Supplementary Materials

Supplementary Methods and Materials

Family and clinical information of participants

A consanguineous Pakistani family, which included two male patients, was recruited in our study. Written informed consent was obtained from all family members before the commencement of this study. Blood was collected from all available family members and hormonal analysis and karyotyping were performed. This study was conducted in accordance with the Ethics Committee of the University of Science and Technology of China (USTC).

Whole-exome sequencing (WES) and data analysis

AI Exome Enrichment Kit V1 (iGeneTech, Beijing, China) libraries were created for exome capture of family members, as directed by the manufacturer. Sequencing was carried out using the Hiseq2000 platform (Illumina, San Diego, CA, USA). Clean reads were mapped to the human reference genome (hg19) using the Burrows-Wheeler Alignment tool. SAM files from each sample were converted to BAM files and sorted and merged using SAM tools (http://samtools.sourceforge.net/) (Li et al, 2009). Polymerase chain reaction (PCR) duplicates were removed using Picard (http://broadinstitute.github.io/picard/). Files were further processed using the Genome Analysis Toolkit (GATK) from the Broad Institute (http://www.broadinstitute.org/gatk/) (Depristo et al, 2011). Indel Realigner was used to process all BAM files. Single-nucleotide and indel variants within the captured coding exonic intervals were defined using Unified Genotyper in GATK.

Filtration of candidate variants

Variant filtration was conducted as follows: 1) Variants potentially affecting protein sequences were retained. 2) Variants with minor allele frequencies (MAF)>0.01 in any of the public databases, 1000 Genome Project (http://www.internationalgenome.org/) (Auton et al, 2015), ESP6500 (http://evs.gs.washington.edu/) (Sukhai et al, 2019), ExAC database (https://exac.broadinstitute.org/) (Karczewski et al, 2017), or GnomAD (http://gnomad.broadinstitute.org) (Tukiainen et al, 2017), and homozygous variants in our in-house WES variant call set generated from 578 fertile men (41 Pakistanis, 254 Chinese, and 283 Europeans) were excluded. 3) Variants potentially predicted to be non-deleterious by more than half the software programs (Adzhubei et al, 2010; Davydov et al, 2010; Dong et al, 2015; Lindblad-Toh et al, 2011; Reva et al, 2011; Schwarz et al, 2014; Shihab et al, 2013; Sim et al, 2012) were excluded. 4) Variants within genes not expressed in testes were excluded. 5) Variants following inheritance patterns were included. 6) Variants within genes that may be functional in spermatogenesis based on SpermatogenesisOnline or literature were included. The Supplementary Figure S1 flow chart shows the variant filtration processes and Supplementary Table S1 shows the candidate variants identified from WES data after filtration.

Sanger sequencing

Sanger sequencing was carried out on all available family members to verify the inheritance pattern of the variants.

Conservation analysis

Amino acid sequences of different species were retrieved from UniProt (https://www.uniprot.org/ink), and multiple sequence alignments were accomplished using Clustal W (https://pubmed.ncbi.nlm.nih.gov/17846036/).

Generation of knockin (KI) mouse model

The mouse model harboring the identified variant was generated by CRISPR/Cas9 genome editing technology, as reported previously (Yang et al, 2013). Guide RNAs were designed to target exon 15 of the *Usp9x* gene. Tail biopsies were used for genomic DNA extraction of founder mice, and genotyping was performed through Sanger sequencing following PCR. Founder mice heterozygous for the missense mutation of *Usp9x* were crossed to obtain homozygous KI mice. All animal experiments were approved by the Institutional Animal Ethics Committee at the University of Science and Technology of China.

Fertility test

The fertility test was carried out by keeping one adult KI male mouse with two wild-type (WT) females (C57BL/6J) for three months. All females were monitored for pups per litter and number of litters.

Sperm count and morphological and motility analysis

Epididymides from adult mice were removed and chopped into small pieces, then incubated at 37 °C for 30 min. Hemocytometer chambers were used for sperm counting. Sperm smear slides were prepared for sperm morphology analysis. The percentages of morphologically normal and abnormal spermatozoa were scored. To assess sperm motility, the caudas of adult control and KI mice were trimmed, rinsed in phosphate-buffered saline (PBS; pH 7.4), and kept in 1 mL of human tubular fluid (HTF) medium (Millipore). The caudas were cut several times and spermatozoa were released into the medium with 10% fetal bovine serum for 5 min incubation at 37 °C. The suspension was gently mixed by swirling and diluted to a concentration of $2–4\times10^5$ spermatozoa/mL, corresponding to 50–120 spermatozoa per microscopic field for computer-aided sperm analysis (CASA; M1130; Nanjing AiBei Biotechnology, China).

