De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development

Christopher Gordon, Shifeng Xue, Gökhan Yigit, Hicham Filali, Kelan Chen, Nadine Rosin, Koh-Ichiro Yoshiura, Myriam Oufadem, Tamara J Beck, Ruth Mcgowan, et al.

De novo mutations in *SMCHD1* cause Bosma arhinia microphthalmia syndrome and abrogate nasal development

Christopher T Gordon[1,2,40], Shifeng Xue[3,4,40], Gökhan Yigit[5,40], Hicham Filali[1,2,6,40], Kelan Chen[7,8,40], Nadine Rosin[5], Koh-ichiro Yoshiura[9], Myriam Oufadem[1,2], Tamara J Beck[7], Ruth McGowan[10], Alex C Magee[11], Janine Altmüller[12–14], Camille Dion[15], Holger Thiele[12], Alexandra D Gurzau[7,8], Peter Nürnberg[12,14,16], Dieter Meschede[17], Wolfgang Mühlbauer[18], Nobuhiko Okamoto[19], Vinod Varghese[20], Rachel Irving[20], Sabine Sigaudy[21], Denise Williams[22], S Faisal Ahmed[23], Carine Bonnard[3], Mung Kei Kong[3], Ilham Ratbi[6], Nawfal Fejjal[24], Meriem Fikri[25], Siham Chafai Elalaoui[6,26], Hallvard Reigstad[27], Christine Bole-Feysot[2,28], Patrick Nitschké[2,29], Nicola Ragge[22,30], Nicolas Lévy[15,21], Gökhan Tunçbilek[31], Audrey S M Teo[32], Michael L Cunningham[33], Abdelaziz Sefiani[6,26], Hülya Kayserili[34], James M Murphy[7,8], Chalermpong Chatdokmaiprai[35], Axel M Hillmer[32], Duangrurdee Wattanasirichaigoon[36], Stanislas Lyonnet[1,2,37], Frédérique Magdinier[15], Asif Javed[32,41], Marnie E Blewitt[7,8,41], Jeanne Amiel[1,2,37,41], Bernd Wollnik[5,13,41] & Bruno Reversade[3,4,34,38,39,41]

Bosma arhinia microphthalmia syndrome (BAMS) is an extremely rare and striking condition characterized by complete absence of the nose with or without ocular defects. We report here that missense mutations in the epigenetic regulator *SMCHD1* mapping to the extended ATPase domain of the encoded protein cause BAMS in all 14 cases studied. All mutations were *de novo* where parental DNA was available. Biochemical tests and *in vivo* assays in *Xenopus laevis* embryos suggest that these mutations may behave as gain-of-function alleles. This finding is in contrast to the loss-of-function mutations in *SMCHD1* that have been associated with facioscapulohumeral muscular dystrophy (FSHD) type 2. Our results establish SMCHD1 as a key player in nasal development and provide biochemical insight into its enzymatic function that may be exploited for development of therapeutics for FSHD.

Congenital absence of the nose (arhinia) is a rare and striking condition with fewer than 50 patients reported thus far[1]. Arhinia is variably associated with absent paranasal sinuses, hypertelorism, microphthalmia, colobomas, nasolacrimal duct abnormalities, midface hypoplasia, high-arched palate, absent olfactory bulbs and defects of the reproductive axis in males. In its most severe presentation, consisting of nasal, ocular and reproductive defects, it is referred to as BAMS (MIM 603457)[1,2]. Arhinia is presumed to result from a specific defect in the nasal placodes or surrounding neural crest–derived tissues during embryonic development, but a genetic cause has not been established.

We investigated 14 unrelated individuals with isolated arhinia or a syndromic presentation compatible with BAMS (**Fig. 1a–l**, **Supplementary Fig. 1** and **Supplementary Table 1**). Trio or quartet whole-exome sequencing for cases 1, 2 and 9–12 led to the identification of *de novo* heterozygous missense mutations in the *SMCHD1* gene (encoding structural maintenance of chromosomes flexible hinge domain containing 1; NM_015295.2) in all six cases (**Fig. 1m**, **Table 1** and **Supplementary Table 2**), which were confirmed by Sanger sequencing (**Supplementary Fig. 2**). Singleton whole-exome sequencing for case 13 also identified an *SMCHD1* mutation. We then performed Sanger sequencing of *SMCHD1* in the seven remaining patients with BAMS. Heterozygous missense mutations were identified in all. In total, 11 of the 14 variants were *de novo*, suggesting germline mutations in parental gametes, while in three cases parental DNA was not available (**Table 1** and **Supplementary Fig. 2**). None of the identified mutations have been reported in the Exome Aggregation Consortium (ExAC), Exome Variant Server (EVS) or dbSNP144 database (accessed via the UCSC Genome Browser, November 2016), all mutations affected highly conserved residues (**Supplementary Fig. 3**) and all were predicted to be damaging by PolyPhen-2 (**Table 1**). All 14 mutations are located in exons 3, 8–10, 12 or 13 of *SMCHD1* (48 exons in total); these exons encode the ATPase domain of SMCHD1 and an associated region immediately C terminal to it. Notably, 6 of the 14 patients had mutations that affected three adjacent amino acids—Ala134, Ser135 and Glu136—while p.His348Arg and p.Asp420Val were identified in three and two independent patients each, suggesting possible mutational hotspots (**Fig. 1m**). Mutations in *SMCHD1* in individuals with arhinia have also been identified in an independent study that included six of the cases analyzed here (cases 2, 4–7 and 13; see the accompanying manuscript[3]).

During craniofacial development, the olfactory placode ectoderm thickens and invaginates to form the olfactory epithelium within the nasal cavity, a process that depends on crosstalk between the placodal

