The epigenetic regulator SMCHD1 in development and disease

Natasha Jansz, Kelan Chen, James M. Murphy and Marnie E. Blewitt*

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

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

*Correspondence: blewitt@wehi.edu.au

Keywords

Smchd1, facioscapulohumoral muscular dystrophy, bosma arhinia and micropthalmia syndrome, epigenetic silencing

Abstract

It has very recently become clear that the epigenetic modifier SMCHD1 has a role in two distinct disorders: facioscapulohumoral muscular dystrophy (FSHD) and Bosma arhinia and micropthalmia (BAMS). In the former, there are heterozygous loss-of-function mutations, while both gain and loss-of-function mutations have been proposed to underlie the latter. These findings have led to much interest in SMCHD1 and how it works at the molecular level. Here we summarise the current understanding of SMCHD1's mechanism of action, its role in these diseases, and what has been learnt from study of mouse models null for Smchd1 in the decade since Smchd1's discovery.

Structural maintenance of chromosomes flexible hinge domain containing 1 (SMCHD1) is a chromatin protein involved in epigenetic silencing. Recently, critical involvement in the pathogenesis of two very different developmental diseases has been attributed to SMCHD1: Facioscapulohumeral muscular dystrophy (FSHD) and Bosma Arhinia Micropthalmia Syndrome (BAMS)[1-3]. SMCHD1 is a large, non-canonical Structural Maintenance of Chromosomes (SMC) family protein, comprising an N-terminal ATPase domain, a vast central domain sharing no homology with other characterised proteins, and a C-terminal SMC hinge domain (Figure 1) [4, 5]. SMCHD1 has been implicated in various epigenetic processes, but the mechanisms by which it elicits transcriptional silencing remain unknown. A detailed understanding of how SMCHD1 functions at the molecular level will enhance our efforts to help patients affected by FSHD, and understanding the etiology of BAMS.

FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY

FSHD is a late-onset, progressive, muscular dystrophy, which first presents in the muscles of the upper extremities and follows a descending progression. In severe cases FSHD can leave patients wheelchair bound (reviewed in [6]). FSHD is the third most common neuromuscular condition, which has been estimated to affect up to 1 in 8000 people worldwide [7]. Although landmark findings have advanced our understanding of the genetic and molecular basis for FSHD in the last decade, treatment for FSHD remains largely aimed at symptomatic improvement.

FSHD is characterised by **chromatin relaxation** (see Glossary) of the D4Z4 **macrosatellite array** (see Glossary) on chromosome 4. D4Z4 is a polymorphic array, comprised of a variable number of copies (up to 100) of the 3.3kb D4Z4 repeat unit (Figure 2A). Each D4Z4 unit contains the *DUX4* **retrogene** (see Glossary), encoding a double homeobox transcription factor. Normally, *DUX4* expression is restricted to germ cells, as in somatic cells D4Z4 is adorned with repressive chromatin modifications, ensuring *DUX4* repression [8]. In FSHD however, **DNA hypomethylation** (see Glossary) of the array, loss of H3K9me3, and a gain in H3K4me3 are associated with **variegated** (see Glossary) DUX4 expression (Figure 2C) [9-11]. In addition to expressing *DUX4,* patients also need to inherit the 4qA haplotype; two allelic variants of the distal region to the D4Z4 array exist, 4qA and 4qB [12]. On the 4qB background, the *DUX4* transcript is not polyadenylated, thus

rapidly degraded. 4qA encodes a polyadenylation (poly-A) signal, which stabilises the *DUX4* transcript. Ectopic expression of *DUX4* initiates a distinct transcriptional profile, which is myotoxic and results in FSHD [13, 14].

FSHD can be classified into two subtypes based on the underlying genetic mutation. FSHD1 patients possess a contracted D4Z4 array, comprising fewer than 10 repeat units (Figure 2A) [15]. FSHD2 patients account for 5% of FSHD cases, and have no repeat contraction, but instead the vast majority harbor heterozygous loss-of-function mutations in *SMCHD1* [1]. Both genetic contractions and *SMCHD1* mutations result in chromatin relaxation of the D4Z4 array on chromosome 4. Thus, for FSHD pathogenesis, patients require digenic inheritance of either D4Z4 array contraction or a mutation in *SMCHD1*, with the 4qA allele.

Notably, FSHD1 patients possessing both a contracted D4Z4 array and a *SMCHD1* mutation present with a more severe clinical phenotype than would be expected for array contraction alone [16]. *SMCHD1* mutations in FSHD patients map to positions along the full-length of the SMCHD1 protein (Figure 1). While most missense mutations have not been functionally characterised, deletions and nonsense mutations are clearly loss-of-function, which suggests that SMCHD1 normally has a role in repressing *DUX4* and maintaining heterochromatin at D4Z4. Accordingly, a non-pathogenic D4Z4 array on chromosome 10 lacking the proximal poly-A signal is derepressed in FSHD2 but not FSHD1 patients [1].

BOSMA ARHINIA MICROPTHALMIA SYNDROME

BAMS is the congenital absence of the nose and a reduction in eye size, which is often accompanied by a series of other malformations [17, 18]. BAMS was formally described by Bosma in 1981 who observed the above symptoms in two unrelated boys with healthy parents [18]. A rare condition, it has since been reported in just 50 patients worldwide. Arhinia poses problems for affected individuals from birth, with extensive surgery often required for BAMS patients from a young age to prevent structural abnormalities from hampering brain development, and to alleviate respiratory problems [19, 20].

