rib from the 8th thoracic element connected to the sternum instead of being the first false rib, as in wild-type embryos (9/12 in *Smchd1*$^{MommeD43/MommeD43}$, 12/17 in *Smchd1*$^{MommeD43/+}$, Fig. 2c). The last phenotype found in some embryos was the presence of a well-formed rib from the vertebral element posterior to T13 (5/12 in *Smchd1*$^{MommeD43/MommeD43}$, 1/17 in *Smchd1*$^{MommeD43/+}$, Fig. 2d). In all cases this expansion of thoracic count was accompanied by a reduction in lumbar element number from 6 to 5, with no change in overall total vertebral number of the animal, indicating this phenotype represents an anterior homeotic transformation. Collectively, these phenotypes point to a dysregulation of *Hox* gene expression and are consistent with *MommeD43* being a gain of function mutation in *Smchd1*.

To investigate *Hox* dysregulation directly, we dissected tailbud tissue that harbours progenitors of the vertebral column from *Smchd1*$^{+/+}$ and *Smchd1*$^{MommeD43/MommeD43}$ embryos at 8 somites (E8.5) and performed RNA-sequencing. In contrast to the upregulation of posterior *Hox* genes observed in *Smchd1* null tail buds at E9.5 found previously[14], here we see a mild decrease in multiple posterior *Hox* genes (Fig. 2e, Supplementary Table 2). We saw the largest effect on *Hoxa6* expression at the 8-somite stage (log2FC -1.3; n=3 pairs of biological replicates), where the *Smchd1*$^{MommeD43/MommeD43}$ embryos had a decrease in expression, suggesting that *MommeD43* is causing a slight delay in the activation of *Hoxa6.* Interestingly, this delay may be sufficient to explain the T1-T2 rib fusion phenotype, as an identical phenotype was observed in *Hoxa6* null embryos[36], while T3-T4 rib fusion

is observed upon *Hox6* paralogous deletion[37]. In a similar manner, the L1->T homeotic transformation might be explained by an activation delay of the second most affected gene, *Hoxd10*. *Hox10* paralogous deletion causes a lumbar to thoracic homeotic transformation in all lumbar vertebrae, but deletion of all paralogues except for one allele of *Hoxa10* causes a bilateral L1->T homeotic transformation of the first lumbar element[38]. Given the upregulation of posterior *Hox* genes observed in *Smchd1* null tail buds at E9.5 found previously[14], these data are consistent with the *MommeD43* mutation causing a gain of function effect on SMCHD1.



	Smchd1$^{MD43/MD43}$	Smchd1$^{MD43/+}$	Smchd1$^{+/+}$
R1-R2 fusion	12/12	3/17	0/8
Extra sternal rib	9/12	12/17	0/8
L1->T homeotic transformation	5/12	1/17	0/8

Figure 2. *MommeD43* has a gain of function effect on *Hox* gene silencing and skeletal development

a. Scoring of the three observed skeletal phenotypes in *Smchd1*$^{MommeD43/MommeD43}$ and *Smchd1*$^{MommeD43/+}$ embryos. b, c, d. Diagrams of the skeletal defects (left) of the three distinct phenotypes observed in *Smchd1*$^{MommeD43/MommeD43}$ E16.5 embryos with corresponding representative images (right). The white arrows point to the defect. e. Heatmap of the log2-fold change between somite-paired *Smchd1*$^{MommeD43/MommeD43}$ and *Smchd1*$^{+/+}$ E8.5 embryos to account precisely for developmental stage. The values used are the mean expression of 3 biological replicates measured by RNA-sequencing. f. Plot of the normalized log2 expression values for the two Hox genes most affected by MommeD43. Each circle represents an individual biological replicate. Bars are mean ± SEM. *MommeD43* abbreviated to *MD43*.

The *MommeD43* mutation does not recapitulate morphological changes observed in BAMS

BAMS is a rare human craniofacial malformation syndrome that causes partial or total loss of the nose and sense of smell, along with reduced eye size and other head morphological abnormalities (MIM:603457). It is caused by heterozygous missense mutations in SMCHD1's extended ATPase domain, many of which are proven or predicted gain-of-function mutations[15; 26], although this remains a matter of debate[16; 39]. Since the effects of *MommeD43* on SMCHD1 on transgene silencing, *Hox* gene expression and skeletal patterning are consistent with a gain-of-function, we examined craniofacial morphology and development in the *MommeD43* mutants.

We collected E14.5 embryos and analysed them by High Resolution Episcopic Microscopy (HREM). After reconstructing a high-resolution three-dimensional model of the embryos' morphology, we quantitated key craniofacial measurements (Fig. 3a-h) and found minimal differences. A slight widening of the nasal capsule in *Smchd1*$^{MommeD43/+}$ embryos was the only parameter found to be significantly altered in mutants. Thus, we decided to test if *MommeD43* might affect gene expression during the development of the nose and snout. For this purpose, we dissected the frontonasal prominence (FNP, Fig. 3i) of somite-matched pairs of *Smchd1*$^{MommeD43/MommeD43}$ and *Smchd1*$^{+/+}$ E10.5 embryos and performed RNA-sequencing. Using duplicate somite-matched pairs at 29 somites, we found 56 differentially expressed genes (False Discovery Rate (FDR)<0.05, sex chromosomes excluded, Fig. 3j). Of these, 53 were downregulated in *Smchd1*$^{MommeD43/MommeD43}$ embryos which is consistent with the *MommeD43* mutation conferring better silencing capacity to SMCHD1 (Supplementary Table 3). Gene Ontology analysis showed that 7 of the 11 most significantly affected biological processes are directly related to development (Fig. 3k, corrected p-value<$6x10^{-9}$). These data suggest that while the *MommeD43* mutation does not cause gross craniofacial abnormalities, it likely influences developmental gene expression programs that could be captured in the RNA-sequencing analysis. Interestingly, when the same region is harvested from *Smchd1*$^{MommeD1/MommeD1}$ embryos, we observe upregulation of a restricted set of genes representing known SMCHD1 targets e.g. clustered protocadherins and genes from the *Snrpn* imprinted cluster (Supplementary Table 4), rather than the set of developmental genes downregulated in the *MommeD43* samples.

We have previously used *SMCHD1* cDNA microinjection in *Xenopus* to assess the effect of BAMS mutations in SMCHD1 on craniofacial development (Supplementary Figure 5a). In this system, BAMS mutations result in a smaller eye phenotype, independent of whether we could detect an increase in ATPase activity in recombinant protein[15; 26]. Here we compared the effect of the *MommeD43* mutation introduced into the human cDNA with that of a BAMS variant, W342S, with wild-type SMCHD1 and uninjected tadpoles as controls. While W342S mutant SMCHD1 resulted in a smaller eye, the *MommeD43* SMCHD1 behaved exactly as wild-type (Supplementary Fig. 5b,c). Taken together with the mouse embryology and the normal nasal morphology, these data suggest that the *MommeD43* variant does not accurately model BAMS in mice or frogs.

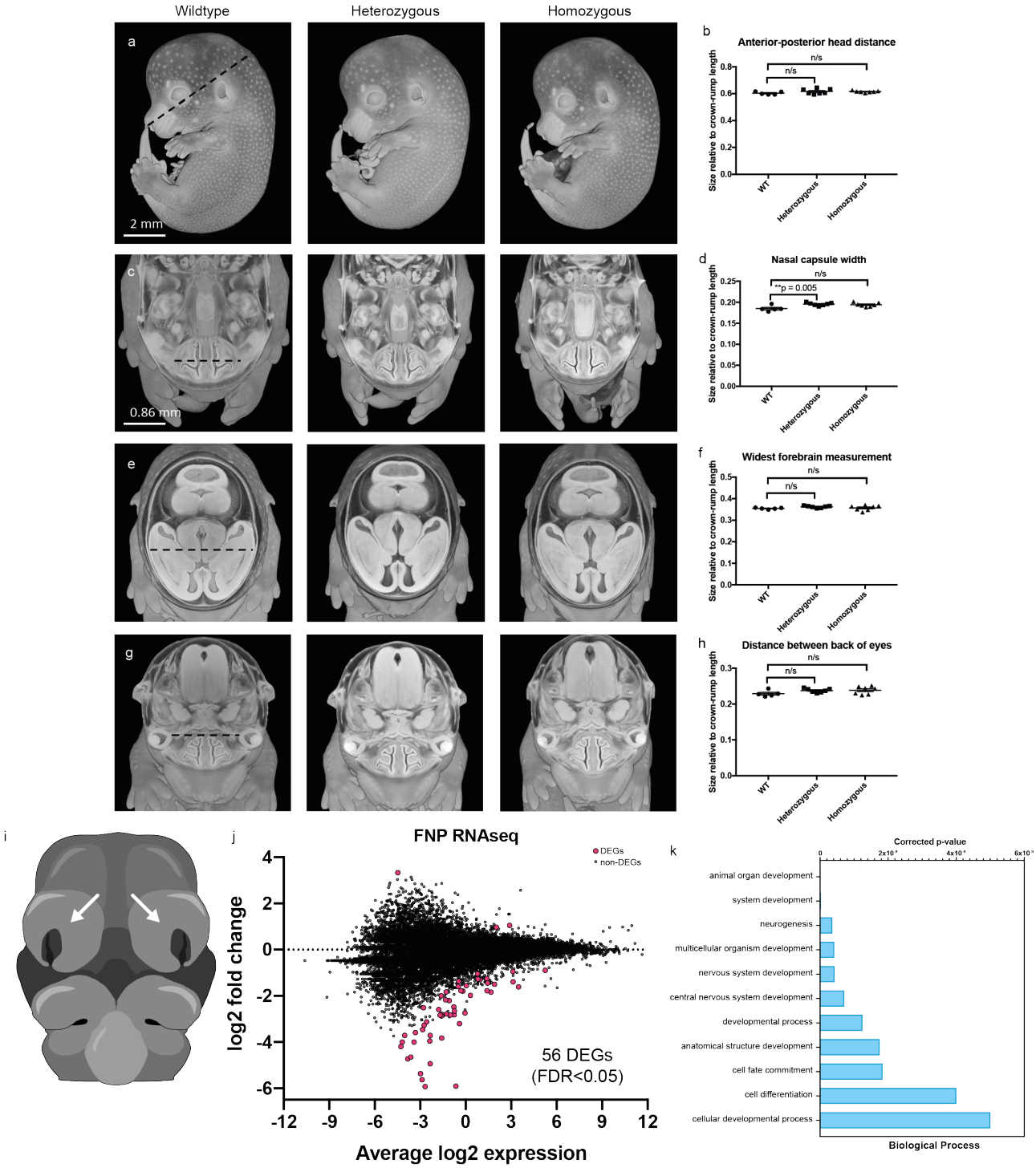


Figure 3. The *MommeD43* mutation alters gene expression in the frontonasal prominences but does not recapitulate morphological changes observed in BAMS

a-h : Cutaways of three-dimensional renderings of E14.5 embryos imaged by HREM oriented to measure various craniofacial features (a, c, e, g) and graphs detailing the measurements normalised to embryo crown-rump length (b, d, f, h). n/s = not significant, **p = significant adjusted p-value. Scale bar in A = 2 mm; scale bar in C = 0.86 mm and relates to E and G. i. Diagram of E10.5 embryo. Arrow points to the FNP collected for RNA sequencing. j. MD plot of log2 fold change of normalized RPKM counts of gene expression in FNP tissue from 2 somite-matched pairs of *Smchd1*$^{MommeD43/MommeD43}$ and *Smchd1*$^{+/+}$ E10.5 embryos showing 53 downregulated and 3 upregulated genes by *MommeD43* (56 total DEGs, FDR<0.05). k. Gene Ontology pathway analysis of the 11 main biological processes affected by the 37 uniquely mapped genes recognized by the GO platform out of the 56 DEGs shown in j. Bars show p-values corrected for multiple testing. *MommeD43* abbreviated to *MD43*.

***MommeD43*'s improved repeat-silencing capability offers therapeutic potential for FSHD**

Since SMCHD1 is required for proper *DUX4* silencing, and the *D4Z4* repeat has a structure somewhat reminiscent of the GFP transgene array used in the ENU mutagenesis screen, we next examined whether *MommeD43* would also provide more efficient silencing of *DUX4*. To test this hypothesis, we crossbred *Smchd1*$^{MommeD43/+}$ mice with hemizygous *D4Z4-2.5* mice carrying a human transgene consisting of 2.5 *D4Z4* repeat units cloned from genomic DNA of an FSHD1 individual[40] (Fig. 4a). This approach was necessary because *D4Z4* is a primate-specific repeat[41]. We previously showed that murine SMCHD1 represses the human *D4Z4* transgene in these mice[42]. Thus, the D4Z4-2.5 mouse model is suitable to evaluate the effect of *MommeD43* on *DUX4* expression and on the chromatin structure of the transgene.

The *D4Z4-2.5/Smchd1*$^{MommeD43/+}$ mice exhibited Mendelian genotype and sex distributions (Supplementary Figure 6a-b). They appeared healthy up to at least two months of age when they were sacrificed for this study. Males showed a slight increase in weight and females were comparable to the other measured genotypes (Supplementary Fig. 6c).

To determine whether *MommeD43* provides more efficient silencing of *DUX4 in vivo* than wild-type SMCHD1, we first quantified the transcript levels of *DUX4*, *Wfdc3* (a murine target of human DUX4[40]), and *Smchd1* in three different skeletal muscles of *D4Z4-2.5/Smchd1*$^{+/+}$ and *D4Z4-2.5/Smchd1*$^{MommeD43/+}$ mice (gastrocnemius, quadriceps, triceps). *DUX4* expression levels were low (Cq values varied between 34 and 38, while Cq values for *Gapdh* and *Rpl13a* were around 14 and 18, respectively), as previously reported for D4Z4-2.5 mice[42], and comparable between *D4Z4-2.5/Smchd1*$^{+/+}$ and *D4Z4-2.5/Smchd1*$^{MommeD43/+}$ mice (Supplementary Fig. 7a). In line with this, we observed no changes in *Wfdc3* transcript levels (Supplementary Fig. 7b). As expected, we saw no change on *Smchd1* transcript levels (Supplementary Fig. 7c) or SMCHD1 protein levels (Supplementary Fig. 7d).