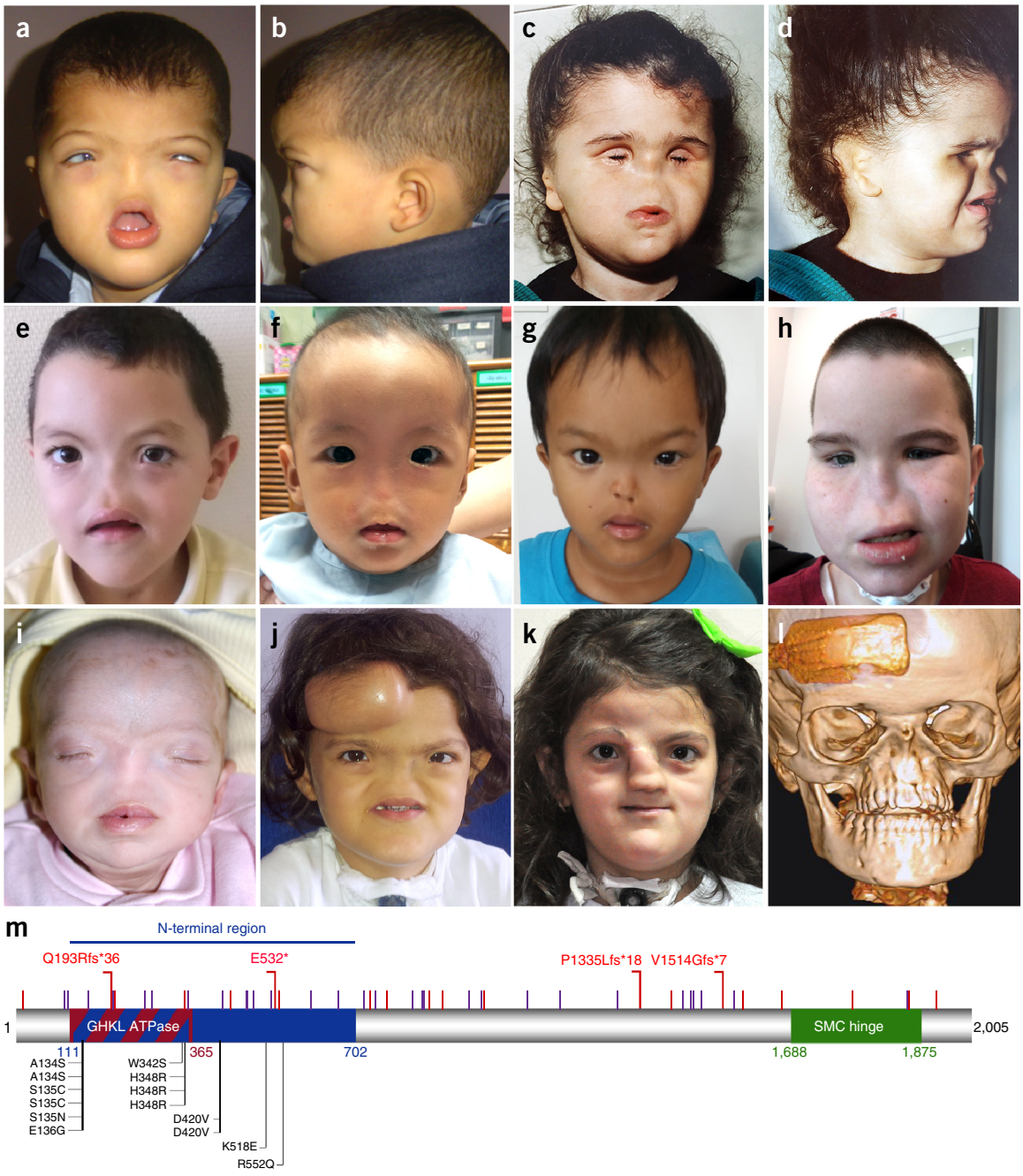

Figure 1 *SMCHD1* is mutated in Bosma arhinia microphthalmia syndrome and isolated arhinia. (**a**,**b**) Case 1. (**c**,**d**) Case 12. (**e**) Case 3. (**f**) Case 9. (**g**) Case 10. (**h**) Case 6. (**i**–**l**) Case 11, with a forehead implant (rectangular box) in preparation for rhinoplasty (**j**), 6 months after the operation (**k**) and in a computed tomography scan of the skull before the operation (**l**). Consent was obtained to publish patient images. (**m**) Positions of BAMS-associated missense variants (black) and heterozygous loss-of-function variants from ExAC (red) in SMCHD1. Short bars represent known missense (purple) and frameshift or nonsense (red) FSHD2-associated variants. See **Supplementary Figure 3** for details on the exact amino acids mutated in FSHD2 in the N-terminal region.

epithelium and the underlying cranial neural crest–derived mesenchyme[4]. For example, ablation of the nasal placode epithelium in chick embryos disrupts the development of adjacent nasal skeletal elements[5]. We observed strong X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) staining in the developing face of mouse embryos expressing *lacZ* from the *Smchd1* locus[6], including in the nasal placodes and optic vesicles at embryonic day (E) 9.5 and in the nasal epithelium at E12.5 (**Supplementary Fig. 4**). Eurexpress *in situ* hybridization data indicate regional expression of *Smchd1* in the nasal cavity in E14.5 mice, while transcriptional profiling of postnatal olfactory epithelium demonstrated that *Smchd1* is specifically expressed in immature olfactory sensory neurons[7]. These data are consistent with roles for *SMCHD1* during early nasal development. Gonadotropin-releasing hormone (GnRH) neurons migrate from the olfactory placode along olfactory axon tracts to the hypothalamus, where they regulate reproductive hormone release from the pituitary gland. Defects in the reproductive axis have occasionally been reported in males with arhinia[1,2,8]; we confirm this finding and

Table 1 ***SMCHD1*** **mutations identified in patients 1–14**

Case	Geographical origin	Nucleotide change[a]	Amino acid change	Predicted functional effect[b]	Mutation origin
1	Morocco	c.407A>G	p.Glu136Gly	0.999	*De novo*
2*	Germany	c.403A>T	p.Ser135Cys	1.000	*De novo*
3	North Africa	c.404G>A	p.Ser135Asn	0.997	*De novo*
4*	Ireland	c.403A>T	p.Ser135Cys	1.000	*De novo*
5*	China	c.1043A>G	p.His348Arg	0.998	*De novo*
6*	Scotland	c.1259A>T	p.Asp420Val	0.877	*De novo*
7*	Japan	c.1655G>A	p.Arg552Gln	1.000	*De novo*
8	Wales	c.1552A>G	p.Lys518Glu	0.976	Unknown (parental DNA unavailable)
9	Thailand	c.1259A>T	p.Asp420Val	0.877	*De novo*
10	Thailand	c.1025G>C	p.Trp342Ser	0.999	*De novo*
11	Turkey	c.400G>T	p.Ala134Ser	0.999	*De novo*
12	Turkey	c.400G>T	p.Ala134Ser	0.999	*De novo*
13*	Norway	c.1043A>G	p.His348Arg	0.998	Unknown (parental DNA unavailable)
14	Ukraine	c.1043A>G	p.His348Arg	0.998	Unknown (parental DNA unavailable)

Individuals also studied by Shaw *et al.*[3] are indicated with an asterisk.

[a]Given with respect to reference sequence NM_015295.2. [b]Based on PolyPhen-2 score using UniProtKB identifier A6NHR9.

also report pubertal delay or anomalies of menarche in all three post-pubertal females in our series (**Supplementary Table 1**). The reproductive axis defects associated with arhinia are likely secondary to a defect in GnRH neuron production in or migration from the olfactory placode.

Smchd1 was identified as a modifier of transgene silencing in mice and was subsequently shown to be involved in X-chromosome inactivation, where it is required for CpG island (CGI) methylation on the inactive X chromosome, CGI-independent silencing of some X-chromosome genes and compaction of the inactive X chromosome[6,9–11]. In addition, Smchd1 functions as an epigenetic repressor at various autosomal loci, with dysregulation of imprinted and monoallelically expressed gene clusters observed in mice mutant for *Smchd1*[10,12,13]. A requirement for *SMCHD1* in the repair of DNA double-strand breaks has also been demonstrated[14,15]. Whereas female mice null for *Smchd1* display midgestation lethality due to derepression of genes on the inactive X chromosome, male mutant mice display perinatal lethality of undescribed causes in certain strains or viability on the FVB/n background[12]. Strikingly, craniofacial abnormalities have not been documented in mice with *Smchd1* loss of function regardless of their sex.

Recently, haploinsufficiency for *SMCHD1* was reported as a cause of FSHD type 2 (FSHD2; MIM 158901)[16]. FSHD has a prevalence of 1 in 20,000, with FSHD type 1 (FSHD1) and FSHD2 accounting for ~95% and ~5% of cases, respectively[17]. FSHD results from pathogenic misexpression of the transcription factor DUX4 (encoded by an array of D4Z4 repeats on chromosome 4q) in skeletal muscle. In FSHD1 (MIM 158900), D4Z4 repeat contraction leads to hypomethylation of the locus and derepression of *DUX4* expression on a permissive haplotype (4qA) that harbors a stabilizing polyadenylation signal for *DUX4* mRNA[17,18]. FSHD2 occurs in individuals harboring loss-of-function *SMCHD1* mutations and the permissive 4qA allele, without the requirement for D4Z4 repeat contraction, although *SMCHD1* mutations can also modify the severity of FSHD1 (refs. 16,19). SMCHD1 is thought to function as a silencer at the 4q locus via binding to the D4Z4 repeats[16]. Over 80 unique, putatively pathogenic *SMCHD1* variants have been reported in patients with FSHD2 (LOVD *SMCHD1* variant database; see URLs). These mutations, which include clear loss-of-function alleles, map throughout the protein and are not clustered in specific domains. Several loss-of-function mutations have also been reported in the ExAC database (**Fig. 1m**), and over 60 deletions affecting *SMCHD1* have been reported in the DECIPHER database (available phenotypic information does not indicate occurrence of arhinia in deletion carriers). We analyzed the methylation status of D4Z4 repeats in peripheral blood leukocytes from patients with BAMS by sodium bisulfite sequencing (**Supplementary Figs. 5–7** and **Supplementary Table 3**). Although a trend for hypomethylation was noted for patients with BAMS relative to controls or their unaffected family members, depending on the site tested within the D4Z4 repeat, some patients with BAMS were normally methylated. A large variability in D4Z4 methylation has also been observed in controls and patients with FSHD[20], and altered methylation is not an absolute indicator of FSHD. Moreover, an important argument against BAMS- and FSHD2-associated mutations acting in the same direction is the absence in the literature (to our knowledge) of reports of BAMS and FSHD co-occurring in the same patient. None of the patients with BAMS reported here have signs of muscular dystrophy, including both the individuals (cases 2 and 12) older than the average age of onset for FSHD2 of 26 years[21], and none of the BAMS-associated missense mutations identified here have been linked to FSHD2.

