

Supplementary Information

Spermatozoa lacking TEX51 display hypofertility and defects in morphology

Sophie Dupuis¹, Morgane Le Beulze¹, Tyler D.R. Vance³, Rémi Pierre¹, Jeffrey E. Lee^{3*}, Ahmed

Ziyyat^{1,2*}, Sandrine Barbaux^{1*}

¹Institut Cochin, Université Paris Cité, INSERM, CNRS, F-75014 Paris, France

²Service d'Histologie, Embryologie Biologie de la Reproduction, AP-HP, Hôpital Cochin, F-75014
Paris, France

³Department of Laboratory Medicine and Pathobiology, Temerty Faculty of Medicine, University of
Toronto, Toronto, ON, Canada

*Co-corresponding authors: sandrine.barbaux@inserm.fr, ahmed.ziyyat@inserm.fr,
jeff.lee@utoronto.ca

Supplementary Table 1: Primary sequences of human TEX51 canonical and computationally mapped isoforms.

Accession number	Primary sequence	Number of residues
A0A1B0GUA7 (canonical)	MLPLLIICLLPAIEGKNCLRCWPELSALIDYDLQILWVTP GPPTELSQNDRHLEEETAKFFTQVHQAIKTLRDDKTVLLE EIIYTHKNLFTERLNKISDGLKEKDIQSTLKVTS CAD CRTH FLSCNDPTFCPARNRRT SLWAVSLSSALLLAIAG DVSFTG KGRRRQ*	166
A0A2R8Y5A7	MLPLLIICLLPAIEGKNCLRCWPELSALIDYDLQILWVTP GPPTELSQSIHSLFLEDNNFLKPWYLDRDHLEEETAKFFT QVHQAIKTLRDDKTVLLEEIIYTHKNLFTERLNKISDGLKE KACNESCDIQSTLKVTS CAD CRTHFLSCNDPTFCPARNR T SLWAVSLSSALLLAIAG GGCFFYWQRKKEAVKVLGKETP TQF*	203
A0A1B0GUP2	MLPLLIICLLPAIEGKNCLRCWPELSALIDYDLQILWVTP GPPTELSQSIHSLFLEDNNFLKPWYLDRDHLEEETAKFFT QVHQAIKTLRDDKTVLLEEIIYTHKNLFTERLNKISDGLKE KDIQSTLKVTS CAD CRTHFLSCNDPTFCPARNRRT SLWAV SLSSALLLAIAG GGCFFYWQRKKEAVKQEQQSPHVFQK*	198
A0A9L9PXS2	MLPLLIICLLPAIEGKNCLRCWPELSALIDYDLQILWVTP GPPTELSQSIHSLFLEDNNFLKPWYLDRDHLEEETAKFFT QVHQAIKTLRDDKTVLLEEIIYTHKNLFTERLNKISDGLKE KACNESCDIQSTLKVTS CAD CRTHFLSCNDPTFCPARNR T SLWAVSLSSALLLAIAG DVSFTGKGRRRQ*	190
A0A9L9PYI6	MLPLLIICLLPAIEGKNCLRCWPELSALIDYDLQILWVTP GPPTELSQNDRHLEEETAKFFTQVHQAIKTLRDDKTVLLE EIIYTHKNLFTERLNKISDGLKEKACNESCDIQSTLKVTSC AD CRTHFLSCNDPTFCPARNRRT SLWAVSLSSALLLAIAG DVSFTGKGRRRQ*	172
A0A1B0GWE4	RDHLEEETAKFFTQVHQAIKTLRDDKTVLLEEIIYTHKNLF TERLNKISDGLKEKACNESCARNRRT SLWAVSLSSALLLAI AG AGTGQPACLPEVNRGRSYHRHKVHSSLGPGDPIPPYP PSWVLGPQSSEA*	132
A0A1B0GUY8	IICLLPAIEGKNCLRCWPELSALIDYDLQILWVTPGPPT LSQSIHSLFLEDNNFLKPWYLDRDHLEEETAKFFTQVHQ IKTLRDGEEAKSPRALGLGLTKAHPPRHSQRPMWPVGG LVKGDECSLSP*	131
A0A1B0GVX2	MLPLLIICLLPAIEGKNCLRCWPELSALIDYDLQILWVTP GPPTELSQNDRHLEEETAKFFTQVHQAIKTLRDDKTVLLE EIIYTHKNLFTERLNKISDGLKEKAPQFLYLLTLPGLQ*	118
A0A1B0GTN5	XKTVLLEEIIYTHKNLFTERLNKISDGLKEKARNRRTSL WAVSLSSALLLAIAGGGRNRAARMSSRSEQRPLPPSQ SSLISGR*	84
A0A1B0GUC9	XCPARNRRTSLWAVSLSSALLLAIAGGCGFFYWQRKKE AVKQEQQSPHVFQK*	52

Colors denote: *signal peptide* (red), *CXXC motif* (blue), and *transmembrane domain* (orange).

