
# Analysis of the role of selected exoribonucleases and poly(A) polymerases in *Mus musculus*

by

Michał Brouze

Doctoral dissertation
Laboratory of RNA Biology
International Institute of Molecular and Cell Biology in Warsaw

Supervisor: Prof. dr hab. Andrzej Dziembowski

Warsaw, October, 2024

**Tytuł pracy w języku polskim**:

Analiza roli wybranych egzorybonukleaz i polimeraz poli(A) u myszy domowej

**The research was initiated in:**

Laboratory of RNA Biology and Functional Genomics, Institute of Biochemistry and Biophysics, Polish Academy of Sciences

**And after moving the laboratory to a new location finalized in:**

Laboratory of RNA Biology, International Institute of Molecular and Cell Biology in Warsaw

**Financial support was provided by:**

Grant ERC StG European Research Council (7FP) "Regulation of Gene Expression by non-canonical poly(A) and poly(U) polymerases" 309419-PAP&PUPs

Subsydium FNP Mistrz „Posttranscriptional regulation of gene expression in metazoans" Mistrz2./2014

Grant FNP TEAM „Mouse Genome Engineering Facility – generation of animal models for biomedical research and preclinical studies" POIR.04.04.00-00-436/17-00

MOlecular Signaling in Health and Disease - Interdisciplinary Centre of Excellence (MOSaIC project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 810425)

## Podziękowania

Chciałbym serdecznie podziękować wszystkim tym, bez których ta praca by nie powstała, a ja nie byłbym tu, gdzie jestem:

Prof. dr hab. Andrzejowi Dziembowskiemu, mojemu promotorowi i mentorowi, który przez lata wspólnej pracy dzielił się ze mną swoim ogromnym doświadczeniem, prowadził mnie przez liczne wyzwania pracy naukowej i ukształtował mnie jako naukowca jakim jestem dzisiaj;

Dr Oldze Gewartowskiej za niezliczone dyskusje, opinie, wspólne eksperymenty, przeczytane drafty i dobre rady, bez których nie poradziłbym sobie w tej pracy i nie obrał nowego kierunku na przyszłość;

Dr hab. Sewerynowi Mroczkowi za dzielenie się niespotykanym doświadczeniem i intuicją w pracy laboratoryjnej;

Mgr Agnieszce Czarnockiej-Cieciurze za wspólną ciężką pracę i wsparcie przy mojej najważniejszej publikacji, bez której ta praca nie byłaby możliwa, ale również za długie rozmowy o wyzwaniach, które oboje musieliśmy pokonywać w czasie doktoratu;

Dr hab. Ewie Borsuk oraz Prof. dr hab. Annie Marii Ciemerych-Litwinienko za przekazane cenne wskazówki i techniki pracy z oocytami i zarodkami;

Dr Pawłowi Krawczykowi i dr Natalii Gumińskiej, bez których bioinformatycznej pracy żadne z odkryć, którymi mogę się pochwalić, nie byłoby możliwe;

Mgr Pauli Kwapisz i dr Zofii Korbut-Mikołajczyk, których codzienne wsparcie i organizacyjne wysiłki pozwoliły mi na oddanie się pracy naukowej w stopniu, który bez nich nigdy nie byłby możliwy;

Całemu zespołowi Pracowni Inżynierii Genomu, za wielki wysiłek w produkcji plazmidów, tworzeniu modeli i prowadzeniu hodowli myszy na potrzeby mojej pracy;

Wszystkim członkom Laboratorium Biologii RNA, których poznałem i z którymi pracowałem przez lata, a których wkładu w moją pracę, których pomocy i udzielonego wsparcia, których wpływu na mój rozwój, na to jakim naukowcem jestem dzisiaj, nie pomieściłyby wszystkie karty tej rozprawy;

Wreszcie chciałbym podziękować mojej mamie i siostrze, które przez lata wspierały mnie i nigdy we mnie nie zwątpiły;

I mojej żonie, Oli, partnerce w pracy i w życiu, za wszystko.

## Abstract

Animal development is characterised by rapid and dynamic differentiation of new cell and tissue types. This requires precise gene expression regulation to control cell fate, often occurring at the transcriptomic level. For this reason, since a long time, multiple stages of animal development are considered useful models to study RNA polyadenylation and selective transcript degradation, as those two mechanisms can be deciding factors in cell differentiation pathway.

In most somatic cells poly(A) tail decides about fate of a transcript– a longer tail allows efficient translation to occur, while its shortening marks transcript for degradation. However, in oogenesis, maintaining short oligo(A) tail stabilizes transcripts stored in the cytoplasm until maturation initiation, when selective polyadenylation allows for their translation, a key process in oocyte's development. Yet mice deficient in GLD2, the only known cytoplasmic poly(A) polymerase in the oocyte, remain fertile, necessitating activity of other, unknown poly(A) polymerases in oogenesis.

On the other hand, RNA quality control and decay mechanisms are known to regulate cell pluripotency and differentiation through selective degradation of mRNAs. Both deadenylation and decapping (with additional uridylation between those steps) can trigger exoribonucleolytic degradation by exosome complex and Xrn1, respectively. Notably, only those exosome-bound exoribonucleases, EXOSC10 and DIS3, were shown to be essential for mouse embryogenesis.

Both of those aspects of transcriptome regulation in development were the focus of my doctoral project. I studied the roles of novel TENT5 family of poly(A) polymerases in oogenesis and the last catalytic subunit of the cytoplasmic exosome complex with an undefined role – DIS3L exoribonuclease - in embryogenesis in mice.

TENT5B and TENT5C noncanonical poly(A) polymerases play essential, but redundant, role in proper folliculogenesis and oocyte development. Knock-out of both genes leads to oocyte degeneration while GFP knock-in of *Tent5b* leads to disruption of chromatin organisation in the oocyte, also leading to infertility. Direct RNA sequencing revealed that TENT5s polyadenylate number of essential, oocyte-specific transcripts encoding secreted proteins, notably ZP3 and GDF9. RNA reporter assay showed that endoplasmic reticulum-leading sequence can be determining factor for TENT5-mediated polyadenylation. In addition, investigation of poly(A) tail composition in wild-type animals also revealed specific

overrepresentation of uridine residues close to 3'-terminus in a subgroup of spermatogenesis-essential transcripts

Meanwhile, DIS3L is essential for embryo development. DIS3L-deficient embryos produce viable embryonic stem cells but fail to develop beyond embryonic day 6,5. RNA sequencing of *Dis3l* knock-out blastocysts showed upregulation of several transcripts. Curiously, this was followed by lowered levels of proteins encoded by those transcripts caused by the overall lowered protein production in *Dis3l* knock-out embryos.

## Streszczenie

Rozwój zwierząt charakteryzuje się szybkim i dynamicznym różnicowaniem nowych typów komórek i tkanek. Wymaga to precyzyjnej regulacji ekspresji genów w celu kontrolowania losu komórek, często na poziomie transkryptomu. Z tego powodu od dłuższego czasu różne etapy rozwoju zwierząt są uważane za użyteczne modele do badania poliadenylacji RNA i selektywnej degradacji transkryptów, bowiem te dwa mechanizmy mogą być decydującymi czynnikami w ścieżce różnicowania komórki.

W większości komórek somatycznych ogon poli(A) decyduje o losie transkryptu – dłuższy ogon umożliwia wydajną translację, podczas gdy jego skracanie stanowi sygnał do degradacji transkryptu. Jednak w oogenezie utrzymywanie krótkiego ogona oligo(A) stabilizuje transkrypty nagromadzone w cytoplazmie do momentu inicjacji dojrzewania, kiedy to selektywna poliadenylacja umożliwia ich translację, co jest kluczowym procesem w rozwoju oocytu. Jednak myszy pozbawione GLD2, jedynej znanej cytoplazmatycznej polimerazy poli(A) w oocycie, pozostają płodne, czyniąc niezbędną aktywność innych, nieznanych polimeraz poli(A) w oogenezie.

Z drugiej strony wiadomo, że mechanizmy kontroli jakości i degradacji RNA mogą regulować pluripotencję i różnicowanie komórek poprzez selektywną degradację mRNA. Zarówno deadenylacja, jak i usuwanie czapeczki (z dodatkową urydylacją pomiędzy tymi etapami) mogą wywoływać egzorybonukleolityczną degradację odpowiednio przez kompleks egzosomu i Xrn1. Co istotne, do dzisiaj wykazano, że tylko te egzorybonukleazy związane z egzosomem, EXOSC10 i DIS3, są niezbędne dla rozwoju zarodkowego myszy.

Oba te aspekty regulacji transkryptomu w rozwoju były przedmiotem mojego projektu doktorskiego. Badałem role nowej rodziny polimeraz poli(A) - TENT5 - w oogenezie oraz ostatniej podjednostki katalitycznej cytoplazmatycznego kompleksu egzosomu o nieokreślonej roli - egzorybonukleazy DIS3L w embriogenezie myszy.

Niekanoniczne polimerazy poli(A) TENT5B i TENT5C odgrywają niezbędną, ale redundantną rolę w prawidłowej folikulogenezie i rozwoju oocytów. Mutacja typu knock-out w obu genach prowadzi do degeneracji oocytów, podczas gdy wprowadzenie sekwencji kodującej białko GFP do genu *Tent5b* prowadzi do zaburzeń w organizacji chromatyny w oocycie, co również prowadzi do bezpłodności. Bezpośrednie sekwencjonowanie RNA ujawniło, że białka TENT5 poliadenylują szereg niezbędnych, specyficznych dla oocytów transkryptów kodujących wydzielane białka, w szczególności ZP3 i GDF9. Test z użyciem reporterów RNA wykazał, że sekwencja kierująca do siateczki śródplazmatycznej może być

czynnikiem determinującym poliadenylację transkryptu przez białka TENT5. Dodatkowo, badanie kompozycji ogona poli(A) u zwierząt typu dzikiego ujawniło specyficzną nadreprezentację urydyny blisko końca 3’ w podgrupie transkryptów niezbędnych do spermatogenezy.

Tymczasem białko DIS3L jest niezbędne do rozwoju zarodka. Z zarodków pozbawionych DIS3L wyprowadzić można zdrowe zarodkowe komórki macierzyste, ale same zarodki przestają rozwijać się po 6,5 dniach od zapłodnienia. Sekwencjonowanie RNA z blastocyst z mutacją typu knock-out w genie *Dis3l* wykazało akumulację kilkunastu transkryptów. Co ciekawe jednak, poziom białek kodowanych przez te transkrypty był obniżony, co spowodowane było globalnym obniżeniem produkcji białek w zarodkach z mutacją genu *Dis3l*.

## Table of Contents

Podziękowania ..... 5
Abstract ..... 6
Streszczenie ..... 8
Introduction ..... 13
Transcriptome regulation in development ..... 13
Oogenesis as a model for cytoplasmic polyadenylation ..... 13
CPEB model of polyadenylation ..... 14
TENT5 family of poly(A) polymerases ..... 14
Embryo development as a model of selective mRNA degradation ..... 16
3'->5' and 5'->3' mRNA decay pathways ..... 16
DIS3 family of ribonucleases ..... 18
Purpose of this work ..... 18
Summary of the results ..... 19
Manuscript 1: Role of TENT5 poly(A) polymerases in oogenesis ..... 19
Manuscript 2: Composition of poly(A) tails in mouse testes and ovaries ..... 20
Manuscript 3: Role of DIS3L exoribonuclease in embryogenesis ..... 20
Bibliography ..... 22
Manuscripts included in the PhD thesis and candidates contribution ..... 28
Manuscript 1. Brouze M., Czarnocka-Cieciura A., Gewartowska O., Kusio-Kobiałka M., Jachacy K., Szpila M., Tarkowski B., Gruchota J., Krawczyk P., Mroczek S., Borsuk E., Dziembowski A. *TENT5-mediated polyadenylation of mRNAs encoding secreted proteins is essential for gametogenesis in mice*. Nat Commun 15, 5331 (2024). https://doi.org/10.1038/s41467-024-49479-4 ..... 29
Manuscript 2. Czarnocka-Cieciura A., Brouze M., Gumińska N., Mroczek S., Gewartowska O., Krawczyk P., Dziembowski A. *Comprehensive analysis of poly(A) tails in mouse testes and ovaries using Nanopore Direct RNA Sequencing*. Scientific Data, accepted, in press (2024) ..... 50
Manuscript 3. Brouze M., Szpila M., Czerwińska A., Antczak W., Mroczek S., Kuliński T., Hojka-Osińska A., Cysewski D., Adamska D., Gruchota J., Borsuk E., Dziembowski A. *DIS3L, cytoplasmic exosome catalytic subunit, is essential for development but not cell viability in mice*. bioRxiv 2022.12.14.520403; doi: https://doi.org/10.1101/2022.12.14.520403 ..... 69

# Introduction

## Transcriptome regulation in development

Development in virtually all organisms is a dynamic and highly complicated process. The complexity rises dramatically with the complexity of the organism undergoing this development. In vertebrates, it involves extremely rapid growth and differentiation of hundreds and thousands of cell types, which will give rise to an entire new organism in a short span of time. This would not be possible without the varied and complex means of gene expression regulation. Mechanisms governing the timing and spatial expression activation are vital for the development of such a great diversity of cells and tissues. Most notably, rapid changes in RNA polyadenylation and selective, dynamic degradation of mRNAs in different developing tissues are key to properly guiding the development. For this reason, multiple stages of this process in vertebrates have been recognised for decades as useful models to study various aspects of RNA regulation mechanisms.

## Oogenesis as a model for cytoplasmic polyadenylation

In entire vertebrate development, oogenesis is a unique and specifically regulated process itself. Starting at early gastrulation stages, with differentiation of primordial germ cells[1], it continues for months or years until the sexual maturation of an adult animal. The final stage of development is marked by oocytes recruitment for ovulation, which initiates rapid growth of transcriptional activity when vast amounts of mRNAs are produced and stored in the oocyte. As the oocyte approaches ovulation, starting at the earliest antral stages, transcription is gradually decreased until near complete halt in fully developed follicle[2,3] and it will not resume until embryonic genome activation (EGA, also known as ZGA, zygotic genome activation) occurs, which, depending on the species, can take up to over a dozen rounds of division[4-7]. Until then, oocytes and early embryos development rely on a pool of mRNAs gathered during oogenesis. This requires unique mechanisms of stabilisation and translation activation of those accumulated transcripts. While in the majority of somatic cells, shortening of a poly(A) tail marks the transcript for degradation, in oocytes maintaining a short tail stabilizes it. Such a transcript is stored until rapid poly(A) tail elongation occurs, triggering its translation[8]. This made the oocyte and its development a popular model for studying mechanisms of deadenylation and polyadenylation and their influence on transcript regulation and gene expression.

*CPEB model of polyadenylation*

Seminal works of Joel Richter, and others, on *Xenopus laevis* oocytes[8-11], allowed for the discovery of the main mechanism of polyadenylation in the oocyte of all vertebrates, centred around the Cytoplasmic Polyadenylation Element Binding (CPEB) protein. In the cytoplasm, CPEB binds to a specific UUUUUAU motif in the 3'UTR sequence of mRNA, called cytoplasmic polyadenylation element (CPE), and recruits a number of other proteins that together can regulate translation from bound mRNA particles. Among them, PARN deadenylase and TENT2 poly(A) polymerase (GLD2 according to old nomenclature) are crucial for poly(A) tail length regulation. When bound together in complex with CPEB, counteracting activity of those two enzymes maintains a short oligo(A) tail of the transcript. In most characteristic instance of this mechanism, hormonal stimuli arriving in oocyte upon maturation initiation leads to PARN phosphorylation, causing its dissociation from the complex, which allows for rapid elongation of poly(A) tail by TENT2 and, in consequence, it triggers translation from the polyadenylated mRNA molecule. This model was later studied in a multitude of different vertebrates and other species, solidifying it as the main mechanism of cytoplasmic polyadenylation and translation activation control [12-14]. Indeed, studies of *Cpeb* knock-out mice showed its crucial role for oogenesis[15]. However, while homologues of TENT2 in non-vertebrates are known to be essential for gametogenesis and embryogenesis[16,17], in mice, surprisingly, knock-out mutation of *Tent2* gene has no visible effect on oocyte development[18]. So far, no other cytoplasmic poly(A) polymerase was identified that would target and regulate translation of key mRNAs required during oogenesis, yet the importance of that regulation dictates such enzyme, or enzymes, to exist along TENT2.

*TENT5 family of poly(A) polymerases*

TENT5 (terminal nucleotidyltransferase 5; formerly FAM46) is a family of non-canonical cytoplasmic poly(A) polymerases consisting of four members: TENT5A, B, C and D. Previously overlooked, their polyadenylation activity was first identified through sequence analysis studies[19], and later confirmed by *in vitro* activity tests[20]. Structural studies identified very similar structure among all TENT5 enzymes, with two domains present: nucleotidyltransferase (NTase) domain (containing catalytical site responsible for TENT5s activity) and PAP-associated domain, but no RNA recognition motifs or RNA binding domains[19,21], prompting the still unanswered question of substrate recognition and specificity.

Several years ago, very little was known about the biological role of those enzymes, apart from TENT5C, which was studied most extensively due to being identified as one of the

most frequently mutated genes in patients with multiple myeloma (MM), a cancer of terminally differentiated B cells known as plasma cells[22]. The past few years of extensive research revealed that in mammals, TENT5 proteins are differentially expressed in a multitude of tissues, and they impact multiple different biological pathways and processes[20,23-30].

TENT5A regulates the expression of several secreted proteins essential for bone mineralization, most notably type 1 collagen. The absence of functional TENT5A leads to congenital *osteogenesis imperfecta* disease in humans, and *Tent5a* KO mice recapitulate its symptoms like lowered bone mineralization, skeletal deformations and frequent bone fractures[23]. Its involvement in adolescent idiopathic scoliosis was also suggested based on its myogenin-stabilizing function influencing muscle fibre maturation[29].

TENT5B was previously shown to be highly and specifically expressed in human preimplantation embryo and pluripotent stem cells[25], with its knock-down postulated to be lethal for the latter. Furthermore, it was also found to inhibit cell proliferation in prostate cancer, or even lead to their apoptosis through ubiquitination of β-catenin and decreasing lactate dehydrogenase (LDH) activity and overall glucose uptake[24,31].

TENT5C is best known as a potent oncosuppressor. Numerous studies identified it as one of the most frequently mutated genes in multiple myeloma patients[32-35]. Consequently, TENT5C is highly expressed in differentiating plasma cells and *Tent5c* KO mutation leads to faster proliferation of B cells, which give rise to plasma cells and myeloma cells[20]. Finally, TENT5C, through its polyadenylation activity, regulates expression on endoplasmic reticulum (ER)-targeted and secreted proteins, most prominently immunoglobulins in B cells[30] and immune response proteins (ex. lysozyme and cathepsins) in macrophages[36]. However, recently TENT5C was also described as necessary for spermiogenesis, as the majority of spermatozoa produced by *Tent5c* KO male mice are headless[26].

TENT5D remains the least understood of the whole TENT5 family. So far, its expression and key role have been attributed, similarly to TENT5C, to sperm development, likely due to its localization on the X-chromosome. Two independent studies identified a number of patients carrying novel *Tent5d* mutations that lead to severe teratozoospermia and infertility[27,28,37]. *Tent5d* KO male mice recapitulated this phenotype, displaying multiple morphological abnormalities and drastically decreased motility in developing sperm. Interestingly though, those sperm were viable for ICSI (intracytoplasmic sperm injection) procedure and could produce viable offspring this way[28].

One of the emerging common traits for all TENT5s is affinity towards targeting mRNAs encoding secreted proteins (as in the case of polyadenylation of collagen-encoding mRNAs by

TENT5A or immunoglobulins-encoding mRNAs by TENT5C). Simultaneously, numerous attempts failed to find stable interactions with other proteins that would facilitate mRNA polyadenylation by TENT5 enzymes. Notably, though, TENT5B and TENT5C are both highly expressed in the oocyte. This positions TENT5 family as a potential new, CPEB-independent, mechanism of cytoplasmic polyadenylation.

**Embryo development as a model of selective mRNA degradation**

Embryo development is characterised by dynamic changes in gene expression. This is dictated by the need to direct cells toward differentiation into a multitude of different tissues[38-41]. In the past decade RNA quality control and decay pathways have emerged as an essential mechanism for regulating appropriate state of the transcriptome necessary for both maintaining pluripotency and execution of differentiation programs in developing embryo, stem cells and pools of self-renewing progenitor cells in adult organisms[42-49]. These mechanisms and degradation pathways are well known in somatic, non-differentiating cells and act the same way in a developing embryo. For this reason, development relies on the selectivity of gene expression and transcript degradation.

*3'->5' and 5'->3' mRNA decay pathways*

In a majority of instances decay of cytoplasmic mRNA is initiated by either deadenylation or decapping, which removes protection of 3' and 5' end, respectively, allowing exoribonucleases to degrade a transcript from that end. Additionally, in some cases endoribonucleases can cleave mRNA molecule in the middle, opening it for degradation by exoribonucleases. Those exoribonucleases, and consequently pathways of degradation, can be divided according to direction of degradation.

5'-3' RNA degradation is preceded by cap removal. The main eukaryotic enzyme responsible for this process is DCP2, first discovered in *S. cerevisiae*[50], which activity *in vivo* relies on enhancers of decapping (EDCs), mainly DCP1A, with which it forms catalytically active complex in mammals[51]. Another EDC, EDC4, was shown to serve as a scaffold for DCP1A-DCP2, but simultaneously also for XRN1 exoribonuclease[52]. XRN1 is the main cytoplasmic nuclease responsible for 5'-3' degradation of decapped mRNA in mammals, for instance in nonsense-mediated decay[53,54].

3'-5' decay relies on deadenylation of mRNA. In mammals, but also eukaryotes in general, two most prevalent deadenylase complexes are CCR4-NOT and PAN2-PAN3. CCR4-NOT complex in mammals has a number of catalytically active subunits which are homologs of those initially identified in yeast, most notably CNOT6 and CNOT6L for yeast Ccr4, and

CNOT7, CNOT8 and CAF1Z for yeast Caf1/Pop2[55-58]. In PAN2/PAN3 complex, PAN2 is the active component, with PAN3 facilitating PAN2 recruitment to mRNA and stimulating its activity through interaction with PABP (poly(A) binding protein)[59].

In addition to that, it was shown that following deadenylation, tails shorter than 25 nucleotides can be selectively uridylated by TUT4 and TUT7 terminal uridylyl transferases. This precedes exonucleolytic degradation of the deadenylated transcripts from 3' end or deadenylation-induced decapping and degradation in 5' to 3' direction[60]. It was then confirmed that selective 3' terminal uridylation by TUT4/7 marks maternal transcripts for degradation in developing oocyte[61] and is responsible for a wave of mRNA degradation essential for meiotic progression in spermatogenesis[62].

Deadenylation allows for transcript degradation by exoribonucleases associated with the RNA exosome complex. Its 9 core proteins are highly conserved across all kingdoms of life[63-65], but in eukaryotes they have lost any catalytic activity [65], and serve as a scaffold for said exoribonucleases, facilitating interaction with supporting proteins and substrate recognition[66-68]. In mammals there are 3 enzymes associated with the exosome that degrade RNA by terminal nucleotides removal from 3' end of the RNA molecule: EXOSC10[65,66], DIS3[67] and DIS3L[67,68].

These mechanisms of degradation are well recapitulated in the embryo development. The best example of this is the complex regulation of RNA degradation during oocyte-to-embryo transition (OET) when maternal mRNAs accumulated in the developing oocyte are gradually degraded and replaced by mRNAs newly transcribed from the zygotic genome after zygotic genome activation (ZGA). In mice, this transition is very rapid, and the first major wave of degradation is initiated with meiosis resumption[69] (followed by the next waves upon fertilization and ZGA), with the 2-cell embryo nearly fully relying on its own transcription[7]. Upon maturation initiation, maternal mRNAs encoding a number of key enzymes involved in 5'-3' and 3'-5' mRNA degradation pathways are recruited for translation. These are DCP1A and DCP2 decapping enzymes as well as CNOT7 and PAN2 deadenylases. Expression of DCP1A and DCP2 leads to rapid decapping and exposure of a vast number of mRNAs for XRN1-mediated 5' to 3' degradation in the zygote[70]. Meanwhile, deadenylation by newly expressed maternal CNOT7 and PAN2 prepares the mRNAs to be degraded by exonucleases associated with the exosome complex, although of those two only CNOT7 expression during oocyte maturation is necessary for proper ZGA in 2-cell embryos[71].

However, the process of selecting transcripts for degradation in subsequent waves during OET remains poorly understood. It is known that the resumption of meiosis triggers the phosphorylation of the RNA binding MSY2 protein by CDK-1. MSY2 has little, if any,

sequence specificity, and, when bound to RNA, both stabilizes it and inhibits its translation[72]. Its phosphorylation and degradation leads to a sudden increase of mRNA instability[73], but it's unknown how the pool of mRNAs stabilized in oocyte by MSY2 is selected.

*DIS3 family of ribonucleases*

DIS3 is one of the two exosomes subunits in the nucleus. Strongly limited to the nucleoplasm, it is the only protein of the family that retained both endo- and exoribonucleolytic activity[67]. It's responsible for degradation of numerous different RNA species such as Promoter Upstream Transcripts (PROMPTs), lncRNAs and premature transcription termination products as well as processing of snoRNAs[74,75]. DIS3 is one of the most frequently mutated genes in multiple myeloma patients[33,35,76], but it was also shown to be essential for embryo development[77], similarly to another nuclear exosome-bound exoribonuclease, EXOSC10[65,78], which mutation leads to embryo failure to develop beyond morula stage[79].

DIS3L exoribonuclease is the only catalytic subunit of the cytoplasmic exosome[67,68]. Our limited knowledge of its activity and role comes from research in human stable cell lines, where it was implicated in degrading 28S RNA[67,68]. We lack, however, any tangible knowledge on its role on the organismal level, with the only reporting showing lack of *Dis3l* conditional knock-out mutations impact on male mice fertility[80].

DIS3L2 is the only member of the family not associated with the exosome, and instead acting as exclusively cytoplasmic exoribonuclease in the free, monomeric form[81]. On the molecular level its responsible for the degradation of ARE-contating[81] transcripts and ncRNAs[82], with its mutations leading to Perlman syndrome and Wilms tumor[83-85].

**Purpose of this work**

Based on known research, both TENT5s in oogenesis and DIS3L in early embryo development offer perspectives to study new, unknown mechanisms of RNA regulation in mammalian development. In both cases, their exact role and importance in development were not previously studied, but there is strong evidence supporting their crucial role in oogenesis and embryogenesis, respectively. In oocyte development, a yet undefined poly(A) polymerase, like TENT5, was expected to act in parallel to the established polyadenylation model based on CPEB, and our preliminary experiments showed strong expression of two TENT5 proteins in the oocyte's cytoplasm. In early embryogenesis, all exosome catalytic subunits studied so far, DIS3 and EXOSC10, were essential for embryo development, prompting the same importance for DIS3L, exclusive exosome-bound cytoplasmic exoribonuclease. Our early attempts at breeding mice with *Dis3l* KO mutation revealed its embryo lethal effect. In the three works

attached, I studied the role of TENT5B and TENT5C in oogenesis, as well as DIS3L in early embryogenesis of mice. By analysing phenotypes of knock-out and knock-in mutations of those genes on different levels (organismal, cellular and transcriptomal), I was able to determine their importance.

**Summary of the results**

*Manuscript 1: Role of TENT5 poly(A) polymerases in oogenesis*

As the essential enzymes guiding cytoplasmic polyadenylation in mammalian oogenesis were still not identified, we investigated the role of poorly understood TENT5 family of cytoplasmic poly(A) polymerases in this process. From all TENT5 genes with knock-out mutations, *Tent5b* and *Tent5c* mutations together lead to severe infertility phenotype in female mice. I determined, that a complete knock-out of both genes results in inhibition of ovarian follicle growth and all oocytes' death at early antral stages, while even one wild-type allele of any of those genes is enough to maintain normal fertility, pointing to their redundant role in the oocyte. Curiously, we observed that C-terminal knock-in of the GFP tag in the *Ten5b* gene also severely affected oogenesis, leading to multiple chromatin organisation-related abnormalities in most ovulated metaphase-II oocytes. In this case, however, the severity of the phenotype is dependent on the number of mutated alleles, with heterozygotes displaying higher fertility than homozygotic mutant females, suggesting dominant gain-of-function effect.

To understand what type of transcripts TENT5s polyadenylate, we employed the 3$^{rd}$ generation Direct RNA sequencing on Nanopore platform. Omitting cDNA synthesis and PCR amplification, sequencing of native RNA particles allowed us to precisely measure the length of poly(A) tails unaltered by any additional procedures. With this method we were able to identify number of oocyte-specific transcripts that had shortened poly(A) tail in *Tent5b/c* dKO and elongated tails in TENT5B-GFP knock-in ovaries. Consistently with our previous results on TENT5s activity in other cell types, the majority of those mRNAs encoded secreted proteins that also contained signal peptide sequence responsible for trafficking synthesized protein to endoplasmic reticulum. Furthermore, I provide experimental evidence, that shortened poly(A) tail in *Tent5b/c* dKO ovaries leads to lowered protein expression for two genes essential for oocyte development: *Gdf9* and *Zp3*. Finally, by injecting GV oocyte with RNA reporters encoding YPET fluorescent protein preceded by signal peptide, I show that the latter might be a determining factor in polyadenylation by TENT5B and TENT5C, as TENT5B and TENT5C targets in the oocyte do not share any other specific sequence motifs, like CPE.

These results provide evidence that TENT5s, unlike GLD2, previously thought to be the main cytoplasmic poly(A) polymerase in developing oocytes, are indispensable for the expression of some of the proteins essential for oogenesis. TENT5B and TENT5C regulate protein levels of ZP3 and GDF9 by polyadenylation of mRNAs encoding them. Both proteins are expressed exclusively in the oocyte and their mutations were shown to recapitulate phenotypes observed by us in both mice[86-89] and humans[90,91]. Additionally, they are both secreted proteins containing ER-leading signal-peptide sequence. This being determining factor for polyadenylation by TENT5s is further strengthened by previous studies on their activity[23,30,36].

*Manuscript 2: Composition of poly(A) tails in mouse testes and ovaries*

As mRNA stability is regulated not only by poly(A) tail length but also its composition, we employed Ninetails, a neural network-based algorithm developed by us, to analyse nucleotide composition of poly(A) tails data obtained previously by Direct RNA sequencing of ovarian and testicular RNA samples. In a fraction of heterogeneous poly(A) tails, uridylation was significantly overrepresented, but only in testicular samples. Within those transcripts we identified a subgroup of spermatogenesis-essential ones with uridine residues present predominantly close to 3'-terminus. This result closely aligns with the previously reported essential character of TUT4/7 mediated 3'-terminal uridylation for spermatogenesis progression[62], suggesting specificity of the non-adenosine residues' pattern observed by us and its important role in posttranscriptional gene expression regulation.

*Manuscript 3: Role of DIS3L exoribonuclease in embryogenesis*

As EXOSC10 and DIS3, catalytic subunits of nuclear RNA exosome complex in mammals, were both deemed essential for embryo development, we aimed at determining whether this is the case for DIS3L – catalytic subunit of cytoplasmic exosome complex of yet undefined role . Our attempts to breed mice with *Dis3l* knock-out mutation quickly led me to the conclusion that it's essential for development beyond day 6,5. I've discovered that *Dis3l* KO blastocysts are morphologically unaffected and display a mendelian ratio of genotype distribution. At days 6,5 and 7,5 I observed a slightly lowered percentage of KO embryos, but their morphological condition and overall viability decreased dramatically in that period, with no viable *Dis3l* KO embryos at day 7,5 of development. We could not rescue this phenotype with the supplementation of wild-type (WT) DIS3L to extraembryonic tissues through the construction of chimeric embryos.

On the other hand, *Dis3l* KO blastocysts could produce healthy embryonic stem cells, although at lowered efficiency than wild-type embryos. Classical RNA sequencing of those cells showed downregulation of a small number of transcripts. This was surprising as striping the exosome from the exoribonucleolytic activity of DIS3L should lead to the accumulation of transcripts that would not be properly degraded. And rightly so, sequencing of RNA from single WT and KO blastocysts showed an accumulation of number of transcripts. However, immunostaining I performed showed that accumulation of some of those transcripts did not lead to higher protein levels. By measuring levels of newly synthesized proteins in blastocysts with the methionine analogue assay I determined that *Dis3l* KO embryos suffer from an overall decrease in the amount of proteins produced.

These results highlight the importance of translational control and overall mechanisms of RNA quality control by the exosome, already shown for other pathways in embryo development. Considering the involvement of DIS3L, and the RNA exosome as such, in numerous mRNA and translation quality control pathways[92-94], upregulation of transcripts in DIS3L-defficient embryos would be the result of accumulation of aberrant transcripts rather than upregulation of certain genes. The fallout effect of inhibited translation and apoptosis observed by us in those pre- and post-implantation embryos, respectively, could be explained by the general stress response mechanisms triggered in response to unresolved ribosome collision events on aberrant transcripts accumulating in the cells[95]. However, the low size of the translation inhibition effect observed by us puts this hypothesis in question. Additionally, while DIS3L is undoubtfully essential for embryo development, this is not the case for cultured cells, like ES cells. The reason for this difference might lie in some embryo-specific transcripts regulated by the RNA exosome, but more in-depth transcriptomic analysis of DIS3L-deficient embryos would be required to fully answer that question.

**Bibliography**


1 Ginsburg, M., Snow, M. H. & McLaren, A. Primordial germ cells in the mouse embryo during gastrulation. *Development* **110**, 521-528, doi:10.1242/dev.110.2.521 (1990).

2 Moore, G. P. & Lintern-Moore, S. Transcription of the mouse oocyte genome. *Biol Reprod* **18**, 865-870, doi:10.1095/biolreprod18.5.865 (1978).

3 Moore, G. P., Lintern-Moore, S., Peters, H. & Faber, M. RNA synthesis in the mouse oocyte. *J Cell Biol* **60**, 416-422, doi:10.1083/jcb.60.2.416 (1974).

4 Lu, X. M., Li, J. M., Elemento, O., Tavazoie, S. & Wieschaus, E. F. Coupling of zygotic transcription to mitotic control at the mid-blastula transition. *Development* **136**, 2101-2110, doi:10.1242/dev.034421 (2009).

5 De Renzis, S., Elemento, O., Tavazoie, S. & Wieschaus, E. F. Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. *PLoS Biol* **5**, e117, doi:10.1371/journal.pbio.0050117 (2007).

6 Mathavan, S. *et al.* Transcriptome analysis of zebrafish embryogenesis using microarrays. *Plos Genetics* **1**, 260-276, doi:ARTN e29 10.1371/journal.pgen.0010029 (2005).

7 Hamatani, T., Carter, M. G., Sharov, A. A. & Ko, M. S. Dynamics of global gene expression changes during mouse preimplantation development. *Dev Cell* **6**, 117-131, doi:10.1016/s1534-5807(03)00373-3 (2004).

8 Kim, J. H. & Richter, J. D. Opposing polymerase-deadenylase activities regulate cytoplasmic polyadenylation. *Molecular Cell* **24**, 173-183, doi:10.1016/j.molcel.2006.08.016 (2006).

9 Hake, L. E. & Richter, J. D. Cpeb Is a Specificity Factor That Mediates Cytoplasmic Polyadenylation during Xenopus Oocyte Maturation. *Cell* **79**, 617-627, doi:Doi 10.1016/0092-8674(94)90547-9 (1994).

10 Barnard, D. C., Ryan, K., Manley, J. L. & Richter, J. D. Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. *Cell* **119**, 641-651, doi:10.1016/j.cell.2004.10.029 (2004).

11 Copeland, P. R. & Wormington, M. The mechanism and regulation of deadenylation:: Identification and characterization of PARN. *Rna* **7**, 875-886, doi:Doi 10.1017/S1355838201010020 (2001).

12 Kwak, J. E., Wang, L., Ballantyne, S., Kimble, J. & Wickens, M. Mammalian GLD-2 homologs are poly(A) polymerases. *Proc Natl Acad Sci U S A* **101**, 4407-4412, doi:10.1073/pnas.0400779101 (2004).

13 Rouhana, L. & Wickens, M. Autoregulation of GLD-2 cytoplasmic poly(A) polymerase. *Rna* **13**, 188-199, doi:10.1261/rna.333507 (2007).

14 Cui, J., Sackton, K. L., Horner, V. L., Kumar, K. E. & Wolfner, M. F. Wispy, the Drosophila homolog of GLD-2, is required during oogenesis and egg activation. *Genetics* **178**, 2017-2029, doi:10.1534/genetics.107.084558 (2008).

15 E, F. *et al.* CPEB3 deficiency in mice affect ovarian follicle development and causes premature ovarian insufficiency. *Cell Death Dis* **13**, 21, doi:10.1038/s41419-021-04374-4 (2021).

16 Sartain, C. V., Cui, J., Meisel, R. P. & Wolfner, M. F. The poly(A) polymerase GLD2 is required for spermatogenesis in. *Development* **138**, 1619-1629, doi:10.1242/dev.059618 (2011).

17 Wang, L. T., Eckmann, C. R., Kadyk, L. C., Wickens, M. & Kimble, J. A regulatory cytoplasmic poly(A) polymerase in Caenorhabditis elegans. *Nature* **419**, 312-316, doi:DOI 10.1038/nature01039 (2002).

18 Nakanishi, T. *et al.* Disruption of mouse poly(A) polymerase mGLD-2 does not alter polyadenylation status in oocytes and somatic cells. *Biochem Biophys Res Commun* **364**, 14-19, doi:10.1016/j.bbrc.2007.09.096 (2007).
19 Kuchta, K. *et al.* FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases. *Nucleic Acids Res* **44**, 3534-3548, doi:10.1093/nar/gkw222 (2016).
20 Mroczek, S. *et al.* The non-canonical poly(A) polymerase FAM46C acts as an onco-suppressor in multiple myeloma. *Nat Commun* **8**, 619, doi:10.1038/s41467-017-00578-5 (2017).
21 Kuchta, K., Knizewski, L., Wyrwicz, L. S., Rychlewski, L. & Ginalski, K. Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic Acids Res* **37**, 7701-7714, doi:10.1093/nar/gkp854 (2009).
22 Barbieri, M. *et al.* Compendium of FAM46C gene mutations in plasma cell dyscrasias. *Br J Haematol* **174**, 642-645, doi:10.1111/bjh.13793 (2016).
23 Gewartowska, O. *et al.* Cytoplasmic polyadenylation by TENT5A is required for proper bone formation. *Cell Rep* **35**, 109015, doi:10.1016/j.celrep.2021.109015 (2021).
24 Liang, T. *et al.* FAM46B Promotes Apoptosis and Inhibits Glycolysis of Prostate Cancer Through Inhibition of the MYC-LDHA Axis. *Onco Targets Ther* **13**, 8771-8782, doi:10.2147/OTT.S258724 (2020).
25 Hu, J. L. *et al.* FAM46B is a prokaryotic-like cytoplasmic poly(A) polymerase essential in human embryonic stem cells. *Nucleic Acids Research* **48**, 2733-2748, doi:10.1093/nar/gkaa049 (2020).
26 Zheng, C. W. *et al.* Non-canonical RNA polyadenylation polymerase FAM46C is essential for fastening sperm head and flagellum in mice. *Biol Reprod* **100**, 1673-1685, doi:10.1093/biolre/ioz083 (2019).
27 Zhang, Y. T. *et al.* Novel variations in TENT5D lead to teratozoospermia in infertile patients. *Andrology-Us*, doi:10.1111/andr.13589 (2024).
28 Sha, Y. *et al.* TENT5D disruption causes oligoasthenoteratozoospermia and male infertility. *Andrology-Us* **11**, 1121-1131, doi:10.1111/andr.13407 (2023).
29 Luo, M. *et al.* Tent5a modulates muscle fiber formation in adolescent idiopathic scoliosis via maintenance of myogenin expression. *Cell Prolif* **55**, e13183, doi:10.1111/cpr.13183 (2022).
30 Bilska, A. *et al.* Immunoglobulin expression and the humoral immune response is regulated by the non-canonical poly(A) polymerase TENT5C. *Nat Commun* **11**, 2032, doi:10.1038/s41467-020-15835-3 (2020).
31 Liang, T. *et al.* FAM46B inhibits cell proliferation and cell cycle progression in prostate cancer through ubiquitination of beta-catenin. *Exp Mol Med* **50**, 1-12, doi:10.1038/s12276-018-0184-0 (2018).
32 Boyd, K. D. *et al.* Mapping of chromosome 1p deletions in myeloma identifies FAM46C at 1p12 and CDKN2C at 1p32.3 as being genes in regions associated with adverse survival. *Clin Cancer Res* **17**, 7776-7784, doi:10.1158/1078-0432.CCR-11-1791 (2011).
33 Chapman, M. A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467-472, doi:10.1038/nature09837 (2011).
34 Manier, S. *et al.* Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol* **14**, 100-113, doi:10.1038/nrclinonc.2016.122 (2017).
35 Lohr, J. G. *et al.* Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell* **25**, 91-101, doi:10.1016/j.ccr.2013.12.015 (2014).

