Supplementary Appendix to

**Variant *PNLDC1*, Defective piRNA Processing and Azoospermia**

Nagirnaja L, Mørup N, Nielsen JE, Stakaitis R, Golubickaite I, Oud MS, Winge SB, Carvalho F, Aston KI, Khani F, van der Heijden GW, Marques CJ, Skakkebaek NE, Rajpert-De Meyts E, Schlegel PN, Jørgensen N, Veltman JA, Lopes AM, Conrad DF, and Almstrup K

## Table of Contents

**Supplementary Information**

## Participating centres of the GEMINI consortium

**Donald F. Conrad** (leading PI**), Liina Nagirnaja**; Department of Genetics, Oregon National Primate Research Center, Oregon Health and Science University, Beaverton, OR, USA

**Kenneth I. Aston** (leading PI), **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

**Moira K. O'Bryan**; School of Biological Sciences, Monash University, Clayton, Victoria, Australia

**Rob McLachlan**; 1) Hudson Institute of Medical Research and the Department of Obstetrics and Gynaecology, Monash University, Clayton, Victoria, Australia; 2) Monash IVF and the Hudson Institute of Medical Research, 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, Susana Seixas**; i3S - Instituto de Investigação e Inovação em Saúde, University of Porto; IPATIMUP - Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Portugal

**Filipa Carvalho, Susana Fernandes, Alberto Barros**; i3S - Instituto de Investigação e Inovação em Saúde, University of Porto; Serviço de Genética, Departamento de Patologia, Faculdade de Medicina da Universidade do Porto, Porto, Portugal

**João Gonçalves**; Instituto Nacional de Saúde Dr Ricardo Jorge, Lisboa, Portugal

**Maris Laan**; Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia

**Margus Punab**; Andrology Center, Tartu University Hospital, 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) Department of Experimental and Clinical Biomedical Sciences, University of Florence, Florence, Italy; 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, University of Toronto, Toronto, ON, Canada

## Description of the cohorts

*GEMINI consortium*

The GEnetics of Male INfertility Initiative (GEMINI; https://gemini.conradlab.org/) is a multi-centre effort to identify rare DNA variations leading to male infertility. Whole exome sequencing (WES) has been performed on the whole blood of 1011 unrelated men diagnosed with idiopathic NOA (azoospermia or sperm concentration ≤1 million/mL) (manuscript in preparation). This sample collection represents both historical and current cases recruited across ten centres for andrology in seven countries (Australia, Canada, Denmark, Estonia, Portugal, Spain and USA). Exclusion criteria included obstruction or absence of vas deferens, varicocele of grade II or higher, history of cryptorchidism, radical pelvic surgery, anejaculation, spinal cord injury, Y chromosome microdeletions or karyotypic abnormalities, history of radiation treatments or chemotherapy. The WES call set of single-nucleotide variants (SNVs) and copy number variants (CNVs) was used to further identify and remove individuals with missed diagnosis of Klinefelter syndrome, CFTR mutations, large structural variants on sex chromosomes or autosomes, as well as cryptic relatedness and sample duplicates. The final GEMINI cohort consisted of 924 unrelated cases which were predicted to be predominantly of European ancestry (91%), followed by South Asia (4%), Africa (3%) and East Asia (2%). Out of 924 men, 72 were found to have >2.6% of autosomes in autozygous state indicative of recent parental consanguinity of $4^{th}$ degree or higher.

*Dutch consortium*

A total of 99 patients who presented with unexplained oligozoospermia (<10 million sperm/mL; n=44) and non-obstructive azoospermia (no sperm in the ejaculate; n=55) were enrolled at the Radboudumc outpatient clinic between July 2007 and October 2017. The average age of included men was 32 years (range 20 to 49 years). Most men were of Dutch descent (93 out of 99 men), one man was from the United Arab Emirates, one male was from Saint Martin (Caribbean Islands) and of four men their descent was unknown.

All males receive routine diagnostic assessment including detailed history, and physical examination, scrotal ultrasound, semen analysis, karyotyping, AZF screening and hormone level evaluation of FSH, testosterone and often inhibin B. Clinical evaluation and genetic testing did not lead to an etiologic diagnosis in these patients.

A total of 56 patients out of all 99 underwent TESE in the Radboudumc and in 31 cases TESE-ICSI was performed. Furthermore, 32 patients with moderate to extreme oligozoospermia underwent ICSI with ejaculated semen. In total, 55 out of 99 men (56%) achieved a pregnancy.

Residual genomic DNA extracted from a blood sample taken at the time of evaluation and treatment at the fertility centre was used for exome sequencing. DNA from all parents was obtained from saliva by using the Oragene OG-500 kit (DNA Genotek, Ottawa, Canada) at home.

*Dutch control cohort*

As a control cohort for the frequency of rare variants in proven fathers, we used an anonymized exome dataset derived from 5,784 Dutch men. These men received routine exome sequencing in the Radboudumc genome diagnostics centre as the healthy parent of a child with a severe illness. Whereas these men fathered a child with, e.g. developmental delay, their fertility is expected to be similar to an unselected sample of the male population.

## Supplementary Methods

### RNA isolation

Testicular RNA from Patient 1, Patient 4, and matched (by fixative) controls were extracted from fixed tissue. For Patient 3 and matched controls, remnants from mechanical and enzymatic TESE were used. The RNA quality was in all cases sufficient for subsequent analyses. In addition, frozen testicular biopsies with complete spermatogenesis, Sertoli cell-only pattern, and commercial RNAs from ovary, testis, skin, liver and ductus deferens were used (Supplementary Table S3). Testicular tissue for RNA extraction was unavailable for Patient 2.

### Gene expression

RNA was reverse transcribed (RT) and qPCR performed using gene-specific primers and conditions, as listed in Supplementary Table S1. For all primer combinations, bands of appropriate size were excised after gel electrophoresis, sequenced, and the identity validated. For ΔCt calculations, *RPS29* was used as a housekeeping gene[1]. When plotting, non-detected signals were imputed according to McCall *et al.*[2] and a Student's t-test was used to test for differences as the sample size was very small[3]. Single-cell RNA sequencing data was obtained from Wang *et al.*[4].

### Small RNA sequencing

Sequencing libraries were prepared with CATS Small RNA-seq kit (Diagenode, Cat. #C05010040) and sequenced on the MiSeq platform using MiSeq Reagent Kit v2 (Illumina, Cat. #MS-102-2001). Reads were trimmed and mapped to hg19 using bowtie v.1.0.1[5] and all small non-coding RNAs other than piRNAs (DASHR v.2.0[6]) were removed. The remaining reads were intersected with 205 piRNA genomic regions previously detected in the adult human testis[7].

