The utilization of advanced RNA sequencing technologies to investigate post-transcriptional mechanisms involved in the regulation of gene expression.

by

Agnieszka Maria Czarnocka-Cieciura

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

Supervisor: Prof. dr hab. Andrzej Dziembowski
Auxiliary supervisor: dr Paweł Krawczyk

Warsaw, March 2025

Podziękowania

W tym miejscu chciałabym serdecznie podziękować osobom, bez których wsparcia nie dokończyłabym tej trudnej podróży jaką było dla mnie zdobywanie stopnia doktora.

Prof. dr hab. Andrzejowi Dziembowskiemu za samozwańcze podjęcie się roli mojego mentora, za czas jaki mi poświęcił, wiedze którą chętnie się dzielił, za nieskończoną cierpliwość i spokój jakie wykazał przy pisaniu wszystkich tych artykułów i dysertacji, oraz za stworzenie mi warunków, w których mogłam rozwinąć się jako naukowiec.

Dr Pawłowi Krawczykowi za pakowanie mnie w coraz to nowe przygody, pomoc w zdobywaniu i doskonaleniu warsztatu bioinformatycznego, wspólne dyskusje, wyzwania, dobre rady i wprowadzenie w tajniki pracy z serwerami.

Dr hab. Agnieszce Tudek za wspólną ciężką pracę nad publikacją na modelu drożdżowym, za jej siłę i niezłomność w obliczu piętrzących się przeciwności.

Dr Michałowi Brouze za jego długie godziny jakie spędził w labie by dostarczyć mi materiału do liczenia, wspólne rozmowy i jego niewyczerpane pokłady poczucia humoru, które nie raz ratowały mnie przed czarnymi myślami i zniechęceniem.

Dr Natalii Gumińskiej za wspólną codzienną pracę, pogodę ducha i emocjonalne wsparcie, gdy otaczająca mnie rzeczywistość przekraczała kolejne poziomy absurdu.

Mgr Pauli Kwapisz za jej niezwykłą skuteczność w organizowaniu wszystkiego co w labie potrzebne dzięki czemu mogłam skupić się wyłącznie na pracy naukowej, za jej wrażliwość na drugiego człowieka i niezliczone drewniane ołówki z gumką.

Wszystkim członkom Laboratorium Biologii RNA, z którymi maiłam przyjemność współpracować, za ich zaangażowanie we wspólne projekty i życzliwość.

Prof. dr hab. Ewie Bartnik za wsparcie w większych i mniejszych kłopotach oraz za wskazanie mi drogi, która zaprowadziła mnie tu, gdzie teraz jestem.

Moim rodzicom, teściowej i wszystkim siostrom za ich niezłomną wiarę w to, że moja długa praca nad doktoratem zakończy się sukcesem.

Na koniec pragnę podziękować Mariuszowi Czarnockiemu-Cieciura, który jest drugą połową mojej duszy za motywację do działania, zdrowy rozsądek, za to, że zawsze jest po mojej stronie i wiele innych rzeczy, których już nie napiszę, bo kończy mi się miejsce na kartce.

Financial support

The research was conducted with financial support from the ERC Starting Grant (European Research Council, 7th Framework Programme) under the grant “Regulation of Gene Expression by non-canonical poly(A) and poly(U) polymerases” 309419-PAP&PUPs.

Contents

Abstract

Antagonistic mechanisms, such as deadenylation and polyadenylation, play critical roles in post-transcriptional regulation of gene expression and transcriptome remodeling in eukaryotic cells. This work explores cytoplasmic mRNA deadenylation and decay in yeast and the role of non-canonical poly(A) polymerases in murine gametogenesis. The dissertation consists of three multiauthor publications in which I was primarily responsible for bioinformatic analysis. Utilizing Nanopore RNA Sequencing (DRS) data, we modeled yeast mRNA deadenylation rates *in vivo* and identified transcript-specific decay dynamics under normal and stress conditions. In murine studies, TENT5 poly(A) polymerases were shown to regulate the stability of key transcripts governing oocyte and spermatocyte maturation. Analyses also revealed unique poly(A) tail compositions in testes and ovaries transcriptomes. The enrichment of non-canonical nucleotides in poly(A) tails of transcripts crucial for germ cell development opens new avenues for future research into the regulatory mechanisms of gametogenesis.

Streszczenie

Deadenylacja i poliadenylacja mRNA są przeciwstawnymi procesami odgrywającymi kluczową rolę w post-transkrypcyjnej regulacji ekspresji genów i kształtowaniu transkryptomów komórek eukariotycznych. Niniejsza praca zawiera wyniki badań nad cytoplazmatyczną deadenylacją i degradacją mRNA u drożdży oraz rolą niekanonicznych polimeraz poli(A) w gametogenezie myszy. Dysertacja zawiera trzy wieloautorskie publikacje, w których odpowiedzialna byłam za analizę bioinformatyczną danych. Wykorzystując technikę bezpośredniego sekwencjonowania RNA opracowaną przez Oxford Nanopore (DRS), otrzymaliśmy dane które umożliwiły nam modelowanie szybkości deadenylacji mRNA drożdży *in vivo* i opisanie specyficznej dla każdego transkryptu dynamiki degradacji w warunkach normalnych i pod wpływem stresu. W badaniach na myszach wykazaliśmy, że polimerazy poli(A) TENT5 regulują stabilność kluczowych transkryptów regulujących dojrzewanie oocytów i spermatocytów. Analiza ujawniła również wzbogacenie niekanonicznych nukleotydów w ogonach poli(A) transkryptów kluczowych dla rozwoju komórek płciowych co otwiera nowe możliwości dla przyszłych badań nad mechanizmami regulującymi gametogenezę.

1. Introduction

Gene expression in eukaryotic cells is a tightly regulated process that converts genetic information into functional proteins or RNA molecules. It involves several stages, including transcription, post-transcriptional mRNA processing, nuclear export, translation, and post-translational modifications, each governed by strict controls. Gene expression is intricately linked to the cell cycle, cell differentiation, and environmental adaptation. Factors regulating RNA turnover are involved in the accompanying processes of transcriptome remodeling. To date, numerous mechanisms have been identified that either stabilize or accelerate the decay of mRNA molecules. This dissertation provides new perspectives on the roles of polyadenylation, deadenylation, and mRNA degradation in shaping transcriptomes in eukaryotic cells. The completion of these studies was made possible by the advancement of sequencing methods enabling the direct reading of RNA sequences. This research utilized data derived from Oxford Nanopore's Direct RNA Sequencing, the analysis of which facilitated the acquisition and interpretation of diverse biological insights.

1.1 Direct RNA Sequencing by Oxford Nanopore as a tool for investigating post-transcriptional gene regulation pathways

The development of next-generation sequencing techniques, such as Illumina, has revolutionized genomic and transcriptomic research. The versatility of these methods in handling diverse biological materials and integrating them into complex experimental protocols has enabled significant advancements in characterizing living organisms and describing the transformation of transcriptomes in cells. The regulation of poly(A) tail length is governed by numerous cis- and trans-regulatory mechanisms and enzymes. Its dynamics result from the interplay of two opposing forces -polyadenylation and deadenylation. Studying these processes requires precise methods to measure poly(A) tail length. Over the years, various experimental approaches have been developed for this purpose (Brouze et al., 2023). The first and most common ones were based on Illumina sequencing (Babaian et al., 2020; Subtelny et al., 2014) and provided the opportunity to conduct global profiling of tail-length dynamics (Eisen et al., 2020). However, these methods suffered from limitations inherent to this technology, specifically PCR amplification bias that affected the accurate reading of homopolymer sequences. This comes from the fact that polymerases during amplification tend to slip on homopolymeric regions (Hommelsheim et al., 2014; Sehn, 2015), conversely introducing errors in poly(A) tail length estimation, even if sophisticated methods to overcome this were introduced. Furthermore, conducting TAILseq or PALSeq analysis requires modifications to the original software of Illumina sequencing machines, which is only feasible for a limited number of models (Brouze et al., 2023). Nevertheless, both methods significantly contributed to the knowledge of poly(A) tails biology.

Another groundbreaking technology appeared in 2019 with the introduction of Direct RNA Sequencing (DRS) by Oxford Nanopore (Garalde et al., 2018). With this technique RNA molecules isolated from biological material are sequenced directly, without converting them into cDNA and any PCR amplification steps. Its ability to produce significantly longer reads has been proven crucial for transcriptomic studies, enabling the analysis of poly(A) tail lengths, transcript isoforms, and alternative polyadenylation sites (Brouze et al., 2023; Mulroney et al., 2022).

In DRS, during the preparation of the sequencing libraries, a DNA adapter is ligated to RNA, optionally followed by reverse transcription. The library is loaded onto a flow cell, which is an electronic device made from artificial membrane coated with protein pores (nanopores). When the voltage is applied, RNA molecules pass through the pore in a 3’ to 5’ direction, driven by the motor protein, causing specific electric current perturbations recorded by very sensitive electrodes linked to each of the protein pore. Each nucleotide (adenine, thymine, cytosine, guanine, or uracil in RNA) produces a distinct disruption in the ionic current, which is measured in real-time and recorded as raw data, including current amplitude, signal events (moves), duration of blockages, and signal noise or fluctuations. Different sequences generate distinct current patterns, captured as variations. Using neural network algorithms, these signals are then converted into nucleotide sequences by base-calling tools. The RNA sequences are base-called with an accuracy of 91-92% (for nanopore chemistry SQK-RNA002) or 96% (nanopore chemistry SQK-RNA004). For nanopore DNA sequencing base-calling accuracy is comparable to Illumina-based methods, reaching 99.9%).

The size of the dataset obtained with nanopore sequencing is variable and depends on sequencing length and RNA quality (Jain et al., 2022). Since DRS generates mostly full-length, strand-specific reads, the sequencing depth required for common bioinformatic analyses is lower than that of Illumina-based methods and there is no need to assemble it from small pieces. Meanwhile, the lack of an amplification step forces the use of more biological material for sequencing. Despite the lack of PCR steps that influence homopolymer length, nanopore base-callers struggle with accurately interpreting long homopolymers like poly(A) tails. In the case of direct RNA sequencing, poly(A) tail, which is always read directly after the adapter, is not even included in the basecalled sequence. Thus, specific data analysis pipelines using raw current read-outs are needed to estimate tail lengths from direct RNA sequencing nanopore data. Popular tools like Nanopolish-polyA (https://github.com/jts/nanopolish) and TailfindR (https://github.com/adnaniazi/tailfindr) segment the nanopore signal into regions corresponding to the sequencing adapter, poly(A) tail, and transcript body, then estimate tail lengths based on poly(A) length transit time through nanopore and rate of the translocation. The exact tail length is estimated individually for each read. Therefore, DRS uniquely enables the study of isoform-specific poly(A) tail length distributions, splicing, and transcript stability (Drexler et al., 2020; Roach et al., 2020; Workman et al., 2019). Its ability to sequence directly from RNA has made it an indispensable tool for investigating post-transcriptional gene regulation pathways. This was also the main technique used in this dissertation.

1.2 Polyadenylation

1.2.1 canonical polyadenylation

The key stages of mRNA turnover are evolutionarily conserved among eukaryotes. Additionally, all mature eukaryotic mRNA molecules share common structural features: a 5’ cap, a 5’ untranslated region (UTR), a coding region, a 3’ UTR, and a poly(A) tail at the 3’ end. The maturation of the mRNA 3’ end begins with recognition of the poly(A) signal sequence (PAS) in the 3’ UTR pre-mRNA, followed by endonucleolytic cleavage and polyadenylation of the upstream cleavage product. These processes are mediated by the large (~1MDa), evolutionarily conserved cleavage, and polyadenylation complex (CPAC), which is composed of multiple structurally and functionally distinct modules. In yeast, CPAC consists of the cleavage factor (CF) and cleavage and polyadenylation factor (CPF), whereas in mammals, it includes the cleavage and polyadenylation specificity factor (CPSF). Polyadenylation of pre-mRNA is initiated by poly(A) polymerase (Pap1 in yeast, PAP in vertebrates), which is part of the CPAC complex in yeast or interacts with it via Fip1/hFip1 in vertebrates. These structural differences affect the dynamics of polyadenylation. Adenosine attachment to the poly(A) tail occurs in discrete intervals due to the instability of binding between the pre-mRNA's 3' terminus and the CPAC complex. In yeast, Pap1 is stably attached to the CPAC, so the polyadenylation rate is constant, and the process of adenosine addition occurs rapidly. However, in vertebrates, polyadenylation is biphasic. Initially, PAP binds CPAC through Fip1 and interacts with the U-rich regions of the pre-mRNA which stabilizes the entire complex. The first 12-15 adenosines are added slowly. PolyA binding protein (PABP) adhesion to poly(A) tail enables a higher polyadenylation efficiency. These proteins bind PAP to the poly(A) tail and a further 200-250 adenosines are added 4 orders of magnitude faster. The growing poly(A) string destabilizes the binding of PAP to CPAC leading to termination of tail elongation (Mitchell and Tollervey, 2000; Wahle, 1999; Zhao et al., 1999).

Poly(A)-binding proteins (PABPs) are a family of proteins that bind to mRNA poly(A) tails. They require a minimum of 12 adenosines for binding but protect 25–27 adenosine residues and can multimerize along the poly(A) tract (Nicholson-Shaw et al., 2022; Sachs et al., 1986). All PABPs contain N-terminal RNA-recognition motifs (RRMs) that enable interaction with poly(A) tails (Burd and Dreyfuss, 1994). PABPs are evolutionarily conserved from yeast to humans. In single-cell eukaryotes, a single gene encodes cytoplasmic PABP (PABPC), whereas metazoans and plants possess multiple PABPC genes. Poly(A)-binding proteins are broadly categorized as nuclear or cytoplasmic based on location and phylogeny (Mangus et al., 2003). In the nucleus, PABPs participate in pre-mRNA maturation by regulating polyadenylation termination and facilitating mRNA export (Dunn et al., 2005; Kwiatek et al., 2023; Rodríguez-Molina and Turtola, 2023; Soheilypour and Mofrad, 2016). In the cytoplasm, they regulate mRNA deadenylation and decay (Wang and Kiledjian, 2000; Wolf et al., 2014). The presence of PABPs on poly(A) tails is not essential for the initiation and termination of translation (Proweller and Butler, 1994; Searfoss and Wickner, 2000). However, mRNA with PABPs on poly(A) tail is translated much more efficiently (Preiss and Hentze, 1998; Wu et al., 2020). In this dissertation, we demonstrate the role of Poly(A)-binding proteins in maintaining the balance between deadenylation, decapping, and mRNA decay.

