**The American Journal of Human Genetics, Volume *108***

## Supplemental Data

# Deleterious variants in X-linked *CFAP47* induce asthenoteratozoospermia and primary male infertility

**Chunyu Liu, Chaofeng Tu, Lingbo Wang, Huan Wu, Brendan J. Houston, Francesco K. Mastrorosa, Wen Zhang, Ying Shen, Jiaxiong Wang, Shixiong Tian, Lanlan Meng, Jiangshan Cong, Shenmin Yang, Yiwen Jiang, Shuyan Tang, Yuyan Zeng, Mingrong Lv, Ge Lin, Jinsong Li, Hexige Saiyin, Xiaojin He, Li Jin, Aminata Touré, Pierre F. Ray, Joris A. Veltman, Qinghua Shi, Moira K. O'Bryan, Yunxia Cao, Yue-Qiu Tan, and Feng Zhang**

## Supplemental Note: Case Reports

All of these male individuals displayed typical MMAF phenotypes (a combination of sperm flagellar defects, including absent, short, bent, coiled, and/or irregular-caliber flagella) with no obvious primary ciliary dyskinesia related symptoms such as sinusitis, bronchitis, pneumonia and otitis media.[1] Moreover, clinical investigations indicated that the development of male external genitalia, bilateral testicular sizes, hormone levels, and secondary sexual characteristics were normal in all of the MMAF-affected men in this study. Men with abnormalities in chromosomal karyotypes or microdeletions on the Y chromosome were also excluded. The study regarding the cohorts was approved by the institutional review boards at all the participating institutes, and signed informed consents were obtained from all subjects participating in the study.

## Supplemental References

1. Knowles, M.R., Zariwala, M., and Leigh, M. (2016). Primary Ciliary Dyskinesia. Clin. Chest Med. *37*, 449-461.

M4 = GRCh37/hg19 Xp21.1(chrX:34147944-37312950) x 0

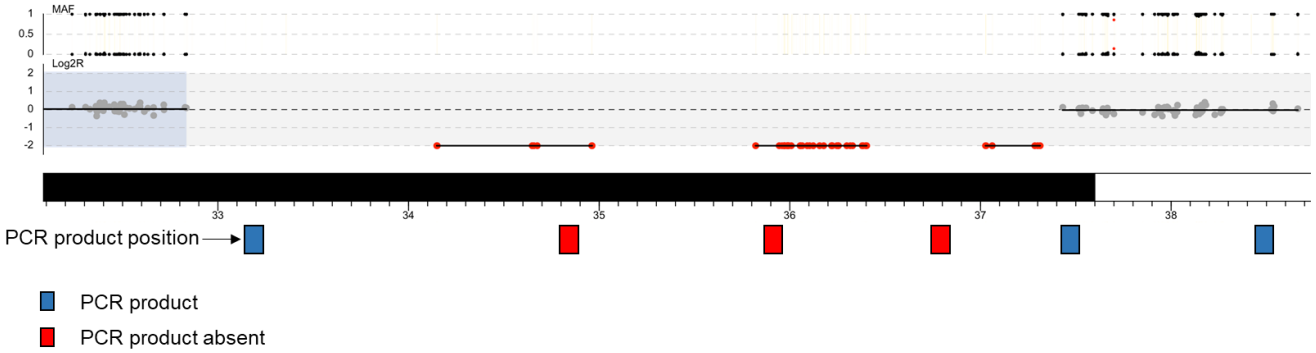



**Figure S1. Verification of the Hemizygous Xp21.1 Deletion Affecting *CFAP47* using PCR Assays.**

Three primer-pairs were designed to amplify the region encompassed by the deletion, while other three primer-pairs were designed to amplify the flanking regions.

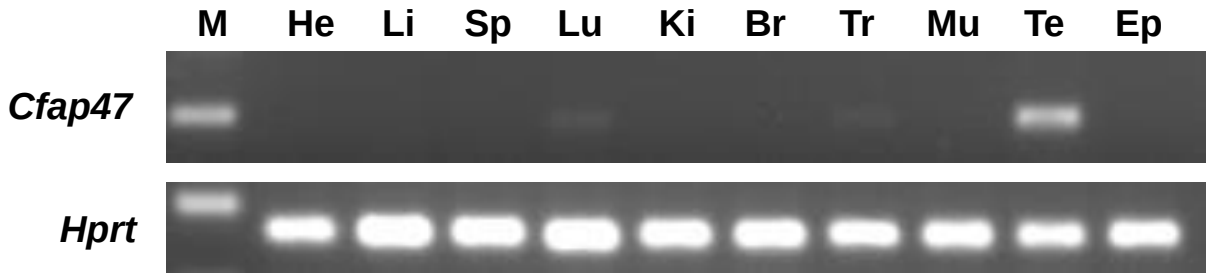



**Figure S2. Gene Expressions of Mouse *Cfap47* in Different Tissues.**

Expressions of *Cfap47* were investigated by RT-PCR in various tissues from adult male mice. *Hprt* was used as an internal control locus.

Abbreviations: M, marker; He, heart; Li, liver; Sp, spleen; Lu, lung; Ki, kidney; Br, brain; Tr, trachea; Mu, muscle; Te, testis; Ep, epididymis.

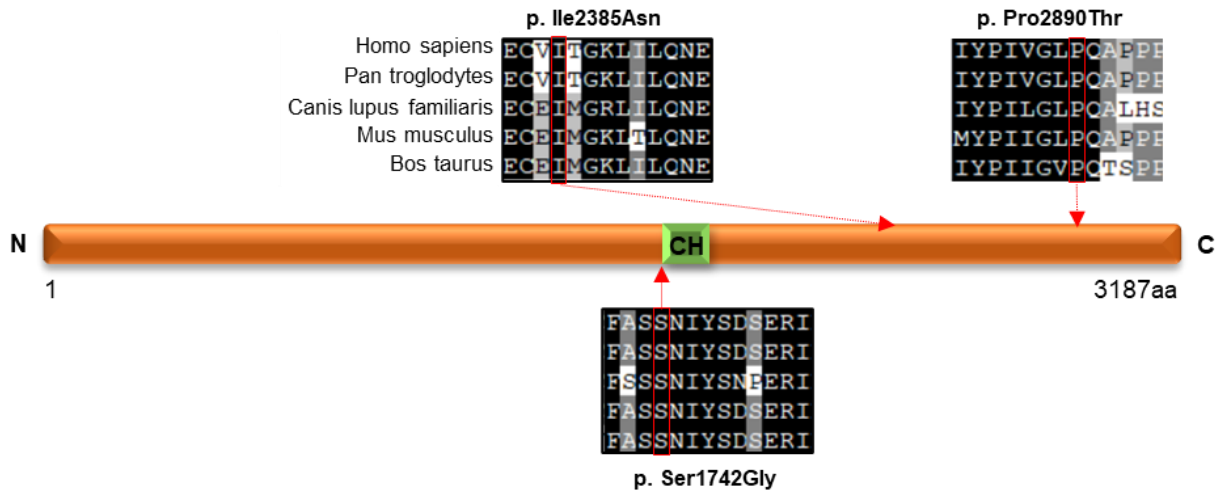



**Figure S3. Phylogenetic Conservation of the Mutated Residues in CFAP47 Protein.**
The CFAP47 protein contains the CH (calponin homology) domain according to the NCBI browser. The mutated residues are indicated by red boxes. The conservativeness analysis was conducted using the ClustalX2 tool. Conserved residues are marked in black (100% conservativeness), dark gray (80% conservativeness), and light gray (60% conservativeness). No shading denotes residues with less than 60% conservativeness.