## Supplementary Figures

## Supplementary Figure S1.

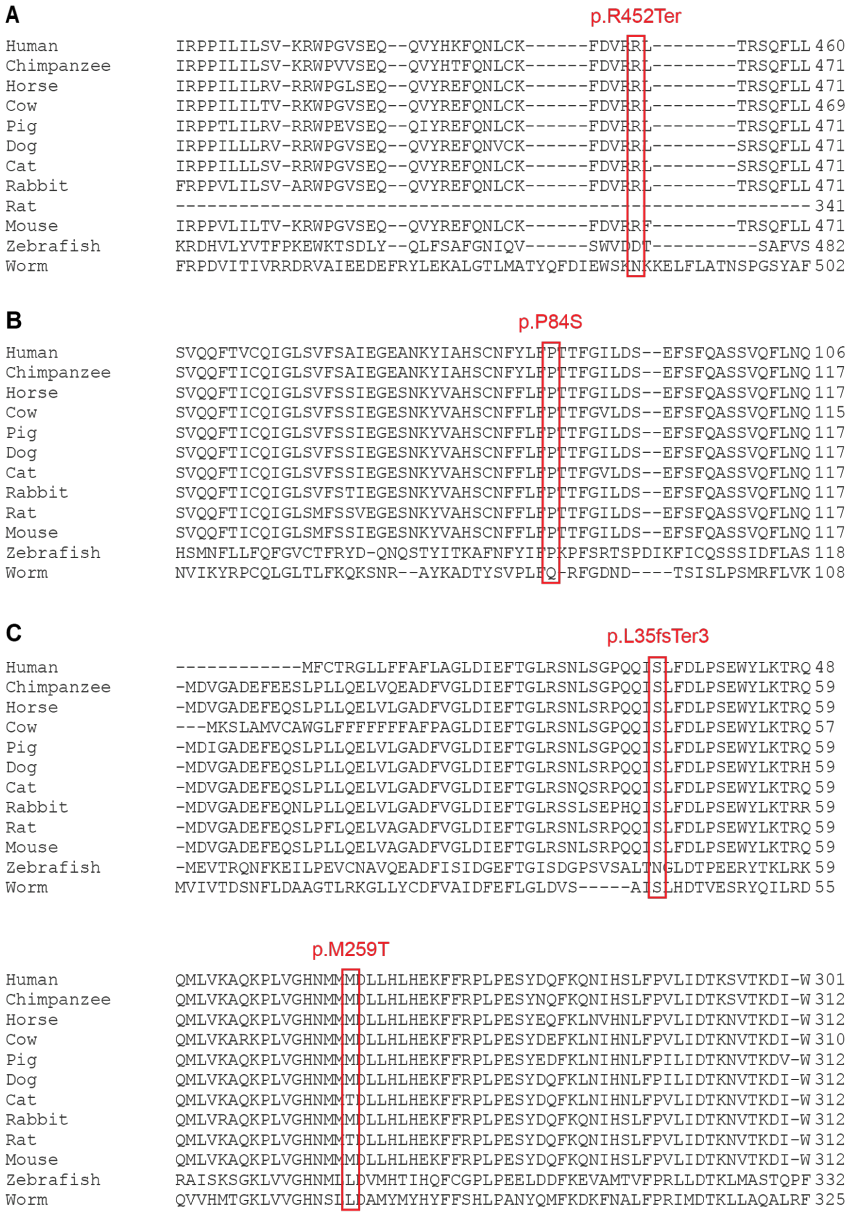


**Figure S1.** Protein sequence alignment of the PNLDC1 gene across species. Mutations identified in NOA Patient 1 from Denmark (**A**), Patient 2 from the US (**B**) and Patient 3 from Portugal (**C**) are highlighted.

**Supplementary Figure S2.**

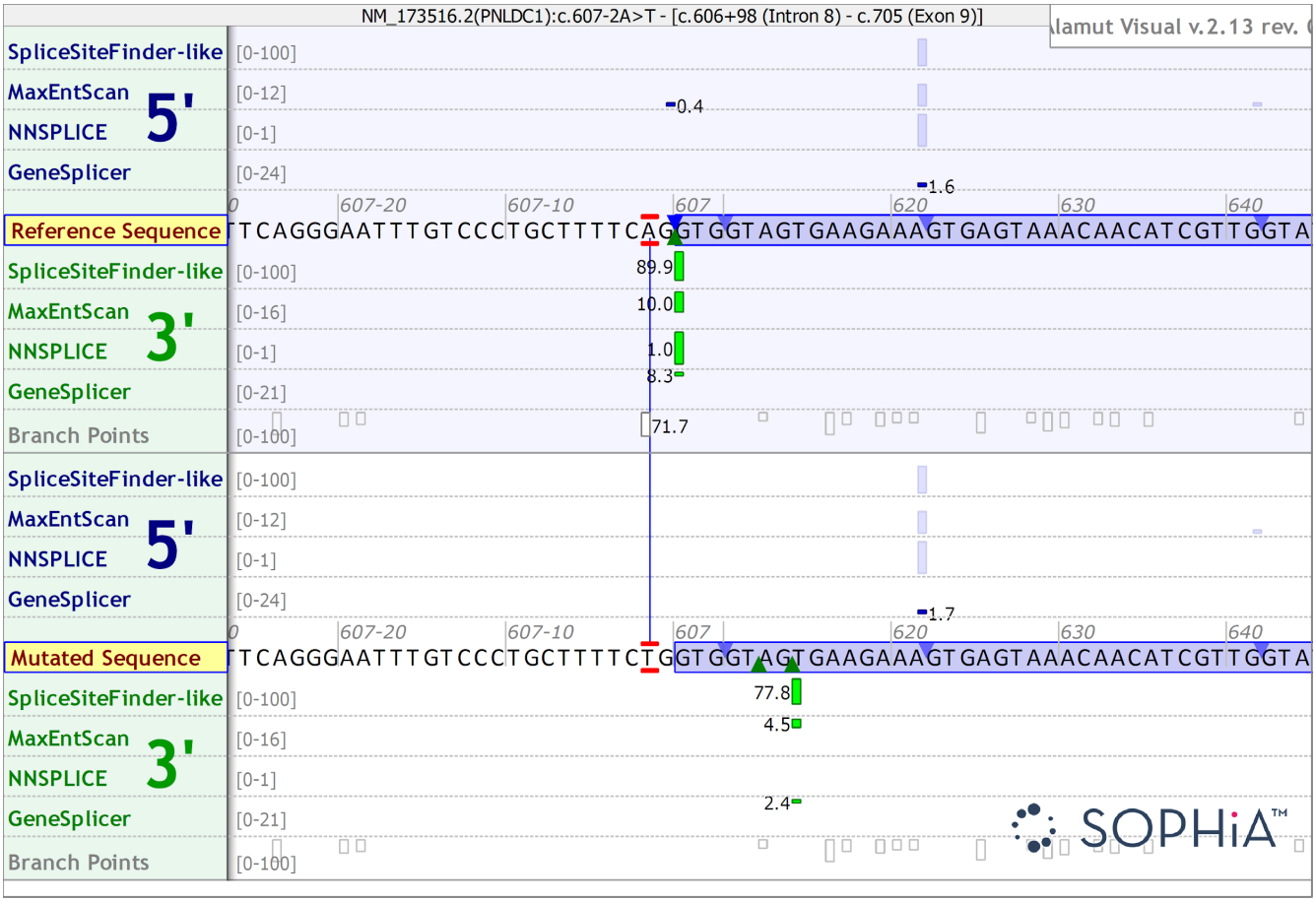


**Figure S2.** Splicing prediction models predict the abolishment of the canonical splice acceptor site by the c.607-2A>T variant in Patient 4.The mutation is located at the boundary of intron 8 and exon 9 of the PNLDC1 gene. Exon 9 is highlighted by a blue box, and the wild-type (reference) *PNLDC1* sequence is given in the top and the alternate sequence at the bottom track of the figure. Predicted splice sites from four different splice prediction models (SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer and Human Splicing Finder) are displayed as green and blue vertical bars for 3' (acceptor) and 5' (donor) sites, respectively. The computed score of each model is presented by the height of the green and blue vertical bars as a proportion of the maximum possible score, and the individual scores are displayed left of the bars. The ortholog alignment and splicing prediction were performed in the splicing module of the Alamut Visual version 2.13 software package (http://www.interactive-biosoftware.com).