1.2.2 Non-canonical polyadenylation

Other evolutionarily conserved proteins, termed non-canonical poly(A) polymerases (ncPAPs), also exhibit poly(A) polymerase activity. In yeast, non-protein coding transcripts such as small nucleolar RNAs (snoRNAs) as well as pervasive transcription products known as cryptic unstable transcripts (CUTs) are polyadenylated by the dedicated enzyme Trf4 protein, a component of the TRAMP4 complex (Trf4-Air1/2-Mtr4 polyadenylation complexes), and subsequently processed by the exosome. Ribosomal RNA precursors (pre-rRNAs) represent a unique class of transcripts, undergoing extensive modifications during transcription. Their quality control and polyadenylation are mediated by the TRAMP5 complex, where Trf5, a paralogue of Trf4, plays a significant role (Tudek et al., 2021).

Unlike *S. cerevisiae*, where polyadenylation occurs exclusively in the cell nucleus, in higher eukaryotes, mRNAs 3’ end may be polyadenylated or polyurydylated also in cytoplasm (Warkocki et al., 2018). Cytoplasmic polyadenylation is a post-transcriptional gene regulatory mechanism critical for animal development. Early investigations concentrated on its function in oocyte differentiation, where maternal mRNA regulation is essential for progression through gamete maturation stages (Read and Norbury, 2002). Based on these findings, a mechanism was identified in which the Cytoplasmic Polyadenylation Element-Binding Protein (CPEB) family governs the translation of specific maternal transcripts. CPEB proteins (CPEB1 and CPEB2) contain two RNA-binding RRM domains within their structure. They recognize the cis-acting cytoplasmic polyadenylation element (CPE) in the 3′UTR, typically with the structure UUUUA(1–2)U. The mechanism regulating polyadenylation dynamics in the cytoplasm involves recruiting the poly(A) polymerase TENT2 (Gld2) to transcripts that contain the CPE motif in their 3′UTR, enabling precise control of mRNA translation (Huang et al., 2023). CPEBs recruit the TENT2 poly(A) polymerase to transcripts with CPE motifs. Both CPEBs and catalytical subunit of TENT2 are evolutionarily conserved across animals. Identified CPEB and Gld2 target transcripts exist in multicellular organisms, suggesting an ancient mechanism that facilitated the evolutionary leap from unicellular to multicellular life (Rouhana et al., 2023).

Further studies of ncPAPS revealed their regulatory roles in multiple cellular processes, not limited to gametogenesis. TENT2 plays role also in synaptic plasticity (Mansur et al., 2021). TENT4 family members, via mixed tailing control the stability of mRNAs (Kim et al., 2020; Lim et al., 2018) and are involved in the viral RNA processing (Kim et al., 2020; Li et al., 2022). Another ncPAPs family, TENT5, plays role in tissue-specific processes, like bone development and immune responses in animals (Bilska et al., 2020; Gewartowska et al., 2021; Liudkovska et al., 2022). This dissertation includes two studies investigating the role of TENT5 family non-canonical poly(A) polymerases in oocyte and spermatocyte maturation processes in mice.

1.3 mRNA deadenylation and decay

1.3.1 Deadenylase complexes

Two conserved complexes carry out the majority of deadenylation in the cytoplasm: Pan2–Pan3 and Ccr4–Not. The exonuclease activity stems from Pan2 (in Pan2–Pan3) and Ccr4 and Pop2 (in Ccr4–Not), while the roles of other subunits remain poorly understood (Boeck et al., 1996; Tucker et al., 2001). Deadenylase complexes can associate with mRNA directly or indirectly via RNA-binding proteins. In Pan2–Pan3, the Pan3 subunit's PAM2 motif (PABP interacting motif 2) interacts with poly(A)-binding protein (PABP), and the mechanism is conserved from yeast to humans (Wolf et al., 2014). The Ccr4–Not complex contains two exonucleases, Ccr4 and Pop2. Ccr4 performs the bulk of deadenylation, but Pop2 partially compensates for Ccr4's inactivation. It has also been shown that Pop2 promotes deadenylation of transcripts with low codon optimality (Webster et al., 2018).

In yeast Pan2–Pan3 and Ccr4–Not may function sequentially. This is supported by the phenotypes observed in single and double *pan2* and *ccr4* mutants in yeast (Tucker et al., 2001). Additionally, it has been shown that the Pan2–Pan3 complex inefficiently removes the final 20–25 adenosines, which corresponds to the approximate length required to bind PABP (Wolf et al., 2014). This suggests that Pan2–Pan3 activity halts or significantly slows after the final PABP has been displaced from the mRNA. In contrast, the Ccr4–Not complex is more efficient in removing the remaining 20–25 adenosines (Webster et al., 2018). Recent studies using nanopore DRS sequencing have challenged the two-step deadenylation model. Although transcripts with elongated poly(A) tails overlap significantly between *ccr4* and *pan2* mutants, unique subsets exist for each deadenylase (Tudek et al., 2021). This indicates a more complex relationship between Pan2/Pan3 and Ccr4-Not complexes, requiring further investigation into cytoplasmic deadenylation dynamics.

In mammals, as in yeast, the CNOT (yeast Ccr4-NOT) and PAN2/PAN3 (yeast Pan2/Pan3) complexes are responsible for most cytoplasmic deadenylation. Structurally, the mammalian CNOT complex resembles its yeast counterpart but includes two additional components, CNOT10 and CNOT11, absent in yeast. Conversely, Caf130 is unique to yeast. Mammalian CNOT6/6L and CNOT7/8 exhibit deadenylase activity, while CNOT4 functions as an E3 ligase. Higher eukaryotes possess other poly(A) tail-shortening enzymes, such as PARN and Angel1/2. PARN belongs to the DEDD exoribonuclease superfamily which members coordinate two metal ions essential to their catalytic mechanism. It is conserved across various eukaryotic species, including *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Xenopus*, and mice, but is absent in *Saccharomyces cerevisiae* and *Drosophila melanogaster*. PARN takes part in the regulation of fertility in worms and frogs. In mammals it is involved in ncRNA poly(A) tail deadenylation (Huynh and Parker, 2023; Łabno et al., 2016). Nocturnin (NOC/Ccr4c) belongs to the exonuclease/endonuclease/phosphatase protein family, which includes CNOT6L, the mammalian orthologue of the yeast deadenylase CCR4.The deadenylases ANGEL1 and ANGEL2 (Ccr4e/Ccr4d) are homologues of the yeast Ccr4 but lack the region required for interaction with the Ccr4-Not complex. Both proteins localize to mitochondria and are believed to be involved in mitochondrial RNA processing. ANGEL2 and its *Drosophila* homologue, DmAngel, are critical for non-canonical RNA processing. Using *Drosophila* and mouse knockout models, it has been shown that the loss of DmAngel or ANGEL2 results in the accumulation of 3′ phosphates on transcripts, preventing their polyadenylation and causing respiratory chain deficiencies (Clemente et al., 2022).

1.3.2 mRNA decay

mRNA decay is a crucial process for regulating gene expression by removing defective or unnecessary transcripts, and it plays a significant role in shaping eukaryotic transcriptomes. The process is very similar across all groups of eukaryotes. For functional, mature mRNA molecules, decay begins with deadenylation, where the poly(A) tail is shortened by the CCR4-NOT or PAN2-PAN3 complexes. Once sufficiently reduced, one of two major RNA degradation pathways is activated: 5'-to-3' decay or 3'-to-5' decay. In the 5'-to-3' pathway, the mRNA cap is removed by the decapping enzyme Dcp1/Dcp2, and the transcript is degraded by the exonuclease XRN1. In the 3'-to-5' pathway, degradation is carried out by the exosome complex DIS3l2 (Łabno et al., 2016). Both pathways contribute to mRNA turnover and cellular homeostasis.

The rate of deadenylation and decay has been a subject of extensive investigation, which has identified factors influencing deadenylation rates (Chen and Shyu, 2011; Eisen et al., 2020; Miller et al., 2011; Neymotin et al., 2014; Presnyak et al., 2015). However, the complete network of dependencies regulating mRNA turnover remains poorly understood. Decapping and Xrn1-mediated degradation occur rapidly, making the rate of poly(A) tail shortening a greater determinant of transcript stability. Although poly(A) tail length influences transcript half-life, this does not directly correlate with expression levels (Tudek et al., 2021). It has also been observed that evolutionarily younger transcripts exhibit shorter half-lives than those encoding housekeeping proteins, suggesting a differing deadenylation rate (Chen et al., 2008; Tudek et al., 2021). Attempts to estimate deadenylation rates using reporter transcripts have produced varying values, with poly(A) shortening rates ranging between 0.5 and 30 adenosines per minute (Alles et al., 2023; Eisen et al., 2020). These estimates were obtained in both *in vitro* and *in vivo* systems. However, each of the experimental approaches faces many limitations. In i*n vitro* studies, it is challenging to replicate the conditions of the cellular environment, and the influence of interactions with proteins not included in the experiment is often overlooked. Consequently, the results provide, at best, an estimate of the enzymatic efficiency of the protein complex under investigation. The advantage of this approach lies in the controlled experimental conditions, which allow for precise measurement of the shortening poly(A) tail of the reporter transcript. In vivo studies face challenges in distinguishing between nuclear and cytoplasmic deadenylation. Additionally, bias is introduced by the use of Illumina-based sequencing methods for measuring poly(A) tail lengths (Brouze et al., 2023). This dissertation presents the first study describing the relationships between deadenylation, decapping, and mRNA decay *in vivo*, alongside a deadenylation model incorporating the regulatory role of Pab1 proteins in yeast. By employing Direct RNA Sequencing by Oxford Nanopore, this work reliably measured poly(A) tail lengths and provided accurate estimations of *in vivo* deadenylation rates.

1.4 Selected aspects of Poly(A) tail length dynamics control

The length of poly(A) tails varies across organisms, tissues, and transcripts (Gewartowska et al., 2021; Lim et al., 2018; Liudkovska et al., 2022; Tudek et al., 2021). In yeast, nuclear-exported mRNAs have an average poly(A) tail length of 50 adenines. However, this mean can differ from 30–40 adenines in mRNAs coding ribosomal proteins up to 80 adenosines in non-coding RNAs. Interestingly, no strong correlation exists between transcript expression levels and poly(A) tail length (Tudek et al., 2021). This indicates the presence of diverse post-transcriptional regulation pathways interacting with transcripts 3'ends. In eukaryotes, cytoplasmic poly(A) tail dynamics are significantly influenced by deadenylation. Both Pan2/Pan3 and CCR4-NOT deadenylation complexes work in distributive mode without inherent transcript specificity (Chen and Shyu, 2011; Webster et al., 2017).

There are mechanisms that direct deadenylase activity toward specific groups of transcripts. For instance, studies in yeast have shown that the CCR4-NOT complex interacts with Puf proteins, which contain a Pumilio domain that binds specific motifs in the 3′UTRs of transcripts (Qiu et al., 2019; Wang et al., 2018). Although the precise mechanism of this interaction remains undefined, experimental methods have identified potential target groups regulated by these factors (Lapointe et al., 2017; Wilinski et al., 2015). In higher eukaryotes, mRNA tail length regulation is even more complex due to the presence of non-canonical poly(A) polymerases, which can also exhibit selective activity (Gewartowska et al., 2021).

The role of poly(A) tail length changes in transcriptome remodeling in eukaryotic cells is well illustrated by stress response experiments. During heat stress in yeast, poly(A) tails shorten globally but lengthen for certain transcripts over time. Similarly, nutrient limitation results in globally shorter poly(A) tails (Tudek et al., 2021). Another example of significant transcriptome rearrangement is cell differentiation. During oocyte maturation and the oocyte-to-embryo transition, widespread changes in protein levels occur, without corresponding mRNA level changes. This process proceeds with the complete silencing of transcription in the nucleus of the differentiating cell (Ermisch and Wood, 2024). Cell reorganization is driven mostly by maternal mRNAs, with Cytoplasmic Polyadenylation Element Binding Protein (CPEB) regulating translation. CPEB was postulated to recruit Maskin (a CPEB-associated factor) to repress translation initially and later promote polyadenylation by binding cytoplasmic poly(A) polymerase alongside Cleavage and Polyadenylation Specificity Factor (CPSF) (Eichhorn et al., 2016). It was shown that poly(A) tail elongation during gametogenesis occurs in waves and is essential for gamete differentiation progression (Morgan et al., 2019). Although TENT2 is the main ncPAP described in the context of gametogenesis, there are works indicating the role of other enzymes, like TENT3A/TENT3B (Morgan et al., 2017) or TENT5s. In my work, I contributed to the demonstration of the role of the non-canonical polyadenylases TENT5B, TENT5C, and TENT5D in regulating the poly(A) tail length of transcripts essential for proper oogenesis and spermatogenesis.

1.5 Poly(A) tail composition

Early evidence of non-canonical nucleotides in poly(A) tails emerged from studies on uridylation, particularly in the context of RNA turnover and degradation pathways. In cells depleted of TUT4/7, the vast majority of mRNAs lose their oligo-U tails, resulting in extended half-lives (Lim et al., 2014). The first global analyses of poly(A) tail composition became possible with the development of the TAIL-seq tool, which allows genome-wide determination of poly(A) tail length and 3' end modifications. Research using TAIL-seq revealed widespread uridylation and guanylation downstream of poly(A) tails, which affects mRNA stability (Chang et al., 2014). Further studies have revealed that uridines incorporated into poly(A) tails by TUTases, including TENT1 (TUT1), TENT3A (TUT4), and TENT3B (TUT7), play a role in the turnover of microRNAs, histone mRNAs, and U6 snRNA (Menezes et al., 2018; Rissland et al., 2007). Guanosines added to poly(A) tails through mixed tailing by TENT4A and TENT4B, inhibits the CCR4-NOT deadenylase complex, thereby enhancing mRNA stability (Lim et al., 2018). Further advancements in tools enabled the comprehensive analysis of non-canonical nucleotides throughout entire poly(A) tails. The single-cell-sensitive method PAIso-seq permits the quantification of poly(A) tail lengths alongside full-length cDNA sequencing while accurately identifying non-adenosine residues within poly(A) tails. This approach quantified isoform-specific poly(A) tail lengths and demonstrated that 17% of mRNAs isolated from mouse GV oocytes contain non-A residues within the body of their poly(A) tails (Liu et al., 2019).

