HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2017 June 16.

Published in final edited form as:

Nat Genet. 2017 February ; 49(2): 238–248. doi:10.1038/ng.3743.

SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome

A full list of authors and affiliations appears at the end of the article.

Abstract

Arhinia, or absence of the nose, is a rare malformation of unknown etiology that is often accompanied by ocular and reproductive defects. Sequencing of 40 people with arhinia revealed that 84% of probands harbor a missense mutation localized to a constrained region of *SMCHD1* encompassing the ATPase domain. *SMCHD1* mutations cause facioscapulohumeral muscular dystrophy type 2 (FSHD2) via a *trans*-acting loss-of-function epigenetic mechanism. We discovered shared mutations and comparable DNA hypomethylation patterning between these distinct disorders. CRISPR/Cas9-mediated alteration of *smchd1* in zebrafish yielded arhinia-relevant phenotypes. Transcriptome and protein analyses in arhinia probands and controls showed no differences in *SMCHD1* mRNA or protein abundance but revealed regulatory changes in genes and pathways associated with craniofacial patterning. Mutations in *SMCHD1* thus contribute to distinct phenotypic spectra, from craniofacial malformation and reproductive disorders to muscular dystrophy, which we speculate to be consistent with oligogenic mechanisms resulting in pleiotropic outcomes.

Arhinia, or the complete absence of an external nose, is a rare congenital malformation reported in only 80 patients without holoprosencephaly in the past century (Supplementary Table 1). This severe malformation can be isolated or accompanied by other craniofacial defects, including anophthalmia, coloboma, cataracts, nasolacrimal duct atresia, choanal atresia and cleft palate (Fig. 1). Seventeen patients with arhinia and ocular defects have been reported with coexistent hypogonadotropic hypogonadism, a triad called Bosma arhinia

---

Correspondence should be addressed to M.E.T. (talkowski@chgr.mgh.harvard.edu), D.R.F. (david.fitzpatrick@igmm.ed.ac.uk) or E.E.D. (erica.davis@duke.edu).

[47]These authors contributed equally to this work.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

M.E.T., D.R.F., E.E.D., N.K., P.J., N.D.S. and H. Brand designed the study. N.D.S., L.P., K.A.W., M.N., S.P., T.K., D.L., A. Silva, S.J., J.C.S., M.F.L., S.S.S., N.P., J.R.L., N.F., A.V., A.R., K. Steindl, I.S., D.S., N.O., C.J., J.T., S.C., L.A.S., B.B., C. Cesaretti, J.E.G.-O., T.P.B., O.P.S., J.D.H., W.M., K.W.R., B.L.L., M.S., A.M.K., C.-H.C., C.C.M., V.v.H., R.B., J.E.H., S.B.S., K.Y., J.M.G., A.E.L., W.F.C. and D.R.F. recruited patients and collected clinical information and samples. Z.A.K., H. Bengani, L.P., S.E., T.I.J., J.R.W., J.R., A. Stortchevoi, C.M.S., Y.A., B.B.C., M.A., R.R.M., J.K.R., M.Z., J.W.J., E.C.L., S.A.M., N.K., P.L.J., E.E.D., D.R.F. and D.S.D. performed molecular genetics and animal modeling studies. H. Brand, K. Samocha, R.L.C., C. Chiang, A.L., M.L., J.F.G., D.G.M. and M.E.T. performed genomic analyses. J.A.M. performed protein modeling. N.D.S., H. Brand, N.K., J.F.G., P.L.J., E.E.D., D.R.F. and M.E.T. wrote the manuscript, which was revised and approved by all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

microphthalmia syndrome (BAM; MIM603457)[1]. The rarity of these malformations and the cross-disciplinary nature of coexistent features have limited efforts to systematically catalog associated phenotypes, though these comorbidities suggest that the genetic architecture underlying this condition has broad developmental implications.

Genetic studies of arhinia have been limited to targeted approaches related to cranial neural crest cells (NCC) or craniofacial placodal development. To date, no causal locus has been identified. Homozygous null mutations in *Pax6* arrest nasal placodogenesis in mice[2] and (in *PAX6*) cause rudimentary or malformed noses in humans[3–5], yet null mutations in *PAX6* also cause aniridia and structural brain abnormalities that are not observed in individuals with arhinia[3–5]. We formed an international consortium to aggregate all available cases and determine the genetic etiology of arhinia. We sequenced 40 individuals with arhinia (38 independent families) and 55 family members without arhinia using a combination of whole-exome, whole-genome and targeted sequencing. These analyses revealed that rare missense variants in *SMCHD1* represent the predominant genetic contributor to arhinia. Notably, *SMCHD1* encodes a protein with established epigenetic repressive activities that has been implicated in FSHD2 (MIM158901), a rare, oligogenic form of muscular dystrophy. Methylation studies in arhinia subjects and complementation testing of arhinia- and FSHD2-associated variants in zebrafish (*Danio rerio*) revealed a common direction of allele effect in these disorders, a surprising observation considering their substantial differences in phenotype. Supporting these results, we observed the same mutations in BAM and FSHD2 probands, as well as at least one individual that met all diagnostic criteria and displayed symptoms for both disorders. Given the known oligogenic architecture of FSHD2, these results suggest that missense variants in an evolutionarily constrained region of *SMCHD1* contribute to the diverse manifestations of arhinia, BAM and FSHD2, probably as a result of alteration to a critical function of the protein and/or interaction with other genomic loci.

RESULTS

Samples, phenotypes and epidemiology of arhinia

Our international consortium established a cohort of 40 subjects (38 probands and 2 affected siblings) that encompassed 24% of all 80 previously reported individuals and 21 new cases, facilitating a comprehensive phenotypic picture of arhinia and its associated comorbidities (Supplementary Tables 1 and 2). Six subjects were also included in an independent analysis by Gordon *et al.*[6]. All subjects had complete arhinia, accompanied in most cases by other craniofacial abnormalities, including high-arched or cleft palate, absent paranasal sinuses, hypoplastic maxilla, nasolacrimal duct stenosis or atresia and choanal atresia (Fig. 1), and 41% had dysmorphic pinnae or low-set ears. Ocular phenotypes included anophthalmia or microphthalmia (77%), uveal coloboma (79%) and cataract (53%), and six subjects had normal eye anatomy and vision. Of the 31 assessable subjects (22 male; 9 female), 97% demonstrated hypogonadotropic hypogonadism (HH), and the seven subjects for whom brain MRI data were available had no olfactory structures. Fate mapping would support arhinia as a primary malformation, with HH representing a developmental sequence[7]. Ocular defects, which are not an obvious part of a developmental sequence with arhinia, were

observed in 26 of 31 individuals, indicating that 84% of subjects met BAM diagnostic criteria.

Gene discovery

We sequenced all 40 arhinia subjects as well as 55 family members representing six multiplex and 32 simplex families. We performed whole-exome sequencing (WES) on 27 arhinia subjects and targeted sequencing in 16 subjects, and we applied both methods to 4 probands (Online Methods and Supplementary Fig. 1). Whole-genome sequencing (WGS) was concurrently performed in four members of a multiplex family (family O) that included a proband and affected sister with BAM, a half-aunt with arhinia and relatives with subphenotypes including anosmia and subtle nasal and dental anomalies[8,9] (Supplementary Fig. 1). Collectively, these analyses identified rare missense variants in *SMCHD1* in 84% of probands (32/38), none of which were present in the Exome Aggregation Consortium[10] (ExAC, $n = 60{,}706$). To test whether this represented an unexpected accumulation of rare missense variants, we compared the rare mutation burden among 22,445 genes in the initial 22 probands with WES to variants observed in ExAC (minor allele frequency (MAF) < 0.1%). Powered by the size of our aggregate cohort, we found only one gene that achieved genome-wide significance: *SMCHD1* ($P = 2.9 \times 10^{-17}$, two-sided Fisher's exact test; odds ratio (OR) of 34.4 (95% CI 18.8–57.9); Fig. 2). This result was significant irrespectively of ethnicity (Supplementary Table 3). Notably, in an independent study, Gordon *et al.*[6] included six of these probands and eight additional probands (Table 1). All *SMCHD1* variants arose *de novo* from the 10 complete simplex trios, and mutations were likewise absent in unaffected family members from the remaining simplex families (Supplementary Fig. 1). Segregation of an *SMCHD1* mutation was observed in all three multiplex families with available parental samples, including one (family T) in which the variant was inherited from a father who had no craniofacial abnormalities and had been contemporaneously diagnosed with muscular dystrophy.

SMCHD1 is highly constrained, or intolerant to loss-of-function variation (probability of being loss-of-function intolerant (pLI) = 1.00)[11], with an estimated prevalence of 1 in 10,000 heterozygous null individuals in ExAC. Notably, all arhinia-associated variants localized to exons 3–13 of *SMCHD1* (Ensembl ENST00000320876.10), spanning a GHKL-type ATPase domain (Table 1 and Supplementary Table 2). We thus considered the possibility that this specific domain is intolerant to missense variation using models of regional constraint. We found that the entire gene is not particularly intolerant to missense variation (81% of expected missense variants observed; $P = 0.016$, two sided Z-test; Z-statistic = 2.14; constraint was taken from the ExAC database[10] and therefore the values to calculate 95% CI are unavailable), but this significance is driven by strong constraint in the 5′ region that harbors all detected arhinia-associated variants and encompasses the ATPase domain (exons 1–19, 61% of expected variants observed, $\chi^2 = 32.40$, $P = 1.26 \times 10^{-8}$; exons 20–48, 95% of expected variants observed, $\chi^2 = 0.86$, $P = 0.36$; Fig. 3). This regional constraint suggests that these alleles may impede protein function, and *in silico* prediction of pathogenicity from the Combined Annotation Dependent Depletion (CADD) database revealed that the 20 arhinia-specific *SMCHD1* variants were more deleterious than rare, nonsynonymous variants in ExAC (MAF < 0.01%, ExAC $n = 378$, $P = 1.27 \times 10^{-5}$, two-sample t-test; $t =$

4.42; Supplementary Fig. 2). However, 19 missense variants in ExAC in exons 3–13 have CADD scores exceeding the median arhinia score (17.03), consistent with the observation in arhinia families that *SMCHD1* variants segregated with subtle dysmorphism, no dysmorphic features and muscular dystrophy. These data support our speculation that deleterious *SMCHD1* variants are not fully penetrant, and such variants alone may not be sufficient to cause arhinia.

Mutational overlap between arhinia and FSHD2

SMCHD1 is an epigenetic regulator of autosomal and X-linked genes[12–15]. The discovery of an association between *SMCHD1* and craniofacial development was unexpected, given that mutations in the gene are associated with FSHD2, a rare, *trans*-acting oligogenic form of muscular dystrophy. In FSHD2, loss of SMCHD1 repressive activity, in combination with a permissive D4Z4 haplotype at 4q35, allows for the ectopic expression of the transcript encoding the DUX4 protein, which is cytotoxic to skeletal muscle[16]. *SMCHD1* mutations in FSHD2 span the entire gene and include missense and truncating variants, whereas all arhinia-associated missense variants clustered tightly around the ATPase domain (Fig. 3), which is thought to control the release of DNA bound by SMCHD1 (ref 17). However, the mutational distribution was not fully distinct between these disorders; several previously reported FSHD2-specific missense variants were localized to exons 3–13, and one causal variant in FSHD2 (p.Gly137Glu) was also detected in an arhinia proband (subject AG1)[18]. Neither the FSHD2 nor the arhinia subject had features of both disorders, indicating that these phenotypes either arise by divergent mechanisms or are influenced by additional loci.

Methylation profiles in arhinia and FSHD2

Haploinsufficiency and dominant-negative loss-of-function models have been invoked for *SMCHD1* mutations associated with FSHD2 (ref. 18). In both models, loss of SMCHD1 repressive activity manifests as a decrease in DNA methylation at SMCHD1 binding sites[16,19–21]. Although SMCHD1 interacts with numerous genetic loci, only the 4q35 D4Z4 macrosatellite array, which contains *DUX4*, and the highly homologous 10q26 D4Z4 array are associated with FSHD2, and hypomethylation of these two loci is assessed during diagnostic evaluation[22,23]. To explore mechanistic overlap between arhinia and FSHD2, we quantified 4q35 D4Z4 methylation in 23 arhinia subjects (19 with *SMCHD1* missense variants) and 22 family members using a bisulfite sequencing (BSS) assay specific for the FSHD2-affected D4Z4 arrays on 4q35 and 10q26 (ref. 23). Of these family members, four harbored *SMCHD1* missense variants, including two individuals with anosmia, one with a hypoplastic nose and one with muscular dystrophy (Fig. 4 and Supplementary Table 4). We observed that 74% of arhinia subjects (and two of the four family members) with an *SMCHD1* variant had D4Z4 hypomethylation characteristic of FSHD2, whereas all four arhinia subjects and 16 of the 18 family members without a missense *SMCHD1* variant had normal methylation patterns. These data confirm that arhinia-specific mutations in *SMCHD1* produce the same methylation patterning at D4Z4 as seen in FSHD2, demonstrating that two completely distinct phenotypes can arise from deleterious changes in the same gene and indeed the same alleles. We thus turned to animal models to probe the *in vivo* functional impact of *SMCHD1* variants.