**Supplementary Figure S3.**

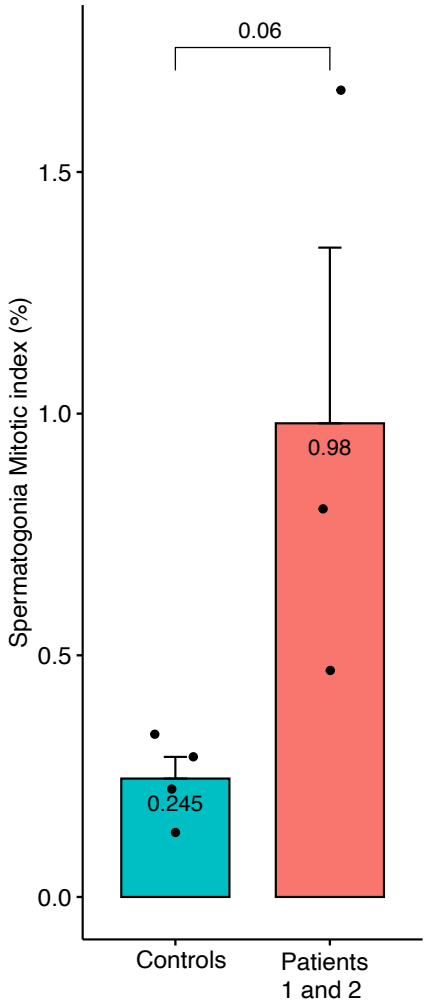


**Figure S3.** Mitotic index of spermatogonia in controls with complete spermatogenesis and in Patient 1 and Patient 2. The P-value was 0.057.

**Supplementary Figure S4.**

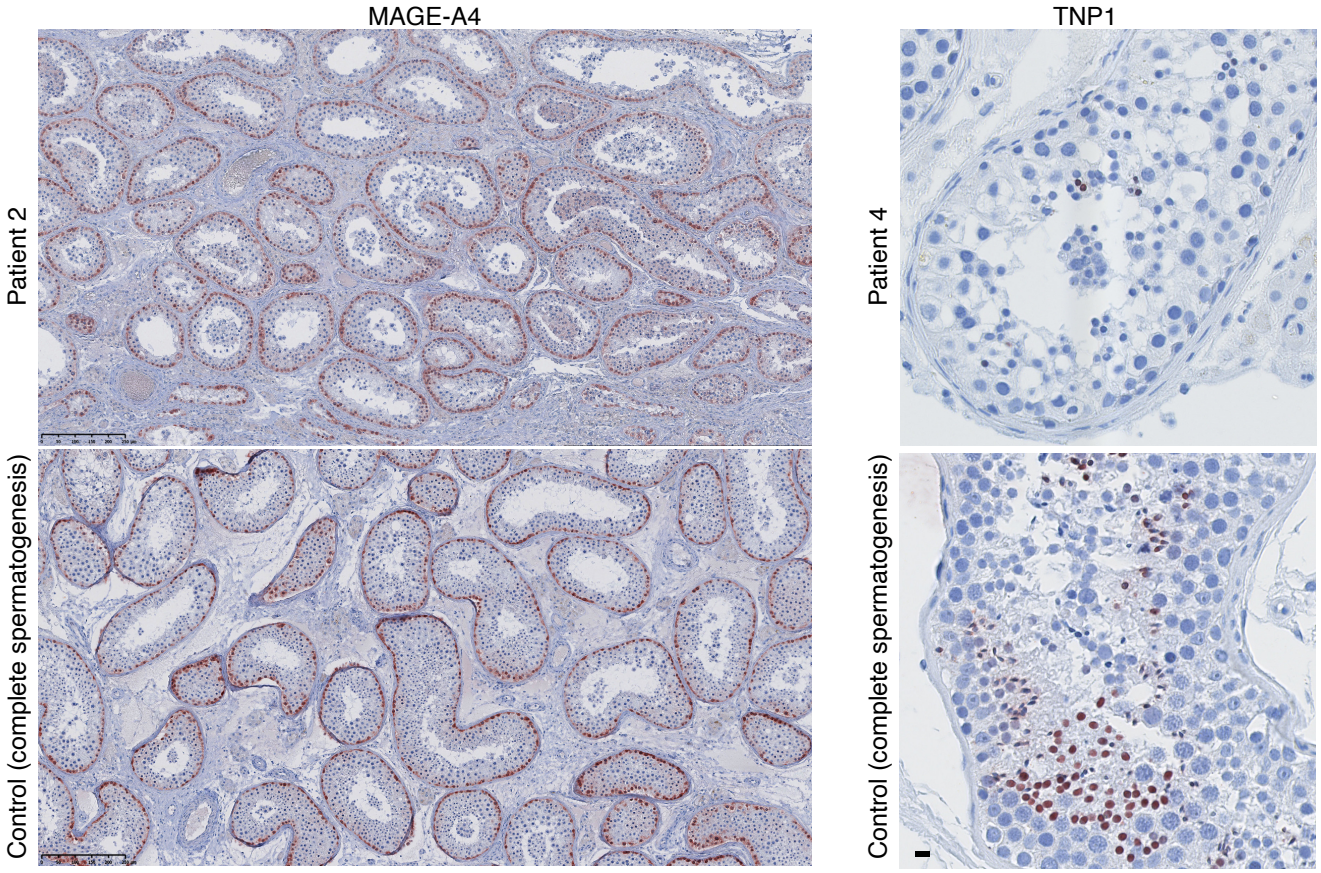


**Figure S4. Left:** MAGE-A4 staining (red/brown) indicating a similar number of spermatogonia in Patient 2 as in the control with complete spermatogenesis. The bar represents 250 microns. **Right**: Transition protein 1 (TNP1) staining showing few spermatids positive for TNP1 in Patient 4 compared to a control with complete spermatogenesis. The bar represents 10 μm.