Both PAIso-seq and TAIL-seq rely on short-read sequencing technologies (Illumina). However, the localization of non-canonical nucleotides in poly(A) tails obtained with these methods is prone to errors caused by DNA polymerase slippage on homopolymeric regions (such as poly(A) tails) during library preparation. More precise measurements are achievable using Nanopore DRS though available basecallers completely omit the poly(A) tail, thus missing also possible non-A residues. This limitation was addressed with the development of the Ninetails tool, which employs a neural network to classify signal anomalies within poly(A) tails into three categories corresponding to the presence of cytidine, guanosine, or uridine nucleotides (https://github.com/LRB-IIMCB/ninetails, Gumińska et al., 2025). The resulting data tables enable the calculation of the frequency of each non-adenosine nucleotide across experimental datasets and provide approximate localisation of non-A residues within poly(A) tails. Studies utilizing Ninetails revealed significant differences in the prevalence of non-A nucleotides across various organisms and tissues, suggesting their role as an additional layer of mRNA regulation in cells (Krawczyk et al., 2025; Natalia Gumińska et al., 2025). This hypothesis was further supported by *in vitro* studies of the deadenylating complexes Pan2/Pan3 and Ccr4-Not, using reporter mRNAs containing U, G, or C in their poly(A) tails. These studies found that the presence of non-canonical nucleotides slowed down deadenylation, potentially increasing transcript half-life as a result (Lee et al., 2024). The presence of non-canonical nucleotides in poly(A) tails is a compelling area of research in the development of novel therapeutics. Studies have shown that uridine incorporation into the poly(A) tails of viral mRNA by TENT proteins accelerates its degradation. Furthermore, poly(A) tail composition is critical for advancements in mRNA vaccine research, highlighting its significance in both antiviral therapies and vaccine development.

In my work, I have shown that the content of non-adenosines in poly(A) tails is higher in germ cells compared to other mouse tissues. However, their role needs further examination.

2. Purpose of this work

In this dissertation, I utilized comprehensive transcriptomic datasets, primarily generated using Direct RNA Sequencing on nanopores, to study:

- the dynamics of mRNA deadenylation and decay in the cytoplasm of *S. cerevisiae*
- mRNA substrates of TENT5 non-canonical poly(A) polymerases during germ cell maturation in mice
- non-canonical nucleotides in poly(A) tails of mRNAs during mouse gametogenesis

These bioinformatic analyses, together with experimental data and modeling results obtained by others, enabled us to enhance our understanding of post-transcriptional gene expression regulatory mechanisms.

3. Results

3.1 Manuscript 1: Modeling of mRNA deadenylation rates reveal a complex relationship between mRNA deadenylation and decay

3.1.1 Description

Poly(A) tails, which decorate the 3′ ends of almost all eukaryotic mRNAs, are crucial for their stability and translation into proteins. Consequently, the removal of the poly(A) tail plays a critical role in regulating gene expression. Determining the rate of mRNA degradation in the cytoplasm is not a trivial task, mainly due to the continuous influx of newly synthesized transcripts from the cell nucleus.

The deadenylation model described in the manuscript is the first to be developed based on data obtained from DRS sequencing. By comparing the poly(A) tail length distributions for depletion mutants of Pab1 (poly(A)-binding protein), Dcp2 (Decapping mRNA 2), and Xrn1 (5′-3′ exoribonuclease), the relationship between deadenylation and decapping, in which the Pab1 protein plays a key role, was characterized. Next, a yeast system with depletion of the Mex67 protein, responsible for exporting transcripts from the cell nucleus, was utilized. Rapamycin-dependent depletion of Mex67 leads to a complete blockade of nuclear mRNA export and, over time, widespread degradation of the accumulated transcripts. Notably, Mex67 depletion does not affect cytoplasmic deadenylation or mRNA decay as was previously shown in Tudek et al., 2021.

A chase experiment was performed, with samples being collected at six time points and the isolated poly(A) fraction subsequently sequenced using the DRS method. The resulting time-course profile of poly(A) tail length changes allowed transcript groups with varying deadenylation rates to be identified and served as the foundation for a mathematical model of deadenylation based on a so-called modified gamma function. Additionally, an alternative method for estimating deadenylation rates based on the quantiles of the poly(A) tail length distribution was proposed. A good approximation of the numerical model is provided by this approach, and the estimation of deadenylation rates for individual transcripts is enabled.

A correlation between deadenylation rate, decay rate, and transcript half-life was further demonstrated using data retrieved from the available literature. A chase experiment was also conducted under conditions of Mex67 depletion combined with thermal stress and thiolutin treatment. It was revealed by these experiments that deadenylation and decay rates accelerate under stress conditions, allowing the transcriptome to be remodeled more rapidly by the cell and adaptation to new environmental conditions to occur. Finally, it was shown that the decay of transcripts encoding ribosomal proteins (RPGs) is independent of deadenylation in heat stress conditions. However, the level of RPG mRNAs is influenced by nuclear export, with the unique regulation of this transcript group being highlighted.

3.1.2 My contribution

In this manuscript, in which I share first authorship with Professor Jarosław Poznański, I was responsible for the comprehensive bioinformatic analysis of all 174 DRS Nanopore sequencing datasets thanks to which the numerical modelling of deadenylation process was possible. It encompassed data quality analyses, statistical analyses of poly(A) lengths, mRNA abundance analyses, functional (GO-terms), characterization and identification of different groups of transcripts, and poly(A) tail length enrichment analyses. I contributed to biological interpretations of obtained results and the overall content of the manuscript. The final version of the manuscript contained more than 100 plots which were evolving and created with the joint efforts of: mine, Professor Jarosław Poznański and dr hab. Agnieszka Tudek. Together with dr hab. Agnieszka Tudek, I compiled, reviewed, and prepared the final figures. In collaboration with dr hab. Agnieszka Tudek and Professor Andrzej Dziembowski, I contributed to writing and refining the final manuscript and supplemental materials.

3.1.3 Citation and manuscript text

Czarnocka-Cieciura A, Poznański J, Turtola M, Tomecki R, Krawczyk PS, Mroczek S, Orzeł W, Saha U, Jensen TH, Dziembowski A, Tudek A. Modeling of mRNA deadenylation rates reveal a complex relationship between mRNA deadenylation and decay. EMBO J. 2024 Dec;43(24):6525-6554. doi: 10.1038/s44318-024-00258-3. Epub 2024 Oct 11. PMID: 39394354; PMCID: PMC11649921

Article

SOURCE DATA
TRANSPARENT PROCESS
OPEN ACCESS
CHECK FOR UPDATES

THE EMBO JOURNAL

Modeling of mRNA deadenylation rates reveal a complex relationship between mRNA deadenylation and decay

Agnieszka Czarnocka-Cieciura[1,6], Jarosław Poznański[2,6], Matti Turtola[3], Rafał Tomecki[2,4], Paweł S Krawczyk[1], Seweryn Mroczek[1,4], Wiktoria Orzeł[1], Upasana Saha[5], Torben Heick Jensen[5], Andrzej Dziembowski[1,4]✉ & Agnieszka Tudek[2]✉

Abstract

Complete cytoplasmic polyadenosine tail (polyA-tail) deadenylation is thought to be essential for initiating mRNA decapping and subsequent degradation. To investigate this prevalent model, we conducted direct RNA sequencing of *S. cerevisiae* mRNAs derived from chase experiments under steady-state and stress condition. Subsequently, we developed a numerical model based on a modified gamma distribution function, which estimated the transcriptomic deadenylation rate at 10 A/min. A simplified independent method, based on the delineation of quantile polyA-tail values, showed a correlation between the decay and deadenylation rates of individual mRNAs, which appeared consistent within functional transcript groups and associated with codon optimality. Notably, these rates varied during the stress response. Detailed analysis of ribosomal protein-coding mRNAs (RPG mRNAs), constituting 40% of the transcriptome, singled out this transcript group. While deadenylation and decay of RPG mRNAs accelerated under heat stress, their degradation could proceed even when deadenylation was blocked, depending entirely on ongoing nuclear export. Our findings support the general primary function of deadenylation in dictating the onset of decapping, while also demonstrating complex relations between these processes.

Keywords mRNA Deadenylation and Degradation; ONT Nanopore Direct RNA Sequencing (DRS); Ccr4-NOT and Pan2/3 Deadenylases; Pab1; Dcp2 Decapping and Xrn1 Degradation

Subject Categories RNA Biology; Translation & Protein Quality

https://doi.org/10.1038/s44318-024-00258-3

Received 28 February 2024; Revised 11 September 2024; Accepted 19 September 2024

Introduction

The 3' end polyadenosine tail (pA-tail) is an essential modification for the mRNA biogenesis in the nucleus and its cytoplasmic lifetime. In budding yeast, Pap1 polyA-polymerase synthesizes de novo 60 adenosine-long pA-tails, in a process governed by the Cleavage and Polyadenylation Factor (CPF) and the nuclear polyA-binding protein Nab2 (Turtola et al, 2021; Aibara et al, 2017; Rodríguez-Molina and Turtola 2022). A pA-tail of at least 40 adenosines is required to safeguard the mRNA from nuclear decay and facilitate its export to the cytoplasm (Dower et al, 2004). The protective function of the pA-tail is mediated by Nab2 (Schmid et al, 2015). The conserved Mex67-Mtr2 hetero-dimer mediates mRNA export to the cytoplasm (De Magistris, 2021). The inhibition of export by depleting Mex67 (Haruki et al, 2008) leads to a rapid decay of newly formed mRNAs, as Nab2 becomes sequestered on non-exported mRNAs that accumulate in the nucleus (Tudek et al, 2018).

Two main complexes, Pan2/3 and Ccr4-NOT, mediate cytoplasmic pA-tail deadenylation. The mechanism of deadenylation has long been a scientific question. In mammals, a two-phase deadenylation model posits that Pan2/3 initiates pA-tail shortening, with Ccr4-NOT completing the process (Chen and Shyu, 2010; Yi et al, 2018). However, in budding yeast, each complex has distinct preferred substrates, showcasing redundancy; yeast Pan2/3 predominantly targets mRNAs of high abundance, whereas Ccr4-NOT is recruited more efficiently to low-abundant mRNAs (Tudek et al, 2021).

Essential to mRNA metabolism, pA-tails in the cytoplasm are coated by the Pab1 polyA-binding protein (Brambilla et al, 2019). In vitro experiments have demonstrated that Pab1 significantly contributes to regulating pA-tail length by stimulating Ccr4-NOT-mediated deadenylation of a 60 A substrate and temporarily inhibiting deadenylation on a 30A RNA (Webster et al, 2018). Other in vitro assays revealed that Pan2/3 more efficiently deadenylates longer pA-tails than the shorter (20–25) ones.

---

[1]International Institute of Molecular and Cell Biology, Księcia Trojdena 4, 02-109 Warsaw, Poland. [2]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Adolfa Pawińskiego 5A, 02-106 Warsaw, Poland. [3]Department of Life Technologies, University of Turku, Biocity, Tykistökatu 6, 205240 Turku, Finland. [4]University of Warsaw, Faculty of Biology, Miecznikowa 1, 02-089 Warsaw, Poland. [5]Aarhus University, Department of Molecular Biology and Genetics—Universitetsbyen 81, 8000 Aarhus, Denmark. [6]These authors contributed equally: Agnieszka Czarnocka-Cieciura, Jarosław Poznański. ✉E-mail: adziembowski@iimcb.gov.pl; atudek@ibb.waw.pl

Importantly, Pab1 can stimulate Pan2/3 activity without being required for deadenylation (Wolf et al, 2014). Despite multiple efforts, the precise in vivo role of Pab1 in mRNA pA-tail metabolism has not been fully established, although it has been implicated in deadenylation based on reporter mRNA assays (Caponigro and Parker, 1995). The coexistence of Pan2/3 and Ccr4-NOT, potentially influenced by Pab1, complicates generalizations based solely on in vitro studies, substantiating the need for a comprehensive in vivo approach. In addition to Pab1, translation was proposed as a major regulator of the deadenylation rate. The translation initiation factor complex is thought to bind to Pab1, forming a loop that may influence translation initiation and mRNA degradation regulation (Tarun et al, 1997; Otero et al, 1999; Archer et al, 2015). Moreover, codon optimality, which is dictated by the relative concentration of specific tRNAs, was shown to additionally regulate the deadenylation and decay rates of reporter transcripts (Presnyak et al, 2015).

Current models of mRNA decay, often derived from reporter systems, suggest that significant pA-tail shortening triggers decapping by Dcp1/2, followed by Xrn1 5'-3' exonuclease-mediated decay (De Magistris, 2021; Chen and Shyu (2010); Yi et al, 2018; Decker and Parker, 1993). Here, our objective was to validate these models using transcriptome-wide in vivo data. To achieve this, we experimentally modeled deadenylation and decay rates under steady-state and stress conditions using data derived from Nanopore Direct RNA Sequencing (DRS). Our approach focused on describing the evolution of pA-tail length distribution across the entire coding transcriptome or specific transcript groups. To this end, we established two analysis methods. The first utilized a modified gamma distribution model combined with Mean Field Theory (Gupta and Groll, 1961; Réka and Barabási; 2002) and determined that the transcriptomic enzymatic deadenylation rate is constant and equal to 10 A/min. The second, a simplified method, calculated deadenylation rates based on the evolution of quantile pA-tail length values. We found that the apparent adenosine half-life of individual transcripts varied from seconds to tens of seconds and strongly correlated with mRNA decay rates, indicating a functional link. Within this correlation, long-lived, mostly abundant mRNAs are deadenylated slowly, whereas short-lived transcripts experienced rapid tail shortening. Our study showed that mRNA decay and deadenylation rates are consistent across large functional groups of transcripts, such as those derived from ribosomal protein genes (RPGs), and are correlated with codon optimality, as previously established (Presnyak et al, 2015). Further analysis of RPG mRNA group metabolism under heat stress conditions showed that although both deadenylation and decay rates increased, deadenylation is not a prerequisite for mRNA decapping and decay but is instead a major stimulating factor. RPG mRNA decay proceeds through complete deadenylation inhibition in a double *ccr4Δ pan2Δ* mutant and is strongly dependent on ongoing mRNA export. Therefore, we conclude that while deadenylation primarily dictates decapping onset, other factors can dominate in significant functional mRNA groups, such as RPGs, which constitute 40% of the coding transcriptome.