It is thought that Arhinia arises due to failed fusion of the maxillary and lateral nasal processes, and associated cribiform plate fusion abnormalities in embryonic development [19]. Although some major chromosomal aberrations have been reported in patients with Arhinia, until recently a genetic cause for BAMS remained elusive [21]. Next generation sequencing has allowed two groups to identify *de novo* *SMCHD1* mutations in BAMS patients [2, 3]. Strikingly all the mutations lie within highly conserved residues of the ATPase domain and the adjacent region of SMCHD1, some of which are predicted to be critical for ATPase activity (Figure 1). Recombinant proteins harboring a selection of patient-derived mutations within the ATP binding pocket of SMCHD1 showed an enhanced capacity to hydrolyse ATP, suggesting these are gain-of-function mutations [2]. Furthermore, injecting *SMCHD1* transcripts encoding mutations found in BAMS patients into *Xenopus* oocytes resulted in a small eye phenotype, reminiscent of that seen in BAMS patients [2]. This phenotype was mimicked by the overexpression of wild-type *SMCHD1*, whereas injection of *SMCHD1* transcripts harbouring FSHD-associated mutations did not manifest this phenotype. While the biochemical and *in vivo* data presented by Gordon *et al.* [2] together suggest mutations found in BAMS patients enhance SMCHD1 function, both this study and that of Shaw *et al.* [3] have shown D4Z4 hypomethylation in some BAMS patients, although to a lesser extent than found in FSHD [2, 3]. In contrast to Gordon *et al.*, Shaw and colleagues found that knocking down or knocking out *Smchd1* in zebrafish resulted in a small-eye phenotype and craniofacial defects, although these defects could not be recapitulated in mouse upon introduction of a BAMS mutation [3]. Moreover, mice both heterozygous and homozygous for a nonsense mutation in *Smchd1* (described in the next section) do not display an overt eye or nose phenotype [22].

Delineating whether *SMCHD1* polymorphisms identified in FSHD and BAMS patients cause loss- or gain-of-SMCHD1 function remains challenging. A major difficulty facing BAMS researchers, as is the case with any rare developmental disorder, is the accessibility and reliability of relevant patient samples and controls. For example, a gene expression analysis was performed in lymphoblastoid cell lines from some arhinia patients and controls, but no significant changes were found [3]. These data are consistent with BAMS arising during a defined window *in utero* with no effects in blood cells, but it does not enhance our understanding of how BAMS

patient mutations found in SMCHD1 alter its function. Similarly, the DNA methylation analysis of D4Z4 was performed in peripheral blood leucocytes, however familial, age-matched controls for young Arhinia patients were not available, and hypomethylation relative to age was not accounted for [2, 3]. To overcome difficulties with using patient samples, both groups drew conclusions using non-mammalian model organisms to characterise patient-derived mutations. Thus, differing cellular or organismal context underlying craniofacial development might also explain the seemingly incongruous data presented by the two groups. It is also worth highlighting that while Gordon and colleagues observed enhanced ATPase activity in proteins encoding some, but not all, mutants the function of SMCHD1's ATPase domain is unknown, and enhanced ATPase activity does not necessarily equate to enhanced silencing capacity.

The genetic and clinical data also highlight the complex relationship between SMCHD1 mutations and disease presentation; on one hand, FSHD2 patients are born without facial abnormalities, arguing that haploinsufficiency for SMCHD1 function alone is not enough to result in BAMS. On the other hand, a Gly137Glu mutation has been identified in both an FSHD and a BAMS patient, and an individual with an Asn139His mutation has been reported to display symptoms of both syndromes [3]. However in neither patient has DUX4 expression been measured, so it is also possible that these patients have another clinically similar disorder, such as limb girdle muscular dystrophy, which is sometimes misdiagnosed as FSHD [23]. So while the weight of evidence suggests FSHD2 is caused by loss-of-function mutations in SMCHD1, it would appear that a loss- versus gain-of-function dichotomy for FSHD versus BAMS could be too simplistic to cover all cases.

Further characterisation of *SMCHD1* missense mutants will help us to understand the pathogenesis of disease, but therapeutic intervention will likely rely on an improved understanding of the molecular mechanisms by which SMCHD1 regulates transcription. Much of our understanding of SMCHD1 function has come from studying model epigenetic processes in mouse since its discovery in 2008. However, recent advances and access to next generation sequencing technology are enabling insights into how Smchd1 may be regulating transcription.

SMCHD1 AS A TRANSCRIPTIONAL REPRESSOR

Much like human SMCHD1 functions at the D4Z4 array, Smchd1 was first implicated in epigenetic control through its role in repeat-induced silencing of a murine multicopy GFP transgene, for which expression is also variegated (Figure 2B) [24]. This strain was used in an **ENU** (see Glossary) mutagenesis screen to find novel epigenetic modifiers. The Modifier of Murine Metastable Epialleles Dominant 1 (MommeD1) line generated in this screen harbors a nonsense mutation in *Smchd1*, resulting in nonsense-mediated decay of the *Smchd1* transcript [22]. MommeD1 mice showed a dose dependent increase in transgene expressing cells, suggesting that Smchd1 is a transcriptional repressor (Figure 2D).

The MommeD1 strain produced heterozygous mutants at expected ratios, and viable homozygous males, albeit at lower than expected numbers, however notably no homozygous females survived beyond mid-gestation [24]. Closer inspection of E10.5 females revealed hypomethylation at the **promoter** (see Glossary) of the X-linked gene *Hprt* in both *Smchd1*$^{MommeD1/MommeD1}$ females, and to a lesser extent in heterozygous females [24]. These observations suggested a role for Smchd1 in X chromosome inactivation (XCI), which was soon validated and has been the focus of research on Smchd1 (see Box 1).