In vivo modeling of *SMCHD1*

We evaluated the functional consequences of *SMCHD1* variation using zebrafish larvae. The zebrafish genome encodes one SMCHD1 ortholog (49% identical, 67% similar to human) with a highly conserved ATPase domain (Supplementary Fig. 3). Of relevance to BAM, eye development is highly conserved between species, making the zebrafish a robust model to study microphthalmia-associated candidate genes[24–26]. Further, *D. rerio* possesses two GnRH paralogs that exist in humans[27], and GnRH neuronal ontogeny is largely conserved between humans and teleosts[28–30]. Cognizant that there is no credible zebrafish structure homologous to the human nose, we evaluated facial cartilage patterning in zebrafish as a potential surrogate phenotype in studies of *smchd1* ablation or ectopic expression.

We designed and validated two morpholino antisense oligonucleotides targeting splice donor sites of two *smchd1* exons encoding the ATPase domain (e3i3 and e5i5, targeting exons 3 and 5, respectively; Supplementary Fig. 3). The e3i3 or e5i5 *smchd1* morpholinos were injected (at 3, 6 or 9 ng per embryo) into *−1.4col1a1:egfp* embryo batches at the one-to-two-cell stage, and larvae were phenotyped quantitatively for aberrant cartilage patterning, ocular development and reproductive axis integrity 1.5–3 d after fertilization (dpf) (Fig. 5). All morphants demonstrated dose-dependent narrowing of the ethmoid plate (Fig. 5a,b, Supplementary Fig. 4a and Supplementary Table 5), a dose-dependent increase in ceratohyal arch angle, delayed (or absent) development of ceratobranchial arches (Fig. 5a and Supplementary Fig. 4b,c) and microphthalmia (tested at the 9-ng dose) (Fig. 5c,d and Supplementary Table 5). Moreover, ventral imaging of wholemount embryos immunostained with a pan-GnRH antibody revealed a prominent phenotype: morphant olfactory bulbs and hypothalami were intact, but the average projection length of the terminal nerve, where GnRH3 neurons reside, was reduced by 45% compared with controls ($P = 1.35 \times 10^{-13}$, Student's *t*-test, reduced by 45%, (95% CI: 39.83–50.17)) (Fig. 5e,f and Supplementary Table 5). The cartilage, eye and GnRH phenotypes were highly specific; each defect was reproduced with both morpholinos tested and rescuable with full-length human wild-type (hWT) *SMCHD1* mRNA (Fig. 5b,d,f, Supplementary Fig. 5 and Supplementary Table 5). To confirm these findings, we used CRISPR/Cas9-mediated genome editing to generate small insertions and deletions into exon 1 of *smchd1*, achieving high mosaic fractions (Supplementary Fig. 6). These F0 mutants recapitulated the craniofacial, ocular and GnRH defects observed in the morphant models (Fig. 5, Supplementary Fig. 5 and Supplementary Table 5).

Having established quantitative *in vivo* assays of disrupted *smchd1* activity, we next tested both gain- and loss-of-function paradigms. To evaluate gain of function, we injected hWT *SMCHD1* mRNA or equivalent doses of human mRNA bearing recurrent arhinia-associated variants (p.Ser135Cys, p.Leu141Phe or p.His348Arg) into zebrafish embryos; none of these overexpression assays yielded craniofacial phenotypes (Supplementary Fig. 7a–c and Supplementary Table 5). Higher doses of mutant mRNA alone and combinatorial injections of mutant and hWT mRNA likewise had little effect, suggesting that, in the context of this assay, a gain-of-function biochemical mechanism is unlikely (Supplementary Fig. 7d–f and Supplementary Table 5). Given that suppression of *smchd1* resulted in three phenotypes relevant to BAM, we next performed *in vivo* complementation, focusing on our most

sensitive assay, the GnRH-positive terminal nerve length. We co-injected the e5i5 morpholino with (i) full-length hWT *SMCHD1* mRNA, (ii) human mRNA encoding one of three recurrent arhinia-associated variants or (iii) human mRNA encoding an FSHD2-associated mis-sense variant (p.Pro690Ser)[16]. Full-length hWT *SMCHD1* mRNA, but none of the mutant mRNAs, rescued the terminal nerve phenotype (Fig. 5 and Supplementary Table 5). Complementation of mRNA with a common, presumably benign, variant from ExAC (p.Val708Ile; rs2270692) also rescued the phenotype, supporting assay specificity. The likely mode of action of the arhinia-associated alleles is therefore loss of function, and we found no foundational differences between the arhinia-specific mutations and FSHD2-associated alleles.

To extend these functional assays to a mammalian system, we introduced two variants, p.Leu141Phe and p.Glu136Asp, into mouse embryos using CRISPR/Cas9 editing. The 83 embryos recovered carried a range of variants including WT, homozygous knock-ins, homozygous knockouts, compound heterozygotes and complex deletions (Supplementary Table 6), but no embryos harbored heterozygous knock-in of the same allele observed in arhinia subjects. Examination of embryos at 13.5 d after conception (dpc) using optical projection tomography[31] revealed no morphological or growth anomalies (Supplementary Fig. 8). These results do not support a simple haploinsufficiency or a null mechanism in mammals but are consistent with previous *Smchd1* knockdown studies in mice that demonstrated no craniofacial phenotype[12,32]. These observations support the idea that mutation of a single copy of *SMCHD1* alone may be insufficient to induce pathology in mammals. DUX4, the gene product responsible for FSHD, is not conserved outside of Old World monkeys and higher primates so neither the mouse nor the zebrafish genome contains orthologous sequences. The partial phenocopy associated with loss of Smchd1 in the zebrafish would therefore appear to preclude developmental overexpression of DUX4 in the olfactory placode as a mechanism of arhinia.

Protein modeling

The structure of the N-terminal region of SMCHD1, where the constrained GHKL ATPase domain resides (amino acids 111–365), is unknown. However, the crystal structure of heat shock protein 90 (Hsp90), a yeast GHKL ATPase protein, is known (PDB 2CG9), and the SMCHD1 and Hsp90 ATPase domains are structurally similar[21]. We generated a structural model of the N-terminal region of SMCHD1 with Phyre2 (ref 33) (Fig. 6a) and found that the top ranking templates were Hsp90 structures (residues 115–573; strongest homology from residues 120–260). The structural model indicates that the arhinia-specific mutations tend to cluster on the protein surface, suggesting that these residues may be part of an interaction surface. This hypothesis is supported independently by sequence-based predictions of solvent accessibility, indicating that arhinia mutations tend to be exposed on the protein surface (Fig. 6b).

Human expression studies and protein abundance

To initially characterize arhinia-associated *SMCHD1* variants in humans, we measured SMCHD1 protein abundance and performed RNA-seq on lymphoblastoid cell lines (LCLs). We extracted protein from 23 total subjects (Supplementary Fig. 1) from 10 families: 10

subjects with arhinia harboring *SMCHD1* variants, 11 unaffected family members without *SMCHD1* mutations and 2 family members with a mutation in *SMCHD1* and anosmia or a hypoplastic nose (AH3 and AH5, respectively). SMCHD1 expression was generally similar in LCLs from all subjects, as assessed by immunoblot analysis using two different anti-Smchd1 antibodies (Bethyl A302–872A-M and Abcam ab122555; Supplementary Fig. 9). We performed RNA-seq on a subset of 10 cases and 10 controls (unaffected family members) from the protein analysis. After confirming all mutations in the expressed transcripts, we found that arhinia subjects showed a slight but not significant decrease in *SMCHD1* mRNA expression compared to controls (fold-change = 0.94, $P = 0.49$, permutated *t*-test) with no average difference in allelic expression of the missense variant compared to the reference allele ($P = 0.50$, paired *t*-test with $t = 0.71$, $P = 0.5$ (95% CI: -6.77–12.97)) (Supplementary Table 7). These results indicate normal message stability in arhinia subjects, at least in the available LCLs.

We next evaluated global patterns of differential gene expression between arhinia subjects and familial controls and, given the limited power of the data set for this rare condition, integrated these data with orthogonal ChIP and RNA-seq data from *Smchd1*-null mouse neural stem cells (NSCs)[13] to identify differentially expressed genes and repeat families with comparable alterations in both data sets (Supplementary Figs. 10 and 11, and Supplementary Table 8). These data revealed an enrichment of differential expression among downregulated (but not upregulated) genes in humans compared to mouse at nominal thresholds (downregulated gene enrichment $P = 0.015$, one-tailed Fisher's exact test; OR = 2.49; Supplementary Table 9). From these analyses emerged a high-confidence set of nine downregulated genes that were differentially expressed in both data sets. Pathway and geneset analyses of these genes revealed that, across all 6,067 human phenotypes evaluated in ToppGene enrichment[34], only one phenotype, 'depressed nasal tip', achieved statistical significance ($P = 6.9 \times 10^{-6}$; Supplementary Table 10). These results were driven by four genes: *DOK7*, *TGIF1*, *KDM6A* and *ICK*. Biallelic mutations in *DOK7* cause fetal akinesia deformation sequence (FADS; MIM208150), which may include depressed nasal bridge[35], though FADS is a deformation, whereas arhinia is a malformation. Heterozygous loss-of-function mutations in *TGIF1* cause holoprosencephaly-4 (MIM142946), which may include arhinia, microphthalmia and cleft palate[36]. Mutations in *KDM6A*, which encodes a histone demethylase and methyltransferase, cause Kabuki syndrome type 2 (refs. [37,38]) (MIM300867), and mutations in *ICK*, which encodes a protein kinase, cause endocrine-cerebroosteodysplasia[39] (ECO; MIM612651). Patients with Kabuki and ECO syndromes display characteristic facies that can include a wide or depressed nasal bridge and cleft palate. These data suggest that mutations in *SMCHD1* result in regulatory changes of genes implicated in craniofacial development, and these four genes are therefore rational mechanistic candidates for modifiers of the arhinia phenotype in the presence of *SMCHD1* mutations.

Predicting comorbid FSHD2

This study identified variants in the 5′ constrained region of *SMCHD1* that were associated with both FSHD2 (ref 18) and arhinia. This overlap included one identical variant (p.Gly137Glu), and tested hypomethylation signatures were largely indistinguishable

between arhinia and FSHD2 probands. To our knowledge, arhinia and FSHD2 have never been reported in the same individual. However, only a small subset of subjects with arhinia, an already rare condition, would be expected to harbor the requisite oligogenic architecture for FSHD2 at the D4Z4 locus, and those that do meet criteria may be too young to be symptomatic or may go undiagnosed because facial weakness might be overlooked in a patient who has undergone reconstructive craniofacial surgery. We addressed this question in subjects with available material in our cohort and identified two arhinia probands (A1 and E1) with *SMCHD1* mutations who met all four clinical criteria for susceptibility to FSHD2: (i) an SMCHD1 pathogenic variant (p.Asn139His and p.Leu141Phe), (ii) D4Z4 hypomethylation (bisulfite sequencing <25%), (iii) a permissive haplotype and (iv) an 11- to 28-D4Z4 repeat unit at the 4q array[16,22,40,41] (Supplementary Table 4). Phenotypic evaluation of A1 and E1 suggested that at least one subject had symptoms of FSHD2, suggesting yet another variant (p.Asn139His) shared by the disorders and implicating a common mechanism between disorders for these identical alleles. Five other arhinia subjects may be at risk for FSHD2 (on the basis of hypomethylation data), but material was not available for confirmatory clinical testing. Overall, these results suggest that at least two alterations (p.Gly137Glu and p.Asn139His), in the presence of a specific genetic background, can manifest as extremely divergent phenotypes.