Histological investigations

The testes and epididymides from adult mice were removed, fixed in Bouin's solution overnight, embedded in paraffin, and sectioned at 5–10 μm thickness per slide. Hematoxylin and eosin (H&E) and periodic acid-Schiff (PAS) staining were performed (Gao et al, 2020). Images were taken using a Nikon ECLIPSE 80i microscope (Japan) with a DR-Ri1 camera and managed with NIS-elements BR software.

Immunofluorescence staining

Paraffin sections of testes from control and KI mice were fixed in 4% paraformaldehyde for immunofluorescence staining. Primary antibodies against LIN28A (1:100, R&D Systems, AF3757, USA) and anti-SOX9 (1:400, Millipore, AB5535, USA) were added to the slides and kept at 4 °C overnight. Anti-mouse IgG cross-adsorbed secondary antibody, Alexa Fluor 488 (1:100, A-21121, USA; Molecular Probes, USA), and donkey anti-goat IgG (H+L) (1:100, ThermoFisher,

D-20698, USA) were added and maintained for 1.5 h at 37 ºC in an incubator. Finally, the slides were mounted by Vectashield along with Hoechst 33342 (Invitrogen, H21492, USA). Images were taken using a digital DS-Ri1 camera equipped with a Nikon Eclipse 80i microscope.

Meiotic prophase I analysis

Spermatocytes from control and KI mice were prepared according to our previous report (Jiang et al, 2018). Primary antibodies of SYCP3 (1:100, Novus, NB300-232, USA) and γH2AX (1:200, Millipore, 05-636, USA) were used. Images were taken with an Olympus BX61 microscope (Japan) with a CCD camera (QImaging, QICAM Fast 1394, Canada) and processed with Image-Pro Plus software (Media Cybernetic).

Statistical analysis

Student's *t*-tests were performed for testis/body weight ratios, litter sizes, and sperm parameters between KI and control mice. Results are presented as mean±SD from at least three mice for each group, and $P<0.05$ was considered statistically significant.

REFERENCES

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods, **7**(4): 248-249.

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, Mccarthy S, Mcvean GA, Abecasis GR. 2015. A global reference for human genetic variation. Nature, **526**(7571): 68-74.

Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol, **6**(12): e1001025.

Depristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, Mckenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet, **43**(5): 491-498.

Djureinovic D, Fagerberg L, Hallström B, Danielsson A, Lindskog C, Uhlén M, Pontén F. 2014. The human testis-specific proteome defined by transcriptomics and antibody-based profiling. Molecular human reproduction, **20**(6): 476-488.

Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet, **24**(8): 2125-2137.

Ernst C, Eling N, Martinez-Jimenez CP, Marioni JC, Odom DT. 2019. Staged developmental mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. Nat Commun, **10**(1): 1251.

Farhi J & Ben-Haroush A. 2011. Distribution of causes of infertility in patients attending primary fertility clinics in Israel. Sat, **4**: 19.

Gao Q, Khan R, Yu C, Alsheimer M, Jiang X, Ma H, Shi Q. 2020. The testis-specific LINC component SUN3 is essential for sperm head shaping during mouse spermiogenesis. J Biol Chem, **295**(19): 6289-6298.

Gershoni M, Hauser R, Barda S, Lehavi O, Arama E, Pietrokovski S, Kleiman SE. 2019. A new MEIOB mutation is a recurrent cause for azoospermia and testicular meiotic arrest. Hum Reprod, **34**(4): 666-671.

Guerri G, Maniscalchi T, Barati S, Busetto GM, Del Giudice F, De Berardinis E, Cannarella R, Calogero AE, Bertelli M. 2019. Non-syndromic monogenic male infertility. Acta Biomed, **90**(10-S): 62-67.

Guo J, Grow EJ, Mlcochova H, Maher GJ, Lindskog C, Nie X, Guo Y, Takei Y, Yun J, Cai L, Kim R, Carrell DT, Goriely A, Hotaling JM, Cairns BR. 2018. The adult human testis transcriptional cell atlas. Cell Res, **28**(12): 1141-1157.

Jiang H, Gao Q, Zheng W, Yin S, Wang L, Zhong L, Ali A, Khan T, Hao Q, Fang H, Sun X, Xu P, Pandita TK, Jiang X, Shi Q. 2018. MOF influences meiotic expansion of H2AX phosphorylation and spermatogenesis in mice. PLoS Genet, **14**(5).

Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, Hamamsy T, Lek M, Samocha KE, Cummings BB, Birnbaum D, Daly MJ, Macarthur DG. 2017. The ExAC browser: displaying reference data information from over 60 000 exomes. Nucleic Acids Res, **45**(D1): D840-D845.

Kishi K, Uchida A, Takase HM, Suzuki H, Kurohmaru M, Tsunekawa N, Kanai-Azuma M, Wood SA, Kanai Y. 2017. Spermatogonial deubiquitinase USP9X is essential for proper spermatogenesis in mice. Reproduction, **154**(2): 135-143.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics, **25**(16): 2078-2079.

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, Ward LD, Lowe CB, Holloway AK, Clamp M, Gnerre S, Alföldi J, Beal K, Chang J, Clawson H, Cuff J, Di Palma F, Fitzgerald S, Flicek P, Guttman M, Hubisz MJ, Jaffe DB, Jungreis I, Kent WJ, Kostka D, Lara M, Martins AL, Massingham T, Moltke I, Raney BJ, Rasmussen MD, Robinson J, Stark A, Vilella AJ, Wen J, Xie X, Zody MC, Baldwin J, Bloom T, Chin CW, Heiman D, Nicol R, Nusbaum C, Young S, Wilkinson J, Worley KC, Kovar CL, Muzny DM, Gibbs RA, Cree A, Dihn HH, Fowler G, Jhangiani S, Joshi V, Lee S, Lewis LR, Nazareth LV, Okwuonu G, Santibanez J, Warren WC, Mardis ER, Weinstock GM, Wilson RK, Delehaunty K, Dooling D, Fronik C, Fulton L, Fulton B, Graves T, Minx P, Sodergren E, Birney E, Margulies EH, Herrero J, Green ED, Haussler D, Siepel A, Goldman N, Pollard KS, Pedersen JS, Lander ES, Kellis M. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. Nature, **478**(7370): 476-482.

Okutman O, Rhouma MB, Benkhalifa M, Muller J, Viville S. 2018. Genetic evaluation of patients with non-syndromic male infertility. J Assist Reprod Genet, **35**(11): 1939-1951.

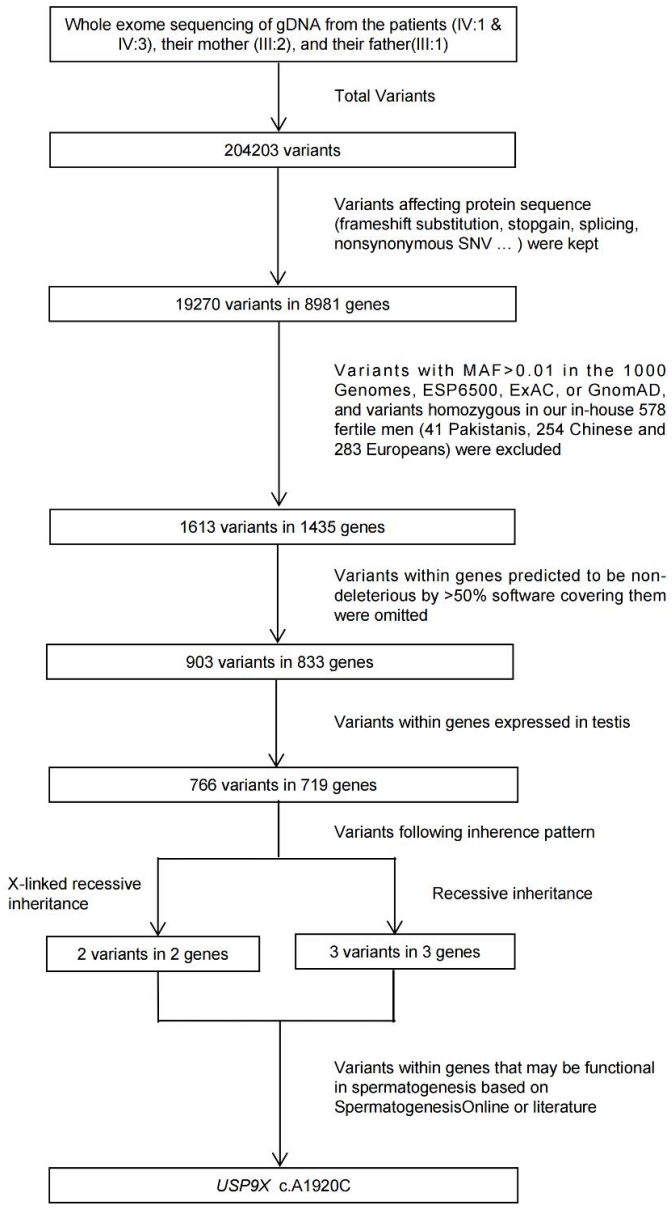
Reva B, Antipin Y, Sander C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res, **39**(17): e118.

