The American Journal of Human Genetics, Volume *102*

Supplemental Data

Absence of *CFAP69* Causes Male Infertility due to Multiple Morphological Abnormalities of the Flagella in Human and Mouse

Frederick N. Dong, Amir Amiri-Yekta, Guillaume Martinez, Antoine Saut, Julie Tek, Laurence Stouvenel, Patrick Lorès, Thomas Karaouzène, Nicolas Thierry-Mieg, Véronique Satre, Sophie Brouillet, Abbas Daneshipour, Seyedeh Hanieh Hosseini, Mélanie Bonhivers, Hamid Gourabi, Emmanuel Dulioust, Christophe Arnoult, Aminata Touré, Pierre F. Ray, Haiqing Zhao, and Charles Coutton

Supplemental data

Supplemental Material and Methods

Whole-Exome Sequencing (WES) and bioinformatic analysis

Genomic DNA was isolated from saliva using the Oragen DNA extraction kit (DNAgenotech®, Ottawa, Canada). Coding regions and intron/exon boundaries were enriched using the “all Exon V5 kit“ (Agilent Technologies, Wokingham, UK). DNA sequencing was undertaken at the Genoscope, Evry, France, on the HiSeq 2000 from Illumina®. Sequence reads were aligned to the reference genome (hg19) using MAGIC[1]. MAGIC produces quality-adjusted variant and reference read counts on each strand at every covered position in the genome. Duplicate reads and reads that mapped to multiple locations in the genome were excluded from further analysis. Positions whose sequence coverage was below 10 on either the forward or reverse strand were marked as low confidence, and positions whose coverage was below 10 on both strands were excluded. Single nucleotide variations (SNV) and small insertions/deletions (indels) were identified and quality-filtered using in-house scripts. Briefly, for each variant, independent calls were made on each strand, and only positions where both calls agreed were retained. The most promising candidate variants were identified using an in-house bioinformatics pipeline, as follows. Variants with a minor allele frequency greater than 5% in the NHLBI ESP6500 [Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA] or in 1000 Genomes Project phase 1 datasets[2], or greater than 1% in ExAC[3] or gnomAD (http://gnomad.broadinstitute.org/) were discarded. We also compared these variants to an in-house database of 94 control exomes obtained from subjects from North Africa (n=60) and the Middle East (n=34), corresponding to the geographical origin of most individuals in this study and which is under-represented in public SNP databases. All these control subjects presented a sperm phenotype clearly different (azoospermia) from those of the MMAF patients. All variants present in homozygous state in this database were excluded. We used Variant Effect Predictor (VEP version 81[4]) to predict the impact of the selected variants. We only retained variants impacting splice donor or acceptor sites or causing frameshifts, in-frame insertions or deletions, stop gain, stop loss or missense variants except those scored as "tolerated" by SIFT[5] (sift.jcvi.org) and as "benign" by Polyphen-2[6] (genetics.bwh.harvard.edu/pph2). To predict the impact of mutations within the 5′ and 3′ splicing consensus, we used the Human Splicing Finder server v 3.0. All steps from sequence mapping to variant selection were performed using the ExSQLibur pipeline.

