Author's accepted manuscript	Gordon CT, Xue S, Yigit G, Filali H, Chen K, Rosin N, Yoshiura KI, Oufadem M, Beck TJ, McGowan R, Magee AC, Altmuller J, Dion C, Thiele H, Gurzau AD, Nurnberg P, Meschede D, Muhlbauer W, Okamoto N, Varghese V, Irving R, Sigaudy S, Williams D, Ahmed SF, Bonnard C, Kong MK, Ratbi I, Fejjal N, Fikri M, Elalaoui SC, Reigstad H, Bole-Feysot C, Nitschke P, Ragge N, Levy N, Tuncbilek G, Teo AS, Cunningham ML, Sefiani A, Kayserili H, Murphy JM, Chatdokmaiprai C, Hillmer AM, Wattanasirichaigoon D, Lyonnet S, Magdinier F, Javed A, Blewitt ME, Amiel J, Wollnik B, Reversade B. *De novo* mutations in *SMCHD1* cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. *Nature Genetics.* 2017 49(2):249-255
Final published version:	doi: 10.1038/ng.3765 https://doi-org/10.1038/ng.3765
Copyright:	

De novo mutations in *SMCHD1* abrogate nasal development

Christopher T. Gordon[1,2]*, Shifeng Xue[3,4]*, Gökhan Yigit[5]*, Hicham Filali[1,2,6]*, Kelan Chen[7,8]*, Nadine Rosin[5], Koh-ichiro Yoshiura[9], Myriam Oufadem[1,2], Tamara J. Beck[7], Ruth McGowan[10], Alex C. Magee[11], Janine Altmüller[12,13,14], 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#], Marnie E. Blewitt[7,8#], Jeanne Amiel[1,2,37#], Bernd Wollnik[5,13#], Bruno Reversade[3,4,34,38,39#]

[1]Laboratory of embryology and genetics of congenital malformations, Institut National de la Santé et de la Recherche Médicale (INSERM) UMR 1163, Institut *Imagine*, Paris, France.

[2]Paris Descartes-Sorbonne Paris Cité University, 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]The Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia.

[8]The University of Melbourne, Melbourne, Australia.

[9]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, 1-12-4, Sakamoto, Nagasaki, Japan.

[10]West of Scotland Regional Genetics Service, Laboratory Medicine Building, 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), UMR S_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, 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]Oxford Brookes University, Oxford, UK.

[31]Hacettepe University Faculty of Medicine, Department of Plastic, Reconstructive and Aesthetic Surgery, 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, 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 & Development, Academic Medical Centre & VU University Medical Center, Amsterdam, the Netherlands.

*co-first authors

[#]co-last authors

Correspondence should be addressed to:

J.A. (jeanne.amiel@inserm.fr), B.W (bernd.wollnik@med.uni-goettingen.de) and B.R. (bruno@reversade.com).

Keywords:

Nasal, nose, arhinia, SMCHD1, *Xenopus*, HSP90, gain-of-function, Bosma Syndrome, facioscapulohumeral muscular dystrophy, FSHD2, chromatin, *de novo*, craniofacial, development, organogenesis, epigenetics.

Introductory paragraph

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 extended ATPase domain of the epigenetic regulator SMCHD1 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* embryos suggest that these mutations may behave as gain-of-function alleles. This is in contrast to 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.

Main text

Congenital absence of the nose (arhinia) is a rare and striking condition with less than 50 patients reported to date[1]. Arhinia is variably associated with absent paranasal sinuses, hypertelorism, microphthalmia, colobomas, nasolacrimal duct abnormalities, mid-face 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 Bosma arhinia microphthalmia syndrome (BAMS) (OMIM 603457)[1,2]. Arhinia is presumed to result from a specific defect of 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 (WES) for cases 1, 2 and 9-12 led to the identification of *de novo* heterozygous missense mutations in the Structural Maintenance of Chromosomes Flexible Hinge Domain Containing 1 (*SMCHD1*; NCBI Reference Sequence: NM_015295.2) gene in all six cases (**Fig.**

1m, Table 1 and **Supplementary Table 2**), which were confirmed by Sanger sequencing (**Supplementary Fig. 2**). Singleton WES for case 13 also identified a *SMCHD1* mutation. We then performed Sanger sequencing of *SMCHD1* in the remaining seven BAMS patients. Heterozygous missense mutations were identified in all. In total, 11 out of 14 variants were *de novo*, suggesting germline mutations in parental gametes, while in three cases parental DNA was not available (**Fig. 1** and **Table 1**). None of the identified mutations have been reported in the ExAC, EVS or dbSNP144 databases (accessed via the UCSC browser, November 2016), all mutations affect highly conserved residues (**Supplementary Fig. 3**) and all are predicted damaging by PolyPhen-2 (**Table 1**). Remarkably, all 14 mutations are located in exons 3, 8-10, 12 and 13 of *SMCHD1* (48 exons total); these exons code for the ATPase domain of SMCHD1 and an associated region immediately C-terminal (see further below). Notably, six of the 14 patients had mutations affecting three adjacent amino acids: Ala134, Ser135 and Glu136, while p.His348Arg and p.Asp420Val were identified in three and two independent patients respectively, suggesting possible hotspots (**Fig. 1m**). Mutations in *SMCHD1* in arhinia patients have also been identified in an independent study that includes six of the cases analyzed here (cases 2, 4, 5, 6, 7 and 13; Shaw *et al*, accompanying manuscript).