Rimoin DL, Connor JM, Pyeritz RE, Korf BR. 2007. Emery and Rimoin's principles and practice of medical genetics. Churchill Livingstone Elsevier.

Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods, **11**(4): 361-362.

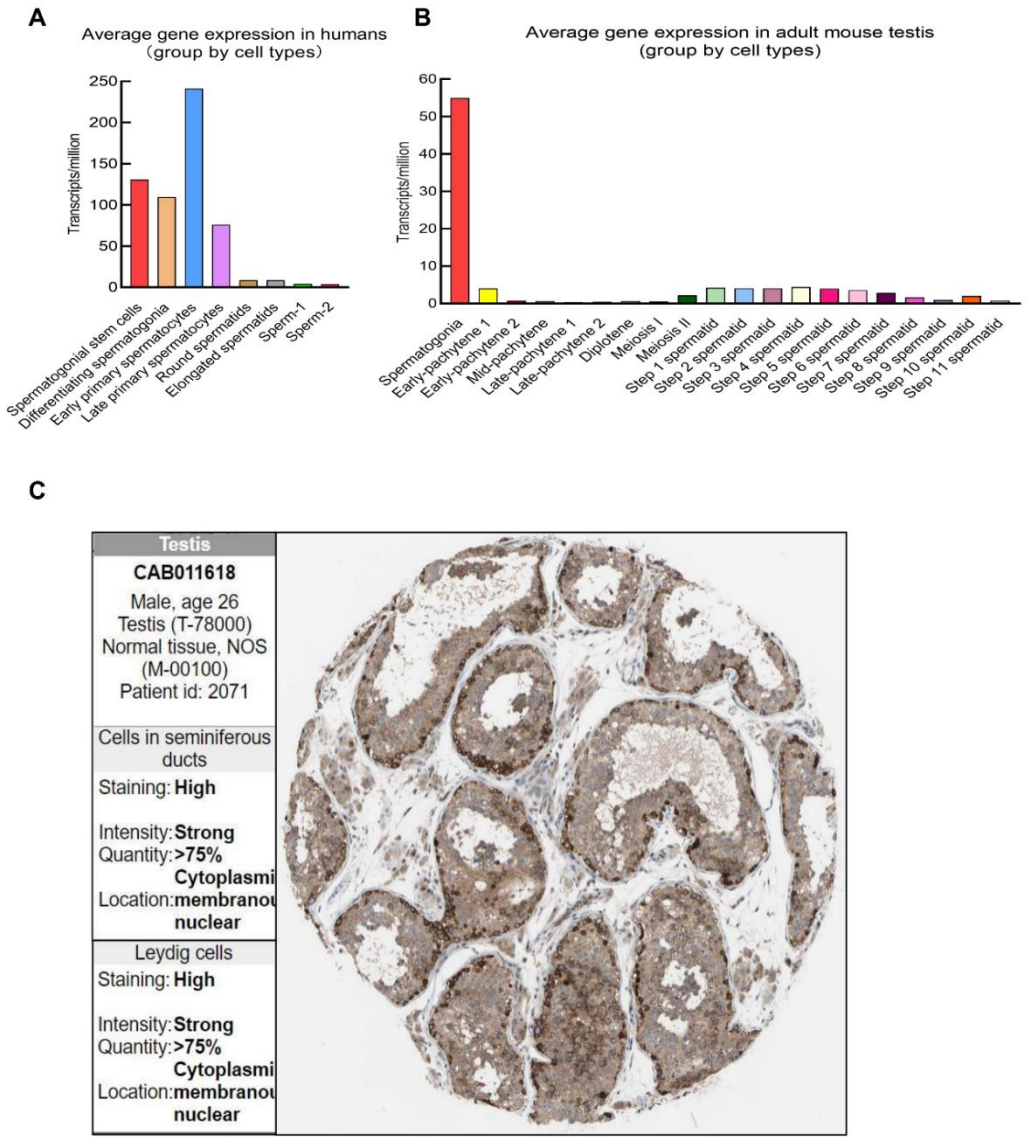
Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, Day IN, Gaunt TR. 2013. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat, **34**(1): 57-65.
Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. 2012. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res, **40**(Web Server issue): W452-457.
Singh P & Schimenti JC. 2015. The genetics of human infertility by functional interrogation of SNPs in mice. Proc Natl Acad Sci U S A, **112**(33): 10431-10436.
Sironen A, Shoemark A, Patel M, Loebinger MR, Mitchison HM. 2020. Sperm defects in primary ciliary dyskinesia and related causes of male infertility. Cell Mol Life Sci, **77**(11): 2029-2048.
Stegeman S, Jolly LA, Premarathne S, Gecz J, Richards LJ, Mackay-Sim A, Wood SA. 2013. Loss of Usp9x disrupts cortical architecture, hippocampal development and TGFβ-mediated axonogenesis. PLoS One, **8**(7): e68287.
Sukhai MA, Misyura M, Thomas M, Garg S, Zhang T, Stickle N, Virtanen C, Bedard PL, Siu LL, Smets T, Thijs G, Van Vooren S, Kamel-Reid S, Stockley TL. 2019. Somatic Tumor Variant Filtration Strategies to Optimize Tumor-Only Molecular Profiling Using Targeted Next-Generation Sequencing Panels. J Mol Diagn, **21**(2): 261-273.
Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, Cummings BB, Castel SE, Karczewski KJ, Aguet F, Byrnes A, Lappalainen T, Regev A, Ardlie KG, Hacohen N, Macarthur DG. 2017. Landscape of X chromosome inactivation across human tissues. Nature, **550**(7675): 244-248.
Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res, **38**(16): 3.
Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell, **154**(6): 1370-1379.
Zhang Y, Zhong L, Xu B, Yang Y, Ban R, Zhu J, Cooke HJ, Hao Q, Shi Q. 2013. SpermatogenesisOnline 1.0: a resource for spermatogenesis based on manual literature curation and genome-wide data mining. Nucleic Acids Res, **41**(Database issue): D1055-1062.