Proteins of the SMC family are involved in chromatid cohesion, condensation of chromosomes and DNA repair. SMCHD1 is considered to be a non-canonical member of the family, with a C-terminal chromatin-binding hinge domain and an N-terminal GHKL (gyrase, Hsp90, histidine kinase and MutL) ATPase domain[22] (**Fig. 1m**). SMCHD1 may potentially use energy obtained from ATP hydrolysis to manipulate chromatin ultrastructure and interactions. Using small-angle X-ray scattering, the purified recombinant mouse Smchd1 ATPase domain and an adjacent C-terminal region (amino acids 111–702 for the two regions combined; denoted N-terminal region in **Fig. 1m**) have been shown to adopt a structural conformation similar to that of Hsp90 (ref. 22). Consistent with this, the Hsp90 inhibitor radicicol decreases the ATPase activity of Smchd1 (refs. 22,23). Mapping of the SMCHDI amino acids mutated in BAMS and FSHD2 to the homology model of Smchd1 on the basis of the Hsp90 crystal structure indicates that the major cluster of variants in BAMS (amino acids 134–136) is situated immediately N terminal to motif I, which is highly conserved among the GHKL ATPases and participates in coordination of the Mg^{2+}–ATP complex during ATP hydrolysis[24] (**Supplementary Figs. 3** and **8**). The finding of other BAMS-associated mutations that map to the region immediately C terminal to the ATPase domain supports the idea that this extended

region has a function intimately associated with that of the ATPase domain. Given that (i) loss of function of *SMCHD1* causes FSHD2, (ii) FSHD is not known to co-occur with arhinia, (iii) there are no visible craniofacial anomalies in *Smchd1*-null mice, (iv) the mutations in patients with BAMS are clustered in the extended ATPase domain and (v) in contrast to *SMCHD1* depletion[14,15], BAMS-associated mutations do not cause alterations of the DNA damage response or impaired non-homologous end joining (**Supplementary Fig. 9**), we hypothesized that the mutations in BAMS might result in a gain rather than a loss of function for the SMCHD1 protein. To test this hypothesis, we conducted ATPase assays using the purified recombinant N-terminal region of mouse Smchd1 harboring BAMS- or FSHD2-associated alterations. In comparison to wild-type protein, the N-terminal region containing the p.Ala134Ser, p.Ser135Cys or p.Glu136Gly substitution had increased protein hydrolysis of ATP, whereas the FSHD2 substitutions p.Tyr353Cys[16] and p.Thr527Met[19] resulted in strongly and slightly decreased ATPase activity, respectively; activity was unchanged with the BAMS-associated substitution p.Asp420Val (**Fig. 2**). The half-maximal inhibitory concentration (IC_{50}) of radicicol was similar for the ATPase activities of the BAMS-associated mutant and wild-type recombinant proteins (**Supplementary Fig. 10**), suggesting that the mutants retain an intact ATP-binding site. These results suggest that BAMS-associated mutations increase the catalytic activity of SMCHD1.

We next sought to validate these biochemical results *in vivo* using full-length SMCHD1 protein. In *Xenopus*, the expression of *smchd1* begins zygotically and increases steadily after gastrulation (**Fig. 3a**). Endogenous *smchd1* expression is strongly enriched in the head region and the neural tube (**Fig. 3b**). To faithfully recapitulate this expression pattern, the two dorsal–animal blastomeres of eight-cell-stage *Xenopus* embryos were microinjected with 120 pg of capped mRNA encoding either wild-type or mutant human SMCHD1 (**Fig. 3c**). Each set of injected embryos was checked to ensure expression of human SMCHD1 protein (**Fig. 3g** and **Supplementary Fig. 11**). Only tadpoles overexpressing *SMCHD1* mRNA with BAMS-associated mutations showed noticeable craniofacial anomalies (**Fig. 3d–f** and **Supplementary Fig. 12**), including microphthalmia and, in severe cases, anophthalmia (**Fig. 3f**, right). At 4 days post-fertilization (d.p.f.), quantification of eye size showed a marked reduction in the diameter of the eye in tadpoles overexpressing BAMS-associated mutants, whereas tadpoles overexpressing wild-type SMCHD1 or Tyr353Cys SMCHD1, an FSHD2-associated mutant, were indistinguishable from control, uninjected embryos (**Fig. 3h**). One of the BAMS-associated mutants with phenotypic effects in this assay, Asp420Val, showed no change in ATPase activity *in vitro* (**Fig. 2**), suggesting that the *in vivo* assay has higher sensitivity. Whole-mount *in situ* hybridization showed a decrease in the size of the eye and nasal placodes, marked by *rx2a* and *six1* expression, respectively, upon overexpression of a BAMS-associated mutant (**Fig. 3i,j**). In contrast, migration of cranial neural crest, marked by *twist1* expression, was largely unaffected. Development of craniofacial anomalies was dose dependent in injections with wild-type SMCHD1 or a BAMS-associated mutant, whereas overexpression of the FSHD2 mutant Tyr353Cys did not have an effect, regardless of dose (**Fig. 3k** and **Supplementary Fig. 12**). The finding that wild-type SMCHD1, when overexpressed at a sufficiently high concentration, acts in the same phenotypic direction as the BAMS-associated mutants suggests that these mutants may, at least in part, act by augmenting the normal activity of the protein. These *in vivo* results, which partially recapitulate the microphthalmia and facial hypoplasia seen in human patients with severe BAMS, further support the notion that, in contrast to FSHD2 alleles, BAMS-associated missense