36 Liudkovska, V. *et al.* TENT5 cytoplasmic noncanonical poly(A) polymerases regulate the innate immune response in animals. *Sci Adv* **8**, eadd9468, doi:10.1126/sciadv.add9468 (2022).
37 Cong, J. *et al.* Deficiency of X-linked TENT5D causes male infertility by disrupting the mRNA stability during spermatogenesis. *Cell Discovery* **8**, 23, doi:10.1038/s41421-021-00369-9 (2022).
38 Briggs, J. A. *et al.* The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, 980-+, doi:ARTN eaar5780 10.1126/science.aar5780 (2018).
39 Wagner, D. E. *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981-+, doi:10.1126/science.aar4362 (2018).
40 Karaiskos, N. *et al.* The Drosophila embryo at single-cell transcriptome resolution. *Science* **358**, 194-199, doi:doi:10.1126/science.aan3235 (2017).
41 Peng, G. *et al.* Molecular architecture of lineage allocation and tissue organization in early mouse embryo. *Nature* **572**, 528-532, doi:10.1038/s41586-019-1469-8 (2019).
42 Lou, C. H. *et al.* Posttranscriptional control of the stem cell and neurogenic programs by the nonsense-mediated RNA decay pathway. *Cell Rep* **6**, 748-764, doi:10.1016/j.celrep.2014.01.028 (2014).
43 Colak, D., Ji, S. J., Porse, B. T. & Jaffrey, S. R. Regulation of axon guidance by compartmentalized nonsense-mediated mRNA decay. *Cell* **153**, 1252-1265, doi:10.1016/j.cell.2013.04.056 (2013).
44 Gong, C. G., Kim, Y. K., Woeller, C. F., Tang, Y. L. & Maquat, L. E. SMD and NMD are competitive pathways that contribute to myogenesis: effects on PAX3 and myogenin mRNAs. *Gene Dev* **23**, 54-66, doi:10.1101/gad.1717309 (2009).
45 Mistry, D. S., Chen, Y. F. & Sen, G. L. Progenitor Function in Self-Renewing Human Epidermis Is Maintained by the Exosome. *Cell Stem Cell* **11**, 127-135, doi:10.1016/j.stem.2012.04.022 (2012).
46 Skamagki, M. *et al.* RNA Exosome Complex-Mediated Control of Redox Status in Pluripotent Stem Cells. *Stem Cell Rep* **9**, 1053-1061, doi:10.1016/j.stemcr.2017.08.024 (2017).
47 Belair, C. *et al.* The RNA exosome nuclease complex regulates human embryonic stem cell differentiation. *J Cell Biol* **218**, 2564-2582, doi:10.1083/jcb.201811148 (2019).
48 Lou, C. H. *et al.* Nonsense-Mediated RNA Decay Influences Human Embryonic Stem Cell Fate. *Stem Cell Rep* **6**, 844-857, doi:10.1016/j.stemcr.2016.05.008 (2016).
49 Lloret-Llinares, M. *et al.* The RNA exosome contributes to gene expression regulation during stem cell differentiation. *Nucleic Acids Research* **46**, 11502-11513, doi:10.1093/nar/gky817 (2018).
50 Dunckley, T. & Parker, R. The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. *EMBO J* **18**, 5411-5422, doi:10.1093/emboj/18.19.5411 (1999).
51 Lykke-Andersen, J. Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay. *Mol Cell Biol* **22**, 8114-8121, doi:10.1128/MCB.22.23.8114-8121.2002 (2002).
52 Chang, C. T., Bercovich, N., Loh, B., Jonas, S. & Izaurralde, E. The activation of the decapping enzyme DCP2 by DCP1 occurs on the EDC4 scaffold and involves a conserved loop in DCP1. *Nucleic Acids Res* **42**, 5217-5233, doi:10.1093/nar/gku129 (2014).
53 Lejeune, F., Li, X. & Maquat, L. E. Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. *Mol Cell* **12**, 675-687, doi:10.1016/s1097-2765(03)00349-6 (2003).

54 Bashkirov, V. I., Scherthan, H., Solinger, J. A., Buerstedde, J. M. & Heyer, W. D. A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. *Journal of Cell Biology* **136**, 761-773, doi:DOI 10.1083/jcb.136.4.761 (1997).
55 Wagner, E., Clement, S. L. & Lykke-Andersen, J. An unconventional human Ccr4-Caf1 deadenylase complex in nuclear cajal bodies. *Mol Cell Biol* **27**, 1686-1695, doi:10.1128/MCB.01483-06 (2007).
56 Lau, N. C. *et al.* Human Ccr4-Not complexes contain variable deadenylase subunits. *Biochem J* **422**, 443-453, doi:10.1042/BJ20090500 (2009).
57 Albert, T. K. *et al.* Isolation and characterization of human orthologs of yeast CCR4-NOT complex subunits. *Nucleic Acids Research* **28**, 809-817, doi:DOI 10.1093/nar/28.3.809 (2000).
58 Draper, M. P., Salvadore, C. & Denis, C. L. Identification of a Mouse Protein Whose Homolog in Saccharomyces-Cerevisiae Is a Component of the Ccr4 Transcriptional Regulatory Complex. *Molecular and Cellular Biology* **15**, 3487-3495 (1995).
59 Uchida, N., Hoshino, S. & Katada, T. Identification of a human cytoplasmic poly(A) nuclease complex stimulated by poly(A)-binding protein. *J Biol Chem* **279**, 1383-1391, doi:10.1074/jbc.M309125200 (2004).
60 Lim, J. *et al.* Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell* **159**, 1365-1376, doi:10.1016/j.cell.2014.10.055 (2014).
61 Morgan, M. *et al.* mRNA 3' uridylation and poly(A) tail length sculpt the mammalian maternal transcriptome. *Nature* **548**, 347-351, doi:10.1038/nature23318 (2017).
62 Morgan, M. *et al.* A programmed wave of uridylation-primed mRNA degradation is essential for meiotic progression and mammalian spermatogenesis. *Cell Res* **29**, 221-232, doi:10.1038/s41422-018-0128-1 (2019).
63 Lorentzen, E. *et al.* The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nature Structural & Molecular Biology* **12**, 575-581, doi:10.1038/nsmb952 (2005).
64 Chekanova, J. A. *et al.* Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. *Cell* **131**, 1340-1353, doi:10.1016/j.cell.2007.10.056 (2007).
65 Liu, Q., Greimann, J. C. & Lima, C. D. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**, 1223-1237, doi:10.1016/j.cell.2006.10.037 (2006).
66 Dziembowski, A., Lorentzen, E., Conti, E. & Seraphin, B. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat Struct Mol Biol* **14**, 15-22, doi:10.1038/nsmb1184 (2007).
67 Tomecki, R. *et al.* The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *EMBO J* **29**, 2342-2357, doi:10.1038/emboj.2010.121 (2010).
68 Staals, R. H. *et al.* Dis3-like 1: a novel exoribonuclease associated with the human exosome. *EMBO J* **29**, 2358-2367, doi:10.1038/emboj.2010.122 (2010).
69 Su, Y. Q. *et al.* Selective degradation of transcripts during meiotic maturation of mouse oocytes. *Developmental Biology* **302**, 104-117, doi:10.1016/j.ydbio.2006.09.008 (2007).
70 Ma, J., Flemr, M., Strnad, H., Svoboda, P. & Schultz, R. M. Maternally Recruited DCP1A and DCP2 Contribute to Messenger RNA Degradation During Oocyte Maturation and Genome Activation in Mouse. *Biol Reprod* **88**, doi:ARTN 11 10.1095/biolreprod.112.105312 (2013).

71 Ma, J., Fukuda, Y. & Schultz, R. M. Mobilization of Dormant Cnot7 mRNA Promotes Deadenylation of Maternal Transcripts During Mouse Oocyte Maturation. *Biol Reprod* **93**, 48, doi:10.1095/biolreprod.115.130344 (2015).
72 Yu, J. Y., Hecht, N. B. & Schultz, R. M. RNA-binding properties and translation repression in vitro by germ cell-specific MSY2 protein. *Biol Reprod* **67**, 1093-1098, doi:10.1095/biolreprod67.4.1093 (2002).
73 Medvedev, S., Yang, J., Hecht, N. B. & Schultz, R. M. CDC2A (CDK1)-mediated phosphorylation of MSY2 triggers maternal mRNA degradation during mouse oocyte maturation. *Dev Biol* **321**, 205-215, doi:10.1016/j.ydbio.2008.06.016 (2008).
74 Davidson, L. *et al.* Rapid Depletion of DIS3, EXOSC10, or XRN2 Reveals the Immediate Impact of Exoribonucleolysis on Nuclear RNA Metabolism and Transcriptional Control. *Cell Reports* **26**, 2779-+, doi:10.1016/j.celrep.2019.02.012 (2019).
75 Szczepinska, T. *et al.* DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts. *Genome Research* **25**, 1622-1633, doi:10.1101/gr.189597.115 (2015).
76 Weissbach, S. *et al.* The molecular spectrum and clinical impact of DIS3 mutations in multiple myeloma. *Br J Haematol* **169**, 57-70, doi:10.1111/bjh.13256 (2015).
77 Kuliński, T. M. *et al.* Somatic DIS3 mutations in Multiple Myeloma arise early in clonal evolution, but are later counterselected due to toxicity. *bioRxiv*, 2023.2007.2027.550471, doi:10.1101/2023.07.27.550471 (2023).
78 Januszyk, K., Liu, Q. S. & Lima, C. D. Activities of human RRP6 and structure of the human RRP6 catalytic domain. *Rna* **17**, 1566-1577, doi:10.1261/rna.2763111 (2011).
79 Petit, F. G. *et al.* EXOSC10/Rrp6 is essential for the eight-cell embryo/morula transition. *Developmental Biology* **483**, 58-65, doi:10.1016/j.ydbio.2021.12.010 (2022).
80 Wang, X. *et al.* Loss of DIS3L in the initial segment is dispensable for sperm maturation in the epididymis and male fertility. *Reprod Biol* **24**, doi:ARTN 100914 10.1016/j.repbio.2024.100914 (2024).
81 Lubas, M. *et al.* Exonuclease hDIS3L2 specifies an exosome-independent 3 '-5 ' degradation pathway of human cytoplasmic mRNA. *Embo Journal* **32**, 1855-1868, doi:10.1038/emboj.2013.135 (2013).
82 Pirouz, M., Du, P., Munafo, M. & Gregory, R. I. Dis3l2-Mediated Decay Is a Quality Control Pathway for Noncoding RNAs. *Cell Reports* **16**, 1861-1873, doi:10.1016/j.celrep.2016.07.025 (2016).
83 Astuti, D. *et al.* Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. *Nature Genetics* **44**, 277-U275, doi:10.1038/ng.1071 (2012).
84 Chang, H. M., Triboulet, R., Thornton, J. E. & Gregory, R. I. A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. *Nature* **497**, 244-+, doi:10.1038/nature12119 (2013).
85 Hunter, R. W. *et al.* Loss of Dis3l2 partially phenocopies Perlman syndrome in mice and results in upregulation of Igf2 in nephron progenitor cells. *Gene Dev* **32**, 903-908, doi:10.1101/gad.315804.118 (2018).
86 Liu, C. *et al.* Targeted disruption of the mZP3 gene results in production of eggs lacking a zona pellucida and infertility in female mice. *Proc Natl Acad Sci U S A* **93**, 5431-5436, doi:10.1073/pnas.93.11.5431 (1996).
87 Rankin, T. *et al.* Mice homozygous for an insertional mutation in the Zp3 gene lack a zona pellucida and are infertile. *Development* **122**, 2903-2910, doi:10.1242/dev.122.9.2903 (1996).

88 Dong, J. W. *et al.* Growth differentiation factor-9 is required during early ovarian folliculogenesis. *Nature* **383**, 531-535, doi:DOI 10.1038/383531a0 (1996).
89 Carabatsos, M. J., Elvin, J., Matzuk, M. M. & Albertini, D. F. Characterization of oocyte and follicle development in growth differentiation factor-9-deficient mice. *Dev Biol* **204**, 373-384, doi:10.1006/dbio.1998.9087 (1998).
90 Chen, T. *et al.* A Recurrent Missense Mutation in ZP3 Causes Empty Follicle Syndrome and Female Infertility. *Am J Hum Genet* **101**, 459-465, doi:10.1016/j.ajhg.2017.08.001 (2017).
91 Zhang, D. *et al.* A novel mutation in ZP3 causes empty follicle syndrome and abnormal zona pellucida formation. *J Assist Reprod Genet* **38**, 251-259, doi:10.1007/s10815-020-01995-0 (2021).
92 van Hoof, A., Frischmeyer, P. A., Dietz, H. C. & Parker, R. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* **295**, 2262-2264, doi:10.1126/science.1067272 (2002).
93 Mitchell, P. & Tollervey, D. An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3 '-> 5 ' degradation. *Molecular Cell* **11**, 1405-1413, doi:Doi 10.1016/S1097-2765(03)00190-4 (2003).
94 Takahashi, S., Araki, Y., Sakuno, T. & Katada, T. Interaction between Ski7p and Upf1p is required for nonsense-mediated 3 '-to-5 ' mRNA decay in yeast. *Embo Journal* **22**, 3951-3959, doi:DOI 10.1093/emboj/cdg374 (2003).
95 Wu, C. C. C., Peterson, A., Zinshteyn, B., Regot, S. & Green, R. Ribosome Collisions Trigger General Stress Responses to Regulate Cell Fate. *Cell* **182**, 404-+, doi:10.1016/j.cell.2020.06.006 (2020).

## Manuscripts included in the PhD thesis and candidates contribution

Manuscript 1. *"TENT5-mediated polyadenylation of mRNAs encoding secreted proteins is essential for gametogenesis in mice"*

In this manuscript I was responsible for the parts devoted to analysis of TENT5 enzymes' role in oogenesis. I performed gross phenotype analysis of *Tent5a* KO, *Tent5b/c* dKO and *Tent5b*$^{GFP}$ female mice, including tissue preparation, histological staining and microscopy. As a part of in-depth analyses of phenotypes, I performed *Tent5a* KO oocyte maturation, ovary immunostaining experiments, oocyte mRNA reporter microinjections, PAT assays and prepared oocyte and ovary RNA sequencing libraries. Finally, I compiled, reviewed, analysed and prepared figures for all the data presented, excluding RNA sequencing experiments. Final version of the manuscript was written by me and prof. Andrzej Dziembowski, with contributions from other authors.

Manuscript 2: "Comprehensive analysis of poly(A) tails in mouse testes and ovaries using Nanopore Direct RNA Sequencing"

In this manuscript I was responsible for preparing ovary RNA sequencing libraries. Final version of this manuscript was written by Agnieszka Czarnocka-Cieciura, Natalia Gumińska, me and prof. Andrzej Dziembowski.

Manuscript 3. *"DIS3L, cytoplasmic exosome catalytic subunit, is essential for development but not cell viability in mice"*

In this manuscript I was responsible for all experiments conducted directly on embryos and ES cells. Specifically, I performed all embryo genotyping and imaging, embryo and ES cells qPCR and embryo immunostaining experiments, isolated RNA and prepared embryo and ES cells RNA sequencing libraries, constructed and transferred chimeric embryos to recipient females. I also performed gene ontology analysis of RNA sequencing results. I reviewed all the data and prepared all the figures. The final version of the manuscript was written by me and prof. Andrzej Dziembowski.

**Manuscript 1. Brouze M., Czarnocka-Cieciura A., Gewartowska O., Kusio-Kobiałka M., Jachacy K., Szpila M., Tarkowski B., Gruchota J., Krawczyk P., Mroczek S., Borsuk E., Dziembowski A.** ***TENT5-mediated polyadenylation of mRNAs encoding secreted proteins is essential for gametogenesis in mice*****. Nat Commun 15, 5331 (2024). https://doi.org/10.1038/s41467-024-49479-4**

nature communications

Article

https://doi.org/10.1038/s41467-024-49479-4

# TENT5-mediated polyadenylation of mRNAs encoding secreted proteins is essential for gametogenesis in mice

Received: 11 September 2023

Accepted: 31 May 2024

Published online: 22 June 2024

Check for updates

Michał Brouze[1,2], Agnieszka Czarnocka-Cieciura[1], Olga Gewartowska[1,3,4], Monika Kusio-Kobiałka[1], Kamil Jachacy[1,4], Marcin Szpila[3,5], Bartosz Tarkowski[1,2], Jakub Gruchota[1,2], Paweł Krawczyk[1,2], Seweryn Mroczek[1,4], Ewa Borsuk[1,5] & Andrzej Dziembowski[1,2,4,5]

Cytoplasmic polyadenylation plays a vital role in gametogenesis; however, the participating enzymes and substrates in mammals remain unclear. Using knockout and knock-in mouse models, we describe the essential role of four TENT5 poly(A) polymerases in mouse fertility and gametogenesis. TENT5B and TENT5C play crucial yet redundant roles in oogenesis, with the double knockout of both genes leading to oocyte degeneration. Additionally, TENT5B-GFP knock-in females display a gain-of-function infertility effect, with multiple chromosomal aberrations in ovulated oocytes. TENT5C and TENT5D both regulate different stages of spermatogenesis, as shown by the sterility in males following the knockout of either gene. Finally, *Tent5a* knockout substantially lowers fertility, although the underlying mechanism is not directly related to gametogenesis. Through direct RNA sequencing, we discovered that TENT5s polyadenylate mRNAs encoding endoplasmic reticulum-targeted proteins essential for gametogenesis. Sequence motif analysis and reporter mRNA assays reveal that the presence of an endoplasmic reticulum-leader sequence represents the primary determinant of TENT5-mediated regulation.

Essentially, every eukaryotic mRNA possesses a poly(A) tail. Bound by poly(A) binding proteins (PABPs), the poly(A) tail facilitates mRNA export from the nucleus and enhances protein synthesis through interactions with translation initiation factors. The poly(A) tail also stabilizes mRNA molecules by preventing exoribonucleolytic decay. Consequently, their deadenylation rate largely determines mRNA half-life. When a poly(A) tail in the cytoplasm is shortened to ~20 nucleotides (nt), PABP is released, rendering mRNA translationally inactive and susceptible to degradation. However, in specific contexts, deadenylated mRNA is stored in a dormant state to be later readenylated in the cytoplasm, to activate protein synthesis[1–4]. Such non-canonical cytoplasmic polyadenylation has been mostly studied in the context of gametogenesis[1–4] and local translation at synapses[5,6]. In these instances, certain mRNAs are rapidly polyadenylated in response to cellular signals, allowing translation to start in a transcription-independent fashion. Our knowledge of cytoplasmic polyadenylation comes primarily from either biochemical analysis in *Xenopus laevis* oocytes or genetic studies in invertebrates. The first and only extensively studied cytoplasmic poly(A) polymerase is GLD2 (or TENT2 according to new nomenclature)[3,5,7–14]. In non-mammalian species, including *Caenorhabditis elegans*, *Xenopus laevis*, and *Drosophila melanogaster*, TENT2 clearly participates in the elongation of poly(A) tails in gametes and early embryos[3,4,10,15]. TENT2 is also important for synaptic plasticity in the fruit fly[5]. On its own, TENT2 does not contain a detectable RNA-

[1]Laboratory of RNA Biology, International Institute of Molecular and Cell Biology, Warsaw 02-109, Poland. [2]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw 02-106, Poland. [3]Genome Engineering Facility, International Institute of Molecular and Cell Biology, Warsaw 02-109, Poland. [4]Institute of Genetics and Biotechnology, Faculty of Biology, University of Warsaw, Warsaw 02-106, Poland. [5]Laboratory of Embryology, Institute of Developmental Biology and Biomedical Research, Faculty of Biology, University of Warsaw, Warsaw 02-096, Poland. e-mail: adziembowski@iimcb.gov.pl

binding domain and in order to select substrates, it cooperates with several RNA-binding proteins such as GLD3 in *C. elegans*[4]. Seminal work on *X. laevis* oocytes has led to a model of cytoplasmic polyadenylation. In this model, cycles of deadenylation and TENT2-mediated polyadenylation are regulated by the cytoplasmic polyadenylation element binding protein 1 (CPEB1), in cooperation with several other factors[1]. Data has suggested that TENT2 is also involved in cytoplasmic polyadenylation in mammals. This is further supported by experiments in which human TENT2, tethered to a reporter mRNA and injected into *X. laevis* oocytes, activated translation. In line with this hypothesis, the knock-down or overexpression of TENT2 in mouse oocytes results in delayed maturation and frequent arrest in metaphase I[16]. However, TENT2 knockout (KO) mice of both sexes are fertile and display no major phenotype[17]. Additionally, oocyte maturation is normal, and the poly(A) tail length of reporter mRNA is not altered in germline or somatic cells[17]. Moreover, TENT2 KO mice do not exhibit any behavioral abnormalities, suggesting that polyadenylation in neurons is also unaffected[7]. This raises the possibility that in mammals, other TENT protein(s), yet to be identified, could be involved in poly(A) tail length regulation. However, all other previously described mammalian non-canonical poly(A) polymerases are either mitochondrial (mtPAP)[18] or mostly nuclear (TENT4A and TENT4B)[18–21], making them unlikely to play a significant role in cytoplasmic polyadenylation. In contrast to TENT2, its potential mammalian regulators, CPEBs (1–4), are more comprehensively studied[22]. Indeed, CPEB1 KO mice display severe gametogenesis defects[23] as well as impaired long-term potentiation in neurons[24]. The three other CPEB proteins contribute to the regulation of various physiological processes, although in many cases, they function as translational repressors rather than activators[25]. Thus, the mechanisms, and impact of cytoplasmic polyadenylation during gametogenesis remain to be established.

A few years ago, we described a previously overlooked family of non-canonical poly(A) polymerases: TENT5 (FAM46)[26]—all members of which were predicted to contain a putative nucleotidyltransferase catalytic domain[26,27]. In mammals, this family of cytoplasmic proteins has four members (TENT5A, TENT5B, TENT5C, and TENT5D)[27] that are differentially expressed in tissues and organs. TENT5A is involved in osteogenesis[28], TENT5C enhances expression of immunoglobulin mRNA in B cells[29], while TENT5A and TENT5C together regulate the expression of innate immune response proteins in macrophages[30]. Here, we analyzed gametogenesis-related phenotypes of all four *Tent5* KO mutations. We demonstrate the involvement of TENT5B and TENT5C in oogenesis. We further show that TENT5C and TENT5D participate in spermatogenesis, which was also recently identified by others[31–34]. Using direct RNA sequencing, we identified TENT5 substrates in both ovaries and testes. The most prominent mRNAs regulated by TENT5 encode secreted proteins, many of which are essential for gametogenesis such as zona pellucida components produced by oocytes. The leader sequence for the endoplasmic reticulum (ER) is also sufficient for TENT5-mediated poly(A) tail length regulation in oocytes, pointing to a simple mechanism of substrate recognition.

## Results

### TENT5 poly(A) polymerases are involved in both spermatogenesis and oogenesis

To understand the physiological role of the TENT5 family of cytoplasmic poly(A) polymerases, we generated and analyzed genetically modified mouse lines carrying constitutive knockout (KO) mutations in all members of the *Tent5* family. The *Tent5a* and *Tent5c* KO mouse lines were described previously[26,28]. We simultaneously developed two *Tent5b* KO mouse lines—one that has an 11 bp deletion, leading to a p.S119EfsX15 frameshift mutation, and another with a deletion spanning the complete catalytic center of the protein (p.L121P, G122_S238del). Both mouse lines exhibited an identical phenotype; hence, they were used interchangeably in this study. The *Tent5d* KO mouse line has a del36bp, ins6bp mutation resulting in the deletion of ten amino acids, including D82 and D84, which are critical for TENT5D activity, and the insertion of two amino acids (p.76-85del, insCL). Furthermore, due to the lack of reliable, commercially available specific antibodies against TENT5 proteins, we generated several new knock-in mouse lines with either endogenous TENT5B or TENT5D tagged with green fluorescent protein (GFP) or FLAG tags. We also used previously generated FLAG- and GFP-tagged TENT5C mouse lines[26,29].

Individual KOs of all *Tent5* genes did not impact the development and birth ratio in matings of heterozygous males and females in each mouse line. In the case of *Tent5b/c* double knockout (dKO), both parents in the tested mating had *Tent5b*$^{-/-}$ *Tent5c*$^{+/-}$ genotype. Since *Tent5d* is located on the X chromosome, homozygous KO females could not be generated due to male infertility described further (Fig. 1A, *Tent5a* data previously reported by Gewartowska, et al.[28]). Additionally, these mutations did not affect overall weight (except for *Tent5a* KO previously reported by Gewartowska, et al.[28]; Supplementary Fig. 1A) or blood morphology (except for *Tent5c* KO previously reported by Mroczek et al.[26]; Supplementary Fig. 1B).

During the initial breeding process, while establishing KO mouse lines, we observed that homozygotic KO mutations of several *Tent5* genes caused partial or complete infertility in both males and females (Fig. 1B, Table 1). This comprised complete male, but not female, infertility in *Tent5c* KO (*Tent5c*$^{-/-}$), male infertility in *Tent5d* KO (*Tent5d*$^{-/null}$), partial male infertility and complete infertility in *Tent5a* KO (*Tent5a*$^{-/-}$) animals, and complete infertility in *Tent5b/c* double KO (dKO; *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$) females (with one wild-type (WT) allele of either gene sufficient for normal fertility). The presence of the GFP tag at the C-terminus of TENT5B (TENT5B-GFP) also impacted fertility—homozygous *Tent5b*$^{gfp/gfp}$ females were completely infertile, while *Tent5b*$^{wt/gfp}$ females had reduced fertility. In contrast, no such effect was observed for N-terminally tagged TENT5B (GFP-TENT5B).

Despite the partial infertility associated with *Tent5a* KO, the morphology of male sperm and testis was normal. Additionally, we observed several cases of fertile *Tent5a* KO males. In females, ovarian morphology (Supplementary Fig. 1C) and histological structure (Supplementary Fig. 1D) remained normal, and they produced a healthy pool of oocytes at the germinal vesicle (GV) stage (Supplementary Fig. 1E). Furthermore, isolated GV oocytes could mature into metaphase II oocytes in vitro (Fig. 1C). Taken together, these results show that the subfertility of *Tent5a* KO animals is not directly related to the process of gametogenesis itself, but rather to the poor general condition of animals, as previously reported[28].

We conclude that TENT5C and TENT5D are essential for spermatogenesis, whereas TENT5B and TENT5C play a crucial but redundant role in oogenesis.

### Lack of TENT5B and TENT5C leads to early oogenesis arrest

As mentioned previously, a complete double KO of *Tent5b* and *Tent5c* results in female infertility, contrasting the lack of phenotypes in mice with disrupted *Tent2* (*Gld2*)[17]. Even one WT allele of either *Tent5b or Tent5c* is sufficient to maintain normal fertility, pointing to their redundant roles. Accordingly, both genes are highly expressed in GV oocytes, as observed in the oocytes of *Tent5c*$^{GFP}$ and *Tent5b*$^{GFP}$ knock-in females expressing respective C-terminal GFP fusions (Fig. 2A). When we examined 8-week-old *Tent5b/c* dKO female mice, we detected markedly abnormal ovaries compared to heterozygous and WT littermates. These ovaries were smaller and showed no signs of significant follicular growth or earlier ovulation (Fig. 2B, Supplementary Fig. 2A)—neither could be restored by hormonal stimulation with pregnant mare's serum gonadotropin (PMSG) and human chorionic gonadotropin (hCG).

https://doi.org/10.1038/s41467-024-49479-4

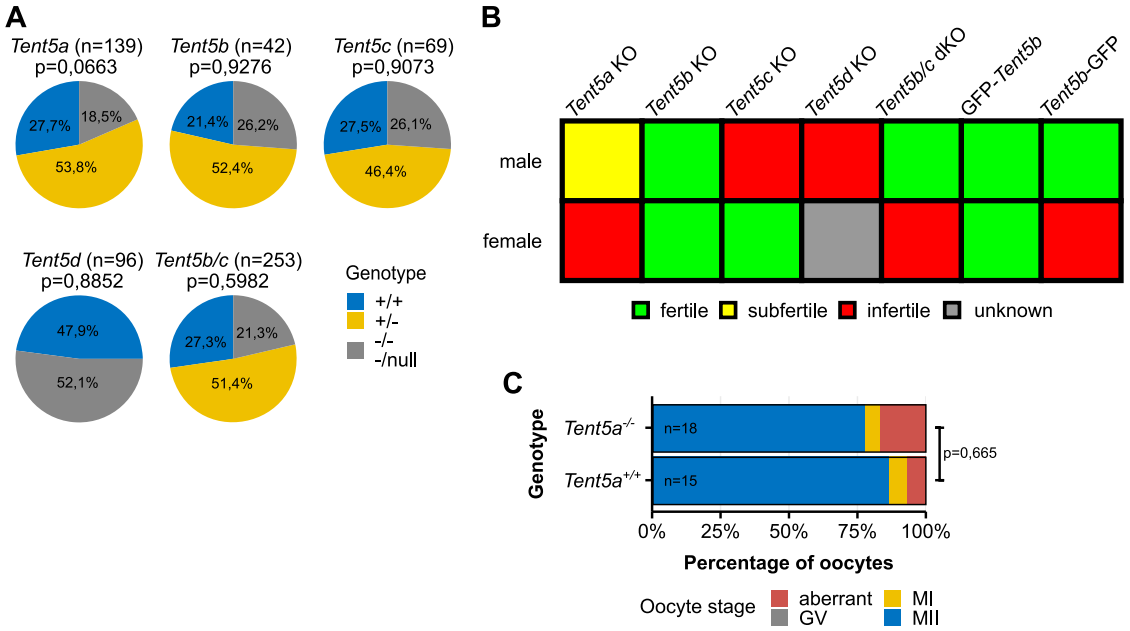



**Fig. 1 | TENT5 poly(A) polymerases are involved in both spermatogenesis and oogenesis. A** Distribution of genotypes in pups of all *Tent5* KO mouse lines from matings of heterozygous parents (only males were accounted for in the *Tent5d* KO line, in the *Tent5*b/c dKO mouse line, both parents had the *Tent5b*$^{-/-}$ *Tent5c*$^{+/-}$ genotype, and for this, only *Tent5c* genotype distribution is presented). P values reported for comparison with expected Mendelian ratios in Fisher's exact test, two-tailed; n values represent the number of pups genotyped. **B** Fertility status of homozygous males and females of all *Tent5* KO and *Tent5b* GFP-tagged mouse lines. **C** In-vitro maturation efficiency of *Tent5a*$^{+/+}$ and *Tent5a*$^{-/-}$ GV oocytes. P value reported for comparing the ratio of metaphase II oocytes to other oocytes after 24 h of in-vitro culture in Fisher's exact test, two-tailed; n values represent the number of oocytes analyzed; GV germinal vesicle, MI metaphase I, MII metaphase II. Source data are provided as a Source Data file.

**Table 1 | Summary of fertility phenotypes discovered for each knock-out and knock-in mutation of *Tent5* genes**

| | Male | Female |
|---|---|---|
| *Tent5a*$^{-/-}$ | Lowered fertility due to poor overall animal condition | Infertility due to poor overall animal condition |
| *Tent5b*$^{-/-}$ | Fertile | Fertile |
| *Tent5c*$^{-/-}$ | Loss of sperm head and DNA material;<br>Fewer total germ cells produced | Fertile |
| *Tent5d*$^{-/null}$ or *Tent5d*$^{-/-}$ | Abnormalities in the development of primary spermatocytes;<br>Degeneration of testes tissue;<br>Lowered blood testosterone levels | Unknown—no homozygous females were analyzed due to *Tent5d* localization on chromosome X |
| *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ | Recapitulated Tent5c$^{-/-}$ phenotype | Inhibited ovarian follicle growth;<br>No ovulation;<br>Degeneration of preovulatory prophase I oocytes |
| *Tent5b*$^{gfp/gfp}$ (N-terminal) | Fertile | Fertile |
| *Tent5b*$^{gfp/gfp}$ (C-terminal) | Fertile | Abnormalities in spindle and chromatin organization in metaphase II oocytes;<br>No ovulation |

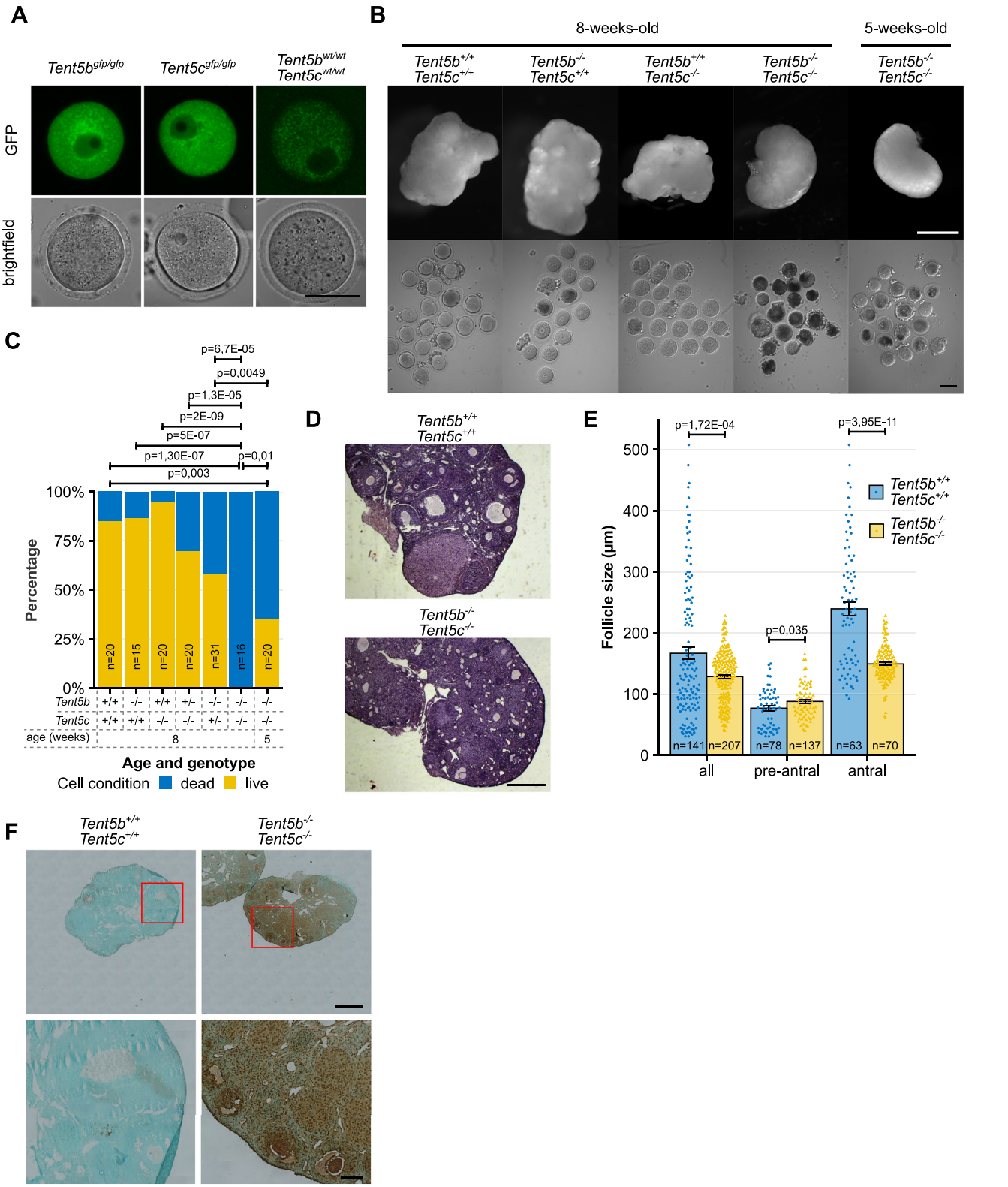
Dissection of these 8-week-old *Tent5b/c* dKO ovaries revealed that all oocytes were in various stages of degeneration and cell death. WT females of the same age maintained a significantly high (85%) population of live oocytes. In comparison, ovaries from single KO females (*Tent5b*$^{-/-}$ or *Tent5c*$^{-/-}$) contained a pool of viable oocytes comparable to that of WT females, while females with a single WT allele of either gene (*Tent5b*$^{+/-}$*c*$^{-/-}$, *Tent5b*$^{-/-}$*c*$^{+/-}$) appeared to have only a slightly lower number of living oocytes. To verify whether this was a congenital or acquired condition, we also dissected 5-week-old *Tent5b/c* dKO ovaries. In these, oocytes showed initial signs of degeneration as seen in adult dKO females, with 35% of oocytes remaining in the living state and the remainder showing various abnormalities—a result that differed significantly from both WT and 8-week-old *Tent5b/c* dKO females (Fig. 2B, C).

Histological staining of individual sections of the ovaries confirmed the arrest of follicular growth in *Tent5b/c* dKO females at the early antral stage. We found neither mature antral follicles ready for ovulation nor the presence of corpora lutea, indicating previous ovulation (Fig. 2D). In *Tent5b/c* dKO females, the observed follicle diameter ranged from 39.8 to 227.8 μm, whereas in WT females, diameters ranged from 30.85 to 507.7 μm (Fig. 2E). Additionally, high levels of apoptosis detected with the TUNEL assay were exclusive to *Tent5b/c* dKO ovaries, observed not only within oocytes but also in surrounding granulosa cells. Ovaries of WT, *Tent5b* KO, and *Tent5c* KO females showed only minor traces of positive TUNEL assay staining (Fig. 2F, Supplementary Fig. 2B), completing the picture of the deterioration of entire *Tent5b/c* dKO ovaries—at morphological, physiological, and cellular levels.

### Dominant effect of TENT5B C-terminal GFP knock-in

A model that provided us with additional insight into the role of TENT5B and TENT5C in oocyte biology was a mouse knock-in line

https://doi.org/10.1038/s41467-024-49479-4




expressing TENT5B with a C-terminal GFP tag (TENT5B-GFP), which, as previously stated, displayed specific infertility in females. On the other hand, N-terminal tagging of TENT5B (GFP-TENT5B) or C-terminal tagging of TENT5C (TENT5C-GFP) did not lead to any oocyte abnormalities. This contrast accompanied a difference in the expression of TENT5B-GFP depending on the GFP fusion site—the mean expression of TENT5B-GFP, measured as the level of GFP fluorescence intensity, was over three times higher than that of GFP-TENT5B (Fig. 3A, B).

The detrimental effect of GFP expression at the C-terminus of TENT5B manifests as an extreme drop in litter sizes, the severity of which depended on the number of GFP alleles in females, thus pointing to a potential dominant effect. To accurately assess this effect, we set up a series of matings with different female and male genotype configurations. We monitored them for 10 weeks and compared the average number of pups birthed by females per week (for the total observation period of 10 weeks). We observed that the negative effect

**Fig. 2 | Lack of TENT5B and TENT5C leads to early oogenesis arrest. A** Expression pattern of TENT5B-GFP and TENT5C-GFP in GV oocytes visible as fluorescence of GFP. Minor fluorescence in WT oocytes represents background autofluorescence in GFP channel. Scale bar = 50 μm. **B** Morphology of ovaries and GV oocytes isolated from adult (8-week-old) and young (5-week-old) female mice. Top scale bar = 1 mm, bottom scale bar = 100 μm. **C** Percentage of live and dead GV oocytes isolated from the ovaries of female mice carrying different combinations of *Tent5b* and *Tent5c* genotypes. N values represent number of oocytes counted; p values reported for comparison of the ratio of live to dead oocytes in Fisher's exact test, two-tailed. **D** Haematoxylin and eosin staining of histological sections from *Tent5b*$^{+/+}$ *Tent5c*$^{+/+}$ and *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ ovaries. Scale bar = 500 μm. **E** Distribution of ovarian follicle size in *Tent5b*$^{+/+}$ *Tent5c*$^{+/+}$ and *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ females, with the analysis of all oocytes further broken down into pre-antral and antral follicle stages. Size value represents the mean width of the follicle at the widest point and width perpendicular to that measurement at the narrowest point, measured on each follicle's widest cross-section available. Individual data points and n values represent individual follicles measured; bars represent mean value; error bars represent SEM; p values reported for follicle size comparison in the t-test, two-tailed. **F** TUNEL assay staining for signs of apoptosis in histological sections from *Tent5b*$^{+/+}$ *Tent5c*$^{+/+}$ and *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ ovaries. Upper scale bar = 500 μm, lower scale bar = 100 μm. See also Supplementary Fig. 2B. Source data are provided as a Source Data file.

of C-terminal GFP on fertility was limited to females: *Tent5b*$^{wt/gfp}$ and *Tent5b*$^{gfp/gfp}$ females suffered from a significant drop in fertility, with a much stronger effect observed in homozygotes (Fig. 3C). Meanwhile, both *Tent5b*$^{wt/wt}$ and *Tent5b*$^{gfp/gfp}$ males parented a similar number of pups when mated with WT females.

Simultaneously, we searched for the cause of this lowered fertility. Knowing the phenotype of *Tent5b/c* dKO females, we focused our investigation on oocyte development. Unlike *Tent5b/c* dKO, oocytes from *Tent5b*$^{GFP}$ (both *Tent5b*$^{wt/gfp}$ and *Tent5b*$^{gfp/gfp}$) females remained alive in the ovaries and progressed further in meiosis through GV to the second metaphase (MII) stage. However, chromosome spreads from these MII oocytes isolated from heterozygous (at the time, homozygous *Tent5b*$^{gfp/gfp}$ females were born at an extremely low rate and were unavailable for these experiments) and control WT females revealed the first signs of the cause underlying observed infertility (Fig. 3D). Although all oocytes of both groups maintained a standard number of 20 pairs of unaltered chromosomes, in most knock-in oocytes analyzed, either one or several chromosomes were separated from the main group of chromosomes, while all chromosomes were gathered in WT oocytes (Fig. 3E). Immunofluorescence staining of chromatin and spindle microtubules in MII *Tent5b*$^{wt/gfp}$ oocytes, of which one-fourth suffered some disruption, indeed revealed disorganized spindles and chromosomes. These stainings additionally revealed parthenogenetic activation as well as large, seemingly empty or fluid-filled follicles within the cytosol (compared to only 2.38% in WT oocytes; Fig. 3F, G).

Together with the observation that the *Tent5b* KO mutation alone does not impact female fertility in any way, our results strongly suggest the dominant-negative or gain-of-function effect of the GFP tag when located at the C-terminus of endogenous TENT5B, as well as the importance of TENT5B activity in the oocyte.