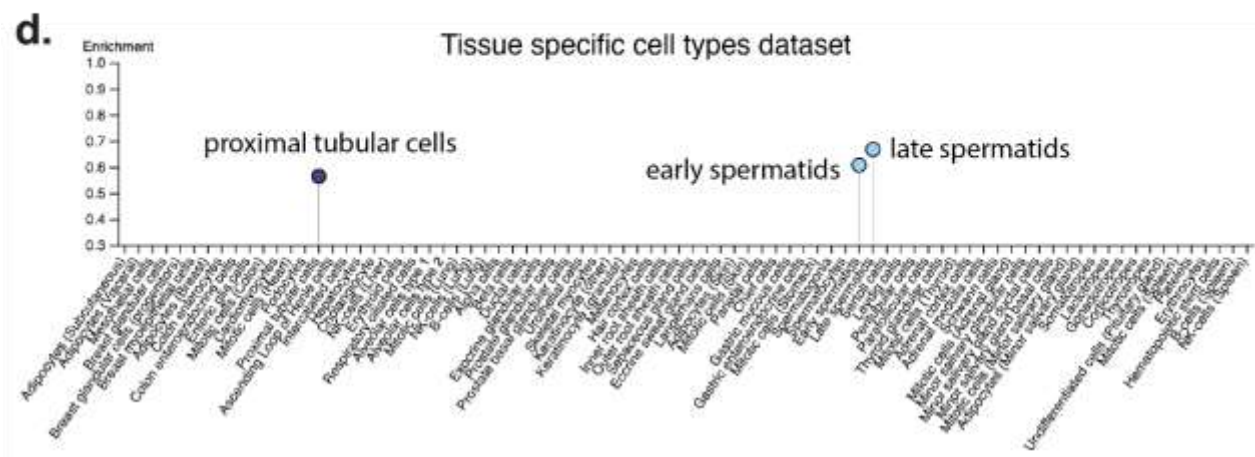
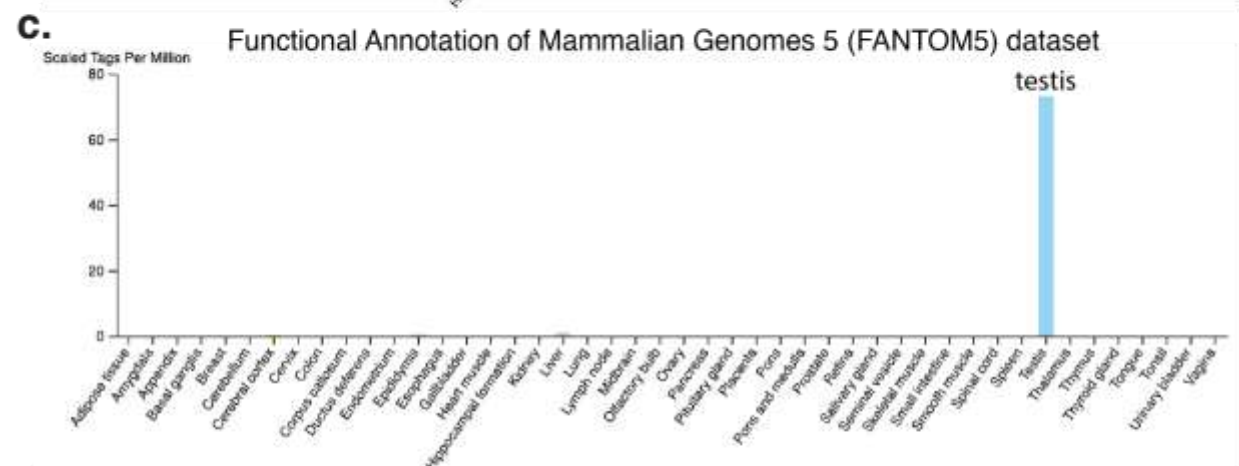
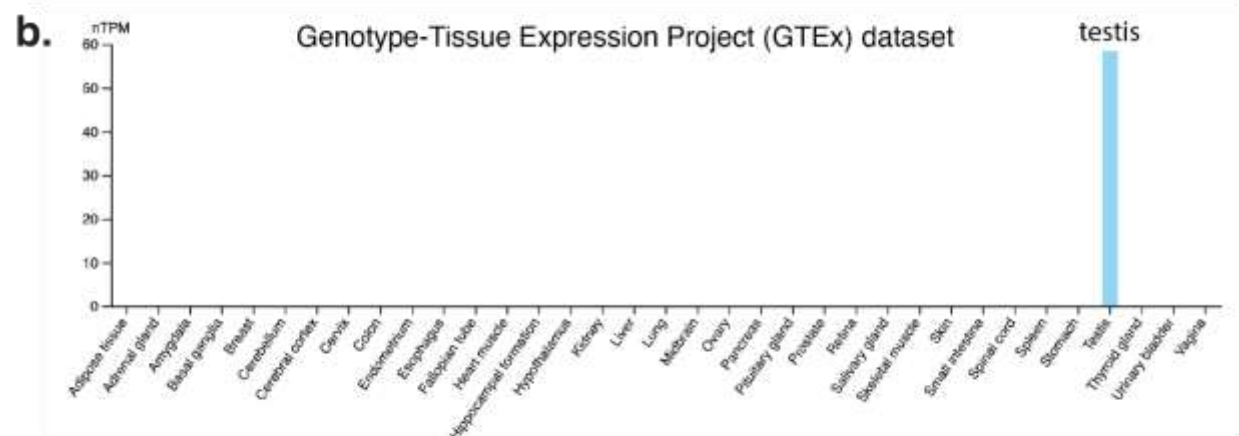
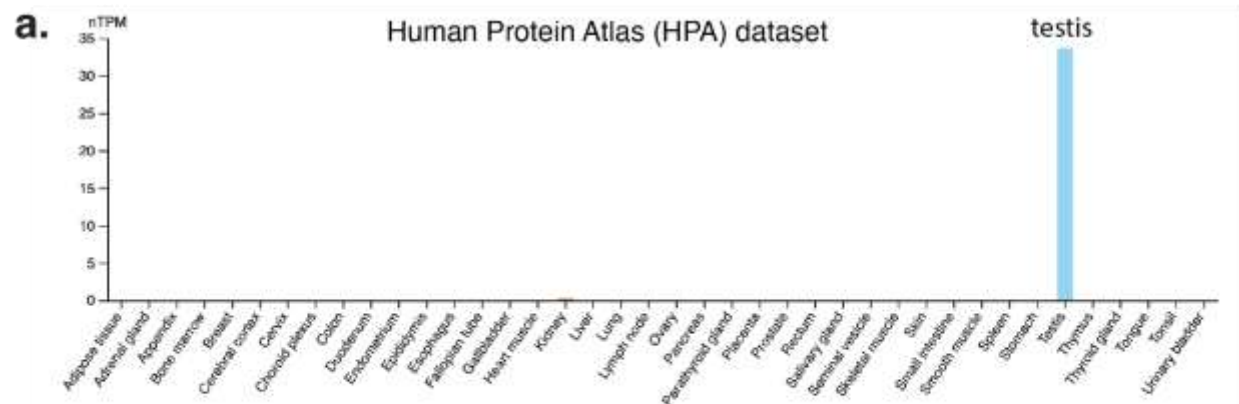
Supplementary Table 2: Computer Assisted Sperm Analysis (CASA) parameters of WT and *Tex51/Gm-KO* sperm.