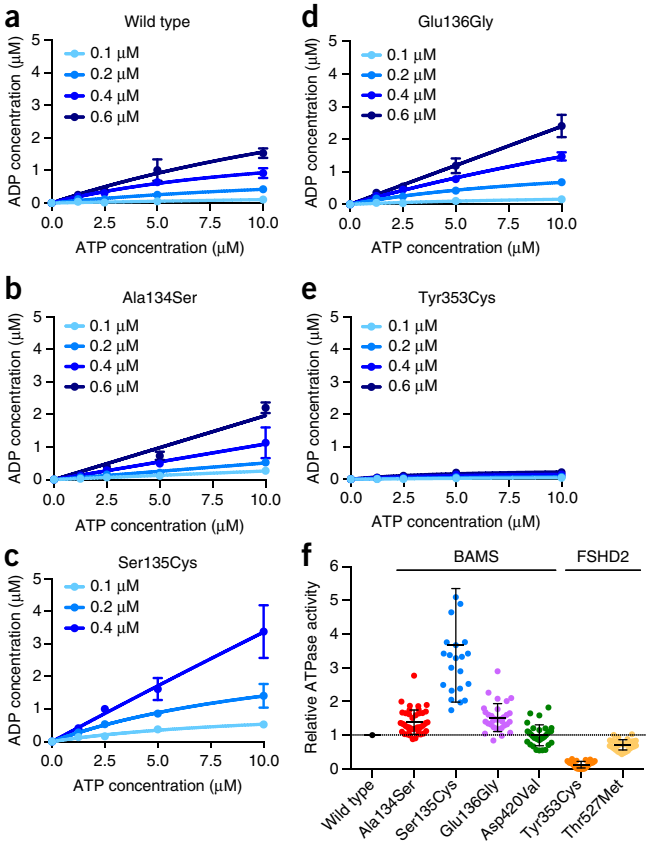

Figure 2 Biochemical assays indicate that BAMS-associated SMCHD1 mutants have increased ATPase activity. (**a**–**e**) ATPase assays performed using recombinant protein encompassing amino acids 111–702 of mouse Smchd1. Results are shown for wild-type (**a**), Ala134Ser (**b**), Ser135Cys (**c**), Glu136Gly (**d**) and Tyr353Cys (**e**) Smchd1. The amount of ADP produced at each protein concentration (0.1, 0.2, 0.4 and 0.6 μM) and ATP concentration (1, 2.5, 5 and 10 μM) was measured as described in the Online Methods. Data are displayed as the means ± s.d. of three technical replicates. Each plot is representative of at least two independent experiments using different batches of protein preparation. (**f**) Relative ATPase activities of the mutant proteins in comparison to wild-type protein. The amount of ADP produced by the mutant proteins was normalized to that produced by wild-type protein at each protein and substrate concentration as in **a**–**e**. Normalized values are plotted as the means ± s.d. from two independent experiments ($n = 44$ for Ala134Ser, $n = 24$ for Ser135Cys, $n = 32$ for Glu136Gly, Asp420Val, Tyr353Cys and Thr527Met). In addition to analyzing normalized fold changes, for each mutant, the mean of the triplicate measures at each protein and ATP concentration was compared to that for wild-type protein using the Wilcoxon matched-pairs signed-rank test; apart from Asp420Val with $P = 0.1776$ (non-significant), all other mutants were different from wild-type protein at $P < 0.0001$ (significant).

mutations may exhibit gain-of-function or neomorphic activity. We have not formally excluded the possibility that BAMS-associated mutants may behave as dominant negatives through heterodimerization with wild-type protein. However, we believe that this is unlikely, given the effects described above for overexpressed wild-type SMCHD1 and the finding that the isolated ATPase domain containing BAMS-associated variants can increase ATPase activity by itself (**Fig. 2**). In addition, a human phenotype associated with a dominant-negative mutation would be expected to present as a more severe disease than that associated with haploinsufficiency of the same gene, with at least some phenotypic overlap, but this is not the case for BAMS and FSHD.

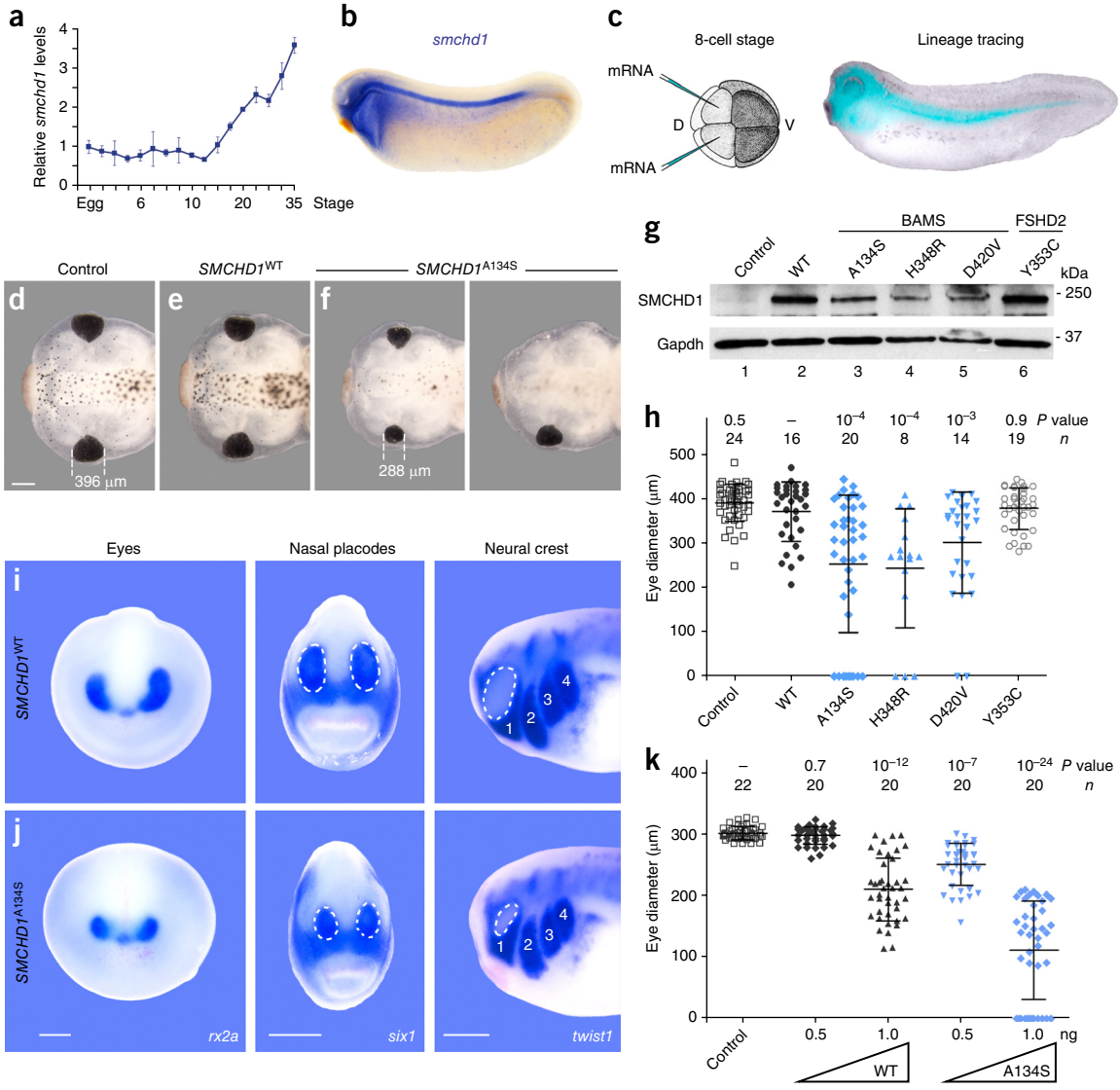

Figure 3 *In vivo* functional assays in *Xenopus* embryos suggest that BAMS-associated mutations behave as gain-of-function alleles. (**a**) Expression of *smchd1* relative to 18S rRNA by qPCR. Data represent means ± s.d. of triplicates. (**b**) In late tailbud stages, *smchd1* expression is restricted to the head region and the neural tube. (**c**) To target the head structures, dorsal–animal blastomeres in eight-cell-stage *Xenopus* embryos were injected with synthesized mRNA (120 pg for all panels except **k**). These cells are fated to give rise to head structures, as shown by dextran lineage tracing. D, dorsal; V, ventral. (**d**–**f**) Representative stage 45 tadpoles that are uninjected (**d**) or injected with *SMCHD1*^{WT} (**e**) or *SMCHD1*^{A134S} (**f**) mRNA. Those injected with *SMCHD1*^{A134S} mRNA display craniofacial anomalies and smaller eyes in comparison to control tadpoles and those injected with *SMCHD1*^{WT} mRNA. Scale bar, 0.3 mm. All images were acquired at the same magnification. (**g**) Immunoblot of stage 12 embryonic extracts from control and injected embryos showing expression of exogenous human SMCHD1. (**h**) Eye diameter is significantly reduced in embryos overexpressing BAMS-associated mutants (blue) relative to siblings overexpressing wild-type SMCHD1 (black) or embryos overexpressing an FSHD2-associated mutant (open circles). (**i**,**j**) *In situ* hybridization for *rx2a*, *six1* and *twist1*, demarcating the eyes, placodes and neural crest, respectively, in embryos injected with *SMCHD1*^{WT} (**i**) or *SMCHD1*^{A134S} (**j**) mRNA. Images are representative of 9 of 10, 7 of 10, and 10 of 10 embryos for each probe. The dotted lines outline nasal placodes (middle) and the eye (right). Numbers label streams of migrating cranial neural crest. Scale bars, 0.2 mm (same magnification for each comparison of **i** to **j**). (**k**) Measurements of eye diameter for *Xenopus* embryos injected with 0.5 or 1 ng of mRNA encoding wild-type SMCHD1 or a BAMS-associated mutant show that SMCHD1 overexpression causes dose-dependent craniofacial anomalies. Biological variation between clutches of tadpoles is seen in the data presented in **h** and **k**. For both **h** and **k**, *n* indicates the number of embryos analyzed, data are shown as means ± s.d. and *P* values were calculated by Kruskal–Wallis test followed by Dunn's post test.