DUX4 expression is typically higher in muscle cell cultures and non-muscle tissues of D4Z4-2.5 mice than muscle tissue from the same animals. So, we turned to muscle cell cultures established from the extensor digitorum longus (EDL) muscle and several non-muscle tissues (cerebellum, heart, spleen, thymus). *DUX4* transcript levels were significantly decreased in both proliferating myoblast cultures and differentiating myotube cultures, in cerebellum and in spleen of *D4Z4-2.5/Smchd1*$^{MommeD43/+}$ mice relative to samples from *D4Z4-2.5/Smchd1*$^{+/+}$ animals (Fig. 4b, myotube differentiation confirmed by *Mef2c* levels, Supplementary Fig. 7k). In line with this, *Wfdc3* levels were also lower in EDL cultures established from *D4Z4-2.5/Smchd1*$^{MommeD43/+}$ mice as well as in spleen (Fig. 4c). SMCHD1 expression and protein levels were not affected in any of the studied cells or tissues (Fig. 4d, Supplementary Fig. 7c,d,g,h).

Although murine *Wfdc3* was previously reported to be a human DUX4 target, it was not verified in all tissues[40]. *Wfdc3* levels were unchanged in cerebellum, which might suggest tissue-dependent responses to *DUX4*. We observed no changes in expression levels of either *DUX4* or *Wfdc3* in heart or thymus (Supplementary Fig. 7e,f). Finally, we found that DUX4 protein levels were significantly decreased in myoblast and myotubes cultures (Fig. 4e). We could not measure DUX4 protein levels in other tissues, since the currently available anti-DUX4 antibodies do not work in tissues.

In humans, silencing of the D4Z4 repeat is achieved by DNA methylation and repressive chromatin marks such as histone 3 lysine 9 trimethylation (H3K9me3) and histone 3 lysine 27 trimethylation (H3K27me3)[43-45]. Since D4Z4 hypomethylation is observed in all tissues tested in carriers of a pathogenic variant in SMCHD1, we tested D4Z4 methylation in tail DNA, as we previously did for

D4Z4-2.5 mice with a *Smchd1* nonsense mutation[42] and found D4Z4 repeat hypomethylation. In tail DNA we observed no difference in DNA methylation levels at the DR1 or FasPas regions (Supplementary Fig. 7i), that are well characterised regions within and just distal to the the *D4Z4* repeat[46; 47] respectively. Next, we measured the permissive chromatin mark H3K4me2 and the repressive histone marks H3K9me3 and H3K27me3 in both fibroblast cultures and spleen tissue. Although we did not observe any significant differences, H3K4me2 levels were slightly decreased in *D4Z4-2.5/Smchd1*$^{MommeD43/+}$ mice compared with *Smchd1* wild-type counterparts, while H3K9me3 and H3K27me3 levels were higher. As a result, the chromatin compaction score (H3K9me3 level corrected for H3K4me2) was significantly increased in the *MommeD43* samples (Fig. 4f). This score is reduced in fibroblasts and myoblasts of individuals with FSHD[48] and in fibroblasts of D4Z4-2.5 mice with the *Smchd1*^{MommeD1} (null) mutation[42].

Our results suggest that *MommeD43* affects *DUX4* levels in specific tissues, perhaps by modulating the chromatin structure of the *D4Z4* repeat transgene. *MommeD43*'s more efficient silencing of DUX4 is not due to altered SMCHD1 binding to the transgene (Supplementary Fig. 7j), consistent with our genome-wide data. As mouse and human SMCHD1 are highly homologous, *MommeD43* may act as a hypermorphic variant in humans and may offer therapeutic potential for individuals with FSHD.

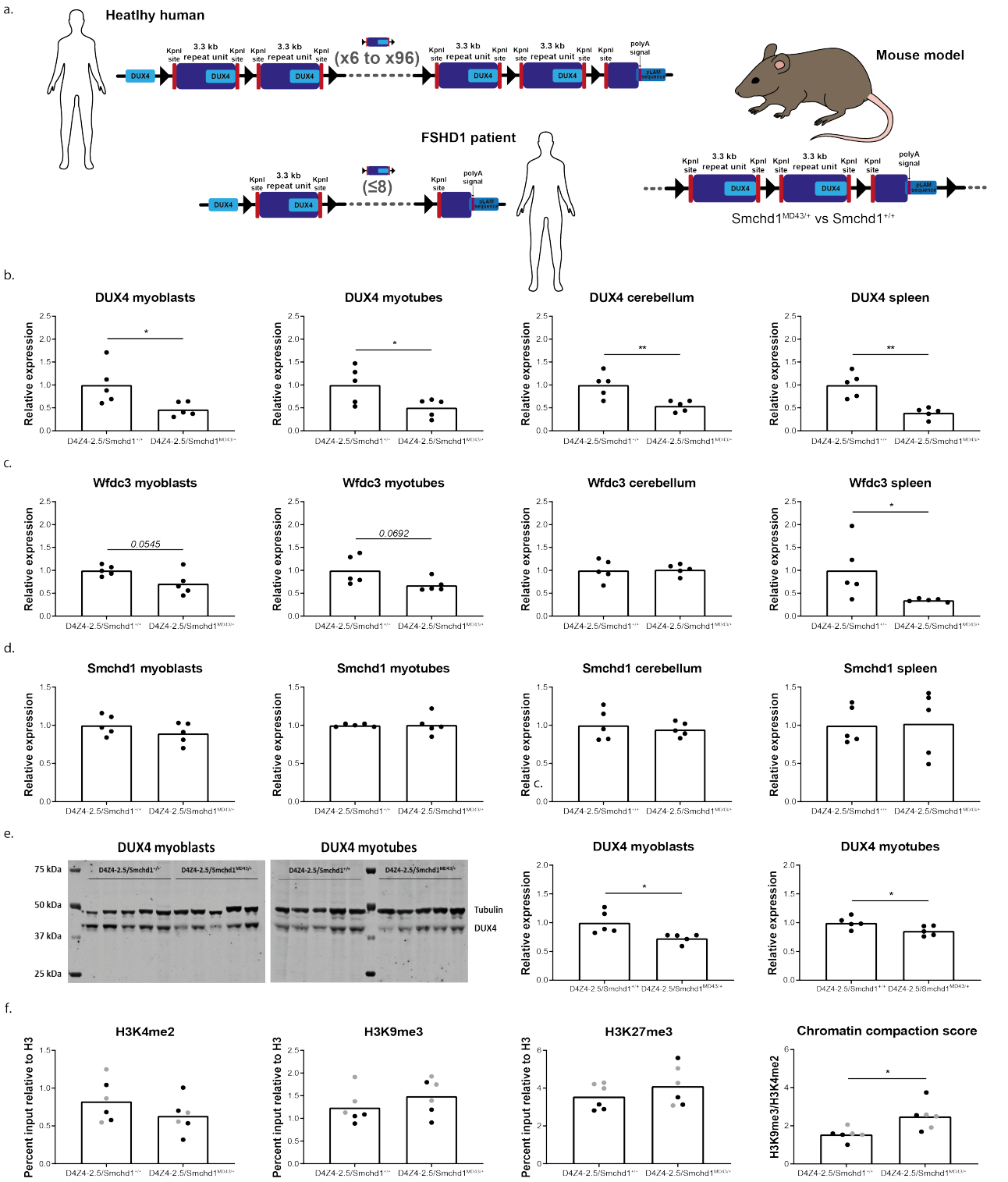


Figure 4: *MommeD43* results in improved silencing of DUX4 in a mouse model of FSHD.

a. Diagram of *D4Z4* repeat in healthy humans, FSHD1 patients and the 2.5-unit transgene repeat used in the murine model (described in [40]). b-d. Relative DUX4 transcript levels, Wfdc3 transcript levels, and Smchd1 transcript levels in myoblasts, myotubes, cerebellum, and spleen. Bars represent the average transcript levels per genotype (average value in D4Z4-2.5 tissue is set as 1); each dot represents a single mouse. Statistical analysis was performed using a Student's t-test. *P<0.05; **P<0.01. e. Relative DUX4 protein levels in myoblasts and myotubes, as measured by DUX4 and tubulin western blot. Bars represent the average transcript levels per genotype (average value in D4Z4-2.5 cells is set as 1); each dot represents a single mouse. Statistical analysis was performed using a Student's t-test. *P<0.05. f. H3K4me2 levels, H3K9me3 levels, H3K27me3 levels, and the chromatin compaction score (H3K9me3 level corrected for H3K4me2 level) in fibroblast cultures and spleen. Bars represent the average levels per genotype; each dot represents a single mouse (black dots are fibroblasts; grey dots are spleen tissue). Statistical analysis was performed using a Student's t-test. *P<0.05. *MommeD43* abbreviated to *MD43*.

The *MommeD43* mutation divorces the role of SMCHD1 in chromatin conformation from its role in gene silencing and chromatin insulation

Our data presented so far are consistent with *MommeD43* being a gain-of-function mutation. Since we have previously shown the effects of loss of SMCHD1 on gene expression, chromatin conformation, and histone marks in neural stem cells[14] (NSCs) we examined these features in *Smchd1*$^{MommeD43/MommeD43}$ NSCs to define the mechanism behind the apparent gain-of-function.

In *Smchd1* null NSCs – either *Smchd1*$^{MommeD1/MommeD1}$ derived from male mice that have never had SMCHD1 or *Smchd1*$^{del/del}$ deleted at least 7 days prior from female *Smchd1*$^{fl/fl}$ NSCs in culture – we found major changes in gene expression by RNA-sequencing (1520 DEGs FDR<0.05 in *MommeD1*[13]; 463 DEGs FDR<0.05 in *Smchd1*$^{del/del}$ cells[14]). Conversely, there were no significant changes in female *Smchd1*$^{MommeD43/MommeD43}$ NSCs in either strain background (FVB/NJ from the ENU screen and C57BL/6 Smchd1-GFP fusion created with CRISPR-Cas9). The clustered protocadherins on chromosome 18, which are well known to be regulated by SMCHD1[10; 11; 13], and whose expression is greatly increased upon SMCHD1 loss, showed interesting changes, albeit not significant. Although barely detectable, their expression is consistently reduced in *Smchd1*$^{MommeD43/MommeD43}$ cells, supporting the notion that *MommeD43* is a gain-of-function mutation (Supplementary Fig. 8).

Another known role of SMCHD1 is in mediating long-range chromatin interactions. We performed *in situ* HiC in female *Smchd1*$^{+/+}$ and *Smchd1*$^{MommeD43/MommeD43}$ NSCs (n=3 each). We have previously shown that *Smchd1* deletion caused major changes in chromatin conformation without altering topologically associated domains (TAD) borders or the distribution of A and B compartments[14]. Similarly, *MommeD43* did not shift TAD borders (Supplementary Fig. 9) or compartments, with the exception of modest differences on the X chromosome (Fig. 5a). The modest changes in chromatin architecture we observed in *MommeD43* were highly correlated to those seen in *Smchd1*$^{del/del}$ cells (r=0.64 at 1 Mb resolution, Fig. 5b-c, r=0.60 at 100 kb resolution, Fig. 5b-c), including those on the X chromosome. Of the 61 significantly different interactions (FDR<0.1, 1 Mb resolution, Supplementary Table 5), 50.8% are on the X chromosome (all strengthened in *Smchd1*$^{MommeD43/MommeD43}$), consistent with what was observed in *Smchd1*$^{del/del}$ female cells and SMCHD1's binding across the inactive X chromosome.

Many changes in the *Smchd1*$^{del/del}$ cells were observed between the *Hoxb* cluster and other clustered gene families on chromosome 11 (Fig. 5d). The most significantly weakened interaction of these was between the *Hoxb* and the keratin gene cluster, which was also noticeably weakened in *Smchd1*$^{MommeD43/MommeD43}$ cells (Fig 5. e,f). We focused on a specific interaction between the *Hoxb* cluster and an olfactory receptor gene cluster approximately 50 Mb away, which was significantly weakened upon *Smchd1* deletion in our HiC data, by using DNA Fluorescence *in situ* hybridization (DNA FISH, Fig. 5g). The results were consistent with our HiC data, showing a significant decrease in the frequency of interaction in *Smchd1*$^{del/del}$ cells (80% decrease, Fig. 5h), and a less pronounced though still significant decrease in *Smchd1*$^{MommeD43/MommeD43}$ cells (63% decrease, Fig. 5i). That same interaction was not significantly weakened in the HiC data in *Smchd1*$^{MommeD43/MommeD43}$ cells, which might be due to HiC being a high-background and low-resolution technique less suited to detecting subtle changes. From these data, *MommeD43* seems to behave as a hypomorphic allele with respect to chromatin conformation, which is in contrast to the gain of function effect we observed by all other measures of SMCHD1 function.

