**The American Journal of Human Genetics, Volume *107***

## Supplemental Data

# Bi-allelic Mutations in *M1AP* Are a Frequent Cause of Meiotic Arrest and Severely Impaired Spermatogenesis Leading to Male Infertility

**Margot J. Wyrwoll, Şehime G. Temel, Liina Nagirnaja, Manon S. Oud, Alexandra M. Lopes, Godfried W. van der Heijden, James S. Heald, Nadja Rotte, Joachim Wistuba, Marius Wöste, Susanne Ledig, Henrike Krenz, Roos M. Smits, Filipa Carvalho, João Gonçalves, Daniela Fietz, Burcu Türkgenç, Mahmut C. Ergören, Murat Çetinkaya, Murad Başar, Semra Kahraman, Kevin McEleny, Miguel J. Xavier, Helen Turner, Adrian Pilatz, Albrecht Röpke, Martin Dugas, Sabine Kliesch, Nina Neuhaus, GEMINI Consortium, Kenneth I. Aston, Donald F. Conrad, Joris A. Veltman, Corinna Friedrich, and Frank Tüttelmann**

**Figure S1. Prioritization scheme for WES data.**

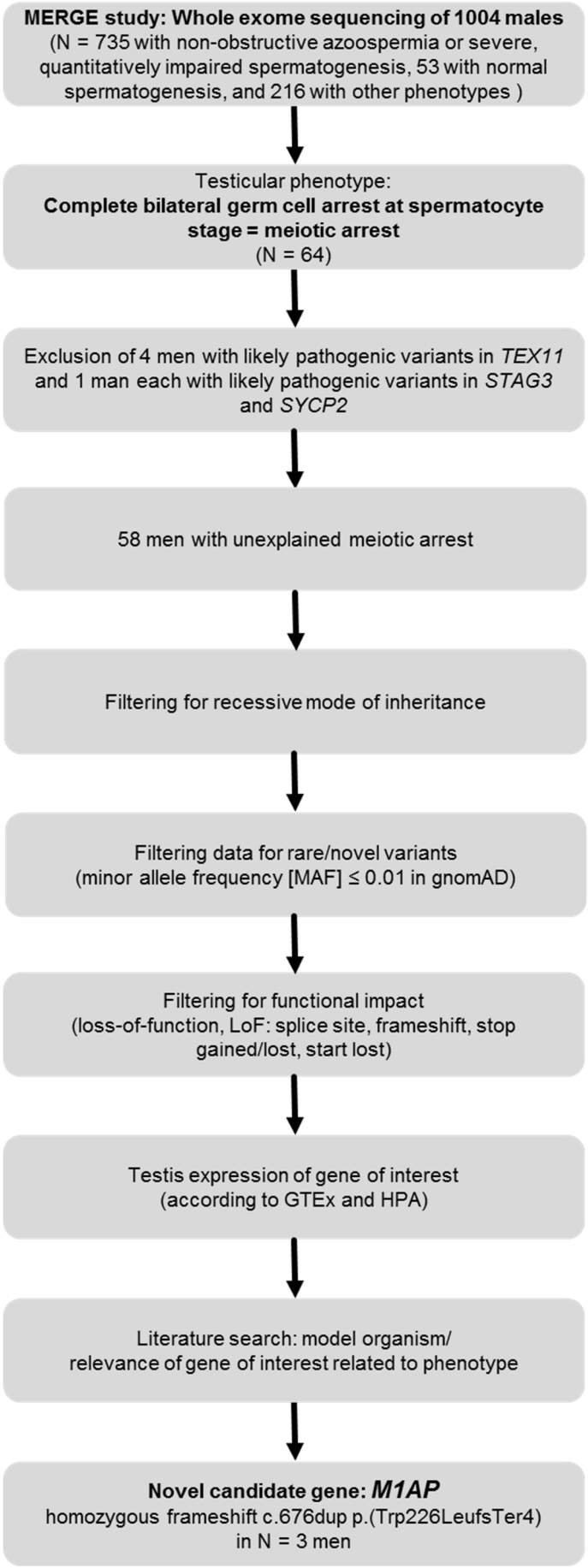
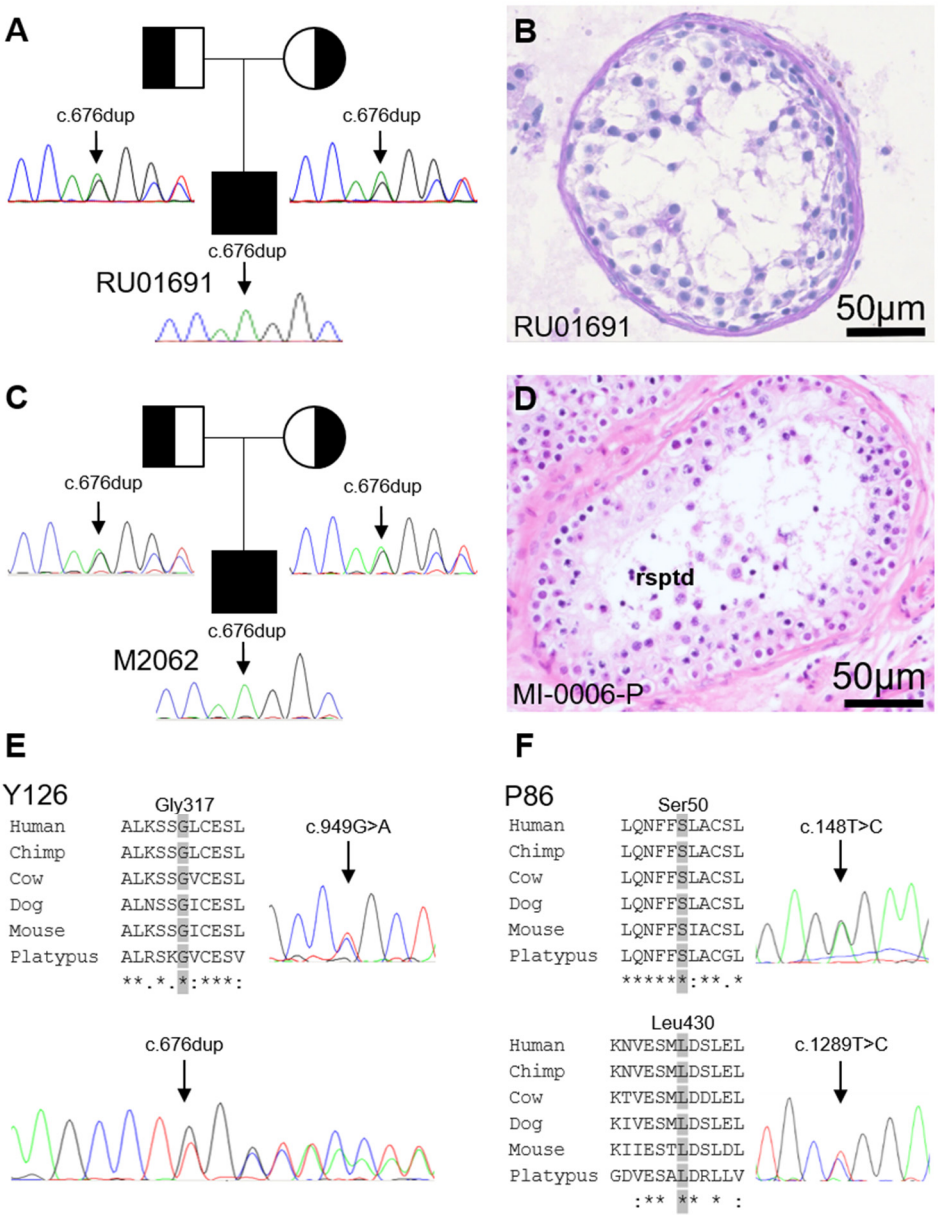
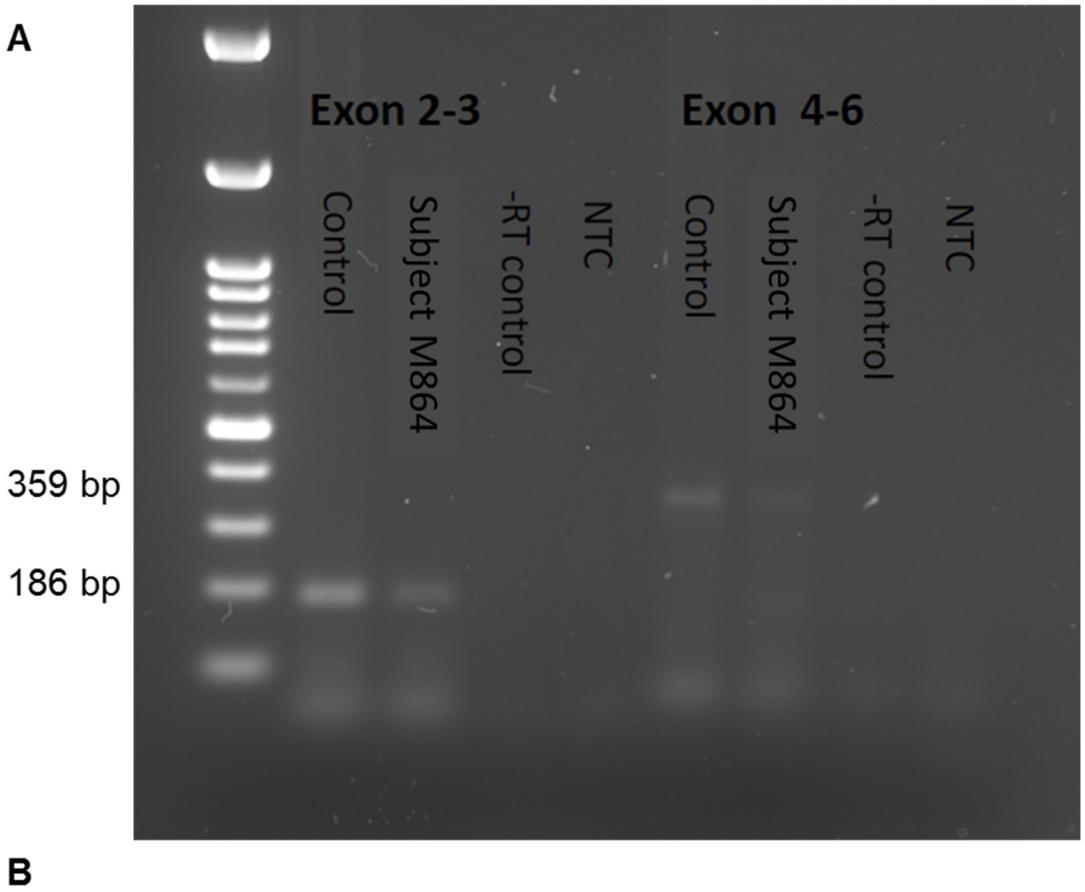