In conclusion, we have identified *de novo* missense mutations restricted to the extended ATPase domain of SMCHD1 as the cause of isolated arhinia and BAMS. It will be of great interest to explore the epistatic relationships between SMCHD1 and known regulators of nasal development, such as the PAX6 protein and fibroblast growth factor (FGF) and bone morphogenetic protein (BMP) signaling[2], as well as to uncover other potential human-specific nasal regulators. Nose shape and size vary greatly across human populations and even more drastically among animal species, with the elephant's trunk being an extreme example. As such, it will be interesting to determine the role of SMCHD1 in controlling nose size from an evolutionary perspective.

Given that loss-of-function mutations in *SMCHD1* are associated with FSHD2, BAMS and FSHD2 represent a rare example of different

functional classes of mutations in the same gene leading to vastly different human disorders, in terms of the affected tissue and age of onset. As FSHD is caused in part by loss of *SMCHD1*, the development of drugs that augment the expression or activity of SMCHD1 in affected muscles as a form of treatment is currently being pursued (for example, by Facio Therapies; see URLs). Our identification of ATPase-activity-augmenting mutations in *SMCHD1* may inform gene therapy approaches or, in combination with future structural studies on the effect of these mutations on the ATPase domain, aid the design of drugs that induce SMCHD1 gain of function, for treatment of FSHD. Importantly, for such an approach, the deleterious consequences of BAMS-associated *SMCHD1* mutations seem to be restricted to a narrow window of human embryonic development.

URLs. Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/; UCSC Genome Browser, http://genome.ucsc.edu/; PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/index.shtml; Facio Therapies, http://www.facio-therapies.com; LOVD *SMCHD1* variant database, http://databases.lovd.nl/shared/variants/SMCHD1/unique; Eurexpress, http://www.eurexpress.org/ee/intro.html; Phyre2, http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index; PBIL server, https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html; NHLBI GO Exome Sequencing Project Exome Variant Server (EVS), http://evs.gs.washington.edu/EVS/; Exome Aggregation Consortium (ExAC) Browser, http://exac.broadinstitute.org/; DECIPHER database, https://decipher.sanger.ac.uk; unabridged *Xenopus* protocols, https://sites.google.com/a/reversade.com/www/protocols/.

ACKNOWLEDGMENTS

We would like to thank all family members and their relatives for their participation and kind contribution to this study. N. Akarsu was instrumental for recruiting patient 11. Support from the Jean Renny Endowed Chair for Craniofacial Research (M.L.C.) is acknowledged. C.D. is the recipient of a fellowship from the French Ministry of Education and Research. H.F. was supported by a postdoctoral grant from INSERM. B.R. is a fellow of the Branco Weiss Foundation, an A*STAR Investigator, an EMBO Young Investigator and a recipient of the inaugural AAA Fellowship in Amsterdam. This work was supported by funding from the Agence Nationale de la Recherche (ANR-10-IAHU-01, CranioRespiro), the Cancer Council Victoria (fellowship to K.C.), the National Health and Medical Research Council (NHMRC) of Australia to M.E.B. and J.M.M. (1098290 and fellowships 1110206 and 1105754), the Scientific and Technological Research Council of Turkey (TUBITAK) to H.K. (grant 112S398, E-RARE network CRANIRARE-2), the Association Française contre les Myopathies (AFM) to F.M., Victorian State Government Operational Infrastructure Support, an NHMRC IRIISS grant (9000220), the German Federal Ministry of Education and Research (BMBF) to B.W. (grant 01GM1211A, E-RARE network CRANIRARE-2), the German Research Foundation (SFB1002, project D02) to B.W., MACS, VICTA and Baillie Gifford grant support to N. Ragge, Mahidol University and Research Career Development Awards from the Faculty of Medicine Ramathibodi Hospital to D. Wattanasirichaigoon, an A*STAR JCO Career Development grant to A.J., an A*STAR BMRC Young Investigator Grant to S.X. and a Strategic Positioning Fund on Genetic Orphan Diseases from the Biomedical Research Council, A*STAR, Singapore, to B.R.

AUTHOR CONTRIBUTIONS

Genetic studies were performed by H.F., C.T.G., M.O., K.Y., C.B.-F., P. Nitschké, P. Nürnberg, C.B., A.S.M.T., A.J., H.T., J. Altmüller and G.Y. Genetic studies were supervised by C.T.G., J. Amiel, B.W., A.M.H. and B.R. The team consisting of B.R., A.J., S.X., H.K. and D. Wattanasirichaigoon independently identified SMCHD1 mutations in patients 9–12 and 14. H.K., D. Wattanasirichaigoon, C.C., G.T., N. Ragge, R.M., A.C.M., N.O., V.V., R.I., S.S., D. Williams, S.F.A., I.R., N.F., M.F., S.C.E., H.R., A.S., S.L., D.M., W.M. and M.L.C. diagnosed patients. K.C., A.D.G., J.M.M. and M.E.B. performed and analyzed the results of ATPase assays. S.X., M.K.K. and B.R. performed and analyzed the results of functional experiments in *Xenopus*. N. Rosin and G.Y. performed DNA damage repair assays, supervised by B.W. C.T.G. and T.J.B. performed analysis of *Smchd1*$^{gt/+}$ embryos. C.D., N.L. and F.M. performed and analyzed the results of methylation studies. The manuscript was written by C.T.G. with contributions from S.X., H.F., J. Amiel and B.R. All authors read and approved its content.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Brasseur, B., Martin, C.M., Cayci, Z., Burmeister, L. & Schimmenti, L.A. Bosma arhinia microphthalmia syndrome: clinical report and review of the literature. *Am. J. Med. Genet. A.* **170A**, 1302–1307 (2016).
2. Graham, J.M. Jr. & Lee, J. Bosma arhinia microphthalmia syndrome. *Am. J. Med. Genet. A.* **140**, 189–193 (2006).
3. Shaw, N.D. *et al. SMCHD1* mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. *Nat. Genet.* http://dx.doi.org/10.1038/ng.3743 (2017).
4. Forni, P.E. & Wray, S. GnRH, anosmia and hypogonadotropic hypogonadism—where are we? *Front. Neuroendocrinol.* **36**, 165–177 (2015).
5. Szabo-Rogers, H.L. *et al.* Novel skeletogenic patterning roles for the olfactory pit. *Development* **136**, 219–229 (2009).
6. Blewitt, M.E. *et al.* SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* **40**, 663–669 (2008).
7. Nickell, M.D., Breheny, P., Stromberg, A.J. & McClintock, T.S. Genomics of mature and immature olfactory sensory neurons. *J. Comp. Neurol.* **520**, 2608–2629 (2012).
8. Tryggestad, J.B., Li, S. & Chernausek, S.D. Hypogonadotropic hypogonadism presenting with arhinia: a case report. *J. Med. Case Rep.* **7**, 52 (2013).
9. Gendrel, A.-V. *et al.* Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev. Cell* **23**, 265–279 (2012).
10. Gendrel, A.-V. *et al.* Epigenetic functions of Smchd1 repress gene clusters on the inactive X chromosome and on autosomes. *Mol. Cell. Biol.* **33**, 3150–3165 (2013).
11. Nozawa, R.-S. *et al.* Human inactive X chromosome is compacted through a PRC2-independent SMCHD1–HBiX1 pathway. *Nat. Struct. Mol. Biol.* **20**, 566–573 (2013).
12. Mould, A.W. *et al.* Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. *Epigenetics Chromatin* **6**, 19 (2013).
13. Chen, K. *et al.* Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. *Proc. Natl. Acad. Sci. USA* **112**, E3535–E3544 (2015).
14. Coker, H. & Brockdorff, N. SMCHD1 accumulates at DNA damage sites and facilitates the repair of DNA double-strand breaks. *J. Cell Sci.* **127**, 1869–1874 (2014).
15. Tang, M. *et al.* Structural maintenance of chromosomes flexible hinge domain containing 1 (SMCHD1) promotes non-homologous end joining and inhibits homologous recombination repair upon DNA damage. *J. Biol. Chem.* **289**, 34024–34032 (2014).
16. Lemmers, R.J.L.F. *et al.* Digenic inheritance of an *SMCHD1* mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* **44**, 1370–1374 (2012).
17. Hewitt, J.E. Loss of epigenetic silencing of the *DUX4* transcription factor gene in facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.* **24**R1, R17–R23 (2015).
18. Lemmers, R.J.L.F. *et al.* A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* **329**, 1650–1653 (2010).
19. Sacconi, S. *et al.* The FSHD2 gene *SMCHD1* is a modifier of disease severity in families affected by FSHD1. *Am. J. Hum. Genet.* **93**, 744–751 (2013).
20. Lemmers, R.J.L.F. *et al.* Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. *Hum. Mol. Genet.* **24**, 659–669 (2015).
21. de Greef, J.C. *et al.* Clinical features of facioscapulohumeral muscular dystrophy 2. *Neurology* **75**, 1548–1554 (2010).
22. Chen, K. *et al.* The epigenetic regulator Smchd1 contains a functional GHKL-type ATPase domain. *Biochem. J.* **473**, 1733–1744 (2016).
23. Brideau, N.J. *et al.* Independent mechanisms target SMCHD1 to trimethylated histone H3 lysine 9–modified chromatin and the inactive X chromosome. *Mol. Cell. Biol.* **35**, 4053–4068 (2015).
24. Dutta, R. & Inouye, M. GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem. Sci.* **25**, 24–28 (2000).