During craniofacial development, the olfactory placode ectoderm thickens and invaginates to form the olfactory epithelium within the nasal cavity, a process that depends on cross-talk between the placodal epithelium and the underlying cranial neural crest-derived mesenchyme[3]. For example, ablation of the nasal placode epithelium in chick embryos disrupts development of adjacent nasal skeletal elements[4]. We observed strong X-gal staining in the developing face of mouse embryos expressing *lacz* from the *Smchd1* locus[5], including in the nasal placodes and optic vesicles at E9.5 and nasal epithelium at E12.5 (**Supplementary Fig. 4**). Eurexpress *in situ* hybridization data indicates regional expression of *Smchd1* in the nasal cavity at E14.5, while transcriptional profiling of post-natal olfactory epithelium demonstrated that *Smchd1* is specifically expressed in immature olfactory sensory neurons[6]. 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 of the

reproductive axis have occasionally been reported in males with arhinia[1,2,7]; we confirm this finding and also report pubertal delay or anomalies of menarche in all three post-pubertal age 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, being required for CpG island (CGI) methylation on the inactive X (Xi), CGI-independent silencing of some X chromosome genes, and Xi compaction[5,8–10]. In addition, Smchd1 functions as an epigenetic repressor at various autosomal loci, with dysregulation of imprinted and monoallelically-expressed gene clusters observed in mutant mice[9,11,12]. A requirement for *SMCHD1* in repair of DNA double-strand breaks has also been demonstrated[13,14]. Whereas female mice null for *Smchd1* display midgestation lethality due to derepression of inactive X chromosome genes, male mutant mice display perinatal lethality of undescribed causes in certain strains or viability on the FVB/n background[11]. Strikingly, craniofacial abnormalities have not been documented in *Smchd1* loss-of-function mice regardless of their sex.

Recently, haploinsufficiency of *SMCHD1* was reported as a cause of facioscapulohumeral muscular dystrophy (FSHD) type 2 (FSHD2) (OMIM 158901)[15]. FSHD has a prevalence of 1/20,000, with FSHD type 1 (FSHD1) and FSHD2 accounting for ~95% and ~5% of cases, respectively[16]. 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 (OMIM 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[16,17]. 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[15,18]. SMCHD1 is thought to function as a silencer at the 4q locus via binding to the D4Z4 repeats[15]. Over 80 unique, putatively pathogenic *SMCHD1* variants have been reported in FSHD2 patients (LOVD *SMCHD1* variant database; see URLs). These mutations, which include clear loss-of-function alleles, occur throughout the protein, and are not clustered in specific domains. Several loss-

of-function mutations have also been reported in ExAC (**Fig. 1m**), and over 60 deletions affecting *SMCHD1* have been reported in the DECIPHER database (available phenotypic information does not indicate arhinia). We analyzed the methylation status of D4Z4 repeats in peripheral blood leukocytes in BAMS patients by sodium bisulphite sequencing (**Supplementary Table 3** and **Supplementary Figs. 5-7**). Although a trend for hypomethylation was noted for BAMS patients relative to controls or unaffected family members, depending on the site tested within D4Z4, some BAMS patients were normally methylated. A large variability in D4Z4 methylation has also been observed in controls and FSHD patients[19], and is not an absolute indicator of FSHD. Moreover, an important argument against BAMS and FSHD2 mutations acting in the same direction is the absence (to our knowledge) of BAMS and FSHD co-occurring in the same patient in the literature. None of the BAMS patients reported here have signs of muscular dystrophy, including both the individuals (2 and 12) older than the average age of FSHD2 onset of 26 years[20], and none of the BAMS missense mutations identified here have been associated with FSHD2.

Proteins of the SMC family are involved in chromatid cohesion, condensation of chromosomes and DNA repair. SMCHD1 is considered 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[21] (**Fig. 1m**). Potentially, SMCHD1 uses energy obtained from ATP hydrolysis to manipulate chromatin ultrastructure and interactions. Using small angle X-ray scattering, the purified 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 Hsp90[21]. Consistent with this, the Hsp90 inhibitor radicicol decreased the ATPase activity of Smchd1[21,22]. Mapping of the SMCHD1 amino acids mutated in BAMS and FSHD2 to the homology model of Smchd1 based on the Hsp90 crystal structure indicates that the major cluster of BAMS mutations (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[23] (**Supplementary Figs. 3** and **8**). The finding of other BAMS mutations in 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 BAMS patients are clustered in the extended ATPase domain and (v) in contrast to *SMCHD1* depletion[13,14], BAMS mutations do not cause DNA damage response alterations or impaired non-homologous end joining (**Supplementary Fig. 9**), we hypothesized that the BAMS mutations may result in a gain- rather than a loss-of-function of the SMCHD1 protein. To test this hypothesis, we conducted ATPase assays using the purified recombinant N-terminal region harboring BAMS or FSHD2 mutations. Compared to wildtype, hydrolysis of ATP was increased for the N-terminal region containing the mutations p.Ala134Ser, p.Ser135Cys or p.Glu136Gly, strongly or slightly decreased for the FSHD2 mutations p.Tyr353Cys[15] or p.Thr527Met[18], respectively, and unchanged for the BAMS mutation p.Asp420Val (**Fig. 2a-f**). The half-maximal inhibitory concentration (IC_{50}) of radicicol was similar for BAMS mutant and wildtype recombinant protein ATPase activities (**Supplementary Fig. 10**), suggesting that the mutants retain an intact ATP-binding site. These results suggest that BAMS-associated mutations elevate the catalytic activity of SMCHD1.