**Figure S2. Variants in *M1AP* in follow-up studies.**



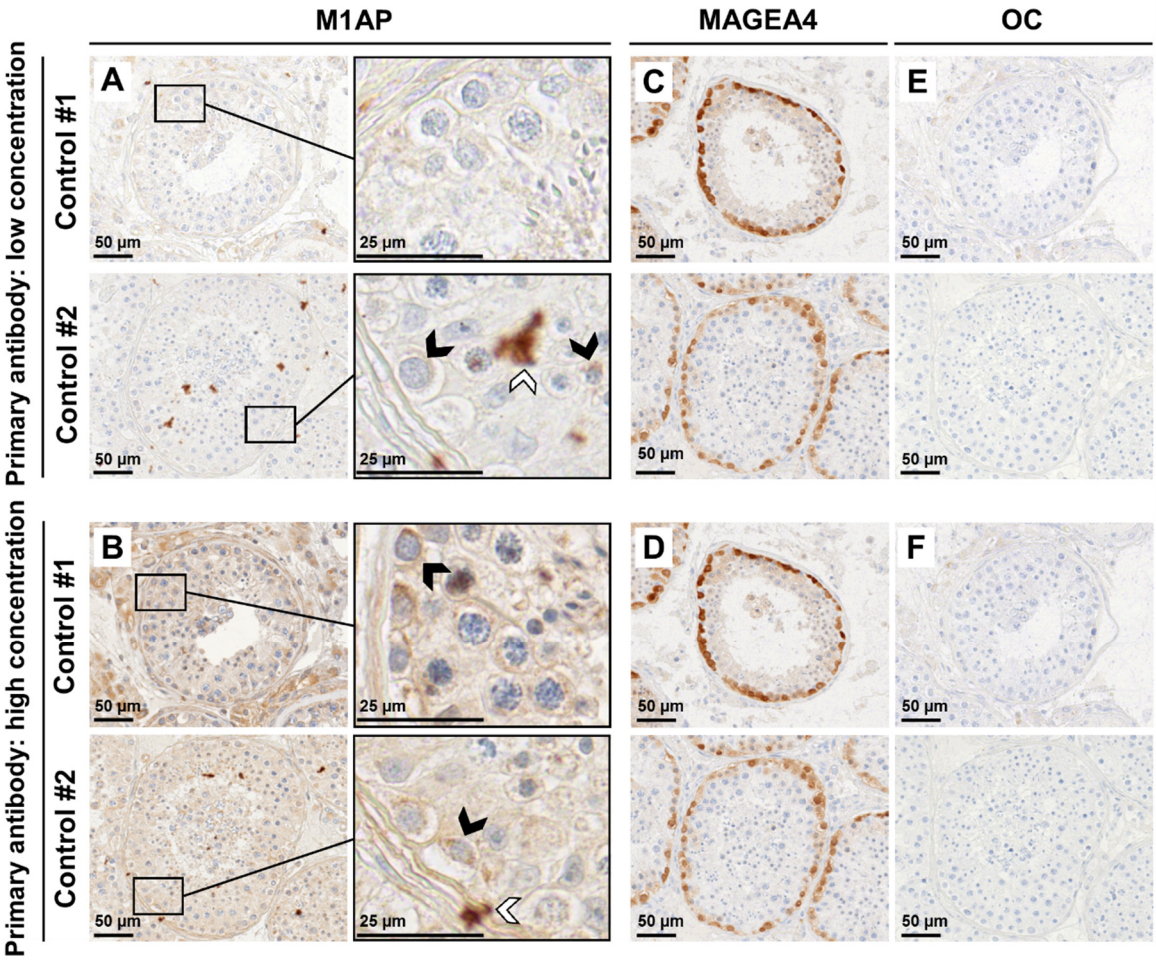

(**A**) Recurrent homozygous duplication c.676dup (p.Trp226LeufsTer4) in *M1AP* in RU01691 from Nijmegen, NL. Both parents are heterozygous for the same variant. (**B**) Testicular histology of RU01691 indicates predominant germ cell arrest at the spermatocyte stage. (**C**) Individual M2062 from Poland with azoo-/cryptozoospermia carries the same homozygous duplication in *M1AP*, also inherited from his heterozygous parents. (**D**) Testicular histology of another homozygous c.676dup subject, MI-0006-P from the UK, revealed a predominant germ cell arrest, but sporadically round spermatids (rsptd) were present in some seminiferous tubules. **(E/F)** Identification of potentially biallelic variants in *M1AP* in Portuguese men from the GEMINI study. (**E**) Y126 carries the recurrent LoF variant c.676dup (p.Trp226LeufsTer4) and the missense variant c.949G>A (p.(Gly317Arg)). (**F**) P86 carries two missense variants c.148T>C(;)1289T>C p.(Ser50Pro)(;)(Leu430Pro). All missense variants affect highly conserved amino acids, as seen from multiple sequence alignments (**E**/**F**).

**Figure S3. Testis RNA analysis of individual M864.**



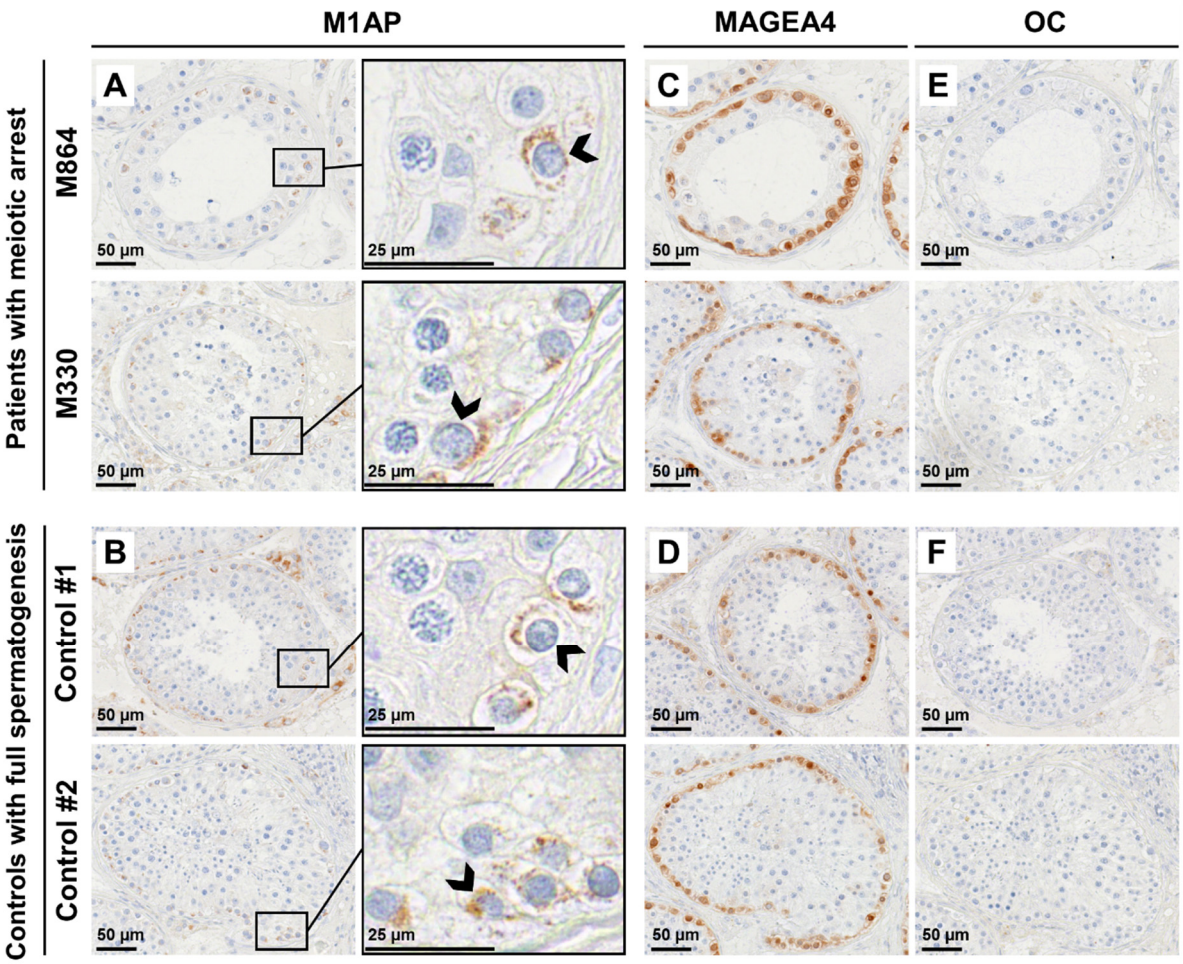

(**A**) PCR of testicular cDNA of exons 2-3 and 4-6 of a control and M864. Exons 2-3 and 4-6 were amplified. Expected band sizes were 186 bp (Exon 2-3) and 359 bp (Exon 4-6). + fertile control DNA, Mutation: M864 DNA, - RT: negative control without reverse transcriptase (RT), NTC: no template control. (**B**) Sanger sequencing of the PCR product amplified from cDNA of M864 was performed according standard procedures and validates the homozygous variant c.676dup in Exon 5.

