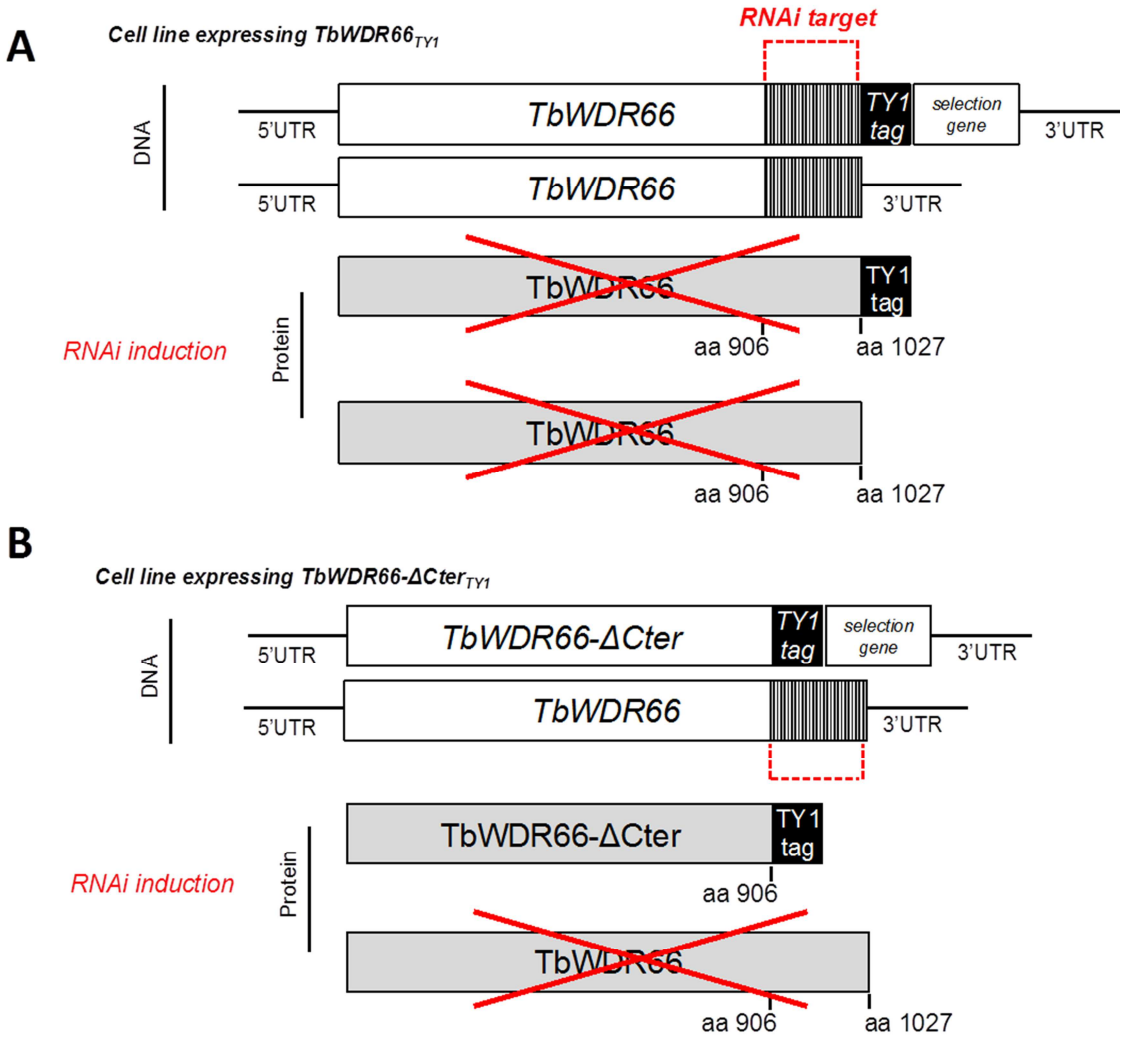
The American Journal of Human Genetics, Volume *103*