We next sought to validate these biochemical results *in vivo* using full-length SMCHD1 protein. In *Xenopus laevis*, the expression of *smchd1* begins zygotically, and rises steadily after gastrulation (**Fig. 3a**). Endogenous *smchd1* 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 8-cell stage *Xenopus* embryos were micro-injected with 120 pg of capped mRNA encoding either wildtype or mutant human SMCHD1 (**Fig. 3c**). Each set of injected embryos was checked to ensure human SMCHD1 protein expression (**Fig. 3g, Supplementary Fig. 11**). Only tadpoles overexpressing *SMCHD1* mRNA with BAMS mutations showed noticeable craniofacial anomalies (**Fig. 3d-f, Supplementary Fig. 12**), including microphthalmia and in severe cases, anophthalmia (**Fig. 3f'**). At 4 days post fertilization, quantification of the eye size showed a marked reduction in the eye diameter in tadpoles overexpressing BAMS mutants whereas tadpoles overexpressing wildtype SMCHD1 or p.Tyr353Cys, an FSHD2 mutation, were indistinguishable from control uninjected embryos (**Fig. 3h**). One of the BAMS mutants with phenotypic effects in this assay, p.Asp420Val, showed no change in ATPase activity *in vitro* (**Fig. 2**),

suggesting higher sensitivity of the *in vivo* assay. Whole mount *in situ* hybridization showed a decrease in the size of the eye and nasal placodes, marked by *rx2a* and *six1* respectively, upon overexpression of a BAMS mutant (**Fig. 3i,j**). In contrast, migration of cranial neural crest, marked by *twist1*, was largely unaffected. Craniofacial anomalies were dose-dependent for both wildtype and BAMS mutant *SMCHD1* injections, while overexpression of the FSHD2 mutant p.Tyr353Cys was without effect regardless of dose (**Fig. 3k**, **Supplementary Fig. 12**). The finding that wildtype SMCHD1, when overexpressed at a sufficiently high dose, acts in the same phenotypic direction as the BAMS 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 severe BAMS patients, further support the notion that, in contrast to FSHD2 alleles, BAMS-associated missense mutations may exhibit gain-of-function or neomorphic activity. We have not formally excluded the possibility that BAMS mutations may behave as dominant negatives through heterodimerization with wildtype protein. However, we believe this is unlikely, given the effects described above for overexpressed wildtype SMCHD1 and the finding that the isolated ATPase domain containing BAMS mutations can increase ATPase activity alone (**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.

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 PAX6 and FGF and BMP signaling[2], as well as to uncover other potential human-specific nasal regulators. Nose shape and size vary greatly between human populations and even more drastically among animal species, 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 a 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, 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 appear 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/; ExAC Browser, http://exac.broadinstitute.org; DECIPHER database, https://decipher.sanger.ac.uk; Unabridged *Xenopus* protocols, http://www.reversade.com-a.googlepages.com/protocols/.

Data Availability Statement. Whole-exome sequencing data has been deposited in the European Genome-phenome Archive (EGA), with accession numbers:

Acknowledgements

We would like to thank all family members and their relatives for their participation

and kind contribution to this study. Prof. 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, A*STAR Investigator and EMBO Young Investigator. 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 number 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 number 01GM0801, E-RARE network CRANIRARE-2), the German Research Foundation (SFB1002, project D02) to B.W., MACS, VICTA and Baillie Gifford grant support to Ni.R., Mahidol University and Research Career Development Awards from the Faculty of Medicine Ramathibodi Hospital to D.W., 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.-I.Y., C.B.-F., Pa.N., P.N., C.B., A.S.M.T., A.J., H.T., Ja.A. and G.Y.. Genetic studies were supervised by C.T.G., J.A., B.W., A.M.H. and B.R.. The team consisting of B.R., A.J., S.X., H.K. and D.W. independently identified *SMCHD1* mutations in patients 9-12 and 14. H.K., D.W., C.C., G.T., Ni.R., R.M., A.C.M., N.O., V.V., R.I., S.S., De.W., 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 ATPase assays. S.X., M.K.K. and B.R. performed and analyzed functional experiments in *Xenopus*. N.R. 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 methylation studies. The manuscript was written by C.T.G. with contributions from S.X., H.F., J.A. and B.R.. All authors read and approved its content.

Competing Financial Interests

The authors declare no competing financial interests.

References

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.* **170,** 1302–1307 (2016).

2. Graham, J. M. & Lee, J. Bosma arhinia microphthalmia syndrome. *Am. J. Med. Genet. A.* **140,** 189–193 (2006).

3. Forni, P. E. & Wray, S. GnRH, anosmia and hypogonadotropic hypogonadism--where are we? *Front. Neuroendocrinol.* **36,** 165–177 (2015).