DISCUSSION

We present genetic, genomic and functional evidence implicating *SMCHD1* as the predominant driver of arhinia in humans. Overall, 84% of subjects harbored a missense mutation in a constrained region encompassing the ATPase domain of *SMCHD1.* Through a large collaborative effort, we were able to combine data from 24% of subjects reported in the literature with data from 21 new subjects, facilitating a uniform evaluation of the clinical phenotype associated with this condition and revealing that most subjects with arhinia who could be assessed presented with the BAM triad. These analyses represent the first evidence of a genetic cause for this rare craniofacial malformation and its associated reproductive phenotype, suggesting a novel and complex role for *SMCHD1* in cranial NCC migration and/or craniofacial placode development.

These observations raise broader questions about the molecular mechanisms by which mutations in the same gene can produce distinct phenotypes. For *SMCHD1*, these include arhinia, BAM and FSHD2. Truncation variants are common in FSHD2, but missense variants have also been reported, whereas all arhinia-associated variants were missense alleles. This suggests that the mutant SMCHD1 protein must be synthesized in arhinia and that these mutations probably interfere with one or more critical protein functions. We found largely identical hypomethylation patterns at the 4q35 D4Z4 locus in arhinia and FSHD2 probands, indicating that neither a loss nor a gain of this particular function of SMCHD1 alone (gene silencing by methylation) explains the difference in phenotype. We also investigated the possibility that arhinia-associated *SMCHD1* mutations inflict genome-wide de-repression of repeat silencing but found no significant differences (Supplementary Fig. 10). Additional factors must therefore be involved in producing these distinct phenotypes, such as interactions with variants at other loci or disruption of SMCHD1 protein interactions that are critical to its epigenetic functions. Indeed, we found that the arhinia-specific variants

tend to cluster on the surface of the protein, potentially disrupting ATP or DNA binding and/or interfering with assembly of the SMCHD1 homodimer[21]. Correspondingly, results in peripheral tissue of arhinia probands and controls precluded models of simple haploinsufficiency or overexpression resulting from these mutations, further suggesting that the bioactivity of the protein, rather than the total amount of protein, is the critical defect in humans.

Distinct findings in two model systems reinforce the complexity suggested by our data in humans. In the zebrafish model, loss of function was sufficient to drive specific phenotypes that were rescued with full-length hWT *SMCHD1* mRNA but not with mRNA containing arhinia or FSHD2 variants, and overexpression conferred no discernible phenotype. In mouse, complete loss of function of *Smchd1* (homozygous exon 23 nonsense mutation) produces hypomethylation, which causes female-specific embryonic lethality[12]. Heterozygosity for this nonsense mutation likewise produces no phenotype in mice, nor did CRISPR/Cas9-mediated induction of homozygous or compound heterozygous deletions in the current study. Unfortunately, we were unable to replicate the heterozygous missense genotypes characteristic of human arhinia in the mouse (Supplementary Fig. 8). The fact that loss of Smchd1 produced distinctive phenotypes in zebrafish but not in mouse supports the importance of genetic background and functional interactions of the mutant protein in the pathogenesis of human arhinia.

The complex oligogenic architecture of FSHD2 suggests that only a small fraction of individuals with arhinia, an exceedingly rare condition on its own, will be at risk for FSHD2. Our analyses identified seven subjects potentially at risk for FSHD2, and at least one appears to be symptomatic. Notably, one-quarter of individuals who meet genetic criteria for FSHD2 are clinically asymptomatic, indicating that some of the factors that mediate FSHD2 are unknown[42]. The absence of arhinia in patients with FSHD2 who harbor *SMCHD1* mutations within the constrained ATPase domain argues that loss of SMCHD1 activity alone is insufficient to produce a craniofacial phenotype, comparably to mutations in *DNMT3B* in FSHD2 patients[43,44]. Indeed, the expressivity of this phenotype is clearly complex, as we observed family members harboring *SMCHD1* mutations with only mild dysmorphism or anosmia and at least one individual without dysmorphic features. Given the epigenetic function of SMCHD1, it is plausible that one or more interacting loci influence susceptibility to arhinia. Variants at these secondary loci need not be ultra-rare: TAR syndrome (MIM274000)[45] and SMAD6-associated craniosynostosis[46] are two examples of severe developmental disorders caused by the combination of a rare, deleterious, frequently *de novo* mutation with a common variant. Disentangling these genetic interactions and biochemical consequences of *SMCHD1* missense mutations will be a critical area of further study.

In conclusion, through an international network of collaborators we discovered that rare missense variants in an evolutionarily constrained region of *SMCHD1* are the predominant cause of isolated arhinia and BAM. Using multigenerational families, we found that phenotypes associated with these mutations can be subtle or even absent in family members. The distributions of these variants and their molecular mechanisms are not fully distinct from those in FSHD2, and the molecular pathways leading from SMCHD1 dysfunction to

either arhinia or FSHD2 remain unknown. Our findings thus emphasize yet another example in a growing list of genes in which mutations can give rise to pleiotropic phenotypes across the spectrum of human anomalies. For *SMCHD1*, these phenotypes—a rare muscle disease and a severe craniofacial and reproductive disorder—are notably diverse. Identifying the genetic modifiers influencing *SMCHD1*-related disease will significantly enhance understanding of the pathogenesis of the arhinia–BAM–FSHD spectra and placode biology and, more broadly, the architecture of oligogenic disorders.

URLs

cBioPortal Mutation Mapper, http://www.cbioportal.org/mutation_mapper.jsp; Ensembl, http://www.ensembl.org/; Leiden Open Variation Database, http://www.lovd.nl/3.0/; Mouse Genome Informatics, http://www.informatics.jax.org/; RefSeq, https://www.ncbi.nlm.nih.gov/refseq/; ExAC, http://exac.broadinstitute.org/; UCSC Genome Browser, https://genome.ucsc.edu/; CHOPCHOP, http://chopchop.cbu.uib.no/; OMIM, http://www.omim.org; Picard Tools (http://picard.sourceforge.net); BISMA (http://services.ibc.uni-stuttgart.de/BDPC/BISMA/).

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

ONLINE METHODS

Research subject enrollment

We collected existing DNA or blood samples from 40 subjects with arhinia (23 male, 17 female). Whenever possible, DNA was also collected from family members. Phenotypic information was obtained via questionnaires completed by patients, parents or referring physicians and confirmed by review of official medical records and consultation with the referring physician. Reproductive axis dysfunction could not be determined in pre-pubertal girls or in pre-pubertal boys without congenital microphallus or cryptorchidism. All research was approved by the Institutional Review Board of Partners Healthcare. Informed consent was obtained from all subjects, of whom a subset consented to the publication of photographs (Fig. 1).

WES

We performed WES on 26 total probands with arhinia (WES was performed after targeted sequencing did not detect an *SMCHD1* variant for four subjects) and 11 family members. The majority of participants ($n = 28$) were sequenced at the Broad Institute, including 21 independent probands and 1 set of affected siblings from a consanguineous family. We also sequenced 6 unaffected family members at the Broad Institute (families A, D and E; Supplementary Fig. 1). We collected data for two families that had previous WES from the University of Zurich (Zurich, Switzerland; one trio (family V) and a mother-proband pair (family U)), as well as a trio (family T) with WES performed by GeneDx in which the affected proband also had a deceased great aunt with arhinia and coloboma. We also

received exome results for a subject (AJ1) with arhinia from the Department of Human Genetics at Nagasaki University. WES data were aligned with BWA-MEM v.0.7.10 to GRCh37 and underwent joint variant calling by GATK[53] following best practices[54,55]. Familial relationships were confirmed by KING v1.4 (ref. 56), and variants were annotated with Annovar v.2016-02-01 (ref. 57) against the RefSeq annotation of the genome[58].

WGS

We obtained samples from four members of multigenerational family O[8,9] (Supplementary Fig. 1) and performed deep WGS to 30× average coverage on an Illumina X Ten platform. Family O had multiple individuals with craniofacial abnormalities beyond the proband's arhinia, including a deceased maternal half aunt with arhinia, a sister with arhinia, a mother with anosmia and subtle nasal and dental anomalies, and a maternal grandmother with mild nasal and dental anomalies. Samples from the affected sister, unaffected brother and unaffected maternal half-aunt were obtained after WGS had been completed and were screened for the p.Gln345Arg variant by targeted sequencing. Variants were aligned with BWA-MEM v.7.7 to GRCh37, and GATK was used to call single nucleotide variants (SNVs) as described above.

Targeted sequencing

SMCHD1 variants discovered by WES and WGS analyses were confirmed by Sanger sequencing (see Supplementary Table 11 for primers) in all subjects, except T1 and AJ1, for whom additional DNA was not available. Initial WES analyses identified rare mutations in *SMCHD1* restricted to exons 3, 8–10, 12 and 13. We therefore performed targeted sequencing of these exons in subsequent subjects ($n = 15$). Targeted screening found rare, missense *SMCHD1* variants in 10/15 subjects. Of the five *SMCHD1*-negative subjects, sufficient DNA was available for four to perform WES, which identified *SMCHD1* variants in adjacent exons for three subjects. Of interest, targeted sequencing identified one variant (p.Leu107Pro) in individual K1 that was missed by initial WES analyses.

Association analyses

We compared the burden of rare, nonsynonymous variants detected by WES in independent arhinia probands from an initial cohort with WES data ($n = 21$) from more than 60,706 controls in ExAC[10]. Analyses were restricted to include variants that passed the following criteria: (i) high quality (GATK Filter = PASS); (ii) rare (ExAC MAF < 0.1%); (iii) mean depth ≥ 10 reads; (iv) a mapping quality ≥ 10; (v) predicted to be nonsynonymous, to alter splicing, or to cause a frameshift. As there was no gender bias among our arhinia subjects to suggest sex-linkage (42.5% female), and we could not ascertain gender from the ExAC database, analyses were restricted to autosomes. Counts between ExAC and the arhinia cohort were compared by Fisher's exact test. Results were visualized as a Manhattan and QQ plot created by the R package qqman[59].

Inheritance testing

For samples with a predicted *de novo* variant from targeted sequencing, we confirmed familial relationships by determining repeat length of 10 STS markers (d15s205, d12s78,

d4s402, d13s170, d4s414, d22s283, d13s159, d2s337, d3s1267, d12s86). Expected parental relationships were confirmed in all families. Inheritance for proband (P1) was similarly confirmed at the University of Edinburgh with the following nine markers: cfstr1, d7s480, dxs1214, amel, nr2e3_22, d4s2366, i1cahd, d5s629, d5s823.

Transcriptome sequencing (RNA-seq)

Total RNA from ~1 million cells was extracted from EBV-transformed lymphoblastoid cell lines (LCLs) using TRIzol (Invitrogen) followed by RNeasy Mini Kit (Qiagen) column purification. We prepared 20 strand-specific RNA-seq libraries (10 cases, 10 controls; Supplementary Table 7) using the Illumina TruSeq kit and the manufacturer's instructions, as described[60,61]. Libraries were multiplexed, pooled and sequenced on multiple lanes of an Illumina HiSeq2500, generating an average of 46 million (46M) paired-end (PE) reads of 75 bp per sample, of which an average 44M PE reads passed vendor quality filters. Further quality control of sequence reads was assessed by fastQC (v.0.10.1). Subsequently, sequenced reads were N-trimmed and quality trimmed using cutadapt[62] (v.1.9.1) with options–trim-N,–quality-cutoff = 14,–minimum length = 70, which resulted in an average of 42.2M PE reads per subject. Post-trimming sequence reads were aligned to human reference genome Ensembl GRCh37 (v.71) using GSNAP[63] (19 December 2014 version) with options −N 1, −B 3,−quality-unk-mismatch = 1. Alignment quality was assessed by a custom script[60] using Picard Tools RNASeQC[64], RSeQC[65] and SamTools[66].

RNA-seq analyses assessed allele-specific expression for *SMCHD1* variant carriers ($n = 10$) using Samtools mpileup (v.1.2) to call variants, requiring base and mapping quality >30. We tested for differences between allele-specific expression patterns across subjects with *SMCHD1* mutations using paired *t*-tests. For differential expression between 10 arhinia cases and 10 unaffected familial controls, gene level counts were tabulated using BedTools's multibamcov algorithm[67] (v.2.17.0) on unique alignments for each library relying on Ensembl gene annotation[68] (GRCh37 v.71). 17,097 genes met the following filtering criteria: (i) > 7 uniquely mapped reads, (ii) gene transcript length ≥ 250 nucleotides, (iii) not classified as rRNA or tRNA. A two-sample permutation test from the perm package[69] (v.1.0) in R[70] (v.3.2.2) was applied to expression values of genes (counts per million), with the following options: alternative = two.sided, method = exact.ce, control = permControl tsmethod = abs. We compared all nominally ($P < 0.05$) differently expressed genes with previously published ChIP-seq and RNA-seq data of the *Smchd1*-null mouse[13]. A complete list of human-to-mouse homologs was retrieved from the Mouse Genome Informatics database[71] as of 9 August 2016. Enrichment of genes overlapping the RNA-seq and the ChIP-seq data was assessed with Fisher's exact test. Pathway analysis of overlapping gene lists was performed using ToppFun in the ToppGene Suite[34], in which genes are represented by unique Ensembl IDs.