Supplemental Data

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

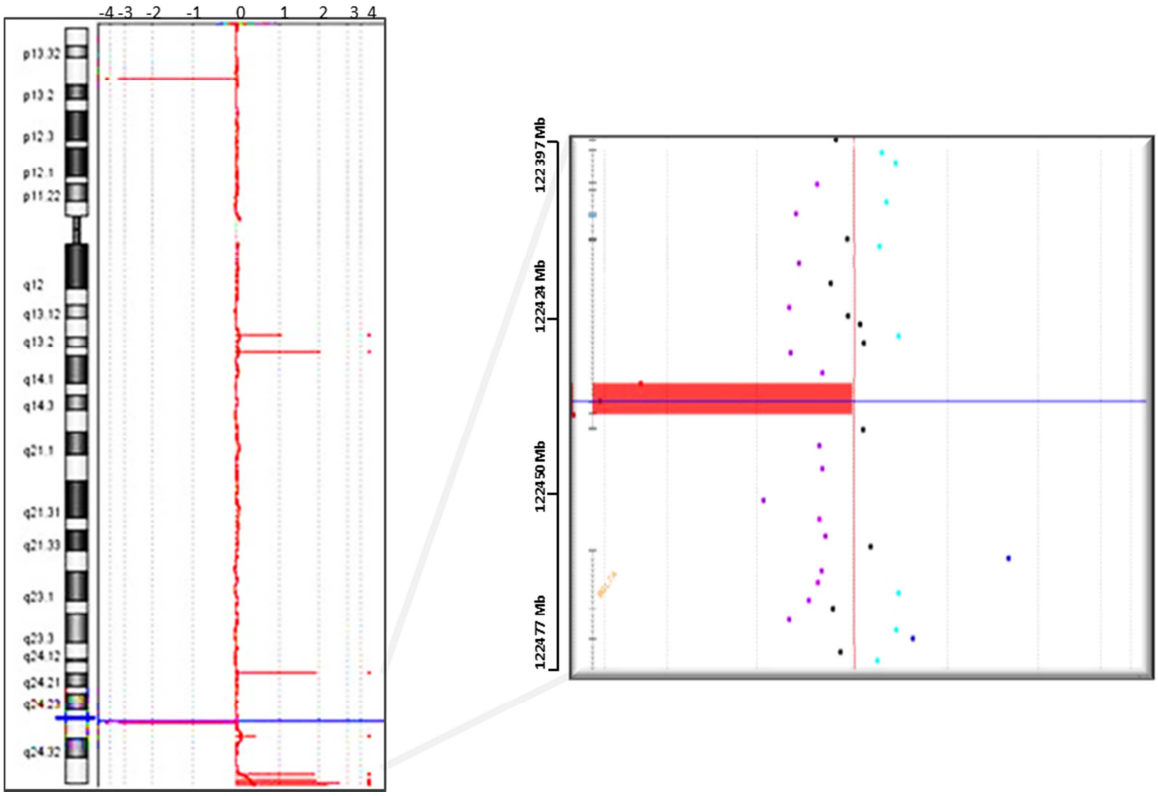
Zine-Eddine Kherraf, Amir Amiri-Yekta, Denis Dacheux, Thomas Karaouzène, Charles Coutton, Marie Christou-Kent, Guillaume Martinez, Nicolas Landrein, Pauline Le Tanno, Selima Fourati Ben Mustapha, Lazhar Halouani, Ouafi Marrakchi, Mounir Makni, Habib Latrous, Mahmoud Kharouf, Karin Pernet-Gallay, Hamid Gourabi, Derrick R. Robinson, Serge Crouzy, Michael Blum, Nicolas Thierry-Mieg, Aminata Touré, Raoudha Zouari, Christophe Arnoult, Mélanie Bonhivers, and Pierre F. Ray

Figure S1. Schematic representation of the RNAi and endogenous tagging strategies.



A. One allele out of two is replaced by a PCR product allowing the insertion of the TY1 tag and a selection gene. The RNAi will target both alleles, and upon induction, both WT and TY1-tagged protein will be knocked down. **B.** One allele out of two is replaced by a PCR product allowing the shortening of *TbWDR66* ORF, the insertion of the C-terminal TY1 tag and a selection gene. The RNAi targets only the WT allele, and as a consequence, only the WT protein will be knocked down, whilst TbWDR66-ΔCter$_{TY1}$ expression is not affected.