## TENT5s polyadenylate mRNAs in oocytes, the tight regulation of which is essential for oogenesis

As described above, TENT5B and TENT5C play essential roles in oogenesis. To study the effect of dysfunction of these poly(A) polymerases at the mRNA level and identify their potential substrates, we performed direct RNA sequencing (DRS) using the nanopore-based MinION platform (Fig. 4A, Supplementary Data 1). DRS enables genome-wide poly(A) tail length profiling and does not suffer from PCR amplification-dependent biases, which can be particularly pronounced for long poly(A) tails often present in developing gametes. The limitation of this sequencing methodology, however, is that it requires a relatively large quantity of material, rendering the sequencing of RNA from groups of oocytes impossible in our setup. Therefore, we instead used total RNA isolated from whole ovaries of WT, *Tent5b/c* dKO, as well as homozygous and heterozygous TENT5B-GFP 30-day-old female mice.

*Tent5b/c* dKO had essentially no effect on the global mRNA poly(A) tail length distribution in ovaries (mean lengths: 103 nt in WT vs 106 nt in *Tent5b/c* dKO; Fig. 4B). However, analysis of the distribution of differential poly(A) tail lengths revealed in *Tent5b/c* dKO samples several mRNAs with significantly shorter tails, encoding proteins involved in oogenesis, such as mRNA encoding Zona Pellucida Glycoprotein 3 (*Zp3*) or Growth Differentiation Factor 3 (*Gdf9*) (Fig. 4C, Supplementary Fig. 3, Supplementary Data 2). We further validated the poly(A) tail length distributions observed in DRS using a PCR poly(A) test (PAT; Fig. 4D).

For many of the mRNAs with shortened poly(A) tails in *Tent5b/c* dKO, we observed elongation in homo- and hetero-zygous *Tent5b*$^{GFP}$ knock-in mice (Fig. 4C, Supplementary Fig. 3, Supplementary Data 3, 4). This is consistent with the elevated expression of TENT5B protein in the presence of a C-terminal GFP tag. In homo- and hetero-zygous *Tent5b*$^{GFP}$ knock-in ovaries, we observed global shortening of poly(A) tails which most probably represents a secondary effect (Fig. 4B).

To specifically examine oocyte-enriched transcripts, we performed an RNA-seq analysis on oocytes isolated from 5-week-old WT and *Tent5b/c* dKO females using the Illumina platform (Fig. 4E). This allowed us to narrow down the number of transcripts to be analyzed to 522 highly expressed in oocytes (Supplementary Data 5), for which we analyzed the poly(A) tail lengths based on DRS data. This revealed 37 mRNAs that had shortened poly(A) tails in *Tent5b/c* dKO and 48 mRNAs with elongated tails in *Tent5b*$^{gfp/gfp}$ (Fig. 4E). Notably, 20 mRNAs were common among these groups (Fig. 4F). From the pool of 75 potential TENT5 targets, we removed those that had shortened tails in *Tent5b*$^{gfp/gfp}$ or lengthened tails in *Tent5b/c* dKO. We also removed targets that had elongated tails in *Tent5b*$^{gfp/gfp}$ and slightly elongated tails in Tent5b/c dKO from the analyses. The combined poly(A) profile of all 55 mRNAs responding to TENT5 dysfunction (Fig. 4G) revealed an opposite effect of KO and TENT5B-GFP knock-in, supporting the hypothesis that GFP at the C-terminus of TENT5B leads to a gain-of-function phenotype. It further indicates that mRNAs with elongated tails in *Tent5b*$^{gfp/gfp}$ or shortened tails in *Tent5b/c* dKO represent direct targets of TENT5s (Fig. 4E).

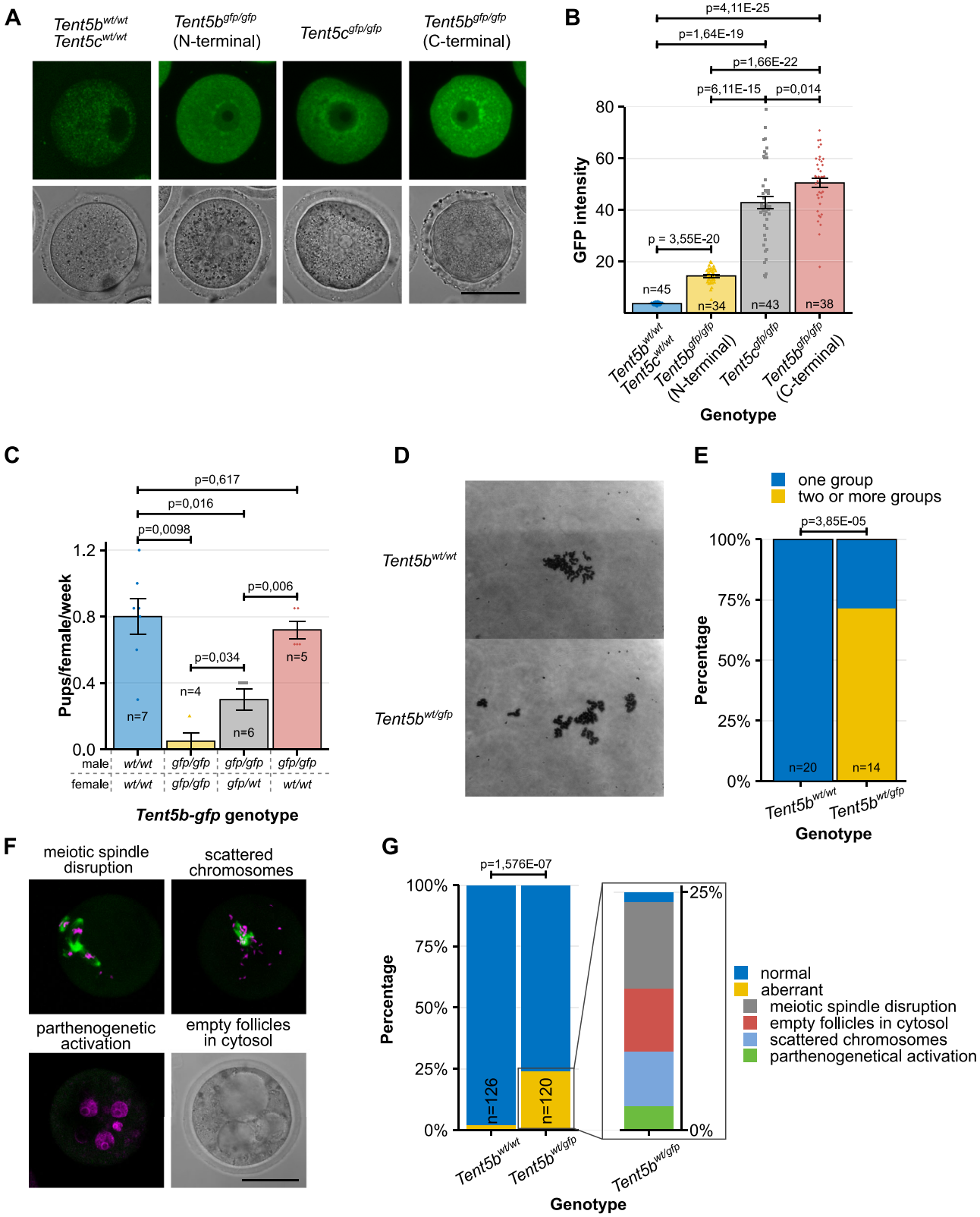
Interestingly, differential expression analysis performed on the WT and *Tent5b/c* dKO Illumina RNA-seq dataset indicated no down-regulation of genes subjected to poly(A) tail shortening in *Tent5b/c* dKO. Nonetheless, immunohistochemistry staining in *Tent5b/c* dKO ovaries (no staining was performed on *Tent5b*$^{gfp/gfp}$ ovaries due to the very low birth rate of those females) revealed a drastic reduction of GDF9 and, to a lesser extent, the level of ZP3 protein in later stages of follicle development. This is in accordance with the observed phenotype of follicle arrest in development at the early antral stage (Fig. 4H, Supplementary Fig. 4). Collectively, these observations point to the previously suggested dominant role of poly(A) tail dynamics in the regulation of translation during gametogenesis.

In summary, TENT5B and TENT5C polyadenylate a group of mRNAs essential for oogenesis, enhancing their expression. Notably, both the lack and excess of TENT5-mediated poly(A) tail extension lead to aberrant oogenesis.

## TENT5C and TENT5D are essential for different stages of spermatogenesis

While neither *Tent5* gene is single-handedly responsible for undisturbed oocyte development, such is not the case for spermatogenesis —as already mentioned, both *Tent5c* and *Tent5d* KO mutations render

https://doi.org/10.1038/s41467-024-49479-4




males sterile[31–34]. However, the expression of these two proteins is markedly different between testis and developing sperm. Immunofluorescence detection of TENT5C-GFP and TENT5D-GFP in histological sections of testis shows high expression of TENT5C close to the lumen of the seminiferous tubules in both round and elongated spermatids for 3- and 8-week-old males. At the same time, TENT5D displays a rather uniform expression across testis tissue in young males, with limited amounts in earlier meiotic stages, and fades from the lumen in adult males (Fig. 5A).

To check the exact pattern of TENT5C and TENT5D expression, we sorted germ cells isolated from males with TENT5C and TENT5D GFP- and FLAG-tag knock-in into different stages of spermatogenesis based on DNA content and chromatin structure (Supplementary Fig. 5) and used them for western blot (FLAG tag lines) and flow cytometry (GFP

**Fig. 3 | Dominant effect of TENT5B C-terminal GFP knock-in. A** Expression pattern of TENT5B-GFP, TENT5C-GFP, and GFP-TENT5B in GV oocytes visible as fluorescence of GFP. Scale bar = 50 μm. **B** Comparison of TENT5B and TENT5C expression levels depending on the presence of the GFP tag at the N- or C-terminus of proteins, measured as a GFP fluorescence intensity. Individual data points and n values represent individual oocytes analyzed, bars represent mean intensity values, error bars represent SEM; p values reported for comparison of mean values in the t-test, two-tailed. **C** Birth rate for different mating configurations depending on male and female C-terminal *Tent5b*$^{GFP}$ genotype, presented as the number of pups born by individual females per week. Individual data points and n values represent individual females observed for 10 weeks, bars represent mean value, error bars represent SEM; p values reported for comparison of the number of pups in the Mann–Whitney–Wilcoxon test, two-sided. **D**, **E** Chromosome scatter preparations showing chromosome conditions and organization in *Tent5b*$^{wt/wt}$ and *Tent5b*$^{wt/gfp}$ MII oocytes. N values represent the number of individual oocytes analyzed; p value reported for comparison of the ratio of oocytes with one group of chromosomes to oocytes with chromosomes separated into two or more groups in Fischer's exact test, two-tailed. **F** Chromatin (magenta) and alpha-tubulin (green) immunofluorescence staining in *Tent5b*$^{wt/gfp}$ MII oocytes. Visible are examples of different chromosomal organization errors and other aberrations occurring after oocyte ovulation. Scale bar = 50 μm. **G** Frequency distribution of different chromosomal segregation and chromatin organization abnormalities in *Tent5b*$^{wt/gfp}$ MII oocytes. N values represent individual oocytes analyzed; p value reported for comparison of the ratio of normal to abnormal oocytes in Fisher's exact test, two-tailed. Source data are provided as a Source Data file.

lines) analyses. Our results confirmed that TENT5C is initially expressed in secondary spermatocytes (spcII), with high protein levels maintained throughout the rest of the spermiogenesis process, while TENT5D expression, already present in primary spermatocytes, peaks in spcII cells to rapidly drop in round- and elongated spermatids (RS and ES, respectively) (Supplementary Fig. 6A–D)

Expression patterns of TENT5C and TENT5D correspond to sperm phenotypes that we and others[31–34] observed—*Tent5c* KO causes a high rate of head or DNA material loss in sperm isolated from the epididymis (Fig. 5B), while in *Tent5d* KO mice, testes become relatively smaller throughout the life of mice as compared to WT males (Supplementary Fig. 6E). This is a result of severe tissue structure deterioration, rendering the testes of 28-week-old males "hollow" (Fig. 5C). Consequently, blood testosterone levels in *Tent5d* KO males were significantly lower than in WT males (Supplementary Fig. 6F).

To gain deeper insight into the spermatogenesis process, we performed a more detailed flow cytometry analysis of the cell cycle. *Tent5c* KO males produce, on average, fewer total germ cells than WT males (Fig. 5D) but accumulate more ES in the lumen of seminiferous tubules (Supplementary Fig. 6G). *Tent5d* KO hinders germ cells from progressing through the meiosis – analysis of germ cell content in males aged 3, 4, 7, 10, and 15 weeks (single male per timepoint) showed that, while in WT males the population of primary spermatocytes (defined in flow cytometry by 4C DNA content) diminishes throughout mice life in favor of a rapidly growing number of spermatids and spermatozoa (collectively identified by 1C DNA content), in KO mice, this process is absent, and the ratio of germ cell populations remains stable through 15 weeks of mice life (Fig. 5E).

Then, we determined whether changes in the germline alone cause abnormal germ cell development and survivability or also depend on surrounding somatic cells. To this end, we co-cultured WT, *Tent5c* KO, and *Tent5d* KO germ- and Sertoli cells in different configurations for 48 h and accounted for live cells following that timepoint. For both genes, KO germ cells exhibited higher mortality than WT cells regardless of the presence of Sertoli cells in culture. Culturing germ cells with Sertoli cells, regardless of their genotype, improved germ cell survivability. This effect was independent of the genotype of Sertoli cells. Mortality and its improvement in an environment of Sertoli cells were particularly visible in the case of *Tent5d* KO (Fig. 5F) and to a lesser extent for *Tent5c* KO (Supplementary Fig. 6H).

All results described above, as well as the differential localization of TENT5s, indicate that the phenotypes are mainly related to germ cells and that the substrates for TENT5C and TENT5D in testes may differ. To identify these substrates, we have performed DRS of mRNA isolated from the testes of adult *Tent5c* KO and 3-week-old *Tent5d* KO males (due to their age-dependent testicular degeneration). *Tent5c* and *Tent5d* KO mutations both had only minor effects on the global poly(A) tail length profile (Fig. 5G, Supplementary Data 6, 7). Such effects can be partially attributed to changes in cell populations. At the same time, the age of mice has quite a dramatic effect on the global poly(A) tail length distribution, which reinforces the importance of comparing males at the same age (a more thorough study of the effect of age on the activity of TENT5 proteins and poly(A) tail dynamics was out of the scope of this work). Importantly, we identified potential TENT5 substrates with significantly shorter poly(A) tails, with no overlap between mRNAs identified in *Tent5c* and *Tent5d* KO samples (Fig. 5H, Supplementary Data 6, 7). Among mRNAs affected by TENT5 dysfunction in sperm, particular attention is owed to several that play important roles in gametogenesis: *Tppp2*, *Insl3* (in *Tent5c* KO), and *Rnaset2* (in *Tent5d* KO) (Fig. 5I–K). We validated the changes in poly(A) tail length distribution for *Tppp2* and *Insl3* in *Tent5c* KO germ cells and whole testes, respectively, by PAT assay (Supplementary Fig. 6I). Finally, quantitative immunochemistry analysis revealed that the shortening of their mRNA poly(A) tails leads to decreased expression at the protein level during stages of spermatogenesis corresponding to KO phenotypes: TPPP2 in *Tent5c* KO spermatids and RNASET2 in *Tent5d* KO spermatocytes. Additionally, INSL3 expression was lowered in *Tent5c* KO Leydig cells (Fig. 5L–N).

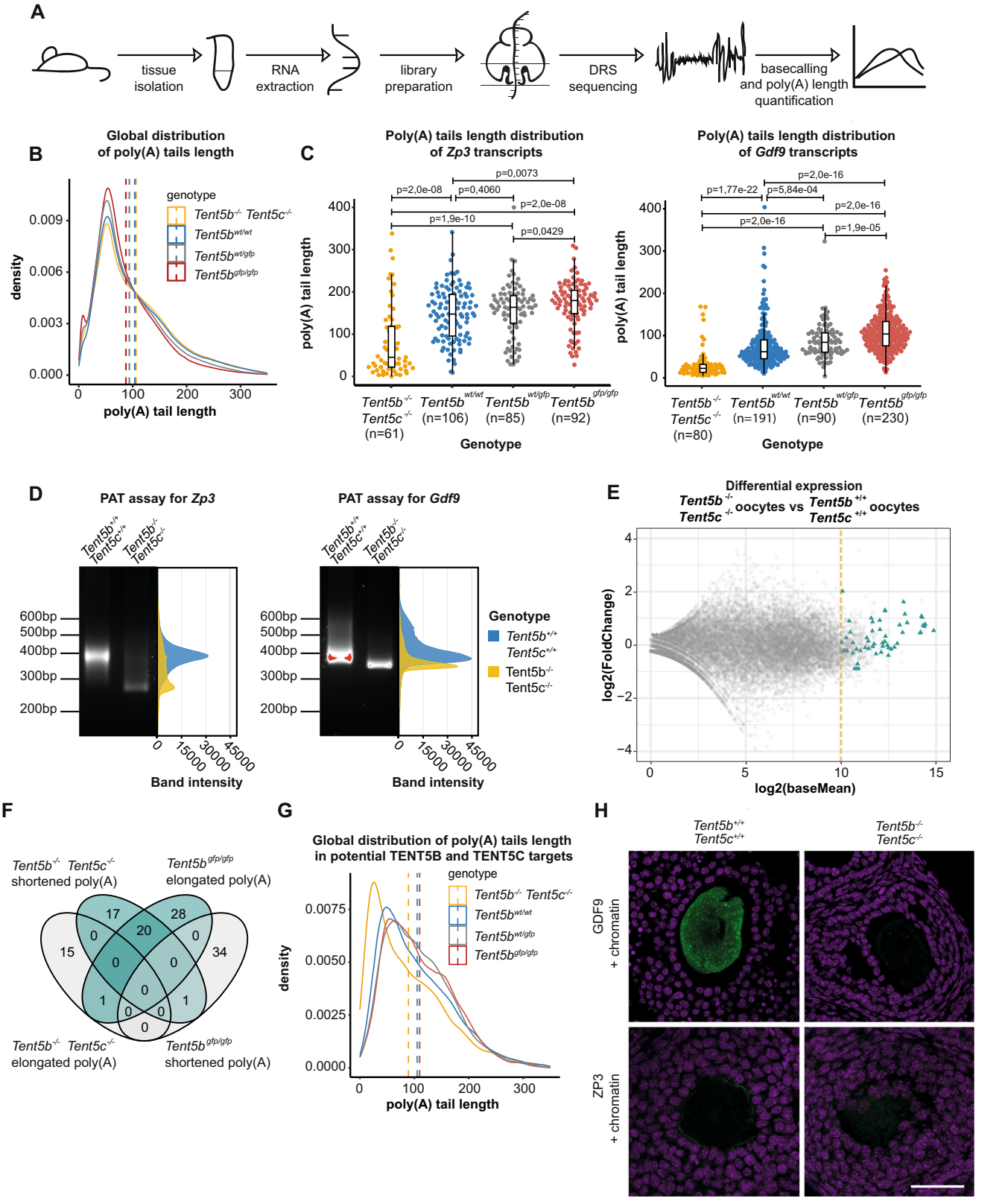
## TENT5s enhance the expression of secreted proteins during gametogenesis

Poly(A) tail profiling using DRS allowed us to identify substrates of TENT5 poly(A) polymerases in testes and ovaries, many of which play essential roles during gametogenesis. This raises the question of the mechanism leading to substrate specificity. To this end, we first searched for sequence features of mRNAs targeted by TENT5B and TENT5C in ovaries, TENT5D in young testes (due to strong deterioration of testicular tissue in adult males), and TENT5C in adult testes. Given that the best-described specificity factors for cytoplasmic polyadenylation are CPEB proteins, we screened the 3′UTR sequences of these transcripts identified by DRS for CPEB1 and CPEB2 motifs. However, no enrichment was found. Moreover, analysis of poly(A) tail length distribution of mRNAs possessing CPEB1/2 motifs showed essentially no effect of TENT5 KO (Fig. 6A, Supplementary Fig. 8A–D). Thus, it is unlikely that CPEB proteins play a role in TENT5-mediated cytoplasmic polyadenylation.

Then, we looked for differences in more general parameters of mRNAs regulated by TENT5s. We first performed 3′UTR and 5′UTR motif analyses on transcripts identified as TENT5 substrates. No specific motif was found except for the canonical polyadenylation signal (Supplementary Fig. 7A, B). Differences in the length of UTR segments, exon length, and GC content (Supplementary Fig. 7C, D), although statistically significant, were negligible.

As sequence analysis did not provide clues to the potential mechanism of substrate selection by TENT5s, we moved on to a functional analysis. Gene ontology (GO) analysis revealed genes involved in reproductive and developmental processes as the most enriched terms (Supplementary Fig. 7E). Moreover, a noticeable fraction of TENT5 substrates were mRNAs encoding secreted proteins (Fig. 6B), which also constituted the majority of TENT5 mRNA substrates in testes. The detection of the endoplasmic reticulum (ER)-

https://doi.org/10.1038/s41467-024-49479-4




targeting signal peptide (SP) confirmed this view (Fig. 6C). SP was found in 6 out of 7 proteins encoded by TENT5D substrates (*Rnaset2a* included) and 8 out of 21 TENT5C substrates in testes. This fraction was smaller in ovaries, with only 9 out of 55 TENT5B/C substrates encoding proteins with SP. Notably, in both testes and ovaries, the effect of TENT5 dysfunction on the lengths of poly(A) tails was more dramatic for mRNAs encoding proteins with ER-targeting signal than for other affected mRNAs, indicating that the former may represent major primary targets of TENT5s (Fig. 6C).

To investigate whether the presence of a SP is a determining factor for TENT5-mediated regulation in the oocyte, we injected GV oocytes with mRNA reporters coding, in 5′–3′ direction: SP (either *Gdf9*- or *Zp3*-specific), YPET fluorescent protein, short ER-retention KDEL motif. Each reporter also had a short oligo-A tail of 20-adenines

**Fig. 4 | TENT5s polyadenylates mRNAs in oocytes, the tight regulation of which is essential for oogenesis. A** Schematic of the DRS-based poly(A) tail length profiling. **B** The global distribution of poly(A) tail lengths of RNA isolated from the ovaries of *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$, *Tent5b*$^{wt/wt}$, *Tent5b*$^{wt/gfp}$, and *Tent5b*$^{gfp/gfp}$ mice. Dashed lines indicate the mean poly(A) tail lengths: *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ = 106 nt; *Tent5b*$^{wt/wt}$ = 103 nt; *Tent5b* $^{wt/gfp}$ = 93 nt; *Tent5b* $^{gfp/gfp}$ = 87 nt. **C** DRS-based poly(A) tail lengths profiling of *Zp3 and Gdf9* mRNAs isolated from ovaries. Median poly(A) tail lengths were marked as a horizontal line on boxplot, borders of box are IQR and black vertical lines correspond to 1.5 IQR. Median calculated for *Zp3*: *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ = 49 nt; Tent*5b*$^{wt/wt}$ = 141 nt; *Tent5b* $^{wt/gfp}$ = 156 nt; Tent*5b*$^{gfp/gfp}$ = 170 nt; p values reported for multiple pairwise comparisons in the Mann–Whitney–Wilcoxon test, two-sided, with Bonferroni Hallberg correction. Median poly(A) tail lengths for *Gdf9*: *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ = 28 nt; Tent*5b*$^{wt/wt}$ = 66 nt; Tent*5b* $^{wt/gfp}$ = 78 nt; Tent*5b*$^{gfp/gfp}$ = 106 nt; p values reported for multiple pairwise comparisons in the Mann–Whitney–Wilcoxon test with Bonferroni Hallberg correction; individual data points represent individual reads analyzed. **D** PAT assay visualizing the distribution of poly(A) tail lengths for *Zp3* and *Gdf9* transcripts from whole mRNA ovary-isolated samples. **E** RNA-seq performed on oocytes isolated from *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ and *Tent5b*$^{wt/wt}$ mice. Yellow dashed line separates 522 genes enriched in oocytes ($log_2$fold change = 2 or more). Most potential TENT5B/C targets, highlighted by triangles, are highly expressed in *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ oocytes. Differential expression calculated using DESeq2. **F** Venn diagram illustrating the overlaps in sets of transcripts with statistically significant changes in the poly(A) tail length when compared with the Mann–Whitney–Wilcoxon test, two-sided, alpha = 0.05. Minimum poly(A) tail length difference is 10 adenosines. **G** Changes in the global distribution of poly(A) tail lengths for the group of transcripts selected as potential TENT5B and TENT5C targets. Dashed lines indicate the mean poly(A) tail lengths. **H** Immunohistochemistry staining of GDF9, ZP3 (green), and chromatin (magenta) in ovaries of *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ and *Tent5b*$^{+/+}$ *Tent5c*$^{+/+}$ female mice. Scale bar = 50 μm. Source data are provided as a Source Data file.

at the 3′end to facilitate further polyadenylation in the cell. Such a reporter, when polyadenylated in the oocyte, would produce full YPET protein, of which fluorescence changes would inform us on the rate of translation. To draw conclusions on polyadenylation rate, we controlled the translation process by co-injecting individual oocytes with in-vitro polyadenylated, mCherry-coding mRNA (which would not require further polyadenylation to be translated). We then normalized YPET fluorescence intensity to mCherry intensity for every oocyte, tying information on fluorescence changes directly to the polyadenylation rate. For this experiment, we used *Tent5b*$^{-/-}$ *Tent5c*$^{+/-}$ oocytes as they remain viable in ovaries but may already display changes at the transcriptomic level.

We observed that when reporter mRNAs encoded SP that recruits them to the ER, their polyadenylation rate was significantly lower in *Tent5b*$^{-/-}$ *Tent5c*$^{+/-}$ oocytes as compared to WT ones. We observed no such difference for reporter mRNAs encoding only YPET (without ER-targeting signal). Our results suggest that only those transcripts recruited to the ER upon translation initiation depend on further polyadenylation by TENT5 proteins (Fig. 6D).

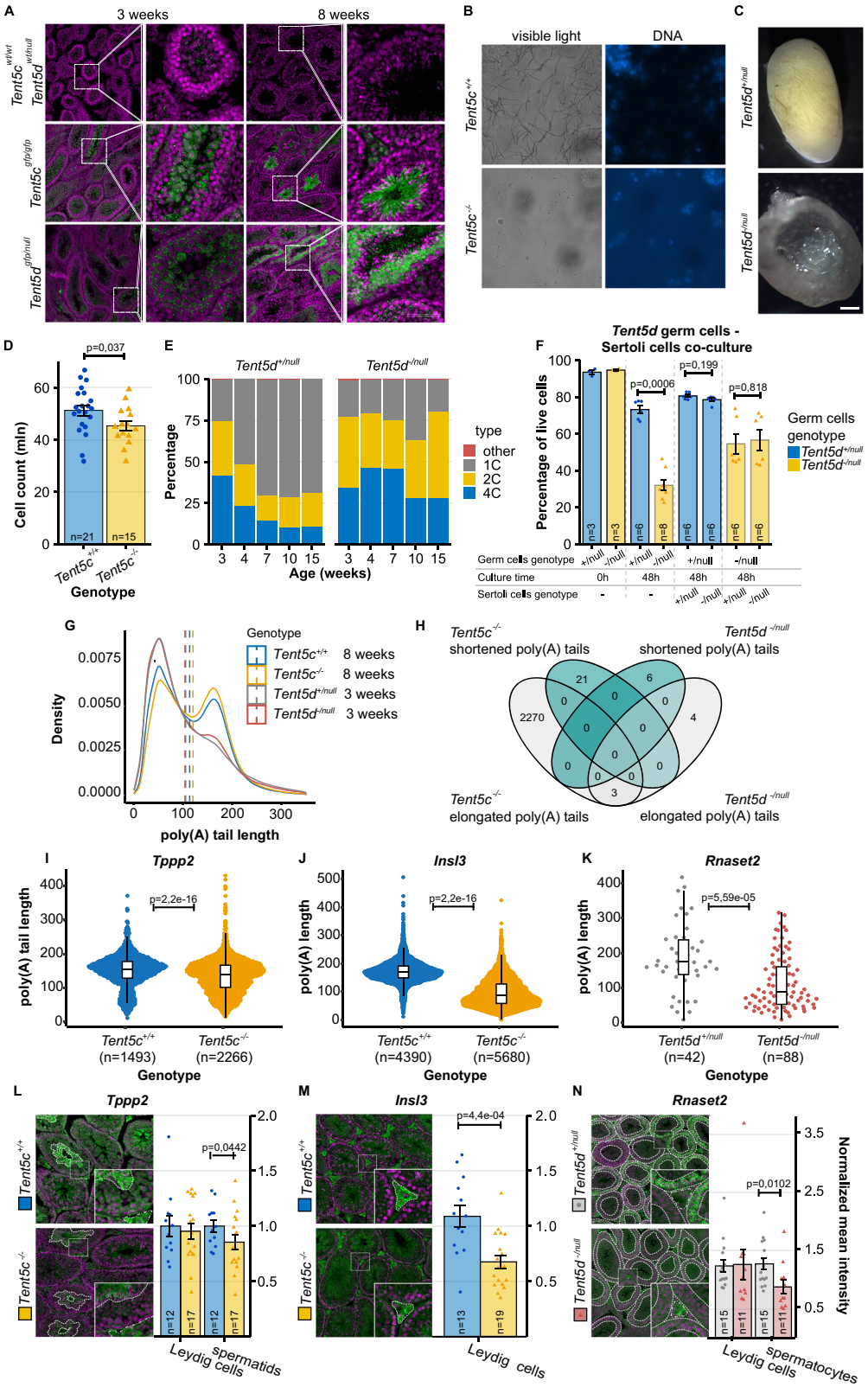
## Discussion

In this study, we demonstrate the essential roles of TENT5B, C, and D poly(A) polymerases during gametogenesis in mice (Fig. 7A, B). In spermatogenesis, in addition to TENT5C and TENT5D, previous studies identified TPAP as a testis-specific poly(A) polymerase, which is present in the cytoplasm of spermatogenic cells; its knockout leads to male infertility[35]. Three distinct cytoplasmic poly(A) polymerases are thus essential for proper spermatogenesis. In contrast, there is redundancy between TENT5B and TENT5C in oogenesis, targeting overlapping pools of transcripts. It is known for years that maintaining the poly(A) tail balance in transcripts accumulated in the oocyte's cytoplasm during its development is crucial for further maturation and fertilization. Until the discovery of TENT5s, the main known mechanism postulated for mRNA poly(A) tail regulation in oocytes was the CPEB protein accompanied by a complex of proteins organized around it[22]. CPEB recognizes specific motifs in the mRNA sequence and was postulated to recruit multiple different effectors like PARN deadenylase and the GLD2 poly(A) polymerase to maintain a short oligo(A) tail of the transcript. When hormonal stimuli arrive to the oocyte, promoting its maturation, PARN is supposed to dissociate from the complex, allowing GLD2 to elongate the transcript's poly(A) tail, thus activating it. The unchanged fertility of *Gld2* KO mice undermines this model and may suggest that TENT5s act redundantly with GLD2[17]. However, the data we provide indicate that TENT5 substrates are largely devoid of CPEB-binding motifs. Additional TENT5B/C act at the stage of oocyte growth when transcription is active rather than during oocyte activation. Thus, the poly(A) polymerase or polymerases responsible for waves of cytoplasmic polyadenylation during oocyte activation remain to be established. These findings reveal a previously unexpected complexity in the regulation of poly(A) tail dynamics during mammalian gametogenesis.

Among TENT5B and TENT5C substrates, there are mRNAs encoding proteins essential for oogenesis. These are oocyte-specific zona pellucida component (ZP3) and signaling protein GDF9. While early studies of murine *Zp3* KO and null mutations showed that it is essential for zona pellucida (ZP) formation and oocyte ovulation along cumulus complexes[36,37], recently described *Zp3* mutations in human patients led to empty follicle syndrome (EFS), where no oocytes could be retrieved from patients' ovarian follicles following hormonal stimulation. This was due to the degeneration of oocytes or their failure to develop properly[38,39]. GDF9 is expressed exclusively in the oocyte[40], but its role extends beyond it, affecting the proliferation of granulosa cells[41,42] and regulating the development of the entire ovarian follicle[43,44], with its deficiency phenotype resembling our observation in *Tent5b/c* dKO female mice. While *Tent5b* KO leads to the shortening of the poly(A) tails of its substrates, the addition of a GFP-tag at the C-terminus of TENT5B leads to the elongation of poly(A) tails, clearly indicating a gain-of-function, possible due to accumulation at the protein level. Interestingly, the unfolded C-terminal region of TENT5B has predicted degron sequences not present in its paralogues (Supplementary Fig. 9)[45]. We thus hypothesize that addition of a GFP-tag at the C-terminus, but not the N-terminus, may inhibit ubiquitination, thereby increasing the stability of TENT5B. From a biological point of view, it is not straightforward to explain the infertility phenotype associated with TENT5B-GFP, manifested by chromosome disorganization during MII, However, at the molecular level, TENT5B-GFP affects proteins that are secreted and appear, at first sight, not to participate in meiosis itself. Further research is needed to elucidate the exact reason for infertility caused by TENT5B gain-of-function. TENT5B is not highly expressed in testes; thus explaining the lack of its effect on spermatogenesis.

The roles of TENT5C and TENT5D in the testis have been previously described by others[31–34]. Yet, the exact reasons for the observed phenotypes remain to be established. These polymerases clearly affect the expression of proteins involved in spermatogenesis. TENT5C regulates the proper expression of TPPP2, a protein vital for sperm motility and its fertilization capacity, and INSL3, a gonad tissue-specific hormone expressed in Leydig cells[46], where it is indispensable for testicular descent[47,48] and promotes germ cell survival[49]. TENT5D controls the expression of extracellular ribonuclease RNASET2 involved in the innate immune response[50], tumor suppression[51], and most importantly for this work, the regulation of sperm motility[52,53]. *Rnaset2* KO mice were shown to recapitulate neurological degeneration symptoms of human RNASET2 deficiency[54]. Other mRNAs affected by TENT5D identified by us and others, such as *Clu, Cst9, Cst12*[32], presumably also contribute to the observed phenotypes. Notably, disruption of TENT5D leads to oligoasthenoteratozoospermia and male infertility in humans[31–34].

https://doi.org/10.1038/s41467-024-49479-4




The mechanism of substrate recognition by TENT5s is far from being understood. However, it becomes apparent that in every biological context, these enzymes are enriched at the ER and, preferably, polyadenylate mRNAs that are translated at the ER (Fig. 7C). Indeed, previous studies identified immunoglobulin-encoding mRNAs as the most affected by *Tent5c* KO. TENT5A in osteoblasts[28] polyadenylates mRNAs encoding constituents of the extracellular matrix essential for bone mineralization, while in macrophages, substrates encode innate immune effectors. Such functions are conserved. In the worm *C. elegans*[30], the only TENT5 enhances the expression of secreted antimicrobial peptides in the gut. Here, we have shown that TENT5B, C, and D proteins are expressed in testes and oocytes, facilitating the expression of secreted proteins. Additionally, the various mRNAs regulated by TENT5s do not display specific sequence motifs. Notably,

**Fig. 5 | TENT5C and TENT5D are essential for different stages of spermatogenesis. A** Immunohistochemistry staining of TENT5C-GFP, TENT5D-GFP (green), and chromatin (magenta) in 3- and 8-week-old testes. Scale bar = 100 μm. **B** DNA staining of *Tent5c*$^{+/+}$ and *Tent5c*$^{-/-}$ sperm isolated from epididymis. **C** Cross section of 28-week-old *Tent5d*$^{+/null}$ and *Tent5d*$^{-/null}$ testes. **D** Total count of germ cells isolated from *Tent5c*$^{+/+}$ and *Tent5c*$^{-/-}$ testes. Individual data points and n values represent individual males used, bars represent mean value and error bars represent SEM; p value reported for t-test, two-tailed. **E** Changes in the distribution of germ cells with different DNA content (representing different stages of spermatogenesis) in *Tent5d*$^{+/null}$ and *Tent5d*$^{-/null}$ testes related to male's age (single male was used for each timepoint). **F** Germ cell survival at 48 h of in-vitro culture depending on the *Tent5d* genotype of germ cells and presence and *Tent5d* genotype of Sertoli cells. Individual data points and n values represent cultures of germ cells isolated from a single male, bars represent mean values, error bars represent SEM; p values reported for comparison in the Mann–Whitney–Wilcoxon test, two-sided. **G** The global distribution of poly(A) tail lengths of RNA isolated from *Tent5c*$^{+/+}$, *Tent5c*$^{-/-}$, *Tent5d*$^{+/null}$, *Tent5d*$^{-/null}$ mice. Dashed lines indicate the mean poly(A) tail lengths: *Tent5c*$^{+/+}$ = 113 nt; *Tent5c*$^{-/-}$ = 123 nt; *Tent5d*$^{+/null}$ = 103 nt; *Tent5d*$^{-/null}$ = 105 nt. **H** Venn diagram illustrating the overlaps in sets of transcripts with statistically significant changes in length of the poly(A) tails, Mann–Whitney–Wilcoxon test, two-sided, alpha = 0.05, minimum poly(A) tail length difference = 10 adenosines. **I**, **J** Changes in lengths of the poly(A) tails in transcripts essential for spermatogenesis. DRS-based poly(A) tail lengths profiling of *Tppp2* and *Insl3* mRNAs isolated from *Tent5c*$^{+/+}$ and *Tent5c*$^{-/-}$ testes. Median poly(A) tail lengths were marked as a horizontal line on boxplot, borders of box are IQR and black vertical lines correspond to 1.5 IQR. Median calculated for *Tppp2*: *Tent5c*$^{+/+}$ = 152 nt; *Tent5c*$^{-/-}$ = 136 nt; individual data points represent reads analyzed (partially obstructed due to the number of reads and resulting density of points); median poly(A) tail lengths for *Insl3*: *Tent5c*$^{+/+}$ = 169 nt; *Tent5c*$^{-/-}$ = 86 nt; p values reported for comparison in the Mann–Whitney–Wilcoxon test, two-sided. **K** Changes in poly(A) tail lengths in transcripts essential for the development of male gametes. DRS-based poly(A) lengths profiling of *Rnaset2* mRNAs isolated from *Tent5d*$^{+/+}$ and *Tent5d*$^{-/-}$ testes. Median poly(A) tail lengths were marked as a horizontal line on boxplot, borders of box are IQR and black vertical lines correspond to 1.5 IQR. Median calculated for *Rnaset2*: *Tent5d*$^{+/null}$ = 175 nt; *Tent5d*$^{-/null}$ = 88 nt; individual data points represent reads analyzed; p values reported for comparison in the Mann–Whitney–Wilcoxon test, two-sided. Immunohistochemistry staining in cross-sections of testes for chromatin (magenta) and proteins (green) encoded by transcripts identified as possible substrates of TENT5C (**L**, **M**) and TENT5D (**N**). Individual data points and n values represent mean values of all ROIs from individual specimens, bars represent mean values, error bars represent SEM, p value reported for the Mann–Whitney–Wilcoxon test, two-sided. Source data are provided as a Source Data file.

the incorporation of ER-targeting leader peptides into YFP-encoding mRNA leads to enhanced expression in WT oocytes but not in oocytes with TENT5B/C dysfunction (Figs. 6D and 7C). Accordingly, our recent data unexpectedly indicates that the Moderna vaccine mRNA-1273, which encodes Spike antigen targeted to the ER, is an efficient substrate for TENT5 polymerases. The Moderna vaccine is completely synthetic and optimized for efficient translation rather than cytoplasmic polyadenylation[55]. This points to a simple model in which efficiently translated mRNA with a leader peptide that are recruited to the ER are polyadenylated by TENT5s, which are enriched at the surface of the ER[29,30,56]. Further research is needed to elucidate how TENT5 poly(A) polymerases are recruited. Yet, the interaction of TENT5s with FNDC3A/B, membrane-bound proteins facing the cytoplasmic site of the ER, suggest that this is the most likely mechanism of such preference[56]. As TENT5s are differentially expressed, polyadenylation occurs only in specific cell types and tissues. For instance, many commonly used cell lines, such as HeLa or HEK293 cells, do not express TENT5s. Interestingly, there are also differences in the effect of TENT5-mediated polyadenylation between somatic cells and during gametogenesis. TENT5 proteins in B cells, macrophages, and osteoblasts mainly regulate mRNA stability. Dysfunction in these proteins thus leads to decreased levels of substrate mRNAs through the regulation of mRNA half-life. In contrast, during gametogenesis, TENT5 KO leads to the shortening of its substrates' poly(A) tails, leading to decreased levels of proteins produced, while mRNA levels remain unchanged. This agrees with previously proposed differences in poly(A) tail metabolism in somatic cells as compared to gametes and early embryos[57], with. In the latter, there are limiting amounts of the poly(A) binding proteins essential for efficient translation. As a result, poly(A) tail extension enhances protein synthesis at the same time in oocytes, and during spermatogenesis, mRNA stability is regulated differently than in somatic cells.

# Methods

## Materials and reagents availability

Plasmids, mouse lines, and reagents generated in this study are available at request from corresponding author.

## Experimental animal models

All experiments were performed using mouse lines generated using CRISPR/Cas9 method in C57BL/6/Tar x CBA/Tar mixed background.

All animal experiments were approved by the Local Ethical Committees in Warsaw affiliated to the University of Warsaw, Faculty of Biology (approval numbers: 176/2026, 917/2019) and Warsaw University of Life Sciences, Faculty of Horticulture and Biotechnology (approval numbers: WAW2/049/2022) and were performed according to Polish Law (Act number 266/15.01.2015).

In this study, females of following genotype and age were used: *Tent5a*$^{-/-}$ (8–12 weeks), *Tent5b*$^{-/-}$*c*$^{-/-}$ (4–12 weeks), *Tent5b*$^{-/-}$*c*$^{+/-}$ (8–12 weeks), *Tent5b*$^{gfp/gfp}$ (C-terminal) (8–12 weeks), *Tent5b*$^{gfp/gfp}$ (N-terminal) (4–12 weeks), *Tent5b*$^{gfp/wt}$ (N-terminal) (4–12 weeks), *Tent5c*$^{gfp/gfp}$ (8–12 weeks), wild-type (4–12 weeks), and males of following genotype and age were used: *Tent5b*$^{gfp/gfp}$ (8–48 weeks), *Tent5c*$^{-/-}$ (8–48 weeks), *Tent5d*$^{null/-}$ (3–48 weeks), *Tent5c*$^{gfp/gfp}$ (8–48 weeks), *Tent5c*$^{flag/flag}$ (8–48 weeks), *Tent5d*$^{gfp/gfp}$ (3 weeks), *Tent5d*$^{flag/flag}$ (8–48 weeks), wild-type (3–48 weeks).