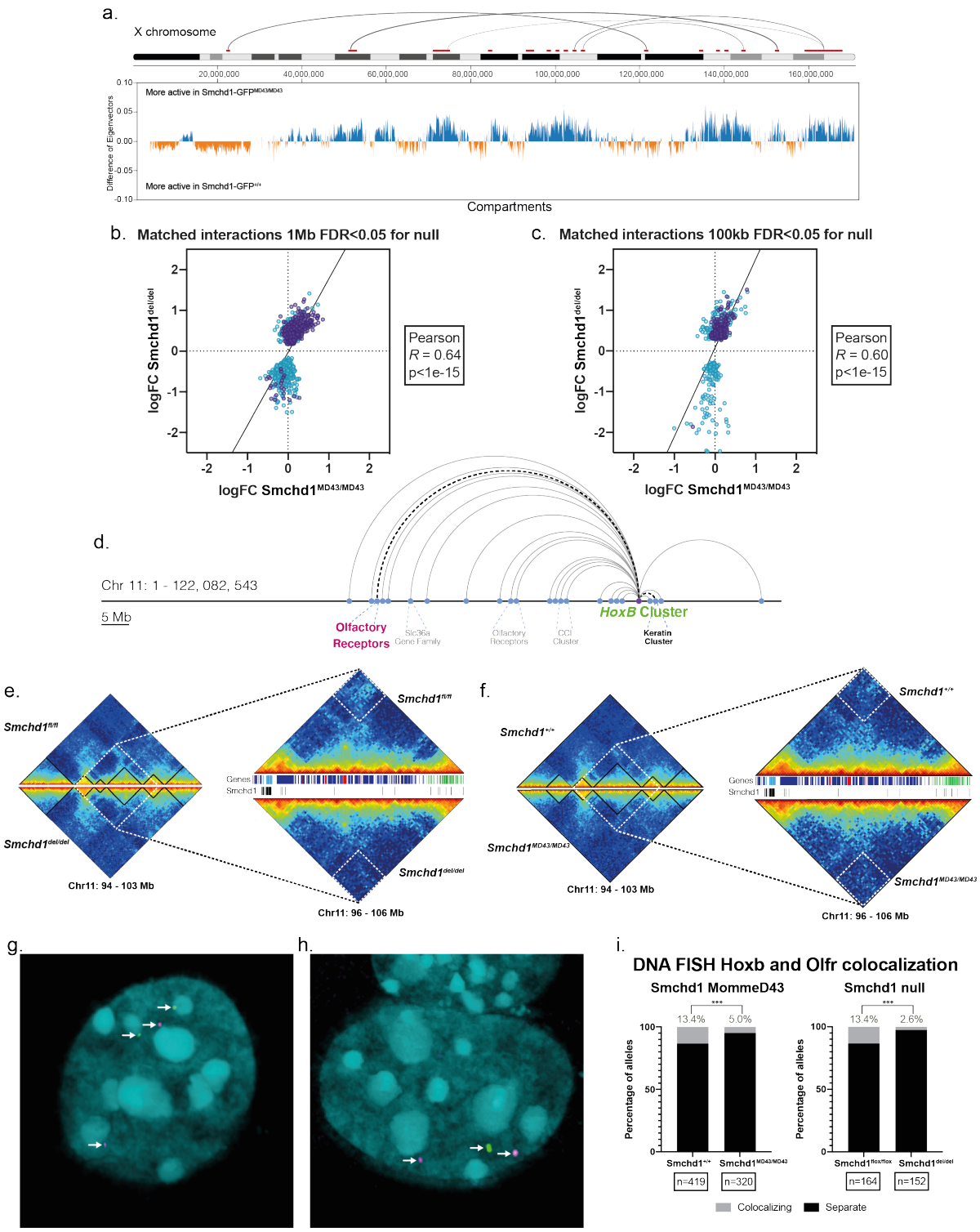


Figure 5: *MommeD43* has a hypomorphic effect on Smchd1-dependent chromatin interactions

a. Diagram of the X chromosome from IGV. On top in red are the significant (FDR<0.1) differential interactions in *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}MommeD43}$ vs *Smchd1*$^{GFP/GFP}$ Hi-C data at 1Mb resolution (all are strengthened). The grey semi-circles on top represent the top 5 interactions by log2 fold change. Underneath is a plot of the difference in Eigenvectors used to determine A and B compartments between *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}MommeD43}$ vs *Smchd1*$^{GFP/GFP}$ (calculated at 100kb resolution). b, c. Correlation plots of the differential interactions caused by loss of Smchd1 between the two datasets shown in b and d at 1Mb (b.) and 100kb (c.) resolution. Each point in the graphs represents one interaction (the link between two genomic regions of length equal to the resolution), where the x coordinate is the log-fold change of that interaction between *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}MommeD43}$ and *Smchd1*$^{GFP/GFP}$ and the y coordinate is the log-fold change between Smchd1$^{del/del}$ and Smchd1$^{fl/fl}$. X-linked interactions shown in purple. d. Diagram of the main Smchd1-dependent interactions on chromosome 11. e, f. The left side panels are heatmaps of normalized in situ Hi-C interactions in *Smchd1*$^{fl/fl}$ (e., top triangle) and *Smchd1*$^{del/del}$ (e., bottom triangle) and *Smchd1*$^{GFP/GFP}$ (f., top triangle) and *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}MommeD43}$ (f., bottom triangle) female NSCs surrounding the Hoxb cluster at 100kb resolution. The right side panels are zoomed in heatmaps of the region 96-106 Mb region at 50kb resolution. Inside the white dotted squares is the most significant interaction lost upon Smchd1 deletion, which is also reduced in *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}}$

^{MommeD43} cells. The tracks between the two triangular heatmaps show the genes in the region (blue and red bars for sense and antisense, respectively). The light blue bars represent the Hoxb genes and in green are the keratin genes whose interaction is inside the white dotted square. The bottom track (black bars) shows previously published Smchd1 peaks from ChIP-seq. g. Representative images of DNA FISH with probes labelling the *Hoxb* cluster (green) and the olfactory receptor gene cluster (magenta) highlighted in a. in the same colours (DAPI DNA stain shown in cyan). The left panel shows non-interacting loci for both alleles, the right panel shows one non-interacting and one interacting allele (colocalization of the magenta and green probes). h,i. Scoring of DNA FISH images such as the ones shown in g. in male and female NSCs of both FVB/NJ and C57Bl/6 backgrounds. The number of scored alleles for each genotype is shown at the bottom of each bar. In black is the frequency of non-interacting loci and in grey the frequency of colocalizing probes. The asterisks show statistical significance (paired t-test) of the reduction of the frequency of interaction between the two observed loci both in *Smchd1*$^{del/del}$ and *Smchd1*$^{MommeD43/MommeD43}$ cells compared to their respective controls. *MommeD43* abbreviated to *MD43*.

Other previously described effects attributable to loss of SMCHD1 during development are DNA hypomethylation at its autosomal targets and on the inactive X chromosome [10; 11; 13; 14] and increased levels and spreading of H3K27me3 on the inactive X chromosome[14]. To examine DNA methylation, we performed Reduced Representation Bisulphite Sequencing (RRBS) in female *Smchd1*$^{MommeD43/MommeD43}$ NSCs. We observed no differences in CpG island methylation levels (only 356 hypermethylated and 81 hypomethylated individual CpG nucleotides found with at least 25% difference $q<0.05$), either at autosomal regions or on the X chromosome. Similarly, H3K27me3 ChIP-seq in NSCs showed no change in peak localisation (Fig. 6a,b, Supplementary Table 6); however, this technique is ill-suited to measure relative levels of histone marks. We turned to immunofluorescence staining to assess relative levels of both SMCHD1 and H3K27me3 in both *Smchd1*$^{del/del}$ and *Smchd1*$^{MommeD43/MommeD43}$ female NSCs, in which we could examine H3K27me3 enrichment specifically on the inactive X chromosome. The X chromosome bears most of the significant alterations in chromatin interactions and shifts in compartment type observed in Hi-C data upon *Smchd1* perturbation (Fig. 5a). After acquiring high-resolution confocal Z-stack images, we defined the nuclear and inactive X volumes by using the boundaries of DAPI (DNA dye) and high H3K27me3 respectively, then used the total fluorescence intensity within each channel normalised to the volume of the region as a proxy for relative levels of SMCHD1 and H3K27me3 on the inactive X (Fig. 6c). *Smchd1*$^{del/del}$ cells showed the expected increase in H3K27me3 levels, whereas *Smchd1*$^{MommeD43/MommeD43}$ cells showed a modest but still significant decrease in H3K27me3 levels, both on the inactive X (Fig. 6d,e) and for the whole nucleus (data not shown). Interestingly, there was a slight yet significant reduction in SMCHD1 levels in *Smchd1*$^{MommeD43/MommeD43}$ nuclei (Fig. 6f) that may have been outside the detection range of the western blots and flow cytometry shown previously. Even with slightly decreased SMCHD1 levels the effect observed with the *MommeD43* variant is opposite to that of loss of SMCHD1, which suggests that the gain-of-function effect seen here outweighs any minor change in protein levels.

We have shown here that *MommeD43* has either no effect or behaves as a gain-of-function SMCHD1 mutation in a context-dependent manner, except in chromatin conformation where it is more akin to a hypomorphic mutation. Therefore, we believe *MommeD43* is a neomorphic mutation with altered function.

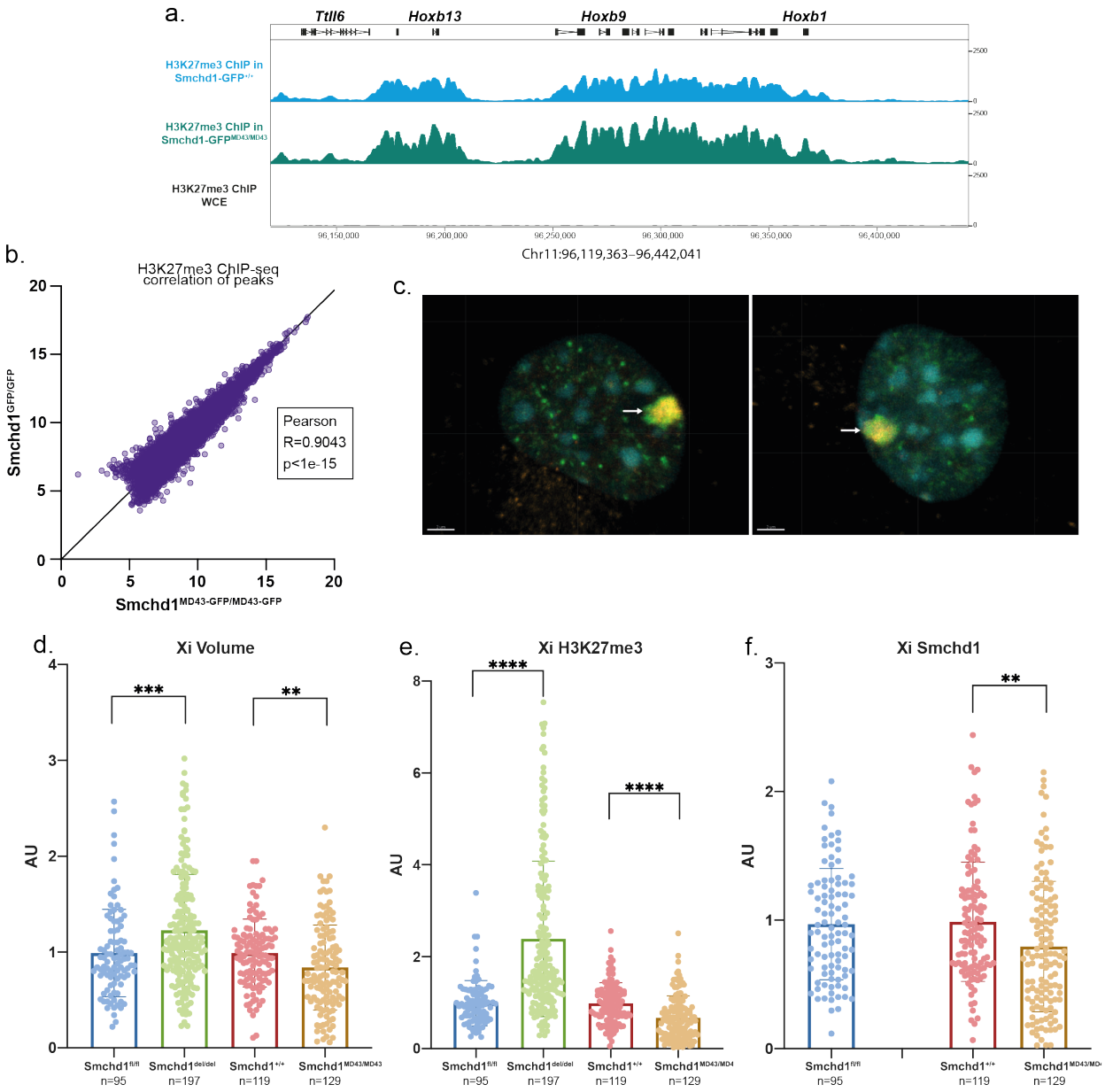


Figure 6: MommeD43 results in depleted H3K27me3 on the inactive X chromosome

a. ChIP-seq track of H3K27me3 in *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}MommeD43}$ vs *Smchd1*$^{GFP/GFP}$ primary female NSCs at the *Hoxb* cluster. The track shows no differences in H3K27me3 localization (ChIP-seq is not a directly quantitative technique, and so the difference in peak height between the two tracks might not be of biological significance). b. Scatter plot of log-transformed normalized H3K27me3 ChIP-seq counts in *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}MommeD43}$ vs *Smchd1*$^{GFP/GFP}$ NSCs around (±2.5kb, then merged if within 1kb) peaks determined from both datasets. The Pearson coefficient indicates very high positive correlation. c. Representative images of immunofluorescence with anti-H3K27me3 (green), anti-SMCHD1 (orange) and DAPI (DNA stain, cyan) in primary female *Smchd1*$^{+/+}$ (left) and *Smchd1*$^{MommeD43/MommeD43}$ (right) NSCs. The arrow points to the inactive X chromosome which is characterized by very high levels of both H3K27me3 and SMCHD1. d, e and f. Scoring of the volume and total levels of H3K27me3 and Smchd1 of the inactive X chromosome in female *Smchd1*$^{fl/fl}$, *Smchd1*$^{del/del}$, *Smchd1*$^{+/+}$ and *Smchd1*$^{MommeD43/MommeD43}$ primary female NSCs. Levels of H3K27me3 and Smchd1 are defined as the total intensity of fluorescence inside the volume of the inactive X, normalized by the total nuclear volume of each scored cell. The inactive X chromosome is defined as the high H3K27me3 region in a semi-automated approach using Imaris to define a closed surface of highest green fluorescence in each nucleus. *MommeD43* abbreviated to *MD43*.