4. Szabo-Rogers, H. L. *et al.* Novel skeletogenic patterning roles for the olfactory pit. *Development* **136,** 219–29 (2009).

5. 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).

6. 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).

7. Tryggestad, J. B., Li, S. & Chernausek, S. D. Hypogonadotropic hypogonadism presenting with arhinia: a case report. *J. Med. Case Reports* **7,** 52 (2013).

8. 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).

9. 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).

10. 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).

11. 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).

12. Chen, K. *et al.* Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E3535–3544 (2015).

13. 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).

14. 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).

15. 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).

16. Hewitt, J. E. Loss of epigenetic silencing of the DUX4 transcription factor gene in facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.* **24,** R17–23 (2015).

17. Lemmers, R. J. L. F. *et al.* A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* **329,** 1650–1653 (2010).

18. 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).

19. 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).

20. de Greef, J. C. *et al.* Clinical features of facioscapulohumeral muscular dystrophy 2. *Neurology* **75,** 1548–1554 (2010).

21. Chen, K. *et al.* The epigenetic regulator Smchd1 contains a functional GHKL-type ATPase domain. *Biochem. J.* (2016). doi:10.1042/BCJ20160189

22. 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).

23. Dutta, R. & Inouye, M. GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem. Sci.* **25,** 24–28 (2000).

Figure Legends

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

Figure 2. Biochemical assays indicate that BAMS-associated mutations in SMCHD1 show increased ATPase activity. (**a-e**) ATPase assays performed using recombinant protein encompassing amino acids 111-702 of Smchd1. (**a**) wildtype, (**b**) p.Ala134Ser, (**c**) p.Ser135Cys, (**d**) p.Glu136Gly, (**e**) p.Tyr353Cys. 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 mean ± s.d. of technical triplicates. Each plot is representative of at least two independent experiments using different batches of protein preparation. (**f**) Relative ATPase activities of the mutant proteins compared to wildtype protein. The amount of ADP produced by the mutant proteins was normalised to that of wildtype protein at each protein and substrate concentration as in (**a-e**). The normalised values are plotted as mean ± s.d. of biological replicates (n=44 for p.Ala134Ser, n=24 for p.Ser135Cys, n=32 for p.Glu136Gly, p.Asp420Val, p.Tyr353Cys and p.Thr527Met). In addition to analyzing normalised fold changes, for each mutant the mean of the triplicates at each protein/ATP concentration was compared to that of wildtype using the Wilcoxon matched-pairs signed rank test; apart from p.Asp420Val with a p-value of 0.1776 (non-significant), all the other mutants had a p-value <0.0001 (significant).

Figure 3. *In vivo* functional assays in *Xenopus* embryos suggest that BAMS mutations behave as gain-of-function alleles. (**a**) Expression of *smchd1* relative to 18S rRNA by qPCR. (**b**) In late tailbud stages, *smchd1* expression is restricted to the head region and the neural tube. (**c**) To target the head structures, the dorsal-animal blastomeres of the 8-cell stage *Xenopus* embryo were injected with synthesized mRNAs (120 pg for all panels except **k**). These cells are fated to give rise to head structures as revealed by Dextran lineage tracing. (**d-f'**) Representative stage 45 tadpoles injected with $SMCHD1^{A134S}$ display craniofacial anomalies and smaller eyes compared to control and $SMCHD1^{WT}$ injected tadpoles. Scale bar represents 0.3 mm. All pictures were taken at the same magnification. (**g**) Western blot of stage 12 embryonic extracts from control and injected embryos shows exogenous human SMCHD1 expression. (**h**) The eye diameter is significantly reduced in embryos overexpressing BAMS mutants (blue) relative to $SMCHD1^{WT}$ overexpressing siblings (black), or embryos overexpressing an FSHD2 mutant (open circles). n = at least 15 embryos for each condition. (**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 $SMCHD1^{A134S}$ mRNA (**j**). Pictures are representative of n= 9/10, 7/10, 10/10 embryos for each probe. Dotted lines outline nasal placodes in middle panels and the eye in the right panels. Numbers label streams of migrating cranial neural crest. Scale bar represents 0.2 mm (same magnification for each **i** to **j** comparison). (**k**) Measurements of eye diameter of *Xenopus* embryos injected with 0.5 ng or 1 ng wildtype or BAMS mutant *SMCHD1* mRNA show that *SMCHD1* overexpression causes dose-dependent craniofacial anomalies. Biological variation between clutches of tadpoles is seen in the data presented in panels **h** and **k**. n = 20 embryos for each condition. Data are shown as means ± s.d.; p values were calculated by Kruskal-Wallis test followed by Dunn's post test.

Table

Table 1. *SMCHD1* mutations identified in patients 1-14. (1) NCBI Reference Sequence: NM_015295.2. (2) UniProtKB identifier: A6NHR9. Individuals also studied by Shaw *et al* are indicated with an asterisk.

