Please cite this article in press as: Kherraf et al., A Homozygous Ancestral SVA-Insertion-Mediated Deletion in *WDR66* Induces Multiple Morphological Abnormalities o..., The American Journal of Human Genetics (2018), https://doi.org/10.1016/j.ajhg.2018.07.014

ARTICLE

A Homozygous Ancestral SVA-Insertion-Mediated Deletion in *WDR66* Induces Multiple Morphological Abnormalities of the Sperm Flagellum and Male Infertility

Zine-Eddine Kherraf,[1,2] Amir Amiri-Yekta,[1,2,3] Denis Dacheux,[4,5] Thomas Karaouzène,[1] Charles Coutton,[1,6] Marie Christou-Kent,[1] Guillaume Martinez,[1,6] Nicolas Landrein,[4] Pauline Le Tanno,[1] Selima Fourati Ben Mustapha,[7] Lazhar Halouani,[7] Ouafi Marrakchi,[7] Mounir Makni,[7] Habib Latrous,[7] Mahmoud Kharouf,[7] Karin Pernet-Gallay,[8] Hamid Gourabi,[3] Derrick R. Robinson,[4] Serge Crouzy,[10] Michael Blum,[9] Nicolas Thierry-Mieg,[9] Aminata Touré,[11,12,13] Raoudha Zouari,[7] Christophe Arnoult,[1] Mélanie Bonhivers,[4,14] and Pierre F. Ray[1,2,14,*]

Multiple morphological abnormalities of the sperm flagellum (MMAF) is a severe form of male infertility defined by the presence of a mosaic of anomalies, including short, bent, curled, thick, or absent flagella, resulting from a severe disorganization of the axoneme and of the peri-axonemal structures. Mutations in *DNAH1*, *CFAP43*, and *CFAP44*, three genes encoding axoneme-related proteins, have been described to account for approximately 30% of the MMAF cases reported so far. Here, we searched for pathological copy-number variants in whole-exome sequencing data from a cohort of 78 MMAF-affected subjects to identify additional genes associated with MMAF. In 7 of 78 affected individuals, we identified a homozygous deletion that removes the two penultimate exons of *WDR66* (also named *CFAP251*), a gene coding for an axonemal protein preferentially localized in the testis and described to localize to the calmodulin- and spoke-associated complex at the base of radial spoke 3. Sequence analysis of the breakpoint region revealed in all deleted subjects the presence of a single chimeric SVA (SINE-VNTR-*Alu*) at the breakpoint site, suggesting that the initial deletion event was potentially mediated by an SVA insertion-recombination mechanism. Study of *Trypanosoma* WDR66's ortholog (TbWDR66) highlighted high sequence and structural analogy with the human protein and confirmed axonemal localization of the protein. Reproduction of the human deletion in *TbWDR66* impaired flagellar movement, thus confirming *WDR66* as a gene associated with the MMAF phenotype and highlighting the importance of the WDR66 C-terminal region.

Introduction

Male infertility is a common reproductive disorder characterized by an extremely heterogeneous pathogeny mostly due to altered sperm production during spermatogenesis. Spermatogenesis is a complex process involving three successive steps: the mitotic proliferation of spermatogonial stem cells, the meiotic division of spermatocytes, and finally spermiogenesis, a key post-meiotic event contributing to major morphological changes to the spermatids, such as chromatin compaction, acrosome biogenesis, and assembly of flagellar structures.[1] Transcriptomic analyses of germ cells has revealed the dynamic transcription of over 4,000 genes during human spermatogenesis,[2] but the genetic basis of male infertility has so far not been extensively studied.[3] Monomorphic sperm defects, such as globozoospermia, sperm macrocephaly, headless spermatozoa, and multiple morphological abnormalities of the sperm flagella (MMAF), have been demonstrated to be caused by genetic defects affecting several testis-specific genes.[4] For affected men, intracytoplasmic sperm injection represents a unique option for conception, but significant efforts should be made to characterize the genetic causes behind their sperm defects and to better understand the nature of their molecular alterations.

In this study, we focused on subjects presenting a typical MMAF phenotype characterized by the presence of immotile sperm with short, bent, coiled, absent flagellum or flagellum with an irregular width (Figure 1). In 2014, our team first identified pathogenic mutations in dynein axonemal heavy chain 1 (*DNAH1* [MIM: 603332]) in 28% of the MMAF individuals analyzed.[5] Recently, whole-exome sequencing (WES) of 78 MMAF individuals allowed us to identify bi-allelic mutations in *CFAP43* (previously known as *WDR96* [MIM: 617558]) and *CFAP44* (previously known as *WDR52* [MIM: 617559]) encoding

[1]Genetic Epigenetic and Therapies of Infertility, Institute for Advanced Biosciences, INSERM U1209, CNRS UMR 5309, Université Grenoble Alpes, Grenoble 38000, France; [2]Centre Hospitalier Universitaire de Grenoble, UM GI-DPI, Grenoble 38000, France; [3]Department of Genetics, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, Academic Center for Education, Culture, and Research, PO Box 16635-148, Tehran, Iran; [4]University Bordeaux, Microbiologie Fondamentale et Pathogénicité, CNRS UMR 5234, Bordeaux, France; [5]Institut Polytechnique de Bordeaux, Microbiologie Fondamentale et Pathogénicité, CNRS UMR 5234, Bordeaux, France; [6]Centre Hospitalier Universitaire de Grenoble, UM de Génétique Chromosomique, Grenoble 38000, France; [7]Polyclinique les Jasmins, Centre d'Aide Médicale à la Procréation, Centre Urbain Nord, 1003 Tunis, Tunisia; [8]Grenoble Neuroscience Institute, INSERM 1216, Grenoble 38000, France; [9]Université Grenoble Alpes, CNRS, TIMC-IMAG, Grenoble 38000, France; [10]Laboratoire de Chimie et Biologie des Métaux, Institut de Recherche en Technologie et Sciences pour le Vivant, CEA iRTSV/LCBM/GMCT, CNRS UMR 5249, Université Grenoble Alpes, Grenoble 38054, Cedex 9, France; [11]INSERM U1016, Institut Cochin, Paris 75014, France; [12]CNRS UMR 8104, Paris 75014, France; [13]Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine, Paris 75014, France
[14]These authors contributed equally to this work
*Correspondence: pray@chu-grenoble.fr
https://doi.org/10.1016/j.ajhg.2018.07.014.
© 2018 American Society of Human Genetics.

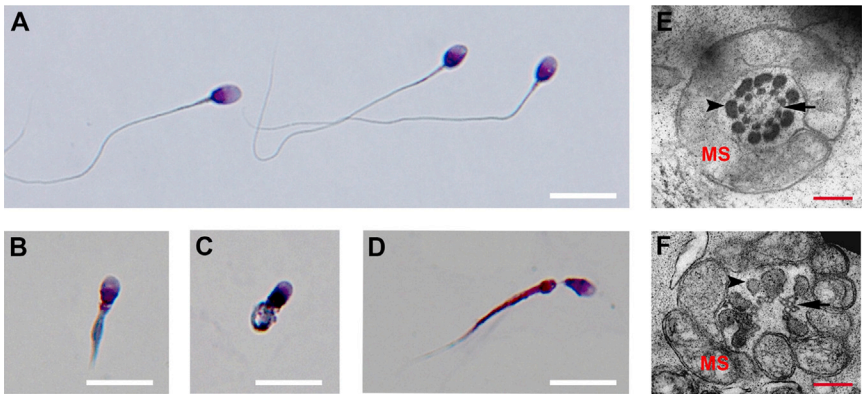

Figure 1. Defective Formation of Sperm Flagella in Individuals with *WDR66* Mutations
(A–D) Light-microscopy analysis of sperm from control (A) and affected (B–D) individuals reveals MMAF in the individuals. Scale bars, 10 μm.
(E and F) TEM micrographs of cross-sections through sperm cell midpieces. (E) Control sperm show a homogeneous ring-like mitochondrial sheath (MS) surrounding nine outer dense fibers (ODFs, arrowheads) and the central axoneme, characterized by the typical presence of nine microtubule doublets (MDs, arrows) and two central microtubule singlets (MSs). (F) Representative sperm from an affected man show a partitioned MS, abnormally positioned and assembled ODFs and MDs, and the absence of the MS. Scale bars, 200 nm.

tryptophan-aspartate (WD)-repeat proteins of unknown functions highly expressed in the testis.[6] Mutations in these genes have also been reported by others[7,8] and account for 20.5% of the MMAF individuals of our cohort.[3] Other rarer homozygous missense mutations have been reported in single familial cases of MMAF in *CEP135* (MIM: 611423)[9] and *AK7* (MIM: 615364),[10] encoding for a centriolar protein and an axonemal adenylate kinase, respectively. These results highlight the genetic heterogeneity associated with this phenotype and, importantly, they indicate that additional genes are likely to be associated with MMAF.

Copy-number variations (CNVs), structural variants defined as large insertions, deletions, or duplications of genomic segments of more than 1 kb,[11] are important elements of human genetic diversity in healthy people[12–14] but are also increasingly recognized as an important etiology of many human diseases, including neuro-developmental diseases, chronic inflammations, and cancers.[15] WES is a cost-effective technique permitting the detection of pathogenic mutations. Its efficiency at identifying single-nucleotide variants (SNVs) and small indels has been proven countless times. It also has recently been used for detecting CNVs,[16,17] but its performance and limitations for detecting CNVs are still undefined. We reanalyzed the WES data from our previously described cohort of 78 sequenced individuals[6] by focusing on CNV detection using ExomeDepth software[18] and identified a homozygous 8.4 kb intragenic deletion encompassing *WDR66* (also known as *CFAP251* [cilia- and flagella-associated protein 251 (MIM: 612573)]) exons 20 and 21 (out of 22) in seven individuals. *WDR66* encodes a WD-repeat protein preferentially localized in the testis. We characterized the extent of the deletion at the nucleotide level and concluded that the deletion was caused by an SVA-insertion-mediated deletion (SIMD) mechanism.[19] Using the *Trypanosoma brucei* model, we confirmed the importance of the deleted region including *WDR66*. Thus, we demonstrate that, in addition to the MMAF-associated genes *CFAP43* and *CFAP44*, *WDR66* is also associated with MMAF and confirm the importance of WD-repeat proteins in human diseases and especially in male infertility.