**Supplementary Figure S5**

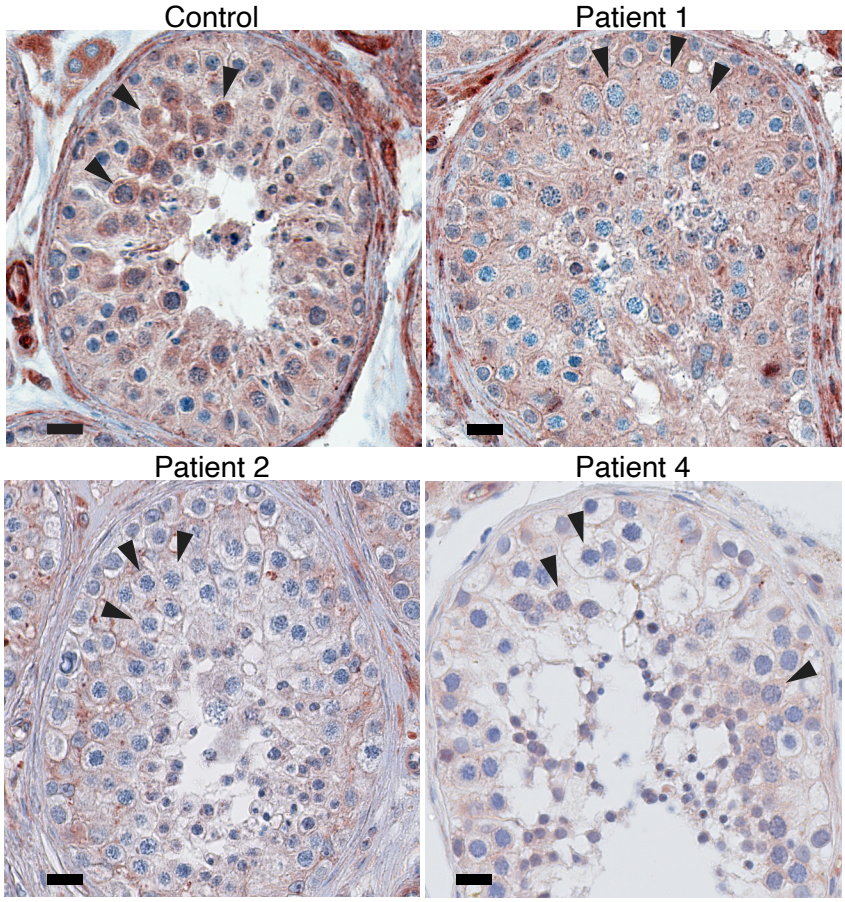


**Figure S5.** Immunohistochemical staining for PNLDC1 in a control biopsy with complete spermatogenesis and in biopsies from the three patients with homozygous pathogenic mutations in *PNLDC1*. Arrows indicate pachytene spermatocytes expressing PNLDC1 in the control but not in the three patients. The bars represent 25 µm.

**Supplementary Figure S6.**

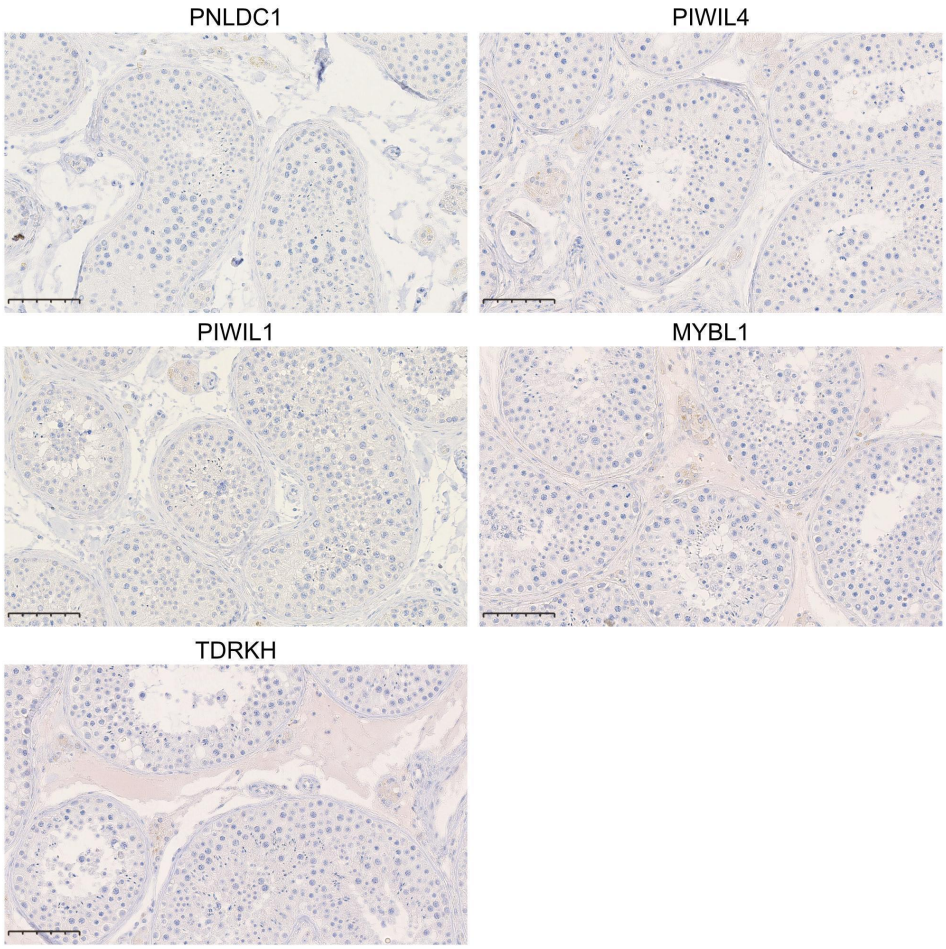


**Figure S6.** Negative controls for antibodies used in the study. Controls were obtained by omitting the primary antibody. The bar represents 100 μm.