Repeat element analysis (RNA-seq)

The GNSAP mapped reads from above were filtered for unique hits using the SAM flag NH:i:1. SAMtools (v0.1.19-44428cd) 'sort' was used to sort reads by coordinate and subsequently to remove potential PCR duplicates using 'rmdup'. The human (hg19) repeat-masker annotation was downloaded from the UCSC Genome Browser using 'Table browser'

and reshaped using custom R scripts (v3.1.0) to the format chr/start/stop/strand/Repeat_name/Repeat_class/Repeat_family/unique_identifier. BEDTools (v2.23.0) interesectBED was used to inner join mapped reads against the hg19 repeatmasker modified table in Unix. Next, the R 'table' function was implemented to count the frequency of unique ID repeat elements within the intersectBed outputs. Counts from controls and patients were concatenated into a master table and analyzed using the R packages edgeR[72] (v3.8.6) and DESeq2 (ref. 73) (v1.18.0). Finally, beeswarm (v0.2.3) was used to plot control and patient median repeat levels across annotated LTR, L1, L2, Satellite and SINE repeats using Student's *t*-test to test for significant differences between groups.

Immunoblotting

Protein was harvested from 1M LCLs in 23 total subjects: 10 subjects with arhinia and harboring *SMCHD1* mutations, 11 unaffected family members without *SMCHD1* mutations and two family members with a mutation in *SMCHD1* and anosmia or a hypoplastic nose (AH3 and AH5, respectively; Supplementary Fig. 9). Protein extraction was performed with the following procedure. Cells were washed in 1× PBS and lysed in 300 μl ice-cold 1× RIPA buffer (bioWorld) supplemented with 5 mM PMSF. After 30 min incubation on ice, cell lysates were cleared by centrifugation (15 × *g*, 15 min, 4 °C) and soluble protein concentration was assayed with a Pierce BCA protein assay kit. Extracted proteins (15–30 μl/sample) were next separated by 8% SDS-PAGE (Bio-Rad MiniProtean 3 Cell, 2 h 15 mA) and transferred onto a polyvinylidene fluoride (PVDF) membrane (Bio-Rad) using liquid transfer system (Bio-Rad Ready Gel Cell) at 4 °C, 10 V for 16 h. Immunoblotting was performed using two sets of antibodies to SMCHD1: Bethyl Laboratories A302-872A-M (anti-SMCHD1, C terminus) and Abcam ab122555 (anti-SMCHD1, N terminus). We used two loading control antibodies: Abcam ab6046 (β-tubulin loading control) and Abcam ab8227 (β-actin loading control). Antibody dilutions were used as recommended by manufacturer. Primary antibodies were diluted in TBS and Tween 20 (TBST) buffer and 1% BSA, secondary HRP-conjugated antibody (1:20,000 dilution) in TBST without BSA. The membrane was cut alongside 75 kDa marker (Bio-Rad Precision Plus Protein standards), and the upper part was used for SMCHD1 (molecular weight (MW) = 250 kDa), and the lower part for β-tubulin (MW = 50 kDa) and β-actin (MW = 42 kDa) controls. Blotting with primary antibody was carried out overnight at 4 °C on a rocking platform, followed by three 10-min washes in TBST at room temperature. Blotting with secondary antibody was carried out at room temperature for 1 h, followed by three 10-min washes in TBST. When re-blotting SMCHD1 with an alternative antibody, the previously used primary antibody was stripped off with mild stripping buffer, as recommended in Abcam's stripping-for-reprobing protocol. Blots were luminesced with ECL reagent (Bio-Rad) and developed with the ChemiDoc MP system (Bio-Rad). Automated protein quantification was done using Image Lab 5.2.1 software (Bio-Rad).

CRISPR/Cas9 genome editing in mouse embryos

To generate mouse embryos carrying the p.Leu141Phe disease-associated missense variant in Smchd1, a double-stranded DNA oligomer that provides a template for the guide RNA sequence was cloned into px461 (see Supplementary Table 11). The full gRNA template sequence was amplified from the resulting px461 clone using universal reverse primer and

T7 tagged forward primers. The guide RNA PCR template was used for *in vitro* RNA synthesis with T7 RNA polymerase (New England Biolabs), and the RNA template was subsequently purified using an RNeasy mini kit (Qiagen) purification columns. Cas9 mRNA was procured from Tebu Bioscience. The wild-type and mutant repair templates for Phe141/Leu141 (chr17:71,463,705–71,463,818 GRCm38) and Glu136/Asp136 (chr17:71,463,701–71,463,822 GRCm38) were synthesized as 114-bp ultramers bearing the desired sequence change from IDT. The injection mix contained Cas9 mRNA (50 ng/μl), guide RNA (25 ng/μl) and repair template DNA (150 ng/μl). Injections were performed in mouse zygotes, and the embryos were later harvested for analysis at 11.5 and 13.5 dpc.

Optical projection tomography

Whole mouse embryos were mounted in 1% agarose, dehydrated in methanol and then cleared overnight in BABB (1:2 benzyl alcohol:benzyl benzoate). The sample was then imaged using a Bioptonics OPT Scanner 3001 (Bioptonics) using tissue autofluorescence (excitation 425 nm, emission 475 nm) to capture the anatomy. The resulting images were reconstructed using Bioptonics proprietary software, automatically thresholded and merged to a single 3D image output using Bioptonics Viewer software.

DNA methylation analysis

The DNA methylation status of the D4Z4 region was assayed as described[22]. Bisulfite conversion was performed on 1 μg genomic DNA using the EpiTect Bisulfite Kit (Qiagen) per the manufacturer's instructions, and 200 ng of converted genomic DNA was used for PCR. Bisulfite sequencing (BSS) analysis of 52 CpGs in the *DUX4* promoter region of the 4q and 10q D4Z4 repeats was performed using primers BSS167F and BSS1036R followed by nested PCR with BSS475F and BSS1036R using 10% of the first PCR product (primer sequences are listed in Supplementary Table 11). PCR products were cloned into the pGEM-T Easy vector (Promega), sequenced, and analyzed using web-based analysis software BISMA (http://biochem.jacobs-university.de/BDPC/BISMA/)[74] with the default parameters. Standard genomic PCR was performed on nonconverted DNA to identify the 4qA, 4qA-L and 4qB chromosome[75]. Specific 4q and 10q haplotypes were identified and assigned as previously described[76,77]. The presence of the *DUX4* polyadenylation site was determined by BS-PCR as previously described[42].

Determination of 4q35 and 10q26 D4Z4 array sizes

Peripheral blood leukocytes or cultured lymphoblasts were embedded in agarose plugs and digested with three different restriction enzymes (EcoRI, EcoRI/BlnI and XapI). Restriction fragments were separated by pulsed-field gel electrophoresis (PFGE) and visualized by Southern blot with a p13E-11 probe. For some subjects, the Southern blot were rehybridized with a D4Z4 probe[78].

Gene suppression and *in vivo* complementation in zebrafish embryos

Splice-blocking morpholinos (MOs) targeting the *Danio rerio smchd1* exon 3 splice donor (e3i3) or exon 5 splice donor (e5i5) were designed and synthesized by Gene Tools LLC (Supplementary Table 11). To determine the optimal MO dose for *in vivo* complementation

studies, we injected increasing doses (3 ng, 6 ng, and 9 ng of each MO; 1 nl MO injected per embryo; 1- to 2-cell stage) into −*1.4col1a1:egfp*[79] embryos harvested from natural mating of heterozygous transgenic adults maintained on an AB background. To determine MO efficiency, we used TRIzol (Thermo Fisher) to extract total RNA from embryos at 1 dpf according to the manufacturer's instructions. Resulting total RNA was reverse transcribed into cDNA using the Superscript III Reverse Transcriptase kit (Thermo Fisher) and used as template in RT-PCR reactions to amplify regions flanking MO target sites. RT-PCR products were gel purified using the QIAquick gel extraction kit (Qiagen), cloned (TOPO-TA; Invitrogen), and plasmid purified from individual colonies was Sanger sequenced according to standard protocols to identify the precise alteration of endogenous transcript. For rescue experiments, a wild-type (WT) human *SMCHD1* ORF (GenBank NM_015295) construct was obtained commercially (OriGene Technologies) and subcloned into the pCS2+ vector. Point mutations were introduced into pCS2+ vectors as described[80], and all vectors were sequence confirmed. WT and variant *SMCHD1* constructs were linearized with NotI, and mRNA was transcribed using the mMessage mMachine SP6 transcription kit (Thermo Fisher). Unless otherwise noted, 9 ng MO (either e3i3 or e5i5) was used in parallel or in combination with 25 pg *SMCHD1* mRNA for *in vivo* complementation studies.

CRISPR/Cas9 genome editing in zebrafish embryos

We used CHOPCHOP[81] to identify a guide (g)RNA targeting a sequence within the *smchd1* coding regions (see Supplementary Table 11). gRNAs were *in vitro* transcribed using the GeneArt precision gRNA synthesis kit (Thermo Fisher) according to the manufacturer's instructions. Zebrafish embryos were obtained from −*1.4col1a1: egfp* embryos harvested from natural mating of heterozygous transgenic adults maintained on an AB background; 1 nl of injection cocktail containing 100 pg/nl gRNA and 200 pg/nl Cas9 protein (PNA Bio) was injected into the cell of embryos at the 1-cell stage. To determine targeting efficiency in founder (F0) mutants, we extracted genomic DNA from 2 dpf embryos and PCR amplified the region flanking the gRNA target site. PCR products were denatured, reannealed slowly and separated on a 15% TBE 1.0-mm precast polyacrylamide gel, which was then incubated in ethidium bromide and imaged on a ChemiDoc system (Bio-Rad) to visualize hetero- and homoduplexes. To estimate the percentage of mosaicism of *smchd1* F0 mutants ($n = 5$), PCR products were gel purified (Qiagen), and cloned into a TOPO-TA vector (Thermo Fisher). Plasmid was prepped from individual colonies ($n = 10–12$ colonies/embryo) and Sanger sequenced according to standard procedures.

Phenotypic analyses in zebrafish

To study craniofacial structures (cartilage or eye development), larval batches were reared at 28 °C and imaged live at 3 dpf using the Vertebrate Automated Screening Technology Bioimager (VAST; software version 1.2.2.8; Union Biometrica) mounted on an AxioScope A1 (Zeiss) microscope using an Axiocam 503 monochromatic camera and Zen Pro 2012 software (Zeiss). Fluorescence imaging of GFP-positive cells on ventrally positioned larvae was conducted as described[82]. In parallel, we obtained lateral brightfield images of whole larvae using the VAST onboard camera. To evaluate gonadotropin-releasing hormone (GnRH) neurons, 1.5 dpf embryos were dechorionated and fixed in a solution of 4% PFA and 7% picric acid for 2 h at room temperature. Embryos were then washed with a solution

of PBS with 0.1% Triton X-100 (PBS-T) and stored at 4 °C until staining. For whole-mount immunostaining, embryos were washed briefly with 0.1% trypsin in PBS, washed in PBS-T and dehydrated at −20 °C in pre-chilled 100% acetone for 15 min. Next, embryos were washed in PBS-T and blocked in a solution of 2% BSA, 1% DMSO, 0.5% Triton X-100 and 5% calf serum for 1 h at room temperature. We used rabbit anti-GnRH antibody (1:500 dilution; Sigma-Aldrich, G8294) for primary detection. After overnight incubation of primary antibody, embryos were washed with blocking solution and incubated with Alexa Fluor 555 anti-rabbit secondary antibody (1:500; Thermo Fisher) for 2 h at room temperature. Images were acquired manually with an AxioZoom. V16 microscope and Axiocam 503 monochromatic camera and were *z*-stacked using Zen Pro 2012 software (Zeiss). Cartilage structure, eye area, and GnRH neuron projection length was measured using ImageJ (NIH); pairwise comparisons to determine statistical significance were calculated using Student's *t*-test. For ceratobranchial pair counts, we used a χ^2 test to determine statistical significance. All experiments were repeated at least twice.