Results

mRNA nuclear export block reveals cytoplasmic deadenylation and decay

Previous studies have estimated bulk yeast mRNA half-lives of 12 min or even lower (Miller et al, 2011; Sun et al, 2012; Neymotin et al, 2014; Presnyak et al, 2015; Chan et al, 2018). The efficient uncoupling of mRNA synthesis and degradation within the shortest possible time frame was critical for simultaneously modeling transcriptome-wide deadenylation and decay. Drawing from our previous work (Tudek et al, 2018, Schmid et al, 2018), we employed the Anchor-Away system (Haruki et al, 2008), to deplete the main cellular export factor, Mex67, thus rapidly inducing massive nuclear degradation of newly synthesized mRNAs. Concurrently, already-exported mRNAs continued their metabolism in the cytoplasm, allowing the specific monitoring of deadenylation and degradation kinetics of this subset of transcripts (Fig. 1A). We generated three biological replicates, each containing one control sample and five test time points to increase modeling power (Fig. 1B). To measure both mRNA abundance and pA-tail lengths, we utilized the Direct RNA Sequencing (DRS) method (Brouze et al, 2023; Tudek et al, 2021).

Transcript degradation was evidenced by reductions in mRNA abundance, as shown for bulk mRNA distributions (Fig. EV1A) and single gene examples (Figs. 1C and EV1B). mRNA half-lives were calculated using a standard continuous exponential function (Dataset EV1; Miller et al, 2011; Sun et al, 2012; Neymotin et al, 2014; Chan et al, 2018). The median transcript half-life was 9.7 min for mRNAs and 2.6 min for ncRNAs (Fig. 1D). These calculated half-lives strongly correlated with those from two previous studies that utilized metabolic labeling (Miller et al, 2011; Chan et al, 2018; Spearman rho 0.74 and 0.62, respectively; Figs. 1E,F and EV1C,D), despite each dataset yielding a different median half-life, likely due to a systemic bias as suggested by Chan et al, 2018. We concluded that Mex67 depletion provides a suitable experimental setting for studying cytoplasmic mRNA decay.

Deadenylation rate differs by transcript abundance group

To describe global changes in mRNA pA-tail length over time, we generated whole-transcriptome distributions derived from the DRS datasets (as sum of all reads indiscriminately of transcript type; Figs. 2A and EV1E). Mex67 depletion resulted in a widespread shortening of pA-tails over time, with the most significant alterations observed in mRNA fractions with long pA-tails, as depicted in a density plot (Fig. 2A). Time-dependent violin plots further illustrated varied dynamics of pA-tail length change across quantiles, with values within the top quantiles decreasing more rapidly (Fig. 2A). Electrophoresis analysis of global pA-tail length distribution showed analogous dynamics following Mex67 depletion, thereby validating our methodology (Fig. 2B).

Subsequently, we separated the transcriptome into classes, which we hypothesized would exhibit varying rates of deadenylation; the main criterion for this classification was mRNA abundance. Consistent with our previous observations (Tudek et al, 2021), we confirmed that highly abundant mRNAs are characterized by shorter mean pA-tails and longer half-lives (Figs. 2C and EV1F,G). Strikingly, the 187 most enriched mRNAs in the DRS datasets (out of ~5000 total) accounted for about 60% of the reads. Within this group, mRNAs produced from ribosomal protein genes (RPG mRNAs) formed the largest functional subgroup, summing to 40% of all mRNA reads (Fig. 2C). These abundant mRNAs largely shape the distribution of the entire transcriptome (Figs. 2A and EV1H). In contrast, low-abundance mRNA clusters

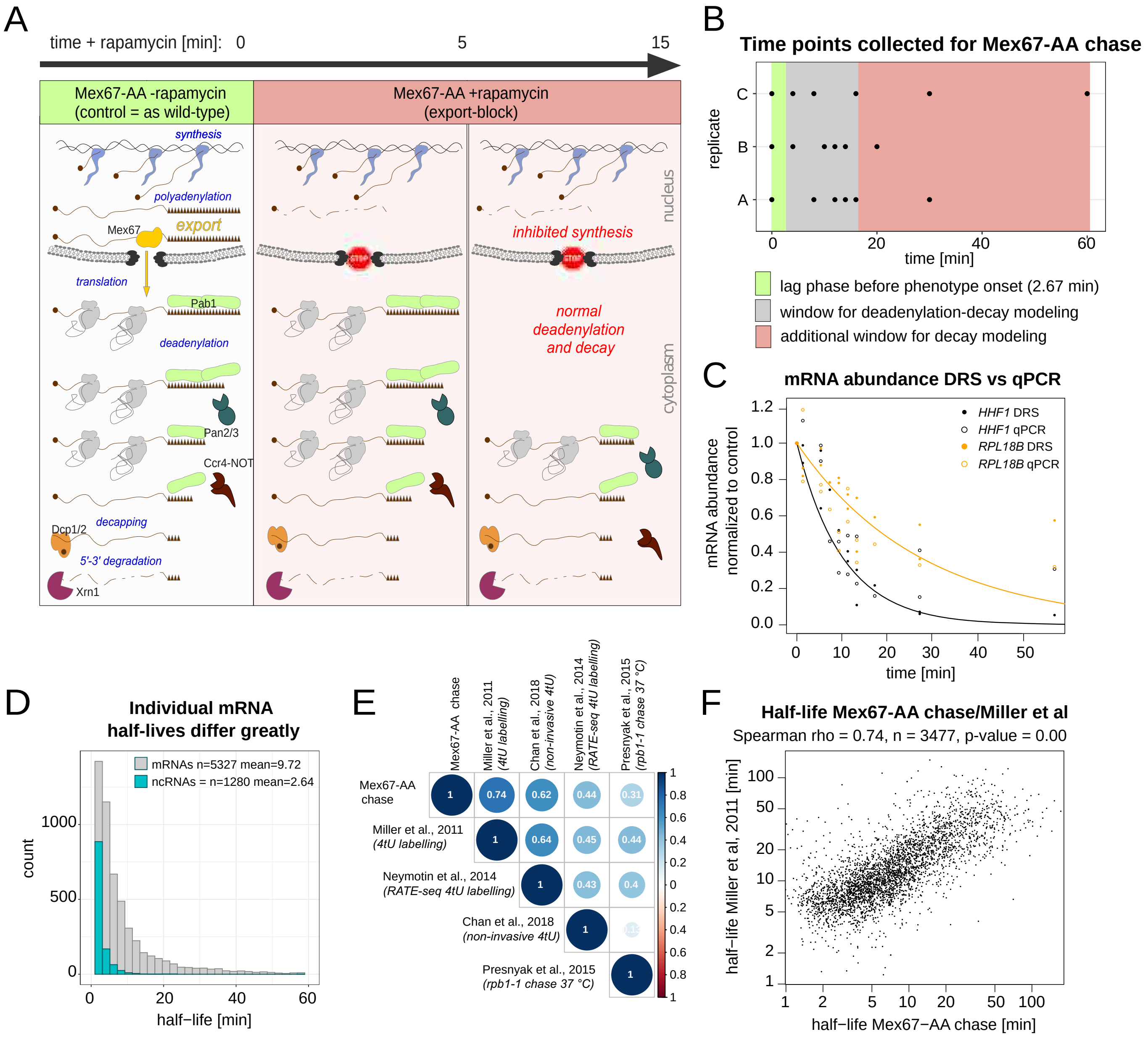


Figure 1. mRNA nuclear export reveals cytoplasmic deadenylation and decay dynamics.

(A) Schematic showing the theoretical experimental set-up. **(B)** Overview of collected time points and their relevance for decay or deadenylation modeling. **(C)** Time-dependent abundance change of *HHF1* or *RPL18B* mRNAs following Mex67 depletion in DRS datasets (dot) or by reverse transcription coupled with qPCR (ring). A fitted line represents the decay factor calculated from DRS data. **(D)** mRNA and ncRNA half-life distribution calculated from the Mex67-depletion time course. **(E)** Comparison of mRNA half-life estimations from various studies visualized as a matrix of Spearman rho coefficients (Miller et al, 2011; Neymotin et al, 2014; Presnyak et al, 2015; Chan et al, 2018). **(F)** Correlation between half-life values in Mex67-depleted sample and half-life values from Miller et al, 2011. Source data are available online for this figure.

exhibited more dynamic alterations in pA-tail lengths (Figs. 2D and EV1I,J), especially pronounced for mRNAs with long pA-tails at the initial time points. Indeed, low-abundance or non-RPGs mRNAs underwent deadenylation more rapidly than did highly abundant transcripts, including RPGs (Fig. 2E). This observation was supported by pA-tail distribution profiles of several individual genes, which exhibited notably diverse half-lives (illustrated in Fig. EV1K). We concluded that deadenylation rates are transcript-specific, with low-abundance mRNAs being deadenylated more rapidly and consequently exhibiting shorter half-lives compared to high-abundance transcripts.

Decapping dominates over deadenylation for mRNAs with pA-tails shorter than 20 As

One notable observation regarding the pA-tail length distribution in Mex67-depleted cells is the absence of tails shorter than 20 adenosines. This can be illustrated by a rapid decline in the

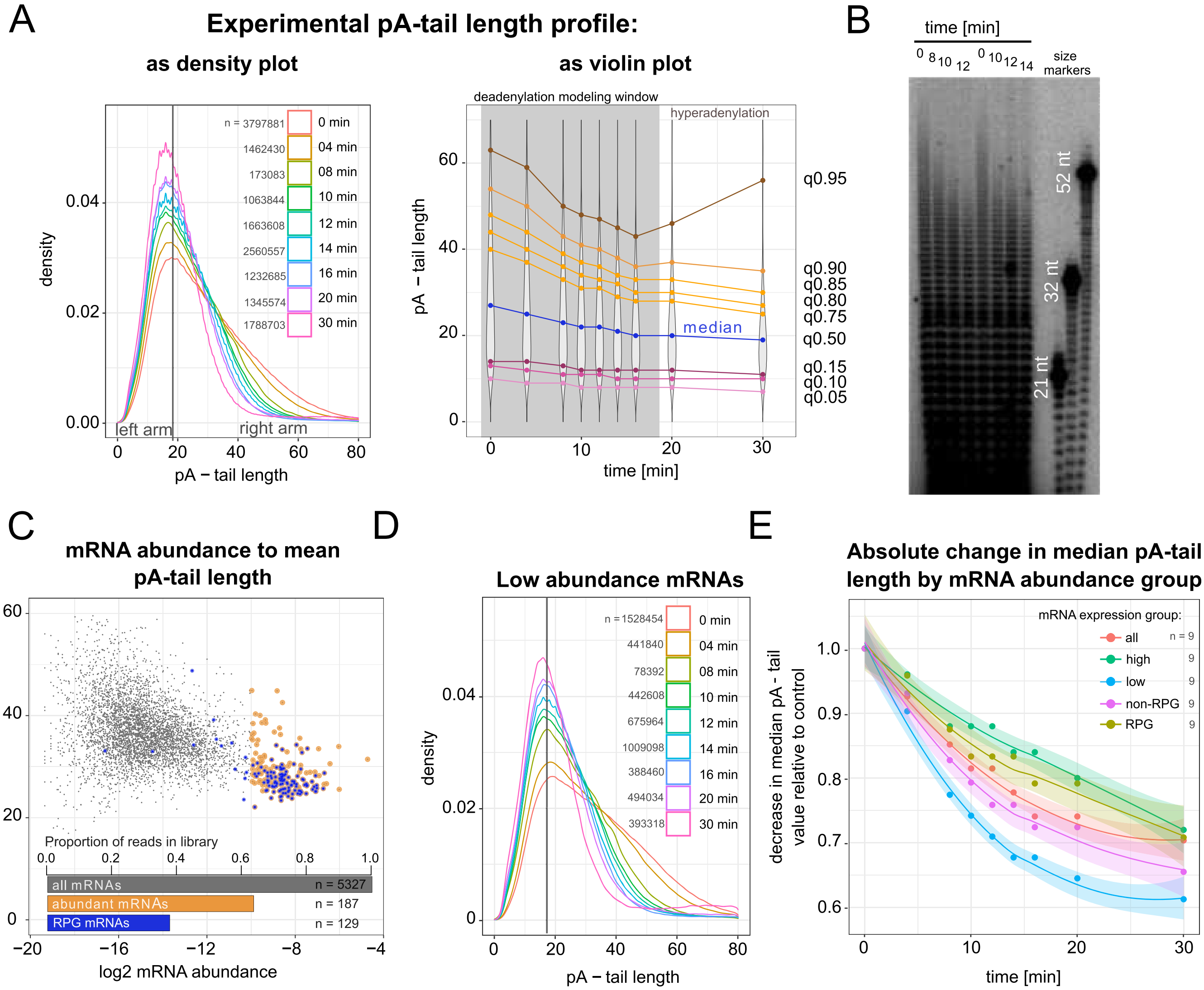


Figure 2. mRNA groups exhibit various deadenylation rates.