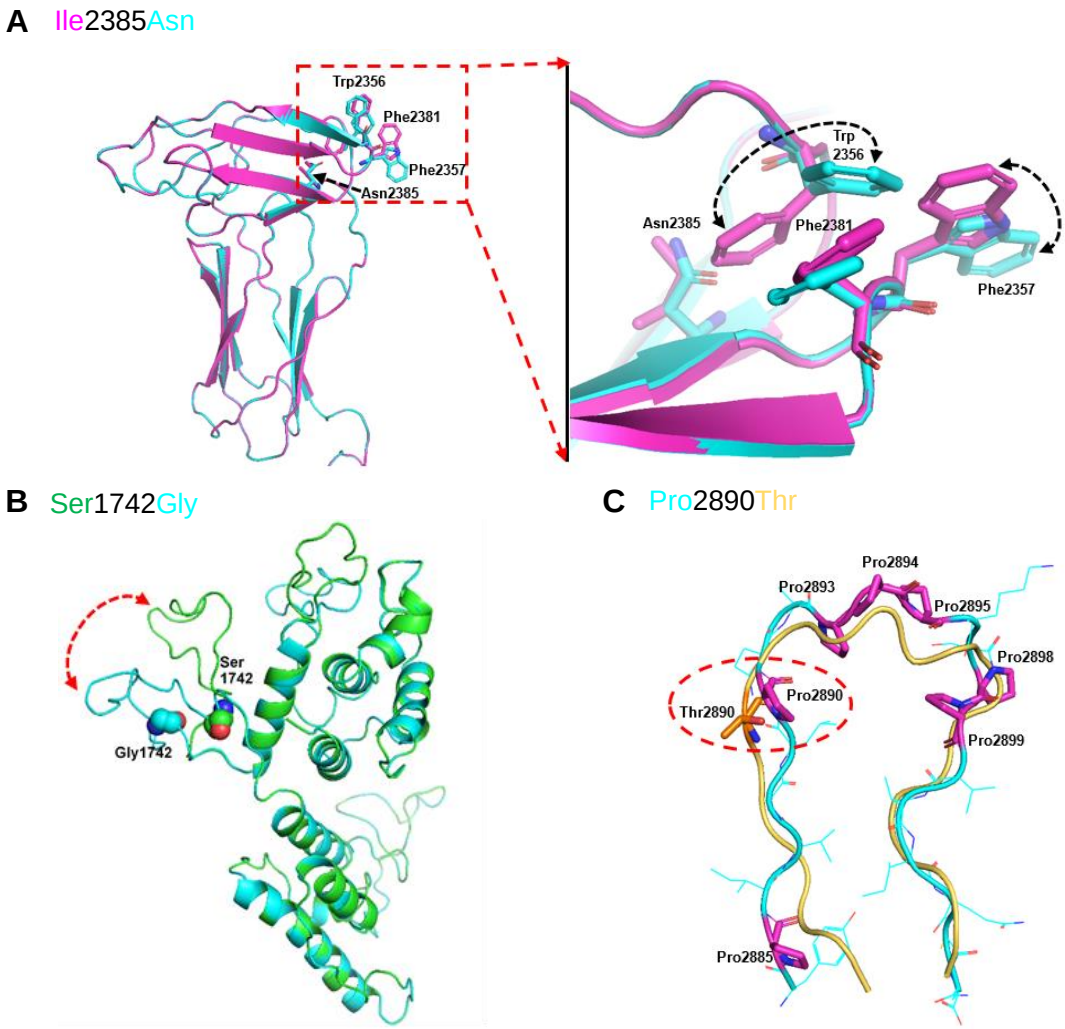



**Figure S4. Effects of *CFAP47* Missense Mutations on the Structure of CFAP47 Protein.**

(A) Structure model of wild-type CFAP47 protein (in pink) and the mutant p.Ile2385Asn (in blue). The mutagenesis of Ile2385Asn introduces the polar interactions and decreases the hydrophobic interactions, which may induce the nearby aromatic residue Trp2356 flipping out and exposing to surface. The corresponding movement of the surrounding aromatic residues Phe2357 and Phe2381 may increase the surface hydrophobic patch for molecular interactions as well as decrease the stability of beta sheets packing structure.

(B) Structure model of wild-type CFAP47 protein (in green) and the mutant p.Ser1742Gly (in blue). The mutagenesis of Ser1742Gly may increase the structure flexibility of backbone and loss the polar interactions, which may induce the possible local conformation change of coli-coli structure.

(C) Structure model of wild-type CFAP47 protein (in blue) and the mutant p.Pro2890Thr (in yellow). The mutagenesis of Pro2890Thr may alter the backbone flexibility, residue charge and hydrophobicity, which may decrease the stability of the loop and change the possible molecular interactions.

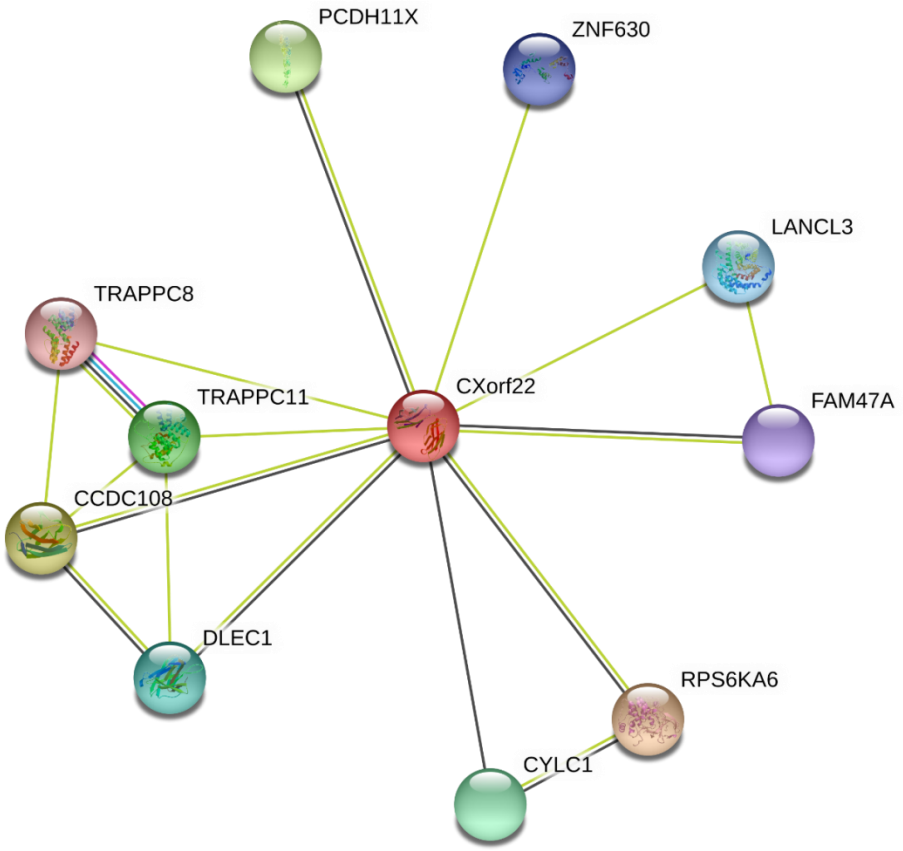



**Figure S5. Protein-Protein Interaction Network for CFAP47 (CXorf22) Predicted by STRING.** The protein interaction network was predicted by *in silico* software String for human CFAP47 (also known as CXorf22 in the figure) protein. A potential interaction was predicted between human CFAP47 and CFAP65 (also known as CCDC108 in the figure). Yellow lines and black lines suggest the predictions from text mining and co-expression, while pink lines and blue lines indicate the known interactions from experimental verification and curated databases.