Supplemental figures

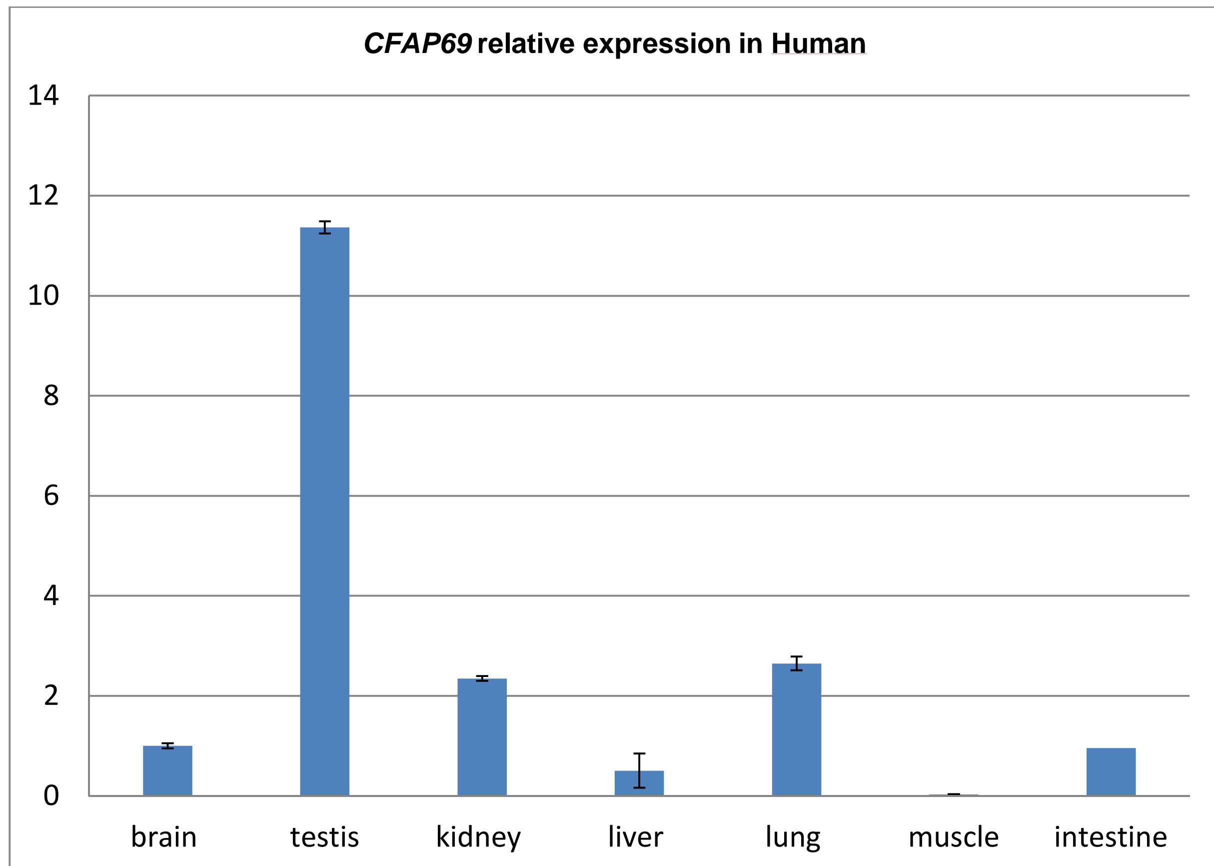

Figure S1. Relative mRNA Expression of human *CFAP69* transcripts. *CFAP69* mRNA levels in a panel of human normal tissues. Results are presented as the mean of triplicates (ratio target gene/ACTB) ± Standard Deviation (SD). RT-qPCR data were normalized using the reference gene ACTB with the -ΔΔCt method. Brain expression is arbitrary set to 1. In human, *CFAP69* has the strongest expression in testis compared to other organs. Unpaired t-test, ***P< 0.001.

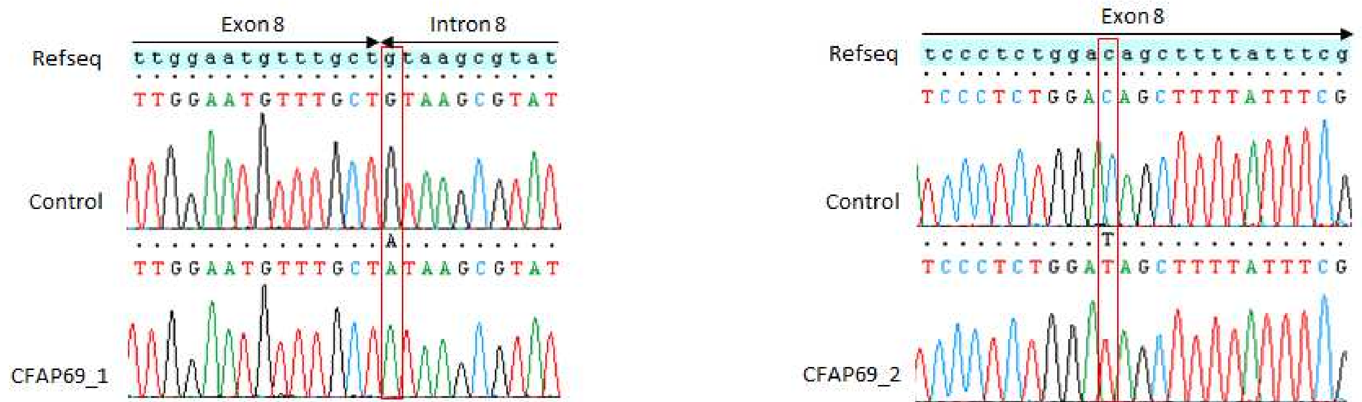

Figure S2. Electropherograms of Sanger sequencing for the two CFAP69-mutated individuals compared to reference sequence.