**Figure S4. Analysis of M1AP protein expression by immunohistochemical staining using #PA5-31627.**



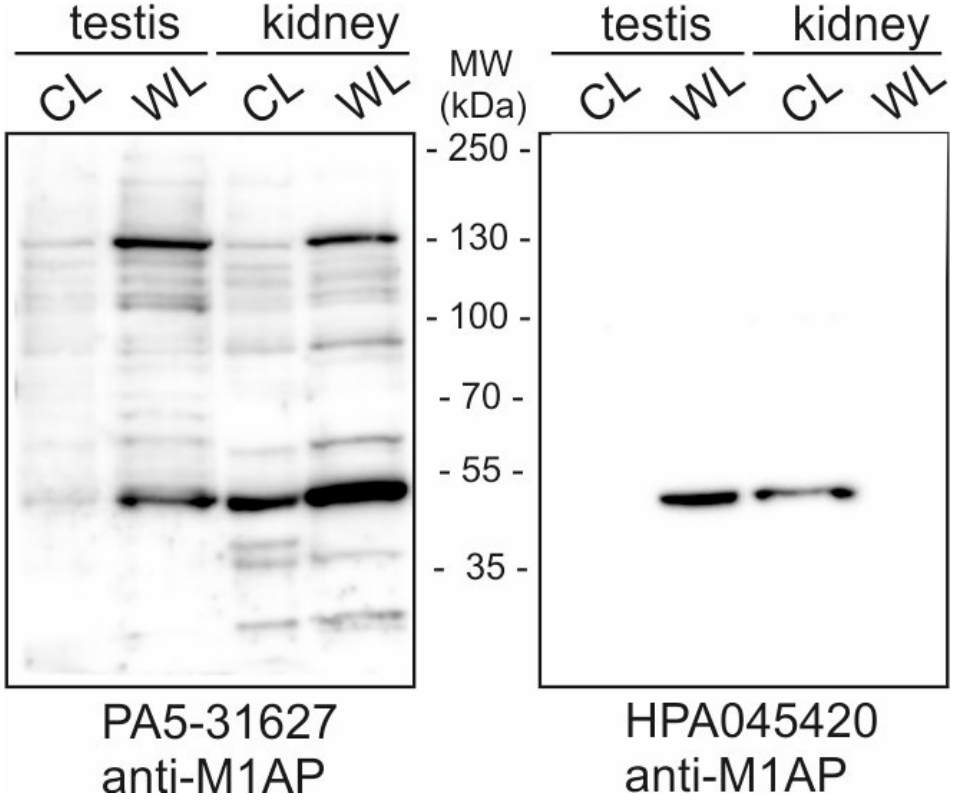

(**A**) Low concentration of M1AP antibody (#PA5-31627, ThermoFisher Scientific, exemplarily depicted for 1:100) was evaluated and black arrow heads indicate spermatogonia, which potentially show weak but specific cytoplasmic protein expression. Additionally, unspecific antibody precipitation in control #2 can be observed (indicated by white arrow head). (**B**) Analysis of higher concentrations of M1AP antibody (exemplarily depicted for 1:50) lead to enhanced unspecific background signal and specific cytoplasmic protein expression primarily in spermatogonia (indicated by black arrow head). (**C**, **D**) MAGEA4 was used as a marker for early germ cells. (**E**, **F**) Omission of primary antibody (OC) is shown as representative technical control. Scale bars are indicated in each micrograph, respectively.

**Figure S5. Analysis of M1AP protein expression by immunohistochemical staining using #HPA045420.**




(**A**) Testis tissue from two individuals M330 and M864 with a homozygous duplication (c.676dup p.Trp226LeufsTer4) in *M1AP* resulting in bilateral meiotic arrest was stained with a commercially available M1AP antibody (#HPA045420, Sigma-Aldrich, exemplarily depicted for 1:100). Representative micrographs show cytoplasmic protein expression specifically in spermatogonia (indicated by black arrow head). (**B**) Sections from men with full spermatogenesis were used as controls. The detected signal did not differ from *M1AP* mutated individuals. Therefore, the antibody does not seem to bind M1AP but pick up a different target instead: both men carry the homozygous frameshift variant p.Trp226LeufsTer4, probably resulting in a truncated protein and consequently leading to a disruption of the supposed epitope of the antibody, which was presumed to reside downstream of the variant and should be specific to M1AP (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). By database search (Ensembl, GTEx), we ruled out that *M1AP* transcript isoforms exist lacking exon 5 and thus none of the isoforms should be escaping the truncation. (**C**, **D**) MAGEA4 was used as a marker for early germ cells. (**E**, **F**) Omission of primary antibody (OC) is shown as representative technical control. Scale bars are indicated in each micrograph, respectively.

**Figure S6. Western blot analysis from tissue lysates.**




Cytosolic (CL) and whole cell (WL) lysates from testis and kidney were obtained from fertile healthy donors. Western blots were stained with two different M1AP antibodies as indicated. Predicted molecular weight of M1AP is 59 kDa. The observed band size of almost 55 kDa is similar to those expected for PA5-31627 antibody by the manufacturer's instructions. In contrast, validation experiments of the Human Protein Atlas of HPA045420 antibody detect protein bands of ~30 kDa in different tissues and cell lines. Wether different band sizes may result from different post-translational modifications remains elusive due to the lack of knowledge about M1AP protein. Furthermore, the Western blot did not only result in detection of a band in testis lysate, but also in kidney lysate where M1AP is putatively not expressed (GTEx, HPA). Taken the results from IHC staining and Western blot analysis together, the specificity of both antibodies is questionable.

**Table S1. Primer sequences.**

| Primer | Sequence Forward | Sequence Reverse |
|---|---|---|
| Validation of c.676dup (MERGE study) | TGGGTCTGGAAATGTTGCTGA | GATTGCTAGAGCCCAGGCAT |
| qPCR of Exon 5 in *M1AP* | TCTGGGAACTGACATTGACCTTC | TGGGTCTGGAAATGTTGCTGA |
| Sanger-Sequencing of all exons of *M1AP* in heterozygous control men | | |
| Exon 2 | TGGATTTTCTCTTCAACAGTACACA | TGTGCACCTGTAGTCCTAGC |
| Exon 3 | CAGTTTTTCCTCATAATTCACTTCAGT | TCCTTTCATGTTTCTGGTAACTCT |
| Exon 4 | CCTCAGTGAAATCTGCTGGC | GCCTGATTGGAAAGGTCCTGT |
| Exon 5 | CTAACTGGCCCTTGCTGGTT | TGCTAGAGCCCAGGCATTTG |
| Exon 6 | CACCATCTGCACATTTGGCC | AACCAGTCAGGCTTTCCTCT |
| Exon 7 | ACAGAATATATATCTAGGGCTTGACAC | GAGTCTGCTTCAACTCTTCCCA |
| Exon 8 | GCCGAAGTTAAATGGCTCTG | TGGCATAATTGCCTATCCTT |
| Exon 9+10 | GGGGGACAGCATCTATTTCA | TTCCCTCTTCAACCCCAACT |
| Exon 11 | CCTTGAGGCTGTCACTCCA | CTTGCTGGAGAAAGGACAGG |
| RNA analyses | | |
| M1AP cDNA Ex2-3 | ACATTGCTCTACCGTCCTGG | TGCAACCTAGCAAAGTTCCCT |
| M1AP cDNA Ex4-6 | CCTAGCCAGAGTCAGGAGGT | ATTCTCAAGGAGCCGTCAGC |
| Mutagenesis Primer | | |
| M1AP mut p.W226L dup | GATTTCTTCAAAGCCTTGGCTACATAA CAGTGG | CCACTGTTATGTAGCCAAGGCTT TGAAGAAATC |