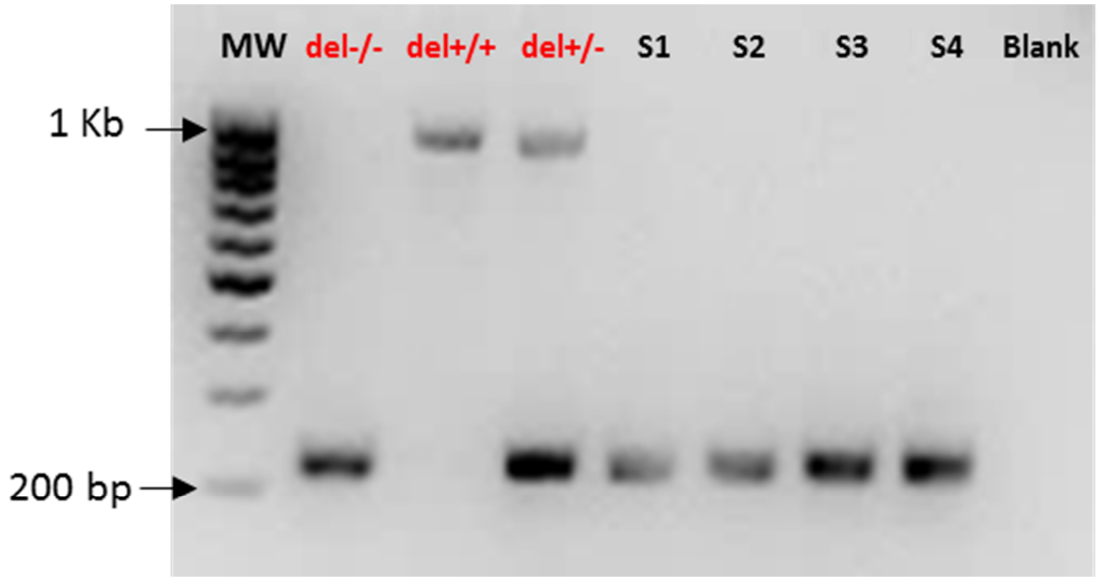
Figure S2: Array-CGH profile of chromosome 12 showing a large homozygous deletion of *WDR66* gene in the patient P1.



Each spot represents an oligonucleotide on the array. Spots are graphically represented according to their corresponding position along the chromosome. Values indicated on the top are expressed as log2 ratios of fluorescence intensity. We observe that three spots are deviated and have ratios between -2.358 and -4.621, indicating that the patient has a homozygous deletion in the *WDR66* gene with an estimated size ranging from 4,735 to 8,304 bp encompassing exons 20 and 21 of the *WDR66* gene. This result specified the location of the breakpoints around the region defined by the coordinates chr12:122,434,940-122,439,675 (GRCh37/hg19).