Statistics and general methods

For zebrafish studies, sample size was based on prior experiments to evaluate similar phenotypes[25,82]. For mouse studies, given that we were looking for a dichotomous effect, we set up the CRISPR experiment with both wild-type and mutant repair template, which is our standard practice for modeling heterozygous variants. The gRNA was very efficient in targeting the locus, so >50 embryos was chosen as a reasonable number. For mouse studies, we scored the phenotype visually prior to genotyping. For zebrafish studies, injections and scoring were accomplished with the investigator blinded to the experimental condition. The phenotype was scored on mouse embryo dissection prior to genotyping. None of the embryos was thought to be craniofacially abnormal. After genotyping, we chose the genotypes to compare and performed optical projection tomography (OPT), which showed no difference.

Data availability

Sequencing data are available under dbGaP accession phs001246.v1.p1. *SMCHD1* variant information has been deposited at ClinVar under accessions SCV000328594–SCV000328618.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Natalie D Shaw[1,2,47], Harrison Brand[1,3,4,5,47], Zachary A Kupchinsky[6], Hemant Bengani[7], Lacey Plummer[1], Takako I Jones[8], Serkan Erdin[3,5], Kathleen A Williamson[7], Joe Rainger[7], Alexei Stortchevoi[3], Kaitlin Samocha[5,9], Benjamin B Currall[3], Donncha S Dunican[7], Ryan L Collins[3,10], Jason R Willer[6], Angela Lek[11], Monkol Lek[5,9], Malik Nassan[12], Shahrin Pereira[13], Tammy Kammin[13], Diane Lucente[3], Alexandra Silva[3], Catarina M Seabra[3,14], Colby Chiang[3], Yu An[3], Morad

Ansari[7], Jacqueline K Rainger[7], Shelagh Joss[15], Jill Clayton Smith[16], Margaret F Lippincott[1], Sylvia S Singh[1], Nirav Patel[1], Jenny W Jing[1], Jennifer R Law[17], Nalton Ferraro[18], Alain Verloes[19], Anita Rauch[20], Katharina Steindl[20], Markus Zweier[20], Ianina Scheer[21], Daisuke Sato[22], Nobuhiko Okamoto[23], Christina Jacobsen[24], Jeanie Tryggestad[25], Steven Chernausek[25], Lisa A Schimmenti[26], Benjamin Brasseur[27], Claudia Cesaretti[28], Jose E García-Ortiz[29], Tatiana Pineda Buitrago[30], Orlando Perez Silva[31], Jodi D Hoffman[32], Wolfgang Mühlbauer[33], Klaus W Ruprecht[34], Bart L Loeys[35], Masato Shino[36], Angela M Kaindl[37], Chie-Hee Cho[38], Cynthia C Morton[5,13], Richard R Meehan[7], Veronica van Heyningen[7], Eric C Liao[39,40,41], Ravikumar Balasubramanian[1], Janet E Hall[1,2], Stephanie B Seminara[1], Daniel Macarthur[5,9,42], Steven A Moore[43], Koh-ichiro Yoshiura[44], James F Gusella[3,4,5,11], Joseph A Marsh[7], John M Graham Jr[45], Angela E Lin[46], Nicholas Katsanis[6], Peter L Jones[8], William F Crowley Jr[1], Erica E Davis[6], David R FitzPatrick[7], and Michael E Talkowski[3,4,5,42]

Affiliations

[1]Harvard Reproductive Endocrine Sciences Center and NICHD Center of Excellence in Translational Research in Fertility and Infertility, Reproductive Endocrine Unit of the Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA [2]National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA [3]Molecular Neurogenetics Unit and Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA [4]Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA [5]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA [6]Center for Human Disease Modeling, Duke University Medical Center, Durham, North Carolina, USA [7]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh Western General Hospital, Edinburgh, UK [8]Department of Cell and Developmental Biology, University of Massachusetts Medical School, Worcester, Massachusetts, USA [9]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA [10]Program in Bioinformatics and Integrative Genomics, Division of Medical Sciences, Harvard Medical School, Boston, Massachusetts, USA [11]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA [12]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota, USA [13]Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Boston, Massachusetts, USA [14]GABBA Program, University of Porto, Porto, Portugal [15]West of Scotland Genetics Service, South Glasgow University Hospitals, Glasgow, UK [16]Faculty of Medical and Human Sciences, Institute of Human Development, Manchester Centre for Genomic Medicine, University of Manchester, Manchester Academic Health Science Centre (MAHSC), Manchester, UK [17]Division of Pediatric Endocrinology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA [18]Department of Oral and Maxillofacial Surgery, Boston Children's Hospital, Boston,

Massachusetts, USA [19]Department of Genetics, Robert Debré Hospital, Paris, France [20]Institute of Medical Genetics and Radiz–Rare Disease Initiative Zurich, Clinical Research Priority Program for Rare Diseases, University of Zurich, Schlieren-Zurich, Switzerland [21]Department of Diagnostic Imaging, Children's Hospital, Zurich, Switzerland [22]Department of Pediatrics, Hokkaido University Graduate School of Medicine, Sapporo, Japan [23]Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan [24]Division of Endocrinology and Genetics, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts, USA [25]Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA [26]Departments of Otorhinolaryngology and Clinical Genomics, Mayo Clinic, Rochester, Minnesota, USA [27]DeWitt Daughtry Family Department of Surgery, University of Miami Leonard M. Miller School of Medicine, Miami, Florida, USA [28]Medical Genetics Unit, Fondazione IRCCS Cà Granda, Ospedale Maggiore Policlinico, Milan, Italy [29]División de Genética, Centro de Investigación Biomédica de Occidente, Instituto Mexicano del Seguro Social, Guadalajara, Mexico [30]Fundación Hospital Infantil Universitario de San José, Bogotá, Colombia [31]Academia Nacional de Medicina de Colombia, Bogotá, Colombia [32]Divisions of Genetics and Maternal Fetal Medicine, Tufts Medical Center, Boston, Massachusetts, USA [33]Department of Plastic and Aesthetic Surgery, ATOS Klinik, Munich, Germany [34]Department of Ophthalmology, University Hospital of the Saarland, Homburg, Germany [35]Center for Medical Genetics, University of Antwerp and Antwerp University Hospital, Antwerp, Belgium [36]Department of Otolaryngology and Head and Neck Surgery, Gunma University Graduate School of Medicine, Gunma, Japan [37]Biology and Neurobiology, Charité–University Medicine Berlin and Berlin Institute of Health, Berlin, Germany [38]Department of Diagnostic, Interventional and Pediatric Radiology, Inselspital, University Hospital of Bern, Bern, Switzerland [39]Center for Regenerative Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA [40]Division of Plastic and Reconstructive Surgery, Massachusetts General Hospital, Boston, Massachusetts, USA [41]Harvard Stem Cell Institute, Cambridge, Massachusetts, USA [42]Center for Mendelian Genomics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA [43]Department of Pathology, University of Iowa Carver College of Medicine, Iowa City, Iowa, USA [44]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan [45]Department of Pediatrics, Cedars Sinai Medical Center, Los Angeles, California, USA [46]Medical Genetics, MassGeneral Hospital for Children and Harvard Medical School, Boston, Massachusetts, USA

Acknowledgments

We thank all participants, family members and clinical staff for their generous contributions of time and materials to this research. We thank T. Gillis, J. Ruliera, C. Hanscom, C. Antolik and M. Anderson for technical assistance. This project was funded by grants from the National Institutes of Health ((NIH) R00MH095867 and R01HD081256 to M.E.T.; P01GM061354 to M.E.T., J.F.G., C.C.M. and E.C.L.; T32HD007396 to H. Brand; P50HD028138 to W.F.C., S.B.S., M.E.T., N.K. and E.E.D.; R01HD043341 and MGH Robert and Laura Reynolds Research Scholar

Award to S.B.S.; K23HD073304-02 and 1SI2ES025429-01 to N.D.S.; P50DK096415 to N.K. and R01AR062587 to P.L.J.); the March of Dimes (FY15-255 to M.E.T.); the Medical Research Council (MR/M02122X/1 to J.A.M.); the German Research Foundation (SFB665 to A.M.K.) and the Berlin Institute of Health (BIH-CRG1 to A.M.K.). D.R.F., R.R.M. (MC_PC_U127574433), D.S.D., H. Bengani, K.A.W., J.R., J.K.R. and J.A.M. are funded by program grants from the Medical Research Council (MRC) Human Genetics Unit award to the University of Edinburgh. M.A. is funded by the University of Edinburgh Institute of Genomics and Molecular Medicine Translational Initiative Fund. S.A.M. is supported by U54-NS053672, which funds the Iowa, Paul D. Wellstone Muscular Dystrophy Cooperative Research Center. N.K. is supported as a Distinguished Jean and George Brumley Professor at Duke University, and M.E.T. is supported as the Desmond and Ann Heathwood MGH Research Scholar.

References

1. Bosma JF, Henkin RI, Christiansen RL, Herdt JR. Hypoplasia of the nose and eyes, hyposmia, hypogeusia, and hypogonadotrophic hypogonadism in two males. J Craniofac Genet Dev Biol. 1981; 1:153–184. [PubMed: 6802865]
2. Hogan BL, et al. Small eyes (*Sey*): a homozygous lethal mutation on chromosome 2 which affects the differentiation of both lens and nasal placodes in the mouse. J Embryol Exp Morphol. 1986; 97:95–110. [PubMed: 3794606]
3. Glaser T, et al. PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. Nat Genet. 1994; 7:463–471. [PubMed: 7951315]
4. Schmidt-Sidor B, et al. Malformations of the brain in two fetuses with a compound heterozygosity for two *PAX6* mutations. Folia Neuropathol. 2009; 47:372–382. [PubMed: 20054790]
5. Solomon BD, et al. Compound heterozygosity for mutations in *PAX6* in a patient with complex brain anomaly, neonatal diabetes mellitus, and microophthalmia. Am J Med Genet A. 2009; 149A: 2543–2546. [PubMed: 19876904]
6. Gordon, CT., et al. *De novo* mutations in *SMCHD1* cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. Nat Genet. 2016. http://dx.doi.org/10.1038/ng.3765
7. Parhar IS. Cell migration and evolutionary signifcance of GnRH subtypes. Prog Brain Res. 2002; 141:3–17. [PubMed: 12508557]
8. Cho CH, Shakibaei M, Merker HJ, Klein M. The rare malformation of nasal aplasia. Mund Kiefer Gesichtschir. 2006; 10:106–117. [PubMed: 16496112]
9. Thiele H, Musil A, Nagel F, Majewski F. Familial arhinia, choanal atresia, and microphthalmia. Am J Med Genet. 1996; 63:310–313. [PubMed: 8723126]
10. Lek M, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536:285–291. [PubMed: 27535533]
11. Samocha KE, et al. A framework for the interpretation of *de novo* mutation in human disease. Nat Genet. 2014; 46:944–950. [PubMed: 25086666]
12. Blewitt ME, et al. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat Genet. 2008; 40:663–669. [PubMed: 18425126]
13. Chen K, et al. Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. Proc Natl Acad Sci USA. 2015; 112:E3535–E3544. [PubMed: 26091879]
14. Gendrel AV, et al. Epigenetic functions of Smchd1 repress gene clusters on the inactive X chromosome and on autosomes. Mol Cell Biol. 2013; 33:3150–3165. [PubMed: 23754746]
15. Mould AW, et al. Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. Epigenetics Chromatin. 2013; 6:19. [PubMed: 23819640]
16. Lemmers RJ, et al. Digenic inheritance of an *SMCHD1* mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet. 2012; 44:1370–1374. [PubMed: 23143600]
17. Hirano T. At the heart of the chromosome: SMC proteins in action. Nat Rev Mol Cell Biol. 2006; 7:311–322. [PubMed: 16633335]
18. Lemmers RJ, et al. Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Hum Mol Genet. 2015; 24:659–669. [PubMed: 25256356]