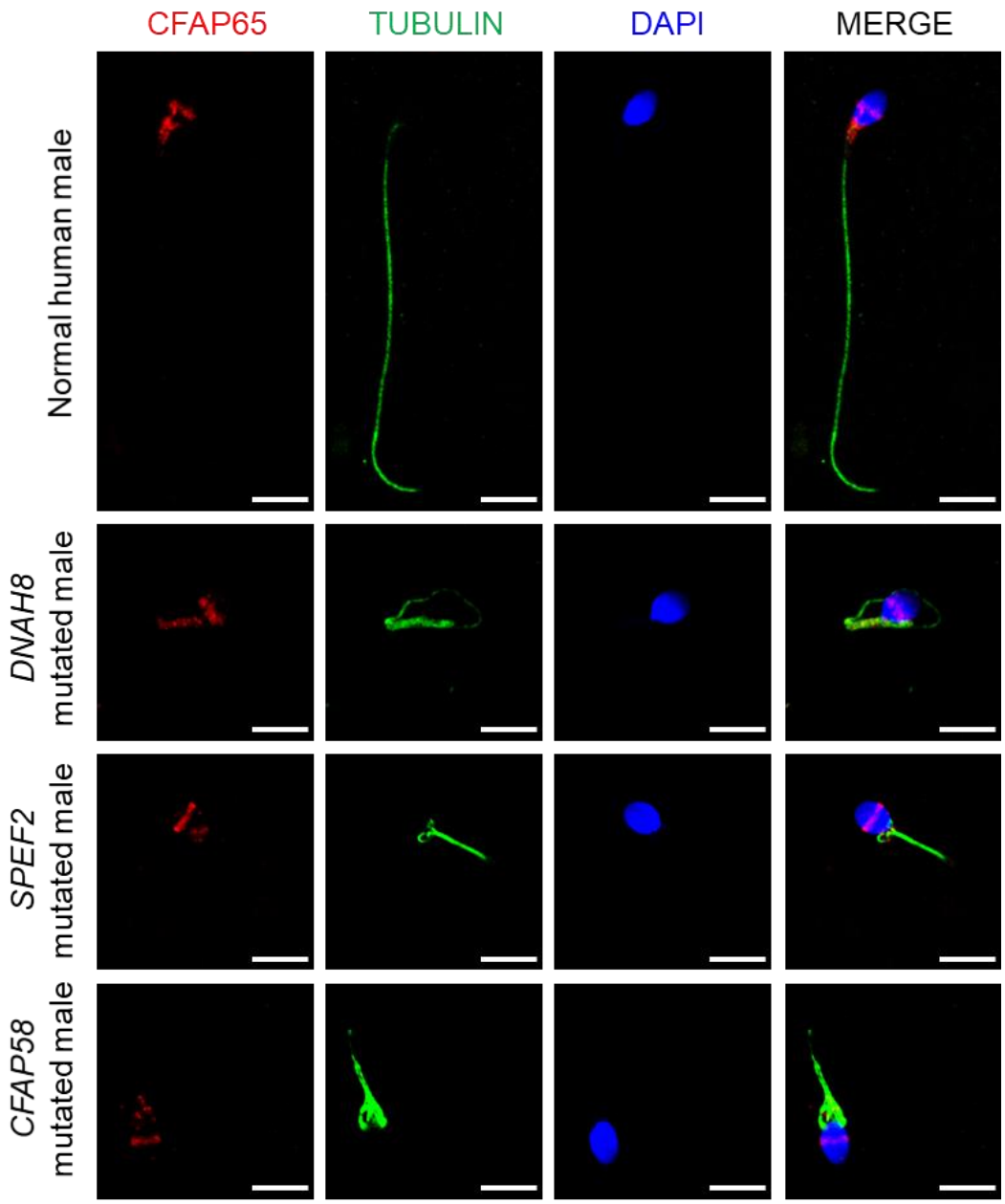



**Figure S6. Immunofluorescence Staining of CFAP65 in the Spermatozoa from a Normal Male Control and MMAF-Affected Men Harboring Bi-allelic Variants in *DNAH8*, *SPEF2* or *CFAP58*.**

Sperm cells were stained with anti-CFAP65 (red) and anti-α-tubulin (green) antibodies. DNA was counterstained with DAPI (4',6-diamidino-2-phenylindole) as a marker of the cell nucleus. CFAP65 staining is concentrated at the equatorial segment of sperm head and the base of flagella from both the control male and men harboring bi-allelic variants of *DNAH8*, *SPEF2* or *CFAP58*.