Discussion

We and others have previously studied SMCHD1 extensively in X chromosome inactivation, development, its effects in gene regulation as well as its role in organising the genome. It has been shown that SMCHD1 plays a role in *Hox* gene regulation[14], clustered protocadherin gene expression[10; 11; 13], in mediating long-range chromatin interactions[14; 31-33], and in canonical and non-canonical imprinting[12]. Moreover, variants in SMCHD1 that have differing outcomes on SMCHD1 function are associated with two different diseases: FSHD and BAMS[15; 27]. In each case, SMCHD1 has been considered an epigenetic repressor, however a molecular mechanism of SMCHD1-mediated silencing has yet to be elucidated. In this study, we report a new single amino acid substitution mouse variant of SMCHD1, *MommeD43*. By studying the consequence of this variant on SMCHD1 molecular function and in models of BAMS and FSHD, we enhance both the understanding of how SMCHD1 brings about gene silencing and identify a potential avenue for treatment of FSHD.

Through our developmental and molecular analyses, we show that, with the level of resolution of the techniques used, there is a counter-intuitive divorce of SMCHD1's activity on gene expression and insulation of the chromatin, versus its effect on chromatin architecture. The *MommeD43* variant was identified because it caused increased transgene silencing. Consistently, we revealed that while its effect on target gene activity is context-dependent, the *MommeD43* variant frequently imparts a gain of function effect on SMCHD1. We observed downregulation, which we believe is a delay in activation, of *Hoxa6* and *Hoxd10* in the *MommeD43* mutants. This is consistent with the anterior homeotic transformation observed in skeletal patterning, and in contrast to the upregulation of posterior *Hox* genes and posterior homeotic transformation in *Smchd1* null embryos[14]. On the inactive X chromosome, we observed a decrease in H3K27me3 enrichment in the *MommeD43* mutant female cells, whereas the *Smchd1* deleted samples show an increase in H3K27me3. These data suggest that SMCHD1's role in insulating against other epigenetic regulators such as PRC2, as proposed by us and others is enhanced in the *MommeD43* variant[13; 14; 32; 33]. While these observations suggested a gain of function effect, the influence of the *MommeD43* variant on chromatin interactions were, however, reminiscent of those found upon loss of SMCHD1. We observed weakened long range interactions between SMCHD1 targets, but to a lesser extent than is observed in the *Smchd1*-null mouse. Our findings suggest that the genome-wide long-range chromatin interactions that require SMCHD1 are not directly linked to SMCHD1's silencing of gene expression. This is consistent with our previous study showing loss of long-range interactions on the inactive X and at *Hox* genes in neural stem cells post *Smchd1* deletion, but no upregulation of these genes when *Smchd1* is deleted after early embryonic development[14].

Many groups have shown the relatively modest effect of altered chromatin interactions on transcription. However, those studies usually rely on the ablation of well characterized members of the cohesin complex or CTCF (e.g. ref[49]), which seem to be predominantly architectural in their function. In contrast, newer findings show that PRC1, a complex that has long been considered as a direct repressor of gene expression, also plays a role in setting long-range chromatin interactions which is independent of its silencing function[50]. SMCHD1 seems to exert similarly independent effects.

In addition to the *MommeD43* effect on gene silencing, we were intrigued by the depletion of H3K27me3 on the inactive X despite no detectable change in gene expression in our RNA-seq data in NSCs. These data suggest that the *MommeD43* variant enhanced SMCHD1's capacity to insulate the chromatin against other epigenetic regulators. Given the loss of long-range interactions on the X chromosome observed in the same cells, SMCHD1-mediated long-range interactions themselves

may not offer the insulation against other chromatin regulators, but instead must arise from another action of SMCHD1.

We studied the effects of the *MommeD43* mutation on SMCHD1 function related to disease for two main reasons: it is a variant that conferred greater transgene silencing capabilities to SMCHD1 where the transgene array was reminiscent of the D4Z4 tandem repeat involved in FSHD; and because this enhanced silencing capacity may inform interpretation of SMCHD1 variants found in BAMS, the other disease where SMCHD1 has a confirmed involvement. Relevant to BAMS, we observed no craniofacial abnormalities in the mouse or *Xenopus* systems; however, we did observe differential expression of some developmentally regulated genes in the frontonasal prominence regions from *MommeD43* embryos, suggesting that the relevant tissue in the embryo is sensitive to the *MommeD43* variant. For reasons we do not yet understand, these changes did not result in phenotypic outcomes, meaning the MommeD43 mouse is not a useful model to relate to BAMS.

The gain-of-function effects of the *MommeD43* variant could provide a potentially elegant solution for treatment of the FSHD phenotype by revealing how to stimulate the wild-type copy of SMCHD1 that is retained in FSHD patients to silence *DUX4*. Indeed, overexpression of SMCHD1, or restoring expression of wild-type SMCHD1 in muscle cell cultures of FSHD2 patients is an efficient way of silencing DUX4[51; 52]. Utilising a mouse model harbouring a short *D4Z4* transgene array, we found that the *MommeD43* variant in SMCHD1 partially rescues aberrant activation of *DUX4* in tissues relevant to FSHD. This finding raises the possibility that gain of function mutations to SMCHD1, similar to *MommeD43*, may be harnessed therapeutically to re-establish the silencing of *DUX4* in FSHD patients. The precise role of ATPase activity in the execution of SMCHD1's gene silencing functions is currently incompletely understood. It remains of outstanding interest to define whether this enzymatic activity could be targeted therapeutically, and which functions beyond this enzymatic activity are conducted by the extended ATPase domain, including the C-terminal region where the *Momme43* substitution resides.

In conclusion, our study of a new mouse variant of SMCHD1 has disentangled SMCHD1's role in chromatin architecture from its role in transcription and insulation, providing greater mechanistic insight into how SMCHD1 may bring about gene silencing. Moreover, we have revealed that modulating SMCHD1's activity is effective in ameliorating failed *D4Z4* silencing in a model of FSHD, with direct relevance to FSHD treatment.

Acknowledgements

We thank Dr Ruth Arkell for valuable discussion of the FNP RNA-seq and the WEHI MAGEC facility for recreating the *MommeD43* mutation on the *Smchd1-GFP* background. We thank mouse technicians at WEHI (Jessica Martin) and QIMR (Joanne Sutton) for their work on this project. We thank SAXS beamline staff at the Australian Synchrotron for assistance with data collection. This work was supported by grants and fellowships from the Australian National Health and Medical Research Council (GNT1098290 to MEB, JMM and MER, fellowship GNT1194345 to MEB, GNT1172929 to JMM, GNT1104924 to MER). NJ and AG were supported by Australian Research Training Program scholarships. MEB was supported by the Bellberry-Viertel fellowship. JCG is funded by a fellowship from the FSHD Society (Sylvia & Leonard Marx Foundation Fellowship FSHS-82017-02). This work was also supported by grants from the Prinses Beatrix Spierfonds (W.OP14-01, W.OR17-04), the US National Institutes of Health (National Institute of Arthritis and Musculoskeletal and Skin Diseases 2R01AR066248), and Spieren voor Spieren. The generation of the MommeD43 mice on the Smchd1-GFP background used in this study was supported by Phenomics

Australia (PA) and the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) program. Additional support was provided by the Victorian State Government Operational Infrastructure Support, Australian National Health and Medical Research Council IRIISS grant (9000653). The Australian Regenerative Medicine Institute is supported by grants from the State Government of Victoria and the Australian Government. SX is supported by NMRC (NMRC/OFYIRG/062/2017) and NUS PYP. B.R. is an investigator of the National Research Foundation (NRF, Singapore), Branco Weiss Foundation (Switzerland) and an EMBO Young Investigator, and is supported by an inaugural Use-Inspired Basic Research Fund from the Agency for Science & Technology and Research (A*STAR) in Singapore. B.H., Y.K., S.M. L.B. and J.G. are members of the European Reference Network for Rare Neuromuscular Diseases [ERN EURO-NMD].

Materials and Methods

Mouse strains and genotyping

All mice were bred and maintained with standard animal husbandry procedures. At WEHI protocols were approved by the WEHI Animal Ethics Committee under animal ethics numbers AEC 2014.026, 2018.004. Mice at QIMR were kept under the approval of the QIMR ethics committee, including the ENU treatment of mice. The approval numbers for this work were QIMR AEC-P1076 and AEC-P1224. At the animal facility of the Leiden University Medical Center all mouse breeding and experiments were performed according to Dutch law and Leiden University guidelines and were approved by the National and Local Animal Experiments Committees. Mice at all three locations were kept in individually ventilated cages with standard rodent chow and water available ad libitum with a standard 12h/12h light/dark cycle.

Smchd1$^{fl/fl}$ mice were maintained on a C57BL/6 background, as described previously[42]. Mice carrying the *Smchd1*^{MommeD1} mutation, as previously described[7], were maintained on the FVB/NJ inbred background. *Smchd1*$^{GFP/GFP}$ mice were maintained on a C57BL/6 background as described[14].

Mice carrying the *Smchd1*^{MommeD43} mutation were produced on the FVB/NJ background homozygous for the *Line3* transgene, exactly as previously described for other Momme mutations[7; 34; 53]. The MommeD43 mutation was backcrossed off the transgenic background and maintained on pure FVB/NJ after mapping of the mutation.

The MommeD43 mutation was recreated on the *Smchd1*$^{GFP/GFP}$ background[14] by injection of Cas9 protein, guide RNA plus repair template into zygotes (sequences for Smchd1 gRNA and repair template in Supplementary Table 8), similarly to what has previously been described[54]. At the same time as introducing the MommeD43 mutation the PAM site was mutated with a silent C to T mutation. This was performed by the WEHI MAGEC facility. The new allele was backcrossed to *Smchd1*$^{GFP/GFP}$ for more than 5 generations before animals were used in experiments. This line was created and maintained on a C57BL/6 background.

D4Z4-2.5 mice were generated before[40]. Female hemizygous D4Z4-2.5 mice on a C57BL6/J background were crossbred with male heterozygous Smchd1$^{MommeD43/+}$ mice on an FVB/NJ background to obtain D4Z4-2.5/Smchd1$^{MommeD43/+}$ mice. Genotypes were determined by PCR analysis using tail DNA.

Both the *MommeD43* and *MommeD1* mutations were genotyped by allelic discrimination (primers in Supplementary Table 8). The *Smchd1*^{fl}, *Smchd1*^{del} and *Smchd1*^{GFP} alleles were genotyped by PCR as previously described[14; 42] (primers in Supplementary Table 8). The D4Z42.5 transgene was genotyped as previously described[42] (Supplementary Table 8 for primer sequences).

Embryo tails or yolk sacs were used to prepare DNA through standard methods. Genotypes were determined by PCR with GoTaq Green (Promega) or allelic discrimination for *MommeD1* or *MommeD43*. Sex was determined by PCR for *Otc* (X chromosome) and *Zfy* (Y chromosome) (Supplementary Table 8 for primer sequences).

ENU mutagenesis screen to identify MommeD43 mutant line

The ENU mutagenesis screen used to identify the MommeD43 mutant line of mice was run exactly as previously described[7; 34]. In brief, homozygous transgenic males were treated with ENU and allowed to recover fertility. They were bred with wild-type transgenic females and the expression of transgene silencing was assessed in the offspring at weaning, by performing flow cytometry for GFP expression, in a drop of whole blood taken from the tail tip. The founder animal for the MommeD43 line displayed reduced GFP expression and was bred further to study inheritance patterns.

Genetic mapping of the *MommeD43* mutation

Mapping of the *MommeD43* mutation was performed as has been reported for other Momme mutations[34]. The *MommeD43* allele was bred down at least 4 generations before mapping was initiated. For mapping the allele was crossed from the FVB/N background to the C57BL/6 background (Line3C, carrying the GFP transgene on the C57BL/6 background) twice to enable positional cloning. These animals were phenotyped for the effect on transgene expression by flow cytometry. The Illumina GoldenGate genotyping assay (Mouse Medium Density Linkage Panel) was performed on 11 phenotypically mutant and 11 phenotypically wild-type littermates. The Mouse Medium Density Linkage panel has 766 measurable SNPs between C57BL/6J and FVB/NJ. Samples were genotyped following the Illumina protocol and genotype calls were made using the Genotyping module of the GenomeStudio v1.1 software. Only samples with a call rate >95 were accepted. Linked intervals were identified based on LOD score (Supplementary Table 7). A LOD score of over 4.5 was found for chromosome 17, 63-98 Mb. Subsequent exome sequencing using Roche Nimblegen capture in *MommeD43* and wild-type controls revealed the G to T mutation at chr17: 71776840, which relates to a C to A mutation in the sense orientation of exon 15 of *Smchd1*. This mutation was verified by Sanger sequencing. No other mutations were found either within the linked interval or in other exons.