Supplementary figure legends



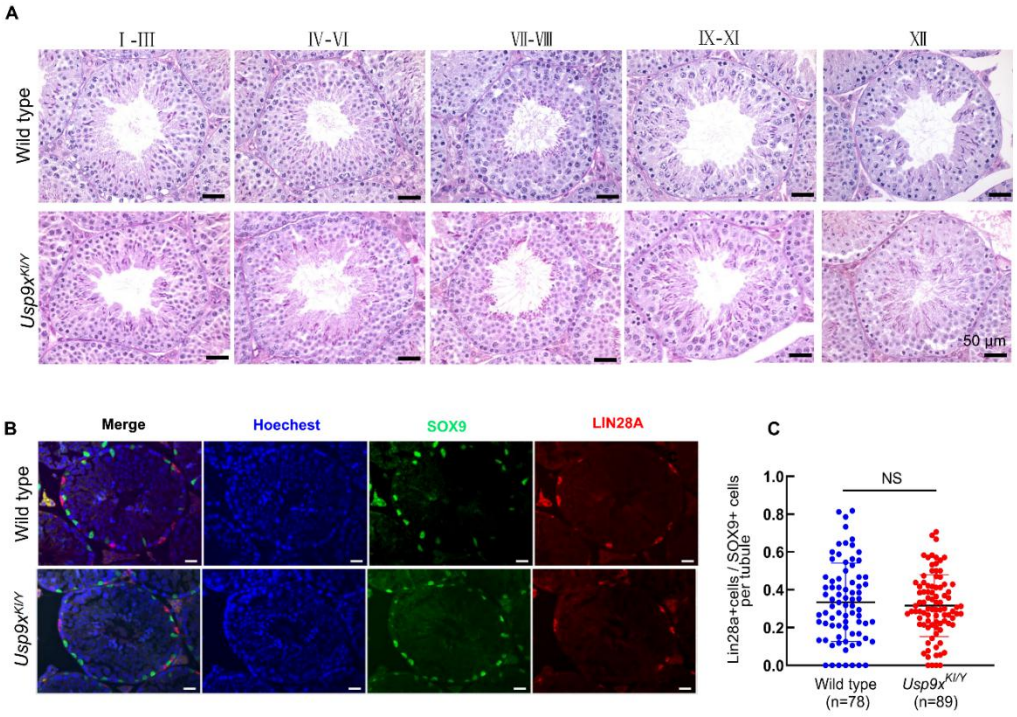
Supplementary Figure S1. Whole-exome sequencing and data filtration strategy

Schematic of screening and filtration of WES for annotating potential variants.