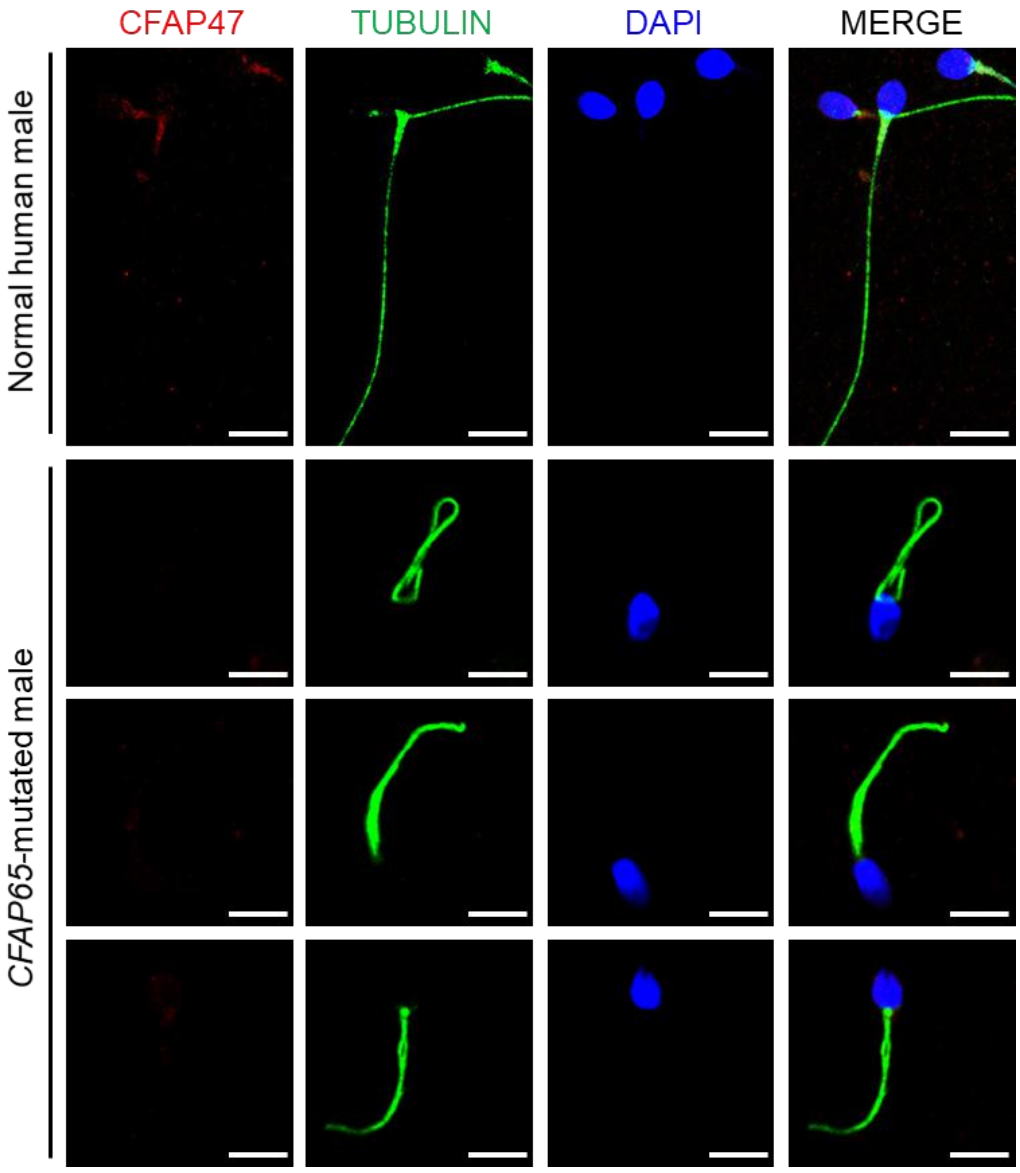



**Figure S7. Reduced CFAP47 Immunostaining in the Spermatozoa from Men Harboring Bi-allelic *CFAP65* Variants.**

Sperm cells were stained with anti-CFAP47 (red) and anti-α-tubulin (green) antibodies. DNA was counterstained with DAPI as a marker of the cell nucleus. In the spermatozoa from a fertile control man, CFAP47 staining appears to be located mainly in the basal body of sperm flagella. In the sperm cells from men harboring bi-allelic *CFAP65* variants, CFAP47 staining is strongly reduced or totally absent. Scale bars: 5 µm.

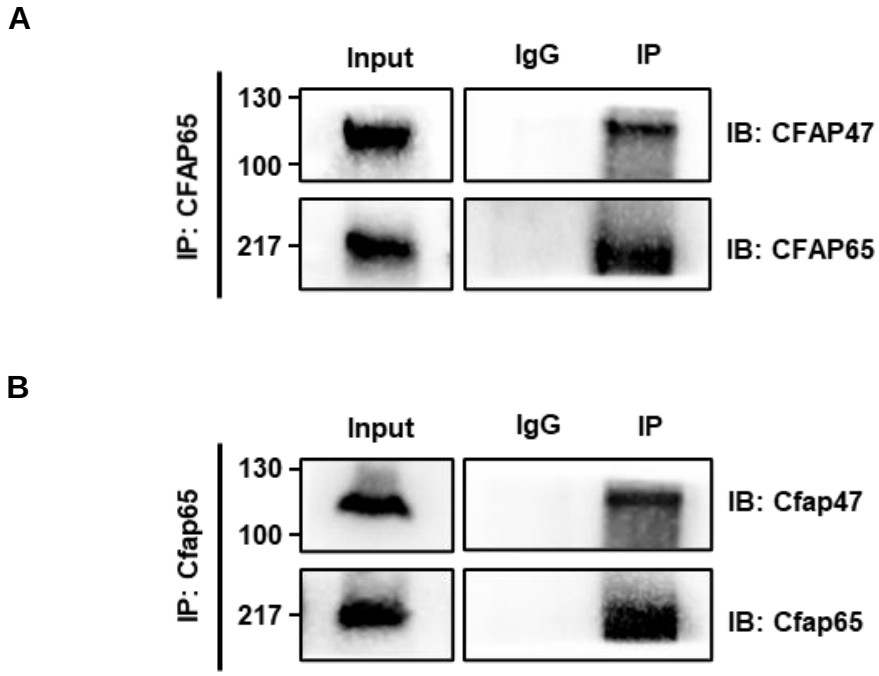



**Figure S8. Co-Immunoprecipitation (Co-IP) Assays Using CFAP47/Cfap47 Antibodies Showed Interactions Between CFAP47/Cfap47 and CFAP65/Cfap65.**

(A) The association between CFAP47 and CFAP65 proteins was shown in the spermatozoa from a normal control man.

(B) The association between Cfap47 and Cfap65 was shown in the testes from wild-type male mice. Input proteins and normal IgG pulldowns were used as positive and negative controls, respectively.