**Table S2. qPCR results to exclude the hemizygosity of LoF variant c.676dup in exon 5 of *M1AP* of men from the MERGE study.**

| Individual | Repeat | Crossing Point Exon 5 *M1AP* | Concentration Exon 5 *M1AP* | Mean concentration Exon 5 *M1AP* | Crossing Point *Albumin* | Concentration *Albumin* | Mean concentration *Albumin* | Ratio mean concentration Exon 5 *M1AP* / mean concentration *Albumin* |
|---|---|---|---|---|---|---|---|---|
| M330 | 1 | 21.58 | 10.2 | | 22.64 | 9.42 | | |
| M330 | 2 | 21.57 | 10.3 | 10.03 | 22.65 | 9.35 | 9.40 | 1.10 |
| M330 | 3 | 21.57 | 10.3 | | 22.64 | 9.43 | | |
| M864 | 1 | 22.62 | 4.96 | | 23.46 | 5.34 | | |
| M864 | 2 | 22.58 | 5.12 | 5.04 | 23.46 | 5.34 | 5.35 | 0.94 |
| M864 | 3 | 22.60 | 5.03 | | 23.45 | 5.37 | | |
| M1792 | 1 | 21.96 | 7.86 | | 23.21 | 6.37 | | |
| M1792 | 2 | 21.97 | 7.79 | 7.80 | 23.12 | 6.77 | 6.78 | 1.16 |
| M1792 | 3 | 21.98 | 7.73 | | 23.03 | 7.20 | | |
| M2062 | 1 | 24.49 | 5.76 | | 22.61 | 11.40 | | |
| M2062 | 2 | 23.84 | 9.02 | 8.56 | 22.68 | 10.80 | 11.1 | 0.77 |
| M2062 | 3 | 23.57 | 10.89 | | 22.64 | 11.10 | | |
| Control | 1 | 21.62 | 10.0 | 10.0 | 22.54 | 10.0 | 10.0 | 1.0 |
| Control | 2 | 21.61 | 10.0 | 10.0 | 22.55 | 10.0 | 10.0 | 1.0 |
| Control | 3 | 21.60 | 10.0 | 10.0 | 22.57 | 10.0 | 10.0 | 10.0 |
| Control | 4 | 23.6 | 10.0 | 10.0 | 22.36 | 10.0 | 10.0 | 10.0 |

**Table S3. Top 50-List of PSAP results of individuals with *M1AP* variants.**

(see respective Excel-file)

**Table S4. Classification of *M1AP* variants.**

| cDNA change | Protein change | Categories | Classification according to ACMG-AMP guidelines[1] |
|---|---|---|---|
| c.676dup | p.Trp226LeufsTer4 | PVS1, PS3, PM2, PP4 | Pathogenic |
| c.148T>C | p.(Ser50Pro) | PM1, PP3, PP4 | Uncertain significance |
| c.797G>A | p.(Arg266Gln) | PP3, PP4 | Uncertain significance |
| c.949G>A | p.(Gly317Arg) | PM2, PP3, PP4 | Uncertain significance |
| c.1166C>T | p.(Pro389Leu) | PM2, PP1, PP3, PP4 | Uncertain significance |
| c.1289T>C | p.(Leu430Pro) | PM1, PP3, PP4 | Uncertain significance |
| c.1435-1G>A[2] | ? | PVS1, PS3, PM2, PP4 | Pathogenic |

**Table S5. Structured clinical validity assessment of *M1AP* according to Smith et al. 2018[3].**

| | Points | Comments/Reference |
|---|---|---|
| **Number of unrelated individuals** (1-2 → 1 pt, 3-4 → 2 pt, 5-9 → 3 pt, 10-24 → 4 pt; > 25 → classification "definitive") | 4 | 11 individuals (M330, M864, M1792, M1943, M2062, Y126, P86, RU01691, MI-0006-P, T1024, F1:II-1[2]) |
| **Other statistical evidence** (AD disease with significant excess of de novos OR AR disease with e.g. LOD score > 3 → 1 pt) | 1 | Consanguineous Turkish family (homozygous variant segregates with infertility, heterozygous carriers fertile, LOD score 3.28) |
| **Number of publications reporting independent probands** (per publication 1 pt, max. 3) | 1 | This publication and Tu *et al.* 2020[2] |
| **Number of pathogenic variants** (per VLP or mutation 1 pt, max. 4) | 2 | c.676dup (p.Trp226LeufsTer4)<br>c.1435-1G>A (p.?)[2] |
| **Gene function** (function/expression consistent with disease → 1 pt | 1 | Testicular expression in mice, expressed in last stages of spermatogenesis,[4] Human Protein Atlas, GTEx |
| …and/or physically interacts with gene characterized for same disease → 1 pt) | 0 | |
| **Gene disruption** (relevant pathology *in vitro* after similar genetic modification → 1 pt | 1 | *in vitro* expression in HEK293T cells demonstrates putatively truncated protein |
| ...and/or determination of mutational mechanism → 1 pt) | 1 | Frameshift variant → truncated protein → LoF (this study: analysis of testis RNA, heterologous expression) |
| **Model organism** (gene function *in vivo* related to pathology of human disease → 1 pt | 1 | Male KO-mouse is infertile[5] |
| …and/or phenotype and genotype match human disease → 1 pt) | 1 | Male knockout mouse exhibits meiotic arrest and severe oligozoospermia[5] |
| **TOTAL POINTS** | 13 | |
| **Classification**<br>no evidence (0-4 pts)<br>limited (2-9 pts)<br>moderate (8-12 pts)<br>strong (13+ pts)<br>definitive (canonical) | strong | |

**Table S6. Members of the GEMINI consortium.**

| Group Leaders | Institution |
|---|---|
| Donald F. Conrad, Liina Nagirnaja | Department of Genetics, Oregon National Primate Research Center, Oregon Health & Science University, Beaverton, OR, USA |
| Kenneth I. Aston, Douglas T. Carrell, James M. Hotaling, Timothy G. Jenkins | Andrology and IVF Laboratory, Department of Surgery (Urology), University of Utah School of Medicine, Salt Lake City, UT, USA |
| Rob McLachlan[1,2] | 1) Hudson Institute of Medical Research and the Department of Obstetrics and Gynaecology, Monash University, Clayton, Victoria, Australia<br>2) Monash IVF and the Hudson Institute of Medical Research, Clayton, Victoria, Australia |
| Moira K. O'Bryan | School of Biological Sciences, Monash University, Clayton, Victoria, Australia |
| Peter N. Schlegel | Department of Urology, Weill Cornell Medicine, New York, NY, USA |
| Michael L. Eisenberg | Department of Urology, Stanford University School of Medicine, Stanford, CA 94305, USA |
| Jay I. Sandlow | Department of Urology, Medical College of Wisconsin, Milwaukee, WI, 53226, USA |
| Emily S. Jungheim, Kenan R. Omurtag | Washington University in St Louis, School of Medicine, St Louis, MO, USA |
| Alexandra M. Lopes[1,2], Susana Seixas[1,2], Filipa Carvalho[1,3], Susana Fernandes[1,3], Alberto Barros[1,3] | 1) i3S - Instituto de Investigação e Inovação em Saúde , Universidade do University of Porto<br>2) IPATIMUP - Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal<br>3) Serviço de Genética, Departamento de Patologia, Faculdade de Medicina da Universidade do Porto, Porto, Portugal |
| João Gonçalves[1,2], Iris Caetano[1], Graça Pinto[3], Sónia Correia[3] | 1) Departamento de Genética Humana, Instituto Nacional de Saúde Dr Ricardo Jorge, Lisboa, Portugal<br>2) ToxOmics, Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Portugal<br>3) Centro de Medicina Reprodutiva, Maternidade Dr. Alfredo da Costa, Lisboa, Portugal |
| Maris Laan | Institute of Biomedicine and Translational Medicine, University of Tartu, 51010 Tartu, Estonia |
| Margus Punab | Andrology Center, Tartu University Hospital, 50406 Tartu, Estonia |
| Ewa Rajpert-De Meyts, Niels Jørgensen, Kristian Almstrup | Department of Growth and Reproduction, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark |
| Csilla G. Krausz[1,2] | 1) Department of Experimental and Clinical Biomedical Sciences, University of Florence, Florence, Italy<br>2) Andrology Department, Fundacio Puigvert, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Spain |
| Keith A. Jarvi | Division of Urology, Department of Surgery, Mount Sinai Hospital and; Institute of Medical Sciences, University of Toronto, Toronto, Ontario, Canada |