19. van den Boogaard ML, et al. Double *SMCHD1* variants in FSHD2: the synergistic effect of two *SMCHD1* variants on D4Z4 hypomethylation and disease penetrance in FSHD2. Eur J Hum Genet. 2016; 24:78–85. [PubMed: 25782668]
20. Lemmers RJ, et al. Hemizygosity for *SMCHD1* in facioscapulohumeral muscular dystrophy type 2: consequences for 18p deletion syndrome. Hum Mutat. 2015; 36:679–683. [PubMed: 25820463]
21. Chen K, et al. The epigenetic regulator Smchd1 contains a functional GHKL-type ATPase domain. Biochem J. 2016; 473:1733–1744. [PubMed: 27059856]
22. Jones TI, et al. Identifying diagnostic DNA methylation profles for facioscapulohumeral muscular dystrophy in blood and saliva using bisulfte sequencing. Clin Epigenetics. 2014; 6:23. [PubMed: 25400706]
23. Jones TI, et al. Individual epigenetic status of the pathogenic D4Z4 macrosatellite correlates with disease in facioscapulohumeral muscular dystrophy. Clin Epigenetics. 2015; 7:37. [PubMed: 25904990]
24. Liu C, et al. A secreted WNT-ligand-binding domain of FZD5 generated by a frameshift mutation causes autosomal dominant coloboma. Hum Mol Genet. 2016; 25:1382–1391. [PubMed: 26908622]
25. Chassaing N, et al. Targeted resequencing identifes *PTCH1* as a major contributor to ocular developmental anomalies and extends the SOX2 regulatory network. Genome Res. 2016; 26:474–485. [PubMed: 26893459]
26. Yahyavi M, et al. ALDH1A3 loss of function causes bilateral anophthalmia/microphthalmia and hypoplasia of the optic nerve and optic chiasm. Hum Mol Genet. 2013; 22:3250–3258. [PubMed: 23591992]
27. Steven C, et al. Molecular characterization of the GnRH system in zebrafish (*Danio rerio*): cloning of chicken GnRH-II, adult brain expression patterns and pituitary content of salmon GnRH and chicken GnRH-II. Gen Comp Endocrinol. 2003; 133:27–37. [PubMed: 12899844]
28. Whitlock KE, Illing N, Brideau NJ, Smith KM, Twomey S. Development of GnRH cells: setting the stage for puberty. Mol Cell Endocrinol. 2006:254–255.
29. Zohar Y, Muñoz-Cueto JA, Elizur A, Kah O. Neuroendocrinology of reproduction in teleost fish. Gen Comp Endocrinol. 2010; 165:438–455. [PubMed: 19393655]
30. Abraham E, Palevitch O, Gothilf Y, Zohar Y. The zebrafish as a model system for forebrain GnRH neuronal development. Gen Comp Endocrinol. 2009; 164:151–160. [PubMed: 19523393]
31. Sharpe J, et al. Optical projection tomography as a tool for 3D microscopy and gene expression studies. Science. 2002; 296:541–545. [PubMed: 11964482]
32. Blewitt ME, et al. An *N*-ethyl-*N*-nitrosourea screen for genes involved in variegation in the mouse. Proc Natl Acad Sci USA. 2005; 102:7629–7634. [PubMed: 15890782]
33. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015; 10:845–858. [PubMed: 25950237]
34. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 2009; 37:W305–W311. [PubMed: 19465376]
35. Hall JG. Pena–Shokeir phenotype (fetal akinesia deformation sequence) revisited. Birth Defects Res A Clin Mol Teratol. 2009; 85:677–694. [PubMed: 19645055]
36. Solomon, BD., Gropman, A., Muenke, M. Holoprosencephaly overview. GeneReviews. https://www.ncbi.nlm.nih.gov/books/NBK1530/ (updated 29 August 2013)
37. Lederer D, et al. Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. Am J Hum Genet. 2012; 90:119–124. [PubMed: 22197486]
38. Lindgren AM, et al. Haploinsuffciency of *KDM6A* is associated with severe psychomotor retardation, global growth restriction, seizures and cleft palate. Hum Genet. 2013; 132:537–552. [PubMed: 23354975]
39. Lahiry P, et al. A multiplex human syndrome implicates a key role for intestinal cell kinase in development of central nervous, skeletal, and endocrine systems. Am J Hum Genet. 2009; 84:134–147. [PubMed: 19185282]
40. de Greef JC, et al. Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology. 2010; 75:1548–1554. [PubMed: 20975055]

41. van Deutekom JC, et al. Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counselling and etiology of FSHD1. Hum Mol Genet. 1996; 5:1997–2003. [PubMed: 8968754]
42. Calandra P, et al. Allele-specific DNA hypomethylation characterises FSHD1 and FSHD2. J Med Genet. 2016; 53:348–355. [PubMed: 26831754]
43. van den Boogaard ML, et al. Mutations in *DNMT3B* modify epigenetic repression of the D4Z4 repeat and the penetrance of facioscapulohumeral dystrophy. Am J Hum Genet. 2016; 98:1020–1029. [PubMed: 27153398]
44. Weemaes CM, et al. Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. Eur J Hum Genet. 2013; 21:1219–1225. [PubMed: 23486536]
45. Albers CA, et al. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit *RBM8A* causes TAR syndrome. Nat Genet. 2012; 44:435–439. [PubMed: 22366785]
46. Timberlake AT, et al. Two locus inheritance of non-syndromic midline craniosynostosis via rare *SMAD6* and common *BMP2* alleles. eLife. 2016; 5:e20125. [PubMed: 27606499]
47. Fokkema IF, et al. LOVD v. 2.0: the next generation in gene variant databases. Hum Mutat. 2011; 32:557–563. [PubMed: 21520333]
48. Cerami E, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012; 2:401–404. [PubMed: 22588877]
49. Gao J, et al. Integrative analysis of complex cancer genomics and clinical profles using the cBioPortal. Sci Signal. 2013; 6:pl1. [PubMed: 23550210]
50. Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct Biol. 2009; 9:51. [PubMed: 19646261]
51. Yang J, et al. The I-TASSER Suite: protein structure and function prediction. Nat Methods. 2015; 12:7–8. [PubMed: 25549265]
52. Heffernan R, et al. Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. Sci Rep. 2015; 5:11476. [PubMed: 26098304]
53. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–1303. [PubMed: 20644199]
54. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–498. [PubMed: 21478889]
55. Van der Auwera GA, et al. From FastQ data to high confdence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013; 43:11.10.1–11.10.33. [PubMed: 25431634]
56. Manichaikul A, et al. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010; 26:2867–2873. [PubMed: 20926424]
57. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38:e164. [PubMed: 20601685]
58. O'Leary NA, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016; 44:D733–D745. [PubMed: 26553804]
59. Turner, SD. qqman: an R package for visualizing GWAS results using Q–Q and Manhattan plots. Preprint at bioRxiv. 2014. http://dx.doi.org/10.1101/005165
60. Blumenthal I, et al. Transcriptional consequences of 16p11.2 deletion and duplication in mouse cortex and multiplex autism families. Am J Hum Genet. 2014; 94:870–883. [PubMed: 24906019]
61. Sugathan A, et al. CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. Proc Natl Acad Sci USA. 2014; 111:E4468–E4477. [PubMed: 25294932]
62. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011; 17:10–12.

63. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics. 2010; 26:873–881. [PubMed: 20147302]
64. DeLuca DS, et al. RNA-SeQC: RNA–seq metrics for quality control and process optimization. Bioinformatics. 2012; 28:1530–1532. [PubMed: 22539670]
65. Wang L, Wang S, Li W. RSeQC: quality control of RNA–seq experiments. Bioinformatics. 2012; 28:2184–2185. [PubMed: 22743226]
66. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]
67. Quinlan AR, Hall IM. BEDTools: a fexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
68. Aken BL, et al. The Ensembl gene annotation system. Database (Oxford). 2016; 2016:baw093. [PubMed: 27337980]
69. Fay MP, Shaw PA. Exact and asymptotic weighted logrank tests for interval censored data: the interval R package. J Stat Softw. 2010; 36:i02. [PubMed: 25285054]
70. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2016.
71. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE. The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. Nucleic Acids Res. 2015; 43:D726–D736. [PubMed: 25348401]
72. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. [PubMed: 19910308]
73. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA–seq data with DESeq2. Genome Biol. 2014; 15:550. [PubMed: 25516281]
74. Rohde C, Zhang Y, Reinhardt R, Jeltsch A. BISMA—fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. BMC Bioinformatics. 2010; 11:230. [PubMed: 20459626]
75. Lemmers RJ, et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. Science. 2010; 329:1650–1653. [PubMed: 20724583]
76. Lemmers RJ, et al. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. Am J Hum Genet. 2007; 81:884–894. [PubMed: 17924332]
77. Lemmers RJ, et al. Worldwide population analysis of the 4q and 10q subtelomeres identifes only four discrete interchromosomal sequence transfers in human evolution. Am J Hum Genet. 2010; 86:364–377. [PubMed: 20206332]
78. Lemmers RJ, O'Shea S, Padberg GW, Lunt PW, van der Maarel SM. Best practice guidelines on genetic diagnostics of facioscapulohumeral muscular dystrophy: workshop 9th June 2010, LUMC, Leiden, the Netherlands. Neuromuscul Disord. 2012; 22:463–470. [PubMed: 22177830]
79. Kague E, et al. Skeletogenic fate of zebrafish cranial and trunk neural crest. PLoS One. 2012; 7:e47394. [PubMed: 23155370]
80. Niederriter AR, et al. In vivo modeling of the morbid human genome using *Danio rerio*. J Vis Exp. 2013; 78:e50338.
81. Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Res. 2016; 44:W272–W276. [PubMed: 27185894]
82. Isrie M, et al. Mutations in either *TUBB* or *MAPRE2* cause circumferential skin creases Kunze type. Am J Hum Genet. 2015; 97:790–800. [PubMed: 26637975]

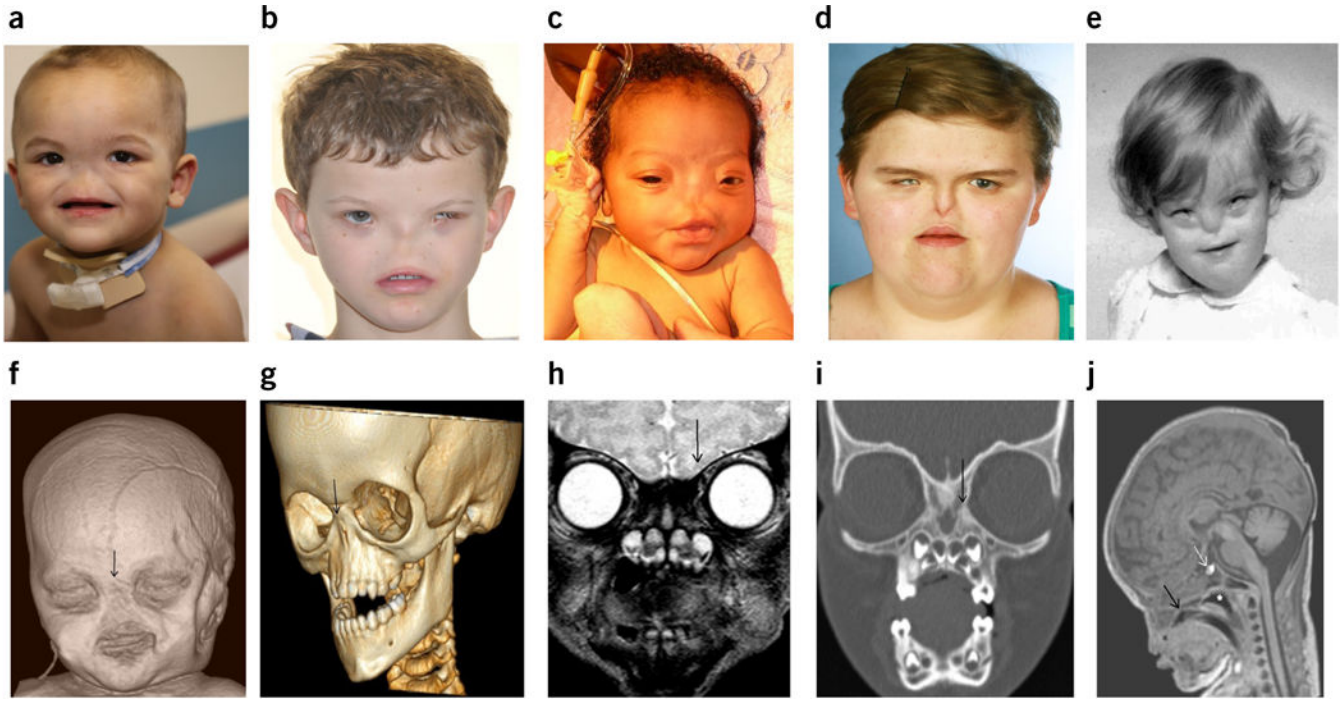

Figure 1.