The mice were kept in conventional polystyrene cages in the animal facility of the Faculty of Biology, University of Warsaw. A 12/12 light cycle was maintained in the room, 15 air exchanges per hour. The relative humidity in the rooms was 55% ± 10%. The temperature in the rooms was 22 °C ± 2 °C. During all procedures and breeding the animals received rodent feed and ad libitum water.

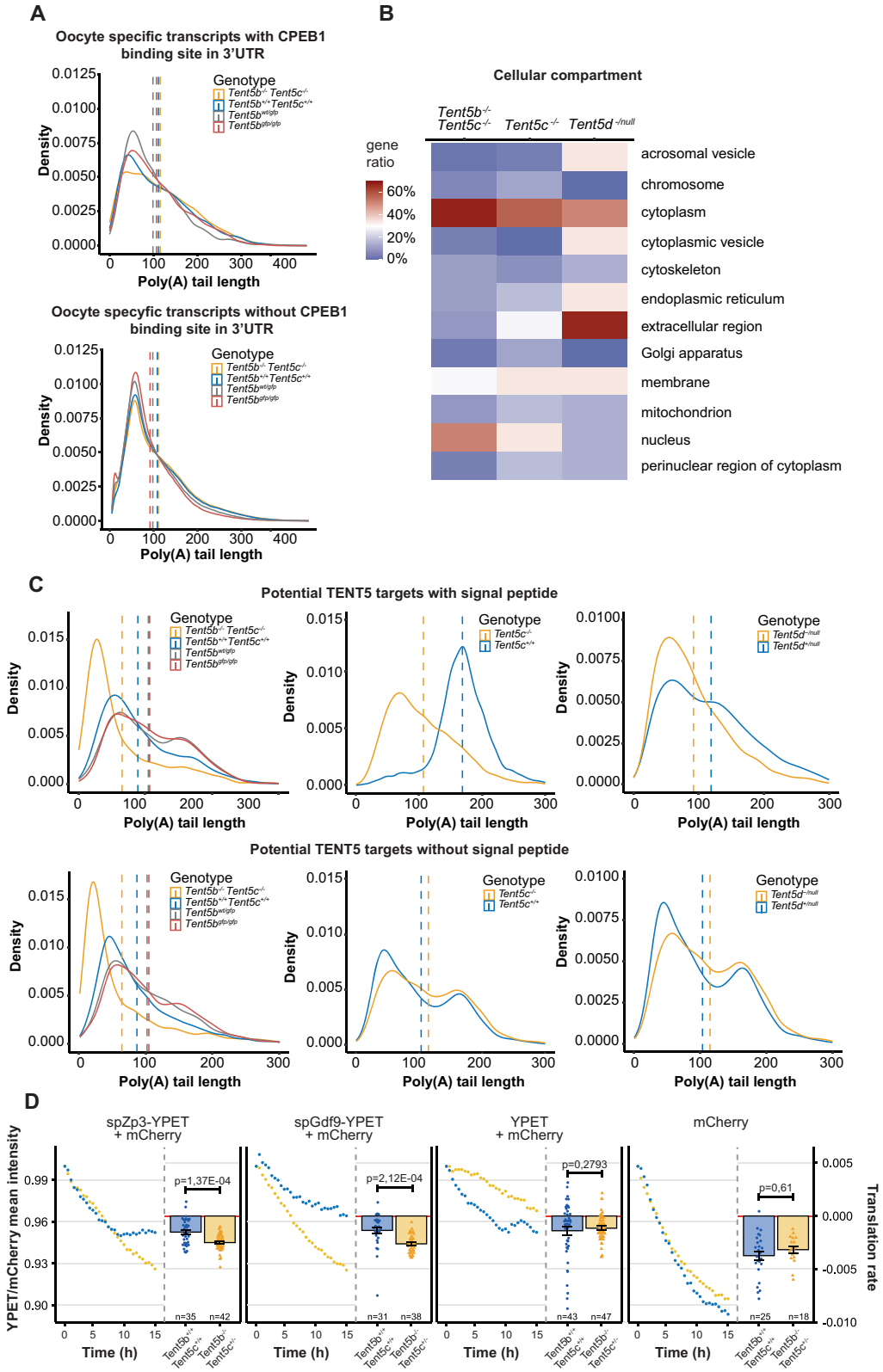
Sex of animals and cells used was determined by the according fertility phenotype connected to particular mutations.

## Mouse line generation

Generation of *Tent5a* KO and *Tent5c* KO mouse lines was described previously[28,29]. Procedure for generation of *Tent5a* KO, *Tent5b* KO, *Tent5d* KO, GFP-*Tent5b, Tent5b*-GFP, *Tent5c*-GFP, *Tent5c*-FLAG, *Tent5d*-GFP, and *Tent5d*-FLAG mouse lines is described below. Chimeric sgRNAs were synthesized by T7 RNA polymerase in-vitro transcription and purified by PAGE. Cas9 mRNA was in-vitro transcribed with T7 RNA polymerase (produced inhouse) and subsequently polyadenylated using *E. coli* poly(A) polymerase (NEB, Cat# M0276L) and m7Gppp5′N Cap was added using Vaccinia Capping System (NEB, Cat# M0280S). For knock-in mouse lines, repair DNA template was purchased as double stranded gene fragments (GeneArt, Thermo Fischer Scientific). See Supplementary Table 1 for sgRNAs, repair templates, and genotyping oligonucleotides sequences.

Zygotes obtained from mated females were microinjected into the cytoplasm using Eppendorf 5242 microinjector (Eppendorf-Netheler-Hinz GmbH) and Eppendorf Femtotips II capillaries with the following CRISPR cocktail: Cas9 mRNA (25 ng/μl), sgRNA (15 ng/μl), and donor dsDNA (7.5 ng/μl). After overnight culture microinjected embryos at 2-cell stage were transferred into the oviducts of 0.5-day p.c. pseudo-pregnant females. Born pups were genotyped by PCR at

https://doi.org/10.1038/s41467-024-49479-4




around 4 weeks. The presence of mutation was confirmed by sequencing in the founder mouse and, after backcrossing, in N1 generation mice. See Supplementary Information for primers' sequences used for genotyping.

### Fertility assessment

Mice were deemed fertile when mating pair produced litters on regular basis within ~4 weeks intervals. Infertile mice never produced litter within span of at least 12 months. Every case of suspected infertility was confirmed by mating homozygous and heterozygous mice with wild-type mates.

### Cells and embryos isolation and culture

To stimulate ovarian follicle growth and ovulation, adult female mice (aged 8–12 weeks) were injected intraperitoneally with pregnant mare's serum gonadotropin (PMSG) and human chorionic

**Fig. 6 | TENT5s enhance the expression of secreted proteins during gametogenesis. A** Differences in poly(A) tail profiles of transcripts containing CPEB1 motif in 3′UTR, compared to the rest of the transcriptome detected by DRS. In ovaries, analyses were performed for 522 oocyte-enriched mRNAs. Two-sided Mann–Whitney–Wilcoxon rank sum test was calculated. Dashed lines indicate the mean poly(A) tail lengths: *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ = 115 nt; *Tent5b*$^{wt/wt}$ = 111 nt; *Tent5b*$^{wt/gfp}$ = 106 nt; *Tent5b*$^{gfp/gfp}$ = 99 nt for transcripts with CPEB1, and *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ = 106 nt; *Tent5b*$^{wt/wt}$ = 104 nt; *Tent5b*$^{wt/gfp}$ = 94 nt; *Tent5b*$^{gfp/gfp}$ = 87 nt for transcripts without CPEB1 **B** Gene ontology (GO) analysis of potential TENT5 protein targets. **C** Distribution of poly(A) tail lengths in potential targets of TENT5 encoding proteins with and without signal peptide (SP). Dashed lines indicate the mean poly(A) tail lengths: *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ = 65 nt; *Tent5b*$^{wt/wt}$ = 94 nt; *Tent5b*$^{wt/gfp}$ = 110 nt; *Tent5b*$^{gfp/gfp}$ = 109 nt; *Tent5c*$^{-/-}$ = 106 nt; *Tent5c*$^{+/+}$ = 168 nt; *Tent5d*$^{-/-}$ = 92 nt; *Tent5d*$^{+/+}$ = 119 nt for substrates with SP, and *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ = 63 nt; *Tent5b*$^{wt/wt}$ = 85 nt; *Tent5b*$^{wt/gfp}$ = 101 nt; *Tent5b*$^{gfp/gfp}$ = 103 nt; *Tent5c*$^{-/-}$ = 103 nt; *Tent5c*$^{+/+}$ = 115 nt; *Tent5d*$^{-/-}$ = 100 nt; *Tent5d*$^{+/+}$ = 109 nt for substrates without SP. P values for TENT5B/C substrates with and without SP were calculated by two-sided Mann–Whitney–Wilcoxon test with Bonferroni Hallberg correction, and all p < 2.2e−12. P values for TENT5C and TENT5D substrates with and without SP were calculated by two-sided Mann–Whitney–Wilcoxon rank sum test, and all p < 2.2e−12 **D** Normalized mean intensity of fluorescence and calculated translation rates of oligoadenylated reporter YPET mRNAs with SP injected into *Tent5b*$^{+/+}$ *Tent5c*$^{+/+}$ and *Tent5b*$^{-/-}$ *Tent5c*$^{+/-}$ GV oocytes. Mean intensity data points represent mean YFP fluorescence intensity normalized to fluorescence of control polyadenylated reporter mRNA encoding mCherry. Individual translation rate data points and n values represent individual oocytes analyzed; bars represent mean values; error bars represent SEM; p values reported for the two-tailed t-test. Source data are provided as a Source Data file.

gonadotropin (hCG), respectively (10 units of hormone in 100 μl of PBS) in 44–46 h interval.

All procedures and culture of germinal vesicle (GV) and metaphase II (MII) stage oocytes as well as zygotes, unless stated otherwise, were performed in pre-warmed M2 or M16 medium (Sigma), additionally supplemented with 150 ng/μl dbcAMP (Sigma) to prevent meiosis resumption of GV prophase I-arrested oocytes when needed. For oocyte isolation, hormonally stimulated females were sacrificed either 48 h after PMSG stimulation for GV oocytes or 16 h after hCG stimulation for MII oocytes. For zygote isolation, hCG stimulated females were mated with males overnight and checked for presence of vaginal plug following day as the insemination indicator and then sacrificed 21–22 h post hCG injection. Ovaries and oviducts were dissected and immediately placed in M2 culture medium. For GV oocytes isolation, cumulus–oocyte complexes were released to medium by follicle puncture and cumulus cells surrounding oocytes were removed by pipetting. For MII oocytes and zygote isolation, oviducts were moved to dish containing warmed hyaluronidase solution (300 μg/ml in M2, Sigma) facilitating cumulus cells detachment, cumulus cells-surrounded oocytes/zygotes were released by ampulla puncturing and pipetted to disperse cumulus cells. Before and between further procedures, denuded oocytes were cultured in 10 μl drops of M2 medium under mineral oil in groups of 10–20 on plastic dishes (35 × 10 Tissue Culture Dishes, Falcon) placed in incubator at 37 °C and 5% $CO_2$ in the air.

For male germ cells isolation from testis, a protocol adapted from the work of Bastos et al.[58] was used. Dissected testes were decapsulated and placed in Falcon with 25 ml of HSBB buffer (Gibco) at room temperature. Next, 1 ml of type XI collagenase (12.5 mg/ml) was added (Sigma) and tubes were put to shake at 32 °C for 20 min in the shaking water bath (120 osc/min). Tubes were manually agitated every 5 min to facilitate the dissociation of the tubules. The tubules were washed once in 25 ml 1× HSBB buffer and resuspended in 10 ml of the HSBB buffer with 200 μl of the stock trypsin (25 mg/ml, Sigma) and 2 μl of the stock DNAse I (5 mg/ml, Sigma) and again put in a water bath with agitation at 32 °C for 15 min (120 osc/min). After the incubation, samples were dissociated for 4 min by pipetting 10 ml serological pipet and centrifuged at 450 × *g* for 3 min. The supernatant was discarded, and cells were resuspended in 1 ml of HSBB with 10% FBS (Gibco) and counted in Thoma chamber. Prepared cells were used for further procedures.

Isolation of Sertoli cells from testis was based on the protocol published by Bhushan et al. and Anway et al.[59,60]. Dissected testes were decapsulated and digested with 10 ml of trypsin (2.5 mg/ml)-DNase I (10 μg/ml) solution (Sigma) for 4 min in water bath at 32 °C (120 oscillations/min). Next, the trypsin digestion was stopped by adding 5 ml of trypsin inhibitor (10 mg/ml) (Sigma), samples were vigorously mixed and incubated for 5 min. After tubules settled down in the tube supernatant was carefully removed and 10 ml of trypsin inhibitor (2.5 mg/ml) was added. Digested tissue was washed nine times to remove all the germ cells. Remaining seminiferous tubules were digested by collagenase (1 mg/ml)–hyaluronidase (1 mg/ml; Sigma)–DNase I (10 μg/ml) mixture followed by hyaluronidase (1 mg/ml)–DNase I (10 μg/ml) digestion. Finally, digested seminiferous tubules were passed ten times through the 18G needle. Isolated Sertoli cells were then seeded in the concentration of $1 \times 10^6$ cells/ml in RPMI medium (Gibco) supplemented in 10% FBS and cultured in 37 °C, 5% $CO_2$ until a monolayer of cells was obtained.

Previously isolated germ cells were seeded in the concentration of $2 \times 10^6$ cells/ml onto the Sertoli monolayer and cultured for 48 h. After 48 h medium containing germ cells were taken, and germ cells were washed with HSBB buffer and prepared for the flow cytometry analysis.

## Morphology, histology, cell visualization

For tissue histology analysis, H&E staining and TUNEL assay tissues were fixed in 10% neutral buffered formalin (Sigma) for 24 h, dehydrated in ethanol dilution series (70% ×1, 95% ×2, 100% ×3, changes every hour with overnight incubation in 4th change of 100% ethanol) and 2 changes of xylene, and incubated in two changes of liquid paraffin followed by paraffin embedding on metal trays using EC 350 Tissue Embedding Center (Myr). Paraffin-embedded tissues were sectioned using semi-automatic microtome (Leica RM2125 RTS) to obtain 10 μm-thick tissue slices transferred on Superfrost Ultra Plus glass slides (Thermo Fisher Scientific) and dried in 37 °C overnight before further processing.

For immunohistochemistry analysis, tissues were fixed in 4% paraformaldehyde (Sigma) in 0.1 M phosphate buffer for 24 h and cryopreserved in two changes of 30% sucrose (Merck) in 0.1 M PB, 24 h each. Tissues were frozen in Killik medium (Bio-Optica), cut into 20 μm-thick sections using cryostat and placed on Superfrost Ultra Plus glass slides. Sections were then stored in 4 °C until further processing.

Oocytes for immunofluorescence were fixed in 4% PFA (Thermo Fisher Scientific), permeabilized with 0.5% Triton X-100 (Sigma), 30 min in RT each, and blocked with 3% BSA (Sigma-Aldritch) in 4 °C overnight.

For nucleus–cytoplasm staining, paraffin-embedded sections on glass slides were first deparaffinized in two changes of xylene (10 min each) and rehydrated in 5 changes of ethanol (100% ×2, 96% × 2, 70% × 1, 2 min each) and distilled water for 2 min. Subsequently, slides were stained in Harris hematoxylin (Kolchem, Poland) for 10 min, rinsed for 2 min in tap water, dipped once in acidic ethanol (1% HCL solution in 70% ethanol), again briefly rinsed in tap water and for 2 min in tap water substitute (2.5 g $NaHCO_3$ and 20 g $MgSo_4 \cdot 7H_2O$ in 1 liter of water) and finally stained in eosin Y (Kolchem, Poland) for 1 min. After that, slides were dehydrated by reversing the rehydration protocol above (with 5 min xylene incubation instead of 10) and sealed using DPX mounting medium (Sigma) and a cover slide and dried overnight.

For apoptosis staining, tissue sections on glass slides were deparaffinized, rehydrated, and sealed as described above for H&E stained and a commercially available TUNEL Assay Kit (Abcam, Cat#

https://doi.org/10.1038/s41467-024-49479-4

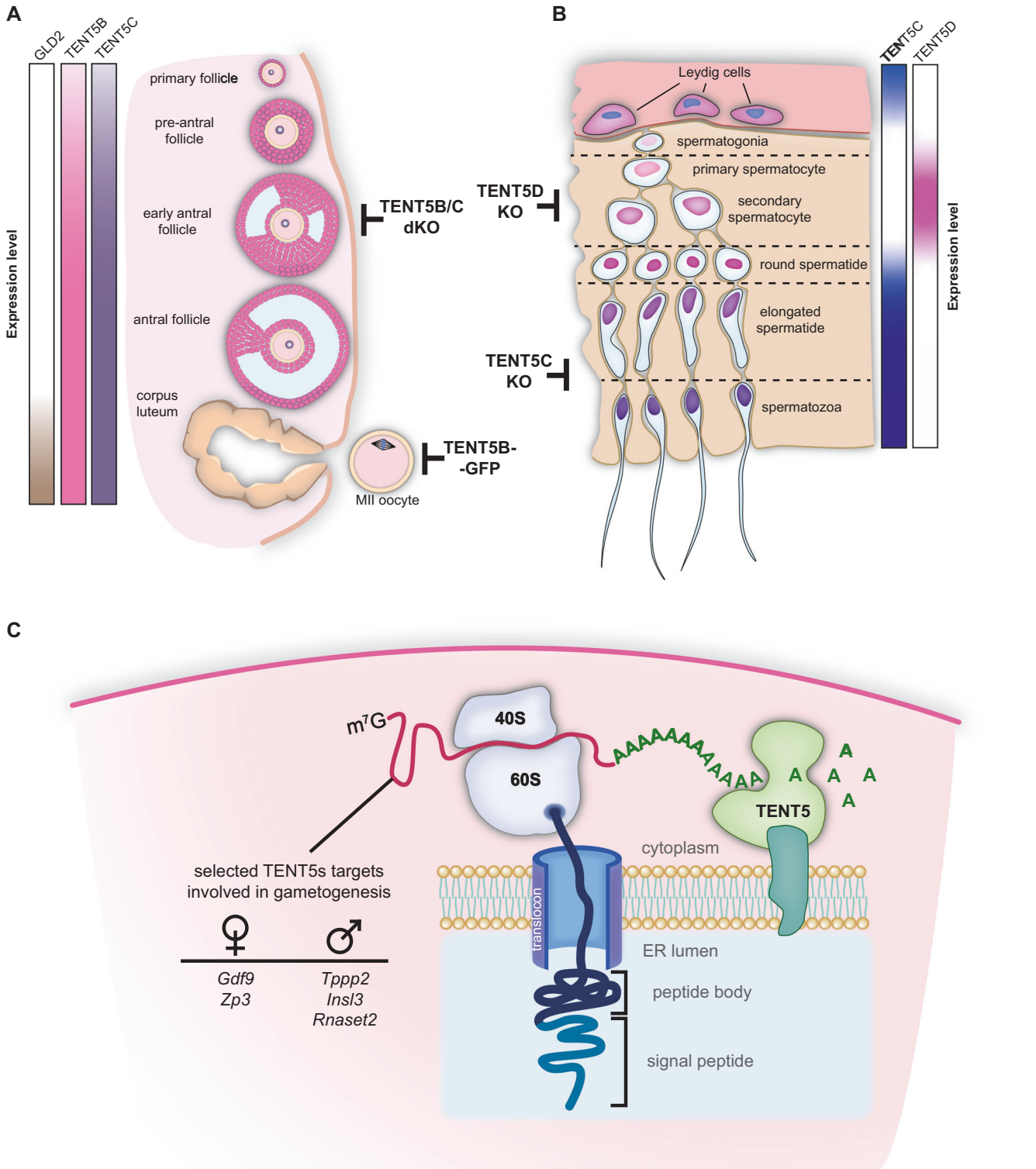



**Fig. 7 | Proposed model for the activity of TENT5 proteins in gametogenesis.** Simplified diagrams of oogenesis (**A**) and spermatogenesis (**B**). Black t-lines with labels indicate stages disrupted by a particular mutation. Gradient boxes indicate approximate expression of *Gld2* and *Tent5b, c*, and *d* genes at different stages of gametogenesis. **C** Proposed model for TENT5-mediated polyadenylation of transcripts encoding secreted proteins. After translation initiation, emerging signal peptide recruits the ribosome–mRNA complex to the ER, where further polyadenylation by TENT5 takes place.

ab206386) was used according to the manufacturer's protocol to detect double stranded breaks of DNA created during apoptosis. Negative control staining was prepared by substituting TdT in reaction mixture with water and positive staining control was prepared by treating sections with 1 μg/μl DNase I in TBS (tris-buffered saline) for 20 min.

MII oocytes used for immunofluorescence were incubated overnight at 4 °C in primary monoclonal antibody against β-tubulin conjugated with FITC (Sigma-Aldritch Cat# F2043, Lot# 108M4766V; 1:50) and then washed twice in PBS for 15 min. To visualize chromatin, oocytes were stained in droplets of propidium iodide (0.01 mg/ml in PBS) at 37.5 °C, for 30 min on glass bottom dishes (MatTek Corp.).

To prepare chromosome spreads, MII oocytes were treated by Acidic Tyrode's Solution (Sigma) for about 30 s at 37 °C to remove zona pellucida. After wash in M2 medium oocytes were placed for 1–5 min in hypotonic solution of 1% sodium citrate. Next, single oocyte was placed on standard glass slide and fixed by several drops of methanol/acetic acid (3:1). Slides were air dried, and chromosomes were stained with 4% Giemsa solution for 10 min. Slides were washed in distillated water, air dried, washed in xylene, and finally mounted using DPX mounting medium.

For immunohistochemistry, frozen sections were dried for 10′ at 55 °C, washed twice for 10′ in TBS, incubated for epitope retrieval in pressure-cooker (70–90 kPa, 160 °C) for 2′ in 20 mM Tris Buffer, pH 9.0 (for INSL3, TPPP2, and ZP3), pH 7.5 (for Rnaset2) or 10 mM citrate buffer, pH 6.0 (for Gdf9), and washed twice briefly with TBS. Next, sections were incubated 1 h at room temperature (RT) in blocking buffer (10% serum from the species that subsequently used secondary antibody was raised in, 1% BSA and 0.3% Triton X-100 diluted in TBS) and then overnight at 4 °C in moisture chamber with primary antibodies (anti-Tppp2 1:100, Abcam, Cat# ab236887; anti-INSL3 1:200, Invitrogen, Cat# PA5-55921 Lot: A96525; anti-RNASET2 1:100, Sigma-Aldrich, Cat# HPA029013, Lot: A115851; anti-GDF9, Abcam, Cat# ab254323, Lot: GR3279299; anti-ZP3, Proteintech, Cat# 21279-1-AP, Lot: 00041356) diluted in the same buffer. Next day slides were washed 3× for 5′ with TBS-T (TBS with 0.025% Triton X-100), incubated 1 h at RT with secondary antibodies (donkey anti-rabbit Alexa 568-conjugated, 1:500, Invitrogen, Cat# A-10042, Lot: 2136776 or for GDF9 staining goat anti-rabbit Alexa 586-conjugated, 1:500, Invitrogen, Cat# A-11011, Lot: 2273773) diluted in blocking buffer with Hoechst 33342 (Sigma, Cat# H3570) 10 μg/ml, washed 3× for 5′ with TBS-T and once with TBS.

To quench autofluorescence all sections except GDF9 staining were incubated with TrueBlack Plus(R) (Biotium) diluted 1× in PBS for 10′, washed 3× for 5′ with PBS. All sections were sealed with cover slide using ProLong Gold Antifade Mountant (Invitrogen).

### Flow cytometry analysis

In the seminiferous tubules of adult mammalians, germ cells in different maturation steps coexist, with 1C (round spermatids, elongating and elongated spermatids, spermatozoa), 2C (several types of G1 spermatogonia, secondary spermatocytes), and 4C (different stages of primary spermatocytes, G2 spermatogonia) DNA content[61]. For the efficient isolation of germ cells, we used protocol established by ref. 58. Considering the various sizes, shapes, and DNA content we employed fluorescence-activated cell sorting (FACS) as a convenient method to analyze cells at different stages of spermatogenesis, resulting in differentiating five populations for downstream analyses: gonial cells (gonia), several stages of primary spermatocytes (4C), secondary spermatocytes (spcII), round spermatids (RS), and elongated spermatids (ES).

For cell cycle analysis germ cells were stained with Vybrant™ DyeCycle™ Violet Stain (Thermo Fisher Scientific, Cat# V35003) 1 μl/1 mln of germ cells[62]. Cells were incubated at 32 °C in a water bath with agitation for 35 min (osc 90/min). Fluorescence was excited by 405 nm laser and DCV Blue fluorescence was detected with 450/50 filters while DCV Red fluorescence was detected with 525/50 filters. For the live/dead staining, the LIVE/DEAD™ Fixable Near-IR Dead Cell Stain Kit (Thermo Fisher Scientific, Cat# L34976) was used. Fluorescence of GFP was excited by 488 nm laser and detected with 530/30 filters. Samples were analyzed with BD LSRFortessa™ and sorted with BD Aria Fusion™ under FACS Diva Software v8.0.1 (BD) software control and analyzed using FlowJo (Data Analysis Software v10)[58,62].

### Imaging

Immunohistochemistry results were scanned as tile arrays on LSM800 (Zeiss) confocal microscope with 20× air objective. Fluorescence intensities were extracted from images using Fiji (Is Just ImageJ)[63] [ImageJ ver. 1.53–1.54] with the help of Grid/Collection Stitching Plugin[64], manually indicated ROIs and self-written macros.

Oocyte live imaging was performed on Opera Phenix High-Content Screening System at 38 °C and 5% $CO_2$ in the air. Oocytes were placed in M16 medium on 384-well plate, one cell per well and were scanned for 15 h in 30 min intervals in brightfield, YFP and mCherry channels with five images gathered in z-axis with 40× water immersion objective every 10 μm with germinal vesicle center as a middle slice. Fluorescence intensity values were gathered from maximum intensity projection of all z-stack images for each oocyte using dedicated Opera Phenix software.

### Western blotting

For western blot analysis, whole tissue or equal amount of cells were lysed with 0.1% NP-40 in PBS supplemented with mix of protease inhibitors and viscolase (A&A Biotechnology) for 30 min at 37 °C with 600 rpm shaking, then Laemmli buffer was added and samples were denatured for 10 min in 100 °C. Samples were separated on 12–15% SDS–PAGE gels, proteins were transferred to Protran nitrocellulose membranes (GE Healthcare), and then membranes were stained with 0.3% w/v Ponceau S in 3% v/v acetic acid to control amount of the protein on the membrane for every sample. Membranes were then incubated with 5% milk or 5% BSA in TBST buffer according to the technical recommendations of the antibodies' suppliers for 1 h followed by incubation with specific primary antibodies diluted 1:1:000 (Flag Tag, Life Technologies, Cat# PA1-984B, Lot: WG319616) or 1:10,000 (GRP 94, Santa Cruz Biotechnology, Cat# sc-11402, Lot: C11616) overnight in 4 °C. Membranes were then washed three times in TBST buffer, incubated with HRP-conjugated secondary antibody: anti-goat (Millipore, Cat# 401393, Lot: 3924034) diluted 1:5000 for 1 h at RT. Membranes were washed three times in TBST buffer and proteins were visualized by enhanced chemiluminescence acquired on X-ray film (Fujifilm) using Clarity Western ECL Substrate (Bio-Rad).

### mRNA reporter oocyte microinjections

Original plasmids carrying mCherry- and YPET reporter-coding sequences, described before[65], were received on courtesy of prof. Marco Conti (UCSF). mCherry sequence-carrying plasmid was used unchanged to produce mCherry reporter mRNA. KDEL (along additional sequences; not used in this work) motif with STOP codon was cloned in ORF downstream of YPET-coding sequence with SLIC method using DNA fragments produced by PCR from mouse oocyte's cDNA (with forward primer containing KDEL motif-coding sequence)[66,67] and signal peptide-coding sequence of Zp3 or Gdf9 were inserted between YPET-coding sequence and T7 reverse transcriptase promoter in ORF by HindIII and BshTI digestion and ligation of oligonucleotides with sticky ends homologous to both enzymes' cutting sites. All plasmids' sequences were verified after cloning by Sanger sequencing. All oligonucleotide sequences are available in Supplementary Table 1.

To synthesize YPET-coding mRNA reporters from prepared plasmids, they were linearized by PCR using Phusion Hot Start II DNA polymerase (Thermo Fisher Scientific) with universal forward primer at YFP sequence and universal reverse primer in Plat 3′UTR sequence with 20 thymine residues overhang (see Supplementary Table 1. for primer sequences). For 50 μl reaction volume 10 ng of output DNA and 0.5 μM final primer concentration was used. The PCR reaction was run in thermal cycler with following program: 98 °C – 3 min, 30 cycles (98 °C – 10 s, 57 °C – 15 s, 72 °C – 30 s), 72 °C – 5 min, 4 °C – hold.

PCR product was run on agarose gel to verify obtained band size and purified using spin-column based Clean-Up kit (A&A Biotechnology).

Plasmid containing mCherry reporter was linearized by restriction digestion with MunI (MfeI) enzyme. For 40 μl of reaction 2.5 μl of DNA

sample was used. Digestion reaction was performed in 37 °C for 2 h and inactivated in 65 °C for 20 min. Digestion product was run on agarose gel to verify obtained band size and purified with 1 volume of AMPure XP (Beckman Coulter) beads and eluted with nuclease-free water.

Resulting DNA fragments where transcribed in-vitro using homemade batch of T7 RNA polymerase with following 40 μl reaction setup for 2 h in 37 °C: 12.5 ng/μl DNA template, 4 μl T-buffer, 10 mM $MgCl_2$, 2.5 mM of ATP, CTP, GTP, and UTP, 1.5 μl Ribolock, and 4 μl of T7 RNA polymerase, $H_2O$. mCherry reporter was additionally polyadenylated in vitro by mixing whole 40 μl of in-vitro transcription reaction with 10 μl of poly(A) buffer (0.5 M Tris-HCl, 2.5 M NaCl), 10 μl 50 nM $MnCl_2$, 4 μl 100 mM ATP, 5 μl 5 U/μl *E.coli* poly(A) polymerase and 331 μl $H_2O$ and incubating reaction in 37 °C for 1 h.

All mRNA reporters were then cleaned by LiCl precipitation. Briefly, reactions were brought up to 400 μl volume, 200 μl of 7 M LiCl, 50 mM EDTA mix were added, and samples were incubated overnight at −20 °C. Then samples were centrifuged for 20 min at 16,000 × *g*, 4 °C, supernatant discarded, and precipitated mRNA washed with 1 ml of 70% ethanol. After centrifuging for 3 min at 14,000 × *g* ethanol was discarded and air-dried pellet resuspended in RNase-free water.

Finally, all mRNA samples were capped using Vaccinia Capping System. 5 μg of mRNA in 30 μl was incubated for 5 min in 65 °C and 5 min in 4 °C before mixing with 4 μl of 10× capping buffer, 2 μl 10 mM GTP, 2 μl 2 mM SAM and 2 μl of Vaccinia Capping Enzyme and incubating for 30 min at 37 °C.

Capped mRNA was cleaned up by LiCl method as described above and resuspended in ~20–50 μl $H_2O$.

For injection, oocytes were isolated and allowed to recover for ~1.5 h. For microinjection, oocytes were transferred into new drops of pre-warmed medium. Microinjections were performed on Axio Observer 5 microscope (Zeiss) equipped with InjectMan 4, TransferMan 4r, CellTram 4r Oil, and FemtoJet 4i (Eppendorf). Mixture of 12.5 ng/μl mCherry mRNA, 12.5 ng/μl YPET reporter, and 0.05% NP-40 (NP-40 in PBS in nuclease-free water) in nuclease-free water was injected into cytoplasm with Femtotip II in minimal possible liquid volume (FemtoJet 4i set up for continuous leak, compensation pressure $p_c$ = 50–150 hPa). After microinjection, oocytes were allowed to recover in culture for ~3 h before imaging.

### RNA isolation

Total RNA from oocytes was isolated using PicoPure RNA isolation kit (Thermo Fischer Scientific, Cat# KIT0204) according to manufacturer's protocol with following adjustments: groups of isolated oocytes were placed in 10 μl of provided extraction buffer in PCR tubes and incubated for 30 min at 42 °C in thermoblock. Samples were then stored in −80 °C. Before proceeding to next step, samples were pooled for total of 30 oocytes per sample and amount of provided ethanol used was adjusted depending on final sample volume. RNA was eluted with 11 μl of elution buffer provided. DNA was removed from final sample by DNase treatment using TURBO DNA-free Kit (Thermo Fisher Scientific, Cat# AM1907) according to manufacturer's protocol. DNA-free RNA was stored in −80 °C.

RNA from whole tissue samples and cultured cells was isolated using TRI Reagent (Sigma) using manufacturer's protocol. Tissues were homogenized in TRI Reagent using glass Dounce homogenizer and cell monolayers were rinsed with TRI Reagent and lysed cells were gathered for further isolation according to manufacturer's protocol. For all RNA samples DNA was removed as described for oocyte samples above.

### cDNA synthesis

cDNA was produced using SuperScript III reverse transcriptase (Thermo Fisher Scientific) according to manufacturer's protocol with oligo(dT) priming and addition of ERCC RNA Spike-InMix (final dilution: 1:100,000; Life Technologies) for first-strand synthesis from oocyte-isolated RNA and RiboLock RNase inhibitor (40 units; Thermo Fischer Scientific) for second strand synthesis in all samples.

Output material was purified in 1 volume of AMPure XP beads, eluted in 20 μl nuclease-free water and second strand synthesis reaction was set up by addition of following components: Second Strand Buffer (NEB), RNase H (10 units; NEB), *E. coli* DNA ligase (10 units; NEB), *E. coli* DNA polymerase (50 units), dNTPs (final conc. 0.4 mM) and nuclease-free water for a final reaction volume of 50 μl. Reaction was incubated at 16 °C overnight. Final cDNA output was purified in 1 volume AMPure XP beads and eluted with 10 μl of nuclease-free water.

### RNA sequencing

Libraries from oocyte RNA for Illumina sequencing were prepared with application of tagmentation reaction, performed following published protocols[68,69] with various steps and amount of enzyme used optimized for use of homemade batch of Tn5 transposase produced in our laboratory. Briefly: 100 μM Tn5ME-A and Tn5ME-B oligonucleotides in annealing buffer (50 mM NaCl, 40 mM Tris-HCl pH = 8) were mixed 1:1 with 100 mM Tn5MErev oligonucleotide and incubated in thermocycler in 95 °C for 5 min and 65 °C for 5 min, and finally cooled to 4 °C and stored at −20 °C (with slow cooling between each step). Tn5 (0.25 mg/ml) was loaded with linker oligonucleotides Tn5ME-A/Tn5MErev and Tn5ME-B/Tn5MErev (see Supplementary Information file for full sequences) by mixing of 10 μl Tn5 with 0.5 μl of both linkers (0.35 μM) and incubating for 45 min in 23 °C with shaking at 350 rpm. Right before the reaction setup loaded Tn5 was diluted 10 times with nuclease-free water. 10 μl of freshly prepared tagmentation buffer (20 mM Tris-HCl pH 7.5; 20 mM $MgCl_2$; 50% dimethylformamide) were mixed with 5 μl of diluted Tn5 and 5 μl od cDNA, incubated for 3 min at 55 °C in preheated thermocycler and cooled to 10 °C for 1 min. Reaction was inactivated by addition of 5 μl 0.2% SDS and incubation for 5 min at RT. Tagmented cDNA was purified in 1.25 volume of AMPure XP beads and eluted in 10 μl of nuclease-free water. For library amplification KAPA HiFi HotStart ReadyMix 2x (Roche, Cat# KK2602) with addition of 5% DMSO was used. 5 μl of tagmented cDNA was used for 15 μl reaction volume. Reaction was run in thermocycler with following program: 72 °C – 15 min, 95 °C – 30 s, 15 cycles: (98 °C – 20 s, 58 °C – 15 s, 72° – 30 s), 72 °C – 3 min, 4 °C – hold. Number of cycles yielding best results was determined experimentally and was further individually adjusted for each batch of prepared cDNA. Libraries were sequenced using Illumina NovaSeq 6000.

Libraries for direct RNA sequencing (DRS) of ovarian and testicular RNA were prepared using Direct RNA sequencing kit (ONT, Cat# SQK-RNA002) following manufacturers protocol, with adjustment to magnetic beads clean-up steps: each time KAPA Pure Beads (Roche) were used and RNA-bead mixture was incubated stationary on bench at RT. To improve sequencing performance and efficiency 90–150 ng of *Saccharomyces cerevisiae* or *Saccharomyces pombe* oligo(dT)-enriched mRNA was added to all samples. Sequencing experiments were performed on MinION device and Flow Cell Type R9.4.1 (ONT, Cat# FLO-MIN106D) with MinKNOW 19.10.1 (MinKNOW core 3.5.5; Bream 4.2.11; GUI 3.5.10) used for data collection, and basecalling with Guppy 6.0.0 (ONT).

### PAT assay

PAT assay on ovarian and testicular RNA was performed using cDNA-PCR Sequencing kit (ONT, Cat# SQK-PCS111) following manufacturer's protocol with following modification allowing for amplification of transcript-specific product: in "Selecting for full-length transcripts by PCR" step, kit-provided cDNA Primer (cPRM) was replaced with 10 mM forward transcript-specific primer and 10 mM universal reverse primer targeting cDNA RT Adapter (CRTA) sequence used in previous steps of the protocol (see Supplementary Table 1 for primer sequences). Amplification was run as described in manufacturers protocol with annealing temperature (57 °C) and extension time (90 s) adjusted to

custom primers and expected product lengths. After amplification PCR products were run on 2% agarose gel and visualized using ChemiDoc Imaging System (Bio-Rad).

### Sequencing data analysis

To figure out poly(A) tail length DRS reads were mapped to the Gencode M26 reference transcript sequences[70] using Minimap 2.17[71] with options -k 14 -ax map-ont –secondary = no and processed with samtools 1.9 (samtools view -b -F 2320)[72] to removed supplementary alignments and reads mapping to reverse strand. All unmapped reads were discarded from analysis. For each read length of poly(A) tail were estimated using the Nanopolish 0.13.2[73] polya function and only reads tagged as PASS were considered in later analyses. Samples from the same condition were analyzed together since they were strongly corelated. Analyses of changes in median poly(A) tail length and mRNA abundance were performed using R. Supplementary Data 2, 6, and 7 files contain the number of counts, mean, median, and geometric mean poly(A) tail lengths. Genes represented by more than 20 reads, and with median poly(A) tail length difference WT:mutant less than −10 nt were selected as potential TENT5BC, TENT5C, and TENT5D targets. We took genes represented by at least 20 reads and with medians calculated for poly(A) tail lengths greater by 10 nt than in the control as potential TENT5B targets in the GFP knock-ins.

Differential expression analysis was performed by mapping reads obtained from Illumina and DRS sequencing to the mouse GRCm39 genome[74] using Minimap 2.17, with options -k 14 -ax splice -uf, features were assigned using Gencode M26. DRS datasets were processed using featureCounts[75] from the subread package in the long read, strand-specific mode (-L -s 1), including only features covered by at least 20% (–fracOverlapFeature 0.2). Illumina datasets were mapped to the mouse GRCm39 genome using STAR[76] on default program settings and processed by featureCounts in the short-read mode (-p –O) including only features covered by at least 20% (–fracOverlapFeature 0.2). R software DESeq2[77] were used to determine differences in mRNA expression levels. The shrinkage approach of DESeq2 was used to implement a regularized logarithm transformation for better visualization and ranking of genes. Supplementary Data 5 contains DESeq statistics for transcripts – Illumina reads dataset.

Motif enrichment analysis was performed independently on three sets of potential TENT5B/C/D targets (Supplementary Data 3, 4, 6, and 7) in comparison to their backgrounds. For TENT5C and TENT5D the background datasets contain the sequences of transcripts identified in the DRS data for WT, represented by more than 20 mapped reads per transcript. For TENT5B/C the background dataset contains transcripts present in WT DRS data assigned to 522 oocyte-specific genes identified based on RNA-seq dataset (Supplementary Data 5). Fasta sequences of 3′ UTRs, exons, and 5′ UTRs of potential TENT5 substrates and their background were obtained using bedtools getfasta tool (v. 2.29.2)[78], using bed files with coordinates downloaded from UCSC Table Browser tool (GENCODE M26 track and known gene table)[79] and GRCm39 genome sequence. Sequence motifs enriched in 3′ UTRs, exons, and 5′ UTRs of TENT5B/C/D potential substrates were identified using the STREME tool[80], run with options --dna --order 4.

CPEB motif identification was performed on the same fasta files like motif enrichment analysis. The motif scanning analysis was performed with the FIMO tool[80] on default settings using published CPEB1 and CPEB2 motif sequences[81]. We obtained groups of transcripts containing and not containing CPEB motifs in TENT5 B/C/D background datasets. We used the same datasets that were used in motif enrichment analysis. Then, we plotted poly(A) tail lengths distributions for those groups. We also plotted poly(A) tail lengths distributions for potential TENT5B/C targets containing and not containing CPEB motifs.

Comparison of the length of structural elements of transcripts identified as potential TENT5 B/C/D targets, were performed on fasta files used in motif enrichment analysis and obtained using sektk comp tool (https://github.com/lh3/seqtk)[82]. Estimation of the statistical significance of differences between the 3′UTR, exons, and 5′UTR lengths of potential TENT5 targets and background were performed using the Mann–Whitney–Wilcoxon test.