Generation of Neural Stem cells

NSCs were derived and cultured exactly as previously described[13; 14]. In brief, cortices of E14.5 embryos were dissociated and grown as an adherent monolayer on tissue culture treated plates. Growth media was NeuroCult NSC Basal Medium (Mouse) containing NeuroCult Proliferation Supplement (Mouse) (StemCell Technologies), 0.2% heparin solution (Stem-Cell Technologies), recombinant human EGF (20 ng/mL), recombinant human basic FGF (20ng/mL, all from Stem Cell Technologies) and 10 ng/ml laminin (Sigma Aldrich). Primary cells were maintained for a maximum of 20 passages.

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

Smchd1$^{del/del}$ NSCs were generated from *Smchd1*$^{fl/fl}$ NSCs exactly as previously described[14]. *Smchd1* deletion was confirmed by immunofluorescence with an in-house made anti-Smchd1 antibody during the immunofluorescence experiments (Mab #8, now available commercially from Merck).

Retrovirus production and transduction

VSVg pseudotyped MSCV-Cre-puromycin retroviral supernatants were produced with calcium phosphate–mediated transient transfection of 293T cells, as previously described[55]. The medium

was collected at 48 h after transfection, centrifuged to remove residual 293T cells and either concentrated with PEG or used unconcentrated.

Transduction of NSCs with PEG-concentrated viral supernatant was performed exactly as previously described[14].

Immunofluorescence

Immunofluorescence was performed on *Smchd1*$^{+/+}$, *Smchd1*$^{MommeD43/MommeD43}$, *Smchd1*$^{fl/fl}$ and *Smchd1*$^{del/del}$ early passage primary NSCs as described[56]. The primary antibodies used were an in-house monoclonal anti-Smchd1 and an anti-H3K27me3 (Cell Signalling technologies, C36B11). After immunofluorescence assays and DAPI stain, cells were mounted in Vectashield H-1000 mounting medium (Vector Laboratories). Cells were visualized on an LSM 880 (Zeiss) microscope. All images were taken using the same settings and the same instrument across multiple sessions. Images for each "treatment" condition and its corresponding control were acquired within the same sessions. The mean value of the corresponding control was used to normalize total intensity values individually for each channel across different acquisition sessions.

Imaris Software (Bitplane) was used to measure the volume of the nucleus (using DAPI staining) or occupied by high levels of H3K27me3 as marker of the Xi within the nucleus. A threshold was manually set to measure the signal only above nucleoplasmic staining or background. Nuclei that were too close together to be defined as separate volumes were manually removed from subsequent analyses. A region of interest was then defined on the basis of DAPI (nuclei) or focal H3K27me3 enrichment (Xi). The volume was then calculated for the region of interest above the threshold. Total fluorescence intensity levels of Smchd1 or H3K27me3 fluorescence were normalized by the nuclear volume of each cell individually, then across each acquisition session.

DNA FISH

DNA FISH was performed on *Smchd1*$^{MommeD43/MommeD43}$ and its *Smchd1*$^{+/+}$ control (FVB/NJ background), *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}MommeD43}$ and its *Smchd1-GFP*$^{+/+}$ control (C57Bl/6 background) and *Smchd1*$^{del/del}$ and its *Smchd1*$^{fl/fl}$ control (C57Bl/6 background) NSCs. 1 µg RP23-196F5 (*HoxB* locus) or RP24-323I2 (*Olfr* locus) BAC DNA (CHORI) was used in a 12-hour nick-translation reaction (Vysis) to generate DNA probes labeled with Green 496 dUTP or Orange 552 dUTP (Seebright), respectively. Approximately 100 ng probe per sample was precipitated in ethanol with 10% NaOAc, 1 µg salmon sperm DNA (Life Technologies) and 1 µg mouse Cot-1 DNA before being resuspended in formamide (Sigma-Aldrich), denatured at 75 °C for 10 min and allowed to compete with Cot-1 DNA for 1 h at 37 °C. Cells were prepared for DNA FISH exactly as described in Chaumeil et al.[56] with 36h of hybridization at 42°C. Cells were mounted in H1000 Mounting Medium (Vectashield) and visualized on an LSM 880 (Zeiss) microscope with Airyscan processing. Images were analysed with the open-source ImageJ distribution package FIJI[57]. Brightness and contrast were manually set for each image for clear scoring, and spectral shift was corrected using an image of Tetraspeck 0.1um beads (ThermoFisher) acquired on each session with the same settings as the experiment images. Overlapping or touching FISH signals for *HoxB* and *Olfr* probes were scored as interacting. 419 alleles were scored for the Smchd1$^{+/+}$ NSCs, 320 alleles for the Smchd1$^{MommeD43/MommeD43}$ NSCs, 164 alleles for the Smchd1$^{flox/flox}$ and 152 alleles for the Smchd1$^{del/del}$.

SAXS *(described in Biochem J 2016, **473**, p. 1733-1744[25])*

SAXS data were collected on the SAXS/WAXS beamline at the Australian Synchrotron, coupled with in-line size exclusion chromatography. 50 µL at 5 mg/mL of recombinant SMCHD1 protein sample was each loaded onto a Superdex-200 5/150 (GE Healthcare) pre-equilibrated in purification buffer [200 mM NaCl, 20 mM Tris-HCl (pH 8.0), 10% (v/v) glycerol, 0.5 mM TCEP] and eluted via a 1.5 mm glass capillary at 16◦ C positioned in the X-ray beam. Diffraction data were collected with a 1M, 170 mm x 170 mm Pilatus detector at 2 s intervals over the course of the elution. Data were processed by the beamline control software, ScatterBrain. 2D intensity plots from the size exclusion chromatography peak of the eluting protein sample were radially averaged and normalised to sample transmission. Scattering profiles from buffer alone were averaged for background subtraction of 1D profiles.

Data analyses were performed with the ATSAS suite [58]. PRIMUS[59] was used to perform Guinier analysis for examining scattering curves at small angles (qRg below 1.3). From this, estimation of two parameters can be obtained: the radius of gyration (Rg) value, which represents the square root of the average distance of each scattering atom from the particle centre, and zero angle intensity (I(0)), which is proportional to the molecular weight and the concentration of the protein. The linearity of the Guinier plot reflects the quality of the scattering data obtained, indicating the absence of high molecular weight aggregates or inter-particle interference. Real-space interatomic distance distribution function, P(r), and maximum dimension of the scattering particle, Dmax, were computed by indirect Fourier transform via GNOM[60]. Low resolution shape envelopes were generated via the ab initio bead-modeling program, DAMMIF[61]. 10 independent models were built from each scattering profile and were further compared in the program DAMSEL, where the mean value of the normalised spatial discrepancy (NSD) of the 10 models were calculated. The most probable models were aligned by DAMSUP, and an averaged model was obtained via DAMAVER. The final model was adjusted to correspond with the experimentally obtained data via the program DAMFILT[62].

Protein expression and purification

The SMCHD1 N-terminal region (residues 111-702) was PCR-amplified from a *Mus musculus* full-length Smchd1 template. The *MommeD43* mutation (A667E) was introduced by oligonucleotide-directed mutagenesis (5' CTGTGCCCATTGAAAAGCTGGAT AGG; 3' CCTATCCAGCTTTTCAATGGGCACAG) and ligated into the pFastBac Htb vector (Life Technologies). Bacmids were prepared using the Bac-to-Bac system, and utilized for protein expression in *Sf*21 insect cells, as described previously[25]. Cells were maintained in Insect-XPRESS protein-free insect cell media with L-glutamine (Lonza) and infected at a density of 3-4 x 10^6 with high-infectivity baculovirus for protein expression.

Purification was performed as previously described[23; 26]. Cells were resuspended in lysis buffer [0.5 M NaCl, 20 mM Tris-Hcl (pH 8.0), 20% (v/v) glycerol, 5 mM imidazole (pH 8.0), 0.5 mM TCEP] supplemented with 1 mM PMSF and 1X cOmplete EDTA-free protease inhibitor (Roche). Sonication was performed on ice for 5 cycles with 1 s on, 0.2 s off, 22 s per cycle at 50% amplitude, using the Bandelin sonicator fitted with the VS 70/T probe. To remove insoluble material, lysates were centrifuged at 45,000 x g for 30 min at 4◦ C. Following cell lysis by sonication, lysate supernatant from cells expressing N-terminal 6- His-tagged proteins were incubated with nickel-nitrilotriacetic acid (Ni-NTA) cOmplete His-tag purification resin (Roche) for 1 h at 4◦ C, on rollers. 1 ml of 50% resin slurry was used per 1 L of cell culture. The resin was pelleted by centrifugation at 1,500 x g for 5 min at 4◦ C and the supernatant was removed as the unbound sample. The resin was washed twice with

5 mM imidazole buffer (pH 8.0), followed by two washes with 35 mM imidazole (pH 8.0) and eluted in 250 mM imidazole buffer (pH 8.0). The 6-His-tag was cleaved by incubation of the pooled elutions with tobacco etch virus (TEV) protease overnight at 4◦ C. Next day, cleaved protein was concentrated with a 30-kDa molecular mass cut-off concentrator (Millipore) by centrifugation for 5 min at 2,500 x g, 4◦ C, then diluted in Buffer A [50 mM NaCl, 25 mM HEPES (pH 7.5), 0.5 mM TCEP, 10% (v/v) glycerol] for ion exchange chromatography. The concentrated protein sample was loaded onto a MonoQ 5/50 GL column (GE Healthcare) pre-equilibrated with Buffer A and exchanged into Buffer B [500 mM NaCl, 25 mM HEPES (pH 7.5), 0.5 mM TCEP, 10% (v/v) glycerol] in a 0-100% gradient over 20 column volumes for protein elution. Fractions of interest were pooled, concentrated and subjected to size exclusion chromatography on a Superdex-200 10/300 GL column (GE Healthcare) pre-equilibrated with either SEC100 buffer [100 mM NaCl, 20 mM HEPES (pH 7.5)]. Fractions containing the recombinant protein of interest were pooled concentrated and snap-frozen in liquid nitrogen for storage at -80◦ C. Protein concentration was measured using a DS-11 FX Spectrophotometer (DeNovix).

ATPase assay

ATPase assays were performed as outlined in Chen *et al.*[25] 10 µL reactions were set up in triplicates in 384-well low flange, black, flat-bottom plates (Corning) containing 7 µL reaction buffer [50 mM Hepes (pH 7.5), 4 mM MgCl2, 2 mM EGTA], 1 µL recombinant protein at concentrations ranging from 0.1-0.6 µM or SEC buffer control, 1 µL nuclease-free water and 1.25-10 µM ATP substrate. Reactions were incubated at 25 C for 1 hour in the dark. Reactions were stopped by the addition of 10 µL detection mix [1X Detection buffer, 4 nM ADP AlexaFluor 633 Tracer, 128 µg/mL ADP[2] antibody] and incubated for another hour in the dark. Fluorescence polarization readings (mP) were measured with an Envision plate reader (PerkinElmer Life Sciences) fitted with excitation filter 620/40 nm, emission filters 688/45 nm (s and p channels) and D658/fp688 dual mirror. Readings from a free tracer (no antibody) control were set as 20 mP as the normalization baseline of the assay for all reactions. The amount of ADP produced by each reaction was estimated by a 12-point standard curve, as outlined in the manufacturer's protocol. Data was plotted and analysed in GraphPad Prism.

Western Blot

Samples were resolved by standard SDS-PAGE analysis on 4-12% Bis-Tris gels (Thermo Fisher Scientific) in MES buffer and transferred to a PVDF membrane (Osmonics, GE Healthcare) by wet transfer at 100 V for 1 h in transfer buffer [25 mM Tris, 192 mM glycine, 20% (v/v) methanol]. Membranes were blocked with a 5% (v/v) skim milk powder in 0.1% (v/v) Tween-20/PBS for 1 h at room temperature. Primary antibody (monoclonal in-house anti-Smchd1, clone 1D6, 1:2000) was added to the membranes in 5 mL blocking buffer and incubated overnight at 4◦ C in a capped tube, on rollers. Membranes were washed for 30 min at room temperature with 0.1% (v/v) Tween-20/PBS, followed by incubation with secondary antibody (anti-rat IgG HRP-conjugated, Southern Biotech, 1:10,000) for 1 h at room temperature, which was diluted in 5 mL blocking buffer. The 30 minute washing step was repeated as before and antibody binding was visualised using the Luminata ECL system (Millipore) following the manufacturer's instructions.

Skeletal preparations and scoring

Whole-mount skeletal staining of E16.5 fetuses was performed as previously described [63]. Each fetus was skinned and organs were removed under a light microscope, in additional to dissolving remaining tissue in acetone. Once stained, skeletons were washed through a graded glycerol/water series before imaging in 100% glycerol using a ZEISS SV11 stereomicroscope. The vertebral phenotype of each skeleton was scored by two independent assessors who were blind to genotype and sex.

Tailbud dissection and somite counting

Tailbud tissue containing the presomitic mesoderm (PSM) was dissected from *Smchd1*$^{+/+}$ and *Smchd1*$^{MommeD43/MommeD43}$ E8.5 embryos as previously described [14; 64]. In brief, embryos were dissected ice-cold DEPC-treated PBS. Tailbud tissue was horizontally dissected at the distance of 1.5 somites below the last segmented somite to ensure the next developing somite from the PSM, S0, was not included in the tailbud dissection. Tailbud tissue was snap frozen on dry ice and stored at -80°C for later RNA extraction using a Zymo Quick-RNA Miniprep Kit. The yolk sac was retained for genotyping. Somites were counted before fixing each embryo in 4% DEPC-treated paraformaldehyde at 4°C overnight. Embryos were washed through a graded methanol/PBT (DECP-treated PBS with 1% Tween (v/v)) series as previously described[65] before brief staining in dilute ethidium bromide solution and imaging under a fluorescence dissection microscope to check somite counting. RNA-sequencing was performed on somite-matched pairs of *Smchd1*$^{+/+}$ and *Smchd1*$^{MommeD43/MommeD43}$ tailbud tissue.