Individual	Geographic origin	Nucleotide change (1)	Amino acid change	Predicted functional effect (Polyphen-2 score) (2)	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)

Online Methods

Subjects

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

Whole-exome sequencing

WES 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 (WES) of case 1, Agilent SureSelect libraries were prepared from 3 µg of genomic DNA from each individual and sheared with a Covaris S2 Ultrasonicator. Exome capture was performed with the 51 Mb 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-, 149- 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)[24], SAMtools[25] and Picard. Variant calls were made with the GATK Unified Genotyper. 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[26]. Variants were filtered against publicly available SNPs plus variant data from more than 7,000 in-house exomes (Institut *Imagine*).

For trio WES 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 HiSeq2000 sequencer at the Cologne Center for Genomics (CCG). Data analysis and filtering of mapped target sequences was performed with the "Varbank" exome and genome analysis pipeline v.2.1 (CCG) and data were filtered for high-quality (coverage of more than 6 reads, a minimum quality score of 10), rare (MAF < 0.5%) autosomal recessive and *de novo* variants.

For cases 9 and 11 trios and 10 and 12 quartets, WES was performed at the Genome Institute of Singapore. Barcoded libraries were prepared for each individual by shearing 1ug of genomic DNA, followed by end-repair, A-tailing, adaptor ligation and PCR enrichment, then pooled and hybridized with NimbleGen SeqCap EZ Human Exome Library v3.0 probes. Captured DNA targets were purified and PCR amplified, then sequenced on Illumina HiSeq 2500 (cases 9 and 11) or HiSeq 4000 (cases 10 and 12) sequencers. Variant calling was performed following GATK (v3.4.46) recommended best practices. Reads were mapped to GRCh37/hg19 using BWA and the aligned files pre-processed by Picard and GATK[24,27,28]. All samples were sequenced at mean coverage of 75X or higher. The variants were called using GATK HaplotypeCaller along with in-house exomes sequenced with the same chemistry. The variants were recalibrated, annotated and filtered against in-house data plus common publicly available databases. Each family was independently analyzed using Phen-Gen[29] for *de novo* and recessive disease inheritance patterns. Variants with alternate allele frequency ≤ 10 or coverage ≤ 20 were not considered.

For case 13 WES, 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 HiSeq2500 (Illumina). Raw data files were converted to FASTQ files with the 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, SNV information in dbSNP[30] build 138 was used for base quality score recalibration. Marking of PCR duplicates and position-wise sorting was performed with Novosort version 3 (Novocraft). Calling of SNVs and small indels was performed using GATK[24,27,28] version 3.4-46. A GATK workflow[31] was used in which local realignment and variant calling were performed by IndelRealigner and HaplotypeCaller, respectively. Low quality SNV and small indel calls were removed using the following criteria: QD<2.0, MQ<40.0, FS>60.0, MQRankSum<-12.5, ReadPosRankSum<-8.0 for SNVs; QD<2.0, ReadPosRankSum<-20.0, FS>200.0 for small indels. SNVs and small indels were annotated with the ANNOVAR software package[32] using the following datasets and programs: gene information from GENCODE[33] (version 19); allele frequencies of the 1000 Genome Project[34] (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[35] and SIFT[36] via dbNSFP[37,38] (version 3.0).

DNA methylation analysis

Sodium bisulfite sequencing. 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 minutes at 37°C in NaOH 0.4N and incubated overnight in a solution of sodium bisulfite 3M pH 5 and 10 mM hydroquinone using a previously described protocol[39]. Converted DNA was then purified using the Wizard DNA CleanUp kit (Promega) following the manufacturer's recommendations and precipitated by ethanol precipitation for 5 hours 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 the MethPrimer software[40] avoiding the presence of CpGs in the primer sequence in order to amplify 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 two minutes, amplification was done at 94°C for 20 seconds, 54°C for 30 seconds and 72°C for one minute for 10 cycles, then at 94°C for 20 seconds, 54°C for 30 seconds, followed by an extension step of 4 minutes and 30 seconds for the first cycle and an increment of 30 seconds at each subsequent cycle for 25 cycles. At the end of the program, a final extension step at 72°C for 7 minutes was performed. PCR products were then 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 PCR amplified directly using T7 or SP6 primers. For each sample and region, at least ten randomly cloned PCR products were sequenced according to Sanger's method by Eurofins MWG Operon (Ebersberg, Germany) with either SP6 or T7 primers. Sequences were analyzed using the BiQ Analyser software[41] and the average methylation score was calculated as the number of methylated CpGs for the total number of CpGs in the reference sequence.

Statistics and subjects. The average methylation level of each group of samples (FSHD2 patients carrying a *SMCHD1* mutation, control individuals and BAMS patients and their relatives) was 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 that have been previously reported[42]. The FSHD2 patients carrying a *SMCHD1* mutation have been previously reported[42,43] and comprise n=8 for the DR1 region and n=15 for each of the 5' and Mid regions, while for the DR1 region 21 additional patients for whom sodium bisulfite sequencing data exists 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 (Protein Homology/Analogy Recognition Engine 2)[44]. Protein sequence of 111-702 aa of mouse Smchd1 was submitted as the input sequence

and intensive modelling mode was selected. The second highest scoring model with the most sequence alignment coverage based on the crystal structure of yeast Hsp90 (PDB: 2CG9) was elected for further evaluation. The model was visualized in PyMOL. The multiple sequence alignment was generated using CLUSTAL W[45] (via the PBIL server) and ESPript 3.0[46].