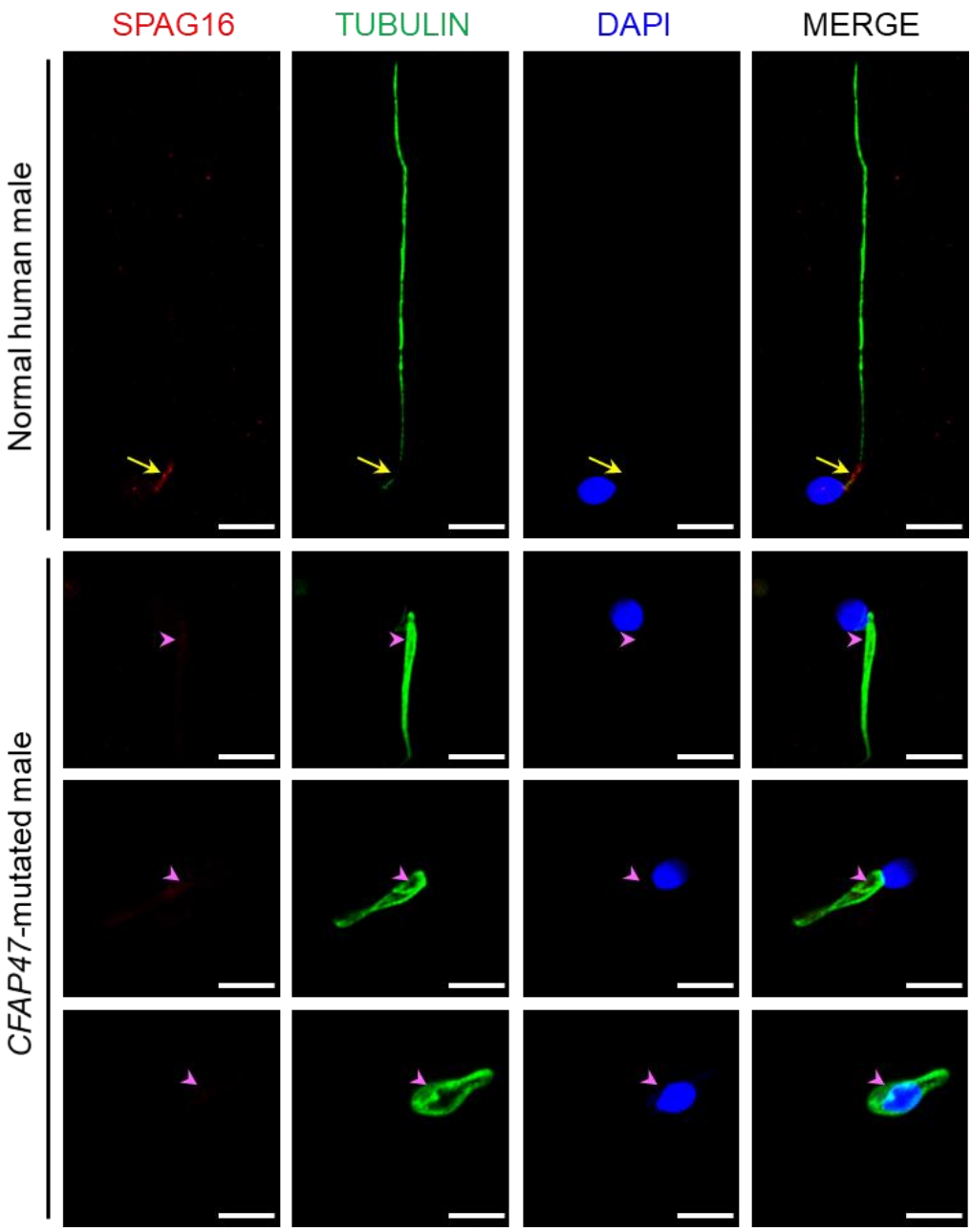
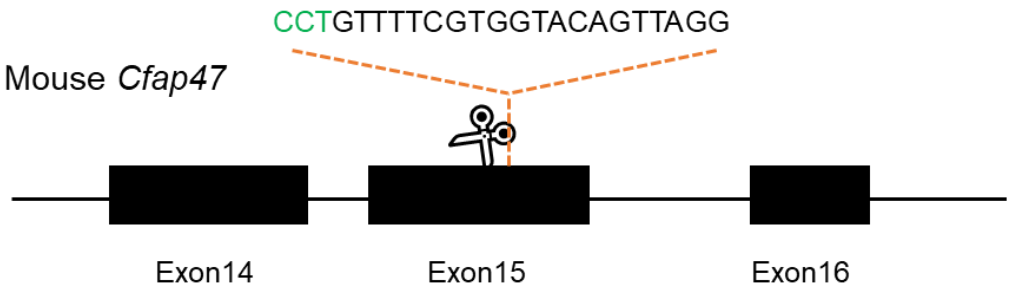



**Figure S9. Immunofluorescence Staining of SPAG16 in the Spermatozoa from a Fertile Male Control and Men Harboring Hemizygous *CFAP47* Variants.**

Sperm cells were stained with anti-SPAG16 (red) and anti-α-tubulin (green) antibodies. DNA was counterstained with DAPI as a marker of the cell nucleus. SPAG16 staining is concentrated at the basal body (as indicated by yellow arrows) of sperm flagella from the normal control, but the signal was almost absent (as indicated by pink arrowheads) in the sperm flagella from men harboring hemizygous *CFAP47* variants. Scale bars: 5 μm.

**A**




**B**


**Figure S10. Schematic Illustration of the Targeting Strategy for Generating *Cfap47*-Mutated Mice.**

(A) The gRNA was targeted in exon 15 of mouse *Cfap47*. The protospacer adjacent motif (PAM) is shown in green.

(B) A frameshift mutation (c.2559insT) was generated in mouse *Cfap47* using CRSIPR-Cas9 technology. The inserted nucleotide was predicted to cause premature translational termination (p.Gly855Profs*8) of mouse *Cfap47*. The termination codon (purple asterisk) was shown in the mutated cDNA. Abbreviation: WT, wild-type.

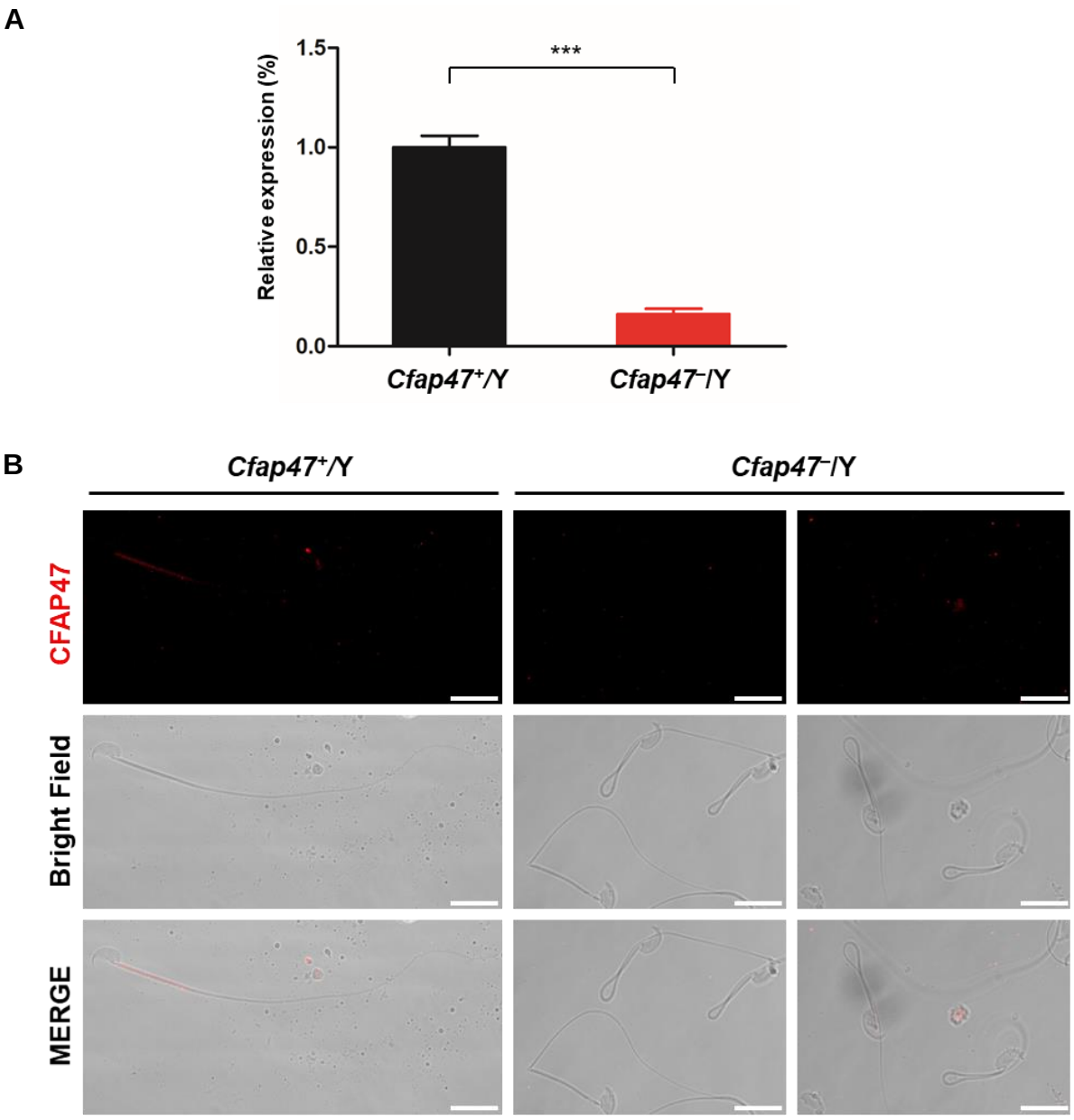



**Figure S11. Expression of *Cfap47* mRNA and Location of CFAP47 Protein in the Spermatozoa from Wild-Type (*Cfap47*$^{+}$/Y) and *Cfap47*-Mutated (*Cfap47*$^{-}$/Y) Male Mice.**

(A) RT-qPCR assay suggested that the level of *Cfap47* mRNA was significantly lower in the spermatozoa from *Cfap47*$^{-}$/Y male mice than that of wild-type male mice. Data represent the means ± SEM of three independent experiments. Two-tailed Student's paired or unpaired *t* tests were used as appropriate (*** $P < 0.001$).