SMCHD1 IS CRITICAL FOR X CHROMOSOME INACTIVATION

In the absence of Smchd1, both random XCI in the embryo and imprinted XCI in the placenta fail [22]. In the embryo, the early stages of XCI proceed normally, indicated by accumulation of *Xist* and H3K27me3 on the inactive X chromosome (Xi). However, there is a failure to properly establish or maintain silencing on the Xi, indicated by the developmental window in which *Smchd1*$^{MommeD1/MommeD1}$ female embryos die [22]. This observation is supported by *in vitro* data; knockdown of *Smchd1* in Mouse Embryonic Fibroblasts (MEFs) results in the reactivation of a GFP transgene on the Xi [25]. The most prominent phenotype in the absence of Smchd1 is a dramatic DNA hypomethylation at CpG islands (CGI) on the Xi, associated with the upregulation of a subset of X-linked genes that are methylated later in the ontogeny of X inactivation [26-28]. However, *Dnmt3b* null embryos display widespread X-linked CGI hypomethylation relative to *Smchd1* null embryos, but do not upregulate the same subset of X-linked genes, suggestive of other mechanisms involved in Smchd1

mediated silencing on the Xi [26]. Accordingly, immunofluorescence assays for Smchd1 in both mouse and human female somatic interphase nuclei have found Smchd1 enriched on the Xi, suggesting a direct and continued role for Smchd1 in the maintenance of XCI [22, 29]. ChIP-seq data in human cells shows that SMCHD1 resides over domains of the Xi enriched for H3K27me3 and *XIST*, and interacts with H3K9me3 domains to bring about compaction of the Xi [29].

SMCHD1 REGULATES AUTOSOMAL MONOALLELIC GENE EXPRESSION

Despite Smchd1's prominent role in XCI, and the viability of *Smchd1* null male mice on some genetic backgrounds, Smchd1's ubiquitous expression in male and female cells suggests a broader role for Smchd1 in regulating transcription [22, 24]. Global expression analyses of cells and embryos derived from male *Smchd1*$^{MommeD1/MommeD1}$ mice have shown that in the absence of Smchd1 autosomal **monoallelic gene expression** (see Glossary) is perturbed at some imprinted clusters, and at the clustered protocadherins (see Boxes 2 and 3, respectively) [27, 28].

Loss of Smchd1 results in biallelic expression of transcripts in the *Snrpn* cluster that are associated with somatic differentially methylated regions (sDMR), whereas expression of imprinted genes that are only under the control of the primary imprint control region (ICR, see Glossary) is not affected. Accordingly, while sDMRs are hypomethylated in the absence of Smchd1, methylation of the ICR is not affected, suggesting Smchd1 is involved in establishing methylation post-implantation [27, 28]. ChIP-seq data has shown that Smchd1 binds at the *Snrpn* locus, suggesting Smchd1 has a direct role in silencing genes in this cluster [30].

At the *Igf2r* imprinted cluster, imprinted expression of *Slc22a3* is lost in placentae lacking Smchd1, while *Igf2r* imprinting is unaffected [28]. The *Igf2r* cluster is under the control of an ICR in exon 2 of *Igf2r*, however *Slc22a3* imprinting is regulated by the expression of the *Airn* lncRNA from the paternal allele, which targets the H3K9 methyltransferase G9a to the locus [31]. Loss of Smchd1 does not alter the differential methylation at the *Igf2r* ICR, raising the possibility that Smchd1 may be involved in *Airn*-H3K9me3 directed silencing [28].

The clustered protocadherins are subject to random, combinatorial monoalleleic gene expression, which is not parent-of-origin specific (Box 3). In *Smchd1* null cells protocadherin genes in all three clusters are upregulated, particularly in the *Pcdh-α* and *Pcdh-β* cluster [27, 28]. Methylation at the CpG islands of the clustered *Pcdh* in Smchd1-deficient cells was significantly reduced, suggesting that Smchd1 is involved in maintaining methylation in this region. Smchd1 directly binds the **enhancer** (see Glossary) and promoters of the clustered protocadherins, suggesting a direct role for Smchd1 in regulating their expression [30].

As described above, studies in mouse models have revealed an important role for Smchd1 in regulating several forms of monoalleic gene expression, which are regulated by diverse mechanisms. However, at all loci examined, the absence of Smchd1 leads to upregulation of genes within stably silenced **facultative heterochromatin** (see Glossary), disturbances to the local chromatin environment, most notably DNA hypomethylation, and alterations to histone modifications. Further investigation into the molecular mechanisms underlying Smchd1 function will enlighten us as to whether the changes to the chromatin are direct or indirect consequences of the loss of Smchd1 at these loci.

MOLECULAR MECHANISMS

Recent advances in our understanding of Smchd1 structure and function, coupled with loss-of-function studies, have provided the first glimpses into the mechanisms underpinning Smchd1 molecular function. Such an understanding would facilitate development of therapeutic strategies to counter FSHD.

Smchd1 is a non-canonical member of the SMC protein family, possessing a C-terminal SMC hinge domain and an N-terminal ATPase domain [4, 5]. SMC proteins heterodimerise to form specific complexes involved in large-scale chromatin organisation, gene regulation, and DNA repair. Heterodimerisation occurs through both the ATPase domains in the presence of ATP, and the SMC hinge domains. Together with Kleisin subunits, SMC proteins form a characteristic ring structure that facilitates interactions with chromatin (reviewed in [32]).

Unlike canonical SMC proteins, the C-terminal hinge domain of SMCHD1 potentially mediates homodimerisation of the full-length protein via a divergent dimeric arrangement flanked by intermolecular coiled-coils [4]. The N-terminal region of Smchd1, containing the GHKL-ATPase domain, appears to adopt an elongated conformation that resembles the full-length structure of Heat shock protein 90 (Hsp90) [5]. Thus, it is possible that the N-terminal region and C-terminal hinge domain are connected via the middle-region to form a head-to-head SMCHD1 homodimer that is arguably reminiscent of the overall topology of the ring-like SMC complexes. Further evidence for this homodimeric conformation comes from negative-staining electron microscopy images of full-length recombinant Smchd1 [33].