	WT #1	WT #2	WT #3	<i>Tex51/Gm-KO</i> #1	<i>Tex51/Gm-KO</i> #2	<i>Tex51/Gm-KO</i> #3
Total cells (10⁶/ml)	42.0	36.0	44.0	45.4	37.1	54.1
Static cells (%)	45.3	49.5	49.5	37.1	54.7	50.6
Motile cells (%)	54.7	50.5	50.5	62.9	45.3	49.4
Progressive cells (%)	30.3	32.6	12.9	29.8	30.0	11.7
Slow cells (%)	20.9	13.6	12.5	28.6	12.0	9.4
VAP (μm/s)	72.8	82.6	66.1	67.3	81.7	75.3
VCL (μm/s)	148.2	175.2	128.0	126.5	168.5	146.8
VSL (μm/s)	46.3	53.6	44.8	41.5	54.4	49.6
ALH (μm)	8.8	10.0	7.7	8.7	10.0	9.1
BCF (beat/s)	28.6	28.6	25.0	23.3	27.1	23.0
LIN (%)	31.1	31.6	36.2	32.2	33.6	34.7
STR (%)	61.5	63.6	66.9	59.5	65.4	64.8

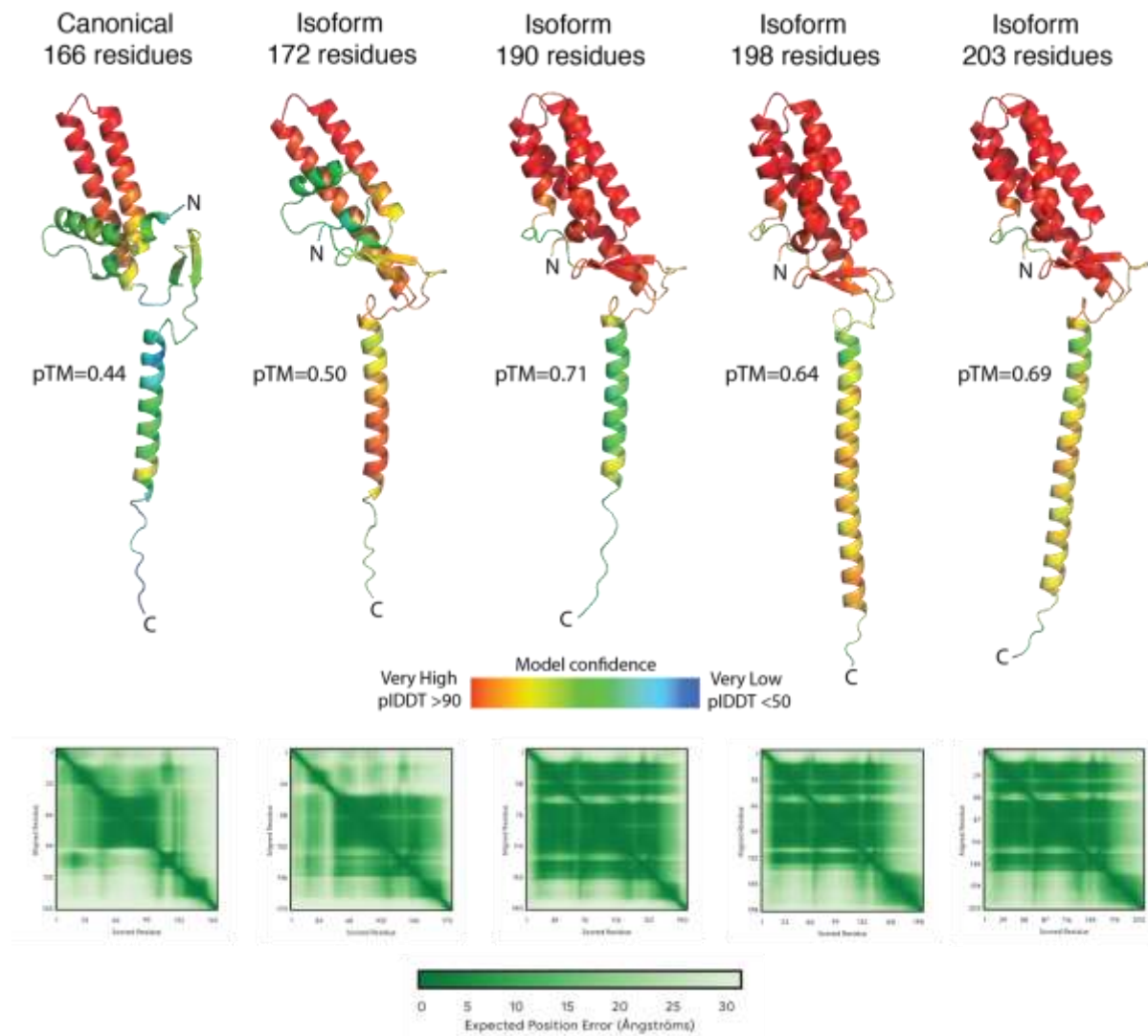
VAP: Velocity along average path; VCL: Curvilinear velocity; VSL: Velocity along straight path; ALH: Amplitude of lateral sperm head displacement; BCF: Beat cross frequency; LIN: Linearity; STR: Straightness

Supplementary Table 3: Sequences, targeted locations and sizes of amplicons for primers and guide RNAs.