Comparison of the GC content of structural elements of transcripts identified as potential TENT5 B/C/D targets, were performed on fasta files used in motif enrichment analysis and obtained using sektk comp tool (https://github.com/lh3/seqtk). Estimation of the statistical significance of differences between the GC contents of potential TENT5 targets and background was performed using the Mann–Whitney–Wilcoxon test.

To identify signal peptide presence in the proteins encoded by potential TENT5 targets, their sequences were downloaded from the UniProt Release 2023_02 database[83] and analyzed for the presence of signal peptides using TargetP Gene ontology analysis[84]. To dataset of potential TENT5D targets we added *Rnase2a*.

Gene ontology analysis was performed using the BioMart R library[82]. For heatmaps we chose 13 more abundant GO-terms assigned to potential Tent5 B/C/D targets.

### Data visualization

All data visualization was performed in R using “ggplot” library[85].

### Translation rate calculation

For every oocyte imagined and analyzed, YPET and mCherry fluorescence values were collected as described above. Oocytes with missing data were discarded altogether from analysis. YPET value was divided by mCherry value for each individual oocyte and timepoint for normalization purposes. For each oocyte, linear regression model was calculated using “lm()” function with “YPET/mCherry intensity ~ timepoint” model in Rstudio software, and beta regression coefficient (describing slopes direction and degree) was used as a translation rate.

### Statistical analysis

All statistical analyses were performed using Rstudio. All results (except sequencing data) were compared using two-sided t-test for normally distributed data, Mann–Whitney–Wilcoxon test for non-normally distributed data (with the exception of follicle size comparison (Fig. 2E) where rank sum test artificially flattens difference between compared genotypes due to small number of significantly bigger follicles in WT ovaries), Fischer exact test for 2 × 2 contingency tables, and Chi-square test for bigger tables. Normality was checked for all datasets by Pearson normality test. P values lower or near assumed significancy value of 0.05 were reported on figures. Full p values, number of samples/specimens analyzed, error bars are reported on figures and fully described in corresponding figure legends.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

Supplementary Figs. 1–9 together with Supplementary Table 1 containing oligonucleotides, dsDNA, and sgRNA sequences, are available in Supplementary Information file. Results of sequencing data analyses are available as Supplementary Data 1–7. All DRS sequencing data generated in this study have been deposited in the European Nucleotide Archive (ENA) database under accession code PRJEB63526. The raw Illumina RNAseq data have been deposited in Gene Expression Omnibus (GEO) database under accession code GSE239661. Additionally, ENA sample accession numbers together with DRS run details are listed in Supplementary Data 1. Any additional information required to reanalyze the data reported in this paper is available from

the corresponding author upon request. Source data are provided with this paper.

## References


1. Kim, J. H. & Richter, J. D. Opposing polymerase-deadenylase activities regulate cytoplasmic polyadenylation. *Mol. Cell* **24**, 173–183 (2006).
2. Kim, K. W., Wilson, T. L. & Kimble, J. GLD-2/RNP-8 cytoplasmic poly(A) polymerase is a broad-spectrum regulator of the oogenesis program. *Proc. Natl Acad. Sci. USA* **107**, 17445–17450 (2010).
3. Sartain, C. V., Cui, J., Meisel, R. P. & Wolfner, M. F. The poly(A) polymerase GLD2 is required for spermatogenesis in *Drosophila melanogaster*. *Development* **138**, 1619–1629 (2011).
4. Wang, L., Eckmann, C. R., Kadyk, L. C., Wickens, M. & Kimble, J. A regulatory cytoplasmic poly(A) polymerase in Caenorhabditis elegans. *Nature* **419**, 312–316 (2002).
5. Kwak, J. E. et al. GLD2 poly(A) polymerase is required for long-term memory. *Proc. Natl Acad. Sci. USA* **105**, 14644–14649 (2008).
6. Udagawa, T. et al. Bidirectional control of mRNA translation and synaptic plasticity by the cytoplasmic polyadenylation complex. *Mol. Cell* **47**, 253–266 (2012).
7. Mansur, F. et al. Gld2-catalyzed 3′ monoadenylation of miRNAs in the hippocampus has no detectable effect on their stability or on animal behavior. *RNA* **22**, 1492–1499 (2016).
8. Chung, C. Z., Jo, D. H. & Heinemann, I. U. Nucleotide specificity of the human terminal nucleotidyltransferase Gld2 (TUT2). *RNA* **22**, 1239–1249 (2016).
9. D'Ambrogio, A., Gu, W., Udagawa, T., Mello, C. C. & Richter, J. D. Specific miRNA stabilization by Gld2-catalyzed monoadenylation. *Cell Rep.* **2**, 1537–1545 (2012).
10. Rouhana, L. et al. Vertebrate GLD2 poly(A) polymerases in the germline and the brain. *RNA* **11**, 1117–1130 (2005).
11. Kim, K. W. et al. Antagonism between GLD-2 binding partners controls gamete sex. *Dev. Cell* **16**, 723–733 (2009).
12. Benoit, P., Papin, C., Kwak, J. E., Wickens, M. & Simonelig, M. PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in *Drosophila*. *Development* **135**, 1969–1979 (2008).
13. Suh, N., Jedamzik, B., Eckmann, C. R., Wickens, M. & Kimble, J. The GLD-2 poly(A) polymerase activates gld-1 mRNA in the *Caenorhabditis elegans* germ line. *Proc. Natl Acad. Sci. USA* **103**, 15108–15112 (2006).
14. Kwak, J. E., Wang, L., Ballantyne, S., Kimble, J. & Wickens, M. Mammalian GLD-2 homologs are poly(A) polymerases. *Proc. Natl Acad. Sci. USA* **101**, 4407–4412 (2004).
15. Barnard, D. C., Ryan, K., Manley, J. L. & Richter, J. D. Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. *Cell* **119**, 641–651 (2004).
16. Nakanishi, T. et al. Possible role of mouse poly(A) polymerase mGLD-2 during oocyte maturation. *Dev. Biol.* **289**, 115–126 (2006).
17. Nakanishi, T. et al. Disruption of mouse poly(A) polymerase mGLD-2 does not alter polyadenylation status in oocytes and somatic cells. *Biochem. Biophys. Res. Commun.* **364**, 14–19 (2007).
18. Tomecki, R., Dmochowska, A., Gewartowski, K., Dziembowski, A. & Stepien, P. P. Identification of a novel human nuclear-encoded mitochondrial poly(A) polymerase. *Nucleic Acids Res.* **32**, 6001–6014 (2004).
19. Sinturel, F. et al. Diurnal oscillations in liver mass and cell size accompany ribosome assembly cycles. *Cell* **169**, 651–663 e614 (2017).
20. Lim, J. et al. Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. *Science*. https://doi.org/10.1126/science.aam5794 (2018).
21. Lubas, M. et al. Interaction profiling identifies the human nuclear exosome targeting complex. *Mol. Cell* **43**, 624–637 (2011).
22. Ivshina, M., Lasko, P. & Richter, J. D. Cytoplasmic polyadenylation element binding proteins in development, health, and disease. *Annu. Rev. Cell Dev. Biol.* **30**, 393–415 (2014).
23. Racki, W. J. & Richter, J. D. CPEB controls oocyte growth and follicle development in the mouse. *Development* **133**, 4527–4537 (2006).
24. Alarcon, J. M. et al. Selective modulation of some forms of schaffer collateral-CA1 synaptic plasticity in mice with a disruption of the CPEB-1 gene. *Learn. Mem.* **11**, 318–327 (2004).
25. Hu, W., Yuan, B. & Lodish, H. F. Cpeb4-mediated translational regulatory circuitry controls terminal erythroid differentiation. *Dev. Cell* **30**, 660–672 (2014).
26. Mroczek, S. et al. The non-canonical poly(A) polymerase FAM46C acts as an onco-suppressor in multiple myeloma. *Nat. Commun.* **8**, 619 (2017).
27. Kuchta, K. et al. FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases. *Nucleic Acids Res.* **44**, 3534–3548 (2016).
28. Gewartowska, O. et al. Cytoplasmic polyadenylation by TENT5A is required for proper bone formation. *Cell Rep.* **35**, 109015 (2021).
29. Bilska, A. et al. Immunoglobulin expression and the humoral immune response is regulated by the non-canonical poly(A) polymerase TENT5C. *Nat. Commun.* **11**, 2032 (2020).
30. Liudkovska, V. et al. TENT5 cytoplasmic noncanonical poly(A) polymerases regulate the innate immune response in animals. *Sci. Adv.* **8**, eadd9468 (2022).
31. Zheng, C. W. et al. Non-canonical RNA polyadenylation polymerase FAM46C is essential for fastening sperm head and flagellum in mice. *Biol. Reprod.* **100**, 1673–1685 (2019).
32. Cong, J. et al. Deficiency of X-linked TENT5D causes male infertility by disrupting the mRNA stability during spermatogenesis. *Cell Discov.* **8**, 23 (2022).
33. Sha, Y. et al. TENT5D disruption causes oligoasthenoteratozoospermia and male infertility. *Andrology* **11**, 1121–1131 (2023).
34. Zhang, Y. T. et al. Novel variations in TENT5D lead to teratozoospermia in infertile patients. *Andrology*. https://doi.org/10.1111/andr.13589 (2024).
35. Kashiwabara, S. et al. Regulation of spermatogenesis by testis-specific, cytoplasmic poly(A) polymerase TPAP. *Science* **298**, 1999–2002 (2002).
36. Rankin, T. et al. Mice homozygous for an insertional mutation in the Zp3 gene lack a zona pellucida and are infertile. *Development* **122**, 2903–2910 (1996).
37. Liu, C. et al. Targeted disruption of the mZP3 gene results in production of eggs lacking a zona pellucida and infertility in female mice. *Proc. Natl Acad. Sci. USA* **93**, 5431–5436 (1996).
38. Chen, T. et al. A recurrent missense mutation in ZP3 causes empty follicle syndrome and female infertility. *Am. J. Hum. Genet.* **101**, 459–465 (2017).
39. Zhang, D. et al. A novel mutation in ZP3 causes empty follicle syndrome and abnormal zona pellucida formation. *J. Assist. Reprod. Genet.* **38**, 251–259 (2021).
40. McGrath, S. A., Esquela, A. F. & Lee, S. J. Oocyte-specific expression of growth/differentiation factor-9. *Mol. Endocrinol.* **9**, 131–136 (1995).
41. Gilchrist, R. B. et al. Molecular basis of oocyte-paracrine signalling that promotes granulosa cell proliferation. *J. Cell Sci.* **119**, 3811–3821 (2006).
42. Dragovic, R. A. et al. Role of oocyte-secreted growth differentiation factor 9 in the regulation of mouse cumulus expansion. *Endocrinology* **146**, 2798–2806 (2005).
43. Dong, J. W. et al. Growth differentiation factor-9 is required during early ovarian folliculogenesis. *Nature* **383**, 531–535 (1996).
44. Carabatsos, M. J., Elvin, J., Matzuk, M. M. & Albertini, D. F. Characterization of oocyte and follicle development in growth differentiation factor-9-deficient mice. *Dev. Biol.* **204**, 373–384 (1998).

45. Szulc, N. A. et al. DEGRONOPEDIA – a web server for proteome-wide inspection of degrons. *Nucleic Acids Res*. https://doi.org/10.1101/2022.05.19.492622 (2022).
46. Zimmermann, S., Schottler, P., Engel, W. & Adham, I. M. Mouse Leydig insulin-like (Ley I-L) gene: structure and expression during testis and ovary development. *Mol. Reprod. Dev*. **47**, 30–38 (1997).
47. Kubota, Y. et al. Leydig insulin-like hormone, gubernacular development and testicular descent. *J. Urol*. **165**, 1673–1675 (2001).
48. Huang, Z., Rivas, B. & Agoulnik, A. I. Insulin-like 3 signaling is important for testicular descent but dispensable for spermatogenesis and germ cell survival in adult mice. *Biol. Reprod*. **87**, 143 (2012).
49. Sagata, D. et al. The insulin-like factor 3 (INSL3)-receptor (RXFP2) network functions as a germ cell survival/anti-apoptotic factor in boar testes. *Endocrinology* **156**, 1523–1539 (2015).
50. Ostendorf, T. et al. Immune sensing of synthetic, bacterial, and protozoan RNA by toll-like receptor 8 requires coordinated processing by RNase T2 and RNase 2. *Immunity* **52**, 591–605 (2020).
51. Acquati, F. et al. Tumor and metastasis suppression by the human RNASET2 gene. *Int. J. Oncol*. **26**, 1159–1168 (2005).
52. Xu, Y. L. et al. RNASET2 impairs the sperm motility via PKA/PI3K/calcium signal pathways. *Reproduction* **155**, 383–392 (2018).
53. Liu, Y. et al. RNASET2 in human spermatozoa and seminal plasma: a novel relevant indicator for asthenozoospermia. *Andrology* **1**, 75–84 (2013).
54. Kettwig, M. et al. Interferon-driven brain phenotype in a mouse model of RNaseT2 deficient leukoencephalopathy. *Nat. Commun*. **12**, 6530 (2021).
55. Krawczyk, P. S. et al. SARS-CoV-2 mRNA vaccine is re-adenylated in vivo, enhancing antigen production and immune response. Preprint at *bioRxiv* https://doi.org/10.1101/2022.12.01.518149 (2022).
56. Fucci, C. et al. The interaction of the tumor suppressor FAM46C with p62 and FNDC3 proteins integrates protein and secretory homeostasis. *Cell Rep*. **32**, 108162 (2020).
57. Xiang, K. H. & Bartel, D. P. The molecular basis of coupling between poly(A)-tail length and translational efficiency. *Elife* **10**. https://doi.org/10.7554/eLife.66493 (2021).
58. Bastos, H. et al. Flow cytometric characterization of viable meiotic and postmeiotic cells by Hoechst 33342 in mouse spermatogenesis. *Cytom. A* **65a**, 40–49 (2005).
59. Bhushan, S. et al. Isolation of Sertoli cells and peritubular cells from rat testes. *J. Vis. Exp*. e53389. https://doi.org/10.3791/53389 (2016).
60. Anway, M. D., Folmer, J., Wright, W. W. & Zirkin, B. R. Isolation of Sertoli cells from adult rat testes: an approach to ex vivo studies of Sertoli cell function. *Biol. Reprod*. **68**, 996–1002 (2003).
61. da Cruz, I. et al. Transcriptome analysis of highly purified mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to postmeiotic-related processes at pachytene stage. *BMC Genomics* **17**, 294 (2016).
62. Hayama, T. et al. Practical selection methods for rat and mouse round spermatids without DNA staining by flow cytometric cell sorting. *Mol. Reprod. Dev*. **83**, 488–496 (2016).
63. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
64. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25**, 1463–1465 (2009).
65. Luong, X. G., Daldello, E. M., Rajkovic, G., Yang, C. R. & Conti, M. Genome-wide analysis reveals a switch in the translational program upon oocyte meiotic resumption. *Nucleic Acids Res*. **48**, 3257–3276 (2020).
66. Jeong, J. Y. et al. One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. *Appl. Environ. Microb*. **78**, 5440–5443 (2012).
67. Li, M. Z. & Elledge, S. J. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol. Biol*. **852**, 51–59 (2012).
68. Hennig, B. P. et al. Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. *G3* **8**, 79–89 (2018).
69. Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res*. **24**, 2033–2040 (2014).
70. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. **47**, D766–D773 (2019).
71. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574 (2021).
72. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**. https://doi.org/10.1093/gigascience/giab008 (2021).
73. Workman, R. E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).
74. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. **44**, D733–D745 (2016).
75. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
76. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
77. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15**. https://doi.org/10.1186/s13059-014-0550-8 (2014).
78. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
79. Karolchik, D. et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*. **32**, D493–D496 (2004).
80. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME suite. *Nucleic Acids Res*. **43**, W39–W49 (2015).
81. Duran-Arque, B. et al. Comparative analyses of vertebrate CPEB proteins define two subfamilies with coordinated yet distinct functions in post-transcriptional gene regulation. *Genome Biol*. **23**, 192 (2022).
82. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc*. **4**, 1184–1191 (2009).
83. Bateman, A. et al. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*. https://doi.org/10.1093/nar/gkac1052 (2022).
84. Armenteros, J. J. A. et al. Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* **2**. https://doi.org/10.26508/lsa.201900429 (2019).
85. Wickham, H. Data Analysis. In: ggplot2. Use R!. (Springer, Cham, 2016). https://doi.org/10.1007/978-3-319-24277-4_9.


## Acknowledgements

We thank the members of the Andrzej Dziembowski lab for their support, prof. Marco Conti (UCSF) for sharing YPET and mCherry mRNA reporter-carrying plasmids, Aleksandra Brouze for critical reading and proof-reading the manuscript, and all members of the Genome Engineering Unit of the International Institute of Molecular and Cell Biology for maintaining the animal colony and for animal genotyping. NGS was performed thanks to the Genomics Core Facility CeNT UW (RRID:SCR_022718), using the NovaSeq 6000 platform financed by the Polish Ministry of Science and Higher Education (decision no. 6817/IA/SP/2018 of 2018-04-10). This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 810425 (AD). The research leading to these results was funded by the Norwegian Financial Mechanism 2014–2021, No. UMO-2019/34/H/NZ3/00733 (AD). This project was

also funded by the National Science Centre, Poland, Grant No. 2019/33/B/NZ2/01773 (AD).

## Author contributions

M.B. and A.D. wrote the final version of the manuscript with contributions from A.C.-C., O.G., and M.K.-K. M.B. measured TENT5B and TENT5C expression in oocytes, performed phenotype analysis of *Tent5a* KO, *Tent5b/c* dKO, and *Tent5b*$^{GFP}$ mice, GDF9 and ZP3 immunohistochemistry staining, oocyte mRNA reporter injection, prepared RNA sequencing libraries for oocytes and ovaries, and performed PAT assays. M.B. and M.S. performed *Tent5a* KO oocyte maturation. O.G. performed blood analysis, body weight measurements, and *Tent5c* and *Tent5d* phenotype analysis. M.K.-K. performed co-cultures of germ cells and Sertoli cells, all flow cytometry experiments, and cell sorting. M.S. performed *Tent5b*$^{GFP}$ mating experiments and chromatin analysis experiments. B.T. and K.J. performed all immunohistochemistry experiments in testes. S.M. prepared RNA sequencing libraries for sperm and testes. J.G. and E.B. established all mouse lines. M.B. reviewed and analyzed all gathered data except sequencing results. A.C.-C. supported by P.K. analyzed all RNA sequencing results. M.B. and A.C.-C. prepared the figures.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-49479-4.

**Correspondence** and requests for materials should be addressed to Andrzej Dziembowski.

**Peer review information** *Nature Communications* thanks Falong Lu, Robin Hobbs and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Manuscript 2. Czarnocka-Cieciura A., Brouze M., Gumińska N., Mroczek S., Gewartowska O., Krawczyk P., Dziembowski A. *Comprehensive analysis of poly(A) tails in mouse testes and ovaries using Nanopore Direct RNA Sequencing*. Scientific Data, accepted, in press (2024)**

# Title

Comprehensive analysis of poly(A) tails in mouse testes and ovaries using Nanopore Direct RNA Sequencing.

# Authors

Agnieszka Czarnocka-Cieciura[1,#], Michał Brouze[1,#], Natalia Gumińska[1,#], Seweryn Mroczek[2,1], Olga Gewartowska[3], Paweł S. Krawczyk[1] and Andrzej Dziembowski[1,4,*]

**Affiliations**

1. Laboratory of RNA Biology, International Institute of Molecular and Cell Biology, Warsaw, 02-109, Poland

2. Institute of Genetics and Biotechnology, Faculty of Biology, University of Warsaw, Warsaw, 02-106, Poland

3. Genome Engineering Facility, International Institute of Molecular and Cell Biology, Warsaw, 02-109, Poland

4. Department of Embryology, Institute of Developmental Biology and Biomedical Sciences, Faculty of Biology, University of Warsaw, Warsaw, 02-096, Poland

# equally contributed

corresponding author: Andrzej Dziembowski (adziembowski@iimcb.gov.pl)

# Abstract

Gametogenesis is a process in which dysfunctions lead to infertility, a growing health and social problem worldwide. In both spermatogenesis and oogenesis, post-transcriptional gene expression regulation is crucial. Essentially, all mRNAs possess non-templated poly(A) tails, whose composition and dynamics (elongation, shortening, and modifications) determine the fate of mRNA. Moreover, gametogenesis, especially oogenesis, represents a unique instance of the complexity of poly(A) tails metabolism, with oocyte-specific waves of cytoplasmic polyadenylation. In this context, we provide a comprehensive transcriptomic dataset focusing on mRNA poly(A) tail composition and dynamics in murine testes and ovaries. It consists of RNA samples isolated from wild-type and transgenic mice lacking TENT5 polymerases, which can extend poly(A) tails in the cytoplasm. TENT5 deficiencies have serious consequences. For instance, the defect of TENT5D causes infertility in humans. The data described here are generated mainly using the Oxford Nanopore Direct RNA Sequencing (DRS) method, which provides ground-truth information about mRNA molecules, including poly(A) tail length and

nucleotide content. For instance, we show the prevalence of uridilated tails in testicular mRNAs.

## Background & Summary

Development and maturation of germ cells constitute a multi-stage process governed by stringent gene expression and mRNA quality control mechanisms. During both oogenesis and spermatogenesis, diploid cells undergo meiotic division, leading to the generation of haploid gametes. Notably, the final stages of these processes occur when transcription is repressed, and regulation of gene expression relies primarily on the post-transcriptional mechanisms. In this context, the metabolism of mRNA poly(A) tails is critical. This poly(A) tail-dependent regulatory pathway involves the activity of various enzymes acting antagonistically at the 3'-end of the mRNA. Deadenylases operating both in the nucleus and the cytoplasm contribute to the shortening of poly(A) tails, while poly(A) polymerases extend them. Interestingly, cytoplasmic polyadenylation is particularly important for germ cell development. Mutations in *Tent5b*, *Tent5c*, and *Tent5d* genes encoding cytoplasmic poly(A) polymerases can affect both spermatogenesis and oogenesis (Figure 1), as we uncovered in our recent work[1]. We described a notable wave of massive cytoplasmic polyadenylation in oogenesis and showed that double knock-out (KO) of *Tent5b/Tent5c* led to early arrest in oocyte development. Additionally, we detailed the effects of *Tent5c* and *Tent5d* KO on groups of transcripts critical to the progression of spermatogenesis. The importance of TENT5D for spermatogenesis is also highlighted by the fact that mutations in *TENT5D* lead to infertility in humans[1-4]. Growing evidence suggests that, in addition to the length of poly(A) tails, their nucleotide composition also affects mRNA stability[5,6]. Notably, the addition of uridines, termed uridylation, has been proven significant in gametogenesis. Conditional knock-outs of poly(U) polymerases *Tent3a* and *Tent3b* in mice have demonstrated their essential role, leading to infertility in both males and females [7,8].

*# Figure 1 should be inserted here.*

Despite the well-established importance of poly(A) tails, their methods of analysis, especially on the whole transcriptome scale, have remained imperfect and largely unchanged for decades. Poly(A) tails, like other homopolymer tracts, are prone to becoming unstable during amplification required in classical sequencing approaches[9,10]. This limitation can be addressed by using the amplification-independent Oxford Nanopore Direct RNA Sequencing (DRS) method. To date, DRS is the only technique that provides ground-truth information about poly(A) tails[9,11-16]. Furthermore, long DRS reads enable investigation of many other post-

transcriptional gene expression regulatory mechanisms, such as alternative splicing and epitranscriptomic modifications. In nanopore sequencing, single-stranded nucleic acid molecules (RNA in the case of DRS) are passed through the protein pore due to applied voltage. This causes subsequent changes in the electric current, which are first recorded as a raw signal and then computationally converted to the corresponding nucleotide sequence. Based on the characteristic patterns (signatures) in the current intensity over time, poly(A) tails can be localized, and their lengths can be precisely estimated[9,12,13]. While the influence of poly(A) tail length on gene expression has been widely studied, the role of non-adenosine (non-A) residues within poly(A) tails remains poorly understood[5,6,9,17]. To fill this gap, we developed the Ninetails, a neural network-based algorithm, for the analysis of the nucleotide composition of poly(A) tails in DRS data (https://github.com/LRB-IIMCB/ninetails). Based on the properties of the raw signal of poly(A) tail region identified by other tools, our pipeline recognizes and quantifies non-As, such as cytosines, guanosines, and uridines, with high precision and recall.

Here, we present poly(A) tail composition DRS-based data derived from testes and ovaries of wild-type (WT) mice and animals with constitutive KO mutations of *Tent5* cytoplasmic poly(A) polymerases. These DRS datasets were initially described in our previous report focusing on poly(A) tail length profiling[1]. In this work, we augment them with the differential expression analysis and outputs of Ninetails software, providing a resource for detailed profiling of non-adenosine nucleotides in poly(A) tails of mRNAs from ovaries and testes without a cDNA proxy (Figure 2). Importantly, our data enable orthogonal validation of previously published findings on the nucleotide composition of poly(A) tails in germline cells obtained with PacBio-based protocols[6,18]. The analysis of the non-A residues in poly(A) tails revealed an interesting observation that uridines are remarkably prevalent near the end of the tails in testes (Figure 3), which ought to prompt further research. To gain more insight, we supplemented DRS with Illumina-based RNA-seq data obtained from a sorted cell population isolated from WT and *Tent5c* KO mice testes. These datasets supply information about the differential expression of genes involved in the maturation of male germ cells. Notably, a comparison of data on poly(A) tail composition with transcriptome changes occurring during differentiation revealed that in the transition from early postmeiotic cells (round spermatids) to later spermiogenesis stages (elongated spermatids), there is an overrepresentation of mRNAs with poly (A) tails rich in uridines among the downregulated mRNAs.

In general, the datasets provided will be useful to the broad scientific community interested in mammalian gametogenesis.

*# Figure 2 should be inserted here.*

# Methods

## Experimental animal models

All material used for producing presented data was obtained from mice lines previously generated using CRISPR/Cas9 method in C57BL/6/Tar x CBA/Tar mixed background and described elsewhere[1,15,16].

All experiments on animals were approved by the Local Ethical Committees at University of Warsaw, Faculty of Biology (approval numbers: 176/2026, 917/2019) and Warsaw University of Life Sciences, Faculty of Horticulture and Biotechnology (approval number: WAW2/049/2022)

For Illumina platform RNA sequencing, organs from following animals were used: 5-week-old *Tent5b*$^{+/+}$/*Tent5c*$^{+/,+}$ and *Tent5b*$^{-/-}$/*Tent5c*$^{-/-}$ females; 8-week-old *Tent5c*$^{+/+}$ and *Tent5c*$^{-/-}$ males.

For Direct RNA sequencing on MinION platform, organs from following animals were used: 30-days-old *Tent5b*$^{+/+}$/*Tent5c*$^{+/+}$, *Tent5b*$^{-/-}$/*Tent5c*$^{-/-}$, *Tent5b*$^{wt/gfp}$ and *Tent5b*$^{gfp/gfp}$ females; 8-week-old *Tent5c*$^{+/+}$ and *Tent5c*$^{-/-}$ males; 3-week-old *Tent5d*$^{+/null}$ and *Tent5d*$^{-/null}$ males.

## Germ cell isolation

Isolation of round and elongated spermatides from testes was performed as described previously[1].

## RNA isolation

Total RNA from whole ovaries, testes and collected spermatides was isolated with TRI Reagent (Sigma) following manufacturer's protocol. Tissues were homogenized in TRI Reagent using glass Dounce homogenizer while collected cell pellets were suspended in TRI Reagent and lysed by pipetting. DNA traces from final RNA samples were removed using TURBO DNA free kit (Thermo Fisher Scientific) before further experiments.

## Library preparation and sequencing

For spermatids and whole testes RNA sequencing, rRNAs were removed from total DNA-free RNA samples with Ribo-Zero Gold rRNA-removal Kit (Illumina). rRNA-free samples were cleaned up by the 3M sodium acetate precipitation method. KAPA Stranded RNA-Seq Library Preparation Kit (KAPA Biosystems) was used for sequencing libraries preparation, with random primers used for cDNA synthesis step. At each step of the procedure, samples were cleaned up using AMPure XP Reagent magnetic beads. Quality and fragment size distribution in the libraries were determined by electrophoresis on Agilent 2100 Bioanalyzer (Agilent

Technologies Inc.). Libraries for Direct RNA Sequencing (DRS) of ovarian and testicular RNA were prepared using the Direct RNA Sequencing kit (ONT, Cat# SQK-RNA002).

To improve the quality and efficiency of sequencing, we added 90-150 ng of *Saccharomyces cerevisiae* or *Schizosaccharomyces pombe* oligo(dT)-enriched mRNA to all samples as a carrier instead of RNA Control Strand supplied in the kit. This step was crucial for achieving desired sequencing stoichiometry without amplification, as the libraries obtained for ovaries and testes were low-input. Sequencing experiments were run on the MinION device loaded with Flow Cell Type R9.4.1 (ONT, Cat# FLO-MIN106D), controlled by the MinKNOW software (ONT).

**Preliminary sequencing data analysis**

All raw DRS data[19-21] were basecalled with Guppy 6.0.0 and mapped to the Gencode M26 reference transcript sequences[22] using Minimap2 2.17[23] with options -k 14 -ax map-ont –secondary=no. To remove supplementary alignments and reads mapping to reverse strand, we processed all datasets with samtools 1.9 (samtools view -b -F 2320)[24] and discarded all unmapped reads. Next, poly(A) tail lengths were estimated using the Nanopolish 0.13.2[12] polya function. The output of this software includes a quality tag indicating the validity of each tail estimation, which assigns reads to one of five categories. For most downstream analyses, we recommend using reads tagged as 'PASS' and 'SUFFCLIP' since these are technically correct and correspond to the reference. The exploratory data analysis was performed using NanoTail package[25].

In addition to the data detailing the composition of poly(A) tails, we also present differential expression data[26], which were not included in our previous work focused on differential adenylation. These results were obtained as described by us previously[1].

Illumina datasets[27] were mapped to the mouse GRCm39 genome using STAR[28] with default settings. Aligned reads were summarized using featureCounts [29] in the short-read mode (-p –O) and including only sequences covered by at least 50% of read (–fracOverlapFeature 0.5). Differential expression analysis was performed using DESeq2[30]. To reduce the background affecting statistical inference and data visualization (i.e., to alleviate potential high levels of variability in lowly expressed genes), $\log_2$ fold changes were shrunken. The tables containing DESeq2 statistics calculated for Illumina reads were deposited in Gene Expression Omnibus (GEO) repository along with the raw data[27].

**Poly(A) tail content analysis**

We conducted an analysis of raw DRS data[19-21] obtained from ovaries and testes to determine the occurrence of non-adenosines within poly(A) tails employing the Ninetails tool (see Code

Availability) with the following parameters: qc=TRUE, pass_only=FALSE. The outputs of the Ninetails poly(A) surveillance pipeline in tsv format are provided[31,32].

## Data record

Here we present previously unpublished RNA-seq data has been deposited in Gene Expression Omnibus (GEO) repository under the accession number: GSE273264[27] along with tables containing raw counts for each sample and results of differential expression analysis[27].

The Table 1 with metadata of all RNA-seq and DRS samples has been uploaded to Figshare[33]. This table comprises 11 columns. The "sample_name" column contains the unique identifier of an independently sequenced sample, consistent across all published resources (e.g., in the corresponding differential expression analyses). The subsequent columns include the replicate number, species name (*Mus musculus*), organ (ovary/testis), and sequencing platform (ONT(DRS)/Illumina). These columns are particularly useful for clustering the data. The "project" and "accession" columns contain the unique identifiers of the projects and sequencing data records in the repositories. The "seq_data_link" column provides a direct hyperlink to each sequencing dataset (sample), while the "seq_source" column indicates whether and where the sequencing data has been previously published or if it is the original data presented herein. The last two columns, "nonA_data" and "diffexp_data", indicate whether the corresponding nucleotide composition of poly(A) tails and differential expression data are provided in this work. In these columns, "Y" denotes the presence of the data, and "n" denotes its absence. The metadata table allows to seamlessly integrate the data from our previous work with the data record presented here.

The raw Ninetails results in tsv format are available on Figshare[31,32]. For each sample analyzed, Ninetails generates two output tables. Files with the suffix “class_data”[32] contain the outcomes of read classification, with each row representing data for a single read. These files indicate whether the poly(A) tails of individual reads meet the quality criteria and whether they are decorated with non-adenosine residues or blank. Files with the suffix “residue_data”[31] store information on specific non-adenosine occurrences. These files detail the type of non-adenosine found at a given position in a given tail and its length, with each row representing a single non-adenosine instance. A detailed explanation of the Ninetails output is available on Wiki (https://github.com/LRB-IIMCB/ninetails/wiki). The Ninetails output files are named according to the metadata table. These output data from Ninetails main analysis module (“check_tails” function) can then be transformed by post-processing functions (e.g., summarized, converted between wide and long format) and visualized with

functions from the Ninetails graphics module. All these functionalities are covered in detail in the built-in help files for Ninetails and in the wiki, under Code Availability.

In tabular data, we avoid spaces and special characters, which facilitates further analysis using tools written in various programming languages.

## Technical Validation

For the 19 DRS source datasets [19-21], the number of reads fluctuates between the samples (Figure 3a), which is dependent on many determinants. One such factor is the amount of the starting material, since the DRS technology does not involve amplification. Notably, this does not affect the overall transcriptome overview, as the enrichment ratios of individual transcripts remain consistent between biological replicates (Figure 3b). The distributions of read quality tags are similar across different replicates (Figure 3c).

*# Figure 3 should be inserted here.*

**Poly(A) tail length and composition**

To determine the poly(A) tail composition of 19 DRS datasets provided[19-21], reads were categorized into five classes based on their quality and their adherence to the criteria required for identifying non-A residues in poly(A) tail. For subsequent analysis, only reads exhibiting a significant change between consecutive k-mers (so-called “move”) and displaying a characteristic anomaly in the nanopore raw signal were considered (Figure 4a). Utilizing a neural network, the Ninetails program assigns the shape of the signal anomaly within a poly(A) tail to one of three categories, each representing anomalies corresponding to the presence of cytidine, guanosine, or uridine nucleotides. Next, we counted the frequency of each non-adenosine nucleotide in all experimental datasets (Figure 4b) and for 522 oocyte-specific genes described in our previous work[1] (Figure 4c). The biological replicates were highly consistent both in terms of the size of the heterogeneous poly(A) tails fraction and in terms of their distribution among classes representing the respective non-adenosine nucleotides.

We observed no significant differences in the distribution of non-A nucleotides between the heterogeneous tails of oocyte-specific genes and all identified ovarian genes (Figure 4b and c). In samples obtained from murine testes, we identified an overrepresentation of uridines within the heterogeneous poly(A) tails (Figure 4b). We further investigated the positions of non-adenosines within the poly(A) tails. As a result, we identified a subset of transcripts in which non-A residues predominantly occurred near the 3’-terminus of poly(A) tails in most reads (see Figure 4d). These transcripts (e.g. *Spem1 and Tdrd5*) have been shown to be crucial

to the progression of spermatogenesis[34-37]. Interestingly, the *Insl3* transcript, whose expression is necessary for the proper development of the urogenital tract and female fertility, has no enrichment in uridine at the end of poly(A) tails. The distribution of data points in the Figure 4e panel is scattered, indicating a random insertion of uridines in the poly(A) tails. A similar pattern is seen for the *Rpl9* transcript (Figure 4f), encoding a ubiquitous protein not involved in gamete differentiation. Thus, we suggest that the presence of non-adenosine nucleotides at the ends of their poly(A) tails is not a coincidence but rather a factor that regulates the expression of these mRNAs.

*# Figure 4 should be inserted here.*

For testicular transcriptome analysis, RNA was isolated from *Tent5c*$^{-/-}$ and *Tent5c*$^{+/+}$ males. The testes of the *Tent5c*$^{-/-}$ exhibit global changes in the expression of numerous genes. Although these differences may initially seem minor, the *Tent5c*$^{-/-}$ mice demonstrate significant aberrations in the terminal stages of male gamete development[1]. Therefore, more substantial transcriptomic alterations are observed in spermatids. Additionally, the integration of data on the composition[26,31,32] of poly(A) tails with changes in gene expression[27] provides a novel perspective for the study of the regulatory mechanisms of spermatogenesis (Figure 5).

*# Figure 5 should be inserted here.*

## Code Availability

Information concerning the software used and its settings is included in the Methods section. The Ninetails tool is available at https://github.com/LRB-IIMCB/ninetails. The detailed user manual (wiki) for this software is available here: https://github.com/LRB-IIMCB/ninetails/wiki. Docker container and scripts used for neural network training with hyperparameter finetuning are provided here: https://github.com/LRB-IIMCB/ninetails_processing. The NanoTail software for exploratory analysis and visualization of poly(A) predictions produced by Nanopolish is available here: https://github.com/LRB-IIMCB/nanotail.

## Acknowledgements

We thank Andrzej Dziembowski lab members for their support and all members of Genome Engineering Facility of International Institute of Molecular and Cell Biology for maintenance of animal colony and animal genotyping. NGS was performed thanks to Genomics Core Facility

CeNT UW (RRID:SCR_022718), using NovaSeq 6000 platform financed by Polish Ministry of Science and Higher Education (decision no. 6817/IA/SP/2018 of 2018-04-10).
The research leading to these results was funded by the Norwegian Financial Mechanism 2014 -2021 (UMO-2019/34/H/NZ3/00733), by the European Union, and European Research Council (ERC AdG 101097317). IIMCB core facilities, the IN-MOL-CELL infrastructure funded by the European Union – NextGenerationEU under National Recovery Plan, co-financed by the European Union under the European Funds for Smart Economy 2021-2027 (FENG) and funded by the European Union

## Author contributions

A.C.-C. supported by P.K. and N.G. performed all analyses, as well as prepared all figures and tables.

M.B. prepared ovarian RNA sequencing libraries.

S.M. prepared testicular RNA sequencing libraries.

A.C.-C., M.B., N.G. and A.D. wrote and revised the manuscript.

## Competing interests

Authors declare no competing interests.

## Figure legends

**Figure 1. Schematic representation of the Tent5 phenotypes in mouse model.** Summary of fertility status of males and females with all analyzed *Tent5* genes' mutations. Subfertile category refers to significantly reduced fertility. Female *Tent5d* KO mice were unavailable for analyses due to *Tent5d* localisation on chromosome X and infertility of *Tent5d* KO males.

**Figure 2. Schematic representation of the non-adenosine analysis workflow.** Nanopore data were basecalled with Guppy, mapped with minimap2, sorted and filtered with samtools. The coordinates of poly(A) tails within the nanopore signals were delimited with Nanopolish. Finally, Ninetails was used to analyze the nucleotide composition of poly(A) tails.

**Figure 3. Quality of underlying DRS sequencing data.**

**a.** Detailed classification of reads based on poly(A) tail length prediction quality tags provided by Nanopolish. The quality of the reads was independently summarized for each sample. **b.** Correlation matrix for all DRS datasets. Spearman correlations were calculated for each pair of samples, taking into account mean transcript abundances (counts). **c.** Frequency of reads labelled with quality tags calculated by Nanopolish for each sample. **d.** Correlation matrix for all DRS datasets. Spearman correlations were calculated for each pair of samples, taking into account mean poly(A) tail lengths. The visualisations were produced with the NanoTail package.

**Figure 4. Distribution of non-adenosine nucleotides in poly(A) tails.**
**a.** Detailed read classification made by Ninetails for poly(A) tail composition analysis. **b.** Frequency of occurrence of cytidine/guanosine/uridine nucleotides in poly(A) tails calculated for whole transcriptome. Only reads classified by neural network (decorated, blank) were included in the analysis. **c.** Frequency of occurrence of cytidine/guanosine/uridine nucleotides in poly(A) tails calculated for 522 oocyte-specific genes determined by Brouze et al. (2024). Only reads classified by neural network (decorated, blank) were included in analysis. **d.** Localization of uridines in poly(A) tails of transcripts (*Spem1*, *Tdrd5*) involved in regulation of spermatogenesis. **e.** Localization of uridines in poly(A) tails of transcript (*Insl3*) involved in regulation of spermatogenesis and oogenesis. **f.** Localization of uridines in poly(A) tails of transcript (*Rpl9*) not involved in regulation of spermatogenesis and oogenesis. The visualisations were produced with the Ninetails software. Differential expression analysis performed on RNA-seq data.

**Figure 5. Differential expression of transcripts with enrichment in uridines at the 3'- end of poly(A) tail.**
**a.** RNA-seq performed on testes isolated from Tent5c$^{-/-}$ and Tent5c$^{+/+}$ mice. Most transcripts with enrichment in uridines at the 3'-end of poly(A) tails, highlighted by triangles, are slightly downregulated in Tent5c$^{-/-}$. Differential expression was calculated using DESeq2. **b.** RNA-seq performed on round spermatids isolated from Tent5c$^{-/-}$ and Tent5c$^{+/+}$ mice. Most transcripts with enrichment in uridines at the 3'-end of poly(A) tails, highlighted by triangles, are slightly upregulated and highly expressed in Tent5c$^{-/-}$. Differential expression was calculated using DESeq2. **c.** RNA-seq performed on long spermatids isolated from Tent5c$^{-/-}$ and Tent5c$^{+/+}$ mice. Transcripts rich in uridines at the 3'-end of poly(A) tails are highlighted by triangles. Differential expression calculated using DESeq2. **d.** RNA-seq performed on round spermatids and long spermatids isolated from Tent5c$^{+/+}$ mice. Most transcripts with enrichment in uridines at the

3'-end of poly(A) tails, highlighted by triangles, are downregulated in long spermatids. Differential expression was calculated using DESeq2.

## References


1 Brouze, M. *et al.* TENT5-mediated polyadenylation of mRNAs encoding secreted proteins is essential for gametogenesis in mice. *Nat Commun* **15**, 5331, doi:10.1038/s41467-024-49479-4 (2024).