(A) Global distribution of mRNA pA-tail lengths during the Mex67-chase experiment. DRS data are presented in a density plot (left) or violin plot (right). Replicates, shown separately in Fig. EV1E, were merged. The number of transcripts (pA-tail estimates) in each density plot is given on the panel. As in Fig. 1B, the violin plot also highlights in gray the timepoints used for deadenylation rate modeling. The latter timepoints were discarded due to the occurrence of mRNA hyperadenylation; a phenotype specific to very few newly made mRNAs in nuclear export-block conditions (Jensen et al, 2001). These scarce species only significantly impact the overall pA-tail distribution after most cytoplasmic mRNA have been deadenylated and degraded. **(B)** Autoradiogram depicting the global distribution of pA-tail lengths in selected Mex67-chase samples used to construct the sequencing library. **(C)** Log2 mRNA abundance compared to mean pA-tail length for control Mex67-AA cells, with highly abundant and RPG mRNAs as gold and blue dots, respectively. A bar plot summarizes the read fraction for each transcript category. **(D)** Global distribution of pA-tail lengths of low-abundance mRNAs. The number of transcripts in each density plot is given on the panel. **(E)** Time-dependent changes in median pA-tail length across the entire coding transcriptome and four large mRNA groups: high-low abundance, and RPG-non-RPG. Local regression trend lines are shown with a 95% confidence interval. The number of pA-tail estimates used to calculate the median for each point is given in (**A**, **D**) and Fig. EV1E,H,J. Replicates were merged to produce nine median estimates for various time points. Source data are available online for this figure.

distribution curve at 20 adenosines and a complete absence of tails shorter than 15 for individual mRNAs (Figs. 2A,D and EV1E,H,J,K). Moreover, even at the later time points of Mex67 depletion, the left side of the distribution curve remained almost unchanged.

To investigate the biological relevance of this observation, we depleted the decapping enzyme Dcp2 or the Xrn1 5'-3' exonuclease for 2 h using the auxin-inducible degron (AID) system (Appendix Fig. S1A,B; Nishimura et al, 2009; Morawska and Ulrich, 2013). Depletion of 5' decapping and decay factors shifted the pA distribution of all mRNAs and individual transcripts towards shorter-tailed values (Fig. 3A,B; Appendix Fig. S1C–E). The DRS method has reduced efficiency in detecting mRNAs with short pA-tails (Fig. 3A; Tudek et al, 2021), potentially underestimating the global pA-tail shortening in Dcp2- or Xrn1-depleted cells. Nonetheless, the observed accumulation of short-tailed mRNAs was significant under Dcp2-/Xrn1-depletion. Since this was not observed in the Mex67-depletion datasets, we can safely conclude that in yeast, deadenylation continues until the pA-tail reaches 20

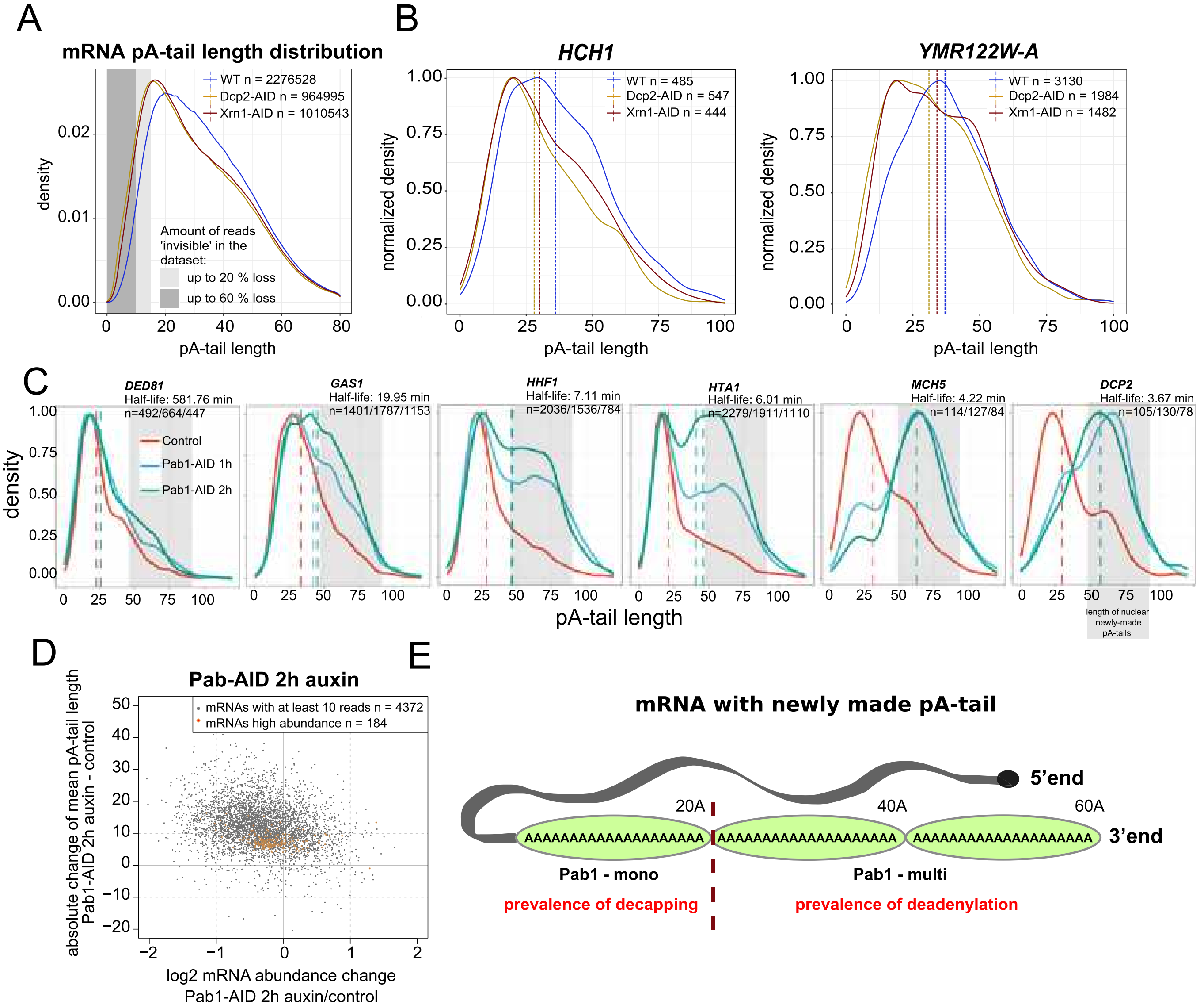


Figure 3. Pab1 controls deadenylation and decapping.

(**A**) Global distribution of pA-tail lengths of RNAs isolated from control, Dcp2- or Xrn1-depleted cells. Gray areas delineate the pA-tail lengths underestimated in the DRS library, as previously reported (Tudek et al, 2021). (**B**) pA-tail length distribution of *HCH1* and *YMR122W-A* mRNAs in control and Dcp2- or Xrn1-depleted cells. (**C**) pA-tail length distribution of *DED81*, *GAS1*, *HHF1*, *HTA1*, *MCH5*, and *DCP2* mRNAs in control cells compared to strains depleted of Pab1 using the AID system for 1 or 2 h. The mRNAs are ranked by half-life time. The number of reads contributing to each distribution is indicated in the panel with *n* = 'control'/'1 h depletion'/'2 h depletion'. The gray boxes show the range of adenosines added de novo in the nucleus by the polyA-polymerase Pap1 on pre-mRNAs (Turtola et al, 2021). (**D**) Comparison of log2 fold change in mRNA abundance to absolute change in mean pA-tail length for 2 h Pab1-depleted cells compared to control. (**E**) Schematic illustrating that one Pab1 can bind from 20 to 30 adenosines, therefore newly synthesized mRNA pA-tail bear di- or trimers of Pab1. Depending on the number of Pab1 attached to the pA-tail, the mRNA is either susceptible to deadenylation of decapping. Source data are available online for this figure.

adenosines in length, whereafter it triggers decapping, consistent with previous findings in mammalian cells (Eisen et al, 2020). To validate our observation, we performed in vitro digestion of a wild-type RNA sample with Xrn1, predicting the removal of uncapped transcripts to reveal the preferred pA-tail length at which decapping occurs (Appendix Fig. S1F). However, the whole mRNA pA-tail distribution did not change, indicating that in a wild-type context, uncapped transcripts are rare (Appendix Fig. S1G,H), which aligns with the belief that exonucleolysis promptly follows decapping, consistent with physical associations between Dcp1/2 and Xrn1 (Braun et al, 2012; Charenton et al, 2017). However, upon closer inspection of individual mRNA p-tail profiles, we observed digestion of a small fraction of short-tailed mRNAs (Appendix Fig. S1I), indicating potential decapping following pA-tail shortening.

Pab1 stimulates deadenylation when bound as a multimer

We examined the role of Pab1 in deadenylation based on several observations. First, previously published in vitro studies suggested that Pab1 regulates deadenylation by physically interacting with Ccr4-NOT and Pan2/3 deadenylases (Webster et al, 2018; Schäfer et al, 2019).

Second, we observed that the peak of the pA-tail distribution for both whole-transcriptome and single mRNAs fell between 20–30 As; a length which was previously shown to be bound by a single Pab1 molecule (Webster et al, 2018; Baer and Kornberg, 1980; Schäfer et al, 2019). Therefore, we reasoned that Pab1 may regulate the balance between decay and deadenylation depending on whether it binds the pA-tail as a monomer or a multimer. To test this hypothesis in vivo, we depleted Pab1 using the AID system for 1 and 2 h, attaining a 40% and 20% decrease in protein, respectively (Appendix Fig. S1J; Nishimura et al, 2009; Morawska and Ulrich, 2013). Inspection of the whole-transcriptome distribution revealed a slight and gradual accumulation of long pA-tailed mRNAs (50–80As) during Pab1 depletion (Appendix Fig. S1K). Next, we inspected single mRNA examples representing a wide range of mRNA half-lives (Fig. 3C; Appendix Fig. S1L). Polyadenosine distributions of mRNAs with long half-lives mimicked the whole-transcriptome distribution. In contrast, mRNAs with short half-lives showed a marked increase in long pA-tailed mRNAs, at times completely replacing the control pA-tail distribution. The length of accumulated pA-tails corresponded to the estimated length of newly synthesized pA-tails (Turtola et al, 2021; Tudek et al, 2021). Since budding yeast lack cytoplasmic adenylases, these accumulated mRNAs must originate from de novo transcription following Pab1 depletion. They gradually replaced mRNAs produced ulteriorly, which was clearly more pronounced for short-lived mRNAs. This suggests a global effect, as all mRNAs displayed an increase in mean pA-tail length (Fig. 3D; Appendix Fig. S1M). Since the long-tailed mRNAs tended to accumulate, we reasoned that they did not undergo deadenylation (Fig. 3E). This highlights the in vivo role of Pab1 in promoting deadenylation, consistent with previous in vitro and in vivo analyses performed on a much smaller scale (Schäfer et al, 2019; Webster et al, 2018; Sachs and Davis, 1989; Caponigro and Parker, 1995). Combining these findings with our previous analyses (Fig. 3A,B), which showed that decapping likely occurs around the most prevalent pA-tail length (20–30 A), we reasoned that mRNAs with a single Pab1 are more susceptible to decapping rather than deadenylation. Conversely, mRNAs with more than one Pab1 are preferentially deadenylated (Fig. 3E). These biological insights were crucial for refining our deadenylation modeling strategies.

A modified gamma distribution describes mRNA pA-tail profiles in yeast

To develop a numerical model for deadenylation, our initial step involved fitting experimental distributions of pA-tails to a single function. This function required adaptable parameters that, when systematically adjusted, would accurately reflect the entire spectrum of experimental distributions. We selected the widely used two-parameter gamma distribution (Gupta and Groll, 1961):

$$p(x) \sim x^{\gamma shape} \cdot e^{(-\gamma rate \cdot x)}$$

Importantly, a variant of the gamma function known as the Erlang distribution can be applied to discrete distributions such pA-tails. However, our model needed to account for specific characteristics of the experimental polyadenosine distribution, which sharply declines at 20 adenosines for all mRNAs, an effect of domination of decapping/decay in this range. This decline aligns with the fact that Pab1 binds minimally to 20 adenosines, and mRNAs lacking Pab1 are highly susceptible to decapping. Therefore, we modified the standard gamma function to consider the protective role of Pab1 against RNA decapping using Mean Field Theory (Réka and Barabási, 2002). To account for the effective impact of each Pab1 RRM binding to the pA-tail, we assumed that each adenosine binds to Pab1 with the same strength and that the binding force is additive, ultimately leading to a strong association of the entire Pab1 molecule with the pA-tail. Thus, interactions involving N_A indistinguishable objects (adenosine residues) were described using the following equation:

$$tanh(\beta \cdot N_A),$$

The β factor, affecting the shape of the distribution, was estimated as 0.096 through iterative adjustment of function fitting to the Mex67-AA depletion dataset. Gratifyingly, this function closely follows the left arm of experimental distributions, rapidly increasing from 0 and saturating at 20 adenosines (Fig. 4A, red line).