Material and Methods

Study Participants

We included 78 subjects who sought treatment for isolated primary infertility between 2008 and 2016 and presented with asthenozoospermia due to a combination of morphological defects of the sperm flagella as follows: absent, short, bent, and/or coiled flagella and/or flagella of an irregular width. Forty-six individuals were of North African origin and consulted the Clinique des Jasmin in Tunis for primary infertility. Ten individuals originated from the Middle East (Iran) and were treated in Tehran at the Royan Institute (Reproductive Biomedicine Research Center) for primary infertility, and 22 individuals were recruited in France (21 at the Cochin Hospital and one at Lille's hospital). Saliva and/or peripheral blood were obtained for all participants. During their medical consultation, all subjects answered a health questionnaire focused on manifestations of primary ciliary dyskinesia (PCD). All subjects had an isolated MMAF phenotype and no mention of any other PCD symptoms. However, ciliary function could not be examined (in nasal or lung cilia), and we cannot exclude the possibility that some of the studied individuals could have a sub-clinical alteration of ciliary function. The detailed description of the cohort can be found in the report by Coutton et al.[6] Signed informed consent was obtained from all the subjects participating in the study according to the local protocols and the principles of the Declaration of Helsinki. The collection was declared to the French Ministry of Health (DC-2015-2580) and the French Data Protection Authority (DR-2016-392).

Study of Sperm Morphology, Spermograms, and Transmission electron Microscopy

Sperm morphology was first analyzed on slides under a microscope after modified Papanicolaou staining. Sperm analysis was carried out in the source laboratories during the course of the routine biological examination of the subjects according to World Health Organization (WHO) guidelines.[20] Then, transmission electron microscopy (TEM) assessment was carried out. Sperm cells from an affected individual with a *WDR66* deletion were fixed in PBS with 4% paraformaldehyde (PFA) for 60 min. They were

pelleted (1,000 × *g* for 10 min) and resuspended in fixation buffer (2.5% glutaraldehyde, 2% PFA, 100 mM phosphate buffer [pH 7.4], and 50 mM sucrose) for 2 hr. Fixed cells were washed in buffer for 10 min, post-fixed in 1% OsO_4 for 1 hr, and washed three times in water, and then samples were stained in 2% uranyl acetate in water at 4°C overnight. Samples were next washed in water, dehydrated in ethanol, embedded in Epoxy resin, and polymerized for 60 hr at 60°C. Sections were stained with 2% uranyl and lead citrate and visualized in a transmission electron microscope (JEOL 1200EX). Trypanosomes were prepared for TEM as above with the exception that they were embedded in Spur resin.

WES and Data Processing

Genomic DNA was isolated from saliva with the Oragen DNA Extraction Kit (DNA Genotech). Coding regions and intron-exon boundaries were enriched with the All Exon V5 Kit (Agilent Technologies). WES and initial data analyses were performed as described previously.[6] CNVs, which had not been analyzed previously, were analyzed with the ExomeDepth software package.[18] We retained only overlapping CNVs found in at least two subjects and not described as benign in Cathagenia. All data are based on GRCh37 (UCSC Genome Browser hg19).

RT-PCR and qRT-PCR

qRT-PCR was performed on a cDNA panel from testis, lung, kidney, and heart tissues (Clontech) with a StepOnePlus Real-Time PCR System (Life Technologies) using Power SYBR Green PCR Master Mix (Life Technologies) according to the manufacturer's protocol. Quantification of the fold change in gene expression was determined by the relative quantification (RQ) method ($2^{-\Delta\Delta CT}$) with *GAPDH* as a reference. Data are shown as the mean RQ ± standard deviation. Primers are described in Table S1.

Immunofluorescence and Confocal Microscopy on Human Sperm Cells

Immunostaining was carried out on samples from individuals P7 and P9. Sperm was fixed in PBS and 4% PFA for 7 min. After being washed in PBS, the sperm suspension was spotted onto slides precoated with 0.1% poly-L-lysine (Thermo Fisher Scientific). After attachment, sperm was permeabilized with 0.1% (v/v) Triton X-100-DPBS (Triton X-100, Sigma-Aldrich) for 15 min at room temperature (RT). Slides were then blocked in 5% normal serum DPBS (normal goat or donkey serum; GIBCO, Invitrogen) and incubated overnight at 4°C with primary antibodies. Polyclonal rabbit anti-WDR66 antibody was purchased from Sigma-Aldrich and diluted at 1:50. Monoclonal mouse anti-acetylated α-tubulin antibody was purchased from Sigma-Aldrich and diluted at 1:2,500. The slides were then washed and incubated with secondary antibody diluted at 1:800 (Dylight-549-conjugated goat anti-mouse immunoglobulin G [IgG] and Dylight-488-conjugated goat anti-rabbit IgG, Jackson ImmunoResearch) and Hoechst 33342 (Sigma-Aldrich) for 1 hr at RT, rinsed, and mounted with Dako Fluorescence Mounting Medium (Life Technology). Images were acquired on a Zeiss LSM 710 confocal microscope and processed with Zen 2009 software. The list of antibodies used in this study is provided in Table S2.

Breakpoint Investigation

We conducted a walking PCR approach to get closer to the deleted fragment. We performed long-range PCR to amplify the CNV breakpoint junctions and analyzed the product by Sanger sequencing to refine the breakpoints. To investigate the potential mechanism causing the deletion, we aligned the sequences around the breakpoints with each other by using the Basic Local Alignment Search Tool (BLAST) and then analyzed them for different repetitive elements by using RepeatMasker software. PCR primers and protocols used for each individual are listed in Table S3. Sequencing reactions were carried out with BigDye Terminator v.3.1, sequence analysis was carried out on ABI 3130XL, and sequences were analyzed with SeqScape software (all Applied Biosystems).

T. brucei Cell Lines, Culture, Transfection, and Western Blotting

The trypanosome cell lines used in this study were derived from the parental (wild-type [WT]) procyclic form (PCF) *T. brucei* 427 29-13 (present in the insect host) and bloodstream form (BSF) *T. brucei* 427 90-13 strains (present in the mammalian host), both co-expressing the T7 RNA polymerase and tetracycline repressor.[21] Cells were cultured, transfected, and cloned as previously described[22,23] in medium supplemented with hygromycin (25 μg/mL for PCF and 5 μg/mL for BSF), neomycin (10 μg/mL for PCF and 2.5 μg/mL for BSF), and blasticidin (20 μg/mL for PCF and BSF) for constitutive translation of 10TY1-tagged proteins and with phleomycin (5 μg/mL for PCF and 2.5 μg/mL for BSF) for RNAi-inducible cells. RNAi was induced with tetracycline (10 μg/mL).

To produce transgenes at their endogenous loci encoding proteins tagged at the C terminus with a 10TY1 tag ($TbWDR66_{TY1}$ or $TbWDR66\text{-}\Delta Cter_{TY1}$), we transfected WT cells with tagging cassettes obtained by PCR by using (1) a forward primer containing 80 nucleotides from the 3′ end of the *TbWDR66* or *TbWDR66-ΔCter* open reading frame (ORF) and a sequence specific to the pPOTv7-10TY1-neo vector[24] to amplify the TY1 tag and (2) a reverse primer containing 80 nucleotides from the 3′ UTR of *TbWDR66* and a sequence specific to the vector to amplify a selection gene.[24] For RNAi, a fragment of the *TbWDR66* ORF (base pairs 2,719–3,070, corresponding to amino acids 903–1,023 in the C-terminal domain) was cloned into p2T7tiB[25] and transfected into the $TbWDR66_{TY1}$ and $TbWDR66\text{-}\Delta Cter_{TY1}$ background cell lines. The primer sequences are presented in Table S4, and a schematic of the constructs is provided in Figure S1. RNAi was induced by tetracycline (10 μg/mL). For western blotting, samples were prepared and treated as previously described.[6]

Immunofluorescence of *T. brucei* Cells

Cells were collected, washed, and processed for immunolabeling as previously described.[3] In brief, cells were pelleted and washed, loaded on poly-L-lysine-coated slides, and air dried. After rehydration, cells were fixed and permeabilized in methanol at −20°C, rehydrated in PBS, and then incubated with primary antibodies in two steps: (1) anti-TY1 (BB2) followed by Alexa-594-conjugated anti-mouse IgG1 (Thermo Fisher Scientific, A21125) and (2) anti-PFR2 (rabbit polyclonal) and anti-tubulin TAT1 mouse IgG2a primary antibodies followed by anti-rabbit ATTO-N647-conjugated (Sigma 40839) and anti-mouse IgG2a Alexa-conjugated (Thermo Fisher Scientific A21131) secondary antibodies. Nuclei and kinetoplasts were stained with DAPI (10 μg/mL) in epifluorescence experiments. Epifluorescence images were acquired on a Zeiss Imager Z1 microscope with a Photometrics Coolsnap HQ2 camera equipped with a Zeiss 100× objective (numerical aperture [NA] 1.4) and Metamorph software (Molecular Devices) and were analyzed

with ImageJ. Stimulated emission depletion microscopy (STED) images were acquired on a Leica DMI6000 TCS SP8 X-STED microscope with an HC PL APO CS2 100×/1.40 oil objective and deconvolved with Huygens Pro 16.10. All antibodies used are listed in Table S5.