**Supplementary Figure S7.**

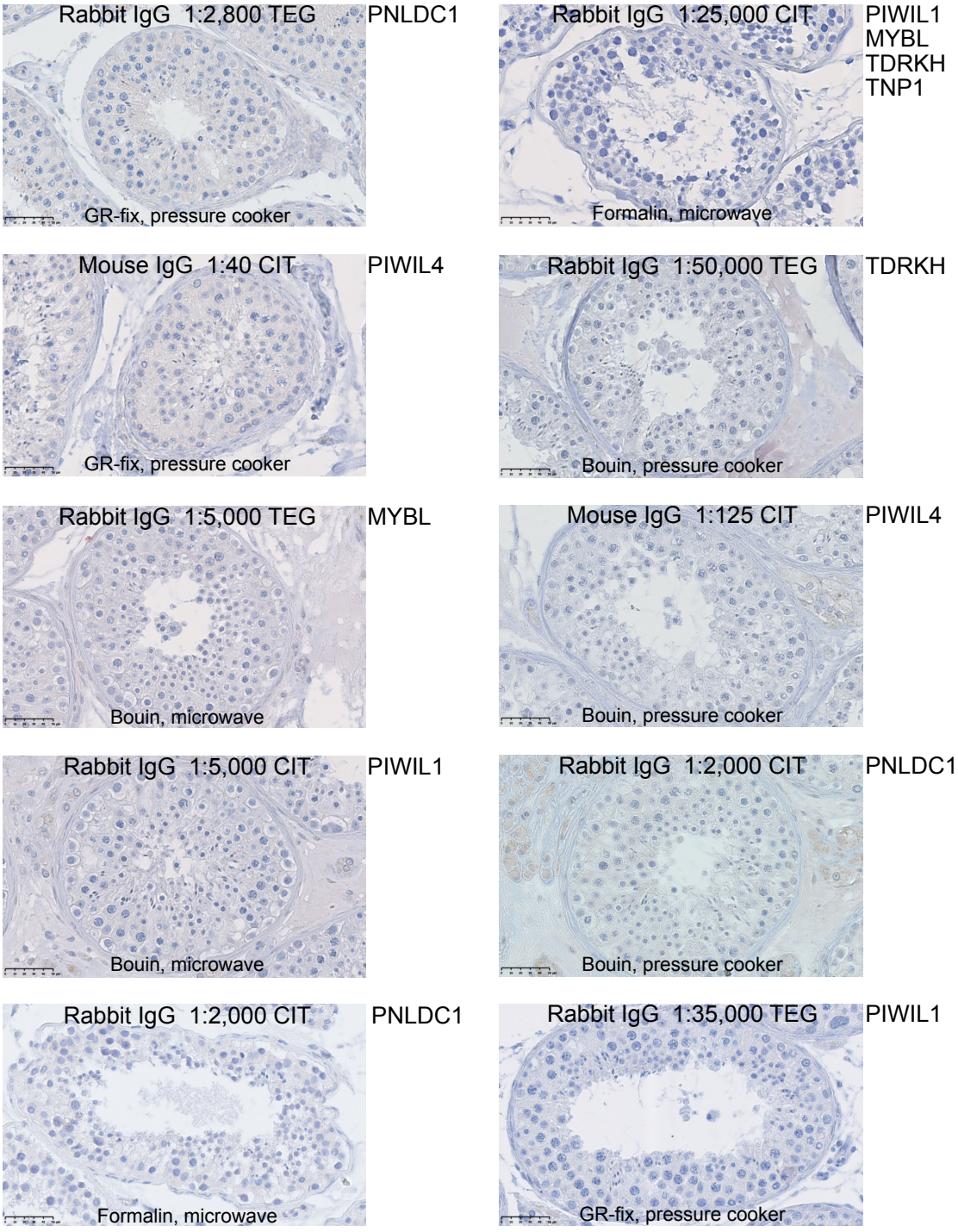


**Figure S7.** Negative controls for antibodies used in the study. Controls were obtained by substituting the primary antibody with an antibody of the same class but keeping all other conditions similar. The bar represents 50 µm.

**Supplementary Figure S8**

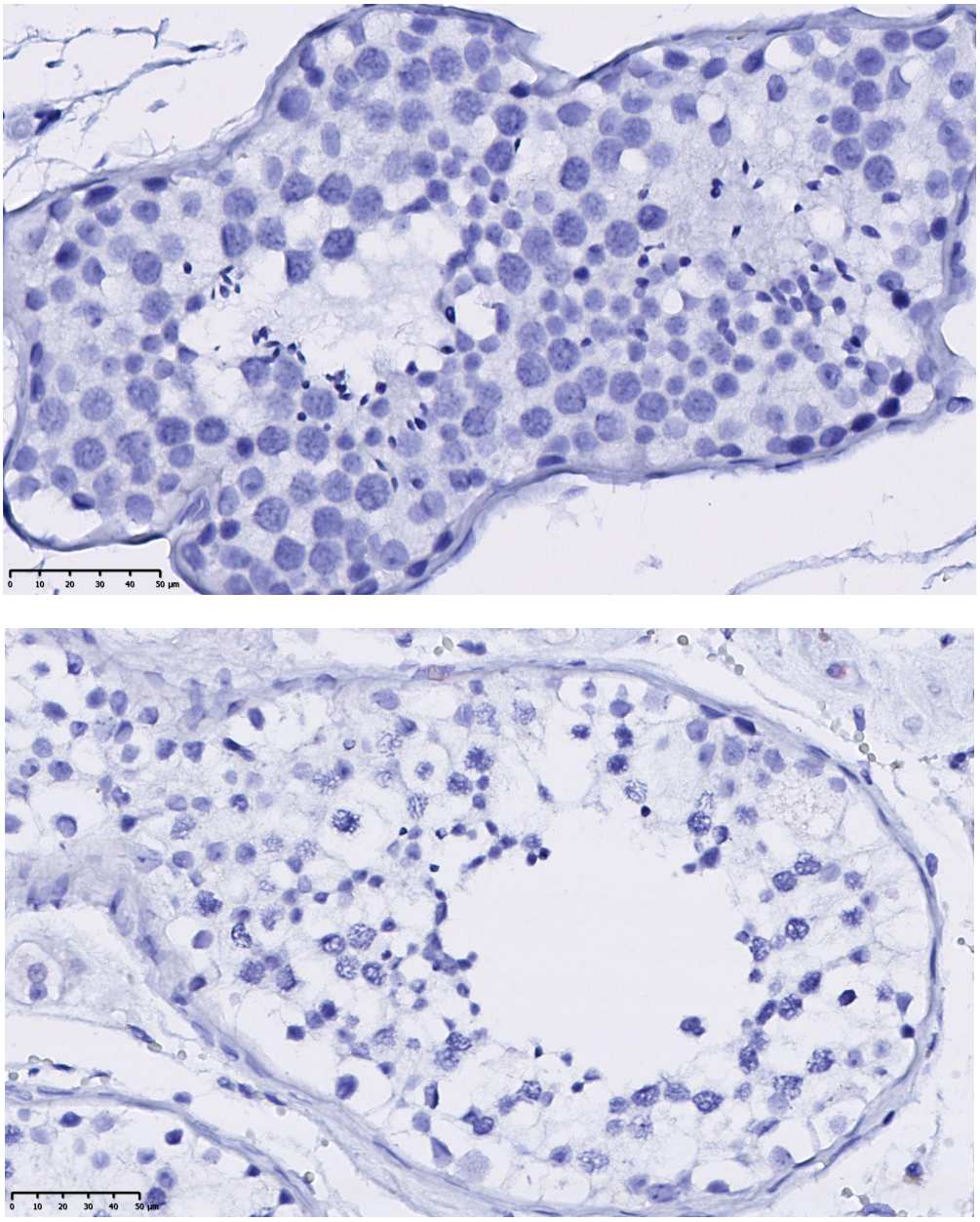


**Figure S8.** Negative control for *in situ* hybridization of *PNLDC1*. *In situ* hybridization with probes targeting *DapB* in a testicular biopsy from a control with complete spermatogenesis (top) and a testicular biopsy from Patient 4 (bottom). *DapB* is a bacterial RNA and hence serves as a negative control. There is no expression of *DapB* in neither case nor control. The bar represents 50 µm.