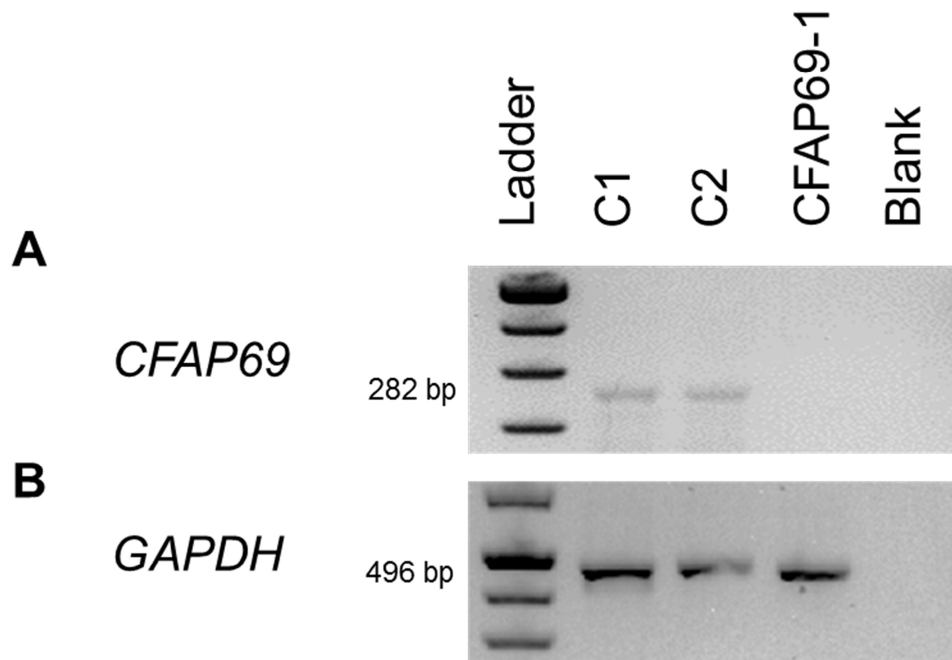

Figure S3. RT-PCR analyses on peripheral whole blood cells from CFAP69_1 patient showing mRNA decay. RT–PCR analysis of CFAP69_1 patients with the c.860+1G>A and control subjects from the general population (C1 and C2). (A) Electrophoresis showing the RT–PCR amplification of CFAP69 exons 12-13. C1 and C2 controls yield a normal fragment of 282 bp, whereas patient CFAP69_1 shows no amplification. There is no amplification from the RT-negative control (Blank). (B) Electropherogram showing the amplification of the same cDNAs amplified with GAPDH primers. Bands of equivalent intensity are obtained from all samples except the RT-negative control (Blank).