(B) CFAP47 immunostaining showed that the CFAP47 staining (red) is concentrated in the mid-piece of the sperm flagella from wild-type male mice, but the CFAP47 signal is almost absent in the sperm flagella from *Cfap47*$^{-}$/Y male mice. Scale bars: 10 μm.

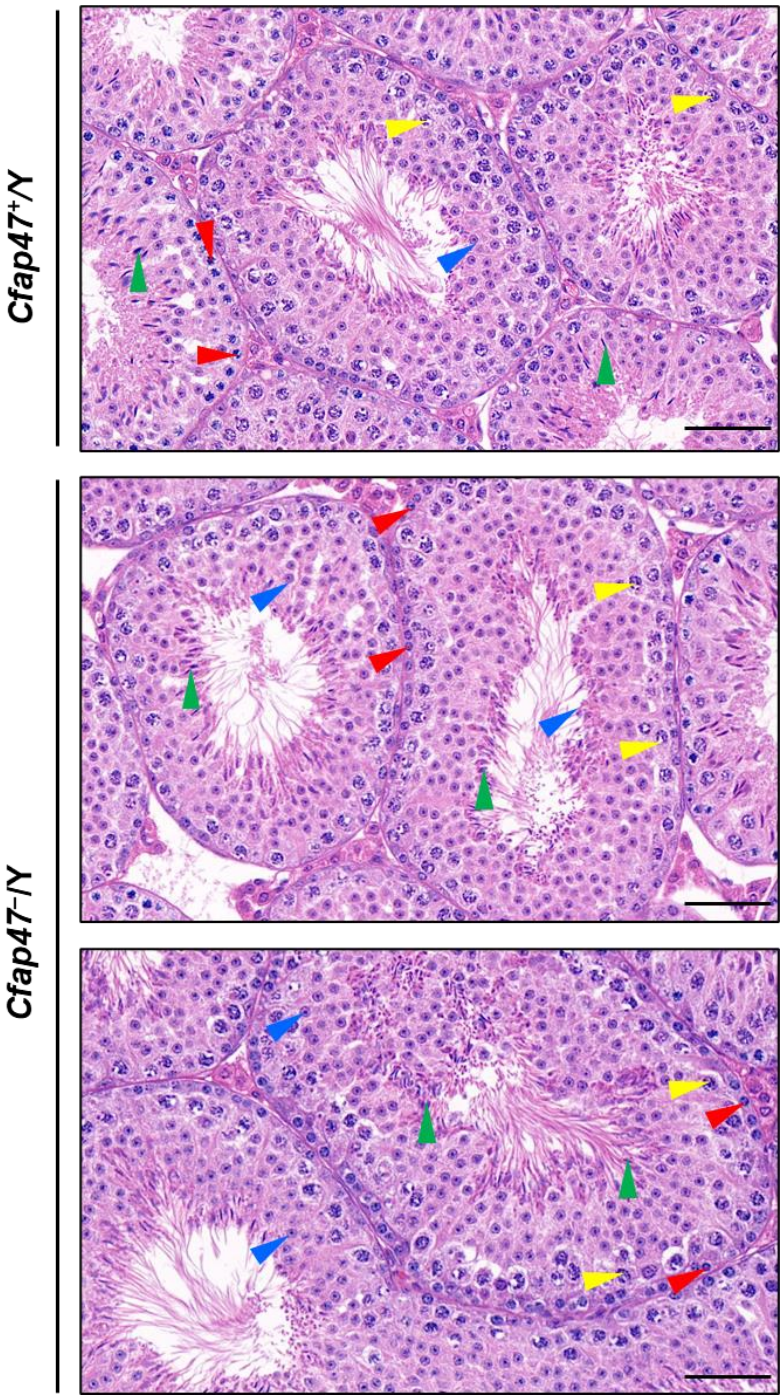



**Figure S12. H&E staining of the Testicular Tissue Sections Obtained from Wild-Type (*Cfap47*$^{+}$/Y) and *Cfap47*-Mutated (*Cfap47*$^{-}$/Y) Male Mice.**

H&E staining of testicular tissues showed no inter-group difference in the overall morphology of germ cells. A variety of cell types can be seen in the testes from both wild-type and *Cfap47*$^{-}$/Y male mice, including spermatogonia (red arrowheads), spermatocytes (yellow arrowheads), round spermatids (blue arrowheads) and spermatozoa (green arrowheads). Scale bars: 50 μm.