The identification of Smchd1 binding sites genome-wide by ChIP-seq suggests that Smchd1 directly interacts with chromatin to regulate transcription [30]. Similarly, ChIP data in humans shows SMCHD1 enriched at the D4Z4 locus, and such enrichment is diminished in FSHD patient samples, concomitant with D4Z4 derepression [1]. SMCHD1's chromatin association is also evident by its localization to the Xi, as shown by immunofluorescence and ChIP [22, 29]. Furthermore, *in vitro* data generated using recombinant protein show that the hinge domain of Smchd1 has the capacity to directly bind synthetic oligonucleotides, and that this capability was abrogated when a mutation found in an FSHD2 patient was introduced into the recombinant protein [30]. It has also been shown that the hinge domain is required for Smchd1 to remain bound to the chromatin in a cellular context [33]. Taken together these data suggest a direct interaction between the hinge domain of Smchd1 and the chromatin is required for silencing at target loci.

The question then remains as to how Smchd1 is directed to target loci. Several models have been proposed, but further investigation is required to determine whether these models are exclusive or there are in fact multiple mechanisms by which Smchd1 is targeted to chromatin. Studies in both mouse and human cells have found that Smchd1 interacts with ligand-dependent nuclear receptor-interacting factor 1(LRIF1) and its human homologue HBiX1, and that through heterochromatin protein 1 (HP1), this interaction directs Smchd1 to bind H3K9me3 marked chromatin [29, 33]. It has been proposed that this interaction is necessary for targeting on autosomes, but not for

SMCHD1 targeting to the Xi despite the interaction between SMCHD1 and LRIF1 still existing on the Xi [33]. Indeed, when HBiX1 is knocked down in human cells, SMCHD1 accumulation over H3K9me3 enriched regions of the Xi is lost, however SMCHD1 continues to interact with regions enriched for H3K27me3 and *XIST*, suggesting SMCHD1 can interact with the chromatin of the Xi independently of H3K9me3 [29]. Furthermore, while H3K9me3 is found throughout the genome, Smchd1's genome-wide occupancy is limited to 227 high-confidence bindings sites in male neural stem cells, suggesting that H3K9me3 enrichment alone is not sufficient for Smchd1 targeting [30].

The colocalisation of SMCHD1 with *XIST* on the Xi has highlighted that SMCHD1 could interact with *XIST*, and more generally with long non-coding RNA (**lncRNA** see Glossary) for chromatin targeting. Notably, lncRNAs are involved in regulating transcription at many loci where Smchd1 functions. SMCHD1 localisation to the Xi is *XIST* dependent, and Smchd1 was also identified as an interaction partner of *Xist* in a screen performed in MEFs [29, 34]. On the other hand, Smchd1 enrichment on the Xi is delayed relative to *Xist* upregulation in differentiating mESC, suggesting that *Xist* does not directly, or at least immediately recruit Smchd1 to the Xi, consistent with Smchd1 not being found in other screens for *Xist* interactors performed early during XCI [35, 36]. However, in support of a direct binding model, Smchd1 recombinant hinge domain has been shown to interact with synthetic RNA oligonucleotides *in vitro* [30]. The potential interaction between *Xist* and Smchd1 needs to be directly addressed, however it is an intriguing possibility that Smchd1 may be targeted to the chromatin by interactions with RNA.

A role for SMCHD1 in higher order chromatin organisiation was first proposed by Nozawa *et al.* in response to an observed decompaction of the Xi upon SMCHD1 knock down [29]. This decompaction was attributed to a lost interaction between SMCHD1 and HBiX, disturbing the bridge between H3K27me3 and H3K9me3 domains of the Xi. More recently, motif analysis of Smchd1 ChIP peaks found that Smchd1 binding overlaps with CCCTC-binding factor (Ctcf) occupancy at promoters and distal cis-regulatory elements [30]. Chen *et al.* demonstrated that Smchd1 and Ctcf have opposing effects on expression of the clustered protocadherins, raising the compelling possibility that functional antagonism exists between Smchd1 and Ctcf. It

remains to be seen whether this opposing effect on transcription can be seen genome-wide; however half of the Smchd1 binding sites found in neural stem cells overlap with Ctcf sites, and it has been found that CTCF has the opposite effect of SMCHD1 in FSHD [30, 37]. These data raise the possibility that Smchd1 may be involved in mediating long-range chromatin interactions to regulate transcription given CTCF has been shown to orchestrate chromatin looping [38]. It is possible that Smchd1 is involved in ATP-dependent chromatin remodelling like other SMC proteins, given it contains a functional ATPase domain [5, 33]. Considering a number of BAMS mutations can elevate recombinant Smchd1's ATPase activity, it would be interesting to investigate how Smchd1 utilises its ATPase activity to engage with chromatin or other chromatin proteins at the target sites to elicit epigenetic control.

MODEL

Smchd1 functions at loci that are subject to stable and heritable silencing, which employ multiple epigenetic mechanisms to ensure silencing is maintained. When Smchd1 is lost from these loci, there are widespread changes to the local chromatin environment, most markedly a dramatic loss of DNA methylation. At many Smchd1 target enhancers and promoters, Smchd1 and Ctcf appear to have opposing roles. Indeed, while Ctcf preferentially binds unmethylated sequences, Smchd1 had a preference for the methylated sequence [30]. Taken together, these data suggest that Smchd1 might have a role in establishing and/or maintaining a repressive chromatin structure, potentially by keeping distal enhancers away from associated promoters in an energy dependent fashion. When Smchd1 binding is diminished, perturbations to these chromatin interactions could lead to a destabilisation of the local chromatin environment. The resulting chromatin relaxation could create a transcriptionally permissive state where enhancers and promoters could interact (Figure 3). Studying the higher order chromatin organisation in the absence of Smchd1, integrated with chromatin profiling, will help to illuminate Smchd1's mechanistic role in these processes.