**Supplementary Figure S9.**

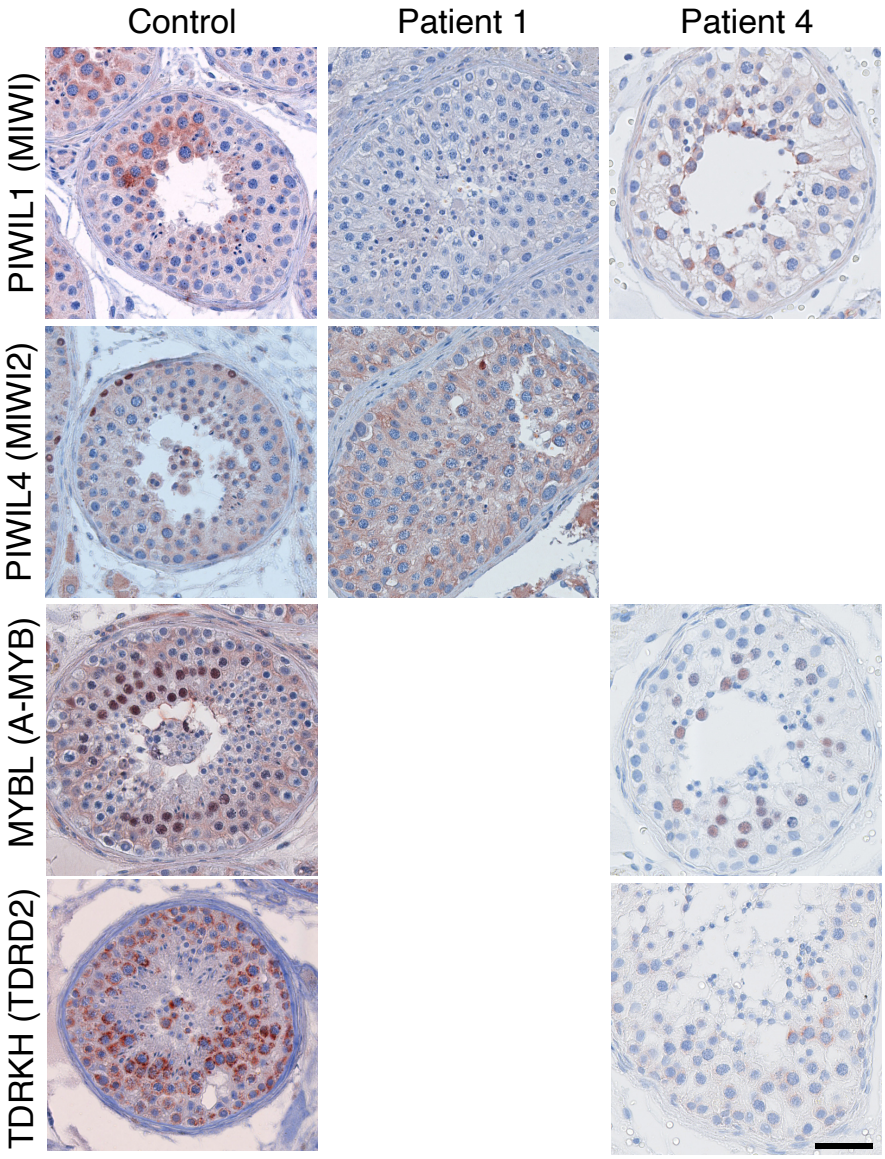


**Figure S9.** Staining of key piRNA processing components in sections from controls (same as in Figure 4A) and from Patient 1 and Patient 4. Due to the limited amount of tissue available and differences in used fixatives, it was not possible to get all antibodies to work on all biopsies. The bar represents 50 µm and all images are in same magnification.

## Supplementary Figure S10.

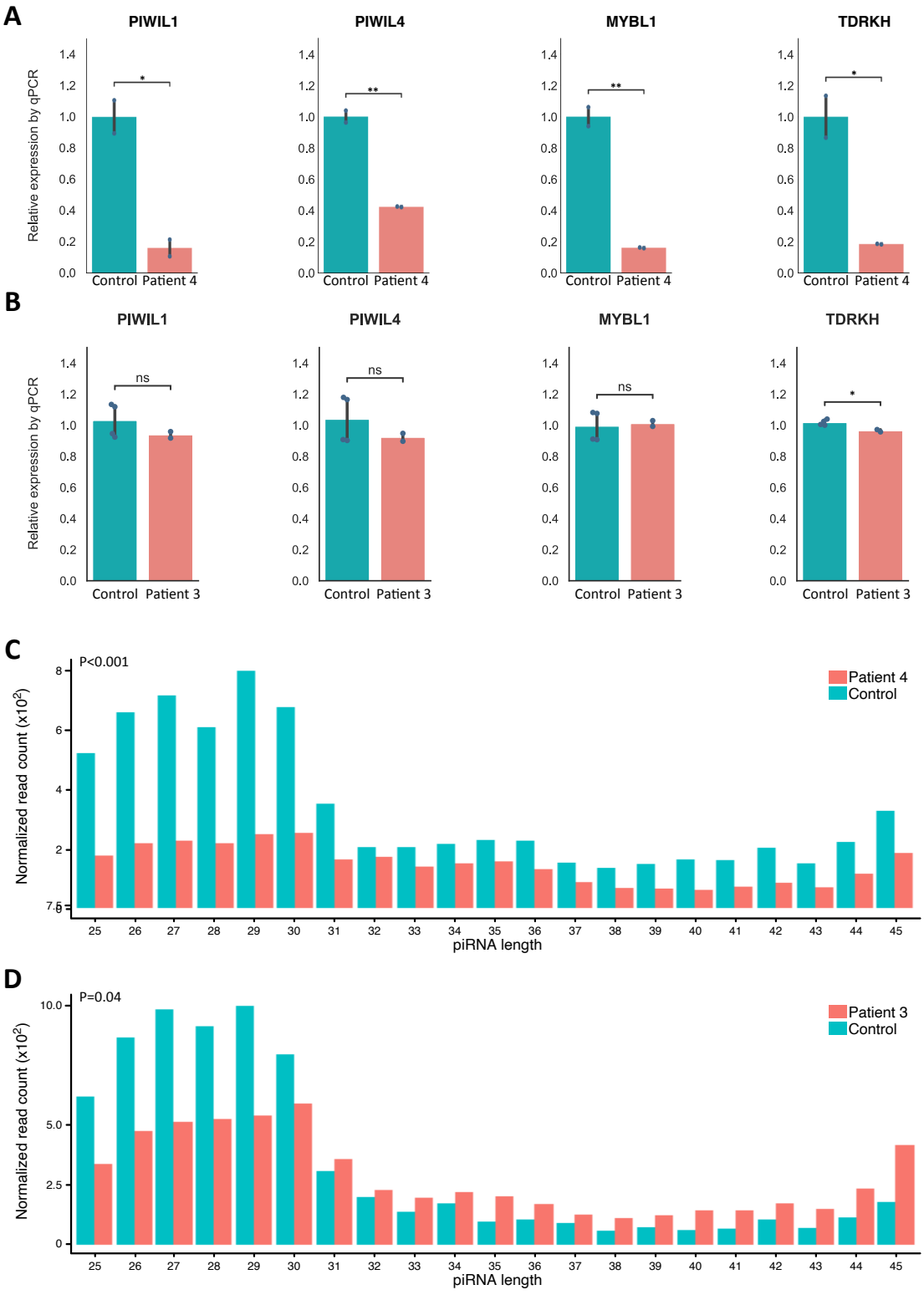

**Figure S10.** RT-qPCR and small RNA sequencing applied to the testicular sample of Patient 4 and the compound heterozygous Patient 3. **A**) Relative expression analysis of components of the piRNA biogenesis in Patient 4 and matched controls. *PNLDC1* (Figure 3C), *PIWIL1*, *PIWIL4*, *MYBL1*, and *TDRKH* all showed a significant down-regulation in Patient 4 compared to controls with complete spermatogenesis similar to Patient 1 (Figures 3 and 4). **B**) Relative expression analysis of components of piRNA biogenesis indicated a mild impact of the heterozygous compound mutations p.M259T and p.L35fsTer3 identified in Patient 3. **C**) Small RNA sequencing on the testicular tissue of Patient 4 demonstrated a significant loss of piRNAs compared to the control ($P<0.001$), specifically at the expected length of around 30 bases, as also observed for the Patient 1 (Figure 4). **D**) Small RNA sequencing confirmed the partially functional piRNA biogenesis in Patient 3, with a milder reduction in the amounts of piRNAs compared to the control ($P=0.04$) and a shift in the distribution of piRNA lengths towards longer transcripts. All read counts of piRNAs are normalized to spike-in controls added to the RNA before library preparation. ns: non-significant, *: $0.01< P <= 0.05$, **: $0.001 < P <= 0.01$, ***: $P <= 0.001$.