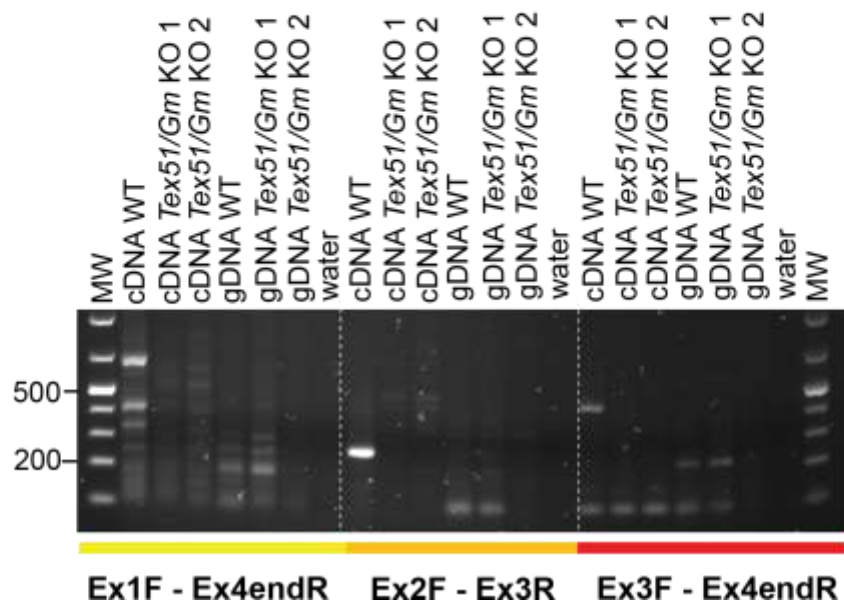
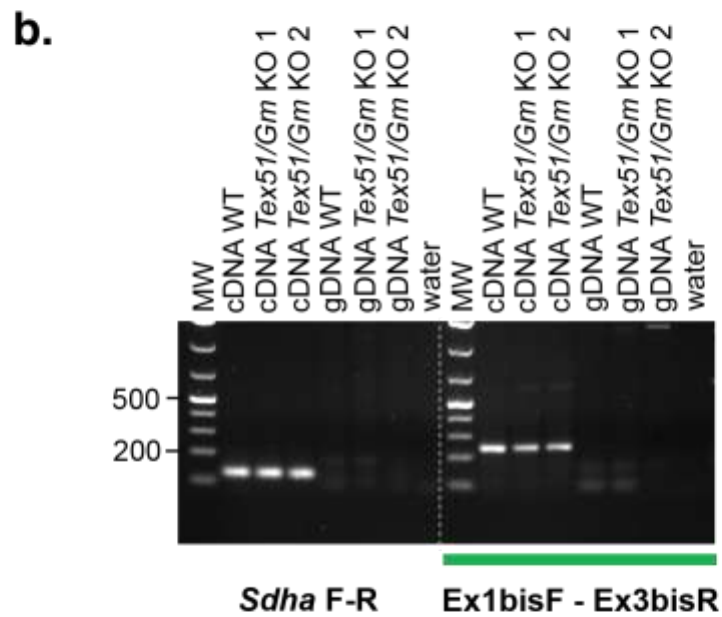
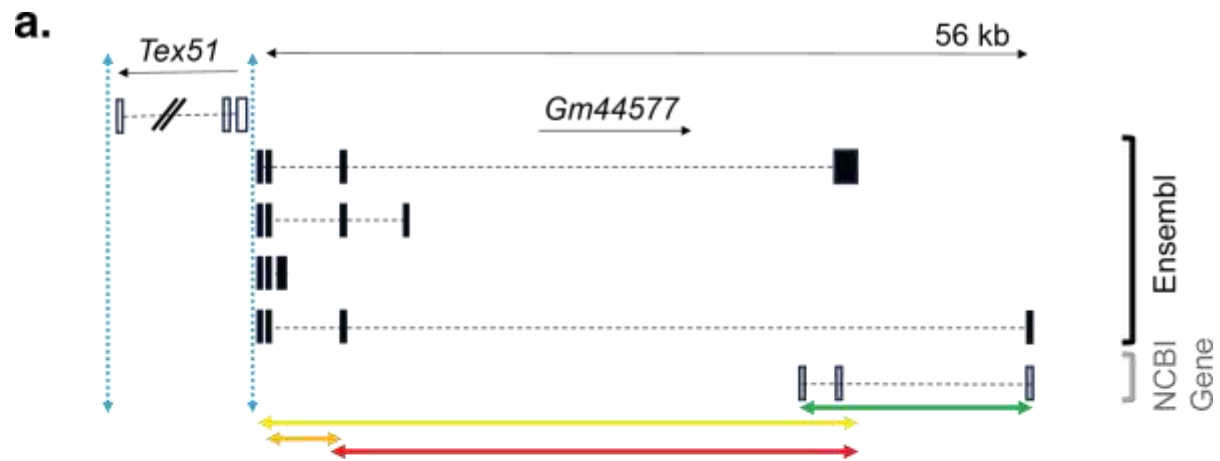
Forward primer	Reverse primer	Gene	Location	Amplicon size (bp)	Application
5F: TCCCTCCATGTTGGCTGTGGTGCC	5R: TGGCCCTGGACTGCCCCAGAGCACC	<i>Tex51</i>	5'UTR-Ex1	281	WT genotyping
3F: CTCTGGCCCTGGTGA CTCTCTATTG	3R: GGGGGACACCAGCCTAGGTGATTTA	<i>Tex51</i>	Ex8-3'UTR	302	WT genotyping
Ex6F: CCAGGGTCACATGGAGCTACAAGTG	Ex8R: GACAGCATGGGCTCTGGCAGCTG	<i>Tex51</i>	Ex6-Ex8	597	WT genotyping
5F: TCCCTCCATGTTGGCTGTGGTGCC	3R: GGGGGACACCAGCCTAGGTGATTTA	<i>Tex51</i>	5'UTR-3'UTR	>4 kb (WT), 238 or 190	Deletion genotyping
Ex1F: GGTGCTCTGGGGCAGTCCAGGG	Ex2R: TCCTCATCCAAATTGTCTCGAG	<i>Tex51</i>	Ex1-Ex2	120 (209 gDNA)	RT-PCR
Ex6F: CCAGGGTCACATGGAGCTACAAGTG	Ex8R: GACAGCATGGGCTCTGGCAGCTG	<i>Tex51</i>	Ex6-Ex8	205 (597 gDNA)	RT-PCR
Sdha F: GGAACACTCCAAAAACAGACCT	Sdha R: CCACCACTGGGTATTGAGTAGAA	<i>Sdha</i>	Ex2-Ex3	106 (1458 gDNA)	RT-PCR
Ex1bis F: AACCATTCCAGAGCTGATTC	Ex3bis R: TGCAGTGATCAAATACCCTGTTT	<i>Gm44577</i>	Ex1bis-Ex3bis	276 (>11kb gDNA)	RT-PCR
Ex1 F: AAAAGAGACTGGAGAAAGTGACG	Ex4end R: TACTCTGAATGTGCTACTGGGA	<i>Gm44577</i>	Ex1-Ex4 (black)	667 (>45kb gDNA)	RT-PCR
Ex2 F: TCATATACCGATACATCCCAGAGG	Ex3 R: CATTGGTTGGGTGCAAATAACTG	<i>Gm44577</i>	Ex2-Ex3 (black)	256 (>6.5kb gDNA)	RT-PCR
Ex3 F: CGCTATTACTGAGGCTATGGAGT	Ex4end R: TACTCTGAATGTGCTACTGGGA	<i>Gm44577</i>	Ex3-Ex4 (black)	459 (>39kb gDNA)	RT-PCR
Guides					
Sg TEX51 UP 1	5' AGAACAAAGGTAAATTCGT3'				
Sg TEX51 UP 2	5' CAATCGGAATGCTGTTGCCC3'				
Sg TEX51 DOWN 1	5' ACAAAGGAGTATGTATTAGG3'				
Sg TEX51 DOWN 2	5' GAGATCTGGTGTAAGTGAT3'				