HOW UNDERSTANDING MECHANISMS MAY HELP DISEASE

SMCHD1 is an interesting case in which mutations that alter SMCHD1 function drive divergent human diseases, characterised by distinct disease onset and affected tissues. While BAMS is a congenital disorder where treatment to inhibit SMCHD1 would not

be of therapeutic benefit, FSHD can be diagnosed at the early stages of disease progression. The discovery of mutations that enhance the ATPase activity of SMCHD1 in BAMS patients raises the possibility that SMCHD1 has the potential to be activated, which could be exploited in developing a treatment for FSHD. FSHD2 patients are heterozygous for mutations in SMCHD1, thus the wild-type copy of SMCHD1 in these patients could be targeted for activation, in an attempt to overcome the effects of SMCHD1 haploinsufficiency or potential dominant negative mutations driving disease. As SMCHD1 is a modifier of disease severity in FSHD1, activating SMCHD1 also has the potential to relieve FSHD1 patients. Further studies in BAMS patients and in mouse models are needed to understand the molecular consequences of enhancing the ATPase activity of SMCHD1, particularly in regards to differential gene expression and the local chromatin environment. These studies will not only teach us about how SMCHD1 functions normally and in disease, but will inform us as to whether enhancing the ATPase activity of SMCHD1 could enhance its silencing capacity, and potentially reverse the effects of loss of SMCHD1 function at the D4Z4 locus in FSHD, and ultimately guide the development of treatments for FSHD patients.

X CHROMOSOME INACTIVATION BOX

X chromosome inactivation (XCI) is a dosage compensation mechanism that has evolved to equalize expression of X linked genes between males (XY) and females (XX) (reviewed in [39]). In therian mammals there are two forms of X chromosome inactivation: imprinted and random X chromosome inactivation. Imprinted XCI occurs in metatherian mammals (marsupials) and in the extraembryonic tissues of the developing mouse, whereas random XCI occurs in the cells that proceed to form the embryo proper of eutherian mammals [40, 41]. In each case, XCI consists of very well defined, but overlapping, stages of initiation, establishment and maintenance of gene silencing. Silencing of one of the two X chromosomes in female cells is initiated by the upregulation of the long noncoding RNA *Xist* from the X chromosome that is to be inactivated, which triggers large-scale chromatin remodeling that results in gene silencing [42, 43]. Initially *Xist* coats the future inactive X (Xi) *in cis*, and RNA polymerase II is excluded from the region of the Xi, establishing a silent nuclear compartment, into which repeats and genes are recruited [44]. Active histone modifications, such as histone H3 lysine 4 acetylation (H3K4ac) are removed, while repressive histone modifications, such as histone H2A lysine 119 mono-ubiquitylation (H2AK119ub1) and H3K27 trimethylation (H3K27me3) catalysed by polycomb repressive complexes 1 and 2 respectively, and H3K9me3 catalysed by Setdb1, are accumulated [25, 45-47]. The maintenance phase of XCI is marked by a shift into late replication timing of the Xi, enrichment of the histone variant macroH2A, Smchd1 recruitment, and the accumulation of DNA methylation [22, 48-50]. The many redundant chromatin changes that occur on the Xi ultimately result in mitotically heritable silencing of the whole chromosome, which is stable for the lifetime of the organism, maybe a hundred years in some mammals.

GENOMIC IMPRINTING BOX

Genomic imprinting is an epigenetic phenomenon that describes the monoallelic expression of genes in a parent-of-origin specific fashion (reviewed here [51, 52]). There are around 150 imprinted genes in both humans and mice. Imprinted genes often encode proteins that are important in growth and development, and loss of imprinting in humans results in various developmental syndromes, and contributes to the development of cancer [53, 54]. Imprinted genes exist in clusters throughout the genome, and each of these clusters is regulated by discrete elements, called imprint control regions (ICR) [55]. The primary imprint is almost universally associated with DNA methylation at the ICR. Genomic imprints are established during primordial germ cell development, allowing for specific parent-of-origin marks to be established in the sperm and egg [56, 57]. After fertilization, the parent-of-origin imprints are maintained in somatic cells of the embryo and throughout the life of the organism, allowing for differential expression between maternal and paternal alleles [58]. Interestingly, methylation of ICRs does not always associate with a particular transcriptional outcome; each ICR functions via different mechanisms, so the outcome of differential DNA methylation will differ in each cluster [59]. Imprinted expression can also vary between developmental stages and different tissues, for example there are many genes that are subject to genomic imprinting the placenta, which are biallelically expressed in the developing and adult mouse [60-62]. There are also regions within imprinted clusters known as somatic differentially methylated regions (sDMR), which are established in the post-implantation embryo [63]. They are still established in a parent-of-origin manner, directed by the primary ICR, and in a similar fashion are stably maintained through mitosis. It has been suggested that sDMRs evolved to allow the coordinate imprinted expression of genes within the same cluster.

CLUSTERED PROTOCADHERINS (PCDH) BOX

The clustered Pcdh are neuronal membrane molecules, which are thought to be important for neuronal circuit assembly. In mice, a total of 58 Pcdh isoforms are encoded in a large cluster on chromosome 18, arranged into three groups; *Pcdh-α*, *Pcdh-β* and *Pcdh-γ*, encoding 14, 22 and 22 members respectively [64]. In a single neuron, one isoform from the *Pcdh-α* cluster, and two from each of the *Pcdh-β* and *Pcdh-γ* clusters are expressed from each allele. Each isoform is composed of a single large variable exon, and in the case of the *Pcdh-α* and *Pcdh-γ* families, three constant exons. The variable exons encode the extracellular and transmembrane domains of the protein, while the three constant exons encode the intracellular domain [64, 65]. Expression of each isoform is monoallelic, and occurs in a random (that is, not parent-of-origin specific), combinatorial fashion. This enables each neuron to express a unique combination of isoforms, such that individual neurons possess a unique cell surface identity. In this way, expression of the clustered Pcdh is thought to provide the molecular basis for neuronal diversity [66].