[1]Laboratory of Embryology and Genetics of Congenital Malformations, INSERM UMR 1163, Institut Imagine, Paris, France. [2]Paris Descartes, Sorbonne Paris Cité Université, Institut Imagine, Paris, France. [3]Human Genetics and Embryology Laboratory, Institute of Medical Biology, A*STAR, Singapore. [4]Institute of Molecular and Cell Biology, A*STAR, Singapore. [5]Institute of Human Genetics, University Medical Center Göttingen, Göttingen, Germany. [6]Centre de Génomique Humaine, Faculté de Médecine et de Pharmacie, Mohammed V University, Rabat, Morocco. [7]Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria, Australia. [8]Department of Medical Biology, University of Melbourne, Melbourne, Victoria, Australia. [9]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan. [10]West of Scotland Regional Genetics Service, Queen Elizabeth University Hospital, Glasgow, UK. [11]Northern Ireland Regional Genetics Service, Belfast City Hospital, Belfast, UK. [12]Cologne Center for Genomics (CCG), University of Cologne, Cologne, Germany. [13]Institute of Human Genetics, University of Cologne, Cologne, Germany. [14]Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany. [15]Aix Marseille Université, INSERM, Génétique Médicale et Génomique Fonctionnelle (GMGF), UMRS 910, Marseille, France. [16]Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany. [17]Praxis für Humangenetik, Cologne, Germany. [18]Plastische und Ästhetische Chirurgie, ATOS Klinik München, Munich, Germany. [19]Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Izumi, Osaka, Japan. [20]Institute of Medical Genetics, University Hospital of Wales, Cardiff, UK. [21]Département de Génétique Médicale, Hôpital Timone Enfant, Assistance Publique–Hôpitaux de Marseille, Marseille, France. [22]West Midlands Regional Genetics Service, Birmingham Women's NHS Foundation Trust, Birmingham, UK. [23]Developmental Endocrinology Research Group, University of Glasgow, RHC, Glasgow, UK. [24]Service de Chirurgie Plastique Pédiatrique, Hôpital d'Enfants, CHU Ibn Sina, Mohammed V University, Rabat, Morocco. [25]Service de Neuroradiologie, Hôpital des Spécialités, CHU Ibn Sina, Mohammed V University, Rabat, Morocco. [26]Département de Génétique Médical, Institut National d'Hygiène, Rabat, Morocco. [27]Neonatal Intensive Care Unit, Children's Department, Haukeland University Hospital, Bergen, Norway. [28]Genomic Platform, INSERM UMR 1163, Institut Imagine, Paris, France. [29]Bioinformatic Platform, INSERM UMR 1163, Institut Imagine, Paris, France. [30]Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, UK. [31]Department of Plastic, Reconstructive and Aesthetic Surgery, Hacettepe University Faculty of Medicine, Ankara, Turkey. [32]Cancer Therapeutics and Stratified Oncology, Genome Institute of Singapore, A*STAR, Singapore. [33]University of Washington Department of Pediatrics, Division of Craniofacial Medicine and Seattle Children's Hospital Craniofacial Center, Seattle, Washington, USA. [34]Department of Medical Genetics, Koç University, School of Medicine (KUSoM), Istanbul, Turkey. [35]Plastic and Maxillofacial Surgery, Department of Surgery, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand. [36]Division of Medical Genetics, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand. [37]Département de Génétique, Hôpital Necker–Enfants Malades, Assistance Publique–Hôpitaux de Paris, Paris, France. [38]Department of Paediatrics, School of Medicine, National University of Singapore, Singapore. [39]Amsterdam Reproduction and Development, Academic Medical Centre and VU University Medical Center, Amsterdam, the Netherlands. [40]These authors contributed equally to this work. [41]These authors jointly directed this work. Correspondence should be addressed to J. Amiel (jeanne.amiel@inserm.fr), B.W. (bernd.wollnik@med.uni-goettingen.de) or B.R. (bruno@reversade.com).

ONLINE METHODS

Subjects. In all cases, informed consent was obtained from the families for genetic testing. For patients in **Figure 1**, consent to publish photographs was obtained.

Whole-exome sequencing. Whole-exome sequencing was conducted in accordance with approved institutional ethical guidelines (Comité de Protection des Personnes Ile-de-France II; Ethics Committee of the University Hospital Cologne, Germany; National University of Singapore Institutional Review Board).

For trio whole-exome sequencing of case 1, Agilent SureSelect libraries were prepared using 3 µg of genomic DNA from each individual and sheared with a Covaris S2 Ultrasonicator. Exome capture was performed with 51Mb SureSelect Human All Exon kit v5 (Agilent Technologies). Sequencing was carried out on a pool of barcoded exome libraries using a HiSeq 2500 instrument (Illumina), generating 100 + 100 bp paired-end reads. After demultiplexing, paired-end sequences were mapped to the reference human genome (GRCh37/hg19 assembly, NCBI) using Burrows–Wheeler aligner (BWA). The mean depth of coverage obtained for the three samples from case 1 was 123-fold, 149-fold and 150-fold, and 98% of the exome was covered by at least 15-fold. Downstream processing was performed using the Genome Analysis Toolkit (GATK)[25], SAMtools[26] and Picard. Variant calls were made with the GATK UnifiedGenotyper. All calls with read coverage ≤2-fold or a Phred-scaled SNP quality score of ≤20-fold were removed from consideration. Variant annotation was based on Ensembl release 71 (ref. 27). Variants were filtered against publicly available SNPs plus variant data from more than 7,000 in-house exomes (Institut Imagine).