Video Microscopy of *T. brucei* Cells

Video microscopy was carried out as previously described.[26] In brief, BSF cells (WT cells and cells induced with RNAi for 48 hr) were washed, and cell mobility was recorded by phase contrast on a Zeiss AxioImager with a 40× objective (NA 1.3). Fifty seconds of digital video from separate regions was captured and analyzed with Metamorphoh software (Molecular Devices) and video edited with ImageJ (setting frame rate was 10 fps).

Electron Microscopy of *T. brucei* Cells

Cells were fixed in culture medium for 60 min through the addition of glutaraldehyde to a final concentration of 2.5%. They were pelleted (1,000 × *g*) for 10 min and resuspended in fixation buffer (2.5% glutaraldehyde, 2% PFA, 100 mM phosphate buffer [pH 7.4], and 50 mM sucrose) for 2 hr. Fixed cells were washed in water for 10 min, post-fixed in 1% OsO4 for 1 hr, washed three times in water, and then en bloc stained in 2% uranyl acetate in water at 4°C. Samples were next washed in water three times for 15 min, dehydrated in ethanol, embedded in Spurr resin, and polymerized overnight at 60°C. Sections (50–60 nm) were stained with UranyLess for 1 min and then washed in water followed by lead citrate. Sections were visualized on a TECNAI 12 transmission electron microscope.

Results

Identification of the Identified CNV

Using the ExomeDepth software package,[18] we re-analyzed the WES data obtained from the 78 previously described MMAF individuals.[6] Only one homozygous deleterious deletion was identified in several subjects and affected a testis-expressed gene. It was found in seven apparently unrelated North African individuals and removed exons 20 and 21 out of 22 of *WDR66*. The deletion was found in 7 of the 46 Tunisian MMAF subjects analyzed here, indicating that in this population, the *WDR66* deletion is responsible for 15% of cases. The deletion was not observed in the subjects recruited at the other centers (France and Iran). We confirmed the presence of the deletion in all affected individuals by using the Integrative Genomics Viewer (IGV) (Figure 2A). The presence of the deletion in the subjects was also confirmed by comparative genomic hybridization array (Figure S2). Consistent with a negative selection of homozygotes throughout evolution, the identified variant was absent from public databases, including DECIPHER,[27] the Database of Genomic Variants,[28] and Copy Number Variation in Disease,[29] and was found only in one heterozygous non-Finnish European (NFE) from the 60,706 unrelated individuals aggregated in the ExAC Browser. We note, however, that all affected subjects were North African, a population that is not well represented in these databases.

WDR66 Description, Transcript Expression, and Protein Localization

WDR66, also known as *CFAP251*, is located approximately 11 Mb from the telomere of 12q (chr12: 122,355,768–122,441,833 [Ensembl: ENSG00000158023]) and contains 22 exons. Transcription of *WDR66* leads to two isoforms: T1 (GenBank: NM_144668), which encodes 1,149 amino acids (UniProt: Q8TBY9-1), and T2 (GenBank: NM_001178003), which encodes 941 amino acids (UniProt: Q8TBY9-2). Because the deleted exons 20 and 21 are present only in T1, it is important to measure the expression of both transcripts in tissues of interest to confirm that T1 is expressed in the testis (Figure 2B). We therefore measured the expression levels of the transcripts by performing qRT-PCR in different human tissues by using specific primers for each isoform. Results showed that T1 was highly and predominantly expressed in the testis and moderately expressed in other ciliated tissues (especially the lung), whereas T2 had negligible expression in all tested tissues (Figure 2C). These results indicate that T1, the transcript affected by the identified deletion, is the main physiological transcript and is highly expressed in the testis (Figure 2C). In this article, we will therefore refer to this transcript (Ensembl: ENST00000288912.8) and to the protein it encodes. The consequence of the deletion on the transcript is c.3007_3337del.

Using the InterPro domain annotation program, we found that nine WD40-repeat domains are present in both isoforms T1 and T2, and one calcium-binding EF-hand domain is present only in isoform T1 in the C terminus (Figure 2D and Table S7), indicating that the activity of this isoform could be modulated by calcium. In the protein encoded by T1 (GenBank: NP_653269), the first amino acid affected by the identified deletion is predicted to be amino acid 1,003. Deletion of exons 20 and 21 induces a frameshift, and translation of exon 22 is predicted to be incorrect for the first 25 amino acids before a stop codon is induced 26 nucleotides after the first abnormally coded amino acid: p.Ile1003Lysfs26. Data on genetic variants described here are available under accession number ClinVar: SUB3375641.

We then wanted to determine whether the deleted transcript is present in sperm from individuals carrying the genomic deletion or subjected to mRNA decay. Unfortunately, we could not obtain enough biological material to perform a reliable RT-PCR. However, mRNA decay has been shown to be elicited by abnormally long 3′ UTRs and to be more efficient when a premature stop codon is introduced near the 5′ end of the transcript.[30] The deletion identified here removes the two penultimate exons and creates a new stop codon in the last exon near the transcript's 3′ end. We therefore do not expect this mechanism to be active here. Out of the four commercially available antibodies, we observed that only one provided a specific staining (Table S2). This antibody has been raised against a 101 amino acid epitope overlapping with the subject's deletion. We observed that, although the staining was

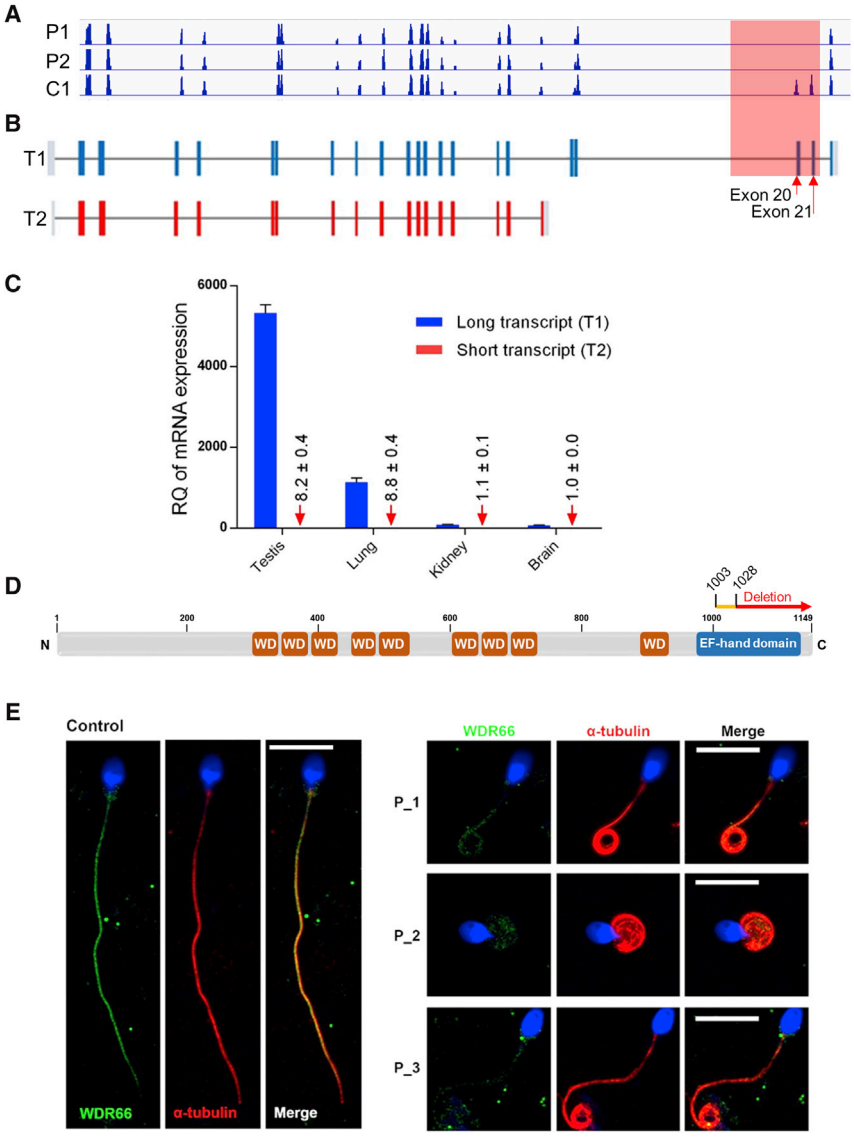

Figure 2. Confirmation of the Presence of the *WDR66* Deletion in Affected Individuals and Validation of Its Effect on mRNA and Protein

(A) Compared with a control individual (C1), affected individuals P1 and P2 show a complete absence of sequence coverage of exons 20 and 21, as illustrated by IGV.

(B) In humans, *WDR66* encodes two transcripts: T1 (Ensembl: ENST00000288912.8; GenBank: NM_144668), in which 22 exons code for a 1,149 amino acid protein, and T2 (Ensembl: ENST00000397454.2; GenBank: NM_001178003), in which 18 exons code for a 941 amino acid protein. Only T1 is affected by the identified deletion (black double arrow).

(C) Relative expression ($2^{-\Delta\Delta Ct}$) of both transcripts (T1 and T2) was performed on a human panel of cDNAs from two tissues expressing 9 + 2 secondary motile axonemes (testis and lung) and two tissues expressing 9 + 0 primary immotile axonemes (kidney and lung). *GAPDH* was used as a reference gene, and the T2 expression level in the brain was used as a reference sample. Data are expressed as mean ± SD.