In the end, the combination of the classical gamma distribution with the saturable pA-tail:Pab1 interaction led to the final equation:

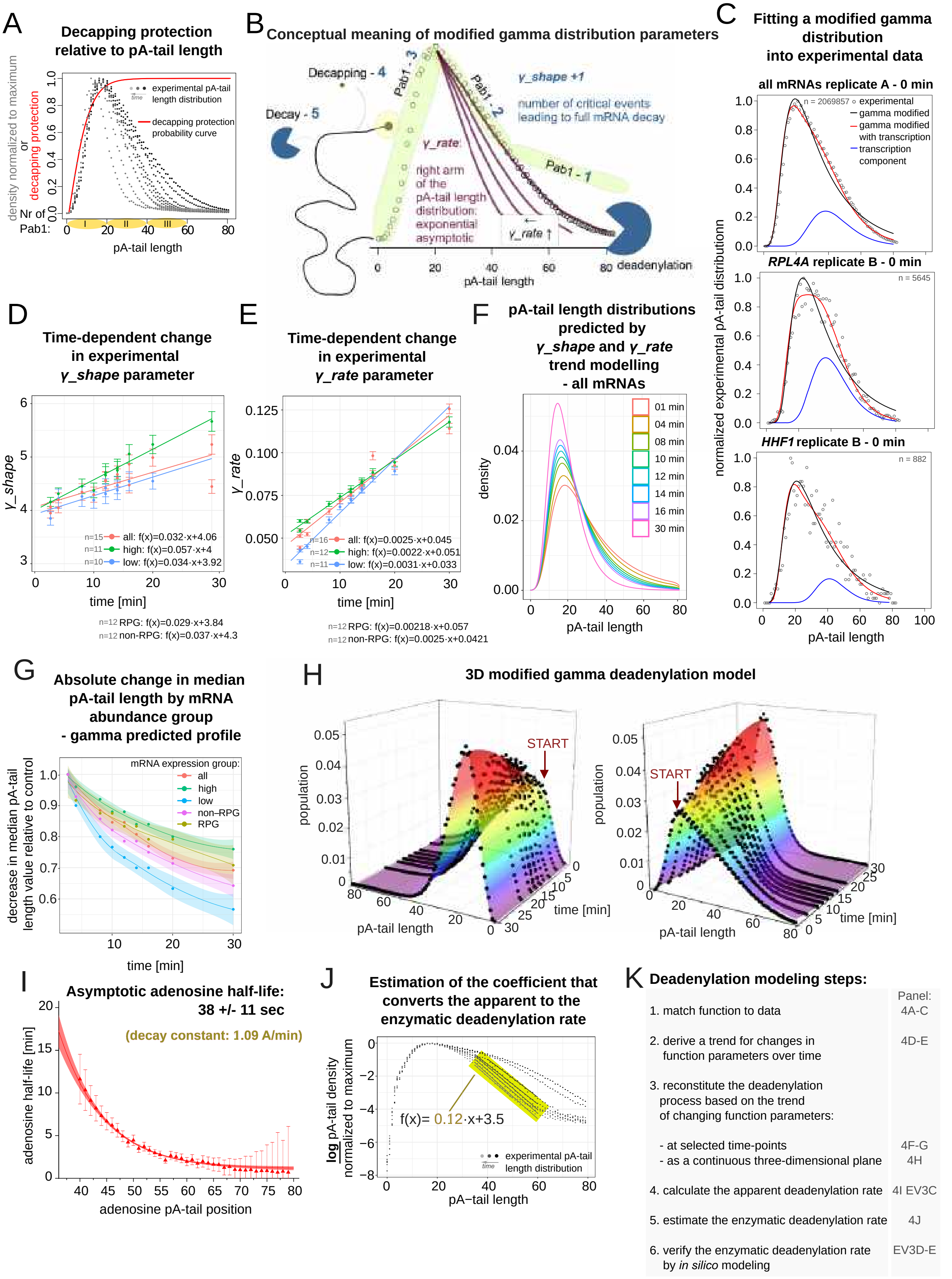
$$p(N_A) = tanh(0.096 \cdot N_A)^{\gamma shape} \cdot e^{(-\gamma rate \cdot N_A)}$$

This equation produced distributions that matched experimental data (Fig. EV2A). The form of the modified gamma distribution is governed by the two standard parameters, γ_rate and γ_shape (Fig. 4B).

The γ_rate can be conceptually linked to the balance between rates of nuclear adenylation and cytoplasmic deadenylation, as it primarily dictates the position and shape of the distributions' right arm, which represents the quantity of long pA-tailed mRNAs (Figs. 4B and EV2A). The model implies that the relative abundance of two RNA populations differing by a single adenosine in the pA-tail is constant ($e^{-\gamma\_rate}$), independent of the pA-tail length. The latter is supported by the asymptotic log-linear dependence observed in the experimental p(N_A) distribution.

The γ_shape parameter lacks a direct biological interpretation, but within our experimental conditions, the interplay between γ_shape and γ_rate dictates the position of the distribution maximum. However, the Erlang distribution interprets the $shape + 1$ value as the number of critical events following the Poisson distribution required to complete a process. In the biological system we studied, these critical events correspond to the removal of three Pab1 molecules (each binding to 20 As of a 60 A newly made pA-tail) during deadenylation, followed by decapping and rapid mRNA decay mediated by Xrn1. Therefore, the total number of known events leading to mRNA disintegration sums to five. Accordingly, the experimental γ_shape values of the control pA-tail distributions ranged between 3.5 and 4, reflecting a number of critical events ($\gamma\_shape + 1$) close to five (Figs. 4B and EV2A).

Having established a function to describe pA-tail distributions, we fitted it to experimental data from each time point of the Mex67-depletion experiment (Figs. 4C and EV2B). The distributions observed after Mex67 depletion aligned well with the model. However, we observed noticeable deviations from the modified



Deadenylation modeling steps:

	Panel:
1. match function to data	4A-C
2. derive a trend for changes in function parameters over time	4D-E
3. reconstitute the deadenylation process based on the trend of changing function parameters:	
- at selected time-points	4F-G
- as a continuous three-dimensional plane	4H
4. calculate the apparent deadenylation rate	4I EV3C
5. estimate the enzymatic deadenylation rate	4J
6. verify the enzymatic deadenylation rate by *in silico* modeling	EV3D-E

◀ **Figure 4. The modified gamma model predicts mRNA pA-tail length distributions.**

(**A**) The probability of protection against decapping in function of pA-tail length is represented by the red line. Gray dots show experimental pA-tail length distributions from all Mex67-chase time points. The number of reads summing up to form the distribution is indicated in Fig. 2A. The presumed number of Pab1 subunits on a pA-tail of a given length is displayed. (**B**) Conceptual scheme illustrating the modified gamma distribution parameters (see main text). *γ_rate* corresponds to the exponential coefficient of the right arm of the pA-tail distribution, whereas *γ_shape* + 1 represents the number of critical events leading to complete mRNA decay. (**C**) Using gray dots, three density plots show experimental pA-tail length distribution for whole mRNA transcriptome in control cells for replicate A and in control samples for *RPL4A* and *HHF1* mRNAs. Black lines represent fitted modified gamma distributions. Estimation of new mRNA production (by log-fitting) is shown using blue lines. The red line represents the sum of the modified gamma and transcription estimate distributions. the *n* number indicates the number od pA-tail estimations (reads) used to produce the experimental distribution. (**D**, **E**). Changes in the value of *γ_shape* (**D**) and *γ_rate* (**E**) parameters of the modified gamma probability distribution over time for all mRNAs or those of low and high abundance separately (RPG and non-RPG are omitted for clarity). The parameters are given as full-colored dots supplemented with vertical standard error bars. Each estimate was derived from distributions shown in Figs. 2A, D and EV1A, H–J (refer to those panels for the number of reads). Continuous functions, which were fitted are shown with the equations given. The number of dataset points considered for the estimation of the functions' equation is indicated on each panel. Those functions were used in subsequent modeling (**F**, **G** and Fig. EV2F–J). (**F**, **G**) Time-dependent evolution of whole-transcriptome pA-tail length distribution (**E**) or median (**G**) based on functions described in (**D**, **E**). (**H**) Modified gamma model of whole-transcriptome deadenylation over time in three-dimension shown as a continuous surface colored according to population density. Black dots show experimental pA-tail distributions. The number of reads summing up to form the distribution is indicated in Fig. 2A. (**I**) As red triangles with vertical error bars are plotted terminal adenosine half-lives calculated independently for each position relative to their rank in the pA-tail, calculated from distributions adjusted for mRNA decapping (Fig. EV3B). A continuous exponential function is fitted with a 0.95 confidence interval. (**J**) Experimental pA-tail distributions (same as in (**A**)) displayed on a normal logarithmic scale (gray dots). The linear distribution between pA-tail lengths 25–60 (yellow box) is highlighted. The slope of the function represents the correction factor used to calculate microscopic enzymatic deadenylation speed from the apparent one. (**K**) Table summarizing the steps of the deadenylation modeling performed using the modified gamma distribution. Source data are available online for this figure.

gamma profile specifically in control samples, particularly within segments containing longer pA-tails. We hypothesized that ongoing transcription may distort the modified gamma profile. To determine the form of the pA-tail fraction responsible for distorting the control samples, we subtracted the modified gamma distribution from the experimental data, obtaining distributions clustering around 40–60 As. These pA-tail lengths corresponded perfectly with the previously published estimates of newly synthesized mRNA pA-tail lengths, which were experimentally determined using various orthogonal methods (Tudek et al, 2021; Turtola et al, 2021). Thus, we conclude that the modified gamma profile effectively describes the experimental yeast pA-tail length distributions, and that any major deviations from the modified gamma distribution are primarily due to increased synthesis of new mRNA. Accordingly, the mean and variance of the fitted modified gamma distributions correlated well with the experimental data from Mex67 depletion (Fig. EV2C,D).

To ultimately validate our approach, we examined the time-dependent change in the *γ_shape* and the *γ_rate* parameters of the fitted modified gamma distributions across the entire transcriptome, its sub-groups (Figs. 4D,E and EV2E), and individual transcript examples (Fig. EV2F,G). Notably, these parameters exhibited coordinated changes over the course of Mex67 depletion. This allowed us to fit predictive functions for their evolution. Consequently, we generated theoretical pA-tail distributions for all mRNAs, as well as those with high and low abundance, which strongly resembled experimental distributions (compare Figs. 4F to 2A; compare EV2H to 2D and EV1H). Specifically, the model accurately replicated the time-dependent changes observed in experimental pA-tail distributions, including their mean, variance (Fig. EV2I), and, most importantly, the median (compare Figs. 4G to 2E). As expected, theoretical pA-tail distributions for individual gene examples (*HHF1*, *RPL4A*, *GAS1*, and *RPL36A*; Fig. EV2J) deviated more significantly from experimental data (likely due to lower read counts, affecting accurate estimation of pA-tail distributions; compare with Fig. EV1K). Nonetheless, these deviations clearly highlighted differences in deadenylation dynamics for each transcript.

We conclude that the modified gamma distribution accurately describes experimental mRNA pA-tail profiles, particularly for highly expressed mRNA groups, thereby providing a reliable numerical model to predict the evolution of pA-tail distributions in budding yeast.

Estimation of the whole-transcriptome microscopic (enzymatic) deadenylation rate from the modified gamma model

To visualize dynamic changes in the deadenylation process over time, we developed a three-dimensional model (Fig. 4H) based on the modified gamma functions fitted to experimental whole-transcriptome data data (Fig. EV3A). It is important to note that the model does not directly provide a precise deadenylation-decapping rate because distributions are internally normalized in each snapshot. To incorporate decapping-decay into our calculations, we instead used a coefficient derived from the decreasing amount of pA^+ RNA recovered from Mex67-depleted cells in each chase time point (Fig. EV3B). To calculate the deadenylation rate, we analyzed changes in the levels of each adenosine relative to the preceding nucleotide (N in relation to N + 1) using a series of 80 coupled differential equations (see "Methods" section). Within the range of 40–80 adenosines, where the contribution of decapping is minimal, the levels of each pA-tail length followed an exponential decay pattern (Fig. EV3C). Importantly, control samples matched this trend when shifted 160 s forward on the time scale (or 2.67 min; Fig. EV3C). We interpret this discrepancy as the time required for Mex67 export from the nucleus and the establishment of the export-block phenotype, marking the biological starting point of our depletion experiment. Therefore, our previous mRNA half-life calculations were adjusted to reflect this time difference (highlighted in Fig. 1B).

Then we plotted the half-lives of individual adenosines against their positions in the pA-tail, revealing an exponential relationship (Fig. 4I). Initially, this suggests that the rate of adenosine removal slows down as the pA-tail shortens. However, the decrease in adenosine half-life at position N is simply a consequence of the same chain length (*N*) being formed from the decay of a longer tail ($N + 1$). The half-lives of individual adenosines at positions 40–80, where the $N + 1$ contribution was minimal, converged to an asymptote. Importantly, this indicated that the deadenylation rate at the transcriptome level remains constant along the length of the pA-tail. The transcriptomic adenosine half-life equaled the value of the asymptote and was estimated at 38 +/− 11 s, translating to a disintegration constant of 1.09 (confidence limits 0.84–1.54) adenosine per minute (Fig. 4I). It is important to note that this rate represents an apparent deadenylation rate, describing changes across the entire pA-tail distribution. As stated above, in a mixed population of pA-tail lengths (X_{NA}), the microscopic (enzymatic) rate of removing the terminal adenosine from a pA-tail of a given length N_A is seemingly diminished by the deadenylation of a polyadenosine chain that is one adenosine longer (X_{NA+1} of $N_A + 1$, respectively). In other words, the loss of the total number of pA-tails of 39 adenosines (due to their deadenylation to 38 As) is apparently reduced by the products of decay in pA-tails of 40 adenosines to 39 As. The difference between the enzymatic (*R*) and apparent deadenylation rate (R_{app}) is encapsulated by the first-order reaction formula widely used in molecular biology:

$$\begin{aligned} -R_{app} \cdot X_{NA} = dX_{NA}/dt &= -R \cdot X_{NA} + R \cdot X_{NA+1} \\ &= -R \cdot X_{NA} \cdot (1 - X_{NA+1}/X_{NA}). \end{aligned}$$

We demonstrated that the distribution of pA-tail lengths (X_{NA}) is described by the continuous modified gamma distribution (Fig. 4C–H). Within the range of pA-tail lengths long enough to saturate interaction with at least one Pab1 molecule (i.e., longer than 20 As, or else the length of pA-tail protected by one Pab1 in footprint experiments, as reported in Webster et al, 2018, or shown to bind Pab1 in structural studies by Schäfer et al, 2019) the ratio X_{NA+1}/X_{NA} equals $e^{-\gamma\_rate}$, and the apparent deadenylation rate can be simply expressed as:

$$R_{app} = R \cdot (1 - e^{-\gamma rate}),$$

where *γ_rate* is a parameter of the modified gamma distribution estimated for the combined RNA population (Fig. EV3A). For short experimental times, *γ_rate* values were estimated to be ~0.1 (Fig. EV3A), implying a microscopic deadenylation rate of ~10 times higher, ~10.9 A/min (confidence interval: 8.4–15.4). Alternatively, the rate adjustment coefficient ($1 - e^{-\gamma\_rate}$) can also be roughly estimated as the average slope value of the linear functions fitted to the right arm of the experimental distributions, where normalized density is displayed on a log-normal scale (Fig. 4J; "Methods"; note that this factor is actually estimated from the global three-dimensional model).