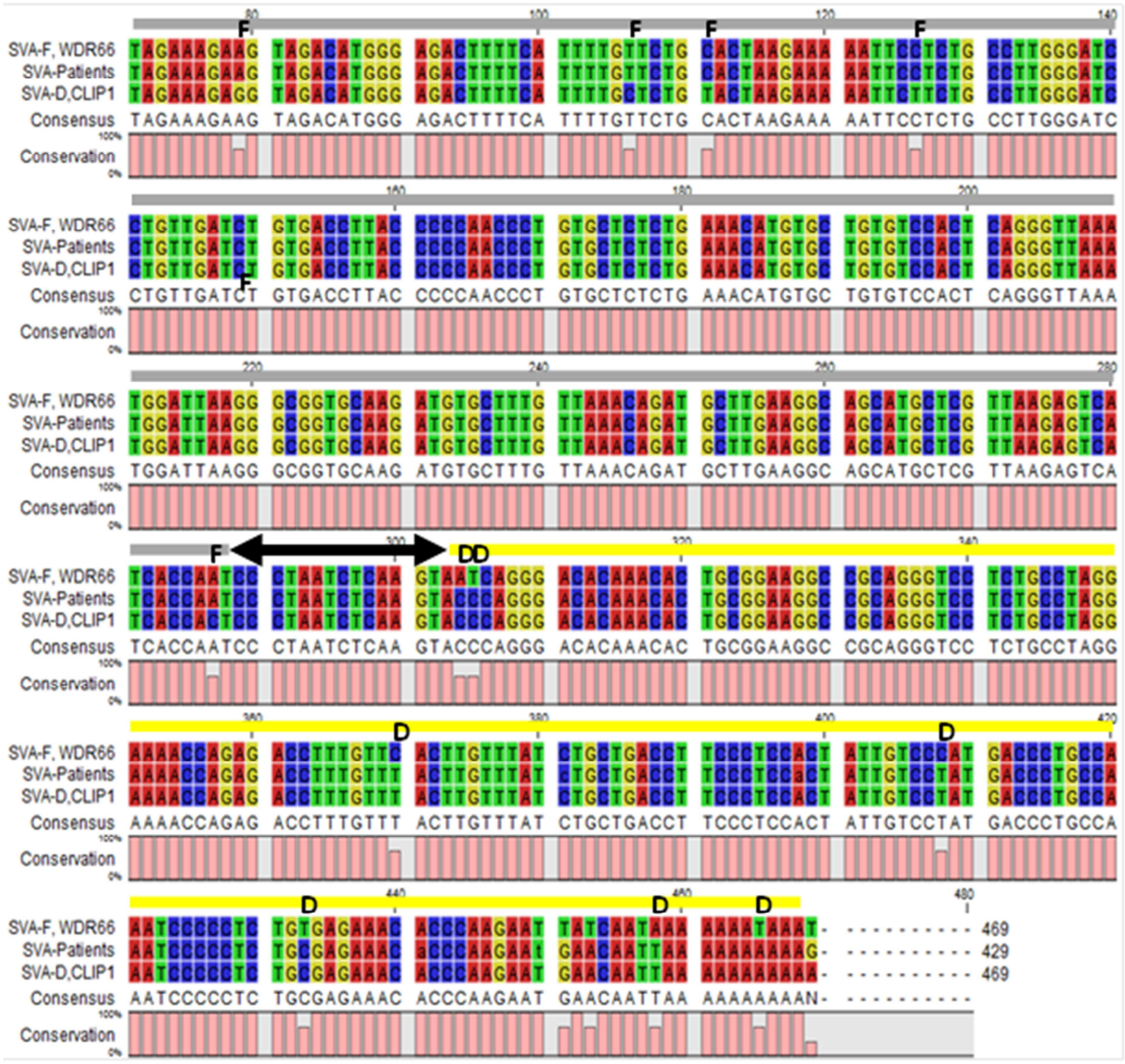
CNV analysis was conducted using an Agilent high-density oligonucleotide-based CGH micro-arrays (kit SurePrint G3 Human CGH Microarray Kit 1x1M Agilent Technologies®, Santa Clara, CA) following the manufacturer's instructions. The average spacing of the probes is of 2.1 kb and length of the oligonucleotides is 60 mer. DNA from the individual P7 was compared with DNA from a male control subject. One microgram of DNA from each DNA sample was added to the digestion master mix containing the enzymes AluI (10 UI/µl) and RsaI (10 UI/µl), albumin (Acetylated BSA). After digestion, genomic DNA was labeled according to the random priming method with nucleic acids bound to fluorescent Cyanine-5dUTP for the individual and to Cyanine-3dUTP for the control. Labelled DNA was then purified on Agilent® columns; the two samples were pooled and mixed with a blocking agent, a hybridization buffer and Human ADN Cot-1 (1.0 mg/ml). The mix was loaded on the array and hybridized for 40 h at 67°C. The micro-array was then washed twice and scanned using an Agilent Scanner Control™ software. Results were extracted using the Feature Extraction 9.5.3.1 software and then interpreted with the Cytogenomics software 3.02.11 (Agilent®), according to the default analysis method parameters: aberration algorithm ADM-2, threshold 6.0, fuzzy zero, centralization and moving average window 0.5 Mb. Coordinates of all variations or probes are based on the UCSC GRCh37/hg19 assembly.

Figure S3: gel electrophoresis illustrating the duplex PCR used to screen 100 North-African individuals for WDR66 deletion