Isoform choice is both transcriptionally and post-transcriptionally regulated by stochastic promoter choice and pre-mRNA spicing events, respectively [38, 65]. In the case of the α-cluster, a variable exon is expressed due to alternative promoter choice, and is then spliced to the most proximal constant exon. It has been shown that Ctcf and the cohesin complex are important in regulating alternative promoter choice; each *Pcdh* gene shares a conserved promoter, within which is a Ctcf binding site. Ctcf interacts with cohesin to mediate chromatin looping events that bring together a *Pcdh* promoter with distal enhancer elements (e.g. the hypersensitive site co-occupied by Smchd1), permitting expression of the variable exon [30, 38].

Trends Box

- The transcriptional repressor Smchd1 hydrolyses ATP through its N-terminal GHKL ATPase domain and directly binds oligonucleotides through its C-terminal hinge domain.
- Loss-of-function mutations throughout SMCHD1 underlie the late-onset, progressive muscular dystrophy FSHD2 and modify disease severity in FSHD1 patients.
- Missense mutations within, or proximal to, the SMCHD1 ATPase domain have been found in Bosma Arhinia Micropthalmia Syndrome patients.
- Whether mutations in BAMS enhance or suppress SMCHD1 function remains a matter of controversy.
- Smchd1 occupies distinct loci genome wide, and loss of Smchd1 results in altered chromatin modifications, most markedly DNA hypomethylation, and changes to gene expression.
- Smchd1 shares binding sites with Ctcf, and at one characterised locus, the clustered protocadherins, Smchd1 and Ctcf mediate opposing effects on gene expression.

Outstanding Questions

What are the molecular mechanisms by which SMCHD1 is involved in mediating transcriptional repression?

Is the observed DNA hypomethylation a direct effect of loss of SMCHD1 function in both human and mouse, or an indirect consequence due to disruptions to chromatin organisation?

Can chromosome conformation capture techniques be used to address whether Smchd1 has a role in higher order chromatin organisation, which may explain the functional antagonism observed between Smchd1 and Ctcf at the clustered protocadherins?

Would depletion of Smchd1 in the oocyte, which has not been studied due to the female specific lethality in the absence of Smchd1, result in disruption of imprinted genes under the control of a primary ICR?

What effect does enhanced SMCHD1 ATPase activity have on gene expression?

How does enhanced ATPase activity alter the chromatin landscape surrounding Smchd1 binding sites, including chromatin modifications and chromatin conformation?

How do mutations in the N-terminal region of SMCHD1 alter SMCHD1's function and result in BAMS?

What is the developmental window in which enhanced SMCHD1 function is detrimental to craniofacial development?

Could modulation of the ATPase activity of SMCHD1 rescue loss of SMCHD1 function in FSHD patients?

What are the respective contributions of catalysis or scaffolding by the ATPase domain to SMCHD1 function?

Besides HBiX1 and *Xist*, does SMCHD1 nucleate interactions with other proteins and nucleic acids?

Are there other SMCHD1 target genes that play tissue- and context-dependent roles in development and disease yet to be identified?

Could small molecules be identified to augment SMCHD1's ATPase activity?

Is the central region connecting the N-terminal region and C-terminal hinge domain of SMCHD1 a structured domain and what function does it serve?

GLOSSARY

Chromatin relaxation: The loss of heterochromatic chromatin modifications, such as DNA methylation, and histone 3 lysine 9 tri-methylation (H3K9me3), and the simultaneous accumulation of euchromatic chromatin modifications, such as histone 3 lysine 4 trimethylation (H3K4me3), which is often accompanied by transcriptional upregulation.

DNA hypomethylation: the loss of the methyl group from 5-methylcytosines.

Enhancers: Regulatory sequences of DNA that can be bound by transcription factors and epigenetic modifiers and enhance transcription of an associated gene.

ENU: N-ethyl-N-nitrosourea is a potent mutagen often used for forward genetic screens. ENU targets all proliferating cells, including spermatogonial stem cells, meaning offspring of mice treated with ENU can be screened for a phenotype of interest, and a candidate gene can then be mapped and identified.

Facultative heterochromatin: Heterochromatin that differs by cell type or stages in the cell cycle.

Imprint control region (ICR): A region of DNA found within a cluster of imprinted genes that controls the imprinted expression of the cluster by means of alternate DNA methylation states on the maternal versus paternal allele.

Long non-coding RNA (lncRNA): a diverse group of transcripts that are longer than 200 nucleotides, and do not encode a protein. lncRNAs have diverse, and critical roles in regulating gene expression.

Macrosatellite array: An array of tandemly repeated DNA. The number of tandem repeat units is variable amongst individuals. Macrosatellites have large repeat units (on average around 3000 kb) and span several hundred kilobases. While macrorepeats are not usually coding, occasionally a gene has retrotransposed into the repeat, e.g. *DUX4* in the D4Z4 repeat array.

Monoalleleic gene expression: Expression from a single allele in a diploid cell.

Promoter: The regulatory region at the start of a gene, from where transcription is initiated.

Retrogene: A DNA fragment that has been inserted into the genome, following the reverse-transcription of an mRNA, a process known as retrotransposition.

Variegation: Mosaic expression of a gene between cells of the same type, i.e. a particular gene will be expressed in one cell, but at a lower level or not at all in another cell in the same population.

Figure 1: Schematic representation of a SMCHD1 homodimer

The N-terminal GHKL ATPase domain is shown in green and white, the uncharacterised central domain in depicted as a black, dashed line, the C-terminal hinge domain is in blue, and the flanking N- and C- terminal coiled coil regions are depicted in purple and yellow, respectively, forming intermolecular coils. Based on data from [4, 5, 33]. Mutations found in FSHD2 patients are shown above Smchd1, and mutations found in BAMS patients are shown below. Asterisk indicates Gly137Glu mutation identified in both an FSHD and a BAMS patient. FSHD mutations from the LOVD database [67], and BAMS mutations from [2, 3].