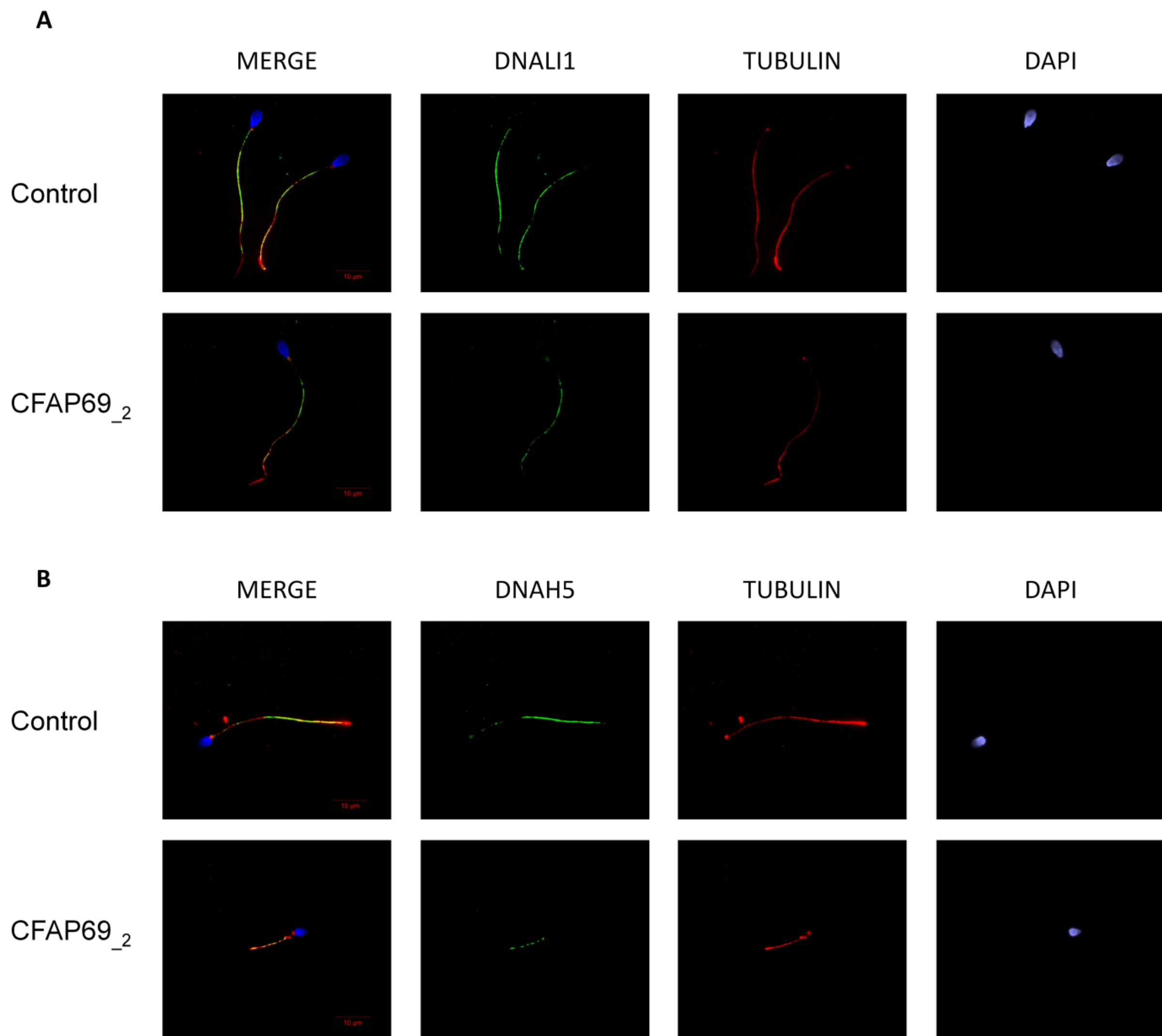

Figure S4. Axonemal inner and outer dynein arms are not affected by the absence of CFAP69. (A) Sperm cells from fertile controls and CFAP69$_{_2}$ stained with anti-DNALI1 (Green), which detects a protein located in the inner dynein arm, and anti-acetylated tubulin (red) antibodies. DNA was counterstained with Hoechst 33342. (B) Sperm cells from a fertile control and CFAP69$_{_2}$ stained with anti-DNAH5 (Green), which detects a protein located in the outer dynein arm, and anti-acetylated tubulin (red) antibodies. DNA was counterstained with Hoechst 33342. Immunostaining for DNALI1, DNAH5 were comparable with controls, suggesting that outer dynein arms (ODAs) and inner dynein arms (IDAs) respectively were not directly affected by mutations in *CFAP69*. Scale bars 10µm.