## Supplemental methods

*Individual's consent and study approval*

All participants gave written informed consent for the evaluation of their clinical data and analysis of their DNA samples. The study protocol was approved by the respective Ethics Committees/Institutional Review Boards (Ref. No. Münster: 2010-578-f-S, Gießen: 26/11, GEMINI consortium: 201502059, Porto: PTDC/SAU-GMG/101229/2008, Nijmegen: NL50495.091.14 version 4, Newcastle: REC Ref: 18/NE/0089, Bursa: 05.01.2015/04).

*Whole exome sequencing (WES) and bioinformatics analysis (MERGE study)*

Genomic DNA was extracted from peripheral blood leukocytes via standard methods. WES sample preparation and enrichment were carried out in accordance with the protocols of either Agilent's SureSelect$^{QXT}$ Target Enrichment kit or Twist Bioscience's Twist Human Core Exome kit. Agilent's SureSelect$^{XT}$ Human All Exon Kits V4, V5 and V6 or Twist Bioscience's Human Core Exome plus RefSeq spike-in's were used to capture libraries. For multiplexed sequencing, the libraries were index tagged using appropriate pairs of index primers. Quantity and quality of the libraries were assessed with the ThermoFisher Qubit and Agilent's TapeStation 2200, respectively. Sequencing was conducted on the Illumina HiScan®SQ, NextSeq®500, or HiSeqX® systems using the TruSeq SBS Kit v3 - HS (200 cycles), the NextSeq 500 V2 High-Output Kit (300 cycles) or the HiSeq Rapid SBS Kit V2 (300 cycles), respectively.

After trimming, Cutadapt v1.15 was used to remove the remaining adapter sequences and primers.[6] Sequence reads were aligned against the reference genome GRCh37.p13 using BWA Mem v0.7.17.[7] We excluded duplicate reads and reads that mapped to multiple locations in the genome from further analysis. Small insertions/deletions (indels) and single nucleotide variations were identified and quality-filtered by GATK toolkit v3.8 with HaplotypeCaller, in accordance with the best practice recommendations.[8] Ensembl Variant Effect Predictor was used to annotate called variants.[9]

*WES and bioinformatic analysis in individual RU01691 and MI-0006P*

RU01691 came from a cohort of 171 offspring-parent trios and 112 singleton cases (N = 283) of men with severe oligozoospermia (<5 million sperm/ml; N = 69) or non-obstructive azoospermia (N = 214). MI-0006P belongs to a cohort of 16 offspring-parent trios and 32 singleton cases (N = 48) of men with unexplained azoospermia (N = 36) or oligozoospermia (N = 12). WES samples were prepared and enriched following the manufacturer's the protocols

of either Illumina's Illumina's Nextera DNA Exome Capture kit or Twist Bioscience's Twist Human Core Exome Kit. All sequencing was performed on the NovaSeq 6000 Sequencing System (Illumina) at an average depth of 72x (Illumina's Nextera Kit) and 99x (Twist Bioscience's Kit). Sequenced reads were aligned to GRCh37.p5 using BWA Mem v0.7.17, Picard and GATK v4.1.4.1. Following best practice recommendations single nucleotide variantions and small indels were identified and quality-filtered using GATK's HaplotypeCaller. Ensembl's Variant Effect Predictor (VEP) was used to fully annotate detected variants.

*WES and bioinformatic analysis in individual Y126 and P86*

Genetics of Male Infertility Initiative (GEMINI; https://gemini.conradlab.org/) is a multi-center consortium dedicated to identifying and describing the underlying genetic causes of male infertility. To date, N = 930 men with non-obstructive azoospermia have been studied using the whole-exome sequencing (WES) approach. In short, WES was performed at the McDonnell Genome Institute of Washington University (genome.wustl.edu) on Illumina HiSeq 4000 and using an in-house exome targeting reagent which captures 39.1 Mb of exome at an average coverage of 80x. The sequence reads were aligned to hg38 using bwa-mem, Picard and Genome Analysis Toolkit (GATK; https://software.broadinstitute.org/gatk) in an alternate contig-aware manner. Genotype calling was performed jointly for all samples using GATK tools and following their procedures of best practices. The resulting genotype callset was subjected to thorough quality control procedures, including but not limited to removing positions with high missingness rate (>15%) and removing samples with low coverage (<30x), high contamination (VerifyBamID freemix >5%) or low call rate (<85%). Individual genotypes with read depth (DP) <10x and genotype quality (GQ) <30 were further excluded from the data. All sequenced cases were screened for the known infertility causes such as Klinefelter syndrome, deleterious CFTR mutations, Y-chromosome microdeletions and large structural variation on sex chromosomes utilizing the WES genotype dataset.

In order to prioritize the deleterious lesions most likely disrupting the function of the respective genes and potentially leading to the disease phenotype, a modified version of the population sampling probability (PSAP) software (https://github.com/conradlab/PSAP) was applied to the WES genotype callset. The genomic coordinates of identified variation were lifted over to hg19 for the PSAP analysis. The list of prioritized variants was subsequently filtered by only including mutations with PSAP popScore values less than $10^{-4}$ and minor allele frequency <1% across all populations in the gnomAD database (v2.1.1, https://gnomad.broadinstitute.org/). As the study aims to identify rare DNA lesions most likely observed in few, if not singleton, cases, genes enriched for and positions found to be commonly affected by rare deleterious variation among the cases were excluded from the study.

*WES and bioinformatic analysis in individual T1024*

Genomic DNA was extracted from peripheral venous blood using the QIAamp® DNA Mini Kit (QIAGEN, Ankara, Turkey). SureSelectXT Library Prep Kit was used for target enrichment. All procedures were carried out according to the manufacturer's protocols. Paired-end sequencing was performed on an Illumina NovaSeq system with a read length of 151. Base calling and image analysis were conducted using Illumina's Real-Time Analysis software. The BCL (base calls) binary is converted into FASTQ utilizing Illumina package bcl2fastq. All bioinformatics analysis performed on Sophia DDMTM platform which includes algorithms for alignment, calling SNPs and small indels (Pepper), calling copy number variations (Muskat) and functional annotation (Moka). Raw reads were aligned to the human reference genome (GRCh37/hg19). Variant filtering and interpretation performed on Sophia DDMTM. Integrative Genomics Viewer (IGV)16 was used to bam file visualization. In families with consanguineous marriages, the homozygosity mapping was carried out with HomSI.