Figure 2: SMCHD1 is involved in repeat induced silencing in both human and mouse

A. The polymorphic D4Z4 macrosatellite array on chromosome 4. Genetic and epigenetic contexts underlie different patient phenotypes. SMCHD1 shown as homodimeric protein binding D4Z4. D4Z4 repeat units are depicted by green (normal) and blue (diseased) circles. Permissive 4qA or non-permissive 4qB alleles are indicated by black boxes. CpG methylation (black circles) or hypomethylation (white circles) is indicated. Normally, D4Z4 arrays of >10 units are CpG methylated, and *DUX4* is silenced. Array contraction to <10 D4Z4 units results in hypomethylation and *D4Z4* expression, but only results in FSHD1 on a 4qA haplotype, as the poly-A signal stabilises the *DUX4* transcript. Hypomethylation of D4Z4 in combination with reduced SMCHD1 binding, allows *DUX4* expression on a 4qA haplotype, consequently FSHD2. Mutations in SMCHD1 modify severity of disease in FSHD1 patients. Schematic representation of data from [1].

B. Expression from the GFP11 transgene array linked to the α-globin promoter is regulated by Smchd1 [24]. GFP monomers depicted in green. Smchd1 binding to *GFP11* has not been shown.

C. Chromatin relaxation of D4Z4 on the 4qA allele results in variegated expression of DUX4. Human muscle cells shown with blue nuclei. Red nuclei represent DUX4 expressing cells, and intensity of red represents DUX4 expression levels.

D. Mutations in Smchd1 reduce transgene variegation in a dose dependent fashion. The GFP11 transgene is expressed in 55% of *Smchd1*$^{+/+}$ erythrocytes, 82% of *Smchd1*$^{MommeD1/+}$ erythrocytes and up to 99% of *Smchd1*$^{MommeD1/MommeD1}$ erythrocytes. Representation of data in [22].

Figure 3: Our proposed model for SMCHD1 function

Based on data from [30]. A schematic representation of a spectrum of SMCHD1 function, in which altered SMCHD1 function could create either a transcriptionally repressive or permissive chromatin environment, depending on genetic and biochemical background. At one end of the spectrum, interactions between promoters and distal regulatory elements are prevented, directly mediated by Smchd1. When SMCHD1 is lost, either through reduced expression or abrogation of DNA binding capabilities, a transcriptionally permissive chromatin environment is created, allowing distal enhancers to interact with their promoters. SMCHD1 is represented as homodimeric protein in green and blue, bound to a promoter depicted in purple, and its enhancer in blue. DNA is shown in black, and DNA methylation (black circles) or hypomethylation (white circles) is indicated. Heterochromatic histone modifications are shown in red, and euchromatic in green. Other chromatin proteins are shown as a solid blue or green ellipse. Active transcription from a promoter is depicted as a pointed arrowhead, and a silent promoter as a flat arrowhead. Wavy lines represent nascent transcripts.

ACKNOWLEDGEMENTS

NJ was supported by an Australian Postgraduate Award from the University of Melbourne. KC was supported by a Cancer Council Victoria Postdoctoral Fellowship. JMM and MEB were supported by National Health and Medical Research Council of Australia (NHMRC) RD Wright Fellowships (APP1105754 and APP1110206 respectively). The studies from the authors' labs reviewed herein were enabled by NHMRC Project grants 1045936 and 1098290, with additional support from NHMRC IRIISS 9000220 and Victoria Government Operational Infrastructure Support. We thank colleagues for valuable discussions and apologise to those whose studies we were unable to discuss here due to space limitations.

REFERENCES

1 Lemmers, R.J., *et al.* (2012) Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nature genetics* 44, 1370-1374

2 Gordon, C.T., *et al.* (2017) De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. *Nature genetics*

3 Shaw, N.D., *et al.* (2017) SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. *Nature genetics*

4 Chen, K., *et al.* (2016) The hinge domain of the epigenetic repressor Smchd1 adopts an unconventional homodimeric configuration. *The Biochemical journal* 473, 733-742

5 Chen, K., *et al.* (2016) The epigenetic regulator Smchd1 contains a functional GHKL-type ATPase domain. *The Biochemical journal* 473, 1733-1744

6 Daxinger, L., *et al.* (2015) Genetic and epigenetic contributors to FSHD. *Current opinion in genetics & development* 33, 56-61

7 Deenen, J.C., *et al.* (2014) Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* 83, 1056-1059

8 Snider, L., *et al.* (2010) Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS genetics* 6, e1001181

9 Balog, J., *et al.* (2012) Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. *Epigenetics* 7, 579-584

10 Zeng, W., *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 genetics* 5, e1000559

11 Lemmers, R.J., *et al.* (2015) Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. *Human molecular genetics* 24, 659-669

12 Lemmers, R.J., *et al.* (2002) Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nature genetics* 32, 235-236

13 Kowaljow, V., *et al.* (2007) The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscular disorders : NMD* 17, 611-623