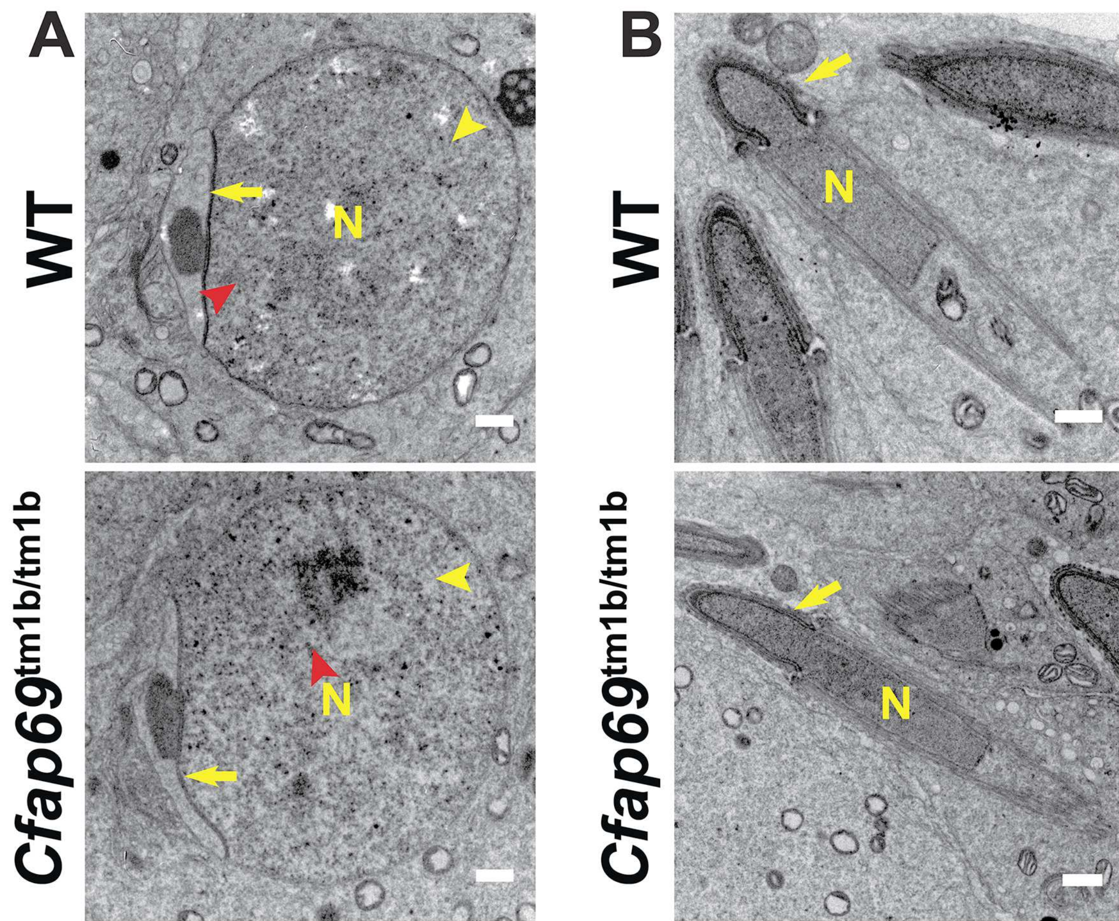

Figure S5. Normal acrosome development and nuclear elongation can be observed in testes of *Cfap69* KO mice. Transmission electron micrographs of spermatids during spermiogenesis. Yellow arrows indicate the acrosome, and yellow “N” indicates the nucleus. Scale bars: C and D, 1 μm.