*Attempt at 3D modelling of M1AP protein*

The longest coding transcript of *M1AP* (GENCODE: ENST00000290536.5, RefSeq: NM_001281296.1, NM_138804.4) translates to the protein sequence with UniProt identifier Q8TC57. This sequence was used to perform a basic local alignment search tool (BLAST) to the sequences of known protein structures in the protein data bank (PDB). Four structures were obtained that partially match the M1AP sequence.

1. PDB ID: 5CK3 (Chain B, D, and F) with sequence identity: 29%
2. PDB ID: 5CK4 (Chain A, and B) with sequence identity: 29%
3. PDB ID: 5CK5 (Chain A, B, C, and D) with sequence identity: 29%
4. PDB ID: 6EGC (Chain A) with sequence identity: 24%

The highest sequence identity of these was 29%. Using these template structures will create unreliable results for a homology model, as templates with below 30% identity are likely to lead to serious mispredictions.

*Quantitative PCR analysis*

Quantitative PCR (qPCR) was carried out in 96-well plates on the LightCycler 480 using the manufacturer's default settings. The ALB (albumin) gene was used for normalization. The reactions were performed in triplicates using the SensiMix Real-Time PCR Kit (Bioline). The

PCR consisted of an initial incubation step at 95°C for 10 min followed by 40 cycles of 95°C for 15 sec, 60°C for 30 sec and 72°C for 15 sec. Baseline and threshold values were automatically detected using the LightCycler software. Primers are provided in Table S1.

*RNA analysis*

For RNA isolation, snap frozen testicular material of a individual M864 carrying the *M1AP* variant and a control proband with full spermatogenesis was used. RNA isolation was conducted with the miR-Neasy Micro kit (217084, Qiagen, Hamburg, Germany) according to the manufacturer's protocol. 500 ng of RNA were used as starting material for cDNA synthesis employing the iScript cDNA Synthesis Kit (Biorad). Subsequently, 4 µl of cDNA were amplified via PCR reaction using the Qiagen Taq Polymerase Kit. Reactions were performed in 20 µl of total volume with primer concentrations of 20 pmol (sequences in Table S1) and 1 mM dNTPs. The PCR included an initial incubation step at 94°C for 2 min followed by 35 cycles at 94°C for 30 sec, 56°C for 45 sec and 72°C for 1 min as well as a final step at 72°C for 10 min. PCR products were evaluated using a 2% agarose gel.

*Attempt at M1AP immunohistochemical localization*

Testicular tissue sections of 3 µm thickness were prepared (SM2010R sliding microtome, Leica Biosystems, Nussloch, Germany) from two human control samples with complete spermatogenesis or men with a homozygous duplication (c.676dup p.Trp226LeufsTer4) in *M1AP*, resulting in bilateral meiotic arrest. Sections were deparaffinized, rehydrated in a descending ethanol row, and rinsed with tap water. Heat-induced antigen retrieval was performed in citrate buffer (pH 6). After cooling to room temperature (RT), sections were washed with 1X Tris-buffered saline (TBS) prior to incubation with 3% $H_2O_2$ for 15 min at RT to inactivate endogenous peroxidases. Washing steps with distilled water and TBS followed. Nonspecific binding sites were blocked by applying 25% goat serum (#G6767-100ML, Sigma-Aldrich, Munich, Germany) in TBS containing 0.5% bovine serum albumin (BSA) for 30 min at RT in a humid chamber. Primary antibody incubation was performed overnight at 4°C in a humid chamber and different concentrations were evaluated for both M1AP antibodies (1:20, 1:50, 1:100, 1:200, 1:500, 1:1000 diluted in blocking solution, respectively). MAGE-A4, a marker for early human germ cells, was used as a positive control (kindly provided by Prof. G. C. Spagnoli, University Hospital of Basel, Switzerland). Respective IgG (#I5006, Sigma-Aldrich, Munich, Germany) and omission of primary antibody controls were included for each staining. On the following day, sections were washed again in TBS and incubated with corresponding secondary antibodies (goat anti-rabbit Biotin, #ab6012, Abcam, Cambridge,

USA - 1:100, 1:200) for 1 h at RT in a humid chamber. After further washing with TBS, Streptavidin conjugated with horse-radish peroxidase (#S5512, Sigma-Aldrich, Munich, Germany) was diluted in blocking solution (1:500) and sections were incubated for 45 min at RT in a wet chamber. Subsequently, sections were washed with TBS and incubated with 3,3'-Diaminobenzidine tetrahydrochloride (DAB, #D3757, AppliChem, Darmstadt, Germany) for visualization of antibody binding. Staining was validated by microscopical acquisition and distilled water was used to stop the reaction. Counterstains were conducted using Mayer's hematoxylin (#109249, Merck Millipore, Darmstadt, Germany). Sections were rinsed with tap water, dehydrated, and mounted using Merckoglas® mounting medium (#103973, Merck Millipore, Darmstadt, Germany). Images were captured using the PreciPoint M8 Scanning Microscopy System (PreciPoint, Freising, Germany).

*Tissue preparation and Western blotting analysis*

Cytosolic lysates were prepared by mincing biopsy samples from human testis or kidney in IP buffer (1% Triton-X 100, 20 mM Tris-HCl [pH 7.5], 25 mM NaCl, 50 mM NaF, 15 mM Na4P2O7, and 1.5 mM EDTA) containing protease inhibitors (Complete; Roche, Mannheim, Germany). Samples were centrifuged at 10,000×g for 45 min at 4°C to pellet cell debris and nuclei. Cytosolic supernatants were removed and stored at −80°C until further use. For total cell lysates, tissue was minced in Laemmli buffer ((20% Glycerin, 125 mM Tris-HCl pH 6.8, 10% SDS, 0.2% Bromphenol blue, 5% β-Mercaptoethanol), boiled for 5 min at 95°C and then stored at -20°C. Sodium dodecyl sulfate polyacrylamid gel electrophoresis (SDS-PAGE) and semidry Western blotting analysis was performed using standard techniques. Primary anti-M1AP antibodies were from ThermoFisher (PA5-31627) and Sigma (HPA045420), respectively. Signals were detected using secondary peroxidase-coupled antibodies and enhanced chemiluminescence (ECL).

*Generation of M1AP variant-specific plasmid DNA*

The LoF variant (c.676dup) was introduced in wildtype (WT) *M1AP* cDNA in the mammalian expression vector pcDNA3.1 (+) (Genscript, Leiden, NL) using site-directed-mutagenesis (SDM) according to manufacturer's instructions (QuikChange II XL Site-Directed Mutagenesis Kit, Agilent Technologies, Waldbronn, Germany). Clones carrying the respective mutations were verified by Sanger sequencing. For follow-up validation, a DYK-tag was integrated at the N-terminus of M1AP. Primer sequences can be found in Table S1.