ATPase assays

Cloning, expression and purification of recombinant mouse Smchd1 protein was performed as previously described[21], and the primers used for cloning and mutagenesis are provided in **Supplementary Table 5**. The purity of the protein preparations was judged by migration of samples on 4-20% (w/v) Tris-Glycine reducing SDS/PAGE gels followed by staining with SimplyBlue SafeStain (Thermo Fisher Scientific, USA) (**Supplementary Figure 13**). The ATPase assay was performed with the Transcreener ADP2 fluorescence polarization assay kit (BellBrook Labs) as previously described[21]. Briefly, 10 µl reactions in triplicate were set up in 384-well (low volume, black) plates, containing 7 µl reaction buffer (50 mM HEPES pH 7.5, 4 mM $MgCl_2$ and 2 mM EGTA), 1 µl of recombinant Smchd1 111-702 aa protein at concentrations ranging from 0.1-0.6 µM or buffer control, 1 µl of radicicol or solvent control and 1 µl of 10 µM ATP substrate or nuclease-free water control. 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 hour in the dark before addition of 10 µl of detection mix (1X Stop & Detection Buffer B, 23.6 µg/ml ADP2 antibody) for a further hour of incubation. 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 by using the standard curve following the manufacturer's instructions.

Mouse embryo dissection and X-gal staining

Mice were housed and mouse work approved under the Walter and Eliza Hall Institute of Medical Research Animal Ethics Committee approval (AEC 2014.026). Embryos were produced from C57BL/6 *Smchd1*$^{gt/+}$ congenic strain sires mated with C57BL/6 dams, with embryo ages ranging from embryonic day 8.5 to embryonic day 12.5[5]. 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 laevis were used according to guidelines approved by the Singapore National Advisory Committee on Laboratory Animal Research. Protocols for fertilization, injections and whole mount *in situ* hybridization are available at the Reversade lab's protocol website (see URLs). Human *SMCHD1* (Origene) was cloned into expression vector pCS2+, linearized with NotI and transcribed with mMESSAGE mMachine SP6 transcription kit (Thermo Fisher). Transcribed mRNA was column purified and its concentration measured using a Nanodrop. 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 8-cell stage with the synthesized mRNA. Embryos were allowed to develop at room temperature until stage 45-46 (4 days post fertilization) and 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. mRNAs injected for **Fig. 3k** did not contain a poly A signal and were polyadenylated *in vitro*, hence requiring higher RNA concentration to produce a phenotype (in other panels in **Fig. 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 Western blot with anti-SMCHD1 (Atlas HPA039441) and anti-GAPDH antibodies (clone 0411, Santa

Cruz). cDNA was made from RNA extracted from *Xenopus laevis* embryos of various stages using iScript reverse transcriptase (Bio-Rad). qPCR was performed using the following primers, xsmchd1_qPCR_F 5'- CAGTGGGTGTCATGGATGCT, xsmchd1_qPCR_R 5'- TCCATGGCTAGACCACTTGC, XL_18S_F 5'- GCAATTATTTCCCATGAACGA, XL_18S_R 5'- ATCAACGCGAGCTTATGACC. In situ hybridization probe for *smchd1* was amplified from stage 20 cDNA using primers 5'-CGAATGCAAAGTCCTTGGGC and 5'-GCATCCATGACACCCACTGA, cloned into pGEM-T, linearized and transcribed using DIG-labelling mix (Roche) according to manufacturer's guidelines.

DNA damage response assays

Cell lines and cell cultures. *XRCC4*-deficient cells[47] and primary fibroblast cell lines established from cases 1 and 2 were cultured in Dulbecco's modified Eagle medium (DMEM, Gibco) supplemented with 10% fetal calf serum (FCS, Gibco), and antibiotics. Testing for mycoplasma contamination was negative. For H2AX activation, cells were either irradiated with 100 J/m^2 UV-C or treated with 50 µM etoposide (Sigma-Aldrich, USA) for 1 hour. Drugs were then washed out, fresh media was added, and cells were incubated for 6 hrs and then subjected to Western blot analysis.

Protein isolation and analysis. Cells were solubilized by 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; protease inhibitors P 2714 [Sigma-Aldrich, USA]). The total protein concentration of extracts was determined using the BCA Protein Assay Kit (Thermo Fisher Scientific, USA). 10 µg of total cell lysates were separated by 4-12 % SDS-PAGE (Invitrogen, Germany) and blotted onto nitrocellulose membranes (GE Healthcare, Germany). Protein detection was performed using antibodies specific for phosphorylation of H2AX at Ser139 (γH2AX) (clone 20E3, Cell Signaling Technology, USA). Anti-β-Actin antibodies were purchased from Sigma-Aldrich (clone AC-74). Secondary antibodies conjugated to peroxidase (Santa Cruz Biotechnology Inc., USA) were used and blots were developed using an enhanced chemiluminescence system, ECL Plus (GE Healthcare), followed by detection on autoradiographic films.

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

Online Methods References

24. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20,** 1297–303 (2010).

25. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–9 (2009).

26. Flicek, P. *et al.* Ensembl 2013. *Nucleic Acids Res* **41,** D48–55 (2013).

27. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–498 (2011).

28. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al* **43,** 11.10.1–33 (2013).