For trio whole-exome sequencing of case 2, exonic and adjacent intronic sequences were enriched from genomic DNA using the NimbleGen SeqCap EZ Human Exome Library v2.0 enrichment kit and probes were run on an Illumina HiSeq 2000 sequencer at the Cologne Center for Genomics (CCG). Data analysis and filtering of mapped target sequences were performed with 'Varbank' exome and genome analysis pipeline v.2.1 (CCG), and data were filtered for high-quality (coverage of more than 6 reads, minimum quality score of 10), rare (MAF < 0.5%) autosomal recessive and *de novo* variants.

For trios of cases 9 and 11 and quartets of cases 10 and 12, whole-exome sequencing was performed at the Genome Institute of Singapore. Barcoded libraries were prepared for each individual by shearing 1 µg of genomic DNA, followed by end repair, A-tailing, adaptor ligation and PCR enrichment, and libraries were then pooled and hybridized with NimbleGen SeqCap EZ Human Exome Library v3.0 probes. Captured DNA targets were purified and PCR amplified, and were then sequenced on Illumina HiSeq 2500 (cases 9 and 11) or HiSeq 4000 (cases 10 and 12) sequencers. Variant calling was performed according to the recommended best practices of GATK (v3.4.46). Reads were mapped to GRCh37/hg19 using BWA, and the aligned files were preprocessed by Picard and GATK[25,28,29]. All samples were sequenced with mean coverage of 75× or higher. Variants were called using GATK HaplotypeCaller along with in-house exomes sequenced with the same chemistry. Variants were recalibrated, annotated and filtered against in-house data plus common publicly available databases. Each family was analyzed independently using Phen-Gen[30] for *de novo* and recessive disease inheritance patterns. Variants with alternate allele frequency of ≤10 reads or total coverage of ≤20 reads were not considered.

For whole-exome sequencing of case 13, a library was prepared using the SureSelect XT Human All Exon v5 kit (Agilent Technologies) according to the manufacturer's instructions, followed by sequencing on a HiSeq 2500 (Illumina) sequencer. Raw data files were converted to fastq files with bcl2fastq software package version 1.8.4 (Illumina). fastq files were mapped by Novoalign version 3 (Novocraft) to the hg19 human reference genome sequence. In this step, single-nucleotide variant (SNV) information in dbSNP[31] build 138 was used for base quality score recalibration. Marking of PCR duplicates and position-wise sorting were performed with Novosort version 3 (Novocraft). Calling of SNVs and small indels was performed using GATK[25,28,29] version 3.4-46. A GATK workflow[32] was used in which local realignment and variant calling were performed by IndelRealigner and HaplotypeCaller, respectively. Low-quality calls for SNVs and small indels were removed it they met the following criteria: QD < 2.0, MQ < 40.0, FS > 60.0, MQRankSum < −12.5 or ReadPosRankSum < −8.0 for SNVs; QD < 2.0, ReadPosRankSum < −20.0 or FS > 200.0 for small indels. SNVs and small indels were annotated with the ANNOVAR software package[33] using the following data sets and programs: gene information from GENCODE[34] (version 19); allele frequencies from the 1000 Genome Project[35] (version August 2015), ExAC (version 0.3; see URLs), EVS (release ESP6500SI-V2; see URLs) and an in-house database; and predictions of protein damage by PolyPhen-2 (ref. 36) and SIFT[37] via dbNSFP[38,39] (version 3.0).

DNA methylation analysis. DNA methylation was analyzed at single-base resolution after sodium bisulfite modification, PCR amplification, cloning and Sanger sequencing. Briefly, 2 µg of genomic DNA was denatured for 30 min at 37 °C in 0.4 N NaOH and incubated overnight in a solution of 3 M sodium bisulfite (pH 5) and 10 mM hydroquinone using a previously described protocol[40]. Converted DNA was then purified using the Wizard DNA CleanUp kit (Promega) following the manufacturer's recommendations and precipitated by ethanol precipitation for 5 h at −20 °C. After centrifugation, DNA pellets were resuspended in 20 µl of water and stored at −20 °C until use. Converted DNA was then amplified using primer sets (**Supplementary Table 4**) designed with MethPrimer software[41], avoiding the occurrence of CpGs in primer sequences to allow amplification of methylated and unmethylated DNA with the same efficiency. Amplification was carried out using High-Fidelity Taq polymerase (Roche) according to the manufacturer's instructions. After initial denaturation at 94 °C for 2 min, amplification was carried out at 94 °C for 20 s, 54 °C for 30 s, and 72 °C for 1 min for 10 cycles and then at 94 °C for 20 s, 54 °C for 30 s, and 72 °C for 4.5 min for the first cycle, with the extension step extended by 30 s for each subsequent cycle, for 25 cycles. At the end of the program, a final extension step at 72 °C for 7 min was performed. PCR products were purified using the Wizard SV Gel and PCR Purification System (Promega), resuspended in 50 µl of water and cloned using the pGEM-T Easy Vector cloning kit (Promega). Colonies were grown overnight at 37 °C with ampicillin selection, and randomly selected colonies were directly PCR amplified using T7 or SP6 primer. For each sample and region, at least ten randomly cloned PCR products were sequenced according to Sanger's method by Eurofins MWG Operon with either SP6 or T7 primer. Sequences were analyzed using BiQ Analyser software[42], and the average methylation score was calculated as the number of methylated CpGs out of the total number of CpGs in the reference sequence.

Statistics and subjects. The average methylation levels of the groups of samples (patients with FSHD2 carrying an *SMCHD1* mutation, control individuals, and patients with BAMS and their relatives) were compared using the Kruskal–Wallis non-parametric multiple-comparisons test followed by a Dunn's comparison and Bonferroni correction, with $\alpha = 0.05$. Control individuals ($n = 8$) were healthy donors whom have been previously reported[43]. The patients with FSHD2 carrying an *SMCHD1* mutation have previously ben reported[43,44] and comprise $n = 8$ with methylation data for the DR1 region and $n = 15$ each with methylation data for the 5′ and Mid regions, while for the DR1 region 21 additional patients for whom sodium bisulfite sequencing data exist in the LOVD *SMCHD1* variant database (see URLs) were included.

Smchd1–Hsp90 structure modeling and multiple-sequence alignment. A homology model of the N-terminal region of Smchd1 was generated using the online server Phyre2 (ref. 45). The protein sequence corresponding to amino acids 111–702 of mouse Smchd1 was submitted as the input sequence, and intensive modeling mode was selected. The second highest scoring model with the most sequence alignment coverage based on the crystal structure of yeast Hsp90 (Protein Data Bank (PDB) 2CG9) was elected for further evaluation. The model was visualized in PyMOL. The multiple-sequence alignment was generated using CLUSTAL W (ref. 46; via the PBIL server) and ESPript 3.0 (ref. 47).

ATPase assays. Cloning, expression and purification of recombinant mouse Smchd1 protein was performed as previously described[22]; the sequences of the primers used for cloning and mutagenesis are provided in **Supplementary Table 5**. The purity of the protein preparations was assessed by migration of samples on 4–20% Tris-glycine reducing SDS–PAGE gels followed by staining with SimplyBlue SafeStain (Thermo Fisher Scientific) (**Supplementary Fig. 13**). ATPase assays were performed with the Transcreener ADP2 Fluorescence Polarization assay kit (BellBrook Labs) as previously described[22]. Briefly, 10-µl reactions in triplicate were set up in 384-well (low-volume, black) plates, containing 7 µl of reaction buffer (50 mM HEPES (pH 7.5), 4 mM $MgCl_2$ and 2 mM EGTA), 1 µl of recombinant $Smchd1_{111–702}$ protein at concentrations

ranging from 0.1 to 0.6 µM or buffer control, 1 µl of radicicol or solvent control, and 1 µl of 10 µM ATP substrate or nuclease-free water control. The Hsp90 inhibitor radicicol (Sigma-Aldrich) was dissolved in 70% ethanol and further diluted to a final concentration ranging from 0.1 nM to 10 µM. A 12-point 10 µM ADP/ATP standard curve was set up in parallel. Reactions were incubated at room temperature for 1 h in the dark before addition of 10 µl of detection mix (1× Stop and Detection Buffer B, 23.6 µg/ml antibody to ADP2) and incubation for a further hour. Fluorescence polarization readings were performed with an Envision plate reader (PerkinElmer Life Sciences) following the manufacturer's instructions. The amount of ADP present in each reaction was estimated using the standard curve, following the manufacturer's instructions.