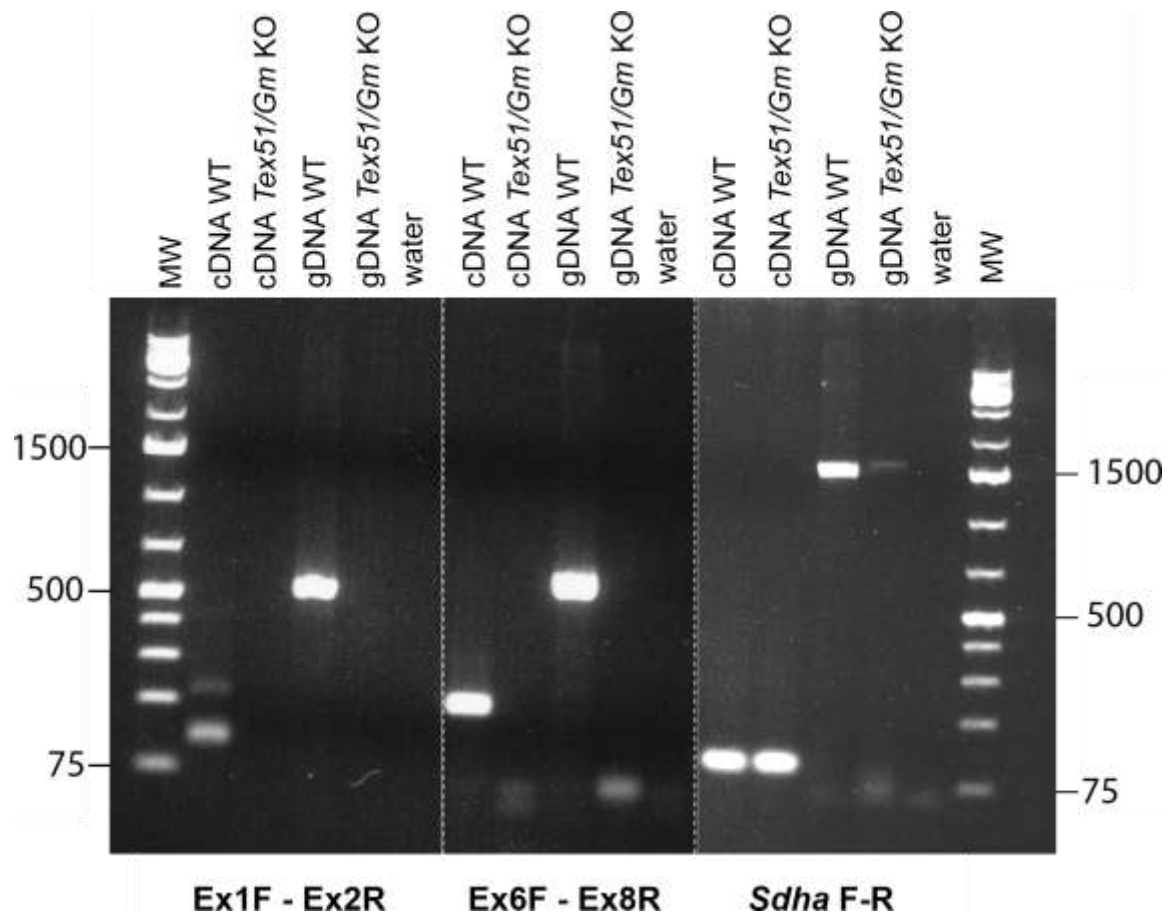
Supplementary Figure 1: *Tex51* expression is restricted to testis tissue based on Human Protein Atlas RNA-seq transcriptomic data. (a) Analysis of Human Protein Atlas (HPA) RNA-seq data for *Tex51* expression in healthy human tissue. (b) Analysis of Genotype-Tissue Expression (GTEx) Project RNA-seq data which surveyed postmortem human tissues for *Tex51* expression. (c) Analysis of Functional Annotation of Mammalian Genomes 5 (FANTOM5) transcriptome database using Cap Analysis of Gene Expression (CAGE) for *Tex51* expression. (d) Analysis of Genotype-Tissue Expression (GTEx) Project RNA-seq data to predict enrichment score for cell type from various tissues. Data from the Human Protein Atlas is available at: <http://www.proteinatlas.org>



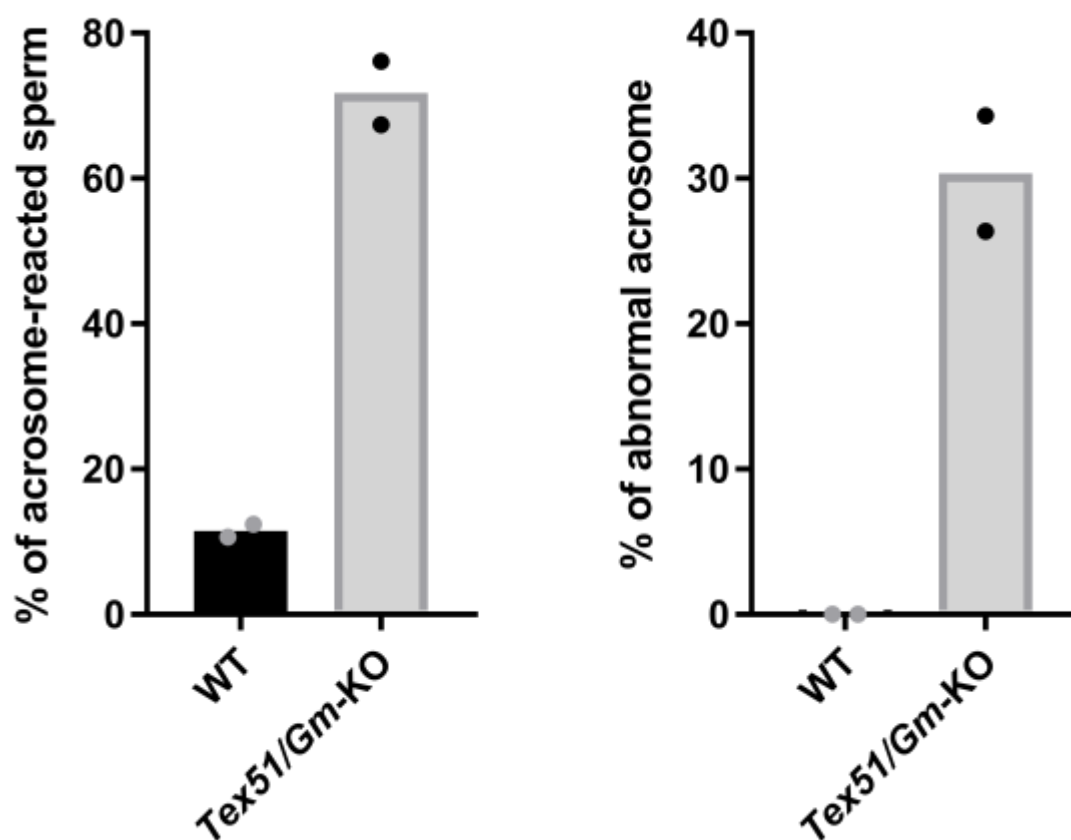
Supplementary Figure 2: TEX51 AlphaFold structural predictions. AlphaFold 3 structural predictions of the Uniprot canonical and computationally predicted isoforms of human TEX51. Structural models of all isoforms include a 4HB, pair of CXXC motifs, and a transmembrane helix. The canonical isoform and the isoform with 172 amino acids have two helices that run perpendicular to a central pair of helices. The TEX51 isoforms with at least 190 amino acids all have a predicted antiparallel 4HB that is similar to the IST superfamily proteins (IZUMO1, SPACA6 and TMEM95). Signal peptides have been removed from the models. Expected positional error graphs also show that the three longest TEX51 isoforms have lowest positional error and greatest model confidence.