Phenotypic spectra associated with arhinia. (**a–e**) Five representative subjects demonstrate complete arhinia and variable ocular phenotypes: Subject V1 (age 2) with left-sided iris coloboma (**a**), subject AC1 (age 10) with left-sided microphthalmia and bilateral nasolacrimal duct stenosis (**b**), subject U1 (as a newborn) with normal eye anatomy and vision (**c**), subject 04 (age 16) with right-sided microphthalmia (**d**) and subject A1 (age 1) with bilateral colobomatous microphthalmia, cataracts and nasolacrimal duct atresia (**e**). (f–j) Craniofacial radiographic images from subject V1: surface rendering reconstruction from a MRI 3D T1 weighted sequence showing complete absence of the nose (arrow) (**f**), 3D volume rendering technique (VRT) reconstruction from spiral CT showing complete absence of nasal bones (arrow) (**g**), coronal reconstruction from CT showing absence of nasal septal structures (**h**; the maxilla articulates with the nasal process of the frontal bone (arrow)), coronal MRI T2 weighted sequence showing absence of the olfactory bulb and olfactory sulcus (arrow) (**i**) and midline MRI sagittal T1 weighted sequence (**j**) showing a high-arched palate (cleft not visible on this image) and decreased distance between the oral cavity and the anterior cranial fossa (black arrow). The rudimentary nasopharynx (**j**, asterisk) is blind and air filled. The pituitary gland (**j**, white arrow) appears normal.

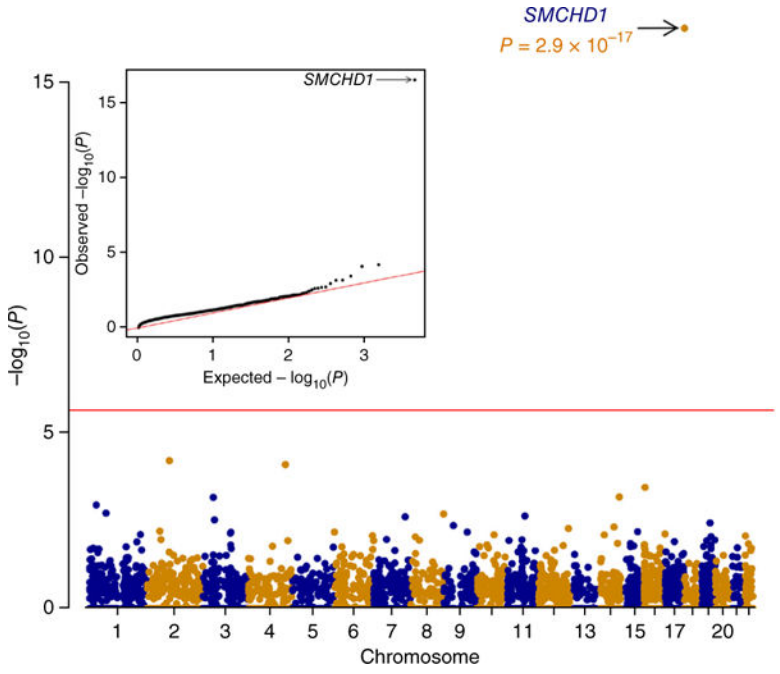

Figure 2.

Association analyses for rare mutation burden in arhinia Manhattan plot and quantile-quantile (q-q) plot demonstrating the significant accumulation of rare *SMCHD1* mutations in subjects with arhinia compared to the ExAC cohort ($P = 2.9 \times 10^{-17}$, Fisher's exact test; OR = 34.4 (95 CI: 18.8–57.9)). Analyses established a variant count at each gene for arhinia subjects compared to ExAC controls (who presumably do not have arhinia; $n = 60{,}706$) after filtering for allele frequency (MAF < 0.1%), quality (mean depth ≥10; mapping quality ≥10) and predicted function (nonsynonymous, splice site and frameshift mutations). Any gene with at least one mutation passing these criteria was included in the analysis ($n = 22{,}445$ genes). Genome-wide significance threshold was $P < 2.2 \times 10^{-6}$ after Bonferroni correction (red line), and only *SMCHD1* exceeded this threshold.

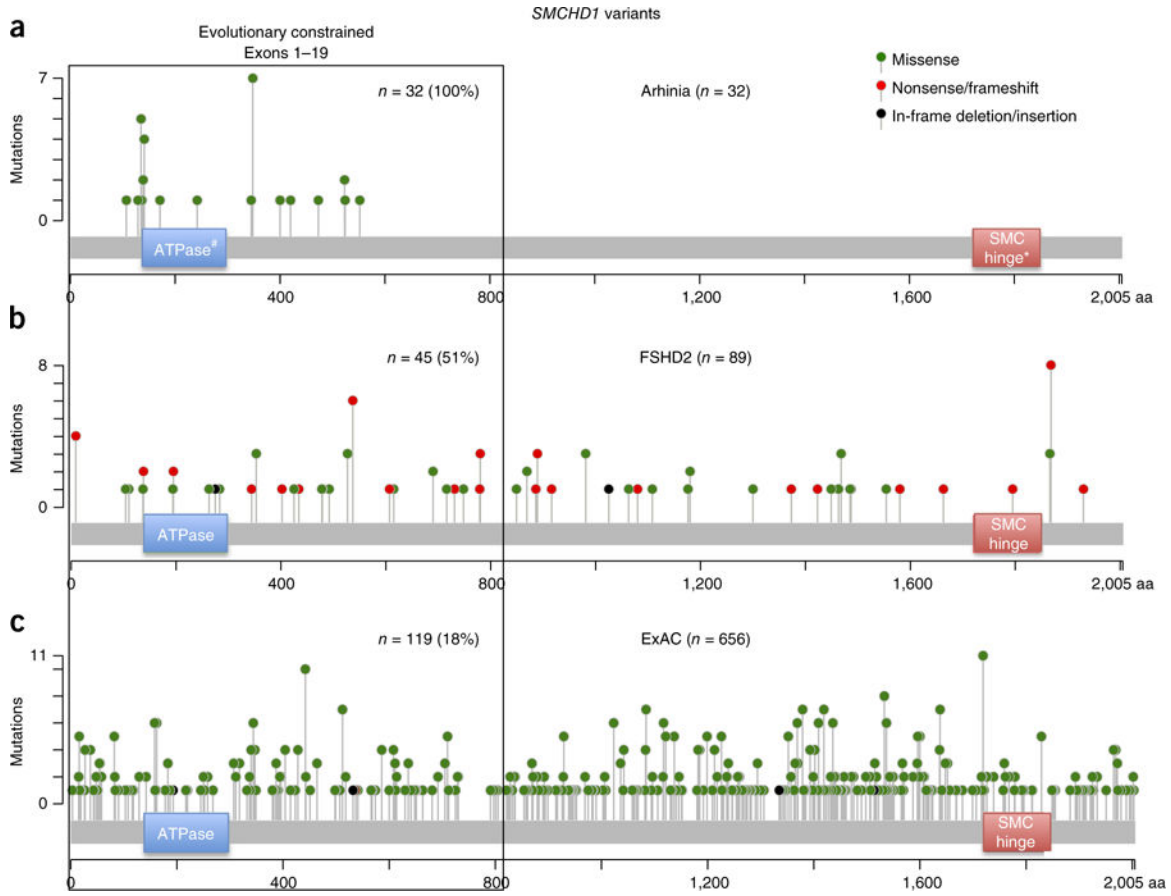

Figure 3.

Arhinia-associated mutations occur near the 5′ GHKL-type ATPase domain. (**a–c**) The distribution of arhinia-associated mutations across (**a**) *SMCHD1* is tightly clustered between exons 3 and 12 of the gene compared to the distribution of variants observed in *FSHD2* subjects (**b**) and ExAC controls (**c**). *FSHD2* variants were taken from the Leiden Open Variation Database (LOVD 3.0)[47]. Constraint analysis revealed that the gene displays slight overall intolerance to missense mutations ($P = 0.016$), but this significance is driven by regional constraint across the first 19 exons of *SMCHD1* (black box; $\chi^2 = 37.73$; $P = 8.12 \times 10^{-10}$), which includes the GHKL-type ATPase domain (ATPase), whereas the region from exons 20–48 is not constrained ($\chi^2 = 0.87$; $P = 0.35$). SMC hinge, structural maintenance of chromosomes flexible hinge domain. Figures were modified from the cBioPortal Mutation Mapper software v1.0.1 (refs. [48,49]).

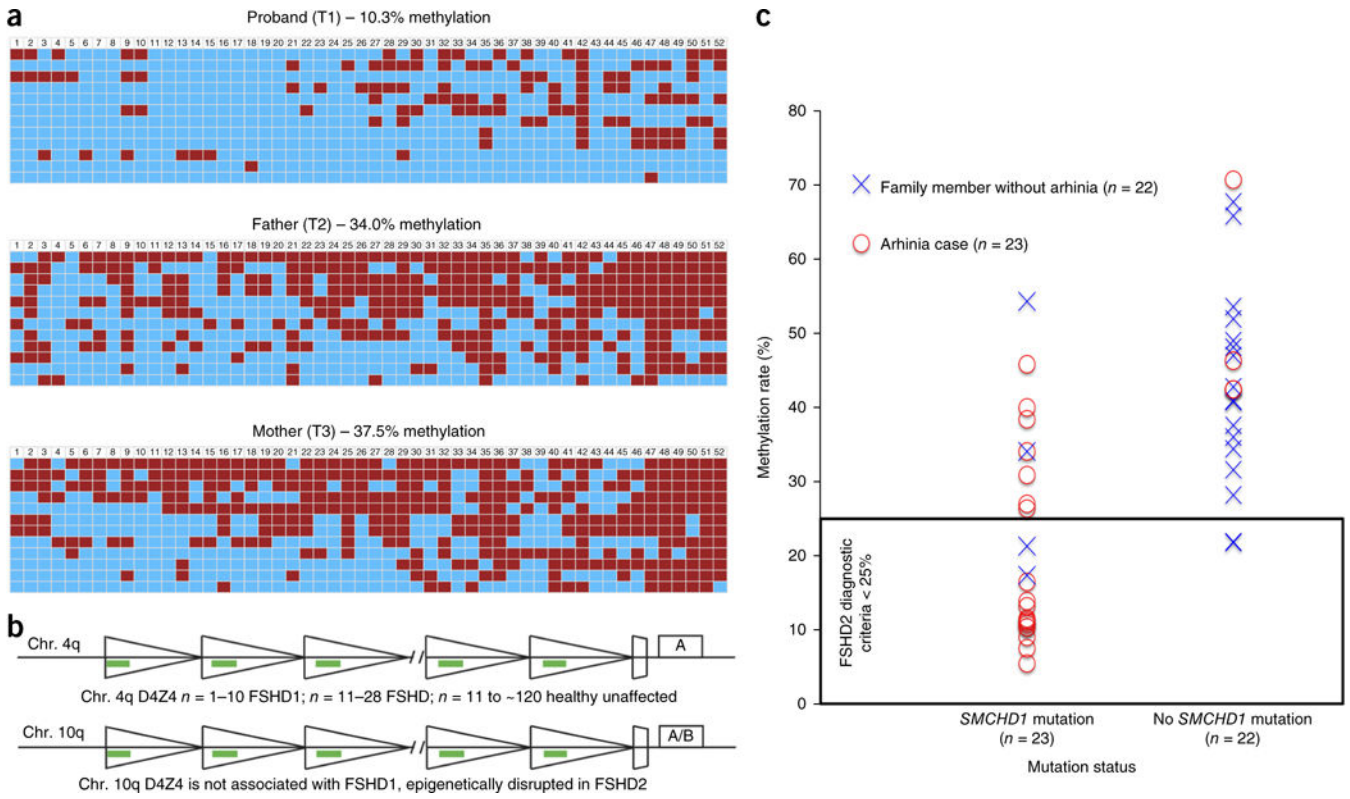

Figure 4.

DNA methylation analysis of D4Z4 repeats. (**a**) Bisulfite sequencing (BSS) analysis of DNA hypomethylation at chromosome 4q and 10q D4Z4 repeats. DNA hypomethylation was consistent with dominant *SMCHD1* hypomorphic mutations found in FSHD2 patients. A total of 52 CpGs were analyzed, arranged linearly from left to right, for 12 clones arranged top to bottom, each representing an independent chromosome analyzed. Each predicted CpG is represented by a box, with red boxes indicating methylated CpGs and blue boxes indicating unmethylated CpGs. (**b**) Cartoon of the chromosome (chr.) 4q and 10q D4Z4 macrosatellites that vary in repeat units (RU) from 1 to ~120 RUs. The region analyzed by BSS in each RU is indicated by a green bar. FSHD2 requires a mutation in *SMCHD1* combined with at least one chromosome 4q D4Z4 array ranging in size from 11 to 28 RUs and a permissive A-type 4q sub-telomere. (**c**) Methylation rate determined by BSS in arhinia probands with *SMCHD1* mutations for which material was available for analysis. 74% had D4Z4 hypomethylation characteristic of FSHD2, whereas the single proband tested without a *SMCHD1* mutation showed a normal methylation pattern. BSS was measured from the lowest quartile as previously described[22], and a methylation rate of <25% was considered consistent with hypomethylation observed in FSHD2. See supplementary table 4 for further details on individual methylation status.