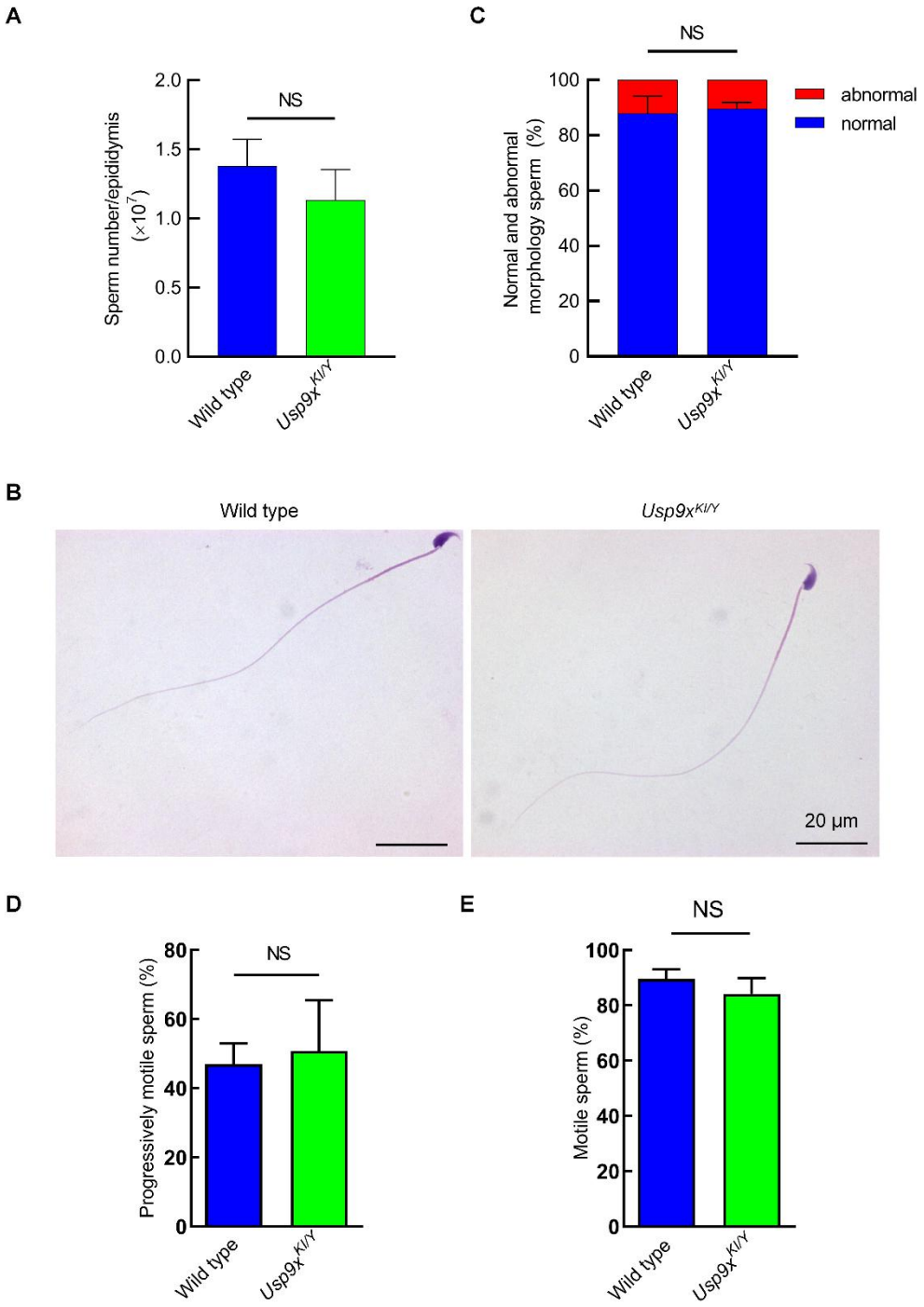
Supplementary Figure S2. Expression patterns of *USP9X* in various databases

(A-B) Expression portfolio of *USP9X* from single-cell RNA sequencing (scRNA-seq) of human (**A**) and mouse **(B)** testicular cells. **(C)** Expression pattern of USP9X in human testicular sections (from Human Protein Atlas database).