2 Zhang, Y. T. *et al.* Novel variations in TENT5D lead to teratozoospermia in infertile patients. *Andrology-Us*, doi:10.1111/andr.13589 (2024).

3 Sha, Y. *et al.* TENT5D disruption causes oligoasthenoteratozoospermia and male infertility. *Andrology-Us* **11**, 1121-1131, doi:10.1111/andr.13407 (2023).

4 Cong, J. *et al.* Deficiency of X-linked TENT5D causes male infertility by disrupting the mRNA stability during spermatogenesis. *Cell Discovery* **8**, 23, doi:10.1038/s41421-021-00369-9 (2022).

5 Lee, Y. S., Levdansky, Y., Jung, Y., Kim, V. N. & Valkov, E. Deadenylation kinetics of mixed poly(A) tails at single-nucleotide resolution. *Nat Struct Mol Biol* **31**, 826-834, doi:10.1038/s41594-023-01187-1 (2024).

6 Liu, Y., Nie, H., Liu, H. & Lu, F. Poly(A) inclusive RNA isoform sequencing (PAIso-seq) reveals wide-spread non-adenosine residues within RNA poly(A) tails. *Nat Commun* **10**, 5292, doi:10.1038/s41467-019-13228-9 (2019).

7 Morgan, M. *et al.* mRNA 3' uridylation and poly(A) tail length sculpt the mammalian maternal transcriptome. *Nature* **548**, 347-351, doi:10.1038/nature23318 (2017).

8 Morgan, M. *et al.* A programmed wave of uridylation-primed mRNA degradation is essential for meiotic progression and mammalian spermatogenesis. *Cell Res* **29**, 221-232, doi:10.1038/s41422-018-0128-1 (2019).

9 Brouze, A., Krawczyk, P. S., Dziembowski, A. & Mroczek, S. Measuring the tail: Methods for poly(A) tail profiling. *Wiley Interdiscip Rev RNA* **14**, e1737, doi:10.1002/wrna.1737 (2023).

10 Shi, H. *et al.* Bias in RNA-seq Library Preparation: Current Challenges and Solutions. *Biomed Res Int* **2021**, 6647597, doi:10.1155/2021/6647597 (2021).

11 Soneson, C. *et al.* A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nature Communications* **10**, doi:ARTN 3359

10.1038/s41467-019-11272-z (2019).

12 Workman, R. E. *et al.* Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nature Methods* **16**, 1297-+, doi:10.1038/s41592-019-0617-2 (2019).
13 Krause, M. *et al.* tailfindr: alignment-free poly(A) length measurement for Oxford Nanopore RNA and DNA sequencing. *Rna* **25**, 1229-1241, doi:10.1261/rna.071332.119 (2019).
14 Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods* **15**, 201-+, doi:10.1038/Nmeth.4577 (2018).
15 Bilska, A. *et al.* Immunoglobulin expression and the humoral immune response is regulated by the non-canonical poly(A) polymerase TENT5C. *Nat Commun* **11**, 2032, doi:10.1038/s41467-020-15835-3 (2020).
16 Gewartowska, O. *et al.* Cytoplasmic polyadenylation by TENT5A is required for proper bone formation. *Cell Rep* **35**, 109015, doi:10.1016/j.celrep.2021.109015 (2021).
17 Legnini, I., Alles, J., Karaiskos, N., Ayoub, S. & Rajewsky, N. FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. *Nature Methods* **16**, 879-+, doi:10.1038/s41592-019-0503-y (2019).
18 Liu, Y. S. *et al.* Remodeling of maternal mRNA through poly(A) tail orchestrates human oocyte-to-embryo transition. *Nature Structural & Molecular Biology* **30**, 200-+, doi:10.1038/s41594-022-00908-2 (2023).
19 ENA European Nucleotide Archive https://identifiers.org/ena.embl:PRJEB63526 (2024).
20 ENA European Nucleotide Archive https://identifiers.org/ena.embl:PRJEB46685 (2024).
21 ENA European Nucleotide Archive https://identifiers.org/ena.embl:PRJEB45063 (2021).
22 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research* **47**, D766-D773, doi:10.1093/nar/gky955 (2019).
23 Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572-4574, doi:10.1093/bioinformatics/btab705 (2021).
24 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, doi:10.1093/gigascience/giab008 (2021).
25 Krawczyk, P. S., Tudek, A., Mroczek, S. & Dziembowski, A. Transcriptome-Wide Analysis of mRNA Adenylation Status in Yeast Using Nanopore Sequencing. *Methods Mol Biol* **2723**, 193-214, doi:10.1007/978-1-0716-3481-3_12 (2024).

26 Czarnocka-Cieciura, A. *et al.* DESeq2 analysis results of the mouse ovarian and testicular samples sequenced with nanopore Diect RNA Sequencing. *Figshare*, https://figshare.com/account/projects/221179/articles/27052504?file=49264906 .

27 GEO Gene Expression Omnibus https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE273264 (2024).

28 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

29 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).

30 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, doi:10.1186/s13059-014-0550-8 (2014).

31 Czarnocka-Cieciura, A. *et al.* Profiling of non-adenosine residues in poly(A) tails of the mouse ovarian and testicular samples sequenced with nanopore Diect RNA Sequencing and analyzed with Ninetails software. *Figshare*, https://figshare.com/account/projects/221179/articles/27051895?file=49264342 .

32 Czarnocka-Cieciura, A. *et al.* Classification of poly(A) tails based on presence/absence of non-adenosine residues in the mouse ovarian and testicular samples sequenced with nanopore Diect RNA Sequencing and analyzed with Ninetails software. *Figshare*, https://figshare.com/account/projects/221179/articles/27051787?file=49264159.

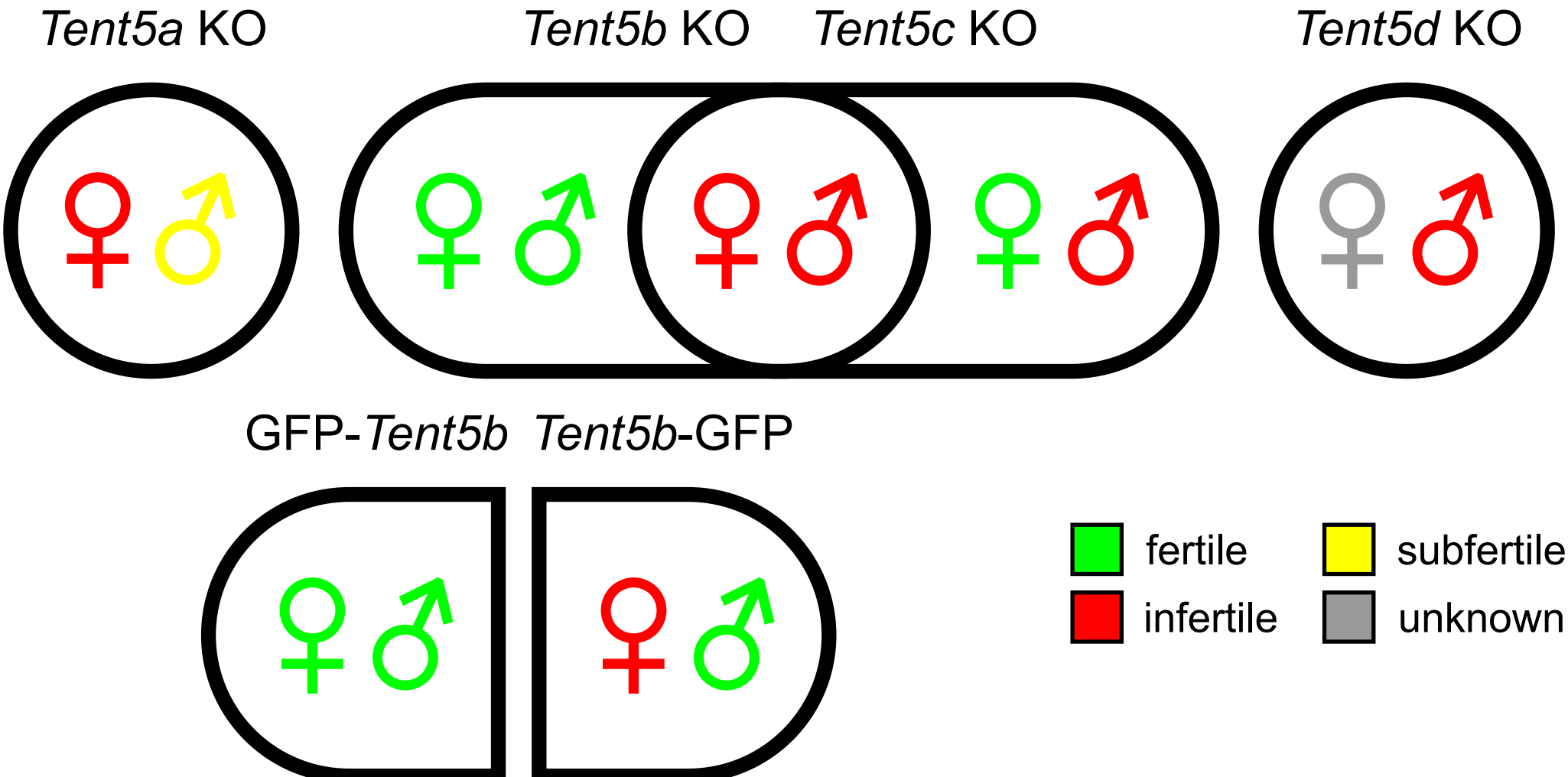
33 Czarnocka-Cieciura, A. *et al.* Metadata of the mouse ovarian and testicular samples sequenced with nanopore Diect RNA Sequencing and Illumina RNA-seq. *Figshare*, https://figshare.com/articles/dataset/Table1_metadata/27051682?file=49264060 .

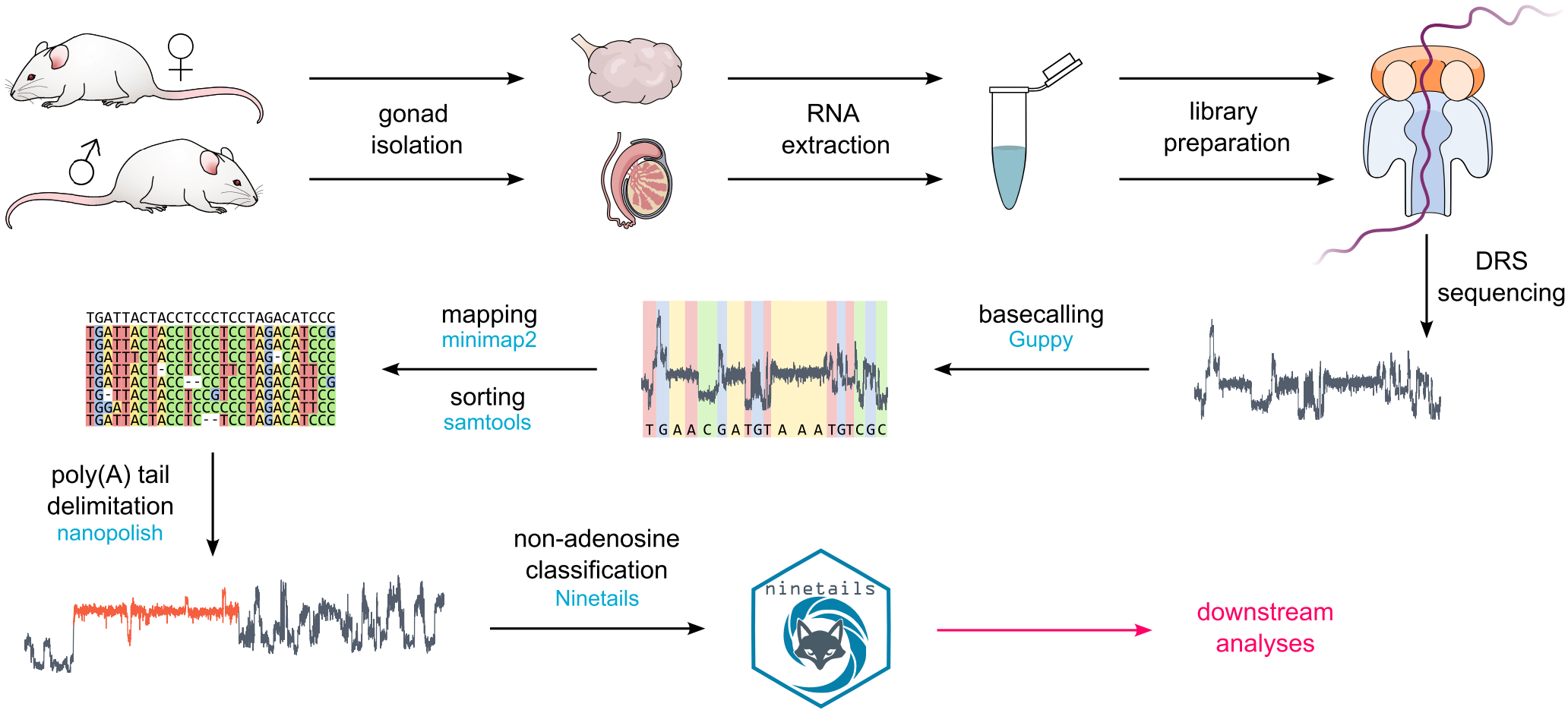
34 Zheng, H. *et al.* Lack of Spem1 causes aberrant cytoplasm removal, sperm deformation, and male infertility. *Proc Natl Acad Sci U S A* **104**, 6852-6857, doi:10.1073/pnas.0701669104 (2007).

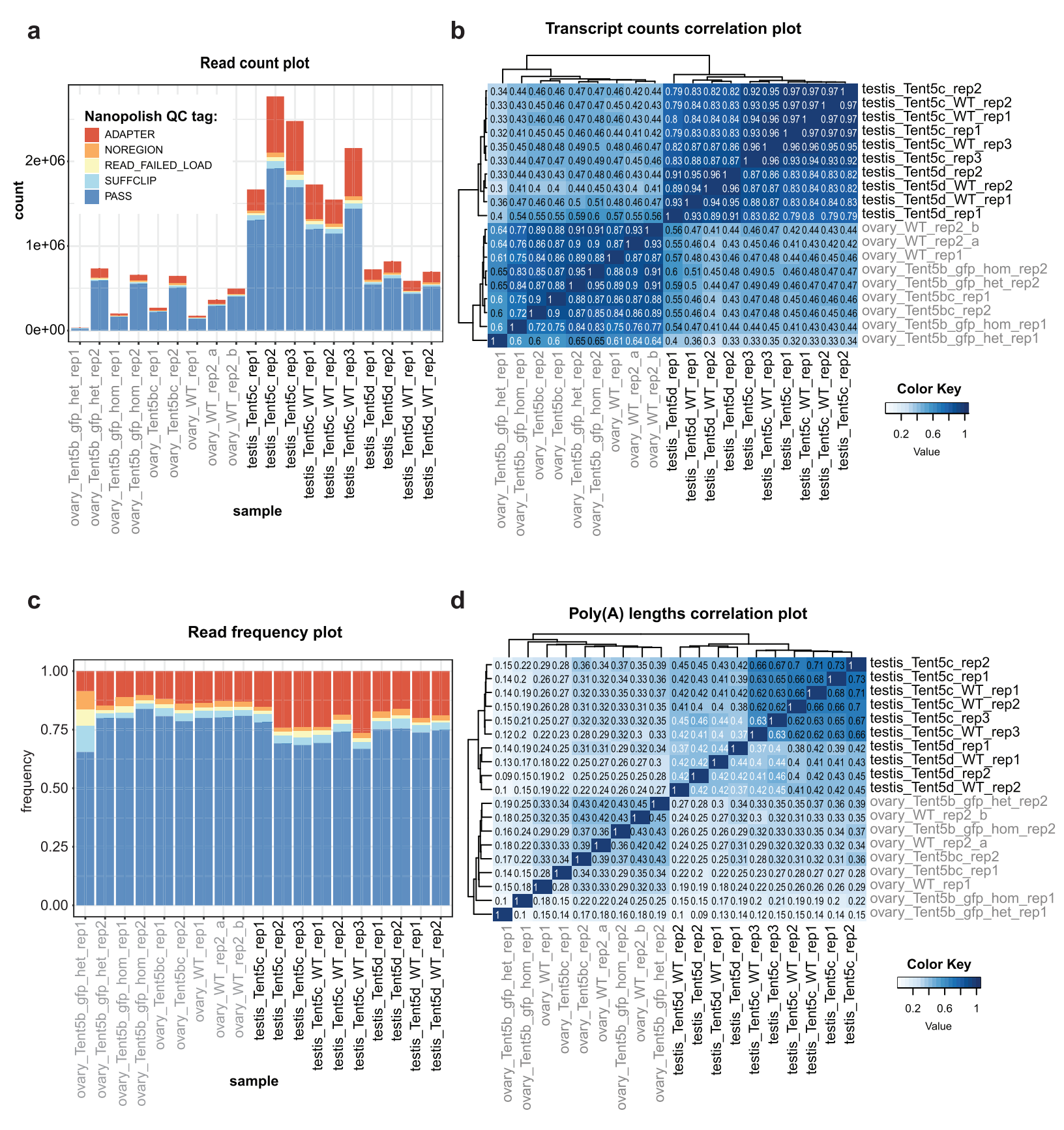
35 Mannowetz, N., Wandernoth, P. & Wennemuth, G. Basigin interacts with both MCT1 and MCT2 in murine spermatozoa. *J Cell Physiol* **227**, 2154-2162, doi:10.1002/jcp.22949 (2012).

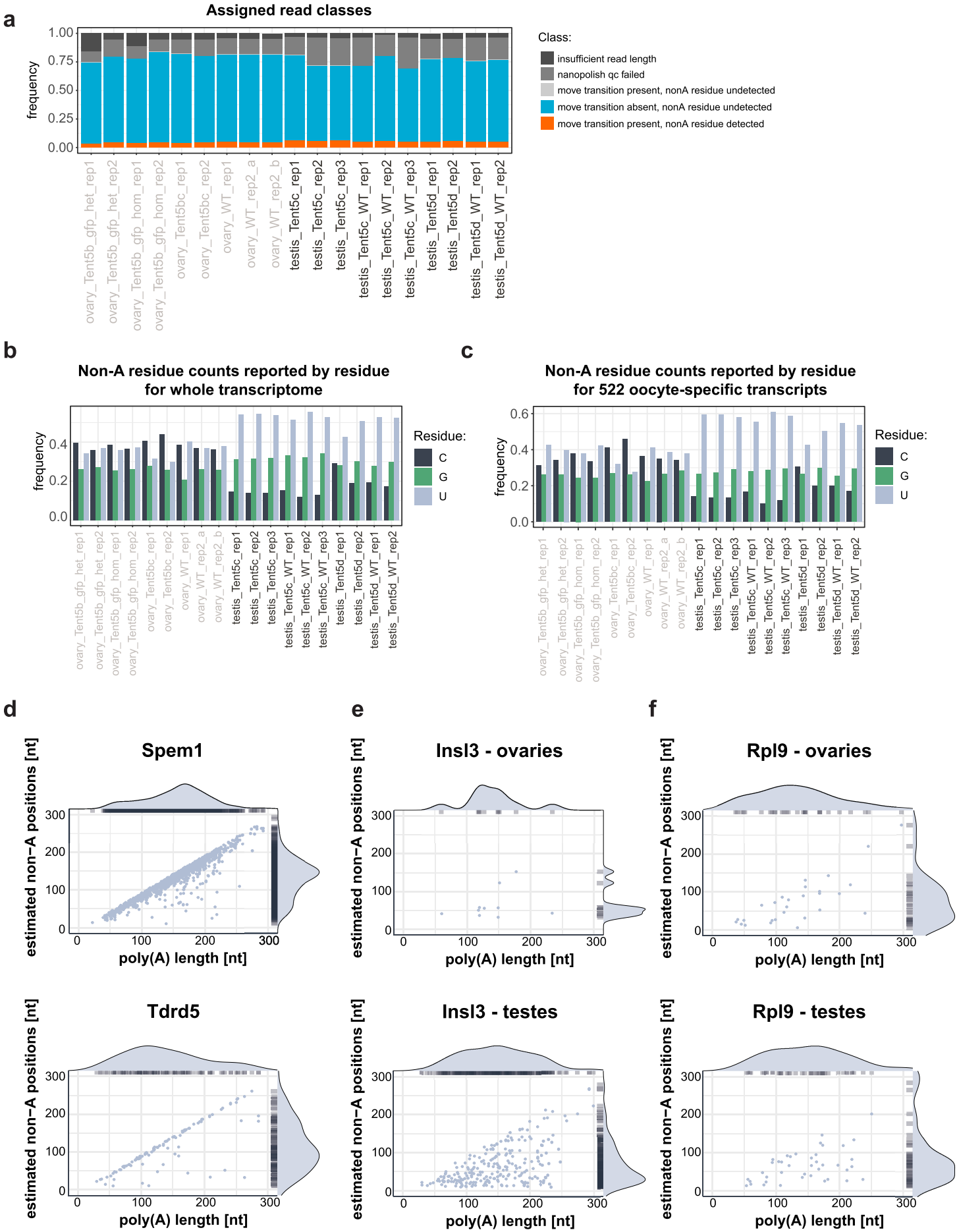
36 Miki, K. *et al.* Targeted disruption of the Akap4 gene causes defects in sperm flagellum and motility. *Dev Biol* **248**, 331-342, doi:10.1006/dbio.2002.0728 (2002).

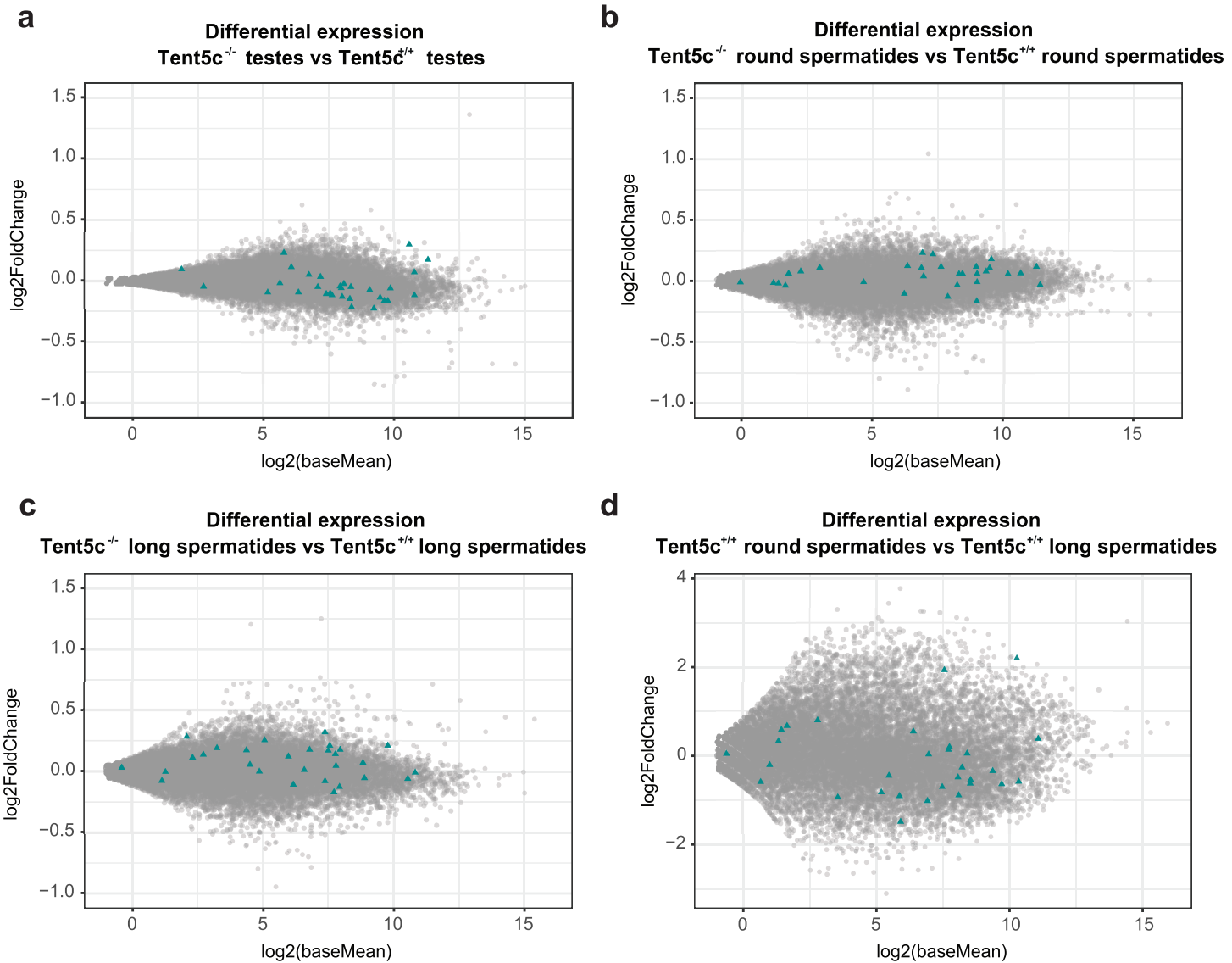
37 Yabuta, Y. *et al.* TDRD5 is required for retrotransposon silencing, chromatoid body assembly, and spermiogenesis in mice. *Journal of Cell Biology* **192**, 781-795, doi:10.1083/jcb.201009043 (2011).


*Tent5a* KO
*Tent5b* KO
*Tent5c* KO
*Tent5d* KO
GFP-*Tent5b*
*Tent5b*-GFP
fertile
subfertile
infertile
unknown


gonad
isolation
RNA
extraction
library
preparation
DRS
sequencing
basecalling
Guppy
mapping
minimap2
sorting
samtools
TGATTACTACCTCCCTCCTAGACATCCC
TGATTACTACCTCCCTCCTAGACATCCG
TGATTACTACCTCCCTCCTAGACATCCC
TGATTTCTACCTCCCTCCTAG-CATCCC
TGATTACT-CCTCCCTTCTAGACATTCC
TGATTACTACC--CCTCCTAGACATTCG
TG-TTACTACCTCCGTCCTAGACATTCC
TGGATACTACCTCCCCCCTAGACATTCC
TGATTACTACCTC--TCCTAGACATCCC
T G A A C G A T G T A A A T G T C G C
poly(A) tail
delimitation
nanopolish
non-adenosine
classification
Ninetails
ninetails
downstream
analyses


a
Read count plot
Nanopolish QC tag:
ADAPTER
NOREGION
READ_FAILED_LOAD
SUFFCLIP
PASS
count
2e+06
1e+06
0e+00
ovary_Tent5b_gfp_het_rep1
ovary_Tent5b_gfp_het_rep2
ovary_Tent5b_gfp_hom_rep1
ovary_Tent5b_gfp_hom_rep2
ovary_Tent5bc_rep1
ovary_Tent5bc_rep2
ovary_WT_rep1
ovary_WT_rep2_a
ovary_WT_rep2_b
testis_Tent5c_rep1
testis_Tent5c_rep2
testis_Tent5c_rep3
testis_Tent5c_WT_rep1
testis_Tent5c_WT_rep2
testis_Tent5c_WT_rep3
testis_Tent5d_rep1
testis_Tent5d_rep2
testis_Tent5d_WT_rep1
testis_Tent5d_WT_rep2
sample
b
Transcript counts correlation plot
testis_Tent5c_rep2
testis_Tent5c_WT_rep2
testis_Tent5c_WT_rep1
testis_Tent5c_rep1
testis_Tent5c_WT_rep3
testis_Tent5c_rep3
testis_Tent5d_rep2
testis_Tent5d_WT_rep2
testis_Tent5d_WT_rep1
testis_Tent5d_rep1
ovary_WT_rep2_b
ovary_WT_rep2_a
ovary_WT_rep1
ovary_Tent5b_gfp_hom_rep2
ovary_Tent5b_gfp_het_rep2
ovary_Tent5bc_rep1
ovary_Tent5bc_rep2
ovary_Tent5b_gfp_hom_rep1
ovary_Tent5b_gfp_het_rep1
Color Key
0.2
0.6
1
Value
c
Read frequency plot
frequency
1.00
0.75
0.50
0.25
0.00
sample
d
Poly(A) lengths correlation plot
Color Key
0.2
0.6
1
Value


a
Assigned read classes
frequency
1.00
0.75
0.50
0.25
0.00
Class:
insufficient read length
nanopolish qc failed
move transition present, nonA residue undetected
move transition absent, nonA residue undetected
move transition present, nonA residue detected
ovary_Tent5b_gfp_het_rep1
ovary_Tent5b_gfp_het_rep2
ovary_Tent5b_gfp_hom_rep1
ovary_Tent5b_gfp_hom_rep2
ovary_Tent5bc_rep1
ovary_Tent5bc_rep2
ovary_WT_rep1
ovary_WT_rep2_a
ovary_WT_rep2_b
testis_Tent5c_rep1
testis_Tent5c_rep2
testis_Tent5c_rep3
testis_Tent5c_WT_rep1
testis_Tent5c_WT_rep2
testis_Tent5c_WT_rep3
testis_Tent5d_rep1
testis_Tent5d_rep2
testis_Tent5d_WT_rep1
testis_Tent5d_WT_rep2
b
Non-A residue counts reported by residue for whole transcriptome
frequency
0.4
0.2
0.0
Residue:
C
G
U
c
Non-A residue counts reported by residue for 522 oocyte-specific transcripts
frequency
0.6
0.4
0.2
0.0
Residue:
C
G
U
d
Spem1
e
Insl3 - ovaries
f
Rpl9 - ovaries
Tdrd5
Insl3 - testes
Rpl9 - testes
estimated non−A positions [nt]
poly(A) length [nt]
0
100
200
300


a
Differential expression
Tent5c-/- testes vs Tent5c+/+ testes
b
Differential expression
Tent5c-/- round spermatides vs Tent5c+/+ round spermatides
c
Differential expression
Tent5c-/- long spermatides vs Tent5c+/+ long spermatides
d
Differential expression
Tent5c+/+ round spermatides vs Tent5c+/+ long spermatides
log2FoldChange
log2(baseMean)

**Manuscript 3. Brouze M., Szpila M., Czerwińska A., Antczak W., Mroczek S., Kuliński T., Hojka-Osińska A., Cysewski D., Adamska D., Gruchota J., Borsuk E., Dziembowski A.** ***DIS3L, cytoplasmic exosome catalytic subunit, is essential for development but not cell viability in mice.*** **bioRxiv 2022.12.14.520403; doi: https://doi.org/10.1101/2022.12.14.520403**

# DIS3L, cytoplasmic exosome catalytic subunit, is essential for development but not cell viability in mice.

Michał Brouze[1,2], Marcin Szpila[1,3], Areta Czerwińska[1,3], Wiktor Antczak[1,3], Seweryn Mroczek[1,4], Tomasz M. Kuliński[1,2], Anna Hojka-Osińska[1,2], Dominik Cysewski[2,6], Dorota Adamska[2,5], Jakub Gruchota[1,2], Ewa Borsuk[1,3]and Andrzej Dziembowski[1,2,3*]

1. International Institute of Molecular and Cell Biology in Warsaw, 02-109, Warsaw, Poland
2. Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-106, Warsaw, Poland
3. Institute of Developmental Biology and Biomedical Sciences, Faculty of Biology, University of Warsaw, 02-096, Warsaw, Poland
4. Institute of Genetics and Biotechnology, Faculty of Biology, University of Warsaw, 02-096, Warsaw, Poland
5. Genomics Core Facility, Centre of New Technologies, University of Warsaw, 02-097, Warsaw, Poland
6. Clinical Research Centre, Medical University of Bialystok, 15-276, Białystok, Poland

* To whom correspondence should be addressed. Email: adziembowski@iimcb.gov.pl

## Abstract

Among numerous enzymes involved in RNA decay, processive exoribonucleases are the most prominent group responsible for the degradation of the entire RNA molecules. The role of mammalian cytoplasmic 3'-5' exonuclease DIS3L at the organismal level remained unknown. Herein we established knock-in and knock-out mouse models to study DIS3L functions in mice. DIS3L is indeed a subunit of the cytoplasmic exosome complex, which disruption leads to severe embryo degeneration and death in mice soon after implantation. These changes could not be prevented by supplementing extraembryonic tissue with functional DIS3L through the construction of chimeric embryos. Preimplantation *Dis3l*$^{-/-}$ embryos were unaffected in their morphology and ability to produce functional embryonic stem cells showing that DIS3L is not essential for cell viability. There were also no major changes in the transcriptome level for both embryonic stem cells and blastocysts, as revealed by RNA sequencing experiments. Notably, however, DIS3L knock-out led to inhibition of the global protein synthesis. These results point to the essential role of DIS3L in mRNA quality control pathways crucial for proper protein synthesis during embryo development.

## Introduction

In recent years RNA decay pathways emerged as the RNA control mechanisms of importance comparable to transcription regulation or posttranscriptional modifications. While RNA degrading enzymes consist of different types and families, processive exoribonucleases are the most prominent group, as they are responsible for the degradation of entire transcripts during mRNA turnover or in quality control pathways. The mammalian genomes encode six processive exoribonucleases, divided into three families. The first family of 5' to 3' exonucleases consists of XRN1 and XRN2, expressed, respectively, in P-bodies scattered across cytoplasm (1) and nucleus (including nucleolus) (2). The second family consists of three hydrolitic 3' to 5' exoribonucleases. Two of them, nuclear DIS3 (which also retained endoribonucleolytic activity) and cytoplasmic DIS3L, are parts of the ring-shaped exosome complex (3,4), which by itself remains inactive in eukaryotes. Additionally, DIS3L2 exoribonuclease is also expressed in the cytoplasm, albeit in free, monomeric form (5). Finally, phosphorolytic 3' to 5' ribonuclease PNPase is responsible for RNA decay in mitochondria (6). While enzymes mentioned above are responsible for the bulk of RNA decay, plenty of less potent enzymes contribute to RNA metabolism, including nucleolar enriched, exosome associated distributive exoribonuclease EXOSC10/RRP6 (7,8). The substrates and mechanisms of recognition appearing in

concert with activatory complexes are quite well established for those enzymes thanks to years of biochemistry, structural biology, and functional studies using cell lines. However, the knowledge about their role *in vivo* is more fragmentary.

DIS3 family of proteins, specifically, were shown to be associated with cancer. Mutation of DIS3L2, a nuclease responsible for the degradation of ARE-containing transcripts (5) and oligouridylated ncRNAs in their quality control pathway (9), is the leading cause of the Perlman syndrome and, consequently, Wilms tumor (10-12). Exosome-associated DIS3, which in the nucleus targets pervasive transcription products (promoter upstream transcripts and enhancer RNAs), snoRNAs and premature termination products (13), is among most frequently mutated genes in multiple myeloma patients (14-16). Unlike DIS3L2, however, we showed that DIS3 is an essential gene in mice, as evidenced by the analysis of both knock-out and point catalytic mutations of the catalytic domain (unpublished). Interestingly, this is also the case for the second nuclear exosome-bound exoribonuclease, RRP6, which was recently identified to be essential for mouse embryo development past the morula stage (17). Notably, essentially nothing is known about the role of DIS3L at the organismal level.

Here we present the results of our study of mammalian DIS3L *in vivo* through the analysis of knock-out mutation phenotype of *Dis3l* gene in mice. We demonstrate that DIS3L is essential for mouse embryo development beyond day 6,5. Lethal phenotype in *Dis3l* knock-out (KO) embryos could not be rescued by supplementation of wild-type (WT) cells to extraembryonic tissue. Although preimplantation embryos lacking functional DIS3L were able to produce embryonic stem (ES) cells, albeit, with reduced efficiency, they also displayed accumulation of a number of transcripts, as shown by RNA sequencing. Curiously, the accumulation of some of those transcripts does not produce higher protein levels. Further examination showed overall impaired protein production in those KO preimplantation embryos.

## Materials and Methods

### Animals

Mice lines were generated using CRISPR/Cas9 method in C57BL/6/Tar x CBA/Tar mixed background mice, following the procedure described previously (18). The loss-of-function *Dis3l*$^{-/-}$ mutation was generated by random insertion of 169 bp fragment through non-homologous end joining repair mechanism in exon 10 in 1259 nt position of cDNA sequence of *Dis3l* gene (sgRNA sequence: TCCAGGTTGCCGTTATTCA). Knock-in *Dis3l*$^{GFP/GFP}$ mutation was introduced in the ORF of *Dis3l* gene at the 3'-end through homologous recombination by insertion of GFP coding single-stranded DNA donor (sequence in Supplementary Figure 1.) with 60 bp homology arms on both ends (sgRNA sequence: GACAAAGGTCTTTAATGACA). All animal experiments were approved by the Local Ethical Committee in Warsaw affiliated to the University of Warsaw, Faculty of Biology (approval numbers: 527/2013, 176/2016) and were performed according to Polish Law (Act number 266/15.01.2015).

### Embryos isolation and culture

Superovulation was artificially stimulated in females by intraperitoneal injection of 10 units of pregnant mare's serum gonadotropin (PMSG, BioVendor) and human chorionic gonadotropin (hCG, MSD Animal Health), respectively, in 48 h interval. Stimulated females were mated with males in the evening and terminated by cervical dislocation and dissected in following days to obtain either 1-, 2-, 4-, 8-cell or blastocysts stage embryos. Zygotes were isolated by puncturing oviducts ampulla in hyaluronidase solution (Sigma-Aldrich) in PBS (300 µg/ml) to dissociate cumulus cells. Other embryos were isolated by flushing oviducts and uterus with M2 medium (Sigma-Aldrich). After isolation all embryos were rinsed in M2 medium and used for further procedures. Postimplantation embryos were isolated from unstimulated females 6 or 7 days after mating with males.

All embryos were cultured in cell culture incubator with 5% $CO_2$ in the atmosphere and 38°C. Embryos were placed in drops of either M2 (for short-term culture) or M16 (for multi-day culture; Sigma-

Aldrich) medium under mineral oil (Sigma-Aldrich) on 35-mm Falcon Easy-Grip tissue culture dishes (Thermo Fisher Scientific).

**Generation and in vitro culture of ES cells and EBs**

For the derivation and culture of ES cells, feeder cells - inactivated mouse embryonic fibroblasts (MEFs) - were prepared according to Robertson (19). Blastocysts collected 96 h after hCG injection were transferred to single wells of culture plates coated with 0,2% gelatin (Sigma-Aldrich) with feeder layer of inactivated MEFs. Composition of the derivation medium was designed based on 2i method (20) (full composition in Supplementary Table 1). After 3-4 days of culture in cell culture incubator in 38°C and with 5% $CO_2$ in the atmosphere, blastocysts formed outgrowths which were disaggregated by 5 min. incubation in 0,25% trypsin-EDTA (Gibco) followed by mechanical pipetting. Resulting cell suspensions were transferred into the separate culture plate wells onto fresh layer of inactivated MEFs and inspected daily for appearance of primary colonies. Cultures containing ES cells were expanded, processed for genotyping, and frozen for further investigation. Established ES cell lines were cultured in ES cells medium composed of KnockOut Dulbecco's modified Eagle's Medium (KnockOut DMEM, Gibco) supplemented with 15% heat inactivated FBS (Performance Plus, Gibco) with addition of nonessential amino acids, L-glutamine, β-mercaptoethanol, penicillin and streptomycin, and 500 IU/ml LIF (Gibco).

To gather pure ES cells population MEFs were removed from cultures by pre-plating, i.e. incubating cells with 0.05% trypsin-EDTA (Sigma-Aldrich) for 3-5 min and seeding of suspended cells onto culture dish covered with gelatin for 20 min, allowing MEFs to attach to the dish. MEF-free suspension of ES cells was cultured on a new gelatin-covered culture dish until reaching confluency. ES cells were then detached by incubation in 0,05% trypsin-EDTA for 3-5 min and washed in PBS (BioShop). Centrifugated dry pellets were frozen at -80°C until further processing.

To induce differentiation of ES cells, embryoid bodies (EBs) were generated. First, ES cells' colonies were disaggregated using trypsin and pre-plated for 20 minutes (as described above). Part of collected cells suspension (day 0 of EB culture) was kept for RNA isolation by freezing a dry pellet after wash in PBS. The second part of cells suspension was centrifuged, and cells were cultured in differentiation medium (ES cells culture medium without LIF) in the concentration of 800 cells per 30 µl in hanging drops to stimulate EBs formation. After 2 days of culture in hanging drops EBs were either collected for analysis or washed and placed in low-adhesive dishes (Medlab) allowing for their culture in suspension. After 5 days of differentiation, the rest of EBs were collected for analysis.

**Immunoprecipitation**

Livers from 10-20-week-old *Dis3l*$^{GFP/GFP}$ knock-in mice were homogenized in lysis buffer (10 mM Tris-HCl pH 8; 125 mM NaCl; 1 mM DTT; 1 mM PMSF; 0,02 µM pepstatin A; 0,02 ug/ml chymostatin; 0,006 µM leupeptin; 20 uM benzamidine hydrochloride) and processed as described in (3) and (21) and analysed by mass spectrometry by the Mass Spectrometry Laboratory, IBB PAS. Identified proteins that were considered non-specific contaminants (haemoglobin subunits, immunoglobulins, keratins and trypsinogens) were excluded from the final analysis of enriched proteins.

**DNA isolation and genotyping**

For genotyping of mice, postimplantation embryos and ES cells adjusted HotShot method was used (22). DNA from fixed and stained embryos was isolated using the protocol described in (23). Preimplantation embryos were repeatedly frozen and warmed several times in a minimal amount of RNase and DNase-free water in PCR tubes. Obtained DNA was used for genotyping by PCR reaction using Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific). Reactions of 20 µl volume were set up according to the manufacturer's protocol, with no DMSO and 10 µM primers used (primer sequences in Supplementary Table 2). Additionally, blastocysts from which RNA was extracted for sequencing and RT-qPCR experiments, were genotyped as described for gDNA above, with cDNA generated during those experiments as a PCR template.