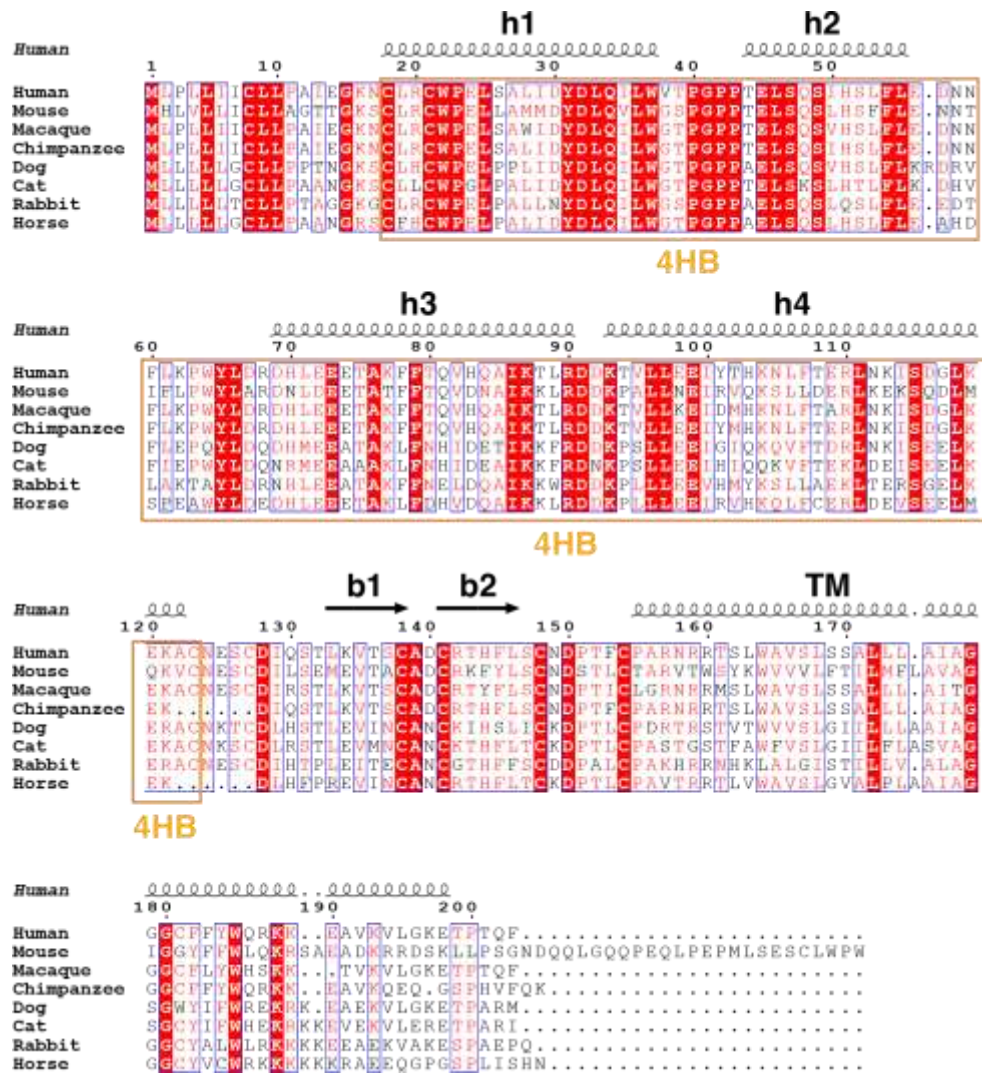
Supplementary Figure 3: A long non-coding RNA is located close to *Tex51*. (a) Map of the mouse *Gm44577* lncRNA gene. The *Tex51* gene (exons in white boxes) and *Gm44577* are in a head-to-head disposition. Presently, different databases report discordant genomic structures (main exons described by Ensembl in black boxes, isoform described by NCBI Gene in grey boxes). Blue dotted arrows show the deletion borders. (b) Expression of a housekeeping gene and *Gm44577* in testis samples. cDNA and genomic DNA (gDNA) were amplified from different regions of *Gm44577* as depicted in color arrows in (a). The Ensembl isoforms are not expressed whereas the NCBI isoform is still detected despite the deletion of *Tex51*. MW: Molecular weight marker