*Heterologous expression of M1AP in HEK293T cells*

Human embryonic kidney 293T (HEK293T) cells were cultivated in 6 well plates (500.000 cells/well) in DMEM medium (Dulbecco's modified Eagle's medium, Sigma-Aldrich, Munich, Germany) supplemented with 1% PenStrep and 10% FCS at 37 °C, 5% $CO_2$. After 24 h post seeding, heterologous expression was induced by transfecting 2 µg plasmid DNA with XTR9 transfection reagent (Sigma-Aldrich, Munich, Germany) according to manufacturer's instructions. Transfection and subsequent analysis were performed in independent triplicates.

24 h post-transfection whole cell lysates were generated as following: cells were washed with 1X PBS and collected in a microcentrifuge tube. After centrifugation (4°C, 5 min, 200 g) and additional washing, the supernatant was discarded, and cells were re-suspended with lysis buffer (120 mM NaCl, 25 mM HEPES, 1% Triton X-100, 2 mM EDTA, 25 mM NaF, 1 mM $NaVO_3$, 0.2% SDS, 1x cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich, Munich, Germany), 1 mM PMSF). Cells were incubated on ice for 15 min and collected by centrifugation (4°C, 15 min, full speed). The protein-containing supernatant was transferred into a fresh tube and stored at -20°C until further processing.

For Western blot analysis, protein samples were separated on Mini-PROTEAN® TGX Stain-Free™ Precast gels (Bio-Rad, Feldkirchen, Germany) and transferred to a PVDF membrane using a Trans-Blot Turbo Mini Transfer Pack kit (Bio-Rad, Feldkirchen, Germany) according to manufacturer's instructions. Membranes were blocked with 5% milk powder TBST solution overnight at 4 °C prior to primary antibody (anti-OctA FLAG®, 1:1000, sc-166355, Santa Cruz and anti-α smooth muscle actin, 1:5000, ab5694, abcam) incubation for 2 h at RT. After washing, secondary antibody incubation (1 h, RT, anti-mouse IgG HRP, 1:1000, sc-2357, Santa Cruz)) followed. Chemiluminescence was detected with the Clarity™ Western ECL Substrate kit (Bio-Rad, Feldkirchen, Germany) and the ChemiDoc MP Imaging System (Bio-Rad, Feldkirchen, Germany).

**Supplemental Note: M1AP missense variant description and interpretation**

We identified another missense variant, c.797G>A, in individual M1943 which results in the substitution of arginine, a positively charged amino acid, with a neutrally charged glycine at position 266 (p.(Arg266Gly)).[10] The substituted amino acid is the only residue found at this position and can therefore be classified as highly conserved (up to zebrafish; Table 2). A mutation at this position is usually damaging for the protein, especially if being exchanged with such a different amino acid like glycine. As a smaller, more flexible and hydrophobic amino acid, glycine could for example disturb the required rigidity of the resulting protein or disturb correct protein folding.[10] The effect of this mutation is predicted to be damaging, deleterious or disease causing by all used *in silico* programs.

In addition to the heterozygous frameshift c.676dup, individual Y126 carries the substitution c.949G>A. The missense variant replaces the highly conserved (up to platypus; Figure 3C) neutral and nonpolar amino acid glycine with the larger, positively charged amino acid arginine (p.(Gly317Arg)). Based on conservation information, the variation in this position is highly likely to impair M1AP protein function.[10] The introduction of a charge may cause the repulsion of interaction partners or of other positively charged residues. In addition, the altered torsion angles may have an influence on the correct conformation and disturb the local structure of the protein. Correspondingly, the amino acid change p.(Gly317Arg) is predicted to be deleterious by all *in silico* algorithms.

Individual P86 carries two missense variants in *M1AP*. The variant c.1289T>C leads to a substitution of the hydrophobic amino acid leucine, which is predicted to be located in an alpha helix, with the less hydrophobic amino acid proline (p.(Leu430Pro)).[10] Because proline is an alpha helix breaker, the alteration likely has severe effects on protein structure. This change is predicted to be disease causing by all *in silico* programs. Furthermore, leucine at position 430 is a highly conserved amino acid (up to platypus; Figure 3D). The substitution c.148T>C replaces the polar amino acid serine at position 50 with the nonpolar amino acid proline (p.(Ser50Pro)). Again, the wildtype residue is predicted to be located in an alpha helix, and, therefore, its substitution by a proline likely has severe effects on protein structure and function.[10] Although the *in silico* algorithms SIFT and MutationTaster predict this change as being tolerated and as a polymorphism, the variant has not previously been described in any population, which supports a possible pathogenic impact. Moreover, this amino acid is likewise highly conserved (up to platypus; Figure 3D).

The substitution c.1166C>T identified in a consaguinous Turkish family leads to a replacement of the highly conserved and weakly hydrophobic proline with the more hydrophobic leucine at position 389 (p.(Pro389Leu), Figure 4C). Proline is known to have a rigid structure, giving a protein a specific conformation, which could be disrupted by substitution with leucine.[10]

Accordingly, the amino acid exchange is predicted to be pathogenic by all *in silico* programs and clinical variant interpretation classifies it as variant of uncertain significance.

**Supplemental references**


1. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. doi:10.1038/gim.2015.30
2. Tu C, Wang Y, Nie H, et al. An M1AP homozygous splice-site mutation associated with severe oligozoospermia in a consanguineous family. *Clin Genet*. 2020;(December 2019):741-746. doi:10.1111/cge.13712
3. Smith ED, Radtke K, Rossi M, et al. Classification of Genes: Standardized Clinical Validity Assessment of Gene–Disease Associations Aids Diagnostic Exome Analysis and Reclassifications. *Hum Mutat*. 2017;38(5):600-608. doi:10.1002/humu.23183
4. Arango NA, Huang TT, Fujino A, Pieretti-Vanmarcke R, Donahoe PK. Expression analysis and evolutionary conservation of the mouse germ cell–specificD6Mm5e gene. *Dev Dyn*. 2006;235(9):2613-2619. doi:10.1002/dvdy.20907
5. Arango NA, Li L, Dabir D, et al. Meiosis I Arrest Abnormalities Lead to Severe Oligozoospermia in Meiosis 1 Arresting Protein (M1ap)-Deficient Mice1. *Biol Reprod*. 2013;88(3):1-6. doi:10.1095/biolreprod.111.098673
6. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 2011;17(1):10. doi:10.14806/ej.17.1.200
7. Li H, Durbin R. Fast and accurate long-read alignment with Burrows – Wheeler transform. *Bioinformatics*. 2010;26(5):589-595. doi:10.1093/bioinformatics/btp698
8. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010. doi:10.1101/gr.107524.110
9. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122. doi:10.1186/s13059-016-0974-4
10. Venselaar H, te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics*. 2010;11(1):548. doi:10.1186/1471-2105-11-548