Validation of the calculated microscopic deadenylation rate using in silico simulation of whole-transcriptome deadenylation

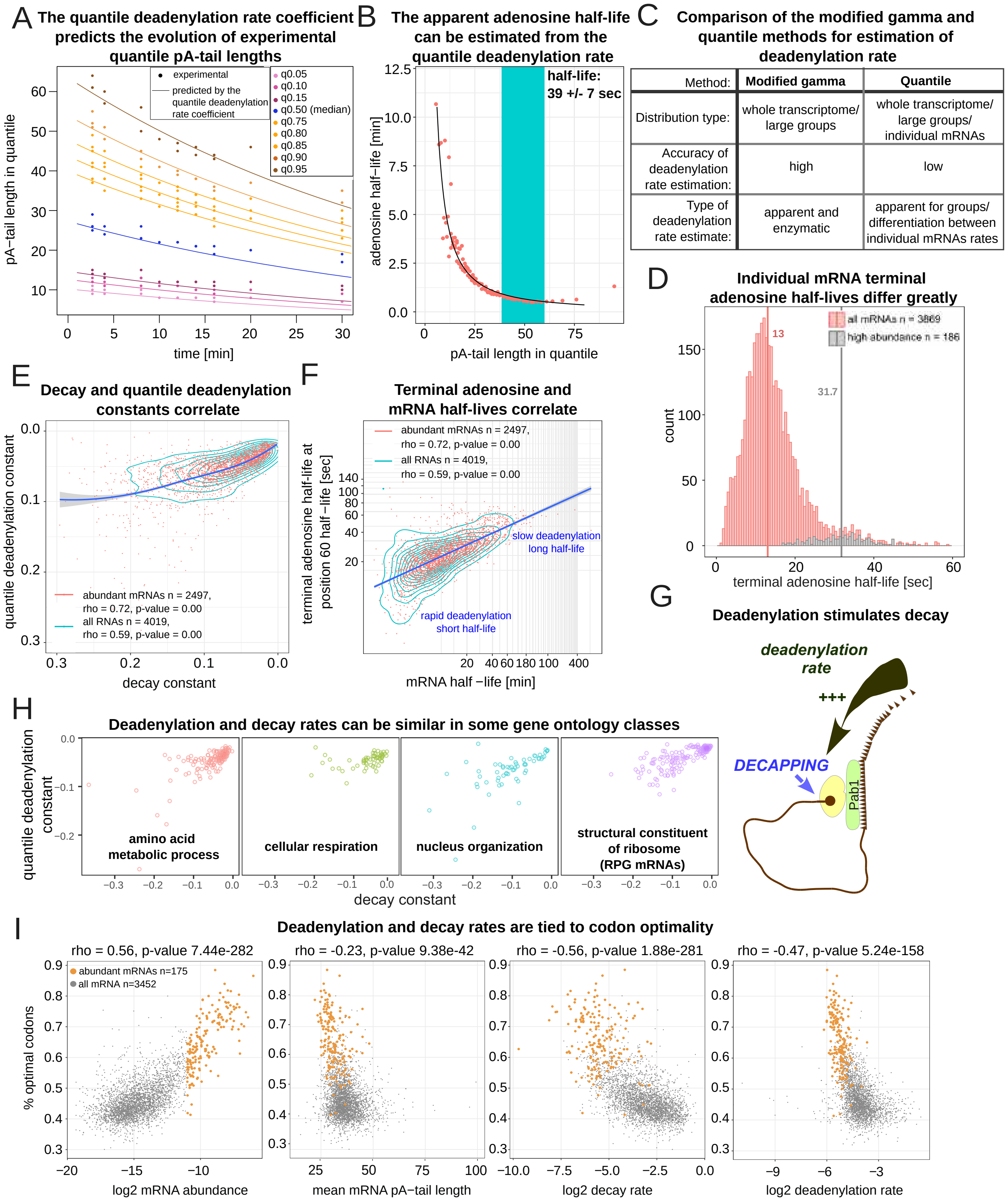
To further validate our model, we conducted deadenylation simulations (Fig. EV3D,E). We assumed that deadenylation is distributive and occurs at a constant speed (i.e., microscopic deadenylation rate). These simulations were based on a key factor, *α*, which is a dimensionless scaling factor that frames one round of deadenylation. Starting with control distribution data for the entire transcriptome (simulation in Fig. EV3D uses one of the control samples in EV1E), we iteratively applied multiple rounds of virtual deadenylation procedures (Fig. EV3D). Next, we compared these simulated results to actual experimental data collected at various time points after Mex67 depletion, aiming to identify the simulated distribution that best matched experimental observations. This comparison revealed a strong Pearson correlation ranging from 0.87 to 0.96 between the experimental time and the number of virtual deadenylation rounds applied in the simulation. (Fig. EV3E). This supports the notion that the microscopic deadenylation rate is constant and does not depend on the pA-tail length, at least within the range of pA-tails longer than 20 As (see "Discussion").

In sum, our modeling strategy yielded a common modified gamma equation that can describe deadenylation-related changes of yeast pA-tail distributions. This allowed us to reconstitute the deadenylation process and led to the calculation of the constant transcriptomic enzymatic deadenylation rate, which we have cross-verified using in silico simulation (Fig. 4K).

An alternative method to estimate deadenylation rates with small datasets

Estimating deadenylation rates using the modified gamma distribution model is effective only for pA-tail profiles with high read coverage because it relies on terminal adenosine half-life estimates. However, for most mRNA profiles, quantitative representation is insufficient. Therefore, to calculate the deadenylation rate of individual transcripts, we developed a method that provides a good approximation of the previous model. In this method, we describe the distribution of polyadenosine tail lengths by measuring specific quantile values, which significantly decrease over time during Mex67 depletion. This decrease is particularly noticeable in the upper quantiles (from the 75th to the 95th percentile). For more details, see the combined distribution in Fig. 2A and the individual data points in Fig. 5A. We fitted exponential functions to quantile pA-tail lengths (Figs. 5A and EV3F), obtaining deadenylation coefficients and calculating adenosine half-lives for each pA-tail length. Adenosine half-lives relative to quantile pA-tail lengths followed an exponential function (Fig. 5B), similar to the findings of the modified gamma model. Within the terminal adenosines, the average adenosine half-life was 39 −/+ 7 s (Fig. 5B), and the associated disintegration constant to 1.3 adenosine/min (Fig. EV3G), consistent with the apparent deadenylation rate derived from the gamma model.

In conclusion, we developed two separate methods to estimate the apparent deadenylation rate. The first method, using the modified gamma distribution, is precise and enables calculation of the enzymatic (microscopic) deadenylation rate. However, it requires large datasets and is unsuitable for analyzing individual transcripts. The second method, based on quantiles, works with a smaller number of reads (at least 10) but offers less precision. This quantile-based method does not directly measure the enzymatic deadenylation rate, but instead provides a numerical value useful for comparing the deadenylation rates among individual mRNAs, as shown in Fig. 5C.



Method:	**Modified gamma**	**Quantile**
Distribution type:	whole transcriptome/ large groups	whole transcriptome/ large groups/ individual mRNAs
Accuracy of deadenylation rate estimation:	high	low
Type of deadenylation rate estimate:	apparent and enzymatic	apparent for groups/ differentiation between individual mRNAs rates

Figure 5. At steady-state, deadenylation and decay rates correlate.

(**A**) Time-dependent change in quantile pA-tail lengths (as color-coded by quantile dots: upper—75-80-85-90-95th; median—50th; and lower—15-10-5th) for the entire coding transcriptome distribution in each replicate. Fitted continuous lines represent the quantile deadenylation coefficient calculated based on the upper quantiles. (**B**) Distribution of adenosine half-lives relative to pA-tail length in each quantile. The terminal adenosine half-life of 39 +/− 7 s was calculated by averaging the half-lives obtained from quantiles 75-95th (blue area). (**C**) Table summarizes key differences between the modified gamma and quantile methods for calculating deadenylation rate. (**D**) Histogram displaying the distribution of single mRNA terminal adenosine half-lives for the entire coding transcriptome (light-red bars) or mRNAs of high abundance (gray bars). Vertical lines indicate the group median. (**E**, **F**) Comparison of decay and quantile deadenylation rates (**E**) or log2-scaled mRNA half-life to terminal adenosine half-life (**F**). Density plot represents all transcripts, whereas mRNAs with at least 70 reads in replicate A are shown as individual dots with a blue regression line. Spearman's rho correlations and p-values were calculated separately for each set using the rstatix package in R. (**G**) Schematic highlighting the strong, potentially causal link between deadenylation and decapping inferred from data presented in (**E**, **F**). (**H**) Comparison of decay and quantile deadenylation rates for various gene ontology groups (see also Fig. EV4D). (**I**) Percentage of optimal codons compared to: log2 mRNA abundance, mean mRNA pA-tail length, log2 decay, or quantile deadenylation rates. Source data are available online for this figure.

Deadenylation and decay rates are positively correlated, and are linked to transcript functions and codon optimality

Our next step was to compare deadenylation rates among individual mRNAs. We averaged the deadenylation coefficients derived from the upper quantiles (75–80–85–90–95th), as these provided the highest estimates across the whole transcriptome (Fig. EV3F). This method accurately predicted changes in pA-tail length across all quantiles for single mRNA species (Fig. EV4A). We calculated the quantile half-lives of terminal adenosines for single mRNAs (Dataset EV1) and found a median of ~13 s (Fig. 5D). Remarkably, this median half-life was 32 s for the 186 most abundant mRNAs (Fig. 5D). We concluded that the transcriptomic deadenylation rate of 38 s is primarily dictated by a small subset of transcripts (see bar plot in Fig. 2C). Indeed, there was a strong correlation between quantile deadenylation rate and mRNA abundance (Fig. EV4B), though not with the mRNA mean pA-tail length (Fig. EV4C). Finally, we observed a strong correlation between decay and quantile deadenylation rates, with a Spearman rho coefficient as high as 0.72 (Fig. 5E,F), suggesting a causal relationship between both processes (Fig. 5G). In summary, we found that deadenylation rates, similar to mRNA half-lives, varied greatly between transcripts, with highly abundant mRNAs generally undergoing slower deadenylation and decay compared to less abundant transcripts, which were predominantly short-lived and underwent rapid tail removal.

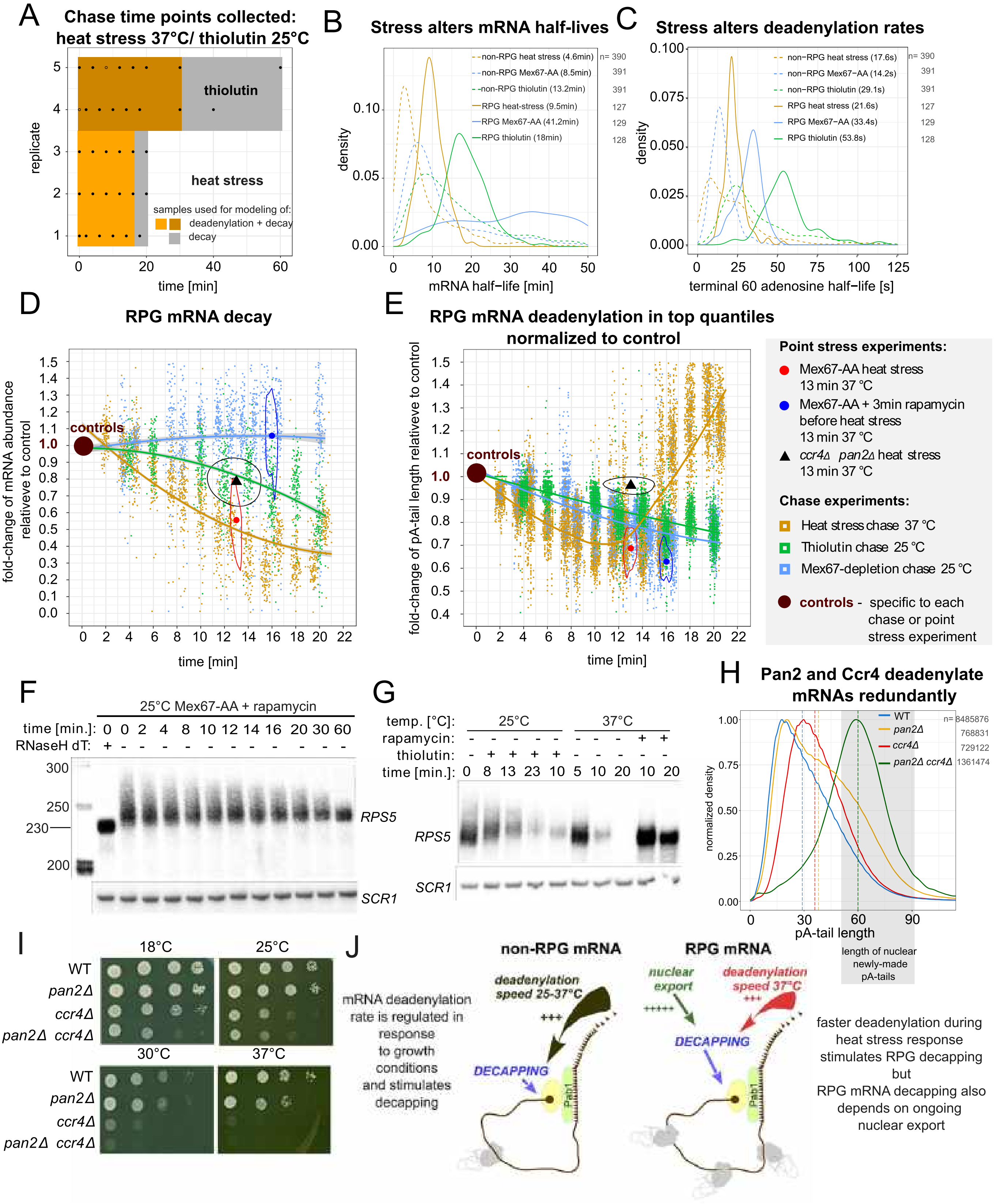
Subsequently, we explored how the distributions of decay and quantile deadenylation rates were shaped across different functional gene ontology (GO) clusters. mRNAs coding for housekeeping genes such as RPGs, factors responsible for amino-acid metabolic process, nucleus organization, and cellular respiration exhibited notably low decay and quantile deadenylation rates (Figs. 5H and EV4D). In contrast, most other GO term groups displayed more varied and generally rapid deadenylation and decay rates (Fig. EV4D), particularly mRNAs involved in regulating mitotic and meiotic events. This illustrates that deadenylation and decay processes in distinct functional mRNA groups can be coordinated.