**RNA isolation and processing**

RNA from ES cells was extracted by lysis of 1 million cells using TRI Reagent (Sigma-Aldrich), following manufacturer's protocol. RNA from EBs was isolated using High Pure RNA Isolation Kit (Roche), following manufacturers protocol. RNA from living embryos was extracted using PicoPure RNA Isolation Kit (Thermo Fisher Scientific) following manufacturers protocol with following modifications: single embryos were incubated for 30 min at 42°C in Extraction Buffer and then mixed with ethanol in 1:1 ratio. All isolated RNA samples were stored at -80°C until further use. DNA was removed from all RNA samples using TURBO DNA-free Kit (Thermo Fisher Scientific) following manufacturers protocol for routine DNase treatment, with additional 2 µl of RiboLock RNase Inhibitor (Thermo Fisher Scientific) added to every reaction. For RNA sequencing and RT-qPCR experiments cDNA was synthesized. For single blastocysts RNA sequencing, RT-qPCR and ES cells RT-qPCR experiments, SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) and oligo$(dT)_{20}$ primer was used to generate cDNA from total RNA, following modified manufacturer's protocol. Reverse transcription reactions were cleaned up using AMPure XP Reagent magnetic beads (Beckman Coulter). For EBs RT-qPCR, 250 ng of total RNA was used for reverse transcription reaction with RevertAid First-Strand cDNA Synthesis kit (Thermo Fisher Scientific), following manufacturer's protocol.

**Chimeric embryos**

To produce tetraploid (4n) WT embryos, 2-cell diploid (2n) WT embryos were isolated from superovulated females 30 h after hCG administration. Early 2-cells embryos were electroporated using home-made electroporation chamber and programmable square pulse generator. Embryos were rinsed several times and placed in electroporation buffer (100 mM $MgSO_4*7H_2O$, 100 mM $CaCl_2*2H_2O$ and 0,045 g/ml glucose in $H_2O$) warmed to 38°C in a glass petri dish between two straight electrodes of platinum wire attached to the dishes bottom approximately 100 µm apart, with embryo cleavage plane parallel to the electrodes. Two 40 V pulses lasting 40 µs in 100 µs interval were generated to induce fusion of the blastomeres. 2-cell embryos were washed once in M2 medium and placed in culture for 24 h. Within the first hour from electroporation embryos were inspected for survival and blastomere fusion and sorted adequately to trace their development back to 2-cell stage.

Simultaneously to 4n WT embryos reaching 2-cell stage, 4-to-8-cell 2n embryos were isolated from *Dis3l*$^{+/-}$ females previously mated with males. All embryos had their zona pellucida removed by short (ca. 10 s) rinsing in Tyrode's Acidic Solution (Sigma-Aldrich) warmed to 38°C. Two 4n and one 2n embryos were rinsed and then moved to fresh aggregation solution: 0,3mg/ml phytohemaglutynin (Sigma-Aldrich) in M2 medium without BSA (home-made). Embryos were manually attached together using mouth pipette. Chimeric embryos prepared this way were rinsed with M2 medium and cultured overnight in M16 medium. Chimeric embryos were then transferred to pseudopregnant females on the first day of pseudopregnancy, following the same procedure as for the mice line generation.

**Embryo fixation and staining**

Preimplantation embryos were fixed in 4% PFA (Thermo Fisher Scientific) for 30 min and permeabilised in 0,5% Triton X-100 (Sigma-Aldrich) for 20 min. After fixation embryos were placed in drops of blocking solution (3% BSA (Santa Cruz Biotechnology) in PBST) on plastic dishes overnight before immunostaining or were stored that way until further procedures.

All immunofluorescence labelling was performed by incubating fixed embryos in blocking solution with appropriate primary antibody dilution (Supplementary Table 3) at 4°C overnight. Embryos were washed once in PBS and twice in blocking solution and incubated with secondary antibody (Supplementary Table 3) for 2 h at room temperature. After 30 min washing in PBS blastocysts were ready for imaging or additional staining. Methionine analogue labelling was performed by incubating embryos in 25 µM Click-iT AHA (L-Azidohomoalanine, Thermo Fisher Scientific) solution in M16 medium for 2 h and then fixing them as described above. AHA incorporated in proteins was detected using dedicated Click-iT Cell Reaction Buffer Kit (Thermo Fisher Scientific) and Alexa Fluor 594 Alkyne (Thermo Fisher Scientific) in 1 µM final concentration according to manufacturer's protocol. After brief

washing in 3% BSA and chromatin staining embryos were ready for imaging. Chromatin in embryos was stained with Hoechst 33343 (1:5000 dilution in PBS, 2 µg/ml, Sigma-Aldrich) for 15 minutes.
All stained embryos were imaged on uncoated 35 mm plastic dishes with glass bottom (MatTek) in drops of M2 medium using LSM 800 Confocal Laser Scanning microscope (Zeiss). Obtained images were analysed using ImageJ and Python software. Cell number in blastocysts was counted in confocal sections manually using ImageJ built-in tools. SLC37A2 and AHA intensities were gathered from maximum projection of individually scanned blastocysts using Python and ImageJ, respectively.

**RNA sequencing libraries**

For ES cells' RNA sequencing, total DNA-free RNA was ribodepleted using Ribo-Zero Gold rRNA-removal kit (Illumina). rRNA-free RNA samples were cleaned-up by precipitation with 3 M sodium acetate. Sequencing libraries were prepared using KAPA Stranded RNA-Seq Library Preparation Kit (KAPA Biosystems). Between each step samples were cleaned up using AMPure XP Reagent magnetic beads. Library quality and fragment size distribution were verified by electrophoresis in Agilent 2100 Bioanalyzer (Agilent Technologies Inc.).
RNA-seq libraries from single blastocysts' DNA-free RNA samples were prepared by tagmentation reaction following published protocols (24,25) with various steps and amount of enzyme optimized for use of a home-made batch of Tn5 transposase produced in our laboratory. Briefly: Tn5 (0,25 mg/ml) was loaded with linker oliguncleotides Tn5ME-A/Tn5Me-rev and Tn5ME-B/Tn5Me-rev by mixing of 10 µl Tn5 with 0,5 µl of both linkers (0,35 µM) and incubating for 45 min in 23°C with shaking at 350 rpm. Right before the reaction setup, loaded Tn5 was diluted 10 times with nuclease-free water. 10 µl of freshly prepared tagmentation buffer (20 mM Tris-HCl pH 7,5; 20 mM $MgCl_2$; 50% dimethylformamide (Sigma-Aldrich) were mixed with 5 µl of diluted Tn5 and 5 µl of cDNA, incubated for 3 min at 55°C in a preheated thermocycler and cooled to 10°C for 1 min. Reaction was inactivated by addition of 5 µl of 0,2% SDS (Sigma-Aldrich) and incubation for 5 min at room temperature (RT). Tagmented cDNA was purified with AMPureXP Reagent magnetic beads using 1:1,25 cDNA to beads ratio.
For library amplification KAPA HiFi HotStart ReadyMix PCR Kit (Roche) with addition of 5% DMSO was used. Number of cycles yielding the best results was determined experimentally and was further individually adjusted for each batch of prepared cDNA. Library quality and fragment size distribution were verified by electrophoresis using Agilent 2100 Bioanalyzer.
ES cells' RNA libraries were sequenced on NextSeq 500 instrument (Illumina) in 2x75 cycles pair-end mode in Next Generation Sequencing Unit of the International Institute of Molecular and Cell Biology in Warsaw. Single blastocysts' RNA libraries were sequenced on NovaSeq 6000 (Illumina) instrument in 2x100 cycles pair-end mode in Genomics Core Facility of Centre of New Technologies, University of Warsaw.

**RNA sequencing bioinformatic analysis**

Biological triplicates were analysed for each genotype and reads were mapped to the reference mouse genome (GRCm38) using the STAR short read aligner (version STAR v2.7.10a) (26). Quality control, read processing and filtering, and visualization of the results were performed using custom scripts and elements of the RSeQC, BEDtools and SAMtools packages. Reads were counted to the Genecode v M6 basic annotation using featureCounts of the subread package (v2.0.1). Differential expression analyses were performed using the DESeq2 (version 1.34.0) Bioconductor R package (27).
Functional enrichment analysis was performed with g:Profiler using gprofiler2 package for R (database version: e106_eg53_p16_65fcd97) with following settings: g:SCS multiple testing correction; 0,05 significance threshold (28).

**RT-qPCR**

RT-qPCR experiments on ES cells and blastocysts were performed using Platinum SYBR Green qPCR SuperMix-UDG (Invitrogen) in LightCycler 480 Instrument (Roche) according to the manufacturer's protocol. All ES cells cDNA samples were diluted 10 times before using as a reaction template. All

primers used are described in Supplementary Table 2. All biological samples were run in technical triplicates. Number of biological samples is denoted in the Results section.
RT-qPCR experiment on EBs was performed using TaqMan Gene Expression Master Mix (Life Technologies) in LightCycler 96 instrument (Roche), according to the manufacturer's protocol. All RNA samples were diluted 4 times. All samples were run in technical duplicates. Specific TaqMan probes against 4 different transcripts were used: Mm00443081_m1 (*Pax6*), Mm01976556_s1 (*Foxa2*), Mm01318252_m1 (*Tbxt*), Mm01205647_g1 (*Actb*).
Cp values of all technical replicates of one biological sample results were averaged. These results were analysed using 2^-ddCp method (final formula used: $2^{-ddCp} = 2^{-[(Cp_{GOI} - Cp_{housekeeping})_{KO} - (Cp_{GOI} - Cp_{housekeeping})_{WT}]}$). Obtained Cp values of genes of interest from every sample were normalized to either actin b (EBs) or GAPDH (ES cells and blastocysts) Cp value. For blastocysts RT-qPCR analysis, only samples with technical replicates' standard deviation below 1 were chosen.

**Statistical analysis**
All statistical analyses were performed using RStudio. All data plots present individual data points (where applicable), as well as mean value with standard error of mean (SEM) indicated as error bars (where applicable). All experiments were performed in at least two individual repetitions (with exception of immunoprecipitation experiment with one repetition) and in each experiment biological samples were obtained from at least 3 different animals. Distribution ratios were compared by Fisher exact test for 2x2 contingency tables (viability of embryos) or Chi-square test for 2x3 tables (genotype distributions). Mean values were compared using either parametric two-sided t-test for sample sets with normal distribution or non-parametric two-sided Mann-Whitney-Wilcoxon test for sample sets without normal distribution. Distribution normality was tested with Shapiro-Wilk test. Value "n" in figures and figure descriptions denotes the number of individual embryos analysed. P-values for every test are reported on plots.

## Results

**DIS3L is a part of the cytoplasmic exosome in mice**
The composition and functionality of the mammalian cytoplasmic exosome complex and its catalytic subunit, DIS3L, was previously analysed only in highly modified cell lines (HEK293 and HeLa). Thus, we have generated a knock-in mouse line with DIS3L C-terminally tagged with GFP in the endogenous locus using the CRISPR/Cas9 system to study it in a more physiologically relevant system. These mice in the homozygous state were healthy, with no apparent developmental, morphological, or physiological abnormalities. Then, we used protein extracts from three $DIS3L^{GFP/GFP}$ mice livers, to determine DIS3Ls association with cytoplasmic exosome complex in mammals by performing immunoprecipitation on anti-GFP resin. Extracts from one WT mouse liver served as a control. Quantitative mass spectrometry analysis showed 90 proteins significantly enriched in the averaged results of three *Dis3l*$^{GFP/GFP}$ samples. DIS3L co-precipitated with all nine core exosome proteins with the highest abundance as well as one of the cytoplasmic exosome's cofactors – HBS1L (Figure 1A, B).
The majority of other proteins co-precipitating with the DIS3L are associated with the protein synthesis apparatus. This includes 32 ribosomal proteins and 9 of translation initiation factors of the EIF family as well as GYGIF2, a factor involved in translation quality control. This is also supported by the functional annotation analysis of identified proteins – Molecular Function, Biological Process and Cellular Component gene ontology terms identified with the highest p-value are all associated with the translation process and its regulation (Figure 1C; Supplementary Figures 2-4).
Thus, DIS3L is indeed a subunit of the exosome complex, which in the cytoplasm is associated with the translation machinery.

***Dis3l*$^{-/-}$ embryos die at E6,5**

Having established that DIS3L in mice is indeed a part of the cytoplasmic exosome complex, we aimed to determine its role in shaping mice transcriptome and its potential influence on mice physiology. Using CRISPR/Cas9 system, we generated a mice line bearing c.1258_1259ins169 frameshift mutation in the *Dis3l* gene, leading to disruption of the protein functionality. In our efforts to breed generated *Dis3l*$^{-/-}$ mouse line, we were unable to obtain mice bearing homozygotic mutation. Therefore, we attributed it to embryolethality of homozygotic *Dis3l* mutation. Genotyping embryos at different developmental stages from breeding *Dis3l*$^{+/-}$ males and females further confirmed the lethality.

Preimplantation development was unaffected in the progeny of those matings. The genotype ratio of obtained blastocysts did not differ significantly from the expected mendelian ratio, with 28,95% *Dis3l*$^{-/-}$ and 31,58% *Dis3l*$^{+/+}$ among them (Figure 2A). Lack of the functional cytoplasmic exosome also did not disturb the overall development of a preimplantation embryo. Both wild-type and KO blastocysts had similar mean total cell count (84,8 ± 6,27 and 77,8 ± 8,13 respectively, Figure 2B) and the ratio of all three cell lineages (i.e. trophectoderm, epiblast, and primitive endoderm) present at that stage of development (Figure 2C, D).

We observed the first signs of impaired development early after implantation. On days 6,5 and 7,5 of embryo development, we identified only 14,29% and 16,67% *Dis3l*$^{-/-}$ embryos, respectively, although the entire genotype distribution on both stages remained within a range similar to the mendelian ratio (Figure 2A). However, the lack of functional DIS3L has already affected embryo morphology and developmental potential. While on day 6,5, more than 85% of *Dis3l*$^{+/-}$ and *Dis3l*$^{+/+}$ embryos were properly developed, 2 out of 3 identified KO embryos were smaller and deformed compared to their WT counterparts. This effect became stronger at day 7,5 when all 6 identified KO embryos were highly degenerated and non-viable for further development, without any change in the viability of embryos of other genotypes (Figure 2E, F).

It is concluded that the *Dis3l* KO mutation leads to embryo lethality at the implantation stage, but the cause of it remains to be established.

**Embryolethality of *Dis3l* KO cannot be rescued by supplementing WT extraembryonic tissues**

One of the possibilities for the cause of *Dis3l* KO early embryolethality is the indispensability of *Dis3l* for the development of extraembryonic tissues responsible, among other functions, for proper implantation. To verify this, we constructed chimeric embryos. Individual 4-8-cell embryos (of unknown genotype) from *Dis3l*$^{+/-}$ x *Dis3l*$^{+/-}$ mating were joined with two 2-cell tetraploid WT embryos (Figure 3A). In such a model, tetraploid cells contribute only to the development of extraembryonic tissue – compensating for the lack of DIS3L in KO embryos and potentially rescuing the phenotype. Chimeric embryos were then transferred to recipient females previously mated with vasectomized males. All females were euthanised 18 days after an embryo transfer to account for all implanted embryos and resorption sites in the uterus.

Out of the total of 98 chimeric embryos transferred to recipient females, 63% of embryos were successfully implanted in the uterus. However, only 44% of those embryos developed to term, and the other 56% (35 out of 62) died and were subsequently resorbed, leaving only implantation sites (Figure 3B). Genotyping of developed embryos revealed that our attempt did not produce any *Dis3l*$^{-/-}$ embryos, resulting in 67% of *Dis3l*$^{+/-}$ and 33% of *Dis3l*$^{+/+}$ pups developed to term(Figure 3C), all of them without any morphological signs of disturbed development. These results exclude the notion that the implantation process and, more broadly, the development of trophectoderm is the only cause of embryolethality.

***Dis3l* KO does not affect cell viability**

Having established that *Dis3l*$^{-/-}$ embryo cannot develop properly after implantation and degenerates soon after, even with DIS3L supplemented in extraembryonic tissues, we assumed that it may be related to changes at the transcriptome level at this developmental stage. Thus, we next focused on analysing the effect of *Dis3l* KO mutation on the developmental potential of the preimplantation embryo.

To test the potential of the inner cell mass (ICM) of the preimplantation embryo to develop the embryo body, we first derived ES cell lines from blastocysts obtained from *Dis3l*$^{+/-}$ x *Dis3*$^{+/-}$ matings. We observed that *Dis3l*$^{-/-}$ ES lines are obtained with a slightly lower ratio compared to *Dis3l*$^{+/-}$ and *Dis3l*$^{+/+}$ embryos. Of 28 blastocysts from which we derived ES cells, 14,29% were *Dis3l*$^{-/-}$ (Figure 4A). Notably, both KO and WT ES cells were also able to form embryoid bodies. However, out of three tested KO cell lines, two produced embryoid bodies that, on days 2 and 5 of culture, displayed a decrease in expression of *Foxa2* and *Tbxt* transcripts coding proteins responsible for differentiation of endodermal and mesodermal organs, respectively. In addition, all three lines also displayed an atypical expression pattern of Pax6, involved in ectoderm development, with a spike of transcript accumulation on day 2 of culture and a drop below starting level on day 5 (Figure 4B).

To assess the state and possible changes in the transcriptome of ES cells lacking functional cytoplasmic exosome, we have performed a total RNA sequencing experiment on the Illumina platform, using RNA isolated from initial cultures. Comparing three *Dis3l*$^{-/-}$ and three *Dis3l*$^{+/+}$ ES cells lines, we identified 28 differentially expressed transcripts (out of all 13009 identified). 23 were downregulated and 5 upregulated (differential expression analysis, padj < 0,05) in KO cells. Of those, 20 had at least 2-fold expression difference ($|\log_2 \text{foldchange}| > 1$) (Figure 4C). Gene ontology analysis revealed developmental and differentiation processes as the most enriched terms among those deregulated transcripts (Figure 4D). Among those downregulated, 3 genes (*Pramel6*, *Pramel7* and *Lama3*) are known to play a significant role in the regulation of ES cells and embryo development. *Pramel6* and *Pramel7* are essential for maintaining pluripotency of ES cells (29) and *Lama3* is a component of basal membrane during postimplantation development although it was only identified in decidua and Reichert's membrane (30). To confirm their downregulation in KO samples, we compared the level of their transcripts in three WT and three KO samples by RT-qPCR. High variability of results in each biological sample was observed. For every gene tested, some samples showed upregulation, while others – downregulation of a given transcript (Figure 4E). This may be due to the process of differentiation being initiated at the time of cell collection.

It is concluded that although it is possible to derive and maintain a culture of *Dis3l* KO ES cells and those cells have very subtle changes on the transcriptome level, they display some abnormalities affecting differentiation factors which are difficult to explain mechanistically.

***Dis3l* KO mutation negatively affects global protein production in blastocysts.**

Since the analysis of ES cells did not reveal the cause of the lethality-inducing effect of *Dis3l* KO mutation, we turned back to embryos at the blastocyst stage. Using a modified single-cell RNA sequencing protocol, we sequenced RNA of three *Dis3l*$^{-/-}$ and three *Dis3l*$^{+/+}$ blastocysts. Similarly to ES cells sequencing, we identified only a small number of differentially expressed transcripts (15 out of 17368 all identified). This time, however, only 2 of them were downregulated, and 13 – were upregulated (differential expression analysis, padj < 0,05). All these transcripts were also changed at least 2-fold ($|\log_2 \text{foldchange}| > 1$) (Figure 5A), and the upregulated ones may represent direct substrates of the cytoplasmic exosome. Out of four Biological Process terms that identified deregulated transcripts were annotated by gene ontology analysis to, three regarded cell death (the same seven transcripts annotated to each). Additionally, three transcripts were successfully annotated to five different Human Phenotype terms (one transcript annotated to five terms, one to four and one to three), out of which four are related to developmental disorders (Figure 5B).

Notably, two genes upregulated in *Dis3l*$^{-/-}$ blastocysts are directly involved in preimplantation and early postimplantation development. Testin (*Tes*) is known for interacting with the cytoskeleton and affects actin dynamics to facilitate cell movements and migration (31,32). In blastocysts, it's highly expressed in ICM and polar trophectoderm and after implantation at E6,5 in epiblast, anterior visceral endoderm precursors and extraembryonic ectoderm (33). SLC37A2 is a glucose-6-phosphate antiporter responsible for its transport from cytoplasm to ER lumen (34) – one of the steps in the metabolism of the primary blastocyst energy source – glucose (35,36). RT-qPCR analysis of those two transcripts' levels from seven *Dis3l*$^{-/-}$ and ten *Dis3l*$^{+/+}$ blastocysts revealed that while *Slc37a2* transcript was accumulated in most of KO samples, with mean fold change of 1,31 ± 0,22, mean fold change of Testin-

coding transcripts was only 1,05 ± 0,2 (Figure 5C). Thus, we focused on *Slc37a2*. Surprisingly immunofluorescence staining of SLC37A2 protein in blastocysts revealed that an elevated amount of transcript did not produce higher protein levels. On the contrary, standardized z-score of SLC37A2 protein amount in *Dis3l* KO blastocysts significantly decreased compared to WT embryos (mean z-score -0,35 ± 0,10 and 0,57 ± 0,28, respectively) (Figure 5D, E). All this suggests that the lethality of *Dis3l* KO is not related to changes in the transcript levels.

Decreased levels of SLC37A2 protein despite increased mRNA expression and the fact that cytoplasmic exosome is connected with the cytoplasmic quality control pathways may suggest global problems with protein synthesis resulting from stress responses. To measure the global protein synthesis levels, we performed metabolic labelling with clickable methionine analogue L-Azidohomoalanine (AHA). Quantification of AHA incorporation in newly synthesised proteins in embryos revealed the negative impact of *Dis3l* KO mutation on global protein synthesis in the preimplantation embryo (mean z-score -0,41 ± 0,22 and 0,23 ± 0,18 for WT embryos; Figure 5F, G).

It is concluded that *Dis3l* KO leads to the inhibition of protein synthesis in early embryos with a substantial effect on mRNA levels

## Discussion

In this paper, we studied the role of mammalian DIS3L processive exoribonuclease at the organismal level. DIS3L, which co-precipitates with all nine core exosome proteins (8), is essential for mouse embryo development. *Dis3l* KO mutation causes severe embryo degeneration and death between day 6,5 and 7,5 of its development. This is due not only to the implantation process mediated by extraembryonic tissue, but mostly to the changes in subcellular processes that occur before implantation. While KO preimplantation embryos do not display any changes in morphology or differentiation of first embryonic cell lineages and ability to produce ES cells from ICM, the translation rate was reduced, despite small changes at the transcriptome level. This presents a picture of the highly deregulated state of the early embryo at the level of subcellular processes.

Despite the disruption of global protein production, *Dis3l*$^{-/-}$ embryos are still able to implant in the uterus and to function there for a brief period, but their fate is most certainly determined at the preimplantation stage. The state of both relies heavily on the mRNA and translation quality control pathways in the cell, in many of which DIS3L, as a catalytic subunit of cytoplasmic RNA exosome complex, is intrinsically involved (most notably no-go (NGD), nonsense-mediated (NMD) and non-stop (NSD) decay pathways of RNA degradation (37-39)). Thus, the upregulation of transcripts observed by us in blastocysts with mutated *Dis3l* would rather stem not from the elevated expression of certain genes but rather from the accumulation of aberrant transcripts that the functional exosome would normally degrade. As for proteome, translational machinery could still be recruited to those lingering transcripts but produced protein products would be incomplete or incorrect and, in return, would also have to be degraded by translation control mechanisms. Many of those mechanisms are triggered by ribosome stalling and collision events on abnormal mRNA particles. In 2020, two groups independently showed that SKIV2L in Ski complex, responsible for unwinding and feeding RNA molecules into the exosome channel (40-42) is required for the recruitment of the exosome in mRNA quality control pathways through its interaction with stalled ribosomes. This allows for their extraction through degradation of stalling-causing mRNA particles, making those ribosomes available for dissociation and recycling by Pelota/Hbs1/ABCE1 complex (43). Hence, the debilitating effect of the cell's inability to degrade those problematic mRNAs would be two-fold: occurring ribosome collision events could not be properly resolved, and the number of stalling-causing mRNA particles would grow in time, causing new collision events.

While local mechanisms recruited to colliding ribosomes counteract negative consequences of these events, like EDF1 and GIGYF2 working together to inhibit translation from mRNA particles on which collision occurred (44,45), they can be overloaded by the global scale of the occurring problems. In such a case, more general stress response mechanisms are triggered, leading to global translation initiation block by the action of GCN2 and eIF2α phosphorylation or, in a case of stress being

unresolved by the cell, apoptosis activated by MAPKKK ZAKα (46). As we present, this is the case for *Dis3l*$^{-/-}$ preimplantation embryos, where the global protein amount is lowered despite the accumulation of multiple transcripts. Additionally, recently SKI2VL was also shown to be essential for mouse development, with whole-body deletion leading to embryolethality before day E13,5 (47). Thus, we propose that DIS3L, while dispensable for cell viability, is essential for embryo development as a part of RNA quality control response for translational stress, with its mutation possibly causing translation inhibition across the preimplantation embryo, and, further on, due to accumulation of the stress-inducing events, wide apoptosis after the implantation, leading to embryo death after day E6,5.

## Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consotrium via the PRIDE (48) partner repository with the dataset identifier PXD038745.
RNA sequencing data been deposited in NCBI's Gene Expression Omnibus (49) and are accessible through GEO Series accession number GSE220800.
Other numerical data underlying results presented are available in the article and its online supplementary materials.

## Funding

This work was supported by National Science Centre (grant numbers UMO-2013/10/M/NZ4/00299; UMO-2016/22/A/NZ4/00380), Foundation for Polish Science (grunt number TEAM TECH CORE FACILITY/2017-4/5) and European Union's Horizon 2020 research and innovation program (grant agreement no. 810425).

## Conflict of Interest

Authors report no conflict of interest.

## Author Contribution

M.B. performed embryo genotyping, blastocyst and ES cells qPCR experiments and prepared RNA-sequencing libraries, M.B. and M.S. constructed and transferred chimeric embryos, A.C. derived ES cells and performed EB experiments, M.B and W.A. performed SLC37A2 and AHA labelling experiments, S.M. performed IP experiment, A.H.-O. and T.K. performed differential expression analyses, D.C. performed and analysed mass spectrometry, D.A. performed sequencing experiments, J.G. and E.B. produced all mice lines, M.B. and A.D. wrote the manuscript, A.D. conceived the study.

## Acknowledgments

We thank Andrzej Dziembowski lab members for their support and fruitful discussions, Anna Ciemerych for the support in the experiments involving ES cells and postimplantation embryos, Olga Gewartowska for multiple suggestions and critical reading of the manuscript, Aleksandra Brouze for critical and proofreading of the manuscript and all members of Genome Engineering Unit of International Institute of Molecular and Cell Biology in Warsaw for maintenance of animal colony and animal genotyping. NGS was performed thanks to Genomics Core Facility CeNT UW (RRID:SCR_022718), using NovaSeq 6000 platform financed by Polish Ministry of Science and Higher Education (decision no. 6817/IA/SP/2018 of 2018-04-10).

## References

1. Bashkirov, V.I., Scherthan, H., Solinger, J.A., Buerstedde, J.M. and Heyer, W.D. (1997) A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. *Journal of Cell Biology*, **136**, 761-773.
2. Wang, M.S. and Pestov, D.G. (2011) 5'-end surveillance by Xrn2 acts as a shared mechanism for mammalian pre-rRNA maturation and decay. *Nucleic Acids Research*, **39**, 1811-1822.

3. Tomecki, R., Kristiansen, M.S., Lykke-Andersen, S., Chlebowski, A., Larsen, K.M., Szczesny, R.J., Drazkowska, K., Pastula, A., Andersen, J.S., Stepien, P.P. *et al.* (2010) The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *EMBO J*, **29**, 2342-2357.
4. Staals, R.H., Bronkhorst, A.W., Schilders, G., Slomovic, S., Schuster, G., Heck, A.J., Raijmakers, R. and Pruijn, G.J. (2010) Dis3-like 1: a novel exoribonuclease associated with the human exosome. *EMBO J*, **29**, 2358-2367.
5. Lubas, M., Damgaard, C.K., Tomecki, R., Cysewski, D., Jensen, T.H. and Dziembowski, A. (2013) Exonuclease hDIS3L2 specifies an exosome-independent 3 '-5 ' degradation pathway of human cytoplasmic mRNA. *Embo Journal*, **32**, 1855-1868.
6. Piwowarski, J., Grzechnik, P., Dziembowski, A., Dmochowska, A., Minczuk, M. and Stepien, P.P. (2003) Human polynucleotide phosphorylase, hPNPase, is localized in mitochondria. *J Mol Biol*, **329**, 853-857.
7. Januszyk, K., Liu, Q.S. and Lima, C.D. (2011) Activities of human RRP6 and structure of the human RRP6 catalytic domain. *Rna*, **17**, 1566-1577.
8. Liu, Q., Greimann, J.C. and Lima, C.D. (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell*, **127**, 1223-1237.
9. Pirouz, M., Du, P., Munafo, M. and Gregory, R.I. (2016) Dis3l2-Mediated Decay Is a Quality Control Pathway for Noncoding RNAs. *Cell Reports*, **16**, 1861-1873.
10. Astuti, D., Morris, M.R., Cooper, W.N., Staals, R.H.J., Wake, N.C., Fews, G.A., Gill, H., Gentle, D., Shuib, S., Ricketts, C.J. *et al.* (2012) Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. *Nature Genetics*, **44**, 277-U275.
11. Chang, H.M., Triboulet, R., Thornton, J.E. and Gregory, R.I. (2013) A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. *Nature*, **497**, 244-+.
12. Hunter, R.W., Liu, Y.J., Manjunath, H., Acharya, A., Jones, B.T., Zhang, H., Chen, B.B., Ramalingam, H., Hammer, R.E., Xie, Y. *et al.* (2018) Loss of Dis3l2 partially phenocopies Perlman syndrome in mice and results in upregulation of Igf2 in nephron progenitor cells. *Gene Dev*, **32**, 903-908.
13. Szczepinska, T., Kalisiak, K., Tomecki, R., Labno, A., Borowski, L.S., Kulinski, T.M., Adamska, D., Kosinska, J. and Dziembowski, A. (2015) DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts. *Genome Research*, **25**, 1622-1633.
14. Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adli, M. *et al.* (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature*, **471**, 467-472.
15. Lohr, J.G., Stojanov, P., Carter, S.L., Cruz-Gordillo, P., Lawrence, M.S., Auclair, D., Sougnez, C., Knoechel, B., Gould, J., Saksena, G. *et al.* (2014) Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell*, **25**, 91-101.
16. Weissbach, S., Langer, C., Puppe, B., Nedeva, T., Bach, E., Kull, M., Bargou, R., Einsele, H., Rosenwald, A., Knop, S. *et al.* (2015) The molecular spectrum and clinical impact of DIS3 mutations in multiple myeloma. *Br J Haematol*, **169**, 57-70.
17. Petit, F.G., Jamin, S.P., Kernanec, P.Y., Becker, E., Halet, G. and Primig, M. (2022) EXOSC10/Rrp6 is essential for the eight-cell embryo/morula transition. *Developmental Biology*, **483**, 58-65.
18. Gewartowska, O., Aranaz-Novaliches, G., Krawczyk, P.S., Mroczek, S., Kusio-Kobialka, M., Tarkowski, B., Spoutil, F., Benada, O., Kofronova, O., Szwedziak, P. *et al.* (2021) Cytoplasmic polyadenylation by TENT5A is required for proper bone formation. *Cell Rep*, **35**, 109015.
19. Robertson, E.J. (1987) *Teratocarcinomas and embryonic stem cells : a practical approach*. IRL Press, Oxford ; Washington, D.C.
20. Tamm, C., Galito, S.P. and Anneren, C. (2013) A Comparative Study of Protocols for Mouse Embryonic Stem Cell Culturing. *Plos One*, **8**.

21. Mroczek, S., Chlebowska, J., Kulinski, T.M., Gewartowska, O., Gruchota, J., Cysewski, D., Liudkovska, V., Borsuk, E., Nowis, D. and Dziembowski, A. (2017) The non-canonical poly(A) polymerase FAM46C acts as an onco-suppressor in multiple myeloma. *Nat Commun*, **8**, 619.
22. Truett, G.E., Heeger, P., Mynatt, R.L., Truett, A.A., Walker, J.A. and Warman, M.L. (2000) Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques*, **29**, 52-+.
23. Artus, J., Vandormael-Pournin, S., Frodin, M., Nacerddine, K., Babinet, C. and Cohen-Tannoudji, M. (2005) Impaired mitotic progression and preimplantation lethality in mice lacking OMCG1, a new evolutionarily conserved nuclear protein. *Mol Cell Biol*, **25**, 6289-6302.
24. Hennig, B.P., Velten, L., Racke, I., Tu, C.S., Thoms, M., Rybin, V., Besir, H., Remans, K. and Steinmetz, L.M. (2018) Large-Scale Low-Cost NGS Library Preparation Using a Robust Tn5 Purification and Tagmentation Protocol. *G3 (Bethesda)*, **8**, 79-89.
25. Picelli, S., Bjorklund, A.K., Reinius, B., Sagasser, S., Winberg, G. and Sandberg, R. (2014) Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res*, **24**, 2033-2040.
26. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.
27. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**.
28. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H. and Vilo, J. (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, **47**, W191-W198.
29. Casanova, E.A., Shakhova, O., Patel, S.S., Asner, I.N., Pelczar, P., Weber, F.A., Graf, U., Sommer, L., Burki, K. and Cinelli, P. (2011) Pramel7 mediates LIF/STAT3-dependent self-renewal in embryonic stem cells. *Stem Cells*, **29**, 474-485.
30. Gersdorff, N., Muller, M., Otto, S., Poschadel, R., Hubner, S. and Miosge, N. (2005) Basement membrane composition in the early mouse embryo day 7. *Dev Dyn*, **233**, 1140-1148.
31. Garvalov, B.K., Higgins, T.E., Sutherland, J.D., Zettl, M., Scaplehorn, N., Kocher, T., Piddini, E., Griffiths, G. and Way, M. (2003) The conformational state of Tes regulates its zyxin-dependent recruitment to focal adhesions. *Journal of Cell Biology*, **161**, 33-39.
32. Coutts, A.S., MacKenzie, E., Griffith, E. and Black, D.M. (2003) TES is a novel focal adhesion protein with a role in cell spreading. *J Cell Sci*, **116**, 897-906.
33. Crompton, L.A., Du Roure, C. and Rodriguez, T.A. (2007) Early embryonic expression patterns of the mouse Flamingo and Prickle orthologues. *Dev Dyn*, **236**, 3137-3143.
34. Pan, C.J., Chen, S.Y., Jun, H.S., Lin, S.R., Mansfield, B.C. and Chou, J.Y. (2011) SLC37A1 and SLC37A2 are phosphate-linked, glucose-6-phosphate antiporters. *PLoS One*, **6**, e23157.
35. Gardner, D.K. and Leese, H.J. (1986) Non-invasive measurement of nutrient uptake by single cultured pre-implantation mouse embryos. *Hum Reprod*, **1**, 25-27.
36. Martin, K.L. and Leese, H.J. (1995) Role of glucose in mouse preimplantation embryo development. *Mol Reprod Dev*, **40**, 436-443.
37. van Hoof, A., Frischmeyer, P.A., Dietz, H.C. and Parker, R. (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science*, **295**, 2262-2264.
38. Mitchell, P. and Tollervey, D. (2003) An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3 '-> 5 ' degradation. *Molecular Cell*, **11**, 1405-1413.
39. Takahashi, S., Araki, Y., Sakuno, T. and Katada, T. (2003) Interaction between Ski7p and Upf1p is required for nonsense-mediated 3 '-to-5 ' mRNA decay in yeast. *Embo Journal*, **22**, 3951-3959.
40. Halbach, F., Reichelt, P., Rode, M. and Conti, E. (2013) The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell*, **154**, 814-826.
41. Zhu, B., Mandal, S.S., Pham, A.D., Zheng, Y., Erdjument-Bromage, H., Batra, S.K., Tempst, P. and Reinberg, D. (2005) The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes Dev*, **19**, 1668-1673.

42. Brown, J.T., Bai, X.X. and Johnson, A.W. (2000) The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo. *Rna*, **6**, 449-457.
43. Pisareva, V.P., Skabkin, M.A., Hellen, C.U.T., Pestova, T.V. and Pisarev, A.V. (2011) Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. *Embo Journal*, **30**, 1804-1817.
44. Sinha, N.K., Ordureau, A., Best, K., Saba, J.A., Zinshteyn, B., Sundaramoorthy, E., Fulzele, A., Garshott, D.M., Denk, T., Thoms, M. *et al.* (2020) EDF1 coordinates cellular responses to ribosome collisions. *Elife*, **9**.
45. Juszkiewicz, S., Slodkowicz, G., Lin, Z.W., Freire-Pritchett, P., Peak-Chew, S.Y. and Hegde, R.S. (2020) Ribosome collisions trigger cis-acting feedback inhibition of translation initiation. *Elife*, **9**.
46. Wu, C.C.C., Peterson, A., Zinshteyn, B., Regot, S. and Green, R. (2020) Ribosome Collisions Trigger General Stress Responses to Regulate Cell Fate. *Cell*, **182**, 404-+.
47. Yang, K., Han, J., Asada, M., Gill, J.G., Park, J.Y., Sathe, M.N., Gattineni, J., Wright, T., Wysocki, C.A., de la Morena, M.T. *et al.* (2022) Cytoplasmic RNA quality control failure engages mTORC1-mediated autoinflammatory disease. *J Clin Invest*, **132**.
48. Perez-Riverol, Y., Bai, J.W., Bandla, C., Garcia-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M. *et al.* (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research*, **50**, D543-D552.
49. Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*, **30**, 207-210.

## Figures captions

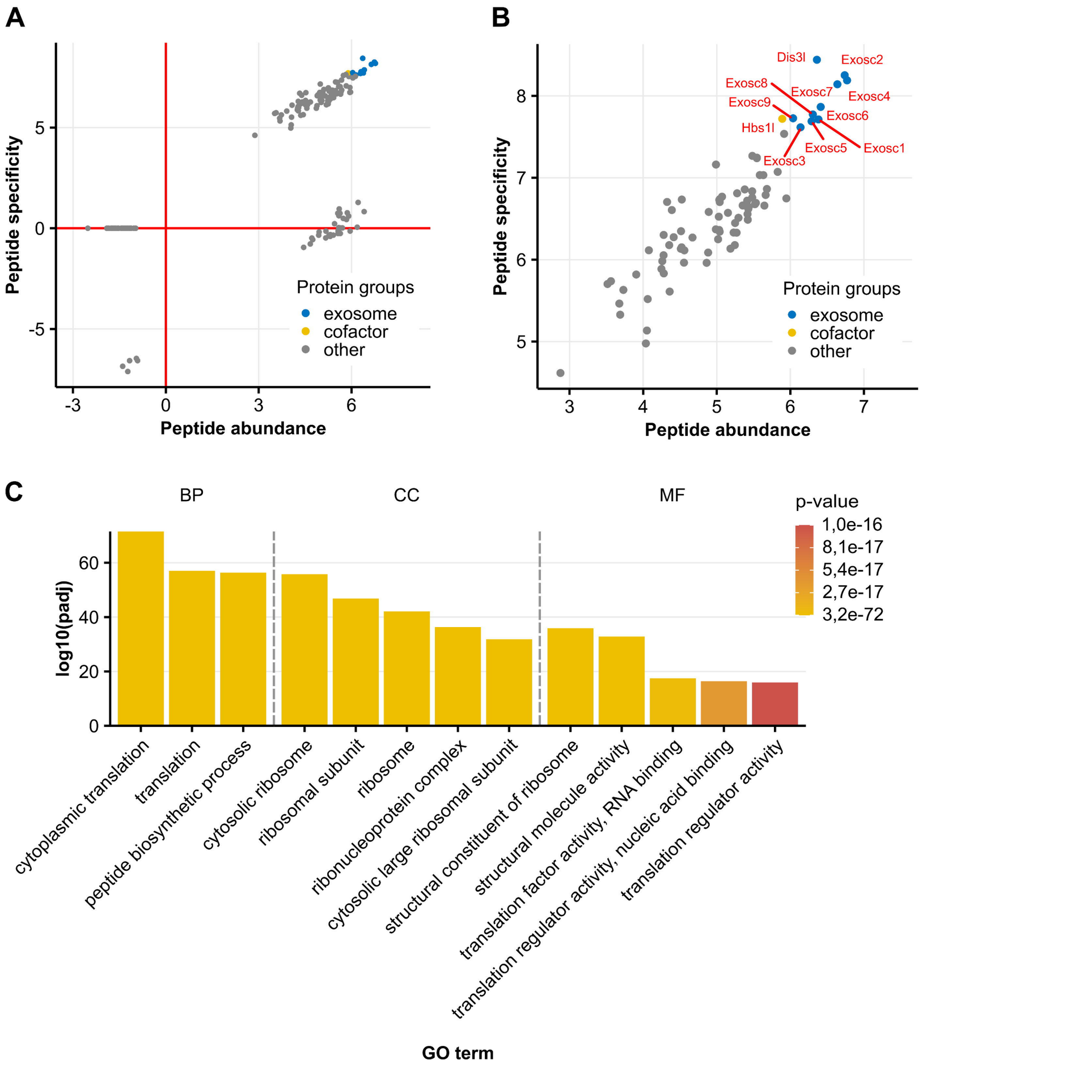
**Figure 1.** (**A**) Distribution of proteins identified by mass spectrometry analysis of DIS3L-GFP co-precipitated peptides from *Dis3l*$^{GFP/GFP}$ protein liver extracts presented as peptide abundance relative to peptide specificity. (**B**) All nine mammalian core exosome proteins co-precipitated with DIS3L-GFP as the most specific and abundant proteins detected in the experiment. (**C**) Functional enrichment analysis of proteins co-precipitated with DIS3L-GFP. Only several terms with the highest p-value are presented. Identified terms are grouped by data source sub-groups: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC). A full list of identified terms is available in Supplementary Figures 2-4.