Frontonasal prominence collection

E10.5 embryos were dissected in ice-cold PBS and the frontonasal prominence (FNP) removed by dissecting along the boundary which separates the FNP from the maxillary prominence using fine forceps. Dissected tissue was immediately snap frozen on dry ice and the remaining embryo placed in a fresh well of PBS for somite counting. Yolk sacs were collected for genotyping purposes.

HREM

E14.5 embryos were collected and fixed in Bouin's fixative overnight followed by extensive washing and storage in PBS. Following dehydration in a graded methanol series, samples were incubated for 3 days in JB-4 (Sigma)/Eosin (Sigma)/Acridine orange (Sigma) dye mix before embedding and imaging as previously described[66; 67] (https://dmdd.org.uk/hrem/). To measure key craniofacial features, samples were aligned to equivalent orientations in three dimensions and the distance measured then divided by crown-rump length using OsirixMD. Statistics were calculated using unpaired t-tests, followed by Benjamini-Hochberg procedure to correct for the false discovery rate.

Xenopus embryological assays

Xenopus laevis were used according to guidelines approved by the Singapore National Advisory Committee on Laboratory Animal Research. Injections of human *SMCHD1* mRNA into Xenopus embryos were performed as previously reported[15]. Briefly, two dorsal animal blastomeres were injected at the 8-cell stage with 240 pg of *in vitro* transcribed human *SMCHD1* mRNA containing various mutations. Embryos were allowed to develop at room temperature until stage 45-46 and fixed. Eye diameter was measured using a Leica stereomicroscope with a DFC 7000T camera. No statistical method was used to predetermine sample size. No randomization or blinding was used. Embryos that died before gastrulation were excluded. Statistics were calculated using 1-way ANOVA, followed by Dunn's multiple comparison test.

Isolation and culturing of fibroblasts and EDL-derived muscle cells

For establishing fibroblast cultures, neonatal mice were sacrificed by decapitation after which the skin was removed. The skin was next cut into small pieces, evenly distributed over a 60 mm culture dish, and a small amount of DMEM/F12 medium (Gibco; Thermo Fisher Scientific, Bleiswijk, the Netherlands) supplemented with 1% penicillin-streptomycin (P/S) (Sigma Aldrich, Zwijndrecht, the Netherlands) and 20% heat-inactivated foetal bovine serum (FBS) (Thermo Fischer Scientific, the Netherlands) was added. Twice a week, a small amount of medium was added to the culture dish, and over time fibroblasts grew out of the skin pieces. These cells were either transferred for further expansion using the same medium or harvested for RNA or chromatin isolation.

For establishing muscle cell cultures, mice of two months of age were sacrificed by cervical dislocation followed by removal of the EDL muscle from tendon to tendon. The EDL muscle was next incubated for 105 minutes at 37 °C in 0.2% collagenase (Sigma Aldrich, Zwijndrecht, the Netherlands) in DMEM medium (Gibco; Thermo Fisher Scientific, Bleiswijk, the Netherlands) containing 1% P/S (Sigma Aldrich, Zwijndrecht, the Netherlands). The resulting individuals muscle fibres were dissociated with a Pasteur pipet with a smooth end and, after several washing steps as previously described[68], transferred to a Matrigel-coated (BD Biosciences, Vianen, the Netherlands) 6-well culture plate (150 fibres per well) in DMEM medium supplemented with 30% FBS, 10% horse serum, 1% P/S, 1% chicken embryo extract, and 2.5 ng/ml fibroblast growth factor (all from Thermo Fischer Scientific, Bleiswijk, the Netherlands). After three days, the fibres were removed, and the attached myoblasts were trypsinised and plated in fresh Matrigel-coated plates. Myoblasts were harvested for RNA and protein isolation at ~70% confluency or differentiated by replacing the medium with DMEM with 2% horse serum and 1% P/S. 72 hours after the start of differentiation, myotubes were harvested for RNA and protein isolation.

RNA isolation, cDNA synthesis, and real-time quantitative PCR

Total RNA was isolated using the miRNeasy kit (Qiagen, Venlo, the Netherlands) following the manufacturer's instructions and included a DNase treatment on the column for 30 minutes at room temperature. RNA concentrations were determined using the Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Bleiswijk, the Netherlands). 1-3 µg RNA was reverse transcribed with the RevertAid H Minus First Strand cDNA synthesis kit and Oligo(dT)18 primers (both Thermo Fisher Scientific, Bleiswijk, the Netherlands), following the instructions of the

manufacturer. cDNA was treated with 2 units of RNaseH (Thermo Fisher Scientific, Bleiswijk, the Netherlands) for 20 minutes at 37°C. RT-qPCR analysis was performed with the CFX96 system (Bio-Rad, Veenendaal, the Netherlands using iQ SYBR Green Supermix (Bio-Rad, Veenendaal, the Netherlands), 0.5 pM of each primer (sequences are listed in Supplementary Table 8), and 1:5 or 1:50 diluted cDNA. The following cycling conditions were used: an initial denaturation step at 95 °C for 3 minutes followed by 40 cycles of 10 seconds at 95 °C and 30 seconds at primer Tm. A melting curve analysis from 65 °C to 95 °C (temperature increments of 0.5 °C) was performed to determine the specificity of each reaction. Data was analysed with Bio-Rad CFX Manager version 3.1 (Bio-Rad, Veenendaal, the Netherlands) and normalized to the housekeeping genes Gapdh and Rpl13a.

Western blot in NSCs and D4Z4 samples

The Western blot for SMCHD1 in NSCs was performed exactly as previously described in de Greef et al[42], using in house monoclonal #5 against SMCHD1, which is now available by Merck.

Tibialis anterior muscle, spleen tissue, and scraped myoblasts and myotubes were homogenized in 10 volumes of solubilization buffer (150 mM NaCl, 50 mM Tris, 200 mM PMSF, 100 mM benzamidine, pH 7.4) with 1% Triton X-100. Next, samples were incubated, rotating top-over-top, for 1 hour at 4°C and centrifuged for 30 minutes at 4°C at maximum speed (tabletop centrifuge) to remove non-homogenized material. Protein concentrations were determined using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific, Bleiswijk, the Netherlands). Protein samples were separated on a NuPAGE™ 4–12% Bis-Tris Protein Gel (Thermo Fisher Scientific, Bleiswijk, the Netherlands) and transferred to an Immobilon-FL PVDF membrane (Merck, Amsterdam, the Netherlands). The membrane was blocked for 1 hour at room temperature in 4% skim milk/PBS containing 0.1% Tween-20, followed by an overnight incubation step at 4°C with primary antibodies: anti-Smchd1 antibody (1:250; HPA039441; Sigma Aldrich, Zwijndrecht, the Netherlands), anti-Emerin (1:200; SC-15378; Santa Cruz Biotechnology; Bio-Connect B.V., Huissen, the Netherlands), anti-DUX4 antibody (1:1000; ab124699; Abcam, Cambridge, United Kingdom), and anti-Tubulin (1:2000; T6199; Sigma Aldrich, Zwijndrecht, the Netherlands). The next day, blots were washed with PBS containing 0.1% Tween-20 and incubated with dye-conjugated secondary antibodies (Li-Cor, Bad Homburg, Germany) for 1 hour at room temperature in 4% skim milk/PBS with 0.1% Tween-20. After washing with PBS containing 0.1% Tween-20 and PBS, blots were imaged using the Odyssey CLx imager (Li-Cor, Bad Homburg, Germany).

DNA methylation analysis at *D4Z4*

Genomic DNA (400 ng) from mouse tail was bisulphite converted with the EZ DNA Methylation-Lightning kit (Zymo Research; BaseClear Lab Products, Leiden, the Netherlands) following the instructions of the manufacturer. A PCR reaction of the DR1 region within the D4Z4 repeat transgene (and a PCR reaction of the FasPas region just distal of the D4Z4 repeat transgene was performed using FastStart Taq DNA polymerase (Roche, Woerden, the Netherlands) with the following cycling conditions: initial denaturation for 10 minutes at 95 °C followed by 35 cycles of 20 seconds at 95 °C, 30 seconds at 60 °C and 40 seconds at 72 °C, with a final extension step for 5 minutes at 72 °C. Next, the PCR products were ligated into the TOPO TA vector (Thermo Fisher Scientific, Bleiswijk, the Netherlands), followed by transformation of the ligation products into competent DH5α bacteria. Plasmid DNA from at least 10 individual colonies was isolated and sent for Sanger sequencing.

Chromatin immunoprecipitation followed by real-time quantitative PCR at *D4Z4*

We followed a previously published protocol with minor modifications[69]. Confluent cultures of neonatal fibroblasts and spleen tissue minced into small pieces were used for these analyses. Both cells and tissues were crosslinked with 1% formaldehyde for 10 minutes at room temperature, after which the crosslinking reaction was stopped by adding 125 mM glycine for 5 minutes at room temperature. Next, fibroblasts were lysed in NP buffer (150 mM NaCl, 50 mM Tris-HCL (pH 7.5), 5 mM EDTA, 0.5% NP-40, 0.1% Triton X-100). Spleen tissue was lysed in LB buffer (50 mM Hepes (pH 7.5), 14 mM NaCl, 1 mM EDTA, 10% glycerol, 0.25% NP-40, 0.25% Triton X-100). The resulting chromatin was sheared in NP buffer using a sonicator bath (Bioruptor Pico; Diagenode, Ougrée, Belgium) for 30 minutes at maximum output and 30 seconds on/off cycles. The fragmentation between 200–2000 bp was confirmed through phenol-chloroform extraction followed by agarose gel electrophoresis. DNA concentrations were determined with the Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Bleiswijk, the Netherlands). For the histone ChIP analyses, 3 µg of chromatin was used. For the Smchd1 ChIP analysis, 30 µg of chromatin was used. First, chromatin was precleared for 1 hour at 4 °C with blocked protein A Sepharose beads (GE Healthcare, Eindhoven, the Netherlands. Next, precleared chromatin was incubated overnight at 4 °C with 5 µg of the following antibodies: H3K4me2 (39141; Active Motif, Carlsbad, USA), H3K9me3 (39161; Active Motif, Carlsbad, USA), H3K27me3 (07-449; Merck, Amsterdam, the Netherlands), Smchd1 (ab31865; Abcam, Cambridge, United Kingdom), or mouse IgG (PP64; Merck, Amsterdam, the Netherlands). Immunoprecipitation was performed by incubating the antibody-chromatin mixture with blocked protein A Sepharose beads for 2 hours at 4 °C. Washing of the beads was performed according to the previously published protocol[69]. DNA was isolated with Chelex resin (Bio-Rad, Veenendaal, the Netherlands). Finally, quantitative PCR analysis was performed to amplify the D4Z4 repeat transgene[45]. The Gapdh promoter was amplified as a control.

Statistical analyses

Statistical analyses of the non-genomic data were performed with GraphPad Prism software (version 8; GraphPad Software, Inc., La Jolla, USA). The statistical tests that were performed are described in the figure legends. $P<0.05$ was considered significant.

<u>Genomics</u>

Genomic data have been deposited in the Gene Expression Omnibus and can be accessed under the accession GSE174113.

All of the genomics datasets were analysed from the raw data stage concomitantly with all the other datasets to which they were compared. In the case of ChIP-seq in Smchd1$^{GFP/GFP}$ NSCs, RNA-seq in Smchd1$^{MommeD1/MommeD1}$ or Smchd1$^{fl/fl}$ and Smchd1$^{del/del}$ NSCs, in situ Hi-C in Smchd1$^{fl/fl}$ and Smchd1$^{del/del}$ NSCs, these data were published previously.

ChIP-seq

Chromatin immunoprecipitation for Smchd1-GFP was performed exactly as described in Wanigasuriya, Gouil et al.[12] in $4x10^7$ cells from three independent primary NSC lines for each

genotype *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}MommeD43}$ and *Smchd1*$^{GFP/GFP}$ with 16ug of anti-GFP antibody (Invitrogen A11122).

Chromatin immunoprecipitation for H3K27me3 was performed exactly as described in Jansz et al.[14] "ChIP for histone proteins". 1x10^6 cells from the same cell lines as above were used with 2ug of anti-H3K27me3 antibody (Cell Signalling technologies, C36B11). After pulldown, DNA was purified using a Zymo ChIP DNA clean and concentrator kit.

Libraries were generated with an Illumina TruSeq DNA Sample Preparation Kit. 200- to 400-bp fragments were size-selected with AMPure XP magnetic beads. Libraries were quantified with a D100 tape in a 4200 Tapestation (Agilent). Libraries were pooled and sequenced on the Illumina NextSeq platform, with 75-bp single-end reads.

ChIP-seq analysis

Adapter trimming was performed with Trim Galore! v0.4.4 with Cutadapt v.1.15[70], library QC was assessed with FastQC v0.11.8 before mapping with Bowtie2 v2.3.4.1[71] and Samtools v1.7[72] with default parameters to the GRCm38.p6 version of the reference mouse genome. BAM files were imported into SeqMonk v1.45.1 extending them by 150 bp to better represent size-selected fragments and peaks were called with the MACS style caller within the SeqMonk package (settings for 300 bp fragments, $P < 1 \times 10^{-5}$) by merging all three Smchd1GFP biological replicates and both WCE biological replicates into replicate sets. ChIP-seq tracks for the GFP ChIP were produced with SeqMonk by defining probes with a running window (width, 1000 bp; step 250 bp), doing a read-count quantitation then normalizing by library size before doing a match distribution normalization within each replicate set and smoothing over 5 adjacent probes. For the Smchd1-GFP ChIP scatter plot, the same data was quantified only over the peaks (±5kb) published in Jansz et al.[14], normalized by total library size log2-transformed counts with datasets from merged biological replicates (3 for each genotype). The H3K27me3 plot was produced the same way, using the lists of peaks from both genotypes produced as described above ±2.5kb, then merging any probes closer than 1kb.