Supplementary Figure S3. PAS-staining and LIN28A immunofluorescence

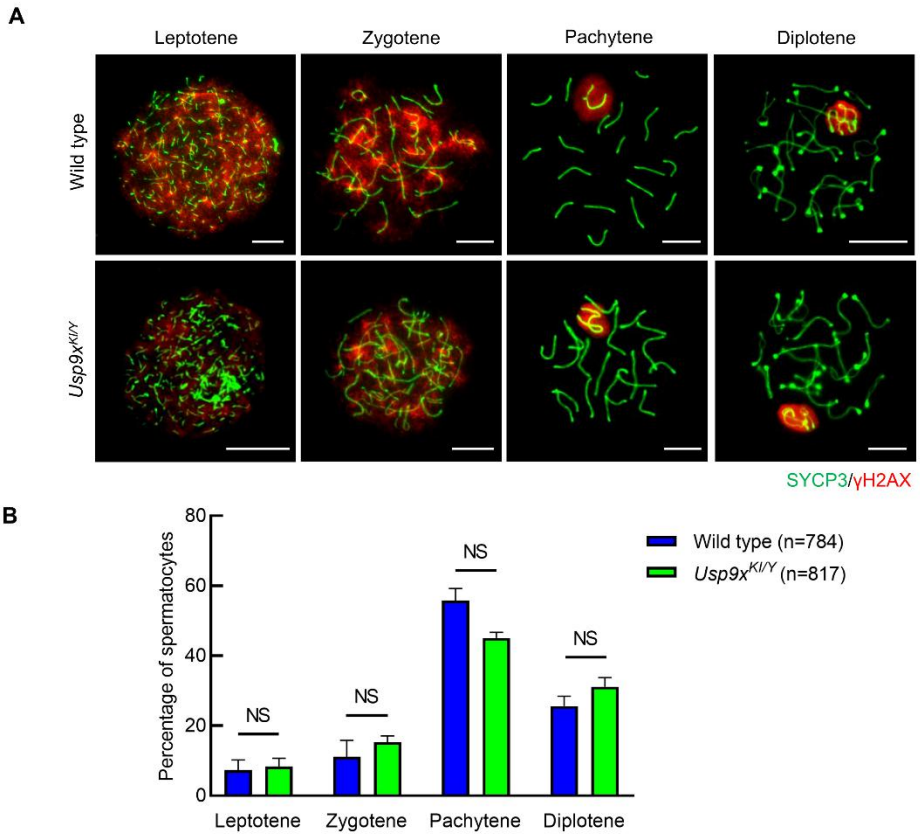
(A) Representative images of PAS-stained testicular sections from adult WT and *Usp9x*$^{KI/Y}$ mice. Scale bars, 50 μm. **(B)** Representative images of testicular sections from adult WT and *Usp9x*$^{KI/Y}$ mice stained with LIN28A (marker for spermatogonia), SOX (marker for Sertoli cells), and Hoechst (DNA). Scale bars, 20 μm. **(C)** Statistical analysis of LIN28A^{+} spermatogonia in WT and *Usp9x*$^{KI/Y}$ mice. NS, no significant difference.



Supplementary Figure S4. Sperm count, morphology, and motility analysis

(A) Sperm number in 8-week-old WT and *Usp9x*$^{KI/Y}$ mice. (**B**) Sperm morphology from adult WT and *Usp9x*$^{KI/Y}$ mice. Scale bars, 20 μm. **(C)** Ratios of normal or

abnormal sperm from 8-week-old WT and *Usp9x*$^{KI/Y}$ mice. For (**A**) and (**C**), data are mean±SD, with at least three mice analyzed per genotype. Student's *t*-test was used for statistical analyses. NS, not significant. **(D-E)** Percentage of progressively and total motile spermatozoa in adult WT and *Usp9x*$^{KI/Y}$ mice. NS, no significant difference.



Supplementary Figure S5. Prophase I progression analysis

(**A**) Surface-spread spermatocyte preparations of WT and *Usp9x*$^{KI/Y}$ testes from 8-week-old mice. Immunofluorescence staining was conducted with antibodies for SYCP3 (green) and γH2AX (red). Scale bars, 10 μm. **(B)** Fractions of leptotene, zygotene, pachytene, and diplotene spermatocytes in prophase I progression analyses. Data are mean±SD, with at least three mice analyzed per genotype. n indicates total number of cells scored.

Supplementary Table S1. Clinical characteristics of patients

	Reference values	IV:1	IV:3
Age (years) [a]	-	1986	1988
Years of marriage	-	2005	2010
Fertility	-	Infertile	Infertile
USP9X mutation	-	c.A1920T/c.A1920T	c.A1920T/c.A1920T
Height/Weight (cm/kg)	-	170/63	166/67
Diagnosis of disease	-	oligo-asthenozoospermia	oligo-asthenozoospermia
Semen parameters [b]			
Semen volume (ml)	>1.5	2.5	4.2
Sperm concentration (millions/ml)	>15	2	5
Motile sperm (%)	>40	13	18
Progressively motile sperm (%)	>32	2	6
Hormone analysis [c]			
FSH (U/L)	1.27-19.26	4.18	7.05
LH (U/L)	1.24-8.62	6.72	4.09
Prolactin (ng/ml)	2.64-13.13	4.41	12.35
Testosterone (ng/ml)	1.75-7.81	5.80	6.08

MT, the mutant allele. FSH, follicle-stimulating hormone. LH, luteinizing hormone. [a] At manuscript preparation (2021). [b] Reference values were published by WHO in 2010. [c] Reference values were suggested by our local clinical laboratory.