(D) Representation of WDR66, encoded by T1. We note the presence of nine WD40-repeat domains and one calcium-binding EF hand (α) in the C-terminal extremity. The first deleted amino acid is number 1,003. Deletion of exons 20 and 21 induces a frameshift. Translation of exon 22 is then incorrect for the first 25 amino acids (orange line) before a stop codon is induced 26 nucleotides after the first abnormal amino acid. The protein encoded by the deleted T1 is therefore expected to stop at amino acid 1,028.

(E) Double immunofluorescence labeling of WDR66 and α-tubulin in human sperm cells. In control sperm, the yellow stain resulting from merged signals indicates the co-localization of both proteins within the flagellar axoneme. Typical sperm from affected subjects P1–P3 displayed abnormally shaped flagella with weak and punctuated WDR66 staining. The images shown are representative of many cells examined in two independent experiments. Scale bar, 10 μm.

weaker than in control sperm, the subject's short and abnormal flagella were immunodecorated along the flagella, suggesting that the truncated protein is present in the affected subjects and localized to the flagella (Figure 2E). The intensity of the signal in sperm from affected individuals is, however, very weak, and in the absence of a negative control we cannot exclude the possibility that it could be due to non-specific staining and thus that the protein could be completely absent in subjects with this deletion.

Breakpoint Characteristics and CNV Mutational Mechanisms

Then, to unequivocally confirm the exact size and position of the deletion, we performed a walking PCR approach. We validated a PCR protocol that specifically amplified a 950 bp fragment in affected individuals by using the primers F1 (located 5,209 bp upstream of exon 20) and R1 (positioned 1,597 bp downstream of exon 21) (Figures 3A and 3B). As could be expected, the 8,541 bp interval in control individuals could not be amplified and yielded no signal, whereas a shorter band (950 bp) was obtained from all affected individuals (Figure 3C). Sanger analysis of the PCR sequence covering the breakpoint region in the affected individuals allowed the characterization of the breakpoint at single-nucleotide resolution (Figure 3D). We observed that the deletion spanned 7,591 bp and exons 20 and 21 of *WDR66* were completely deleted (chr12: 122,432,883–122,440,629). All affected individuals carried the same deletion with the same nucleic acid sequence at the breakpoint locus, suggesting that they had all inherited the same deleted allele from a

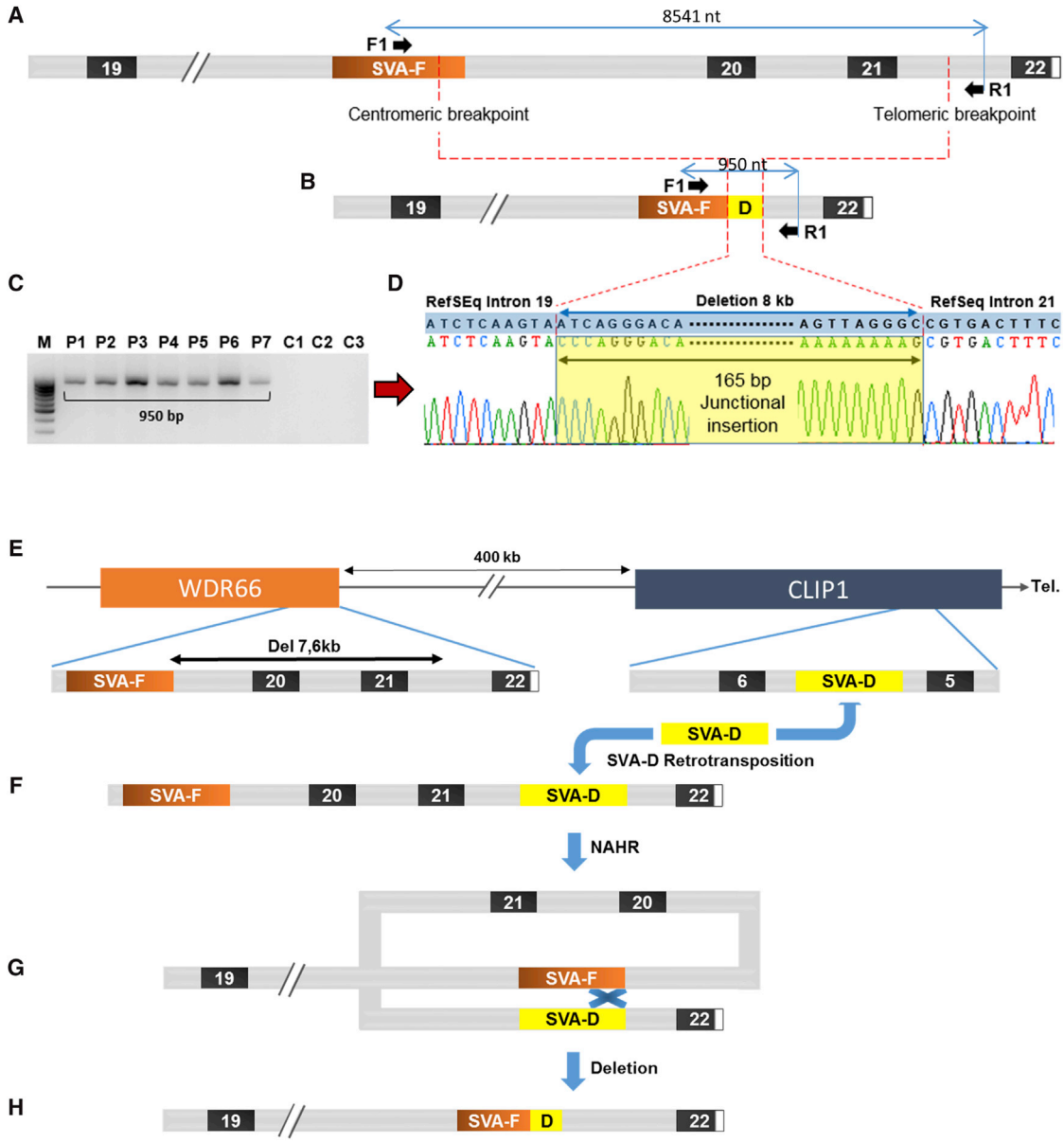

Figure 3. Fine Characterization of the Deletion and Theoretical Description of the Initial Deletion Event
(A) Representation in control subjects of part of *WDR66* (between exons 19 and 22) and localization of the deletion-specific primers F1 and R1.
(B) Representation of the same region in affected subjects. Detailed analysis of the deleted region within *WDR66* highlights the presence of a type F SINE-VNTR-Alu (SVA) retrotransposon (SVA-F) positioned at the centromeric border of the identified *WDR66* deletion (orange box).
(C) PCR of the deleted region. In control subjects (C1–C3), F1 and F2 primers are too distant (8,541 bp) and do not yield any amplification. PCR amplification with F1 and F2 in affected subjects (P1–P7) permits the amplification of a 950 nucleotide fragment.
(D) Sanger sequencing through the deletion breakpoints permits localization of the exact position of the breakpoints, indicating that a total of 7,591 nucleotides were deleted. It also highlights the presence of a 165 nucleotide sequence (yellow) inserted at the junction point.
(E) Blast analysis of the 165 nucleotide inserted sequence shows that an identical sequence (100% identity) is present within an SVA-D sequence located 400 kb telomeric to SVA-F between exons 5 and 6 of *CLIP1* (yellow box).
(F) SVA-D was most likely retrotransposed between *WDR66* exons 21 and 22.
(G) An intra-chromatidien non-allelic homologous recombination (NAHR) was then initiated.
(H) The formation of a hybrid SVA-F/SVA-D sequence preceding the *WDR66* exon 20 and 21 deletions.

common ancestor. The loss of exons 20 and 21 induces a frameshift leading to a premature stop codon at the beginning of the last exon (exon 22). Therefore, the 147 amino acids encompassing the C-terminal region of the protein (12.8% of the full-length protein) are predicted to be lost.

To estimate the allelic frequency of this variant and to confirm that it is not a frequent CNV in the North African

population, we carried out a multiplex PCR allowing the co-amplification of the deleted allele (950 bp of product size) with the normal allele (220 bp), thus including a positive control in all reactions. A total of 100 North African individuals were analyzed (Figure S3). None of the tested control individuals amplified the deletion, indicating that the variant is also rare (allelic frequency $< 1/200$) within the North African population. We then re-analyzed the whole data from WES analysis (SNPs and indels) of the 78 individuals, but we did not identify any other *WDR66* deleterious variants.