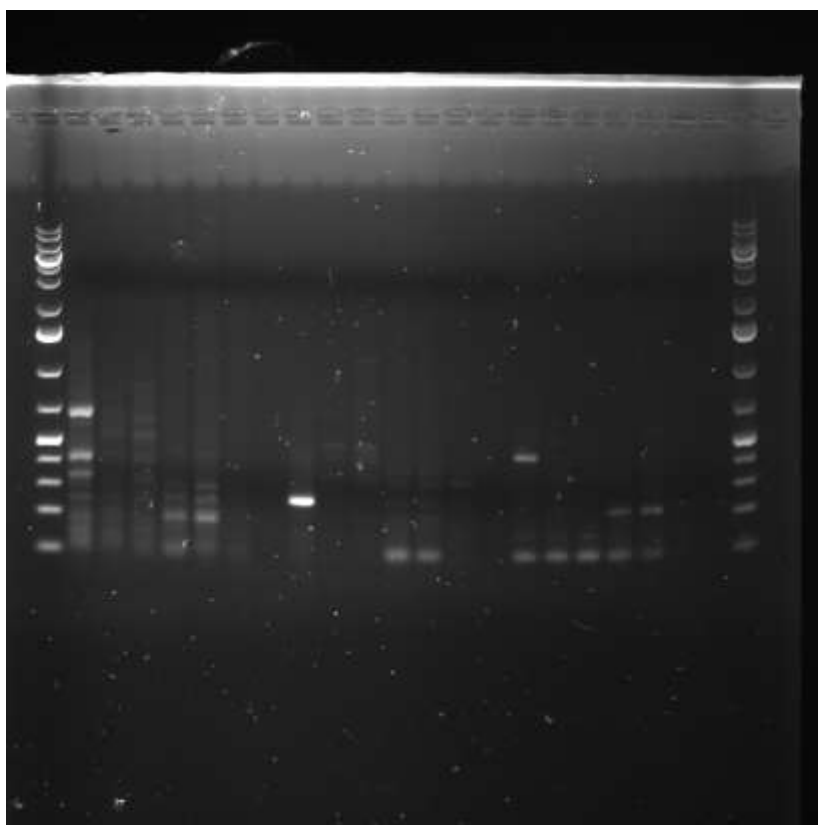
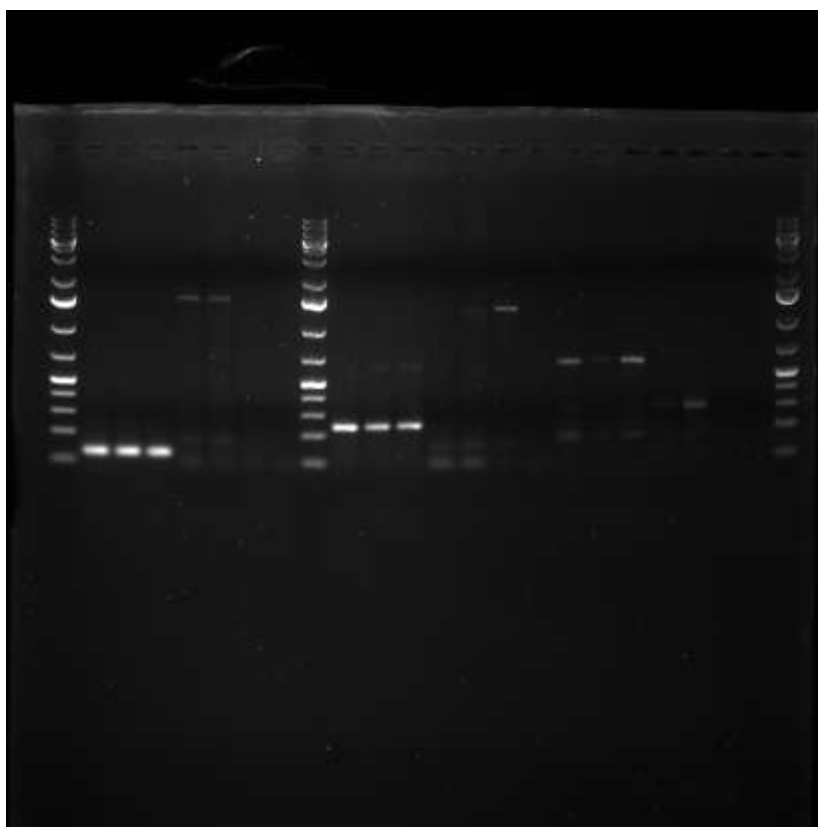
Supplementary Figure 4: Expression of *Tex51* and a housekeeping gene in testis samples. Expression of *Tex51* in testis samples from WT and *Tex51/Gm*-KO males. cDNA and genomic DNA were amplified from different regions of the *Tex51* gene. The *Sdha* gene was used as a control. MW: molecular weight markers.



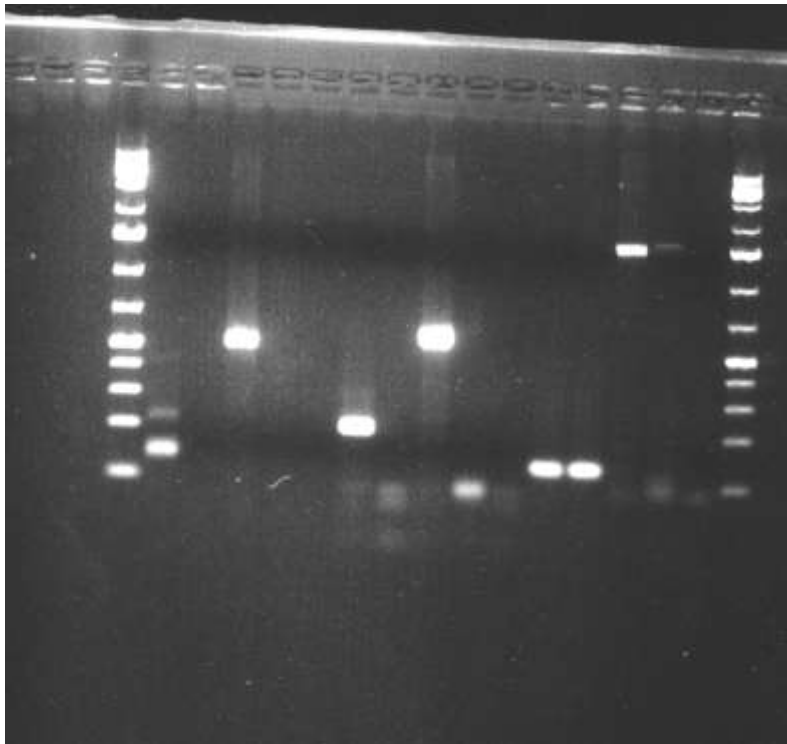
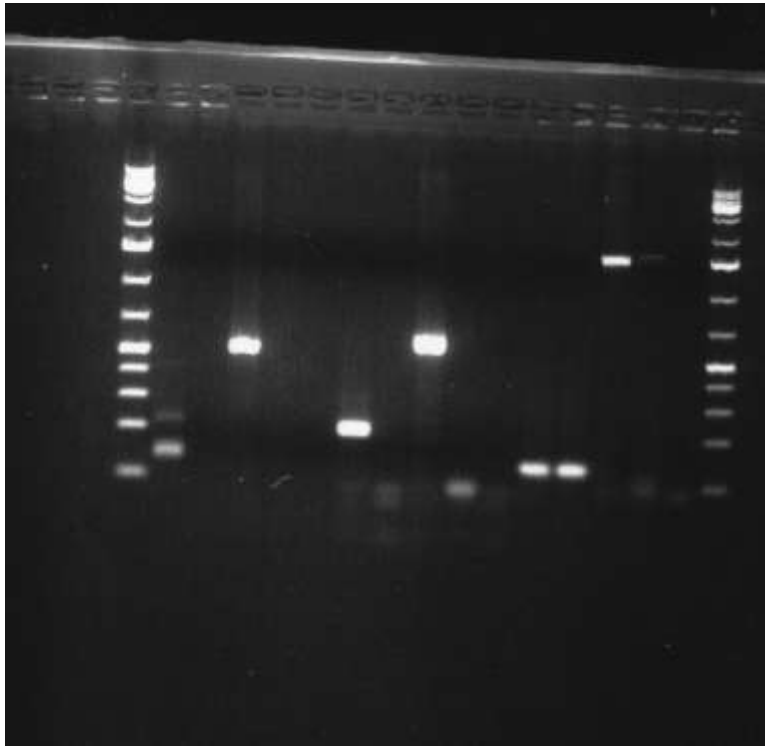
Supplementary Figure 5: *Tex51/Gm-KO* sperm is abnormal. Percentages of acrosome-reacted (left) and acrosome-abnormal sperm heads (right) from WT (black bars) and *Tex51/Gm-KO* (grey bars) males. Data were obtained from 2 independent experiments each involving a total of n=2 *Tex51/Gm-KO* and n=2 WT littermate males. Each sperm sample was split in 2 and treated differently, either as a suspension (this figure) or as a smear (in Figure 6d).