Reduced representation bisulphite sequencing (RRBS)

Genomic DNA was extracted using a Qiagen DNeasy kit. RRBS was performed with 100ng of genomic DNA from each sample of female primary Smchd1$^{MommeD43/MommeD43}$ and Smchd1$^{+/+}$ NSCs (3 independent cell lines of each genotype) with the Ovation RRBS Methyl-seq kit (NuGen) following the manufacturer's instructions. The bisulphite conversion was performed with the EpiTect Fast Bisulfite Conversion kit (Qiagen) following the instructions.

The libraries produced were were cleaned and size-selected with AMPure XP magnetic beads. Libraries were quantified with a D100 tape in a 4200 Tapestation (Agilent). Libraries were pooled and sequenced on the Illumina NextSeq platform with 75-bp paired-end reads[73].

RRBS analysis

The fastq files were processed with TrimGalore! v0.4.4 with Cutadapt v1.15[70] to remove Illumina adapter sequences. Custom adapter sequences from the Ovation RRBS kit were removed with the trimRRBSdiversityAdaptCustomers.py script provided on the manufacturer's website. Trimmed libraries were processed with Bismark v0.19.0[74] with Bowtie2 v2.3.4[71] and Samtools v1.7[72]. They were aligned to the GRCm38.p6 version of the mouse genome.

The bismark coverage files produced were then loaded into R v3.6.1 and processed using the methylKit v1.10 package[75] (bismarkCoverage pipeline for CpG context), filtered by coverage (minimum 10 reads per base, maximum 99.9% percentile). The reported differences in methylated bases were obtained with the getMethylDiff function for a difference of at least 20% in methylation status with q-value<0.05 for significance.

***In situ* Hi-C**

Three independent female NSC cell lines of each *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}MommeD43}$ and *Smchd1*$^{GFP/GFP}$ genotype were used to generate *in situ* Hi-C libraries exactly as previously described[14] based on Rao et al.[76], using MboI as a restriction enzyme. Libraries were sequenced on the Illumina NextSeq platform, with 75-bp paired-end reads.

Primary-data processing was performed with HiCUP v0.7.2[77], mapping reads to the mm10 mouse genome assembly. DIs were identified with diffHiC v1.16.0[78], which uses edgeR statistics[79] (edgeR v3.26.8 and R v3.6.1). In brief, reads mapped and filtered with HiCUP were counted into 100-kb and 1-Mb bin pairs. Noise was removed by filtering out low-abundance reads on the basis of a negative binomial distribution and with interchromosomal counts to determine nonspecific ligation events. Libraries were then normalized with LOESS normalization (the counts from the matrices' diagonals were normalized separately to the rest), and trended biases were removed by fitting libraries to a generalized linear model. EdgeR was then used to test for differential interactions between genotypes, at either 100-kb or 1-Mb resolution, with a quasi-likelihood F test, and then adjustment was performed for multiple testing with FDR. To make a track of significant differential interactions for chromosome X, both sets of anchors were merged into a single 2-D track.

Generation of Hi-C contact matrices.

To construct the Hi-C interaction matrices of the *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}MommeD43}$ and *Smchd1*$^{GFP/GFP}$ genomes, we used TADbit v0.4.39[80] with the original iterative mapping strategy ICE (mm10 reference genome). The minimal size for mapping was set to 25 bp, and the iterative mapping procedure increased in steps of 5 bp until a maximal read length of 75 bp was reached. The reads were filtered with the apply_filter function from TADbit with the following parameters: maximum molecule length adjusted individually for each library to the longest insert in the 99.9% percentile, minimum distance to restriction site to be defined as a random break equal to 1.5 times the maximum molecule length, minimum fragment size of 150bp, too close to restriction site if within 4bp of it and default settings for the other parameters. Once filtered, the read pairs were binned at 100-kb or 1-Mb resolution, and columns containing few interaction counts were removed according to the two-step strategy described in Serra et al.[80]. The remaining bins were further normalized with ICE as implemented in TADbit. After checking the correlation between the 3 biological replicates of each genotype (Spearman correlation > 0.98), we merged the unnormalized reads into a single dataset for each of them. The new datasets were then normalized the same way.

TAD detection on Hi-C contact maps.

The merged interaction maps were used for domain detection at a 100-kb resolution. The TADbit program uses a breakpoint-detection algorithm that returns the optimal segmentation of the chromosome with a BIC-penalized likelihood[81]. TADbit was used to make the TAD alignment diagrams for chromosome 3, 6, 11 and X at 100kb resolution.

Compartment detection on Hi-C contact maps.

The merged interaction maps were used for compartment detection at 1-Mb resolution. The TADbit program was used to search for compartments in each chromosome by computing a correlation matrix on the normalized matrix and by using the first eigenvector to identify compartments. TADbit was used to make the Eigenvector-difference plot for chromosome X.

RNA-sequencing

For RNA-sequencing in primary NSCs of the following genotypes: *Smchd1*$^{MommeD43/MommeD43}$ vs *Smchd1*$^{+/+}$ (FVB/NJ background) and *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}MommeD43}$ vs *Smchd1*$^{GFP/GFP}$ (C57BL/6) cells from the indicated number of independent lines were harvested at an early passage (≤10) and RNA was extracted and treated with DNAse I with the Quick RNA kit (Zymo) as per the instructions. For PSM RNA-sequencing libraries, tissue was dissected from E8.5 embryos and E10.5 embryos respectively, as described in the relevant methods sections, snap-frozen, then treated the same way as the other samples. 100ng of total RNA was used to make libraries with the TruSeq RNA Library Prep kit v2 or the TruSeq stranded mRNA Library Prep kit following the manufacturers' instructions. 200- to 400-bp products were size-selected and cleaned up with AMPure XP magnetic beads. Final cDNA libraries were quantified with a D1000 or D5000 tape in the 4200 Tapestation (Agilent) and pooled for sequencing on the Illumina NextSeq platform, with 75-bp single-end reads.

RNA-sequencing analysis

Reads from the experiments mentioned above as well as previously published data from Chen et al. PNAS 2015 (*Smchd1*$^{MommeD1/MommeD1}$ vs *Smchd1*$^{+/+}$ RNA-seq in NSCs of FVB/NJ background; data available from GEO under accession number GSE65747) were trimmed for adapter sequences with TrimGalore! v0.4.4 with Cutadapt v2.9[70] then mapped with hisat2 v2.0.5[82; 83] to the GRCm38.p6 reference assembly of the mouse genome.

For the E8.5 PSM libraries and the E10.5 FNP libraries, the analysis was performed using SeqMonk v1.47.2 by importing the BAM files and using its RNA quantitation pipeline to create log2-transformed counts normalized by library size and corrected for transcript length to obtain log2 RPKM counts. Heatmaps of the log2 fold-change were made by subtracting the log2 RPKM counts from the control to the test dataset. Differential gene expression analysis on FNP samples was performed with SeqMonk's inbuilt EdgeR-based statistical test[79] with an FDR cut-off of 0.05 after excluding sex chromosomes.

Analysis of NSC RNAseq data from *Smchd1*$^{MommeD43/MommeD43}$ vs *Smchd1*$^{+/+}$, *Smchd1*$^{GFP\text{-}MommeD43/GFP\text{-}MommeD43}$ vs *Smchd1*$^{GFP/GFP}$ and *Smchd1*$^{MommeD1/MommeD1}$ vs *Smchd1*$^{+/+}$ was performed using limma v3.40.6[84] with edgeR v3.26.8[79] in R v3.6.1. Since these data were obtained from multiple experiments, we corrected for batch effects using technical replicates across batches as well as sex (all Smchd1$^{MommeD1/MommeD1}$ cells are male since *MommeD1* is an embryonic lethal mutation in homozygous females) and strain (FVB/NJ or C57BL/6). Linear models were fitted to the log2 CPM count matrices using voom with quality weights[85]. DEGs were determined using empirical Bayes' moderated *t*-tests[86] with an FDR cut-off of 0.05.

Gene Ontology analysis

GO term analysis of the differentially expressed genes found in Smchd1MommeD43/MommeD43 FNP RNA-seq was performed by uploading a list of those 56 genes to the geneontology.org GO Enrichment Analysis platform, looking for affected biological processes in *Mus musculus*.

References

1. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 125(2):315-326.
2. Kile BT, Hilton DJ. 2005. The art and design of genetic screens: Mouse. Nat Rev Genet. 6(7):557-567.
3. Allshire RC, Ekwall K. 2015. Epigenetic regulation of chromatin states in schizosaccharomyces pombe. Cold Spring Harb Perspect Biol. 7(7):a018770.
4. Pikaard CS, Mittelsten Scheid O. 2014. Epigenetic regulation in plants. Cold Spring Harb Perspect Biol. 6(12):a019315.
5. Schotta G, Ebert A, Dorn R, Reuter G. 2003. Position-effect variegation and the genetic dissection of chromatin regulation in drosophila. Seminars in Cell & Developmental Biology. 14(1):67-75.
6. Elgin SC, Reuter G. 2013. Position-effect variegation, heterochromatin formation, and gene silencing in drosophila. Cold Spring Harb Perspect Biol. 5(8):a017780.
7. Blewitt ME, Vickaryous NK, Hemley SJ, Ashe A, Bruxner TJ, Preis JI, Arkell R, Whitelaw E. 2005. An n-ethyl-n-nitrosourea screen for genes involved in variegation in the mouse. Proc Natl Acad Sci U S A. 102(21):7629-7634.
8. Blewitt ME, Gendrel AV, Pang Z, Sparrow DB, Whitelaw N, Craig JM, Apedaile A, Hilton DJ, Dunwoodie SL, Brockdorff N et al. 2008. Smchd1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in x inactivation. Nat Genet. 40(5):663-669.
9. Gendrel AV, Apedaile A, Coker H, Termanis A, Zvetkova I, Godwin J, Tang YA, Huntley D, Montana G, Taylor S et al. 2012. Smchd1-dependent and -independent pathways determine developmental dynamics of cpg island methylation on the inactive x chromosome. Dev Cell. 23(2):265-279.
10. Gendrel AV, Tang YA, Suzuki M, Godwin J, Nesterova TB, Greally JM, Heard E, Brockdorff N. 2013. Epigenetic functions of smchd1 repress gene clusters on the inactive x chromosome and on autosomes. Mol Cell Biol. 33(16):3150-3165.
11. Mould AW, Pang Z, Pakusch M, Tonks ID, Stark M, Carrie D, Mukhopadhyay P, Seidel A, Ellis JJ, Deakin J et al. 2013. Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for x inactivation. Epigenetics Chromatin. 6(1):19.
12. Wanigasuriya I, Gouil Q, Kinkel SA, Tapia Del Fierro A, Beck T, Roper EA, Breslin K, Stringer J, Hutt K, Lee HJ et al. 2020. Smchd1 is a maternal effect gene required for genomic imprinting. Elife. 9.
13. Chen K, Hu J, Moore DL, Liu R, Kessans SA, Breslin K, Lucet IS, Keniry A, Leong HS, Parish CL et al. 2015. Genome-wide binding and mechanistic analyses of smchd1-mediated epigenetic regulation. Proc Natl Acad Sci U S A. 112(27):E3535-3544.
14. Jansz N, Keniry A, Trussart M, Bildsoe H, Beck T, Tonks ID, Mould AW, Hickey P, Breslin K, Iminitoff M et al. 2018. Smchd1 regulates long-range chromatin interactions on the inactive x chromosome and at hox clusters. Nat Struct Mol Biol. 25(9):766-777.
15. Gordon CT, Xue S, Yigit G, Filali H, Chen K, Rosin N, Yoshiura KI, Oufadem M, Beck TJ, McGowan R et al. 2017. De novo mutations in smchd1 cause bosma arhinia microphthalmia syndrome and abrogate nasal development. Nat Genet. 49(2):249-255.
16. Shaw ND, Brand H, Kupchinsky ZA, Bengani H, Plummer L, Jones TI, Erdin S, Williamson KA, Rainger J, Stortchevoi A et al. 2017. Smchd1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and bosma arhinia microphthalmia syndrome. Nat Genet. 49(2):238-248.
17. Lemmers RJ, Tawil R, Petek LM, Balog J, Block GJ, Santen GW, Amell AM, van der Vliet PJ, Almomani R, Straasheijm KR et al. 2012. Digenic inheritance of an smchd1 mutation and an fshd-permissive d4z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet. 44(12):1370-1374.