To investigate the origin of this complex rearrangement, we carried out a series of bioinformatic analyses focusing on the genomic region surrounding the deletion. Clustering the deleted sequence by using RepeatMasker software indicated that the deleted region is rich in transposable elements and prone to genomic rearrangements. In addition, the centromeric breakpoint of the identified deletion was localized at the exact ending of an SVA_F retrotransposable element (Figures 3A, 3B, and 3E). Intriguingly, detailed sequence analysis revealed that in the affected subjects, the telomeric fragment of this element (122,432,706–122,432,883) did not perfectly match the expected SVA-F sequence (Figure S4). Alignment of this sequence with BLAT showed 100% similarity to a distal sequence of an SVA-D element (Figure S4) located in intron 5 of *CLIP1*, a gene located 400 kb telomeric to *WDR66* (Figure 3E). In control individuals, the abnormal telomeric end of the *WDR66* SVA_F sequence perfectly matched the reference sequence (GRCh37, UCSC Genome Browser hg19), indicating that the short SVA_D sequence observed only in affected subjects was specific to the deletion event and most likely came from the *CLIP1* SVA_D retrotransposable element. We conclude that SVA-D was most likely retrotransposed into *WDR66* intron 21 (Figure 3F) and initiated non-allelic homologous recombination (NAHR) with the SVA-F element in intron 19 (Figure 3G) to induce the formation of a single chimeric SVA-F/SVA-D sequence, which preceded the *WDR66* deletion (Figure 3H). This indicates that *WDR66*-SVA-F recombined at the following localization: 122,432,691–122,432,706. Such a mechanism has already been observed and has been described as an SIMD.[19]

Dating of the Identified Mutation

To compute the age of the deletion, we use the size of the shared ancestral haplotypes surrounding the *WDR66* deletion locus (Table S6); if we assume a recombination rate of 1.30 cM/Mb,[31] the size is equal to 0.26 cM. The probability of observing an ancestral haplotype of size c (measured in morgans) or larger around the *WDR66* deletion is equal to $(1-c)^g$, where g is the deletion's age measured in generations.[32] The probability that the $2n$ haplotypes carried by the seven individuals is larger than c is $(1-c)2^{ng}$, which can be approximated by e^{-2ngc}. Therefore, the length of the shared ancestral haplotype can be approximated by an exponential distribution of rate $2ng$. The parameter g can be estimated by $1/2nL$, which provides an estimate of $g = 27$ generations. However, Labuda et al. have shown that dating arising from genetic clock methods is biased downward because these methods do not account for population expansion.[33] To account for an exponential expansion of rate r per generation, the mathematical expression $(1/r)\ln(c.e^r/(e^r-1))$ should be added to the estimates provided by genetic clock methods. If we assume that an upper bound for the expansion rate is given by the expansion $r = 0.4$, as was calculated for the Ashkenazi Jewish populations,[34] the age of the mutation is in the range of 27–43 generations. Given 25 years for the generation time, the mutation would have then arisen 675–1,075 years ago. Heterozygous individuals have apparently not been negatively selected, and the deleted allele now seems relatively frequent in Tunisia given that 7 of the 46 (15%) Tunisian MMAF subjects analyzed here had a homozygous *WDR66* deletion.

Phenotypic Analysis of MMAF Individuals Carrying a Genomic *WDR66* Deletion

The analysis of semen samples from the seven individuals carrying the same mutation showed normal semen volume and concentration, although with large inter-individual variations (Table 1). A total absence of sperm motility was observed for all individuals carrying the *WDR66* deletion (Table 1). Sperm morphology appeared seriously altered under light microscopy, wherein 100% of spermatozoa were abnormal, as illustrated in Figure 1. The most common morphological defects observed in the semen of all individuals concerned the flagellum dimensions (reduced length and irregular caliber). A high rate (56%–95%) of acrosomal abnormalities was also observed in many affected subjects. We compared the sperm parameters of individuals with different genotypes (mutations in *DNAH1*, *CFAP43*, *CFAP44*, and *WDR66*) but observed no significant differences among the groups (Table S8).

Because an insufficient number of sperm cells were collected from the affected individuals, only sperm cells from one individual carrying the *WDR66* deletion could be analyzed by TEM (Figure 1). Longitudinal sections showed severe axonemal and peri-axonemal defects affecting the outer dense fibers (ODFs), the fibrous sheath (FS), and the mitochondrial sheath (MS), which appeared completely disorganized, resulting in aborted flagella or their replacement by a cytoplasmic mass englobing unassembled axonemal components (data not shown). Because oligo-teratozoospermia was observed in this individual, fewer than ten cross-sections were of a sufficient quality for analysis. Among these sections, 100% were abnormal, and the main defect constantly observed concerned the axonemal and the peri-axonemal structures, such as the unassembled FS (Figure 1F).

Functional Study of the *T. brucei* WDR66 Ortholog

To better characterize WDR66 localization and function, we decided to use a tractable model organism, which could

Table 1. Semen Parameters of the Seven MMAF Subjects

	Affected Subjects								
	P1	P2	P3	P4	P5	P6	P7	Mean ± SD	Lower Reference Limits (WHO 2010)
Geographical origin	Tunisia	Tunisia	Tunisia	Tunisia	Tunisia	Tunisia	Tunisia	–	–
Consanguinity	yes	yes	yes	yes	no	yes	yes	–	–
Age (years)	45	54	51	45	31	28	59	44.7 ± 10.6 (n = 7)	–
Sperm volume (mL)	4	1.8	1.7	2.2	3.4	3	5	3.0 ± 1.1 (n = 7)	1.5
Sperm concentration (million/mL)	11.3	6	56	56	60	7,8	15.5	30.3 ± 23.5 (n = 7)	15
Total motility after 1 hr (%)	2	1	5	0	5	5	4	3.1 ± 1.9 (n = 7)	40
Vitality (%)	53	80	73	79	62	NA	47	65.6 ± 12.6 (n = 6)	58
Normal spermatozoa (%)	0	0	0	0	0	0	0	0 (n = 7)	4
Abnormal acrosome (%)	66	56	56	78	80	NA	95	71.8 ± 15.3 (n = 6)	–
Abnormal head morphology (%)	42	10	20	58	8	NA	7	24.1 ± 25.1 (n = 6)	–
Abnormal base (%)	36	2	32	14	2	NA	3	14.8 ± 15.6 (n = 6)	–
Absent flagella (%)	31	6	22	50	24	NA	30.5	27.2 ± 13.1 (n = 6)	–
Short flagella (%)	34	32	59	32	50	NA	25	38.6 ± 11.8 (n = 6)	–
Coiled flagella (%)	20	17	27	34	8	NA	27	21.1 ± 8.3 (n = 6)	–
Flagella of irregular caliber (%)	4	68	37	50	68	NA	33	43.3 ± 22.1 (n = 6)	–
Multiple anomalies index	2.5	2.3	2.7	3.4	2.5	NA	2.56	2.6 ± 0.4 (n = 6)	–

The following abbreviation is used: NA, not available.

be used for forward and reverse genetics. We chose *T. brucei* because this flagellated, pathogenic protozoan is a common model for studying the function of axonemal proteins and has contributed to elucidating the molecular pathogeny of human ciliopathies[35] and more recently of MMAF syndrome.[6] BLASTp analysis of the *T. brucei* genome database[36] with the human WDR66 sequence identified the *T. brucei* ortholog Tb927.3.1670 (named TbWDR66 in this study). TbWDR66 is a 1,027 amino acid protein. The sequence identity calculated with Clustal Omega[37] between the two proteins is 30.55%. Furthermore, alignment of the secondary structures of the human and *T. brucei* proteins showed a near-perfect match (Figure S5 and Table S7), suggesting that the different protein domains (including the C-terminal domain containing an EF-hand motif) have preserved their functionality throughout evolution. Moreover, previous functional genomics and proteomic studies identified TbWDR66 as a flagellar protein.[38,39] In addition, TbWDR66 is the *Tetrahymena* and *Chlamydomonas* CFAP251 ortholog that was shown to play a role in building radial spoke 3 (RS3) in cilia.[40]

Because the localization and function of WDR66 in the trypanosome flagellum is currently unknown, we localized TbWDR66 in BSF (present in the mammalian host) and PCF (present in the insect host) *T. brucei* by using 10× TY1-tagged protein and generating *T. brucei* cell lines expressing endogenous levels of C-terminal-tagged TbWDR66 ($TbWDR66_{TY1}$).[24,41] $TbWDR66_{TY1}$ was found in the axoneme in the BSF, as substantiated by double labeling with antibodies against the paraflagellar rod structure (PFR) and TY1 (Figure 4A). We also used STED and triple PFR-tubulin-TY1 labeling, which confirmed the axoneme localization in the PCF (Figure S6). $TbWDR66_{TY1}$ signal extended throughout the whole length of the flagellum, as demonstrated by the TY1 labeling preceding the PFR at the proximal end of the flagellum and along the flagellum up to its distal end (Figure 4A, white boxes). We also generated BSF and PCF cell lines expressing TbWDR66 without its C-terminal domain (deleted amino acids 907–1,027 [$TbWDR66\text{-}\Delta Cter_{TY1}$]), corresponding to the deletion of exons 20–22 in the human protein, and showed that the absence of the C-terminal domain did not affect protein localization (Figure 4A and Figure S6), similar to our observations in the sperm of affected individuals with the WDR66 C-terminal deletion.

To evaluate the impact of TbWDR66 knockdown on flagellar structure and function, we generated a BSF cell line expressing $TbWDR66_{TY1}$ and inducible for *TbWDR66* by RNAi targeting the sequence encoding the C-terminal domain (Figure S1). Although RNAi induction did not fully deplete the amount of $WDR66_{TY1}$, as shown by immunofluorescence and western-blot detection of the tagged protein (Figures 4A and 4B), it did lead to a decrease in cell proliferation compared with that in WT or non-induced cultures (Figure 4C). More importantly, RNAi induction