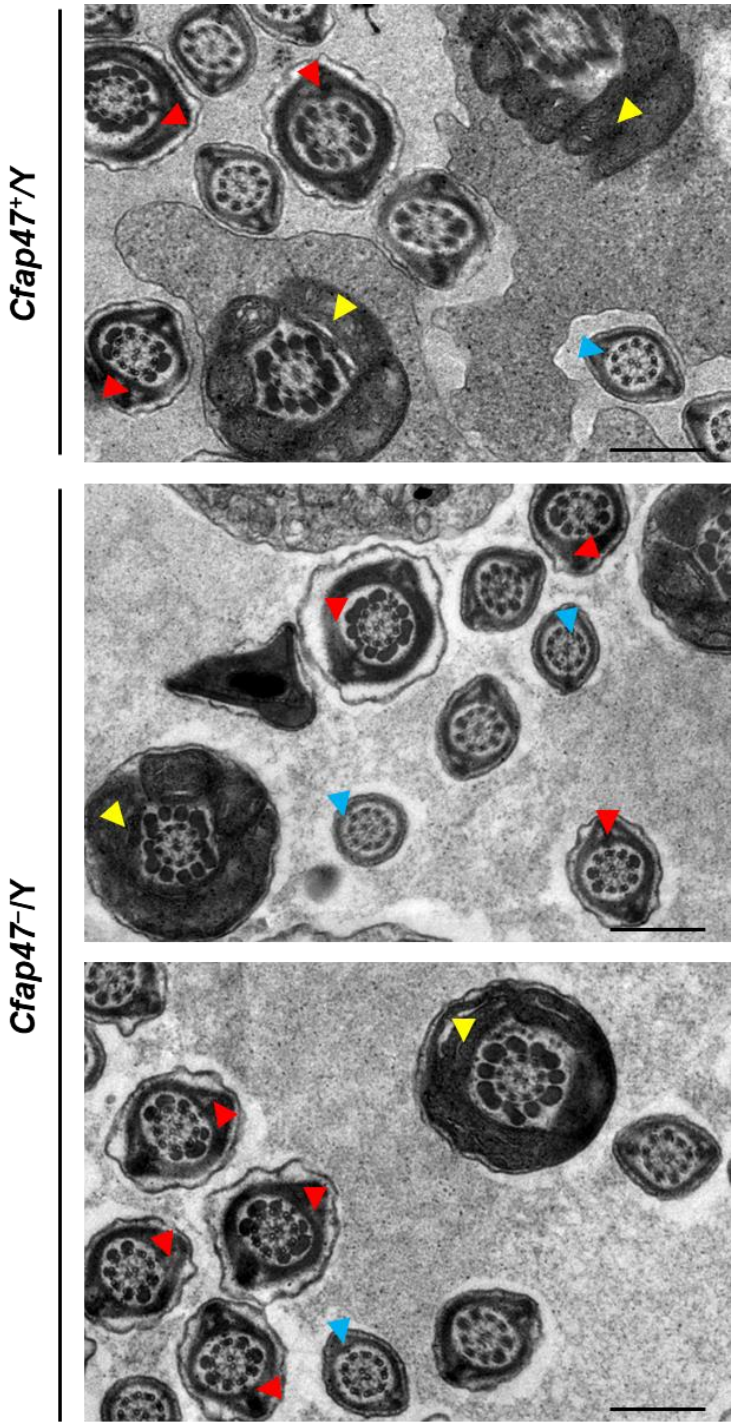



**Figure S13. TEM Cross-Sections of the Spermatozoa from *Cfap47*-Mutated (*Cfap47*$^{-}$/Y) Male Mice Revealed Normal Ultrastructures.**

Yellow arrowheads indicate the mid-piece, red arrowheads indicate the principal piece, and blue arrowheads indicate the end piece of sperm flagella. Scale bars: 200 nm.

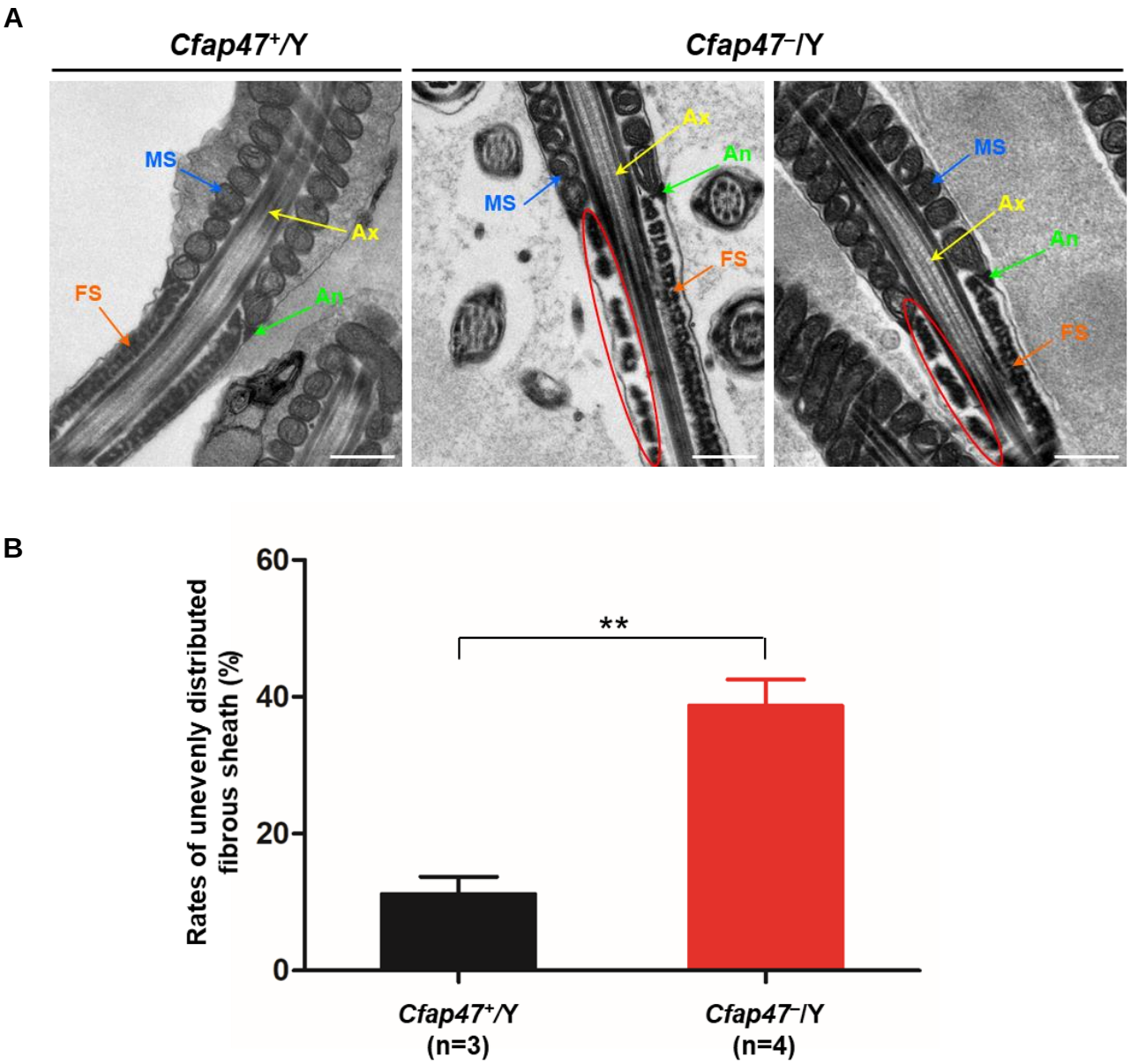



**Figure S14. TEM Vertical Sections of the Spermatozoa from Wild-Type (*Cfap47*$^{+}$/Y) and *Cfap47*-Mutated (*Cfap47*$^{-}$/Y) Male Mice.**

(A) Unevenly distributed fibrous sheathes (as indicated by red circles) were frequently observed in the spermatozoa from *Cfap47*$^{-}$/Y male mice. FS, fibrous sheath (orange arrows); MS, mitochondrial sheath (blue arrows); An, annulus (green arrows); Ax, axoneme (yellow arrows). Scale bars: 0.5 μm.

(B) Rates of unevenly distributed fibrous sheathes at the principal pieces of sperm flagella from *Cfap47*$^{+}$/Y and *Cfap47*$^{-}$/Y male mice. A higher rate of unevenly distributed fibrous sheathes was found in the spermatozoa from *Cfap47*$^{-}$/Y male mice than those in *Cfap47*$^{+}$/Y male mice. At least 50 sections were counted for each mouse. Data are presented as mean ± SEM (**$P < 0.01$). n, the number of male mice analyzed in this assay.