**Supplementary Figure S11.**

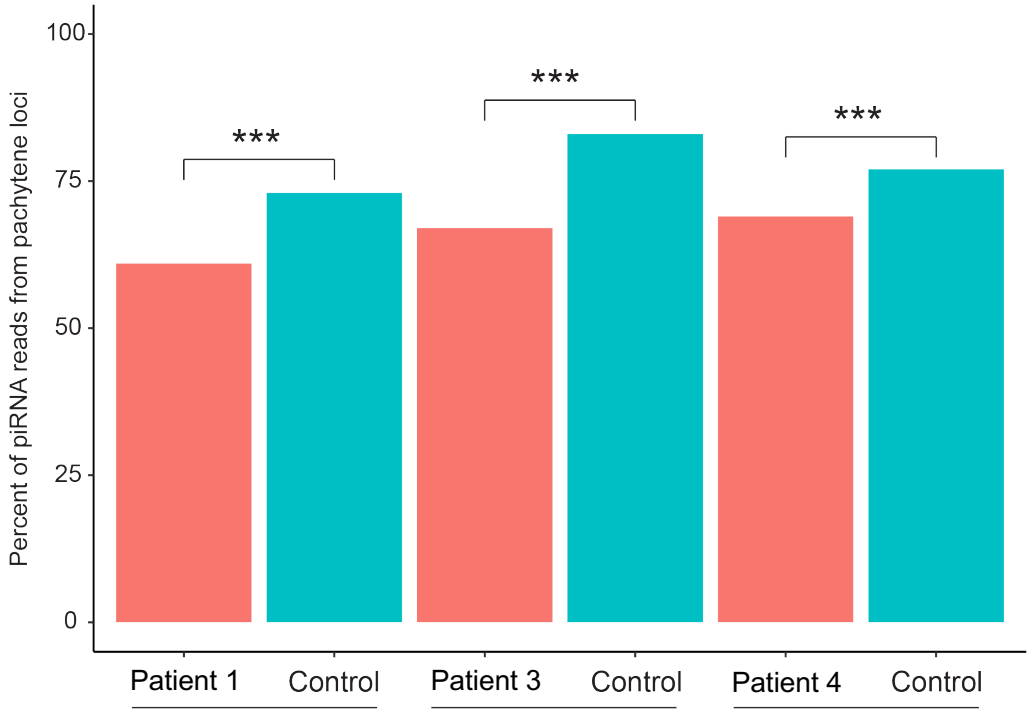


**Figure S11.** Percentage of piRNA reads arising from pachytene piRNA loci. Significantly reduced percentage of reads were produced from pachytene piRNA loci in all sequenced cases with PNLDC1 variants compared to respective controls. ***Fisher's exact test, P<0.001

## Supplementary Tables

**Table S1.** Primers and conditions for Sanger sequencing and qPCR.

| Method | Target | Primer sequences (5' -> 3') | Primer conc. | Annealing temperature | Product size |
|---|---|---|---|---|---|
| Sanger sequencing[#] | PNLDC1 p.P84S | Forw: AGCCTGCACGCCCTAACCTG<br>Rev: CGGGAGGCGAAGATTGCACC | 0.4 µM | Touch down 65°C to 55°C | 491 bp |
| Sanger sequencing[#] | PNLDC1 p.R452Ter | Forw: AGCAGAGCCTTTCCTAGTTTCT<br>Rev: GAGCAGGCAGTTGACGTTTG | 0.4 µM | Touch down 65°C to 55°C | 500 bp |
| Sanger sequencing[§] | PNLDC1 p.S34fs | Forw: GGTTGCGAGTGGTGGTCC<br>Rev: GCCCAACACAAAACAAGCAT | 0.4 µM | 60°C | 294 bp |
| Sanger sequencing[§] | PNLDC1 p.M259T | Forw: GTGATCAGAATTGCAGTGGGT<br>Rev: TGTACCAATCCCTCTGCCAG | 0.4 µM | 60°C | 250 bp |
| Sanger sequencing[§] | PNLDC1 c.607-2A>T | Forw: CCAAATCTGGCCTCTGAAAA<br>Rev: CAGAAAAACCCCTTGCTGAG | 0.33 µM | 60°C | 252 bp |
| RT-qPCR* | RPS29 | Forw: CGCTCTTGTCGTGTCTGTTCA<br>Rev: CCTTCGCGTACTGACGGAAA | 0.67 aM<br>0.67 aM | 62°C | 91 bp |
| RT-qPCR* | MYBL1 | Forw: TGTGTCCCAGCCACTTGCTTTC<br>Rev: TCTGACTTTCTTCCCAGAAGCGGT | 0.2 aM<br>0.2 aM | 62°C | 148 bp |
| RT-qPCR* | PIWIL1 | Forw: GCGGAGGTCAATCGCAG<br>Rev: TTGCAGCTATTCCAAGCCC | 0.2 aM<br>0.2 aM | 64°C | 150 bp |
| RT-qPCR* | PIWIL4 | Forw: TCCAGCTCAAATACCAGCTCA<br>Rev: GTACAGTGCGGTTCATTTCGG | 0.2 aM<br>0.2 aM | 64°C | 91 bp |
| RT-qPCR* | PNLCD1 | Forw: ATCTCCCTGTACCGCTACTG<br>Rev: CAAGGATGAACGCGAGAAGG | 0.2 aM<br>0.2 aM | 62°C | 100 bp |
| RT-qPCR* | TDRKH | Forw: GATCCAGATCGTTGGCTCCC<br>Rev: ACAGTCAAGTCTTCAGGCACA | 0.2 aM<br>0.2 aM | 62°C | 96 bp |
| RT-qPCR* | RPS20 | Forw: AACAAGCCGCAACGTAAAATC<br>Rev: ACGATCCCACGTCTTAGAACC | 0.2 aM<br>0.2 aM | 62°C | 166 bp |

[#] Sequenced by Integrated DNA Technologies (Coralville, IA, USA)

[§] Sequenced in-house with the BigDye Terminator v.3.1 Cycle Sequencing Kit (PATIENT 3; Applied Biosystems, Carlsbad, CA, USA).

* RNA was reverse transcribed (RT) using AffinityScript qPCR cDNA Synthesis Kit (Agilent, Cat. #600559), qPCR performed with the Brilliant II SYBR Green qPCR master mix (Agilent, Cat. #600828) on QuantStudio 3 (Thermo Fisher Scientific, Cat. #A28137) using gene-specific primers and conditions as listed. A 12 cycle pre-amplification step was added.