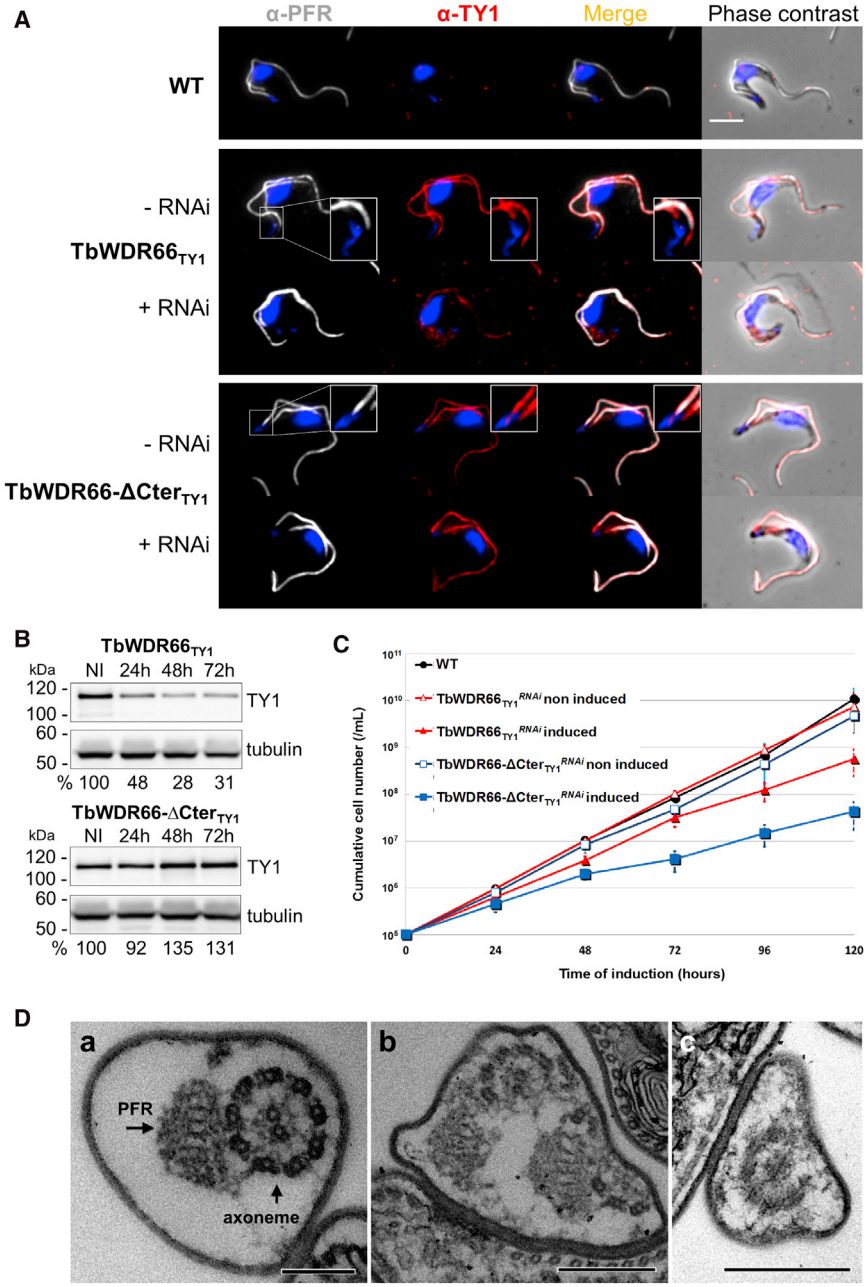

Figure 4. TbWDR66 Is an Axoneme Protein, and Its C-Terminal Domain Is Required for Function in BSF Trypanosome

(A) Control BSF cells (WT) or cells expressing TbWDR66$_{TY1}$ or TbWDR66-ΔCter$_{TY1}$ were immunolabeled with anti-PFR (white) and anti-TY1 (red); anti-TY1 decorates the tagged proteins throughout the full length of the flagellum. Induction of RNAi for 48 hr (+RNAi) showed a decrease in fluorescence intensity of TbWDR66$_{TY1}$ labeling but not of TbWDR66-ΔCter$_{TY1}$ labeling. Nuclei and kinetoplasts (mitochondrial genomes) were labeled with DAPI (blue). Scale bar, 5 μm.

(B) Western-blot analysis of the presence of TbWDR66$_{TY1}$ and TbWDR66-ΔCter$_{TY1}$ without (NI) and with 24, 48, and 72 hr of RNAi induction shows a decreased amount of TbWDR66$_{TY1}$ (indicated by the labeling percentage relative to the non-induced condition), whereas the amount of translated TbWDR66-ΔCter$_{TY1}$ was not affected. Tubulin labeling was used as loading control.

(C) *TbWDR66* RNAi affects cell proliferation. Shown are growth curves for WT cells, and RNAi-non-induced or -induced cells expressing TbWDR66$_{TY1}$ or TbWDR66-ΔCter$_{TY1}$. Cells were counted every 24 hr. The graph represents the cumulative number of cells per milliliter. Error bars represent the standard error from three independent experiments.

(D) Electron microscopy images of thin sections of a WT cell (a) and a cell expressing TbWDR66$_{TY1}$ (b) or TbWDR66-ΔCter$_{TY1}$ (c) and induced with RNAi for 48 hr show that the decreased amount of TbWDR66 induces severe disorganization of the axoneme and of the paraxonemal structure, which is not rescued by the presence of TbWDR66-ΔCter$_{TY1}$. Scale bars, 50 nm.

strongly affected flagellar motility, as shown by video microscopy (Video S1 for WT and Video S2 for cells expressing TbWDR66$_{TY1}$ and induced with RNAi for 48 hr). Anecdotally but notably, electron microscopy revealed some flagellar sections (2%, n > 149) showing a disorganization of the axoneme with displaced doublet microtubules (DMTs), a phenotype not observed in the WT control cells (Figure 4D). Together, these data indicate that TbWDR66 plays an important role in flagellar motility and is also likely to be involved in axoneme structure, two functions that are in agreement with what has been described for CFAP251 in *Tetrahymena*.[40]

To assess the role of the TbWDR66 C-terminal domain in protein function, we generated a cell line expressing TbWDR66-ΔCter$_{TY1}$ and inducible for the RNAi targeting only the endogenous *TbWDR66* (Figure S1B). Cell proliferation was not affected by the expression of TbWDR66-ΔCter$_{TY1}$ in RNAi non-induced cells (Figure 4C). As expected, expression and localization of TbWDR66-ΔCter$_{TY1}$ was not affected by the RNAi as shown by immunofluorescence and western-blot analyses (Figures 4A and 4B). However, flagellum motility was impaired similarly to the phenotypes induced by RNAi in the WDR66$_{TY1}$ background cell line (Video S3) demonstrating that expression of TbWDR66 deleted of its C-terminal domain cannot complement for the downregulation of the full-length TbWDR66. Further, cell proliferation decrease was slightly stronger than in the WDR66$_{TY1}$ background cell line suggesting that, in the context of a protein complex, presence of a non-functional protein might be more deleterious than absence of the protein (Figure 4C). Together, these data demonstrate that TbWDR66 is involved in flagellum

motility and that its C-terminal domain is required for protein function.

Discussion

WDR66 Is Involved in Human Male Infertility and Accounts for 15% of MMAF Cases in Tunisian Men

In the present study, we analyzed 78 MMAF individuals by combining SNPs and indels with CNV analysis of WES data and identified a homozygous intragenic deletion encompassing exons 20 and 21 of *WDR66*, which encodes a WD-repeat protein. We previously showed that two other WD-repeat proteins (CFAP43 and CFAP44) are responsible for the MMAF phenotype and account for 20.5% of the same cohort.[6] In this study, we implicate another WD-repeat protein in 15% of MMAF subjects from Tunisia, indicating that *WDR66* is another important gene for this phenotype. This suggests that WD-repeat proteins are enriched in axonemal and peri-axonemal structures. Interestingly, we observed the exact same deletion in all seven individuals, all of whom were of North African origin, indicating a founder effect. Study of the SNPs present around the deleted regions confirmed this and showed that all affected subjects shared an ancestral 0.26 cM haplotype, suggesting that their common ancestor lived 675–1,075 years ago.

Efficiency of WES to Detect Pathological CNVs and Importance of SVA Retrotransposons Elements

CNVs are a common cause of disease but until now have not been explored in WES data in the field of male infertility, particularly the MMAF phenotype. In the current study, we analyzed WES data by using ExomeDepth software, which is based on the assessment of sequencing read depth. We analyzed the breakpoints to elucidate the mechanisms underlying the deletion. Results showed a high density of *Alu* retrotransposons in the region flanking the deletion breakpoints, suggesting that this region exhibits a high degree of genomic instability. Interestingly, we observed a single chimeric SVA element located precisely adjacent to the centromeric breakpoint, strongly suggesting that the creation of this abnormal SVA and the deletion event are the consequence of the same NAHR between SVA_F in intron 19 of *WDR66* and the SVA_D retrotransposed from intron 5 of *CLIP1* to intron *21* of *WDR66*. This event resulted in the deletion of a portion of each SVA element and of the intervening DNA sequences (Figure 3). This mechanism was previously described as SIMD.[19] SVAs, originally named SINE-R, are composite transposons that make up 0.2% of the human genome and represent the youngest and currently the most active retrotransposable elements of the human genome.[42] SVAs are mobilized by L1 enzymatic machinery throughout the genome by a "copy-and-paste" mechanism.[43] It has been shown that retrotransposition of SVA elements is a source of insertions, deletions, and rearrangements within the target site and leads to various pathologies.[44] The spectrum of diseases that have been reported in relation to SVA transposition, including Marfan syndrome, hemophilia B, neurofibromatosis types 1 and 2, leukemia, and congenital muscular dystrophy, has been increasing progressively.[45–50] In this study, we report a homozygous CNV that is generated by an SVA transposition event and causes MMAF, leading to male infertility.