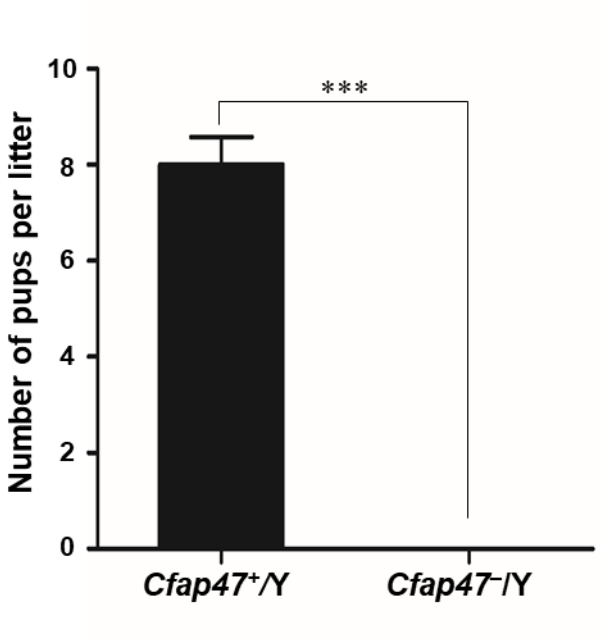



**Figure S15. Fertility of Wild-Type (*Cfap47*$^{+}$/Y) and *Cfap47*-Mutated Male Mice (*Cfap47*$^{-}$/Y).** Wild-type and *Cfap47*$^{-}$/Y male mice were bred with wild-type female mice, and the numbers of pups per litter were counted. Wild-type male mice routinely produced offspring, but *Cfap47*$^{-}$/Y males were sterile. (Unpaired *t* test; ***$P < 0.001$; the error bar represents S.D.)

**Table S1. Primers Used for Amplification and Verification of *CFAP47* Mutations**

| Primer Names | Primer Sequences (5'-3') | Tm |
|---|---|---|
| M1-F | ACAAACGATAAATGGACCTTTCAA | 58 ℃ |
| M1-R | TGTTATGTTCATGATGTTTATGCCT | |
| M2-F | TAGTGCCCTCTTTAATCTCCTGT | 59 ℃ |
| M2-R | GAAAGTAAATGTTGGAGAGGGGTC | |
| M3-F | GCTTTGTGATGATTTTCCAGGGTTT | 57 ℃ |
| M3-R | ACATTCTGACTGGTTTGG | |

**Table S2. Primers Used for the Verification of the Xp21.1 Deletion Affecting *CFAP47***

| PCR product ranges (hg19) | Primer Sequences (5'-3') |
|---|---|
| chrX:33147773-33147947 | F: GCGGAAATTTCATTTGGAGA |
| | R: TAGGCAACCTCCATTTCCAT |
| chrX:34912831-34913051 | F: CTTTGGTCCCTTTTCATGTCA |
| | R: CCAACTTGGTCAATGGAGAA |
| chrX:35912538-35912753 | F: TGAGTCCGAAATTGGAGAGG |
| | R: GGGAGAAGCCAGCAGGTAT |
| chrX:36812497-36812710 | F: TCAGCCATGAGAATGATCCA |
| | R: TCCACTGAGAAAGCACTGAAGA |
| chrX:37431692-37432064 | F: AGGAGGTAAGAGGTAGCCCG |
| | R: GCCTCAGCTTTTGACAGCAG |
| chrX:38512792-38513002 | F: AGTTTTCCATGTGTTTGTCAAGCT |
| | R: TCCCCTCCTTCAAGCAGGTA |

**Table S3. Primers Used for Mouse *Cfap47* Genotyping**

| Primer Names | Primer Sequences (5'-3') | Tm |
|---|---|---|
| M-*Cfap47*-F | CCTTGTCTTGTAAACATCTG | 60 ℃ |
| M-*Cfap47*-R | GTATGCTTTTATGGTTCAGT | |

**Table S4. Primers Used for RT-qPCR and RT-PCR Assays**

| Primer Names | Primer Sequences (5'-3') | Tm |
|---|---|---|
| H-*CFAP47*-F | TTTTCATACGTGATTCTACC | 60 ℃ |
| H-*CFAP47*-R | TCAAGTGTTACTGGCTGGAC | |
| H-*GAPDH*-F | GGAGCGAGATCCCTCCAAAAT | 60 ℃ |
| H-*GAPDH*-R | GGCTGTTGTCATACTTCTCATGG | |
| M-*Cfap47*-F | CCAAGTAGTGGTATAGTAAAGG | 60 ℃ |
| M-*Cfap47*-R | CATGTGAAACAATTCGATGA | |
| M-*Gapdh*-F | AGGTCGGTGTGAACGGATTTG | 60 ℃ |
| M-*Gapdh*-R | TGTAGACCATGTAGTTGAGGTCA | |
| M-*Hprt*-F | TGGATATGCCCTTGACTATAATGAG | 58 ℃ |
| M-*Hprt*-R | TGGCAACATCAACAGGACTC | |

**Table S5. Predicted Effects of *CFAP47* Missense Variants on CFAP47 Protein According to the Online Tool HOPE and the Conservation Analysis at UCSC Genome Browser.**

| | **M1** | **M2** | **M3** |
|---|---|---|---|
| **CFAP47 alteration** | **p.Ile2385Asn** | **p.Ser1742Gly** | **p.Pro2890Thr** |
| **Structure** | The wild-type residue is predicted to be located in a β-strand while the mutant residue prefers to be in another secondary structure, therefore the local conformation will be slightly destabilized. | The mutation introduces a glycine at this position. Glycines are very flexible and can disturb the required rigidity of the protein at this position. | Prolines are known to be very rigid and therefore induce a special backbone conformation which might be required at this position. The mutation can disturb this special conformation. |
| **Amino acid properties** | 1) The mutant residue Asparagine is bigger, which might lead to bumps.<br>2) Hydrophobic interactions will be lost. | 1) The wild-type and mutant amino acids differ in size.<br>2) The mutant residue is smaller, this might lead to loss of interactions. | 1) Hydrophobic interactions will be lost.<br>2) The mutant residue changes a proline into another residue, thereby disturbing the local structure. |
| **Conservation** | Highly conserved | Highly conserved | Highly conserved |

**Table S6. Semen Characteristics and Sperm Morphology in *Cfap47*-Mutated Male Mice**

| | **Wild Type**[a] | ***Cfap47*$^{-}$/Y**[a] |
|---|---|---|
| **Semen Parameter**[b] | | |
| Sperm concentration ($10^6$/mL) | 23.6 (15.3-31.4) | 21.6 (18.3-30.4) |
| Motility (%) | **74.7 (69.3-80.4)** | **47.8 (28.6-65.2)*** ** |
| Progressive motility (%) | **57.8 (51.9-65.9)** | **29.1 (15.8-39.6)*** ** |
| **Sperm Morphology** | | |
| Absent flagella (%) | 5.0 (1.5-7.0) | 13.5 (2.5-15.5) |
| Short flagella (%) | 1.7 (0.5-5.0) | 1.2 (0.5-2.5) |
| Coiled flagella (%) | 3.0 (1.5-8.0) | 3.4 (1.5-6.5) |
| Irregular caliber (%) | 1.1 (0.5-2.5) | 1.3 (0.5-3.0) |
| Bent flagella (%) | **2.2 (0.5-5.0)** | **15.0 (12.0-18.5)** ** |

[a] Values represent the mean (range).

[b] Per single epididymis.

** $P < 0.01$, *** $P < 0.001$. The items with statistical significance are shown in bold.