29. Javed, A., Agrawal, S. & Ng, P. C. Phen-Gen: combining phenotype and genotype to analyze rare disorders. *Nat. Methods* **11,** 935–937 (2014).

30. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–311 (2001).

31. Mishima, H., Sasaki, K., Tanaka, M., Tatebe, O. & Yoshiura, K.-I. Agile parallel bioinformatics workflow management using Pwrake. *BMC Res. Notes* **4,** 331 (2011).

32. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).

33. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22,** 1760–1774 (2012).

34. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

35. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–249 (2010).

730 36. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-
731 synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4,**
732 1073–1081 (2009).

733 37. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: a lightweight database of human
734 nonsynonymous SNPs and their functional predictions. *Hum. Mutat.* **32,** 894–899
735 (2011).

736 38. Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: A One-Stop Database of
737 Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site
738 SNVs. *Hum. Mutat.* **37,** 235–241 (2016).

739 39. Magdinier, F. *et al.* Regional methylation of the 5' end CpG island of BRCA1 is
740 associated with reduced gene expression in human somatic cells. *FASEB J. Off.*
741 *Publ. Fed. Am. Soc. Exp. Biol.* **14,** 1585–1594 (2000).

742 40. Li, L.-C. & Dahiya, R. MethPrimer: designing primers for methylation PCRs.
743 *Bioinforma. Oxf. Engl.* **18,** 1427–1431 (2002).

744 41. Bock, C. *et al.* BiQ Analyzer: visualization and quality control for DNA
745 methylation data from bisulfite sequencing. *Bioinforma. Oxf. Engl.* **21,** 4067–4068
746 (2005).

747 42. Gaillard, M.-C. *et al.* Differential DNA methylation of the D4Z4 repeat in
748 patients with FSHD and asymptomatic carriers. *Neurology* **83,** 733–742 (2014).

749 43. Gaillard, M.-C. *et al.* Segregation between SMCHD1 mutation, D4Z4
750 hypomethylation and Facio-Scapulo-Humeral Dystrophy: a case report. *BMC Med.*
751 *Genet.* **17,** 66 (2016).

752 44. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The
753 Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10,**
754 845–858 (2015).

755 45. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the
756 sensitivity of progressive multiple sequence alignment through sequence weighting,
757 position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,**
758 4673–4680 (1994).

759 46. Robert, X. & Gouet, P. Deciphering key features in protein structures with the
760 new ENDscript server. *Nucleic Acids Res.* **42,** W320–324 (2014).

761 47. Rosin, N. *et al.* Mutations in XRCC4 cause primary microcephaly, short stature
762 and increased genomic instability. *Hum. Mol. Genet.* **24,** 3708–3717 (2015).

763 48. Verkaik, N. S. *et al.* Different types of V(D)J recombination and end-joining
764 defects in DNA double-strand break repair mutant mammalian cells. *Eur. J. Immunol.*
765 **32,** 701–709 (2002).

766

767

768

Supplementary information

Supplementary Figure 1. Computed tomography and magnetic resonance imaging (MRI) in BAMS. (**a-c**) Controls and (**d-f**) patient 1 at four years. Patient 1 displays maxillary hypoplasia and absent nasal bones (**d** and **e**). Olfactory bulbs and sulci (labelled with red and white arrows, respectively, on the left side in the control in **c**) are absent in patient 1 (**f**). (**g-i**) patient 5, with right microphthalmia as shown by MRI (**i**). Skeletal imaging of patients 14 (**j,k**) and 11 **(l)** indicating midface hyploplasia.

Supplementary Figure 2. BAMS pedigrees and Sanger sequencing chromatograms of *SMCHD1* mutations. Individuals submitted for exome sequencing are indicated by a red asterisk. Note Sanger sequencing was unavailable for individual 13.

Supplementary Figure 3. Multiple sequence alignment of vertebrate SMCHD1 orthologues and yeast Hsp90. Residues mutated in BAMS are indicated by pink arrows. Residues mutated in FSHD are indicated by purple arrows. Hs, *Homo sapiens*; Mm, *Mus musculus*; Bt, *Bos taurus*; Gg, *Gallus gallus*; Md, *Monodelphis domestica*; Cm, *Chelonia mydas*; Xt, *Xenopus tropicalis*; Dr, *Danio rerio*; Sc, *Saccharomyces cerevisiae*. FSHD mutation reference: LOVD *SMCHD1* variant database, http://databases.lovd.nl/shared/variants/SMCHD1/unique.

Supplementary Figure 4. X-gal staining of mouse embryos expressing *lacz* from the *Smchd1* locus. E, embryonic day. gt/+, embryos heterozygous for the *Smchd1*^{gt} allele expressing lacz. +/+, wildtype embyros. hf, head folds. npl, nasal placode. ov, optic vesicle. npi, nasal pit. ne, nasal epithelium. **f-i**, coronal sections. **r** and **s**, transverse sections. An asterisk in panel **p** indicates deep nasal staining.

Supplementary Figure 5. Sodium bisulfite sequencing in BAMS patients (individuals 1-6). The position of the three different regions analyzed within D4Z4 is indicated above the corresponding column (left, DR1; middle, 5'; right, Mid). For each sample, at least 10 cloned DNA molecules were analyzed by Sanger sequencing. Each histogram column corresponds to a single CpG. Black corresponds to the global percentage of methylated CpGs and white to the global percentage of unmethylated

CpGs. The percentage of methylated CpGs among the total CpGs in each individual analyzed are given in **Supplementary Table 3**.