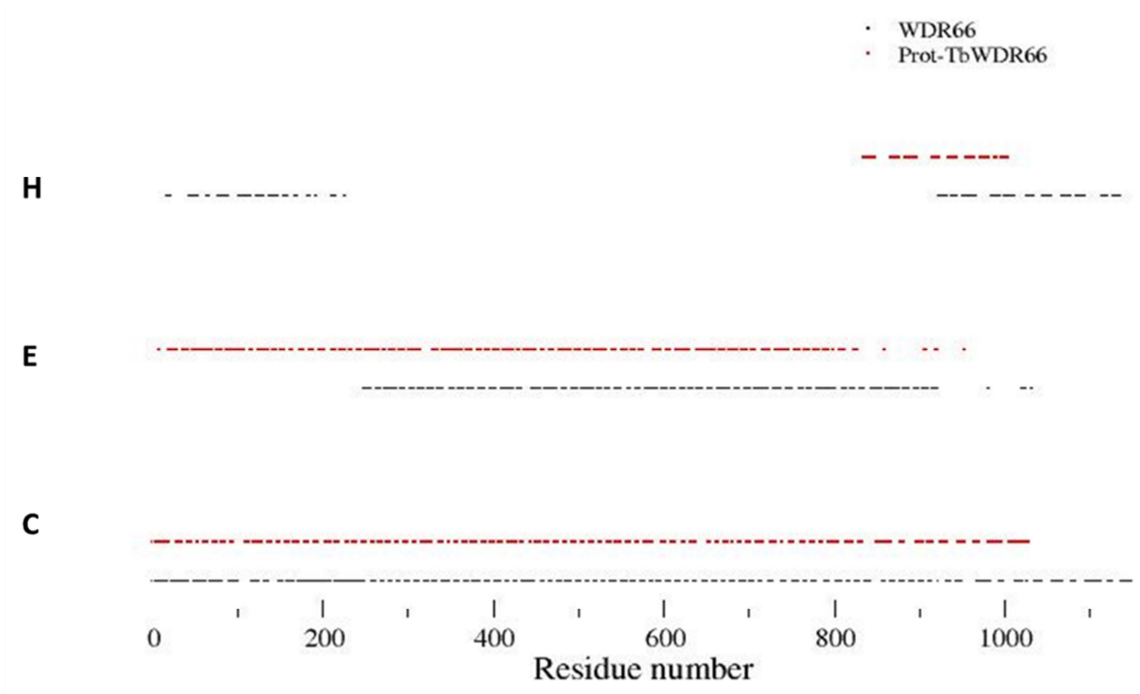
The primer used permit to amplify a 220 nucleotide sequence located within the *WDR66* deletion hence present only in non-deleted alleles and a 950 nucleotide sequence encompassing the deletion breakpoint hence only present in the deleted alleles. In lane 1 (-/-) the amplified DNA was extracted from a non-deleted control, in lane 2 (+/+) from the homozygous deleted patient 1. Lane 3 (+/-) was obtained from mixing deleted and non-deleted DNAs. S1-S4 corresponds to the first 4 tested North African subjects. MW= molecular weight.

Figures S4: Alignment of the SVA-F/D sequence encompassing the centromeric breakpoint in patients with the SVA-F and SVA-D sequences in WDR66 (intron 19) and CLIP1 (intron 5) respectively.



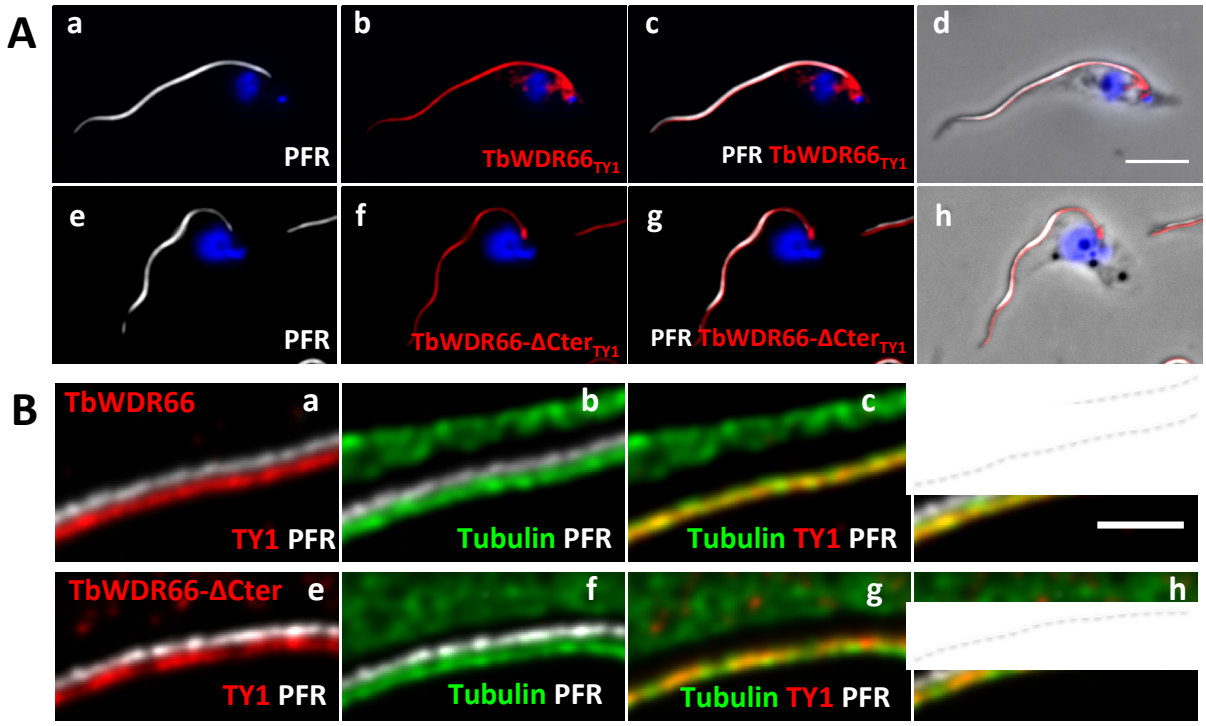
Alignment was performed using CLC sequence viewer. D indicates the variants identical between the patients and the *CLIP1* SVA-D sequence, whereas F indicates the variants identical between the patient and the WDR66 SVA-F sequence. Results show that in patients, the telomeric part of the recombined SVA is homologous to the *CLIP1* SVA-D (yellow line) whereas the centromeric SVA sequence corresponds to the original*WDR66* SVA-F intron 19 (grey line). The 15 nucleotides indicated between the D and the F sequence (double black arrow) corresponds to the minimal region localizing the centromeric breakpoint. The telomeric breakpoint is indicated by a black vertical arrow. In deleted men, the sequence following the telomeric breakpoint corresponds to WDR66 intron 21.