14 Geng, L.N., *et al.* (2012) DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Developmental cell* 22, 38-51
15 Wijmenga, C., *et al.* (1992) Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nature genetics* 2, 26-30
16 Sacconi, S., *et al.* (2013) The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. *American journal of human genetics* 93, 744-751
17 Brasseur, B., *et al.* (2016) Bosma arhinia microphthalmia syndrome: Clinical report and review of the literature. *American journal of medical genetics. Part A* 170A, 1302-1307
18 Bosma, J.F., *et al.* (1981) Hypoplasia of the nose and eyes, hyposmia, hypogeusia, and hypogonadotrophic hypogonadism in two males. *Journal of craniofacial genetics and developmental biology* 1, 153-184
19 Graham, J.M., Jr. and Lee, J. (2006) Bosma arhinia microphthalmia syndrome. *American journal of medical genetics. Part A* 140, 189-193
20 Feledy, J.A., *et al.* (2004) Vertical facial distraction in the treatment of arhinia. *Plastic and reconstructive surgery* 113, 2061-2066
21 Sato, D., *et al.* (2007) Congenital arhinia: molecular-genetic analysis of five patients. *American journal of medical genetics. Part A* 143A, 546-552
22 Blewitt, M.E., *et al.* (2008) SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nature genetics* 40, 663-669
23 Leidenroth, A., *et al.* (2012) Diagnosis by sequencing: correction of misdiagnosis from FSHD2 to LGMD2A by whole-exome analysis. *European journal of human genetics : EJHG* 20, 999-1003
24 Blewitt, M.E., *et al.* (2005) An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proceedings of the National Academy of Sciences of the United States of America* 102, 7629-7634
25 Keniry, A., *et al.* (2016) Setdb1-mediated H3K9 methylation is enriched on the inactive X and plays a role in its epigenetic silencing. *Epigenetics & chromatin* 9, 16
26 Gendrel, A.V., *et al.* (2012) Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Developmental cell* 23, 265-279
27 Gendrel, A.V., *et al.* (2013) Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. *Molecular and cellular biology* 33, 3150-3165
28 Mould, A.W., *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, 19
29 Nozawa, R.S., *et al.* (2013) Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nature structural & molecular biology* 20, 566-573
30 Chen, K., *et al.* (2015) Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. *Proceedings of the National Academy of Sciences of the United States of America* 112, E3535-3544
31 Nagano, T., *et al.* (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322, 1717-1720
32 Uhlmann, F. (2016) SMC complexes: from DNA to chromosomes. *Nature reviews. Molecular cell biology* 17, 399-412

33 Brideau, N.J., *et al.* (2015) Independent Mechanisms Target SMCHD1 to Trimethylated Histone H3 Lysine 9-Modified Chromatin and the Inactive X Chromosome. *Molecular and cellular biology* 35, 4053-4068
34 Minajigi, A., *et al.* (2015) Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* 349
35 Chu, C., *et al.* (2015) Systematic discovery of Xist RNA binding proteins. *Cell* 161, 404-416
36 McHugh, C.A., *et al.* (2015) The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521, 232-236
37 Ottaviani, A., *et al.* (2009) The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in facio-scapulo-humeral dystrophy. *PLoS genetics* 5, e1000394
38 Guo, Y., *et al.* (2012) CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice. *Proceedings of the National Academy of Sciences of the United States of America* 109, 21081-21086
39 Dixon-McDougall, T. and Brown, C. (2016) The making of a Barr body: the mosaic of factors that eXIST on the mammalian inactive X chromosome. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 94, 56-70
40 Takagi, N. and Sasaki, M. (1975) Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature* 256, 640-642
41 Lyon, M.F. (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* 190, 372-373
42 Brown, C.J., *et al.* (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349, 38-44
43 Borsani, G., *et al.* (1991) Characterization of a murine gene expressed from the inactive X chromosome. *Nature* 351, 325-329
44 Chaumeil, J., *et al.* (2006) A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes & development* 20, 2223-2237
45 Plath, K., *et al.* (2003) Role of histone H3 lysine 27 methylation in X inactivation. *Science* 300, 131-135
46 Silva, J., *et al.* (2003) Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Developmental cell* 4, 481-495
47 Plath, K., *et al.* (2004) Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome. *The Journal of cell biology* 167, 1025-1035
48 Costanzi, C. and Pehrson, J.R. (1998) Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* 393, 599-601
49 Sado, T., *et al.* (2000) X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Developmental biology* 225, 294-303
50 Hansen, R.S., *et al.* (1996) Role of late replication timing in the silencing of X-linked genes. *Human molecular genetics* 5, 1345-1353
51 Reik, W. and Walter, J. (2001) Genomic imprinting: parental influence on the genome. *Nature reviews. Genetics* 2, 21-32
52 Barlow, D.P. and Bartolomei, M.S. (2014) Genomic imprinting in mammals. *Cold Spring Harbor perspectives in biology* 6

53 Eggermann, T., *et al.* (2016) Erratum to: Imprinting disorders: a group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. *Clinical epigenetics* 8, 27
54 Jelinic, P. and Shaw, P. (2007) Loss of imprinting and cancer. *The Journal of pathology* 211, 261-268
55 Kaneda, M., *et al.* (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 429, 900-903
56 Bourc'his, D., *et al.* (2001) Dnmt3L and the establishment of maternal genomic imprints. *Science* 294, 2536-2539
57 Webster, K.E., *et al.* (2005) Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 102, 4068-4073
58 Howell, C.Y., *et al.* (2001) Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell* 104, 829-838
59 Li, E., *et al.* (1993) Role for DNA methylation in genomic imprinting. *Nature* 366, 362-365
60 Coan, P.M., *et al.* (2005) Imprinted genes in the placenta--a review. *Placenta* 26 Suppl A, S10-20
61 Yu, S., *et al.* (1998) Variable and tissue-specific hormone resistance in heterotrimeric Gs protein alpha-subunit (Gsalpha) knockout mice is due to tissue-specific imprinting of the gsalpha gene. *Proceedings of the National Academy of Sciences of the United States of America* 95, 8715-8720
62 Wroe, S.F., *et al.* (2000) An imprinted transcript, antisense to Nesp, adds complexity to the cluster of imprinted genes at the mouse Gnas locus. *Proceedings of the National Academy of Sciences of the United States of America* 97, 3342-3346
63 Hanel, M.L. and Wevrick, R. (2001) Establishment and maintenance of DNA methylation patterns in mouse Ndn: implications for maintenance of imprinting in target genes of the imprinting center. *Molecular and cellular biology* 21, 2384-2392
64 Wu, Q. and Maniatis, T. (1999) A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* 97, 779-790
65 Tasic, B., *et al.* (2002) Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Molecular cell* 10, 21-33
66 Kaneko, R., *et al.* (2006) Allelic gene regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic expression in single Purkinje cells. *The Journal of biological chemistry* 281, 30551-30560
67 Fokkema, I.F., *et al.* (2011) LOVD v.2.0: the next generation in gene variant databases. *Human mutation* 32, 557-563