Mouse embryo dissection and X-gal staining. Mice were housed and mouse work was performed under approval from the Walter and Eliza Hall Institute of Medical Research Animal Ethics Committee (AEC 2014.026). Embryos were obtained by mating C57BL/6 *Smchd1*$^{gt/+}$ congenic strain sires with C57BL/6 dams, with embryo ages ranging from E8.5 to E12.5 (ref. 6). All embryos analyzed were female. No randomization or blinding was used during the experimental procedure. Embryos were briefly fixed in 2% paraformaldehyde/0.2% glutaraldehyde and stained in 1 mg/ml X-gal for several hours. Cryosections were cut at 12 µm.

***Xenopus* embryological assays.** *Xenopus* were used according to guidelines approved by the Singapore National Advisory Committee on Laboratory Animal Research. Protocols for fertilization, injection and whole-mount *in situ* hybridization are available at the protocol website for the Reversade laboratory (see URLs). Human *SMCHD1* (Origene) was cloned into the pCS2+ expression vector, the vector was linearized with NotI and the insert was transcribed with the mMESSAGE mMachine SP6 transcription kit (Thermo Fisher). Transcribed mRNA was column purified, and its concentration was measured using a Nanodrop instrument. The mRNA contains a poly(A) signal that allows for polyadenylation *in vivo*. To specifically target the cells destined to contribute to anterior head tissue, the two dorsal–animal blastomeres were injected at the eight-cell stage with the synthesized mRNA. Embryos were allowed to develop at room temperature until they reached stages 45–46 (4 d.p.f.) and were fixed. Eye diameter was measured using a Leica stereomicroscope with a DFC 7000T digital camera. No statistical method was used to predetermine sample size. No randomization or blinding was used. Embryos that died before gastrulation were excluded. Injections were performed on multiple clutches to reduce clutch-specific bias. The mRNAs injected for **Figure 3k** did not contain a poly(A) signal and were polyadenylated *in vitro*, hence requiring higher RNA concentrations to produce a phenotype (in other panels in **Figure 3** and in **Supplementary Figure 12**, the mRNAs contained a poly(A) signal allowing polyadenylation *in vivo*). Embryonic extracts were prepared by lysing stage 12 embryos in CelLytic Express (Sigma) on ice, followed by centrifugation to remove yolk proteins. Extracts were analyzed by immunoblotting with antibodies to SMCHD1 (Atlas, HPA039441; 1:500 dilution) and GAPDH (clone 0411, Santa Cruz Biotechnology; 1:2,000 dilution). cDNA was generated from RNA extracted from *Xenopus* embryos of various stages using iScript reverse transcriptase (Bio-Rad). qPCR was performed using the primers listed in **Supplementary Table 6**. The *in situ* hybridization probe for *smchd1* was amplified from stage 20 cDNA using the primers listed in **Supplementary Table 6** and cloned into the pGEM-T vector. The vector was linearized, and the insert was transcribed using digoxigenin RNA labeling mix (Roche) according to the manufacturer's guidelines.

DNA damage response assays. *Cell lines and cell culture. XRCC4*-deficient cells[48] and primary fibroblast cell lines established from cases 1 and 2 were cultured in DMEM (Gibco) supplemented with 10% FCS (Gibco) and antibiotics. Testing for mycoplasma contamination was negative. To assess H2AX activation, cells were either irradiated with 100 J/m^2 UV-C or treated with 50 µM etoposide (Sigma-Aldrich) for 1 h. Drugs were then washed out, fresh medium was added, and cells were incubated for 6 h and then subjected to immunoblot analysis.

Protein isolation and analysis. Cells were solubilized using ice-cold RIPA buffer (10 mM Tris (pH 8.0), 150 mM NaCl, 1 mM EDTA, 10 mM NaF, 1 mM Na_3VO_4, 10 µM Na_2MoO_4, 1% NP-40, 0.25% SDS and protease inhibitors P 2714 (Sigma-Aldrich)). The total protein concentration of extracts was determined using the BCA Protein Assay kit (Thermo Fisher Scientific). Ten micrograms of total cell lysate were separated by 4–12% SDS–PAGE (Invitrogen) and blotted onto nitrocellulose membrane (GE Healthcare). Protein detection was performed using antibody specific for phosphorylation of H2AX at Ser139 (γH2AX) (clone 20E3, Cell Signaling Technology; 1:1,000 dilution). Antibody to β-actin was purchased from Sigma-Aldrich (clone AC-74; 1:10,000 dilution). Secondary antibodies conjugated to peroxidase (Santa Cruz Biotechnology) were used, and blots were developed using an enhanced chemiluminescence system (ECL Plus, GE Healthcare), followed by detection of autoradiography by film.

Microhomology-mediated end joining assay. Microhomology-mediated end joining (MMEJ) assays using linearized pDVG94 plasmid were performed as previously described[49]. In brief, cells were transfected with 2 µg of pDVG94 linearized with EcoRV (Thermo Fisher Scientific) and AfeI (New England BioLabs), and extrachromosomal DNA was isolated 48 h after transfection. PCR analysis was performed, and PCR products were digested using BstXI, separated by gel electrophoresis and visualized by ethidium bromide staining.

Data availability. Whole-exome sequencing data have been deposited in the European Genome-phenome Archive (EGA) under accessions EGAS00001002193.

25. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
26. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
27. Flicek, P. *et al.* Ensembl 2013. *Nucleic Acids Res.* **41**, D48–D55 (2013).
28. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
29. Van der Auwera, G.A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013).
30. Javed, A., Agrawal, S. & Ng, P.C. Phen-Gen: combining phenotype and genotype to analyze rare disorders. *Nat. Methods* **11**, 935–937 (2014).
31. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
32. Mishima, H., Sasaki, K., Tanaka, M., Tatebe, O. & Yoshiura, K. Agile parallel bioinformatics workflow management using Pwrake. *BMC Res. Notes* **4**, 331 (2011).
33. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
34. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
35. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
36. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
37. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
38. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum. Mutat.* **32**, 894–899 (2011).
39. Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum. Mutat.* **37**, 235–241 (2016).
40. Magdinier, F. *et al.* Regional methylation of the 5′ end CpG island of *BRCA1* is associated with reduced gene expression in human somatic cells. *FASEB J.* **14**, 1585–1594 (2000).
41. Li, L.-C. & Dahiya, R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* **18**, 1427–1431 (2002).
42. Bock, C. *et al.* BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* **21**, 4067–4068 (2005).
43. Gaillard, M.-C. *et al.* Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. *Neurology* **83**, 733–742 (2014).
44. Gaillard, M.-C. *et al.* Segregation between *SMCHD1* mutation, D4Z4 hypomethylation and facio-scapulo-humeral dystrophy: a case report. *BMC Med. Genet.* **17**, 66 (2016).
45. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
46. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
47. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
48. Rosin, N. *et al.* Mutations in *XRCC4* cause primary microcephaly, short stature and increased genomic instability. *Hum. Mol. Genet.* **24**, 3708–3717 (2015).
49. Verkaik, N.S. *et al.* Different types of V(D)J recombination and end-joining defects in DNA double-strand break repair mutant mammalian cells. *Eur. J. Immunol.* **32**, 701–709 (2002).