**Table S2.** Antibodies and conditions used in the study.

| Epitope | Vendor, code, species and class | Bouin's: working dilutions and buffers | GR-fix: working dilutions and buffers | Formalin: working dilutions and buffers | Working protocol | Comments |
|---|---|---|---|---|---|---|
| **PNLDC1** | Acris Antibodies, ADP53365PU-N, Rabbit, IgG | 1:200: CIT | 1:275: TEG | 1:200: CIT | Formalin: Microwave AEC<br>GR / Bouin's: Pressure cooker AEC | Interstitial background |
| **PNLDC1** | Invitrogen, PA5-62535, Rabbit, IgG | 1:100, 1:150: CIT/TEG* | 1:100: CIT/TEG*<br>1:250: TEG* | | Pressure cooker DAB | Suboptimal in our protocols |
| **PNLDC1** | Invitrogen, PA5-62535, Rabbit, IgG | 1:100: CIT/TEG*<br>1:150, 1:300, 1:500: CIT* | | | Microwave DAB | Suboptimal in our protocols |
| **PNLDC1** | Abcam, ab151151, Rabbit, IgG | 1:100, 1:200: CIT/TEG*<br>1:500: CIT* | 1:150: CIT/TEG*<br>1:250: TEG* | | Pressure cooker DAB | Suboptimal in our protocols |
| **PNLDC1** | Abcam, ab151151, Rabbit, IgG | 1:100: CIT/TEG*<br>1:250: CIT* | | | Microwave DAB | Suboptimal in our protocols |
| **PIWIL1 (MIWI)** | Cell Signaling Technology, #2079, Rabbit, IgG | 1:150: CIT | 1:650: TEG | 1:100: CIT | Bouin's / Formalin: AEC Microwave,<br>GR-fix: Pressure cooker AEC | |
| **PIWIL4 (MIWI2)** | Thermo Fisher Scientific, MA5-17151, Mouse, IgG | 1:1500 CIT | 1:450 CIT | | Pressure cooker AEC | |
| **MYBL (A-MYB)** | Sigma Aldrich, HPA008791, Rabbit, IgG | 1:150 TEG | | 1:500: CIT | Microwave AEC | |
| **TDRKH (TDRD2)** | Proteintech,13528-1-AP, Rabbit, IgG | 1:3000 TEG | | 1:5000. CIT | Bouin`s: Pressure cooker AEC<br>Formalin: Microwave | |
| **MAGE-A4** | Gift from G. Spagnoli, Clone 57B, Mouse, IgG | 1:2000 TEG | | | Pressure cooker AEC | |
| **TNP1** | Novusbio NBP2-30567, Rabbit, IgG | | | 1:750: CIT | Microwave AEC | |
| **Rabbit IgG-isotype** | Abcam, Ab37415 | various | various | various | All above | Negative control |
| **Mouse IgG-isotype** | Abcam, Ab91353 | various | various | various | All above | Negative control |

Notes: *Intervals tried. GR-fix, modified Stieve's fixative; CIT, 0,01 M Citrate (pH 7,4); TEG, TEG buffer (10 mM Tris, 1mM EDTA, 0,05% Tween 20, pH 8,5); AEC, 3-amino-9-ethylcarbazole; DAB, 3,3′ diaminobenzidine tetrahydrochloride.

**Table S3.** Sources of RNA used in the study.

| Source | Tissue | Extraction method | Remarks |
|---|---|---|---|
| Patient 1*, # | Testis | RecoverAll Total Nucleic Acid Isolation Kit for FFPE (Thermo Fisher Scientific, Cat. #AM1975) | Fixed (GR-fix, a modified Stieve's fixative) testicular biopsy |
| Patient 3§, # | Testis | Trizol reagent (Thermo Fisher Scientific, Cat. #15596018) | Remnants from testicular sperm extraction (TESE) |
| Patient 4** | Testis | RecoverAll Total Nucleic Acid Isolation Kit for FFPE (Thermo Fisher Scientific, Cat. #AM1975) | Fixed (Formalin) testicular biopsy |
| Sertoli-cell-only pattern (SCO) * | Testis | NucleoSpin® RNA kit (Macherey-Nagel, Cat. #740955) | Frozen testicular biopsy (n=2) |
| Controls with complete spermatogenesis (for Patient 1) *, # | Testis | RecoverAll Total Nucleic Acid Isolation Kit for FFPE (Thermo Fisher Scientific, Cat. #AM1975) | Fixed (GR-fix, a modified Stieve's fixative) testicular biopsy control cases (n=4) with complete spermatogenesis |
| Controls with complete spermatogenesis (for Patient 3) §, # | Testis | Trizol reagent (Thermo Fisher Scientific, Cat. #15596018) | Remnants of TESE from patients with conserved spermatogenesis but undergoing infertility treatment due to anejaculation (n=2). |
| Control with complete spermatogenesis (for Patient 4) | Testis | RecoverAll Total Nucleic Acid Isolation Kit for FFPE (Thermo Fisher Scientific, Cat. #AM1975) | Archived FFPE biopsy of a patient undergoing a TESE after a vasectomy was used. Histological inspection of the biopsy showed complete spermatogenesis. |
| Controls with complete spermatogenesis | Testis | NucleoSpin® RNA kit (Macherey-Nagel, Cat. #740955)<br>Clontech - Acid guanidinium thiocyanate-phenol-chloroform extraction<br>Ambion - Modified version of the ToTALLY RNA Kit (Ambion, Cat. #AM1910) | Frozen testicular biopsy (n=1)*<br><br>Human Testis Total RNA (Clontech, Cat. #636533)<br>Human Testes Total RNA (Ambion, Cat. #AM7972) |
| Ovary | Ovary | A modified version of the ToTALLY RNA Kit (Ambion, Cat. #AM1910) | Human Ovary Total RNA (Ambion, Cat. #AM6974) |
| Somatic | Skin | Modified guanidine thiocyanate technique | Human Skin Total RNA (BioChain, Cat. #R1234218-50) |
| Somatic | Ductus deferens | Modified guanidine thiocyanate technique | Human Ductus Deferens Total RNA (BioChain, Cat. #R1234100-10) |
| Somatic | Liver | Acid guanidinium thiocyanate-phenol-chloroform extraction | Human Liver Total RNA (Clontech, Cat. #636531) |

* Obtained from the archives at Department of Growth of Reproduction, Rigshospitalet.
** Obtained from the pathology archives at Radboudumc.
§ Obtained from Departamento de Patologia, Faculdade de Medicina da Universidade do Porto
# Degradation level and quantity were evaluated with the 2100 Bioanalyzer (Agilent, Cat. #G2939BA) and Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Cat. #Q33216) respectively.

## References


1. Svingen T, Jørgensen A, Rajpert-De Meyts E. Validation of endogenous normalizing genes for expression analyses in adult human testis and germ cell neoplasms. Mol Hum Reprod 2014;20(8):709–18.
2. McCall MN, McMurray HR, Land H, Almudevar A. On non-detects in qPCR data. Bioinformatics 2014;30(16):2310–6.
3. De Winter JCF. Using the Student's t-test with extremely small sample sizes. Practical Assessment, Research, and Evaluation 2013;18(1):10.
4. Wang M, Liu X, Chang G, et al. Single-Cell RNA Sequencing Analysis Reveals Sequential Cell Fate Transition during Human Spermatogenesis. Cell Stem Cell 2018;23(4):599-614.e4.
5. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10(3):R25.
6. Leung YY, Kuksa PP, Amlie-Wolf A, et al. DASHR: database of small human noncoding RNAs. Nucleic Acids Res 2016;44(D1):D216-22.
7. Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 2006;442(7099):199–202.