Figure S5. Secondary structure prediction of human WDR66 and Trypanosoma TbWDR66.



Elements of the secondary structures of WDR66 and tbWDR66 were predicted by Porter 4.0. C, E and H correspond to coils, extended conformations (β-sheets) and helices (α-helices), respectively.

Figure S6. TbWDR66$_{TY1}$ and TbWDR66-ΔCter$_{TY1}$ are axoneme proteins in PCF *Trypanosoma*.



A. Epifluorescence analysis of immunolabeled PCF cells expressing TbWDR66TY1 (a-d) and TbWDR66-ΔCter$_{TY1}$ (e-h) using anti-PFR2 (grey) that labels the paraflagellar rod and anti-TY1 (red) antibodies. Scale bar represents 5 µm. **B.** Deconvolved images of STED confocal microscopy experiments on PCF cells expressing TbWDR66$_{TY1}$ (a-d) or TbWDR66-ΔCter$_{TY1}$ (e-h) and immuno-labeled with anti-PFR2 (grey), anti-TY1 (red), and anti-tubulin (green) that labels the axonemal microtubules as well as the subpellicular microtubules (delimitated by the dotted white lines). Scale bar represents 0.5 µm.

Table S1. List of primers used for qRT-PCR experiment on human sperm extracts.

Primer name	Primer sequence (5'-3')	Tm (°C)	Product length (bp)
GAPDH-F	ACCCACTCCTCCACCTTTGA	60	101
GAPDH-R	CTGTTGCTGTAGCCAAATTCGT	60	
WDR66-T1F	AGCTTTGAGGTGCTCGGTTA	60	132
WDR66-T1R	CAGTGAAGCAAAGCAATCCA	60	
WDR66-T2F	GGTGCAGTGGAAAATCACCT	60	139
WDR66-T2R	CTGTGAATCCCAGCTCCACT	60	

Transcripts' names are listed in Ensembl database (assembly GRCh37/hg19). Their respective transcript ID is ENST00000288912.4 (T1) and ENST00000397454.2 (T2).

Table S2. Antibodies used for immunofluorescence experiments on human sperm cells.

Primary antibodies	Source	Type	Dilution	Reference
WDR66	Rabbit	Polyclonal	1/20	Sigma-Aldrich- HPA040005
α-tubulin	Mouse	Monoclonal	1/200	Sigma-Aldrich-T7451

Table S3. List of primers used for multiplex PCR to amplify differentially wild-type and mutated allele and for deleted allele sequencing.

Primer name	Primer sequence (5'-3')	Tm (°C)	Product length (bp)
WDR66-Ex20F	AGCTTTGAGGTGCTCGGTTA	59	220 (in controls)
WDR66-Int20R	CCTCATGGCTCAAATCTGGT	58	
WDR66-Int19F (F1)	GAATAGAAAGGCGGGAAAGG	56	950 (in mutated patients)
WDR66-Int22R (R1)	AACTCTGGGTGGAAATTGCTTA	58	

Table S4. Endogenous tagging and RNAi primer sequences for *T. brucei* experiments.