WDR66 Function and the Effect of the Identified Deletion

The axoneme, the internal cytoskeleton of motile cilia and flagella, is a highly evolutionary conserved structure that consists of nine DMTs circularly arranged around the central-pair complex of microtubules (9 + 2 structure).[51] The beating of cilia and flagella is orchestrated by multiprotein-ATPase complexes that are located on the peripheral doublets, which provide the sliding force for sperm motility. In addition, the sperm flagellum harbors specific peri-axonemal structures that are not found in other motile cilia: a helical MS in the midpiece, the FS in the principal piece (PP), and ODFs in the midpiece and the proximal part of the PP.[52]

Previous reports have shown that the *WDR66* ortholog (*FAP251*) in *Chlamydomonas* localizes to the CSC (calmoduline- and spoke-associated complex) at the base of RS3.[53] The CSC was recently identified as an axonemal complex that associates the bases of RS2 and RS23 and the N-DRC (nexin-dynein regulatory complex) and is involved in the modulation of dynein activity and therefore in the regulation of flagellar motility.[53] Recent work in *Tetrahymena* has shown that, in the absence of CFAP251, RS3 was either missing or incomplete, and the protozoa had an altered swimming ability with an abnormal ciliary waveform.[40] Here, we studied the localization of the *WDR66* ortholog in *T. brucei* and used this model to confirm the deleterious effect of the identified deletion. Immunofluorescence and STED microscopy confirmed that TbWDR66 is an axonemal protein and that in *T. brucei*, knockdown by RNAi affects flagellar structure and motility. We then generated cell lines with a deletion of the TbWDR66 C-terminal domain, corresponding to the deletion identified in MMAF individuals. We observed that the absence of the C-terminal domain did not affect protein localization (Figure 4A and Figure S6) but severely affected cell mobility, as was observed for the sperm of affected individuals. This work confirms that the WDR66 C-terminal domain, which contains a calcium-regulating EF-hand domain, is critical for maintaining flagellar function probably by regulating flagellar beating; we also observed that WDR66 plays a critical structural role, perhaps by consolidating the links between RS3 and the microtubule doublets.

Impact of MMAF-Related Genes on Flagellar Structure and Ultrastructure in Mammals and *T. brucei*

The seven subjects with a partial *WDR66* deletion presented with a classical MMAF phenotype including its

typical hallmarks: a near absence of motility (3%) and a high proportion of short (39%), absent (27%), and coiled (21%) flagella and of irregularly calibrated flagella (43%) (Table 1 and Figures 1A–1D). These parameters are similar to those found previously in individuals with mutations in other MMAF genes, such as *DNAH1*, *CFAP43*, and *CFAP44* (Table S8), confirming the relative homogeneity of the phenotype and underlining its genetic heterogeneity. Previous ultrastructural analyses of the sperm flagella of subjects with mutations in the three initial genes highlighted a severe axoneme disorganization particularly affecting the central-pair microtubules but also the other doublets and even the extra-axonemal structures. For instance, we have shown that the absence of DNAH1, a protein located in the inner dynein arm facing RS3 (and thus a putative partner of WDR66), induces a strong disorganization of the axonemal structure. Similarly, the absence of CFAP43 and CFAP44, two proteins playing a role in the interaction between the axoneme and peri-axonemal structures, leads to a strong disorganization of the axonemal structure in human and mouse sperm. Here, we showed that the absence of a functional WDR66 also induces a spectacular disorganization of the flagellar ultrastructure (Figure 1 E). Interestingly, we observed that the level of flagellar disorganization was much more pronounced in human and mouse sperm than in a flagellate protozoan depleted of the MMAF orthologs (as illustrated by the study of CAFAP43, CFAP44, and WDR66 orthologs in *T. brucei*). In the studied mammals, the absence of these axonemal proteins seemed to have a profound destabilizing effect during flagellar growth whereby it induced the different morphological anomalies described above, including malformations of the FS and other peri-axonemal structures. The lack of structural defects in *T. brucei* could indicate that axonemal growth and intraflagellar transport (IFT) are more resilient in the parasite, suggesting that IFT differs between sperm and parasite flagella. Moreover, in *T. brucei* the flagellum is attached to the body of the parasite, and we can speculate that the axoneme is not subjected to the same constraints during flagellar formation, thus perhaps also explaining the absence of marked ultrastructural alterations.

Overall, we have shown that the deletion of the WDR66 C-terminal domain is a common genetic cause of male infertility in Tunisian men affected by MMAF. This finding highlights the importance of this domain and of WDR66 for flagellar structure, assembly, and beating. We confirmed that WES is efficient at identifying small CNV events (here, a deletion of two exons) and could eventually replace or at least complement conventional microarray analysis. Importantly, as previously observed for CFAP43 and CFAP44, we showed the high level of conservation of flagellar proteins between humans and *T. brucei*, confirming that this organism is an excellent model for investigating the molecular function and pathogeny of genes associated with flagellar dysfunction in humans.

Accession Numbers

The accession number for the variants reported in this paper is ClinVar: SUB3375641.

Supplemental Data

Supplemental Data include six figures, eight tables, and three videos and can be found with this article online at https://doi.org/10.1016/j.ajhg.2018.07.014.

Acknowledgments

We thank the affected and control individuals for their participation. This work was mainly supported by the Agence Nationale de la Recherche (ANR; programme DS0401-2014, project ANR-14-CE15-0002) through the "MAS-Flagella" project, the Direction Générale de l'Offre de Soins through the 2014 Programme de Recherche Translationnelle en Santé (to P.F.R., M.B., A.T., C.A., and N.T.M.), the Fondation Maladies Rares through "Whole genome sequencing of individuals with Flagellar Growth Defects (FGD)" of the 2012 Séquençage à Haut Débit program (to P.F.R.), and ANR LabEx through the ParaFrap grant (ANR-11-LABX-0024 to D.R.R.; http://www.labex-parafrap.fr/fr). Stimulated emission depletion microscopy was performed at the Bordeaux Imaging Centre, a service unit of the CNRS-INSERM and Bordeaux University and a member of France-BioImaging, supported by the ANR (ANR-10-INBS-04). The help of Patrice Mascalchi is acknowledged. Some of the electron microscopy was performed at the Bordeaux Imaging Centre. We thank Nicolas Biteau (Bordeaux University) for the rabbit anti-PFR2 antibody, Sam Dean (Oxford University) for the pPOTv7 plasmid, Keith Gull (Oxford University) for the anti-tubulin (TAT1) antibody, and Philippe Bastin (Institut Pasteur, Paris) for the anti-TY1 antibody. We also thank the Grenoble Institut des Neurosciences electron microscopy platform and Anne Bertrand as well as the Institut pour l'Avancée des Biosciences microscopy platform, Alexei Grichine, and Jacques Mazzega for their technical help.

Declaration of Interests

The authors declare no competing interests.

Received: February 23, 2018
Accepted: July 18, 2018
Published: August 16, 2018

Web Resources

ClinVar, https://www-ncbi-nlm-nih-gov.gate2.inist.fr/clinvar/
Copy Number Variation in Disease, http://bioinfo.hrbmu.edu.cn/CNVD
Database of Genomic Variants, http://dgv.tcag.ca/
DECIPHER, https://decipher.sanger.ac.uk/
ExAC Browser, http://exac.broadinstitute.org
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
InterPro domain annotation program, http://www.ebi.ac.uk/interpro/
NCBI Gene, https://www.ncbi.nlm.nih.gov/gene
OMIM, http://www.omim.org/
RepeatMasker, http://www.repeatmasker.org
UCSC Genome Browser, http://genome.ucsc.edu
UniProt, http://www.uniprot.org/