Supplementary Table S2. Variants not selected for functional study in family

Gene name	Mutation type	cDNA change	Reasons for exclusion
WI2-2373I1.2	nonsynonymous SNV	A275G	This gene is FOXL3-OT1 (FOXL3 Overlapping Transcript 1), an RNA Gene, and is affiliated with the lncRNA class. An important paralog of this gene is FOXL1. (Genecard ID: GC07P000297). (https://www.genecards.org/Search/Keyword?queryString=GC07P000297). There is currently no clue that long non-coding RNA has a role in spermatogenesis. Additionally, the mutation site is not conserved evolutionarily. In particular, there is no homologous locus in the non-human species, Rhesus.
ADGRA3	nonsynonymous SNV	G1133A	This gene is predominantly expressed in spermatogonial cells. (https://mcg.ustc.edu.cn/bsc/spermgenes2.0/detail.php?sg=SG012494). We also made mouse model carrying this mutation. Although no mice with the target mutation were obtained, we got several mice carrying frameshift mutations near this site. Homozygous mutant mice are fertile and grossly normal. Relevant data is available as request.
TKTL1	nonsynonymous SNV	C356T	Mice knocked-out of this gene are fertile. (http://www.informatics.jax.org/searchtool/Search.do?query=Tktl1).
ENPP5	nonsynonymous SNV	C323T	This mutation has a high population MAF of 0.02 in Bengali (Bangladesh). This should be a polymorphic variant rather than a pathogenic variant in South Asia.

Supplementary Table S3. *In silico* tools used for prediction of variant effects on protein structure

Algorithm	Categorical Prediction	USP9X rs569095263
FATHMM (fathmm-MKL)	D: Deleterious; T: Tolerated	D
GERP++ (gerp++)	higher scores are more deleterious (>3)	4
SiPhy (siphy)	higher scores are more deleterious (>10)	9.23
PolyPhen 2 HDIV (pp2_hdiv)	D: Probably damaging (>=0.957), P: possibly damaging (0.453<=pp2_hdiv<=0.956); B: benign (pp2_hdiv<=0.452)	P (0.884)
MutationTaster (mt)	A: disease causing automatic; D: disease causing; N: polymorphism; P: polymorphism automatic	D (1)
MutationAssessor (ma)	H: high; M: medium; L: low; N: neutral. H/M: functional; L/N: non-functional	L (1.39)
SIFT (sift)	D: Deleterious (sift<=0.05); T: tolerated (sift>0.05)	T (0.095)

Supplementary Table S4. Male mouse fertility assay

Genotype	Mating period (months)	No. of fertile mice (%)	Average litters/ mouse/month	Average pups/litter
Wild type	3	3 (100%)	1.80 ± 0.08	7.04 ± 0.30
Usp9x$^{KI/Y}$	3	3 (100%)	1.78 ± 0.11^{NS}	6.88 ± 0.34^{NS}

Data are presented as mean ± SD. The Student's t-test was used to statistical analyses. NS, not significant.

Supplementary Table S5. Primers used in mouse construction

Information	Name	Sequence	Product size (bp)
Primers for KI mice construction	Usp9x sg1-F	GaaattaatacgactcactatagggagaTGACCCACAAACTGTGAGACGttttagagc	-
	Usp9x sg2-F	GaaattaatacgactcactatagggagaCTTCCCAGTCTCACAGTTTGGttttagagc	
	Usp9x -sg -R	AAAAAAGCACCGACTCGGTG	
	Usp9x-oligo	ATATATATAAAATGCATCTTTTTATTTTTCCTAATTTGCAGACAATGAAG	
		ACTATGACCCACAcACaGTGAGACTGGGAAGTAGATATAGTCATGTTC	
		AAGAAGTCCAAGAACGGCTTAACTTTCTTAGG	
Primers for genotyping	Usp9x genotyping-F	GCCCCAGTATATGAAGTAGG	WT:534; mutant:534
	Usp9x genotyping-R	GGATGGTATACAACCCACTC	