***TbWDR66* RNAi**	**1036**	5'-GCCGGCCGCTCTAGAtgtgatgtgagcagcattccg-3'
	1037	5'-TAAGCTTGCTCTAGAtttcctcttccccaccgtca-3'
***TbWDR66*$_{TY1}$ endogenous tagging**	**1038**	5'-CGGAAAATCTTCTTGGATTTGAAGATTATGATGCTATGGCCCAGGGTGATGACGGTGGGGAAGAGGAAATGAGTTTGCAGggttctggtagtggttcc-3'
	1039	5'-GAGAGAAAACGAGCATTCCACTGATGGTTTGGTACCTTTAGTCTCCAAATTCACCTGCCCATAACTCTCTGTTCCCTCACccaatttgagagacctgtgc-3'
***TbWDR66-ΔCter*$_{TY1}$ endogenous tagging**	**1039**	5'-GAGAGAAAACGAGCATTCCACTGATGGTTTGGTACCTTTAGTCTCCAAATTCACCTGCCCATAACTCTCTGTTCCCTCACccaatttgagagacctgtgc-3'
	1040	5'-CGGAAAATCTTCTTGGATTTGAAGATTATGATGCTATGGCCCAGGGTGATGACGGTGGGGAAGAGGAAATGAGTTTGCAGggttctggtagtggttcc-3'

Table S5. List of primary and secondary antibodies used in immunofluorescence experiments in *Trypanosoma*.

Antibodies	References	Species	Protein localization	Dilution
TAT1 (tubulin)	Described in (Woods et al., 1989)	Mouse IgG2a	Cytoskeleton/flagellum	1/50
PFR2	Gift from Dr N. Biteau*	Rabbit	PFR	1/400
BB2 (TY1)	Described in (Bastin et al., 1996)	Mouse IgG1	TY1-tagged protein	1/500
Anti-mouse IgG2a Alexa-488	Invitrogen A21131	Goat		1/100
Anti-mouse IgG1 Alexa-594	Invitrogen A21125	Goat		1/100
Anti-rabbit ATTO-N647	Sigma 40839	Goat		1/100

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.

Woods, A., Sherwin, T., Sasse, R., MacRae, T.H., Baines, A.J., and Gull, K. (1989). Definition of individual components within the cytoskeleton of Trypanosoma brucei by a library of monoclonal antibodies. J. Cell Sci. *93*, 491–500.

* MFP, University of Bordeaux, CNRS UMR 5234

Table S6. List of all chromosome 12 SNVs obtained by WES of all deleted subjects

The smallest common region of homozygosity was establish by taking into account the SNPs common for at least 6 of the 7 analysed patients (P1-P7) that are different from the control individual (C). The region taken into consideration is surrounded by a black box. The extend of the WDR66 gene is highlighted in yellow. See the excel file **"Table S6"**

Table S7. Protein folds found in the 2 target sequences

Protein	Family	Domains (aa number)	E-value
WDR66	Quinoprotein alcohol dehydrogenase-like (β)	249-407, 460-486, 601-774	$1.26e^{-33}$
	WD40 repeat-like (β)	691-922	$1.83e^{-16}$
	EF-hand (α)	980-1099	$3.33e^{-7}$
TbWDR66	WD40 repeat-like (β)	218-372, 458-630	$2.75e^{-40}$
	YVTN repeat-like/Quinoprotein amine dehydrogenase (β)	20-94, 148-319	$1.65e^{-8}$
	WD40 repeat-like (β)	591-631, 699-796	$6.59e^{-6}$
	EF-hand (α)	863-981	$2.24e^{-2}$

WD40 folds and other β-domains are found in the two proteins by program Superfamily with very good confidence (e-value corresponds to the probability of finding such folds by chance in the sequence). Both proteins also share α-EF-Hand domains in their C-Terminal region.

Table S8. Average semen parameters in the different genotype groups for the 78 MMAF subjects.