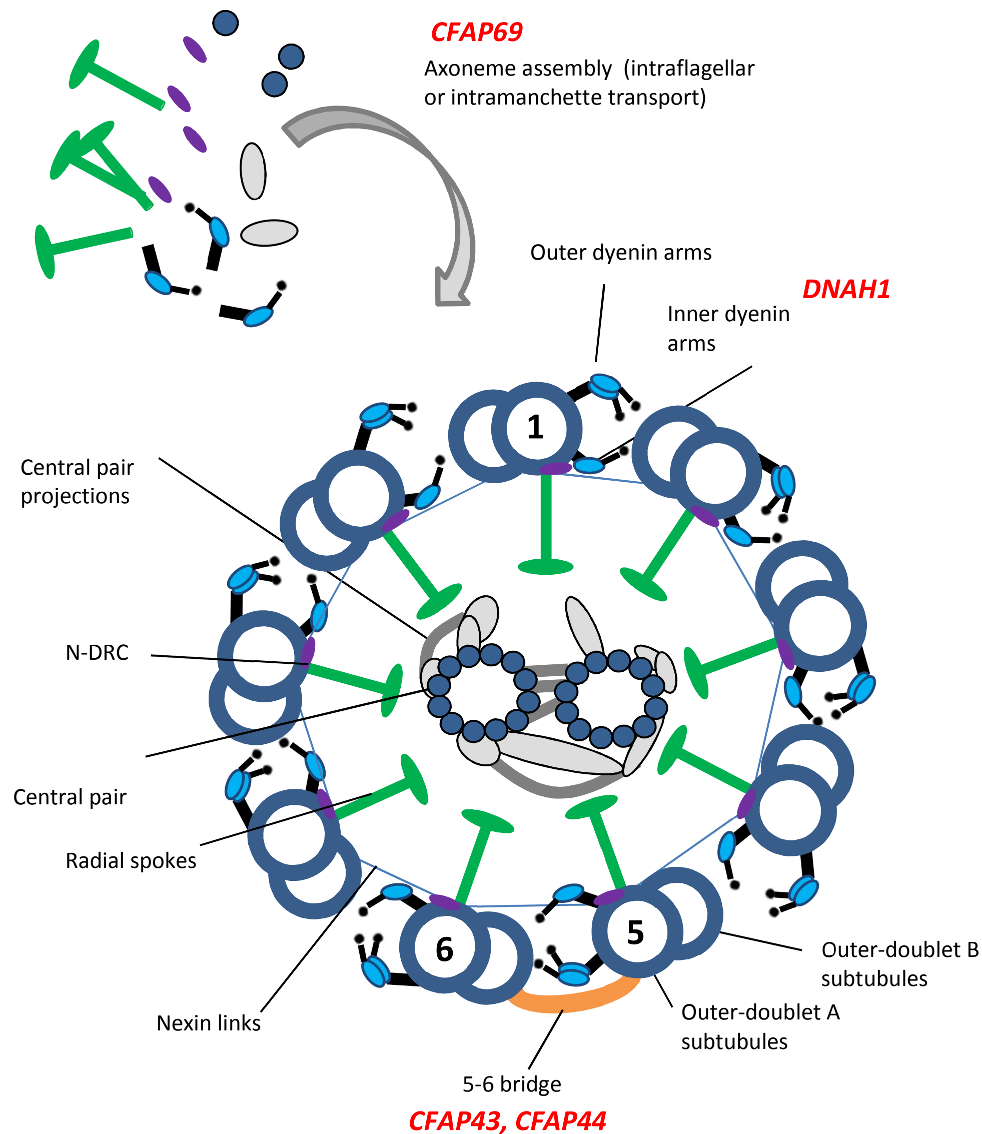

Figure S6. Schematic cross-section of human sperm flagellar axoneme and localization of MMAF-related proteins. The outer and inner dynein arms, nexin-dynein regulator complex (N-DRC) and radial spokes are attached to the nine outer doublet microtubules. The 5-6 bridge likely links the outer doublet microtubules 5-6 and potentially with other extra-axonemal structures. Genes formally identified for the MMAF phenotype in human are reported in red.

Supplementary Tables

Table S1. Average semen parameters in different genotype groups for the 78 included MMAF subjects in the present study.