Supplementary Figure 6: Alignment of TEX51 protein sequences, obtained from UniProt, for human (A0A2R8Y5A7), mouse (A0A140LIV7), macaque (A0A7N9CRR7), chimpanzee (A0A2I3SG66), dog (A0A8I3PMS3), cat (A0A337SRC4), rabbit (U3KP20), and horse (A0A9L0S5F5). For human TEX51, the isoform with 203 amino acid was used. AlphaFold predicted helices and β -strands are shown above the sequences with a coil and arrow, respectively. The 4HB domain is highlighted with an orange outline. The analysis was performed with the Clustal Omega and ESPript 3.0 tools.



Uncropped gel images corresponding to Supplementary Figure 3.



Uncropped gel images corresponding to Supplementary Figure 4 at different exposures.

Supplementary Figure 7: Uncropped images of Supplementary Figures

Supplementary Text: *Tex51* Knock-Out strategy and description of deleted alleles of founders

Five different lots of embryos were treated with Cas9 and different combinations of the guide RNAs reported in Supplemental Table 3. A total of 47 live births was recorded. Among them, 29 were carriers of a *Tex51* gene deletion (29/47=61,7%). The amplification corresponding to the deletion (PCR between primer 5F (upstream exon 1) and 3R (downstream the 3'UTR region) describe in Supplementary Table 3) was sequenced for 14 putative founders. One showed an unreadable sequence probably due to multiple alleles, 2 harbored a complex deletion involving the insertion of additional sequence and 11 showed a clear deletion at slightly different nucleotide breakpoints. Among these 11 satisfactory founders, 2 males were randomly chosen for further breeding.

They were each crossed with WT females, respectively. Heterozygotes were identified in their progeny and mated together in brother-sister crosses. Male homozygotes identified as the result of this last cross were considered as *Tex51*-KO subjects and their WT littermates as controls. Both strains were independently further maintained by crosses between heterozygotes.

The sequence of the *Tex51* gene is shown below, retrieved from Ensembl.org.

In **green letters** are untranslated sequences upstream and downstream the gene.

In capitals are exonic sequences, coding in **BLUE** and non-coding in **BROWN** (underlined letters correspond to SNPs).

Highlighted in **yellow** are shown primers 5F and 5R.

Highlighted in **blue** are shown primers 3F and 3R.

Highlighted in **pink** are shown RNA guides.

The control sequence is shown in A).

In B) and C) are the sequences of the 2 founders used for breeding and generation of experimental animals, together with a chromatogram illustrating the sequencing.

A) WT sequence

5' upstream
sequence

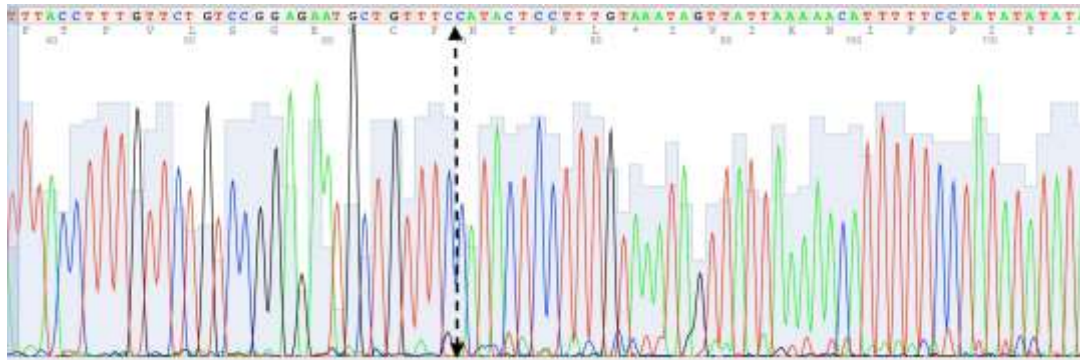
```
agtgcagcaggtcctctgggatgtatcggtatatgaagtcttcaggggagcagaccaact
aggggtctgcaccagtagcaaggaccaggcagaaggggtgggaaggatagaaggaatggg
gggcattaactctttaataaagttttacccaaatagacaatgcacccagtgctccat
ccccagcccctatggccctcagacctctcagtgccagttgacttcttacattgctgttg
acgtcactttctcagtcctcttttctcctccatgttggctgtggtgcctaagcctggt
tcactccagccacccttgctgtgtcccacggaatttaccttgttctgtccggagaatgc
tgttccagggcaacagcattccgattgagactgtgagtacttggaaaggactcggaaag Exon 1
ATGCATCTAGTCTTGCTGATCTGCCTGCTGGCAGGAACCACTGGGAAGAGCTGCCTCCGC
TGCTGGCCAGAGGCTGCTCGCCATGATGGACTATGACCTGCAGGTGCTCTGGGGCAGTCCA
GGGCCACCTACAGAACTCTCACAAAGCCTACACTCTTTTCTGGAGAATAACACCATA
TTCTTGCCCTGGTATCTTG
```

...

Exon 8

```
AGAAGAGACAGCAAGTTACTTTCCTTCAGGAAAGACCAGCAACTAGGGCAGCAACCAGAG
CAGCTGCCAGAGCCCATGCTGTCTGAAAGCTGTCTCTGGCCCTGGTGACTCTCTATTGCA
```


tatgattatcctcacatcctcccagaatcctgccctgtccaggcttgtcccagagctaac
actgatttcct



Supplementary Data 1: SPACA6 DALI AlphaFold human proteome structural database search

See attached Excel spreadsheet; yellow highlighted rows indicate IST superfamily members

Spreadsheet columns are defined as:

Chain- AlphaFold database code and chain

Z- Z-score

rmsd- root mean square deviation (in Å) between SPACA6 and the candidate protein

lali- alignment length (number of residues)

nres- overall number of residues in the candidate protein

%id- percent sequence identity

Supplementary Data 2: Numerical source data.

All source data underlying graphs and charts presented in the main manuscript and Supplementary Information. Numbers between parentheses correspond to mice identity.