Semen parameters	MMAF *WDR66* patients n=7	MMAF *CFAP43* patients n=10	MMAF *CFAP44* patients n=6	MMAF *DNAH1* patients n=6	MMAF all mutations n=29	MMAF with unknown causes n=49	Overall MMAF patients n=78
Mean age (years)	44.7 ± 10.6 (n'=7)	37.7 ± 9.6 (n'=9)	41.3 ± 4.3 (n'=6)	41.5 ± 4.4 (n'=6)	41.0 ± 8.4 (n'=28)	41.9 ± 7.3 (n'=49)	41.6 ± 7.7 (n'=77)
Sperm volume (ml)	3.0 ± 1.1 (n'=7)	3.5 ± 1.4 (n'=8)	3.2 ± 0.87 (n'=6)	3.5 ± 1.2 (n'=6)	3.3 ± 1.1 (n'=28)	3.6 ± 1.5 (n'=48)	3.5 ± 1.4 (n'=75)
Sperm concentration (10^6/ml)	30.3 ± 23.5 (n'=7)	27.2 ± 23.4 (n'=8)	7.9 ± 8.4 (n'=6)	22.9 ± 15.2 (n'=6)	22.3 ± 20.4 (n'= 28)	27.2 ± 37.2 (n'=48)	25.6 ± 32.1 (n'=75)
Motility (a+b+c) 1 h	3.1 ± 1.9 (n'=7)	0 ± 0 (n'=9)	0 ± 0 (n'=6)	2.6 ± 4.2 (n'=6)	1.3 ± 2.5 (n=29)	5.3 ± 6.4 (n'=48)	3.9 ± 5.6 (n'=76)
Vitality	65.6 ± 12.6 (n'=6)	55.5 ± 24 (n'=8)	43.3 ± 22.6 (n'=6)	51.2 ± 23.1 (n'=5)	54.5 ± 21.2 (n'=26)	51.9 ± 20.2 (n'=47)	52.7 ± 20 (n'=72)
Normal spermatozoa	0 (n'=7)	1.25 ± 3.5 (n'=8)	0 ± 0 (n'=6)	0 ± 0 (n'=6)	0.4 ± 1.9 (n'=27)	2.3 ± 4.3 (n'=48)	1.6 ± 2.7 (n'=61)
Abnormal acrosome	71.8 ± 15.3 (n'=6)	41.1 ± 29.8 (n'=6)	85.2 ± 7.9 (n'=5)	47.0 ± 31.3 (n'=5)	60.8 ± 28.3 (n'=22)	63.6 ± 25.8 (n'=44)	62.6 ± 26.5 (n'=66)
Abnormal head morphology	24.1 ± 25.1 (n'=6)	8.0 ± 3.5 (n'=5)	54.6 ± 27.8 (n'=5)	45.0 ± 55.6 (n'=4)	31.9 ± 33.7 (n'=20)	32.3 ± 21.7 (n'=43)	32.2 ± 25.8 (n'=63)
Abnormal base	14.8 ± 15.6 (n'=6)	22.2 ± 14.5 (n'=5)	54.0 ± 29.7 (n'=5)	26.5 ± 22.3 (n'=4)	28.8 ± 24.8 (n'=20)	32.4 ± 19.2 (n'=42)	31.2 ± 21.0 (n'=62)
Absent flagella	27.2 ± 13.1 (n'=6)	21.8 ± 17.6 (n'=5)	36.8 ± 4.1 (n'=5)	25.6 ± 15.9 (n'=5)	26.9 ± 14.3 (n'=22)	17.3 ± 15.5 (n'=45)	20.7 ± 15.7 (n'=66)
Short Flagella	38.6 ± 11.8 (n'=6)	65.3 ± 31.7 (n'=8)	52.2 ± 27 (n'=6)	49.8 ± 24.3 (n'=5)	53.3 ± 25.2 (n'=26)	38.6 ± 27.0 (n'=47)	43.7 ± 27.3 (n'=72)
Coiled Flagella	21.1 ± 8.3 (n'=6)	8.2 ± 6 (n'=6)	14.4 ± 7 (n'=5)	9 ± 6.3 (n'=5)	12.8 ± 9.0 (n'=23)	13.0 ± 9.7 (n'=47)	12.8 ± 9.4 (n'=69)
Flagella of irregular caliber	43.3 ± 22.1 (n'=6)	20.2 ± 19.3 (n'=5)	28.4 ± 16.9 (n'=5)	35 ± 22.7 (n'=5)	32.0 ± 21.8 (n'=22)	31.3 ± 26.6 (n'=46)	31.7 ± 25.1 (n'=67)
Multiple anomalies index	2.6 ± 0.4 (n'=6)	2.3 ± 0.2 (n'=4)	3.4 ± 0.4 (n'=5)	2.3 ± 1.4 (n'=5)	2.6 ± 0.8 (n'=20)	2.7 ± 0.6 (n'=41)	2.7 ± 0.7 (n'=61)

Values are expressed in percent, unless specified otherwise.Values are mean ± SD; n= total number of patients in each group; n'= number of patients used to calculate the average based on available data. We compared statistical differences between MAAF due to *WDR66*, *CFAP43*, *CFAP44* and *DNAH1* mutations versus MAAF due to uncharacterized genetic cause. No significant statistical difference was observed between the two groups.