18. Sacconi S, Lemmers RJ, Balog J, van der Vliet PJ, Lahaut P, van Nieuwenhuizen MP, Straasheijm KR, Debipersad RD, Vos-Versteeg M, Salviati L et al. 2013. The fshd2 gene smchd1 is a modifier of disease severity in families affected by fshd1. Am J Hum Genet. 93(4):744-751.
19. Lemmers RJ, van der Vliet PJ, Klooster R, Sacconi S, Camano P, Dauwerse JG, Snider L, Straasheijm KR, van Ommen GJ, Padberg GW et al. 2010. A unifying genetic model for facioscapulohumeral muscular dystrophy. Science. 329(5999):1650-1653.
20. van Deutekom JC, Bakker E, Lemmers RJ, van der Wielen MJ, Bik E, Hofker MH, Padberg GW, Frants RR. 1996. Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: Implications for genetic counselling and etiology of fshd1. Hum Mol Genet. 5(12):1997-2003.
21. Sacconi S, Briand-Suleau A, Gros M, Baudoin C, Lemmers R, Rondeau S, Lagha N, Nigumann P, Cambieri C, Puma A et al. 2019. Fshd1 and fshd2 form a disease continuum. Neurology. 92(19):e2273-e2285.
22. Ai S, Xiong H, Li CC, Luo Y, Shi Q, Liu Y, Yu X, Li C, He A. 2019. Profiling chromatin states using single-cell itchip-seq. Nat Cell Biol. 21(9):1164-1172.
23. Chen K, Czabotar PE, Blewitt ME, Murphy JM. 2016. The hinge domain of the epigenetic repressor smchd1 adopts an unconventional homodimeric configuration. Biochem J. 473(6):733-742.
24. Brideau NJ, Coker H, Gendrel AV, Siebert CA, Bezstarosti K, Demmers J, Poot RA, Nesterova TB, Brockdorff N. 2015. Independent mechanisms target smchd1 to trimethylated histone h3 lysine 9-modified chromatin and the inactive x chromosome. Mol Cell Biol. 35(23):4053-4068.
25. Chen K, Dobson RC, Lucet IS, Young SN, Pearce FG, Blewitt ME, Murphy JM. 2016. The epigenetic regulator smchd1 contains a functional ghkl-type atpase domain. Biochem J. 473(12):1733-1744.
26. Gurzau AD, Chen K, Xue S, Dai W, Lucet IS, Ly TTN, Reversade B, Blewitt ME, Murphy JM. 2018. Fshd2- and bams-associated mutations confer opposing effects on smchd1 function. J Biol Chem.
27. Jansz N, Chen K, Murphy JM, Blewitt ME. 2017. The epigenetic regulator smchd1 in development and disease. Trends Genet. 33(4):233-243.
28. Lemmers R, van der Stoep N, Vliet PJV, Moore SA, San Leon Granado D, Johnson K, Topf A, Straub V, Evangelista T, Mozaffar T et al. 2019. Smchd1 mutation spectrum for facioscapulohumeral muscular dystrophy type 2 (fshd2) and bosma arhinia microphthalmia syndrome (bams) reveals disease-specific localisation of variants in the atpase domain. J Med Genet. 56(10):693-700.
29. Mul K, Lemmers R, Kriek M, van der Vliet PJ, van den Boogaard ML, Badrising UA, Graham JM, Jr., Lin AE, Brand H, Moore SA et al. 2018. Fshd type 2 and bosma arhinia microphthalmia syndrome: Two faces of the same mutation. Neurology. 91(6):e562-e570.
30. Jansz N, Nesterova T, Keniry A, Iminitoff M, Hickey PF, Pintacuda G, Masui O, Kobelke S, Geoghegan N, Breslin KA et al. 2018. Smchd1 targeting to the inactive x is dependent on the xist-hnrnpk-prc1 pathway. Cell Rep. 25(7):1912-1923 e1919.
31. Wang CY, Colognori D, Sunwoo H, Wang D, Lee JT. 2019. Prc1 collaborates with smchd1 to fold the x-chromosome and spread xist rna between chromosome compartments. Nat Commun. 10(1):2950.
32. Wang CY, Jegu T, Chu HP, Oh HJ, Lee JT. 2018. Smchd1 merges chromosome compartments and assists formation of super-structures on the inactive x. Cell. 174(2):406-421 e425.
33. Gdula MR, Nesterova TB, Pintacuda G, Godwin J, Zhan Y, Ozadam H, McClellan M, Moralli D, Krueger F, Green CM et al. 2019. The non-canonical smc protein smchd1 antagonises tad formation and compartmentalisation on the inactive x chromosome. Nat Commun. 10(1):30.
34. Daxinger L, Harten SK, Oey H, Epp T, Isbel L, Huang E, Whitelaw N, Apedaile A, Sorolla A, Yong J et al. 2013. An enu mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. Genome Biol. 14(9):R96.

35. Preis JI, Downes M, Oates NA, Rasko JEJ, Whitelaw E. 2003. Sensitive flow cytometric analysis reveals a novel type of parent-of-origin effect in the mouse genome. Current Biology. 13(11):955-959.
36. Kappen C. 2016. Developmental patterning as a quantitative trait: Genetic modulation of the hoxb6 mutant skeletal phenotype. PLoS One. 11(1):e0146019.
37. McIntyre DC, Rakshit S, Yallowitz AR, Loken L, Jeannotte L, Capecchi MR, Wellik DM. 2007. Hox patterning of the vertebrate rib cage. Development. 134(16):2981-2989.
38. Wellik DM, Capecchi MR. 2003. Hox10 and hox11 genes are required to globally pattern the mammalian skeleton. Science. 301(5631):363-367.
39. Wilkie AO. 2017. Many faces of smchd1. Nat Genet. 49(2):176-178.
40. Krom YD, Thijssen PE, Young JM, den Hamer B, Balog J, Yao Z, Maves L, Snider L, Knopp P, Zammit PS et al. 2013. Intrinsic epigenetic regulation of the d4z4 macrosatellite repeat in a transgenic mouse model for fshd. PLoS Genet. 9(4):e1003415.
41. Clapp J, Mitchell LM, Bolland DJ, Fantes J, Corcoran AE, Scotting PJ, Armour JA, Hewitt JE. 2007. Evolutionary conservation of a coding function for d4z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. Am J Hum Genet. 81(2):264-279.
42. de Greef JC, Krom YD, den Hamer B, Snider L, Hiramuki Y, van den Akker RFP, Breslin K, Pakusch M, Salvatori DCF, Slutter B et al. 2018. Smchd1 haploinsufficiency exacerbates the phenotype of a transgenic fshd1 mouse model. Hum Mol Genet. 27(4):716-731.
43. Das S, Chadwick BP. 2016. Influence of repressive histone and DNA methylation upon d4z4 transcription in non-myogenic cells. PLoS One. 11(7):e0160022.
44. Haynes P, Bomsztyk K, Miller DG. 2018. Sporadic dux4 expression in fshd myocytes is associated with incomplete repression by the prc2 complex and gain of h3k9 acetylation on the contracted d4z4 allele. Epigenetics Chromatin. 11(1):47.
45. Zeng W, de Greef JC, Chen YY, Chien R, Kong X, Gregson HC, Winokur ST, Pyle A, Robertson KD, Schmiesing JA et al. 2009. Specific loss of histone h3 lysine 9 trimethylation and hp1gamma/cohesin binding at d4z4 repeats is associated with facioscapulohumeral dystrophy (fshd). PLoS Genet. 5(7):e1000559.
46. Hartweck LM, Anderson LJ, Lemmers RJ, Dandapat A, Toso EA, Dalton JC, Tawil R, Day JW, van der Maarel SM, Kyba M. 2013. A focal domain of extreme demethylation within d4z4 in fshd2. Neurology. 80(4):392-399.
47. Calandra P, Cascino I, Lemmers RJ, Galluzzi G, Teveroni E, Monforte M, Tasca G, Ricci E, Moretti F, van der Maarel SM et al. 2016. Allele-specific DNA hypomethylation characterises fshd1 and fshd2. J Med Genet. 53(5):348-355.
48. Balog J, Thijssen PE, de Greef JC, Shah B, van Engelen BG, Yokomori K, Tapscott SJ, Tawil R, van der Maarel SM. 2012. Correlation analysis of clinical parameters with epigenetic modifications in the dux4 promoter in fshd. Epigenetics. 7(6):579-584.
49. Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. 2017. Targeted degradation of ctcf decouples local insulation of chromosome domains from genomic compartmentalization. Cell. 169(5):930-944 e922.
50. Boyle S, Flyamer IM, Williamson I, Sengupta D, Bickmore WA, Illingworth RS. 2020. A central role for canonical prc1 in shaping the 3d nuclear landscape. Genes Dev. 34(13-14):931-949.
51. Balog J, Thijssen PE, Shadle S, Straasheijm KR, van der Vliet PJ, Krom YD, van den Boogaard ML, de Jong A, RJ FL, Tawil R et al. 2015. Increased dux4 expression during muscle differentiation correlates with decreased smchd1 protein levels at d4z4. Epigenetics. 10(12):1133-1142.
52. Goossens R, van den Boogaard ML, Lemmers R, Balog J, van der Vliet PJ, Willemsen IM, Schouten J, Maggio I, van der Stoep N, Hoeben RC et al. 2019. Intronic smchd1 variants in fshd: Testing the potential for crispr-cas9 genome editing. J Med Genet. 56(12):828-837.
53. Ashe A, Morgan DK, Whitelaw NC, Bruxner TJ, Vickaryous NK, Cox LL, Butterfield NC, Wicking C, Blewitt ME, Wilkins SJ et al. 2008. A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. Genome Biol. 9(12):R182.

54. Kueh AJ, Pal M, Tai L, Liao Y, Smyth GK, Shi W, Herold MJ. 2017. An update on using crispr/cas9 in the one-cell stage mouse embryo for generating complex mutant alleles. Cell Death Differ. 24(10):1821-1822.
55. Majewski IJ, Blewitt ME, de Graaf CA, McManus EJ, Bahlo M, Hilton AA, Hyland CD, Smyth GK, Corbin JE, Metcalf D et al. 2008. Polycomb repressive complex 2 (prc2) restricts hematopoietic stem cell activity. PLoS Biol. 6(4):e93.
56. Chaumeil J, Augui S, Chow JC, Heard E. 2008. 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. In: Hancock R, editor. The nucleus: Volume 1: Nuclei and subnuclear components. Totowa, NJ: Humana Press. p. 297-308.
57. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B et al. 2012. Fiji: An open-source platform for biological-image analysis. Nat Methods. 9(7):676-682.
58. Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HD, Konarev PV, Svergun DI. 2012. New developments in the atsas program package for small-angle scattering data analysis. J Appl Crystallogr. 45(Pt 2):342-350.
59. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. 2003. Primus: A windows pc-based system for small-angle scattering data analysis. Journal of Applied Crystallography. 36(5):1277-1282.
60. Svergun DI. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. Journal of Applied Crystallography. 25(4):495-503.
61. Franke D, Svergun DI. 2009. Dammif, a program for rapid ab-initio shape determination in small-angle scattering. J Appl Crystallogr. 42(Pt 2):342-346.
62. Volkov VV, Svergun DI. 2003. Uniqueness ofab initioshape determination in small-angle scattering. Journal of Applied Crystallography. 36(3):860-864.
63. Rigueur D, Lyons KM. 2014. Whole-mount skeletal staining. Methods in molecular biology (Clifton, NJ). 1130:113-121.
64. Wong SF, Agarwal V, Mansfield JH, Denans N, Schwartz MG, Prosser HM, Pourquie O, Bartel DP, Tabin CJ, McGlinn E. 2015. Independent regulation of vertebral number and vertebral identity by microrna-196 paralogs. Proceedings of the National Academy of Sciences of the United States of America. 112(35):E4884-4893.
65. McGlinn E, Mansfield JH. 2011. Detection of gene expression in mouse embryos and tissue sections. Methods in molecular biology (Clifton, NJ). 770:259-292.
66. Mohun TJ, Weninger WJ. 2012. Embedding embryos for high-resolution episcopic microscopy (hrem). Cold Spring Harb Protoc. 2012(6):678-680.
67. Weninger WJ, Maurer-Gesek B, Reissig LF, Prin F, Wilson R, Galli A, Adams DJ, White JK, Mohun TJ, Geyer SH. 2018. Visualising the cardiovascular system of embryos of biomedical model organisms with high resolution episcopic microscopy (hrem). J Cardiovasc Dev Dis. 5(4).
68. Collins CA, Zammit PS. 2009. Isolation and grafting of single muscle fibres. Methods Mol Biol. 482:319-330.
69. Nelson JD, Denisenko O, Bomsztyk K. 2006. Protocol for the fast chromatin immunoprecipitation (chip) method. Nat Protoc. 1(1):179-185.
70. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 17(1).
71. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. Nat Methods. 9(4):357-359.
72. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The sequence alignment/map format and samtools. Bioinformatics. 25(16):2078-2079.

73. Yin D, Ritchie ME, Jabbari JS, Beck T, Blewitt ME, Keniry A. 2016. High concordance between illumina hiseq2500 and nextseq500 for reduced representation bisulfite sequencing (rrbs). Genom Data. 10:97-100.
74. Krueger F, Andrews SR. 2011. Bismark: A flexible aligner and methylation caller for bisulfite-seq applications. Bioinformatics. 27(11):1571-1572.
75. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. 2012. Methylkit: A comprehensive r package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 13(10):R87.
76. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES et al. 2014. A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 159(7):1665-1680.
77. Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, Andrews S. 2015. Hicup: Pipeline for mapping and processing hi-c data. F1000Res. 4:1310.
78. Lun AT, Smyth GK. 2015. Diffhic: A bioconductor package to detect differential genomic interactions in hi-c data. BMC Bioinformatics. 16:258.
79. Robinson MD, McCarthy DJ, Smyth GK. 2010. Edger: A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26(1):139-140.
80. Serra F, Bau D, Goodstadt M, Castillo D, Filion GJ, Marti-Renom MA. 2017. Automatic analysis and 3d-modelling of hi-c data using tadbit reveals structural features of the fly chromatin colors. PLoS Comput Biol. 13(7):e1005665.
81. Le Dily F, Bau D, Pohl A, Vicent GP, Serra F, Soronellas D, Castellano G, Wright RH, Ballare C, Filion G et al. 2014. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev. 28(19):2151-2162.
82. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with hisat2 and hisat-genotype. Nat Biotechnol. 37(8):907-915.
83. Kim D, Langmead B, Salzberg SL. 2015. Hisat: A fast spliced aligner with low memory requirements. Nat Methods. 12(4):357-360.
84. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. Limma powers differential expression analyses for rna-sequencing and microarray studies. Nucleic Acids Res. 43(7):e47.
85. Law CW, Chen Y, Shi W, Smyth GK. 2014. Voom: Precision weights unlock linear model analysis tools for rna-seq read counts. Genome Biol. 15(2):R29.
86. Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 3:Article3.