References

1. Neto, F.T.L., Bach, P.V., Najari, B.B., Li, P.S., and Goldstein, M. (2016). Spermatogenesis in humans and its affecting factors. Semin. Cell Dev. Biol. *59*, 10–26.
2. Jan, S.Z., Vormer, T.L., Jongejan, A., Röling, M., Silber, S.J., de Rooij, D.G., Hamer, G., Repping, S., and van Pelt, A.M.M. (2017). Unraveling transcriptome dynamics in human spermatogenesis. Development. *144*, 3659–3673.
3. Coutton, C., Escoffier, J., Martinez, G., Arnoult, C., and Ray, P.F. (2015). Teratozoospermia: Spotlight on the main genetic actors in the human. Hum. Reprod. Update *21*, 455–485.
4. Ray, P.F., Toure, A., Metzler-Guillemain, C., Mitchell, M.J., Arnoult, C., and Coutton, C. (2017). Genetic abnormalities leading to qualitative defects of sperm morphology or function. Clin. Genet. *91*, 217–232.
5. Ben Khelifa, M., Coutton, C., Zouari, R., Karaouzène, T., Rendu, J., Bidart, M., Yassine, S., Pierre, V., Delaroche, J., Hennebicq, S., et al. (2014). Mutations in *DNAH1*, which encodes an inner arm heavy chain dynein, lead to male infertility from multiple morphological abnormalities of the sperm flagella. Am. J. Hum. Genet. *94*, 95–104.
6. Coutton, C., Vargas, A.S., Amiri-Yekta, A., Kherraf, Z.-E., Ben Mustapha, S.F., Le Tanno, P., Wambergue-Legrand, C., Karaouzène, T., Martinez, G., Crouzy, S., et al. (2018). Mutations in CFAP43 and CFAP44 cause male infertility and flagellum defects in Trypanosoma and human. Nat. Commun. *9*, 686.
7. Tang, S., Wang, X., Li, W., Yang, X., Li, Z., Liu, W., Li, C., Zhu, Z., Wang, L., Wang, J., et al. (2017). Biallelic mutations in *CFAP43* and *CFAP44* cause male infertility with multiple morphological abnormalities of the sperm flagella. Am. J. Hum. Genet. *100*, 854–864.
8. Sha, Y.-W., Wang, X., Xu, X., Su, Z.-Y., Cui, Y., Mei, L.-B., Huang, X.-J., Chen, J., He, X.-M., Ji, Z.-Y., et al. (2017). Novel mutations in *CFAP44* and *CFAP43* cause multiple morphological abnormalities of the sperm flagella (MMAF). Reprod. Sci. https://doi.org/10.1177/1933719117749756.
9. Sha, Y.-W., Xu, X., Mei, L.-B., Li, P., Su, Z.-Y., He, X.-Q., and Li, L. (2017). A homozygous *CEP135* mutation is associated with multiple morphological abnormalities of the sperm flagella (MMAF). Gene *633*, 48–53.
10. Lorès, P., Coutton, C., El Khouri, E., Stouvenel, L., Givelet, M., Thomas, L., Rode, B., Schmitt, A., Louis, B., Sakheli, Z., et al. (2018). Homozygous missense mutation L673P in adenylate kinase 7 (AK7) leads to primary male infertility and multiple morphological anomalies of the flagella but not to primary ciliary dyskinesia. Hum. Mol. Genet. *27*, 1196–1211.
11. Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. Nat. Rev. Genet. *7*, 85–97.
12. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. Nat. Genet. *36*, 949–951.
13. Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., et al.; 1000 Genomes Project (2011). Mapping copy number variation by population-scale genome sequencing. Nature *470*, 59–65.
14. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-scale copy number polymorphism in the human genome. Science *305*, 525–528.
15. Almal, S.H., and Padh, H. (2012). Implications of gene copy-number variation in health and diseases. J. Hum. Genet. *57*, 6–13.
16. Ma, D., Yang, J., Wang, Y., Huang, X., Du, G., and Zhou, L. (2017). Whole exome sequencing identified genetic variations in Chinese hemangioblastoma patients. Am. J. Med. Genet. A. *173*, 2605–2613.
17. Tsuchida, N., Nakashima, M., Kato, M., Heyman, E., Inui, T., Haginoya, K., Watanabe, S., Chiyonobu, T., Morimoto, M., Ohta, M., et al. (2018). Detection of copy number variations in epilepsy using exome data. Clin. Genet. *93*, 577–587.
18. Plagnol, V., Curtis, J., Epstein, M., Mok, K.Y., Stebbings, E., Grigoriadou, S., Wood, N.W., Hambleton, S., Burns, S.O., Thrasher, A.J., et al. (2012). A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics *28*, 2747–2754.
19. Lee, J., Ha, J., Son, S.-Y., and Han, K. (2012). Human genomic deletions generated by SVA-associated events. Comp. Funct. Genomics *2012*, 807270.
20. Jungwirth, A., Giwercman, A., Tournaye, H., Diemer, T., Kopa, Z., Dohle, G., Krausz, C.; and European Association of Urology Working Group on Male Infertility (2012). European Association of Urology guidelines on male infertility: The 2012 update. Eur. Urol. *62*, 324–332.
21. Wirtz, E., Leal, S., Ochatt, C., and Cross, G.A. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *99*, 89–101.
22. Brun, R., and Schönenberger. (1979). Cultivation and in vitro cloning or procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. Short communication. Acta Trop. *36*, 289–292.
23. Schumann Burkard, G., Jutzi, P., and Roditi, I. (2011). Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. Mol. Biochem. Parasitol. *175*, 91–94.
24. Dean, S., Sunter, J., Wheeler, R.J., Hodkinson, I., Gluenz, E., and Gull, K. (2015). A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. Open Biol. *5*, 140197.
25. LaCount, D.J., Barrett, B., and Donelson, J.E. (2002). *Trypanosoma brucei FLA1* is required for flagellum attachment and cytokinesis. J. Biol. Chem. *277*, 17580–17588.
26. Oberholzer, M., Lopez, M.A., Ralston, K.S., and Hill, K.L. (2009). Approaches for functional analysis of flagellar proteins in African trypanosomes. Methods Cell Biol. *93*, 21–57.
27. Bragin, E., Chatzimichali, E.A., Wright, C.F., Hurles, M.E., Firth, H.V., Bevan, A.P., and Swaminathan, G.J. (2014). DECIPHER: Database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic Acids Res. *42*, D993–D1000.
28. MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L., and Scherer, S.W. (2014). The Database of Genomic Variants: A curated collection of structural variation in the human genome. Nucleic Acids Res. *42*, D986–D992.
29. Qiu, F., Xu, Y., Li, K., Li, Z., Liu, Y., DuanMu, H., Zhang, S., Li, Z., Chang, Z., Zhou, Y., et al. (2012). CNVD: Text mining-based copy number variation in disease database. Hum. Mutat. *33*, E2375–E2381.
30. Schweingruber, C., Rufener, S.C., Zünd, D., Yamashita, A., and Mühlemann, O. (2013). Nonsense-mediated mRNA

decay - Mechanisms of substrate mRNA recognition and degradation in mammalian cells. Biochim. Biophys. Acta *1829*, 612–623.

31. Jensen-Seaman, M.I., Furey, T.S., Payseur, B.A., Lu, Y., Roskin, K.M., Chen, C.-F., Thomas, M.A., Haussler, D., and Jacob, H.J. (2004). Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. *14*, 528–538.
32. Slatkin, M., and Rannala, B. (2000). Estimating allele age. Annu. Rev. Genomics Hum. Genet. *1*, 225–249.
33. Labuda, M., Labuda, D., Korab-Laskowska, M., Cole, D.E., Zietkiewicz, E., Weissenbach, J., Popowska, E., Pronicka, E., Root, A.W., and Glorieux, F.H. (1996). Linkage disequilibrium analysis in young populations: Pseudo-vitamin D-deficiency rickets and the founder effect in French Canadians. Am. J. Hum. Genet. *59*, 633–643.
34. Risch, N., de Leon, D., Ozelius, L., Kramer, P., Almasy, L., Singer, B., Fahn, S., Breakefield, X., and Bressman, S. (1995). Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. Nat. Genet. *9*, 152–159.
35. Vincensini, L., Blisnick, T., and Bastin, P. (2011). 1001 model organisms to study cilia and flagella. Biol. Cell *103*, 109–130.
36. Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., et al. (2010). TriTrypDB: A functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. *38*, D457–D462.
37. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.
38. Broadhead, R., Dawe, H.R., Farr, H., Griffiths, S., Hart, S.R., Portman, N., Shaw, M.K., Ginger, M.L., Gaskell, S.J., McKean, P.G., and Gull, K. (2006). Flagellar motility is required for the viability of the bloodstream trypanosome. Nature *440*, 224–227.
39. Subota, I., Julkowska, D., Vincensini, L., Reeg, N., Buisson, J., Blisnick, T., Huet, D., Perrot, S., Santi-Rocca, J., Duchateau, M., et al. (2014). Proteomic analysis of intact flagella of procyclic *Trypanosoma brucei* cells identifies novel flagellar proteins with unique sub-localization and dynamics. Mol. Cell. Proteomics *13*, 1769–1786.
40. Urbanska, P., Song, K., Joachimiak, E., Krzemien-Ojak, L., Koprowski, P., Hennessey, T., Jerka-Dziadosz, M., Fabczak, H., Gaertig, J., Nicastro, D., and Wloga, D. (2015). The CSC proteins FAP61 and FAP251 build the basal substructures of radial spoke 3 in cilia. Mol. Biol. Cell *26*, 1463–1475.
41. Bastin, P., Bagherzadeh, Z., Matthews, K.R., and Gull, K. (1996). A novel epitope tag system to study protein targeting and organelle biogenesis in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *77*, 235–239.
42. Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nat. Rev. Genet. *10*, 691–703.
43. Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., Löwer, J., Strätling, W.H., Löwer, R., and Schumann, G.G. (2012). The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res. *40*, 1666–1683.
44. Hancks, D.C., and Kazazian, H.H., Jr. (2016). Roles for retrotransposon insertions in human disease. Mob. DNA *7*, 9.
45. Brett, M., Korovesis, G., Lai, A.H.M., Lim, E.C.P., and Tan, E.-C. (2017). Intragenic multi-exon deletion in the *FBN1* gene in a child with mildly dilated aortic sinus: A retrotransposal event. J. Hum. Genet. *62*, 711–715.
46. Legoix, P., Sarkissian, H.D., Cazes, L., Giraud, S., Sor, F., Rouleau, G.A., Lenoir, G., Thomas, G., and Zucman-Rossi, J. (2000). Molecular characterization of germline NF2 gene rearrangements. Genomics *65*, 62–66.
47. Nakamura, Y., Murata, M., Takagi, Y., Kozuka, T., Nakata, Y., Hasebe, R., Takagi, A., Kitazawa, J., Shima, M., and Kojima, T. (2015). SVA retrotransposition in exon 6 of the coagulation factor IX gene causing severe hemophilia B. Int. J. Hematol. *102*, 134–139.
48. Takasu, M., Hayashi, R., Maruya, E., Ota, M., Imura, K., Kougo, K., Kobayashi, C., Saji, H., Ishikawa, Y., Asai, T., and Tokunaga, K. (2007). Deletion of entire *HLA-A* gene accompanied by an insertion of a retrotransposon. Tissue Antigens *70*, 144–150.
49. Vogt, J., Bengesser, K., Claes, K.B.M., Wimmer, K., Mautner, V.-F., van Minkelen, R., Legius, E., Brems, H., Upadhyaya, M., Högel, J., et al. (2014). SVA retrotransposon insertion-associated deletion represents a novel mutational mechanism underlying large genomic copy number changes with non-recurrent breakpoints. Genome Biol. *15*, R80.
50. Xiong, H., Wang, S., Kobayashi, K., Jiang, Y., Wang, J., Chang, X., Yuan, Y., Liu, J., Toda, T., Fukuyama, Y., and Wu, X. (2009). Fukutin gene retrotransposal insertion in a non-Japanese Fukuyama congenital muscular dystrophy (FCMD) patient. Am. J. Med. Genet. A. *149A*, 2403–2408.
51. Ishikawa, T. (2017). Axoneme structure from motile cilia. Cold Spring Harb. Perspect. Biol. *9*, a028076.
52. Lindemann, C.B., and Lesich, K.A. (2016). Functional anatomy of the mammalian sperm flagellum. Cytoskeleton *73*, 652–669.
53. Heuser, T., Dymek, E.E., Lin, J., Smith, E.F., and Nicastro, D. (2012). The CSC connects three major axonemal complexes involved in dynein regulation. Mol. Biol. Cell *23*, 3143–3155.