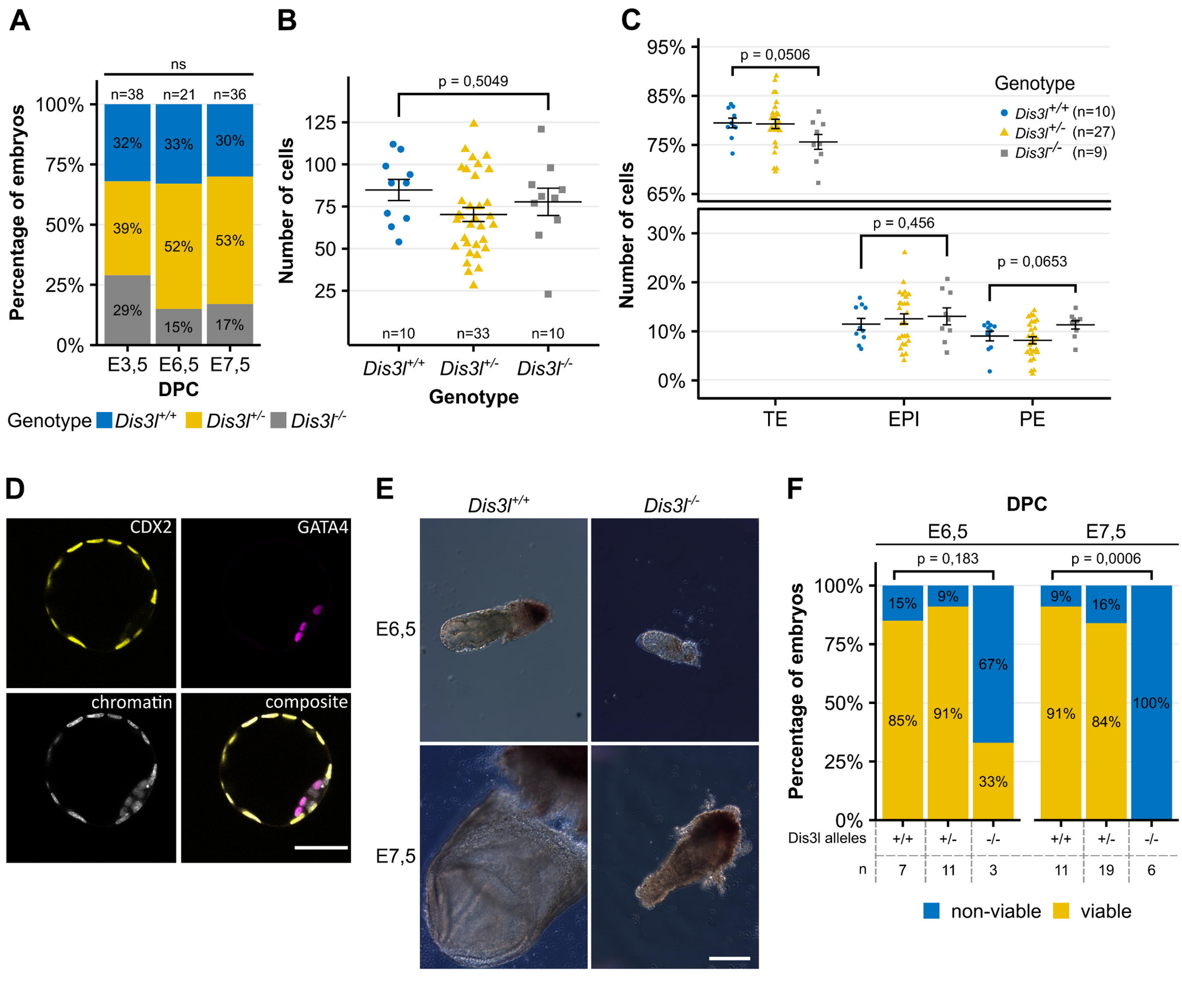
**Figure 2.** (**A**) Genotypes distribution in embryos of different developmental stages obtained from *Dis3l*$^{+/-}$ x *Dis3l*$^{+/-}$ matings. (**B**) Total cell count of blastocysts in relation to the genotype. (**C**) Percentage ratio of trophectoderm (TE), epiblast (EPI) and primitive endoderm (PE) in blastocysts in relation to the genotype. (**D**) Example of staining used to determine the number of differentiated cells in blastocyst. CDX2 marker of trophectoderm in yellow, GATA4 marker of primitive endoderm in magenta, chromatin in grey. Number of epiblast cells was determined by subtracting the number of CDX2 and GATA4 positive cells from total cell number. Scale bar = 50 um. (**E**) Example of incorrectly developing *Dis3l*$^{-/-}$ embryos at E6,5 and E7,5 DPC compared to valid *Dis3l*$^{+/+}$ embryos. Scale bar = 200 µm. (**F**) Ratio of viable to nonviable embryos of different genotypes at E6,5 and E7,5 DPC. Individual data points represent single blastocysts analysed with black summary bars representing mean value and error bars representing SEM.

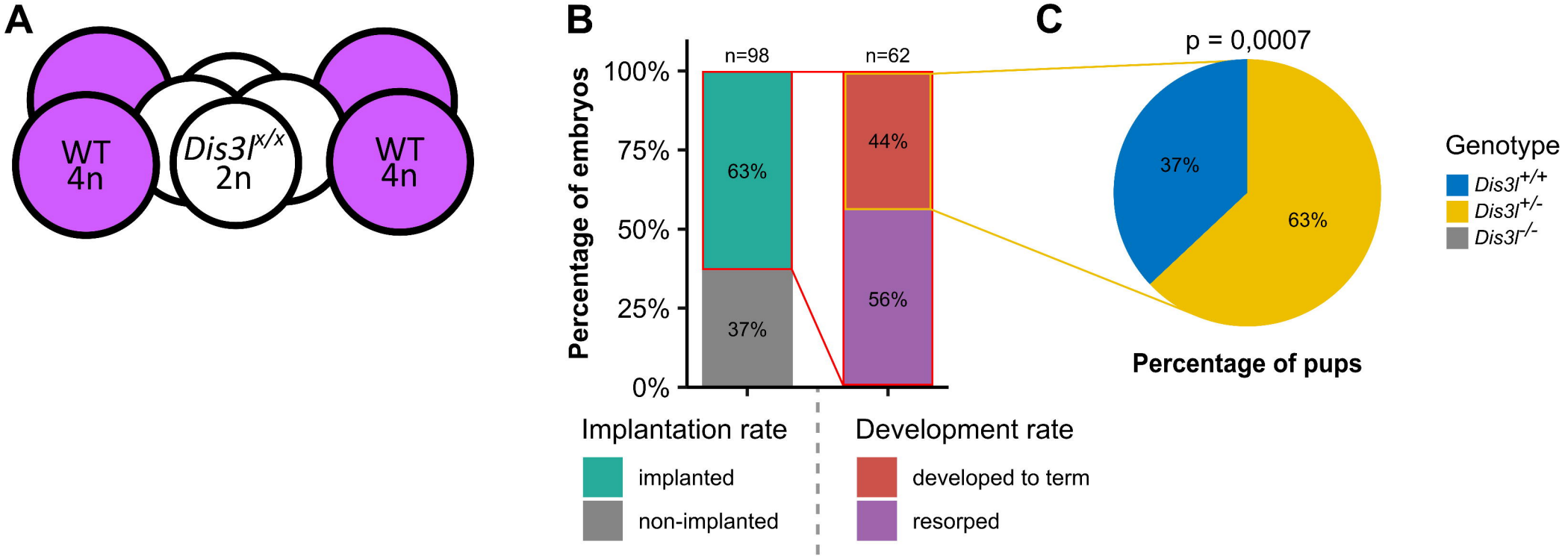
**Figure 3.** (**A**) Schematic representation of chimeric embryo construction. Diploid 4-cell embryos from *Dis3l*$^{+/-}$ x *Dis3l*$^{+/-}$ mating (white) were covered with two previously prepared tetraploid 2-cell wild-type embryos (magenta). (**B**) Chimeric embryos implantation rate after transfer to recipient females and percentage of those implanted embryos that fully developed to term. (**C**) Genotype distribution of chimeric pups developed to term. No *Dis3l*$^{-/-}$ embryos developed fully.

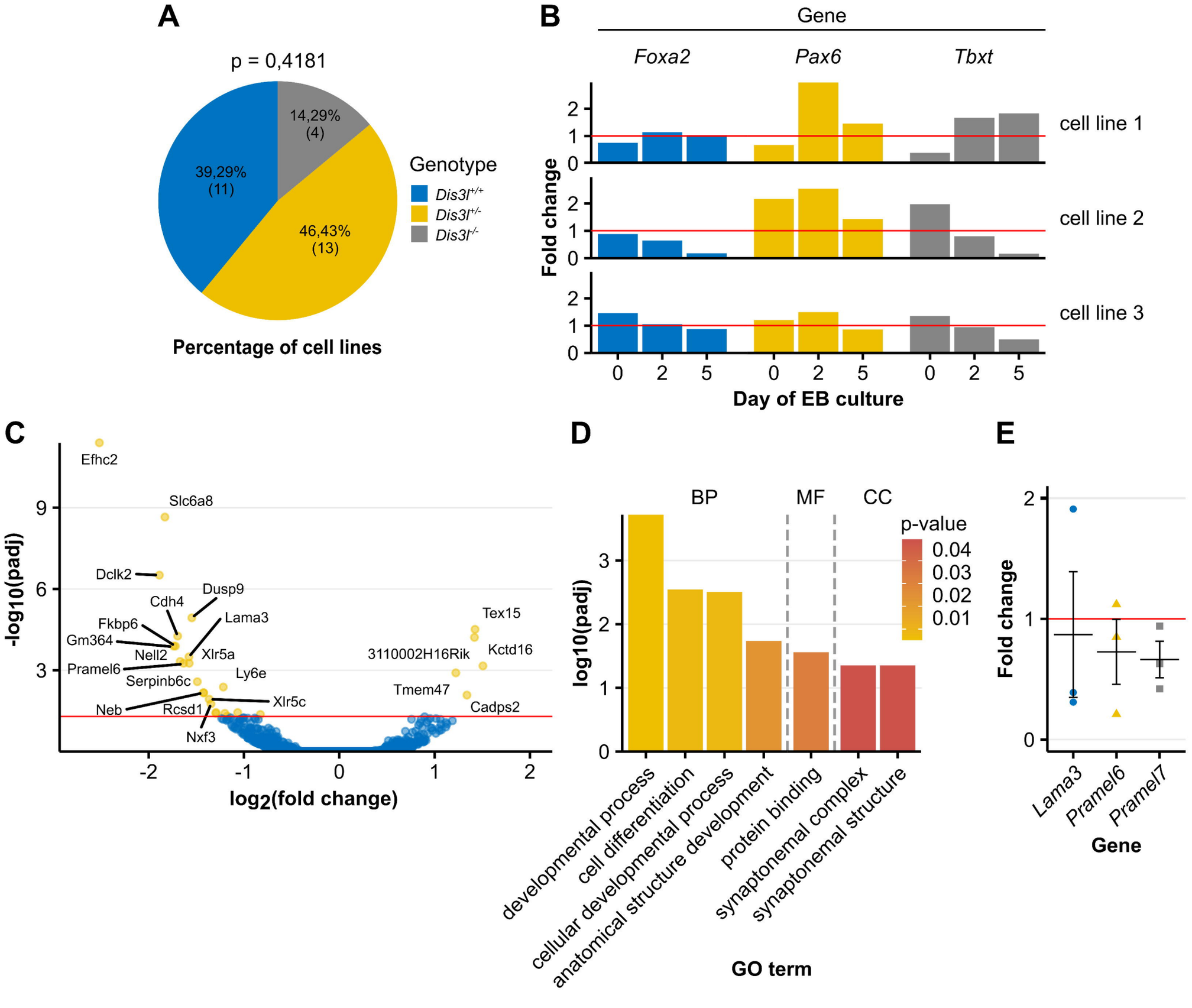
**Figure 4.** (**A**) Genotype distribution of ES cell lines derived from blastocysts from *Dis3l*$^{+/-}$ x *Dis3l*$^{+/-}$ matings. (**B**) Transcript level fold change of three germ layer-specific differentiation factors: *Foxa2*, *Pax6* and *Tbxt*, in embryoid bodies through 5 days of culture in three obtained *Dis3l*$^{-/-}$ ES cell lines related to mean result of three WT cell lines, obtained by RT-qPCR experiment. Red horizontal line marks fold change of 1, meaning no change compared to WT. (**C**) Differential expression analysis of RNA sequencing results of three *Dis3l*$^{-/-}$ and three *Dis3l*$^{+/+}$ ES cell lines. Positive and negative $\log_2$(fold change) value represents up- and downregulation, respectively, of transcript in KO cells. Red horizontal line marks adjusted p-value of 0,05 with transcripts above it considered significantly changed in KO cells compared to WT ones. (**D**) Functional enrichment analysis of significantly deregulated transcripts identified in ES cells' RNA sequencing. Identified terms are grouped by data source sub-groups: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC). (**E**) Fold change of three of significantly downregulated transcripts identified in ES cells' RNA sequencing in *Dis3l*$^{-/-}$ ES cells related to mean result of three WT cell lines, obtained by RT-qPCR. Each data point represents fold change for one KO cell lines. Horizontal red line marks fold change of 1, meaning no change. Summary black bar represents mean fold change of a given transcript and error bars represent SEM.

**Figure 5.** (**A**) Differential expression analysis of RNA sequencing results of three *Dis3l*$^{-/-}$ and three *Dis3l*$^{+/+}$ blastocysts. Positive and negative $\log_2$(fold change) value means up- and downregulation, respectively, of transcript in KO embryos. Red horizontal line marks adjusted p-value of 0,05 with transcripts above it considered significantly changed in KO embryos compared to WT ones. (**B**) Functional enrichment analysis of significantly deregulated transcripts identified in blastocysts RNA sequencing. Identified terms are grouped by data source sub-groups: Biological Process (BP) and Human Phenotype (HP). (**C**) Fold change of two of significantly downregulated transcripts identified in blastocyst's RNA sequencing of *Dis3l*$^{-/-}$ embryos, obtained by RT-qPCR. Each data point represents fold change for one KO blastocyst. Horizontal red line marks fold change of 1, meaning no change. (**D**) Example of SLC37A2 protein immunofluorescence staining used to measure protein level in blastocysts. (**E**) Normalized z-score value of SLC37A2 protein level in blastocysts depending on genotype. Each data point represents one blastocyst. (**F**) Example of protein synthesis labelling with methionine analogue AHA used to determine the amount of newly produced peptides in blastocysts. (**G**) Normalized z-score value of detected AHA-containing proteins levels in blastocysts depending on genotype. Each data point represents one blastocyst. Scale bars = 50 µm. Black summary bars represent mean value and error bars represent SEM.

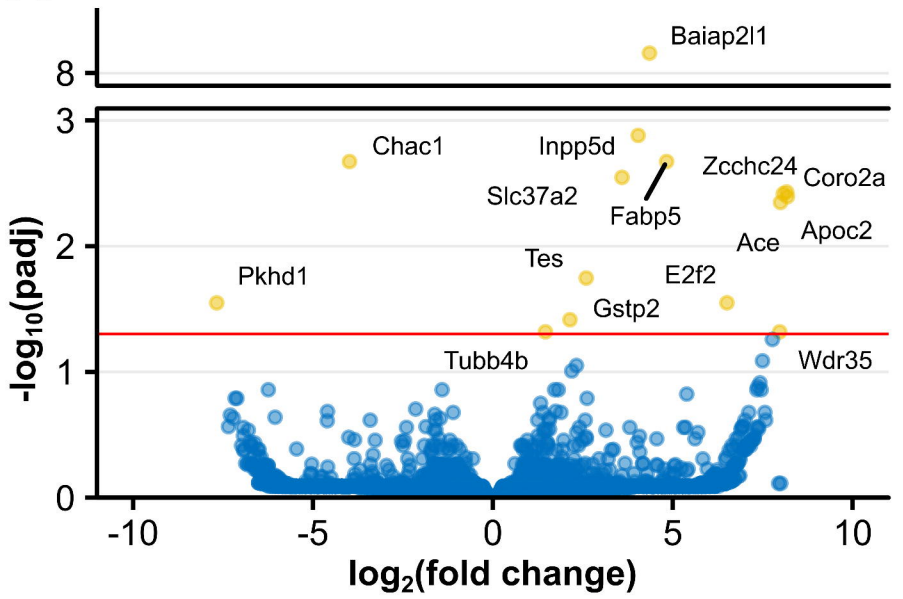

A
Peptide specificity
Peptide abundance
Protein groups
exosome
cofactor
other
B
Peptide specificity
Peptide abundance
Dis3l
Exosc2
Exosc8
Exosc7
Exosc4
Exosc9
Exosc6
Hbs1l
Exosc5
Exosc1
Exosc3
Protein groups
exosome
cofactor
other
C
BP
CC
MF
p-value
1,0e-16
8,1e-17
5,4e-17
2,7e-17
3,2e-72
log10(padj)
cytoplasmic translation
translation
peptide biosynthetic process
cytosolic ribosome
ribosomal subunit
ribosome
ribonucleoprotein complex
cytosolic large ribosomal subunit
structural constituent of ribosome
structural molecule activity
translation factor activity, RNA binding
translation regulator activity, nucleic acid binding
translation regulator activity
GO term


A
ns
n=38
n=21
n=36
100%
75%
50%
25%
0%
Percentage of embryos
32%
39%
29%
33%
52%
15%
30%
53%
17%
E3,5
E6,5
E7,5
DPC
Genotype Dis3l+/+ Dis3l+/- Dis3l-/-
B
p = 0,5049
125
100
75
50
25
Number of cells
n=10
n=33
n=10
Dis3l+/+
Dis3l+/-
Dis3l-/-
Genotype
C
95%
85%
75%
65%
p = 0,0506
Genotype
Dis3l+/+ (n=10)
Dis3l+/- (n=27)
Dis3l-/- (n=9)
30%
20%
10%
0%
p = 0,456
p = 0,0653
Number of cells
TE
EPI
PE
D
CDX2
GATA4
chromatin
composite
E
Dis3l+/+
Dis3l-/-
E6,5
E7,5
F
DPC
E6,5
E7,5
p = 0,183
p = 0,0006
100%
75%
50%
25%
0%
Percentage of embryos
15%
85%
9%
91%
67%
33%
9%
91%
16%
84%
100%
Dis3l alleles
+/+
+/-
-/-
+/+
+/-
-/-
n
7
11
3
11
19
6
non-viable
viable

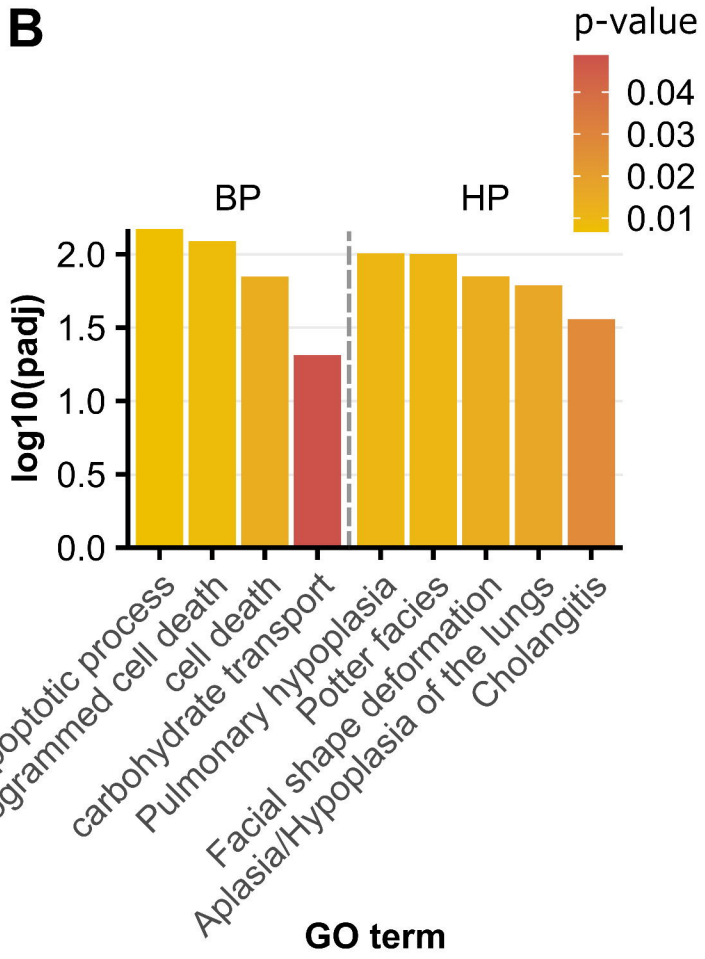

A
WT
4n
Dis3l^x/x
2n
WT
4n
B
n=98
n=62
100%
75%
50%
25%
0%
Percentage of embryos
63%
37%
44%
56%
Implantation rate
implanted
non-implanted
Development rate
developed to term
resorped
C
p = 0,0007
37%
63%
Percentage of pups
Genotype
Dis3l+/+
Dis3l+/-
Dis3l-/-


A
p = 0,4181
14,29%
(4)
39,29%
(11)
46,43%
(13)
Genotype
Dis3l+/+
Dis3l+/-
Dis3l-/-
Percentage of cell lines
B
Gene
Foxa2
Pax6
Tbxt
Fold change
cell line 1
cell line 2
cell line 3
Day of EB culture
C
Efhc2
Slc6a8
Dclk2
Dusp9
Cdh4
Lama3
Fkbp6
Gm364
Nell2
Xlr5a
Pramel6
Serpinb6c
Ly6e
Neb
Rcsd1
Xlr5c
Nxf3
Tex15
Kctd16
3110002H16Rik
Tmem47
Cadps2
-log10(padj)
log2(fold change)
D
BP
MF
CC
p-value
log10(padj)
developmental process
cell differentiation
cellular developmental process
anatomical structure development
protein binding
synaptonemal complex
synaptonemal structure
GO term
E
Fold change
Lama3
Pramel6
Pramel7
Gene

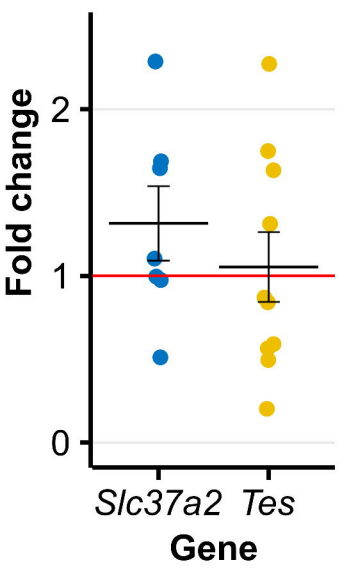
A


Baiap2l1
Chac1
Inpp5d
Zcchc24
Coro2a
Slc37a2
Fabp5
Apoc2
Ace
Tes
E2f2
Pkhd1
Gstp2
Tubb4b
Wdr35
-log10(padj)
log2(fold change)

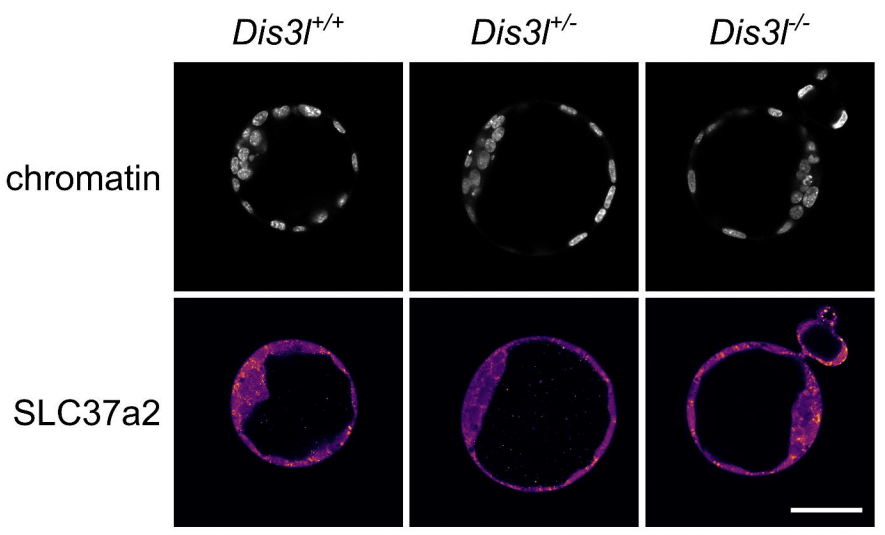
B


p-value
0.04
0.03
0.02
0.01
BP
HP
log10(padj)
apoptotic process
programmed cell death
cell death
carbohydrate transport
Pulmonary hypoplasia
Potter facies
Facial shape deformation
Aplasia/Hypoplasia of the lungs
Cholangitis
GO term

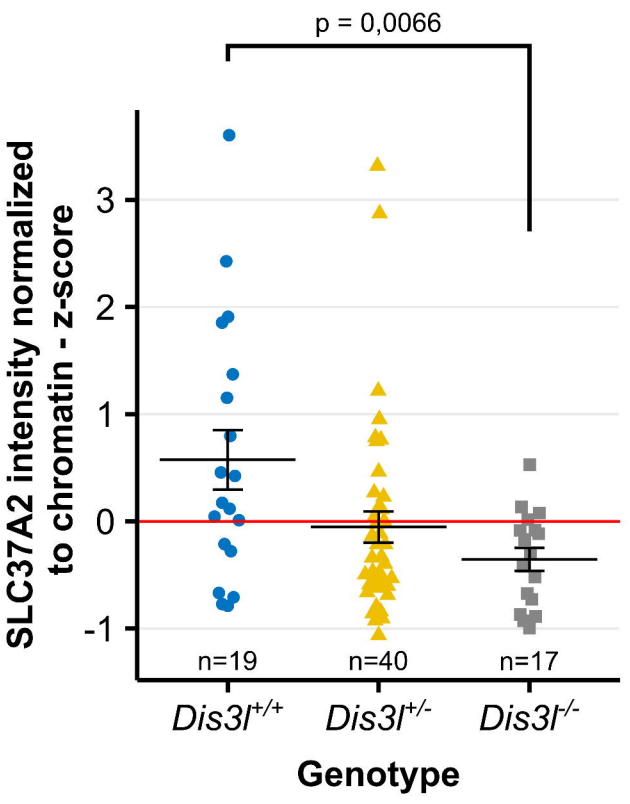
C


Fold change
Slc37a2
Tes
Gene

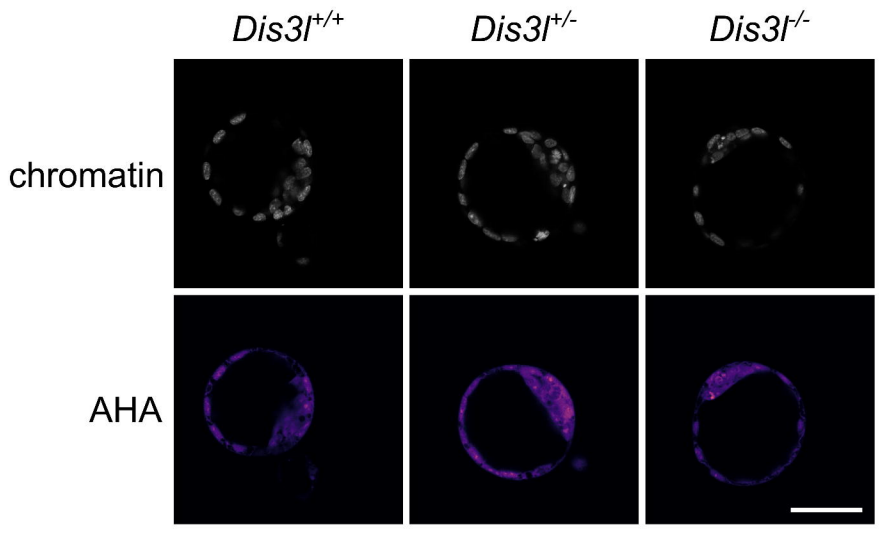
D


Dis3l+/+
Dis3l+/-
Dis3l-/-
chromatin
SLC37a2

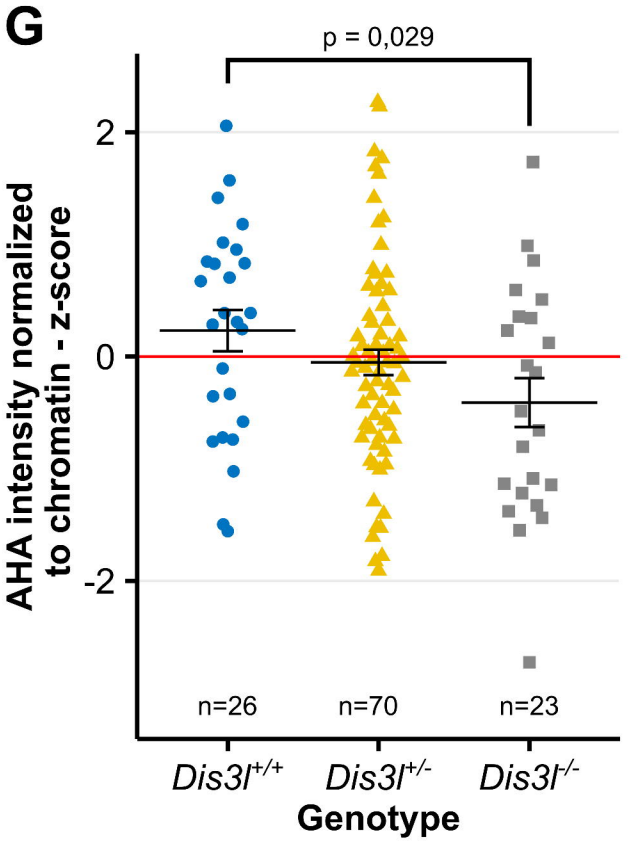
E


p = 0,0066
SLC37A2 intensity normalized to chromatin - z-score
n=19
n=40
n=17
Dis3l+/+
Dis3l+/-
Dis3l-/-
Genotype

F


Dis3l+/+
Dis3l+/-
Dis3l-/-
chromatin
AHA

G


p = 0,029
AHA intensity normalized to chromatin - z-score
n=26
n=70
n=23
Dis3l+/+
Dis3l+/-
Dis3l-/-
Genotype

**Supplementary Table 1.**

| Component | Ratio/concentration | Unit | Manufacturer |
|---|---|---|---|
| Neurobasal | 2 | - | Gibco |
| DMEM | 1 | - | |
| F-12 | 1 | - | |
| N2 | - | - | |
| B27 | - | - | |
| L-glutamine | 2 | mM | |
| Non-essential amino acids | 0,1 | mM | |
| BSA (Bovine Serum Albumine) | 0,1 | % | Sigma-Aldrich |
| β-mercaptoethanol | 0,1 | mM | |
| penicilin | 5000 | Units/ml | Gibco |
| streptomycin | 5000 | Units/ml | |
| LIF (Leukemia Inhibitory Factor) | 1000 | IU/ml | Chemicon International |
| MEK1 inhibitor - PD0325901 | 1 | uM | Sigma-Aldrich |
| GSK3 inhibitor – CHIR99021 | 3 | uM | Tocris Bioscience |

**Supplementary Table 1.** Medium composition for ES cells derivation. Components above thick bar are media composing final medium used, mixed in ratio denoted in Ratio/concentration column. Components below thick bar are additional components supplemented to the medium.

**Supplementary Table 2.**

| Target | Application | Direction | 5' -> 3' sequence |
|---|---|---|---|
| Dis3l | gDNA genotyping | forward | CGATGCTTGGGTTTCTGAAT |
| Dis3l | gDNA genotyping | reverse | TTGGAGGAAACAAACGATGG |
| Dis3l | cDNA genotyping | forward | AGCCCAGATGTGTGAGATGC |
| Dis3l | cDNA genotyping | reverse | AAGTGTGTGACGTCAGCGAT |
| Gapdh | RT-qPCR | forward | AAGGGCTCATGACCACAGTC |
| Gapdh | RT-qPCR | reverse | GGATGACCTTGCCCACAG |
| Slc37a2 | RT-qPCR | forward | GAAGGGAAAGCGGGGATTCA |
| Slc37a2 | RT-qPCR | reverse | ATGAATGACAGGCCCCAGTG |
| Tes | RT-qPCR | forward | AAGTACACCACCCTGATCGC |
| Tes | RT-qPCR | reverse | GGGGAGCCCACTCATAGGTA |

**Supplementary Table 2.** Primer sequences used in the study.

**Supplementary Table 3.**

| Staining | Antigen | Source | Producer | Cat. Number: | Dilution | Fluorochrome |
|---|---|---|---|---|---|---|
| Gata4 | Gata4 | goat | R&D Systems | AF2606 | 1:200 | - |
| Cdx2 | Cdx2 | rabbit | Abcam | ab76541 | 1:5000 | - |
| Slc37a3 | Slc37a2 | rabbit | Abcam | ab223048 | 1:200 | - |
| Gata4 | Goat IgG | donkey | Invitrogen | A-11055 | 1:500 | AF488 |
| Cdx2 | Rabbit IgG | donkey | Invitrogen | A-31573 | 1:500 | AF647 |
| Slc37a2 | Rabbit IgG | donkey | Invitrogen | A-10042 | 1:500 | AF568 |

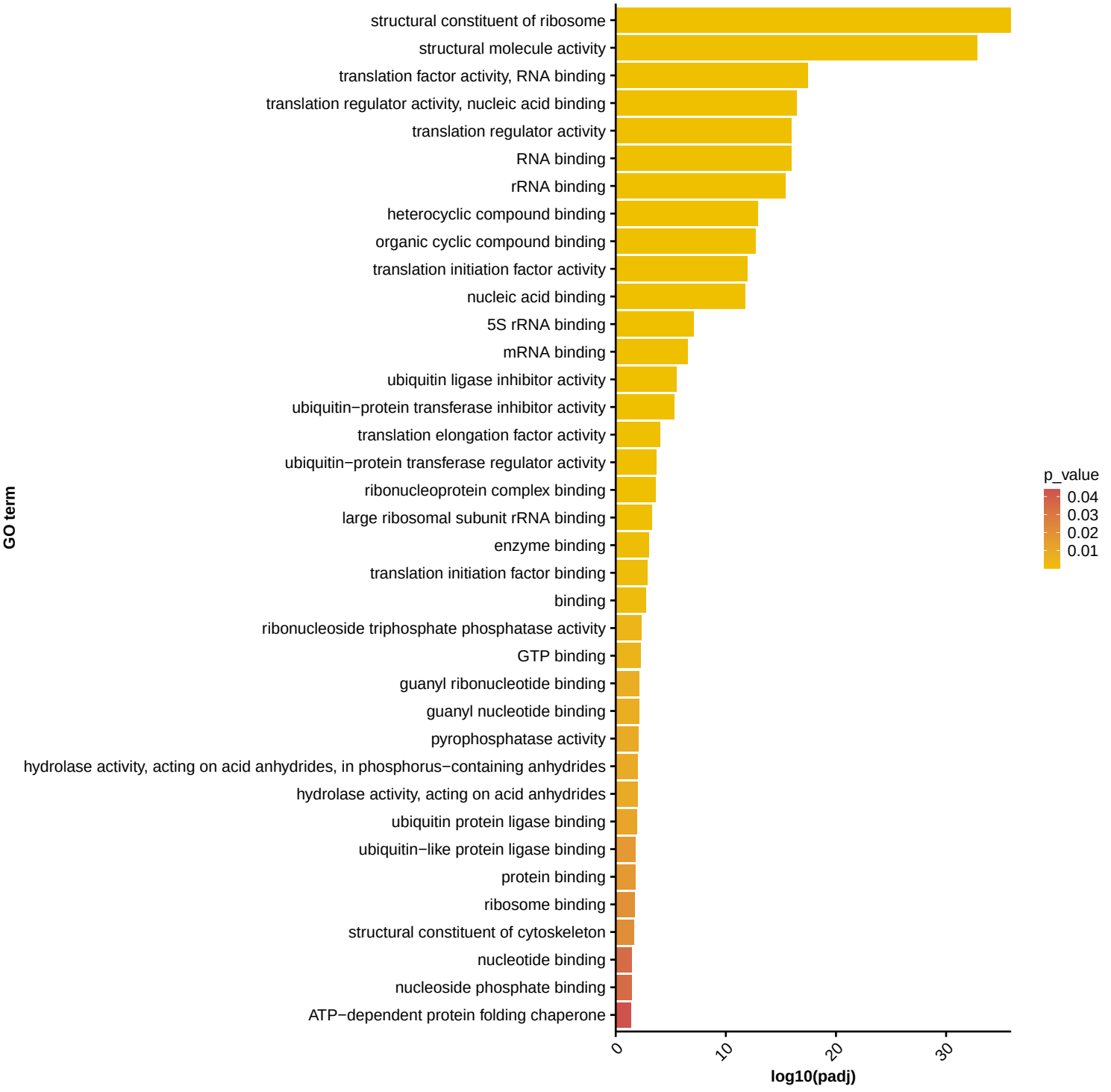
**Supplementary Table 3.** Antibodies used in the study. Only secondary antibodies were conjugated with fluorophore.

**Supplementary Figure 1.**

ACACTTCTGGAGGAGATAAGGGACCTAGCTCTTCTGGATGTCTCTGACAGTTGTGCAATGgagaatttgtattttcag
ggtgatatcatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttc
agcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggccca
ccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaa
ggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaac
cgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatc
atggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactacca
gcagaacaccccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaa
gcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaaTGAAATACTTCCATG
TCATTAAAGACCTTTGTCTTAAGTGGTGTACTTTTTTTTCTTTCT

**Supplementary Figure 1.** GFP donor sequence with homology arms (in uppercase).

**Supplementary Figure 2**

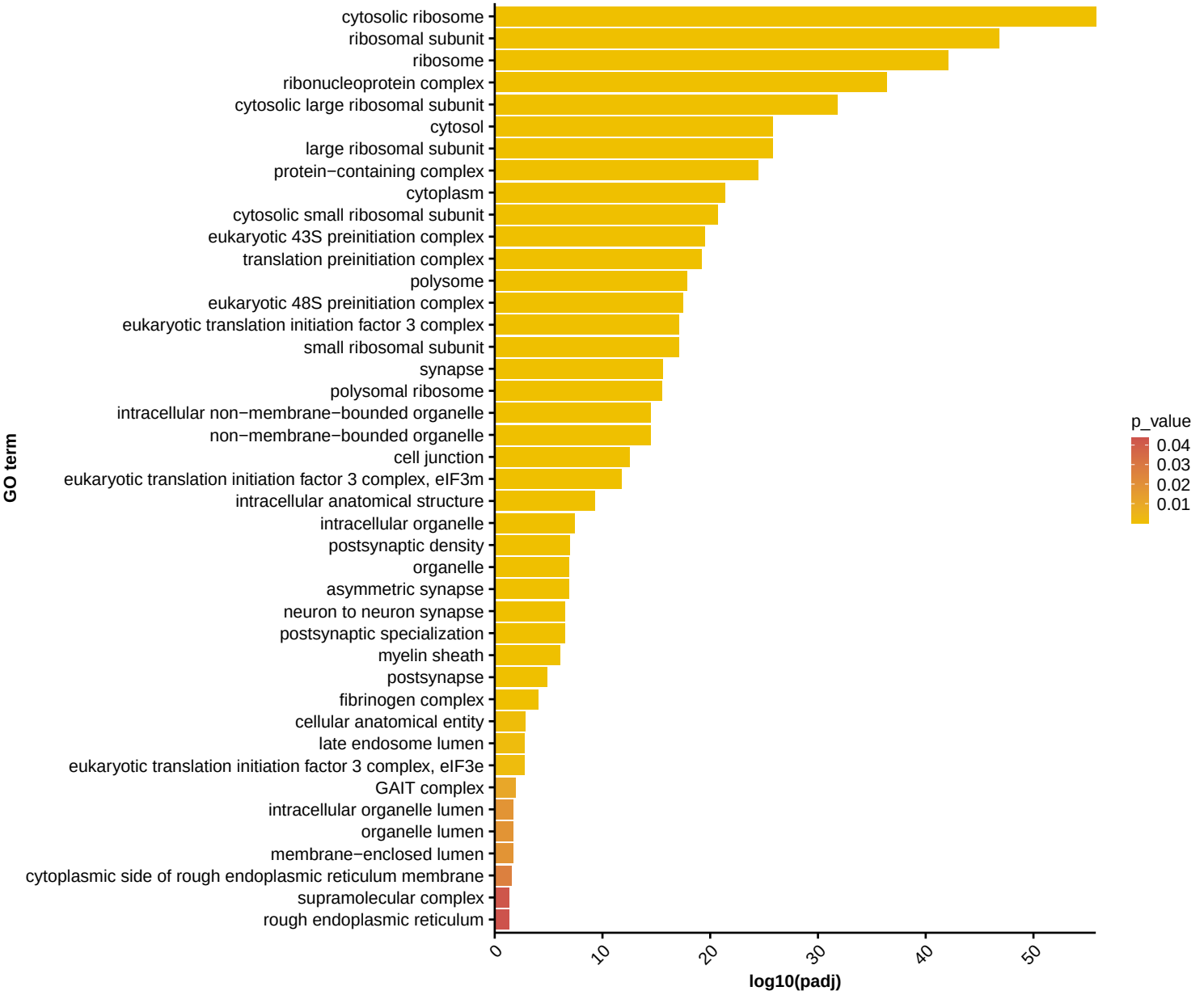


**Supplementary Figure 2.** Gene Ontology Molecular Function terms identified in functional annotation analysis of proteins coprecipitated with DIS3L-GFP.

**Supplementary Figure 3**

**Supplementary Figure 3.** Gene Ontology Biological Proces terms identified in functional annotation analysis of proteins coprecipitated with DIS3L-GFP.

**Supplementary Figure 4**



**Supplementary Figure 4.** Gene Ontology Cellular Component terms identified in functional annotation analysis of proteins coprecipitated with DIS3L-GFP.