Supplementary Figure 6. Sodium bisulfite sequencing in BAMS patients (individuals 8-11 and 14). See legend of **Supplementary Figure 5** for further information.

Supplementary Figure 7. Comparison of D4Z4 methylation in BAMS or FSHD2 patients, BAMS patient relatives and controls. Distribution of methylation for the three different regions within the D4Z4 sequence (DR1, 5' and Mid) in control individuals, patients with FSHD2 carrying a *SMCHD1* mutation and BAMS patients and their relatives. Means ± SEM are shown. A Kruskal-Wallis multiple comparisons test was performed, followed by a Dunn's test and Bonferroni correction, with $\alpha = 0.05$. ***, $p<0.0001$; **, $p<0.001$; *, $p<0.05$. Blue points indicate outliers. Red crosses indicate medians. The level of methylation is statistically significantly different between controls and FSHD2 patients for the DR1 (** ; $p<0.001$) and the 5' (*** ; $p< 0.0001$) regions. The level of methylation is significantly different between controls and BAMS patients for the 5' region (*, $p<0.05$) and between BAMS patients and their relatives for the DR1 (*, $p<0.05$) and 5' (**, $p<0.001$) regions.

Supplementary Figure 8. Smchd1 structure modelling, based on the structure of Hsp90. Residues mutated in BAMS are indicated in pink. Residues mutated in FSHD are indicated in purple.

Supplementary Figure 9. Fibroblasts derived from BAMS patients show no defects in NHEJ or in H2AX activation. (**a**) A microhomology mediated end joining (MMEJ) assay was performed on wildtype (WT), *XRCC4*-deficient and case 1 and 2 fibroblasts. Whereas *XRCC4*-deficient fibroblasts show multiple smaller DNA bands after BstXI digestion indicating defects in NHEJ-mediated DNA repair and leading to preferential use of MMEJ-mediated DNA double strand repair, BAMS patient fibroblasts show no defects in NHEJ-mediated DNA repair pathways compared to wildtype. (**b**) Western blot analysis of UV- and etoposide-induced phosphorylation of H2AX at Ser139 (γH2AX). Wildtype fibroblasts (WT) and fibroblasts derived from cases 1 and 2 were treated with UV-C (UV) or etoposide (Eto) or left untreated as a control (-). Cells were lysed and subjected to Western blot analysis with an antibody against γH2AX. Equal protein loading was confirmed by reprobing of the membrane

with an antibody against β-Actin. Wildtype and BAMS patient fibroblasts did not show significant differences in H2AX activation.

Supplementary Figure 10. ATPase assays performed using recombinant wildtype or mutant Smchd1 protein in the presence of radicicol. Data are displayed as mean ± s.d. of technical triplicates. The data are representative of at least two independent experiments using different batches of protein preparation.

Supplementary Figure 11. Full-length Western blot of the cropped blot image in **Fig. 3g**.

Supplementary Figure 12. *SMCHD1* overexpression in *Xenopus* causes dose-dependent craniofacial anomalies. (**a,b**) Measurements of eye diameter of *Xenopus* embryos injected with 240 pg (**a**) or 500 pg (**b**) *SMCHD1* mRNA. Y353C is an FSHD2 mutation. n = at least 20 embryos for each condition. (**c-f**) Representative *Xenopus* embryos injected with 500 pg of WT or FSHD2 mutant *SMCHD1* or 120 pg of BAMS mutant mRNA show varying degrees of craniofacial abnormalities as compared to uninjected control tadpoles at 4 days post fertilization. Data are shown as means ± s.d.; p values were calculated by Kruskal-Wallis test followed by Dunn's post test. n.s. not significant.

Supplementary Figure 13. Purity of proteins used for ATPase assays. Purified recombinant wild type or mutant proteins were resolved by 4-20% (w/v) Tris-Glycine reducing SDS/PAGE and were stained with SimplyBlue SafeStain. Protein quantities loaded: left gel, 1.4 µg; middle gel, 1.05 µg; right gel, 0.7 µg. Molecular weight (MW) markers are as indicated on the left-hand side.

Supplementary Table 1. Clinical features of 14 BAMS patients.

Supplementary Table 2. Exome variant filtering for cases 1, 2 and 9-13.

Supplementary Table 3. DNA methylation analysis in BAMS probands and family members. Three different regions within the D4Z4 macrosatellite repeat were analyzed: DR1 (as described in Hartweck *et al*., *Neurology*, **80**, 392-399 (2013)); 5' and Mid (as described in Gaillard *et al*., *Neurology*, **83**, 733-742 (2014)). The Mid region corresponds to the *DUX4* promoter. % M+ indicates the percentage of methylated CpGs among the total CpGs for a given region. X indicates samples that

were not analyzed. All samples were obtained from peripheral blood leukocytes except for individual 4's brother and sister and individual 14, which were from saliva.

Supplementary Table 4. Primers used for sodium bisulfite PCR.

Supplementary Table 5. Primers used for cloning and mutagenesis of recombinant murine *Smchd1*.