Semen parameters	**MMAF with *CFAP69* mutation** n=2	**MMAF with other mutations**[a] n=22	**Overall MMAF** n=78
Mean age (years)	46.5 ± 6.3 (n'=2)	39.8 ± 7 (n'=21)	41.6 ± 7.7 (n'=77)
Sperm volume (ml)	4.5 ± 0.7 (n'=2)	3.4 ± 1.2 (n'=20)	3.5 ± 1.4 (n'=75)
Sperm concentration (10^6/ml)	5 ± 1.4 (n'=2)	20.1 ± 18.8 (n'=20)	25.6 ± 32.1 (n'=75)
Total motility 1 h	5.5 ± 6.3 (n'=2)	0.7 ± 2.4 (n'=21)	3.9 ± 5.6 (n'=76)
Vitality	37.5 ± 34.6 (n'=2)	50.5 ± 22.7 (n'=19)	52.7 ± 20 (n'=72)
Normal spermatozoa	6 ± 8.5 (n'=2)	0.5 ± 2.3 (n'=20)	1.6 ± 2.7 (n'=61)
Absent flagella	6.5 ± 7.8 (n'=2)	28.1 ± 14.4 (n'=15)	20.7 ± 15.7 (n'=66)
Short Flagella	46 ± 46.7 (n'=2)	57.1 ± 27.9 (n'=19)	43.7 ± 27.3 (n'=72)
Coiled Flagella	4 ± 4.2 (n'=2)	10.4 ± 6.6 (n'=16)	12.8 ± 9.4 (n'=69)
Bent Flagella	2.5 ± 3.5 (n'=2)	8.7 ± 5.8 (n'=6)	4.2 ± 8.4 (n'=26)
Flagella of irregular caliber	4 ± 4.2 (n'=2)	27.9 ± 19.4 (n'=15)	31.7 ± 25.1 (n'=67)
Tapered head	10.5 ± 13.4 (n'=2)	22.5 ± 29 (n'=13)	16.5 ± 20.2 (n'=68)

Thin head	25.5 ± 16.2 (n’=2)	10.8 ± 13.7 (n’=13)	11.1 ± 13.4 (n’=65)
Microcephalic	0.5 ± 0.7 (n’=2)	3.8 ± 2.4 (n’=14)	4.3 ± 4.8 (n’=67)
Macrocephalic	0 (n’=2)	0.15 ± 0.5 (n’=13)	0.6 ± 1.8 (n’=66)
Mutliple heads	0 (n’=2)	1.9 ± 4 (n’=15)	1.9 ± 3.7 (n’=67)
Abnormal base	21 ± 8.5 (n’=2)	36.3 ± 26.4 (n’=13)	31.2 ± 20.8 (n’=64)
Abnormal acrosomal region	67 ± 4.2 (n’=2)	58.3 ± 31.5 (n’=15)	61.9 ± 26.7 (n’=68)

[a] Other mutations correspond to individuals mutated in CFAP43, CFAP44 and DNAH1[7]. Values are percentages unless specified otherwise. Values are mean +/- SD; n= total number of individuals in each group; n’= number of individuals used to calculate the average based on available data.

Table S2. Primer sequences used for Sanger sequencing verification of *CFAP69* variations and respective melting temperatures (Tm).

Primer names	Primer sequences (5'-3')	Tm
CFAP69-Ex8F	AAAAATGTCAATATTGTAAAGCACAAA	58°C
CFAP69-Int8R	TGTGGCTTGTTATTGTGCAG	

Table S3. Primers used for RT-qPCR of *CFAP69* in human.

Primer names	Primer sequences (5'-3')	Tm
CFAP69-RTqPCR-Ex12F	ATTGACTGGTCTGCAGCACA	60°C
CFAP69-RTqPCR-Ex13R	ACTGTAACGCATCTGGGCAA	
GAPDH-RTqPCR-F	AGCCACATCGCTCAGACAC	60°C
GAPDH-RTqPCR-R	GCCCAATACGACCAAATCC	

Table S4. Primer sequences used in human *CFAP69* RT-PCR and respective melting temperatures (Tm).

Primer names	Primer sequences (5'-3')	Tm
CFAP69-RT-Ex7F	TTCTGCAGCATCTCTCAACTTC	57°C
CFAP69-RT-Ex10R	CAAATCCTTGGTAAAGCCACA	

Table S5. All *CFAP69 (C7orf63)* variations identified by WES.

Gene	Patients	Variant coordinates	Canonical Transcripts	cDNA Variations	Amino acid variations	Exon	Nationality	Allelic status
CFAP69	CFAP69$_{_1}$	chr7:89901273	ENST00000389297	c.860+1G>A	splice_donor_variant	8	Iranian	Homozygous
CFAP69	CFAP69$_{_2}$	chr7:89901175	ENST00000389297	c.763C>T	p.Gln255Ter	8	Tunisian	Homozygous

Supplemental References

1. SEQC/MAQC-III Consortium (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat. Biotechnol. *32*, 903–914.

2. 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.

3. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

4. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122.

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

6. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.

7. 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.