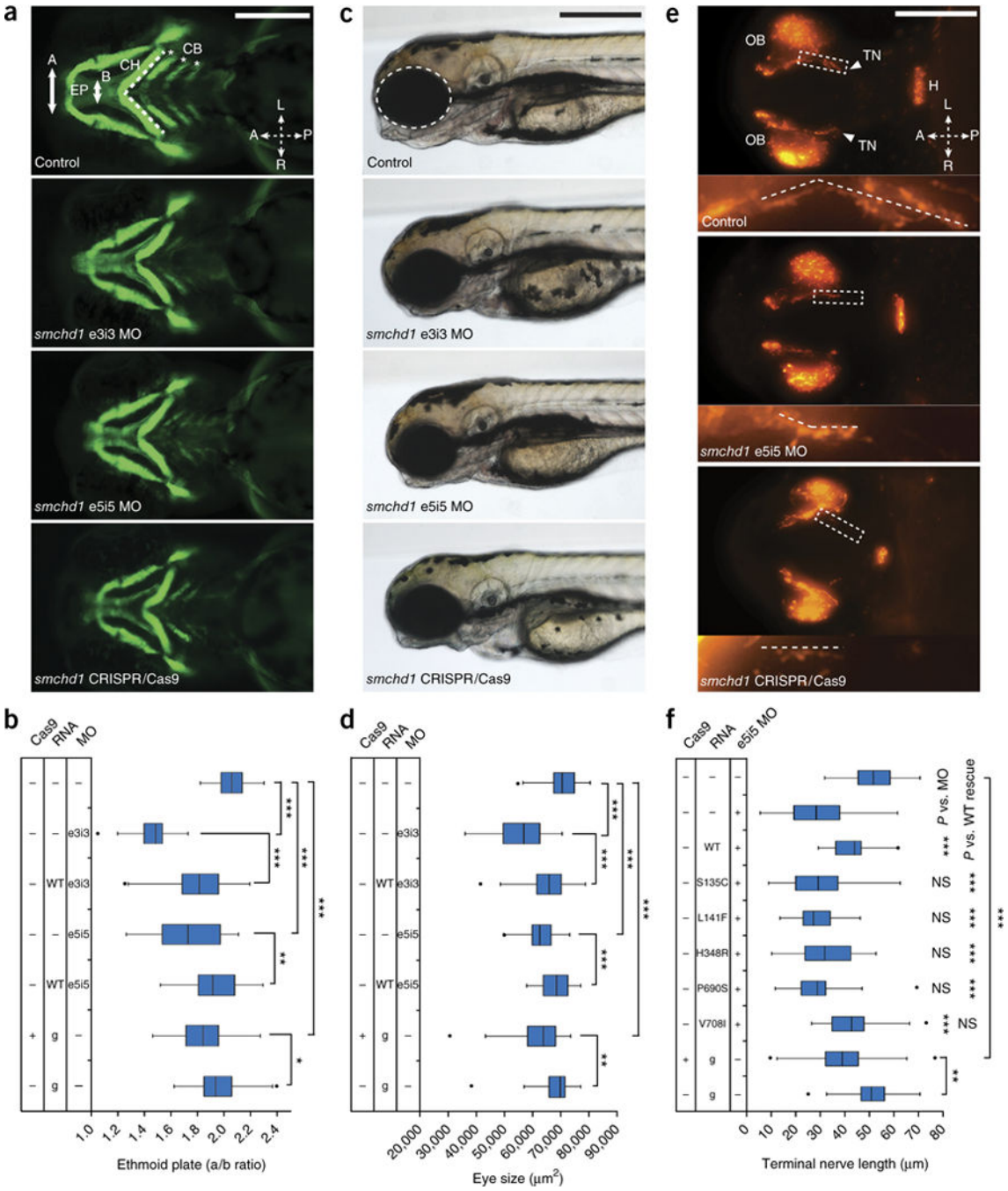

Figure 5.

In vivo modeling of *smchd1* in zebrafish. (**a**) Suppression of *smchd1* results in altered cartilage structures in −*1.4col1a1:egfp* zebrafish larvae at 3 dpf. Representative ventral images of *smchd1* morphants (e5i5 MO and e3i3 MO) and F0 mutant larvae (CRISPR/Cas9) display smaller ethmoid plates (EP, white arrow); a broadened ceratohyal angle (CH, dashed white line) and fewer ceratobranchial arches (CB, asterisks). A, furthest distal width; B, width at the ethmoid plate-trabecula junction. Orientation arrows indicate anterior (A), posterior (P), left (L) and right (R). Scale bar, 200 μm. (**b**) Ethmoid plate width measured on

ventral images in **a**. The furthest distal width (A) was normalized to the width at the ethmoid plate-trabecula junction (B). g, guide RNA. (**c**) Loss of *smchd1* results in smaller eyes (dashed white circle); lateral bright-field images of representative 3 dpf larvae are shown. Scale bar, 300 μm. (**d**) Eye size (area) in larval batches from c. (**e**) Representative immunostaining of GnRH neurons in 1.5 dpf embryos with a pan-GnRH antibody shows shorter terminal nerve (TN, arrowheads and dashed white lines) projection from the olfactory bulb (OB) in *smchd1* models. Ventral views are shown. H, hypothalamus; scale bar, 100 μm; dashed white lines in insets highlight TN projections. (**f**) Complementation assay of missense SMCHD1 variants using GnRH TN length as a readout. p.Ser135Cys (S135C), p.Leu141Phe (L141F) and p.His348Arg (H348R) are recurrent variants in arhinia cases; p.Pro690Ser (P690S) is associated with FSHD2 (ref. 18), and p.Val708Ile (V708I; rs2276092) is a common variant in ExAC. NS, not significant. All experiments were repeated at least twice, with masked scoring; controls were uninjected embryos from the same clutch. See supplementary table 5 for details about embryo numbers and statistical comparisons.

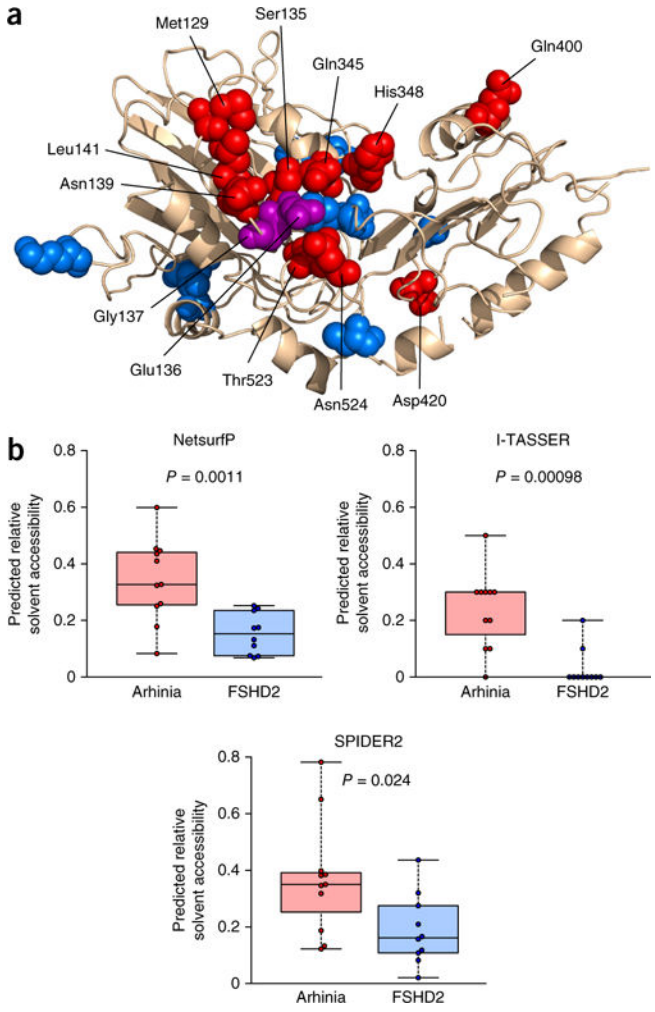

Figure 6.

SMCHD1 protein modeling predicts that arhinia-associated alterations are more likely to occur on the surface of SMCHD1 and disrupt a binding surface than are FSHD2-associated variations. (**a**) Homology model of the N-terminal region of SMCHD1 generated with Phyre2 (ref. 33) with residues altered in arhinia (red; $n = 11$) and FSHD2 (blue; $n = 10$). All of the top 20 structural templates had GHKL domains: 16 were Hsp90 structures, 2 were mismatch repair proteins (MutL, Mlh1) and 2 were type II topoisomerases. Only the residues modeled with high confidence are shown (115–295; 314–439; 458–491; 504–535; 552–573). (**b**) Comparison of predicted relative solvent accessibility values for residues in the N-terminal region of SMCHD1 altered in arhinia and FSHD2. Three predictive methods were used: NetsurfP[50], I-TASSER[51] and SPIDER2 (ref. 52). Residues altered in both disorders (136–137) were excluded from this analysis. *P* values were calculated with the Wilcoxon rank-sum test. Box boundaries indicate interquartile range; center line, median; whiskers, entire range of the distribution.

Table 1

SMCHD1 mutations observed in the arhinia cohort

Chr.	Nucleotide mutation	Exon	Inheritance (sample ID)	Number of subjects	Sample ID[a]	Amino acid alteration	Sex (sample ID)
18	g.2666926T>C	3	N/A	1	K1	p.Leu107Pro	F
18	g.2666992T>A	3	N/A	1	D1	p.Met129Lys	M
18	g.2667009A>T	3	*De novo* (AF1) N/A (M1)	2	M1, AF1	p.Ser135Cys	F (M1, AF1)
18	g.2667010G>A	3	*De novo* (I1) N/A[b] (R1)	2	I1, R1	p.Ser135Asn	F (R1), M (I1)
18	g.2667010G>T	3	*De novo*	1	AK1	p.Ser135Ile	M
18	g.2667014A>C	3	Father[b]	1	T1	p.Glu136Asp	M
18	g.2667016G>A	3	N/A	1	AG1	p.Gly137Glu	F
18	g.2667021A>C	3	*De novo* (A1) N/A (Y1)	2	A1, Y1	p.Asn139His	F (A1, Y1)
18	g.2667029G>C	3	N/A	3	C1, E1, S1	p.Leu141Phe	F (S1), M (C1, E1)
18	g.2667029G>T	3	*De novo*	1	V1	p.Leu141Phe	M
18	g.2674017T>G	5	N/A[b]	1	AB1	p.Phe171Val	M
18	g.2688478C>G	6	*De novo*	1	AA1	p.Ala242Gly	M
18	g.2694685A>G	8	Mother[b]	2	O1, O4[c]	p.Gln345Arg	F (O1, O4)
18	g.2697032A>G	9	*De novo* (X1, AC1, AE1) N/A (F1, L1, N1, Z1)	7	F1, L1, N1, Z1, X1, AC1, AE1	p.His348Arg	F (L1, X1), M(F1, N1, Z1, AC1, AE1)
18	g.2697896A>T	10	Father[b]	1	AH1	p.Gln400Leu	F
18	g.2697956A>T	10	*De novo*	1	P1	p.Asp420Val	M
18	g.2700611G>C	11	N/A	1	W1	p.Glu473Gln	M
18	g.2700837C>A	12	N/A	2	J1, U1	p.Thr523Lys	F (U1), M (J1)
18	g.2700840A>G	12	N/A	1	B1	p.Asn524Ser	M
18	g.2703697G>A	13	N/A	1	AJ1	p.Arg552Gln	M

[a]Samples L1, M1, N1, P1, AF1 and AJ1 overlap with those studied in Gordon *et al*.[6].

[b]Multiplex family.

[c]Sibling. Subjects G1, H1, H2, Q1, AD1, AI and AL1 did not show a rare missense mutation in *SMCHD1*. N/A, parental samples not available; M, male; F, female.