Prior research postulated that translation rates govern the speed of mRNA degradation. In particular, mRNAs containing rare codons are deadenylated rapidly. In contrast, those with optimal codons display slower deadenylation rates (Presnyak et al, 2015; Harigaya and Parker, 2016; Radhakrishnan et al, 2016). Indeed, both decay and quantile deadenylation rates showed nonlinear correlations with codon optimality, with Spearman rho coefficients of 0.56 and 0.47, respectively (Fig. 5I). Subsequently, we compared our findings with the translation model developed by Siwiak and Zielenkiewicz (2010). We observed that quantile deadenylation and decay rates were less correlated with ribosome density than with the time required to translate a single mRNA (Fig. EV4E). This supports the hypotheses that slower translation, due to the presence of rare codons, can accelerate decay (Presnyak et al, 2015; Radhakrishnan et al, 2016; Webster et al, 2018), extending this effect to the deadenylation step. Off note, codon optimality is also positively correlated with mRNA level (Fig. 5I). Thus, abundant, mainly housekeeping transcripts, are easily translated, which slows their deadenylation, extends their productive lifespan, and ultimately increases protein production.

Deadenylation rate is not the dominant factor dictating decay of RPG mRNAs

Under optimal growth conditions, decay and deadenylation rates were strongly correlated (Fig. 5E–G). To explore potential functional and causative relationships further, we examined yeast cells under stress conditions (Fig. 6A), wherein RNA degradation plays a prominent role (Bresson et al, 2017; Marguerat et al, 2014). We conducted two experiments: a heat stress chase and a thiolutin chase. During the heat stress experiment, certain housekeeping genes, including RPGs, are transcriptionally silenced, while others, like those encoding chaperone mRNAs, become active (Vinayachandran et al, 2018). Thiolutin is a transcription inhibitor, which, at low doses, is known to activate the transcription of genes related to stress response (Adams and Gross, 1991). In the heat stress experiment, we quickly raised the temperature from 25 °C to 37 °C and monitored mRNA for up to 20 min (Fig. 6A). We confirmed the expected increase in chaperone mRNAs under both stress conditions by reverse transcription coupled to qPCR and in the DRS datasets (Appendix Fig. S2A,B).

In the Mex67-AA chase experiment, we analyzed approximately 125 RPG mRNA transcripts and observed similar deadenylation and decay rates among them (Fig. 5H). By aggregating the data for all RPG transcripts, we generated common pA-tail density profiles under various stress conditions (Appendix Fig. S2C–E). Significant differences in the overall RPG mRNA deadenylation rates prompted us to calculate decay and deadenylation rates for both RPG and non-RPG mRNAs in two new chase experiments. We presented those as density plots showing the half-lives of mRNA and terminal adenosines (Fig. 6B,C). In addition, we plotted the


A
Chase time points collected: heat stress 37°C/ thiolutin 25°C
replicate
thiolutin
heat stress
samples used for modeling of:
deadenylation + decay
decay
time [min]
B
Stress alters mRNA half-lives
density
non-RPG heat stress (4.6min) n= 390
non-RPG Mex67-AA (8.5min) 391
non-RPG thiolutin (13.2min) 391
RPG heat-stress (9.5min) 127
RPG Mex67-AA (41.2min) 129
RPG thiolutin (18min) 128
mRNA half−life [min]
C
Stress alters deadenylation rates
non-RPG heat stress (17.6s) n= 390
non-RPG Mex67-AA (14.2s) 391
non-RPG thiolutin (29.1s) 391
RPG heat stress (21.6s) 127
RPG Mex67-AA (33.4s) 129
RPG thiolutin (53.8s) 128
terminal 60 adenosine half−life [s]
D
RPG mRNA decay
fold-change of mRNA abundance relativeve to control
controls
time [min]
E
RPG mRNA deadenylation in top quantiles normalized to control
fold-change of pA-tail length relativeve to control
controls
time [min]
Point stress experiments:
Mex67-AA heat stress 13 min 37 °C
Mex67-AA + 3min rapamycin before heat stress 13 min 37 °C
ccr4Δ pan2Δ heat stress 13 min 37 °C
Chase experiments:
Heat stress chase 37 °C
Thiolutin chase 25 °C
Mex67-depletion chase 25 °C
controls - specific to each chase or point stress experiment
F
25°C Mex67-AA + rapamycin
time [min.]: 0 0 2 4 8 10 12 14 16 20 30 60
RNaseH dT: + - - - - - - - - - - -
300
250
230
200
RPS5
SCR1
G
temp. [°C]: 25°C 37°C
rapamycin: - - - - - - - - + +
thiolutin: - + + + + - - - - -
time [min.]: 0 8 13 23 10 5 10 20 10 20
RPS5
SCR1
H
Pan2 and Ccr4 deadenylate mRNAs redundantly
normalized density
WT n= 8485876
pan2Δ 768831
ccr4Δ 729122
pan2Δ ccr4Δ 1361474
pA-tail length
length of nuclear newly-made pA-tails
I
18°C 25°C 30°C 37°C
WT
pan2Δ
ccr4Δ
pan2Δ ccr4Δ
J
non-RPG mRNA
RPG mRNA
mRNA deadenylation rate is regulated in response to growth conditions and stimulates decapping
deadenylation speed 25-37°C +++
DECAPPING
Pab1
nuclear export +++++
deadenylation speed 37°C +++
DECAPPING
faster deadenylation during heat stress response stimulates RPG decapping but RPG mRNA decapping also depends on ongoing nuclear export

◀ **Figure 6. Deadenylation stimulates decay but is not a prerequisite for RPG mRNA decapping under stress conditions.**

(**A**) Time points collected for the thiolutin 25 °C and heat stress 37 °C chase experiments. Colored surfaces designate the utility of each time frame. Dot contours represent data points excluded from the analysis. (**B**, **C**) Distribution of mRNA half-lives (**B**) or terminal adenosine half-lives (**C**) in Mex67 depletion, thiolutin, and heat stress chase datasets for RPG and non-RPG mRNAs transcriptionally down-regulated during heat stress according to Vinayachandran et al (2018). The number of estimates in each group is indicated next to the figure legend. (**D**) Time-dependent change in levels of RPG mRNAs in Mex67 depletion, thiolutin, and heat stress time course normalized to control. Fitted lines show the group trend with a 95% confidence interval in gray. Data points normalized to corresponding controls obtained for 13 min heat stress of a double *ccr4Δ pan2Δ* mutant (black triangle) or Mex67-AA strain treated or not with rapamycin 3 min prior to heat stress for 13 min at 37 °C. For these data, the median is shown along with the data-point density contour. The experiment legend is listed next to panel 6E. (**E**) Time-dependent changes in pA-tail length values of the 95–90–85–80–75th quantiles normalized to the control sample for the same samples as in (**D**). Normalization allows direct comparison of all changes within the group without grouping transcripts by control pA-tail length. (**F**, **G**) Northern blot showing *RPS5* levels during a Mex67-depletion time course (**F**) or heat stress at 37 °C and thiolutin treatment at 25 °C (**G**). (**H**) Normalized pA-tail density plot showing the whole-transcriptome adenylation profiles in wild-type cells compared to single *ccr4Δ* or *pan2Δ* mutants and a double *ccr4Δ pan2Δ* strain. The gray box shows the range of adenylation produced de novo in the nucleus by the polyA-polymerase Pap1 on pre-mRNAs (Turtola et al, 2021). (**I**) Spot tests comparing wild-type cells to *ccr4Δ*, *pan2Δ*, and double *ccr4Δ pan2Δ* mutant at indicated temperatures. (**J**) Scheme illustrating the role of deadenylation in mRNA decapping and decay. For most mRNAs, deadenylation is a rate-limiting factor that dictates the onset of decapping. RPG transcripts are a special group for which deadenylation can accelerate decay in conditions such as heat stress, but ultimately, decapping and decay are activated by an unknown factor linked to nuclear export. Source data are available online for this figure.

raw data underpinning our calculations, including the time-dependent changes in RPG abundance (Fig. 6D) and changes in upper quantile pA-tail lengths, both normalized to control conditions (Fig. 6E). These analyses revealed accelerated deadenylation and decay of RPG and non-RPG mRNAs under heat stress compared to other conditions, as shown by specific examples (Appendix Fig. S2F; Fig. EV5A). Validation of these findings for selected RPG mRNAs was performed using Northern blotting and quantitative PCR (Fig. 6F,G; Appendix Fig. S2F), confirming that the rates of mRNA deadenylation and decay can be modulated in response to changing growth conditions. However, these chase experiments did not conclusively determine whether deadenylation is the primary factor defining the decay rate.

To delve deeper into the functional relationship between deadenylation and decapping/decay, we investigated the relative catalytic contributions of the three deadenylase catalytic subunits expressed in budding yeast cells. The Pan2/3 complex harbors one enzyme, Pan2, whereas the multisubunit Ccr4-NOT complex contains two deadenylases, Ccr4 and Pop2. We examined adenylation changes across the entire transcriptome profile and found that pA-tail lengths exhibited greater alterations in *pan2Δ* or *ccr4Δ* cells (on average by more than seven As for both; Tudek et al, 2021) than in the *pop2Δ* mutant (1.9 As; Fig. EV5B). Importantly, pA-tail length changes in *pop2Δ* and *ccr4Δ*, rather than *pan2Δ* and *ccr4Δ*, were highly correlated (Fig. EV5C), clearly indicating that Pop2 and Ccr4 both target similar transcripts within the same complex, even though Pop2's contribution to deadenylation was minor. Therefore, we conclude that Ccr4 and Pan2 are the primary cellular deadenylases in budding yeast (Tudek et al, 2021).

Next, we investigated the roles of Ccr4 and Pan2 in RPG mRNA deadenylation and decay during heat stress. Under steady-state conditions, RPG mRNA pA-tails were longer in *pan2Δ* cells compared to *ccr4Δ* (Fig. 6H; Tudek et al, 2021), suggesting these mRNAs are often targeted by the Pan2/3 complex. However, in a double *pan2Δ ccr4Δ* strain, mean pA-tail lengths were ~60.73 A, with the 5th and 95th quantiles at 40.08 and 90.89 adenosines, respectively (Fig. 6H). We concluded that very little, if any, deadenylation occurs in the cytoplasm of a double *pan2Δ ccr4Δ* strain because previous research indicated that the newly produced pA-tail in the nucleus are 60 As long with fail-safe mechanisms that limit adenylation by Pap1 to 90 adenosines (Turtola et al, 2021). We further tested this conclusion by comparing pA-tail lengths of a *ccr4Δ pan2Δ* mutants grown under steady-state conditions to those after a 13 min exposure to heat stress. While wild-type RPG pA-tails were significantly shortened compared to controls (Fig. 6E: yellow chase curve), minimal deadenylation levels were observed in the double *pan2Δ ccr4Δ* strain (Fig. 6E: compare black triangle to control; some underlying raw data Fig. EV5D). This confirmed that both deadenylases are required to shorten RPG mRNAs pA-tails during heat stress. Strikingly, RPG mRNAs in heat-shocked *pan2Δ ccr4Δ* cells were still degraded. The decay rate was similar to conditions in which wild-type cells were treated with thiolutin at 25 °C (Fig. 6D; compare the black triangle designating the *ccr4Δ pan2Δ* mutant to the green thiolutin chase curve). This observation suggested that decapping and subsequent decay can occur on long-tailed mRNAs, challenging the strictly causal link between mRNA deadenylation and decay. We also found that RPG mRNA decay was delayed in both single *pan2Δ* or *ccr4Δ* mutant strains, heat-stressed for 13 min (Fig. EV5E), further confirming that both deadenylase complexes contribute independently to the decay process. In contrast, the loss of *POP2* had no notable effects on RPG mRNA decay dynamics, underscoring the minor role of this deadenylase in decapping/decay (Fig. EV5E). Thus, we concluded that in a wild-type strain, accelerated deadenylation by either Pan2 or Ccr4 stimulates RPG mRNA decay under heat stress, however it is not critical for decapping and degradation. Notably, the observation that decay is not completely inhibited when deadenylation is blocked explains the viability of the double *ccr4Δ pan2Δ* mutant (Fig. 6I).

Finally, we explored additional factors potentially triggering RPG mRNA decay beyond deadenylation. Initially, we considered the role of exosome-mediated 3'–5' decay that could be triggered by changes in the translation machinery at high temperatures. However, we ruled this out because the deletion of the exosome-associated helicase Ski2 did not inhibit RPG mRNA decay (Fig. EV5E; Tomecki et al, 2023). Instead, we observed that RPG mRNA decay under heat stress was strongly inhibited by the depletion of Dcp2 (Fig. EV5F) as well as the deletion of the decapping co-factors *DHH1* and *LSM1* (Fig. EV5E). Next, we investigated whether mRNA export affects decay by blocking export 3 min before applying heat stress. We observed that the decay of most RPG mRNAs was completely inhibited under these conditions (Fig. 6D: density around the median blue dot clustered at the level of its control, similarly to the entire Mex67-AA chase marked with a blue line; Figs. 6D and EV5F). This was surprising, as only a few RPG mRNAs decayed efficiently without export (e.g., *RPL4A*; Fig. EV5G). However, deadenylation of RPG mRNAs was accelerated under heat stress regardless of whether cells were export-blocked or not (Fig. 6E: compare density clusters around the blue and red dot representing export-blocked and wild-type cells to the heat stress chase curve marked in yellow).

In sum, we showed that during the heat response for RPG mRNA, both deadenylation and degradation are enhanced. However, under these specific conditions, deadenylation is not essential for decay. This deadenylation-independent pathway is strictly dependent on ongoing mRNA export. Altogether, our results suggest that whereas at the transcriptome-wide level, deadenylation and decay are strongly linked, functionally they can be uncoupled under specific stress conditions (Figs. 5E,F and 6J).

Discussion

The causal link between deadenylation and decapping is axiomatic to the cytoplasmic mRNA decay model. There are many transcriptomic estimates of mRNA decapping and degradation rates in yeast (Miller et al, 2011; Neymotin et al, 2014; Chan et al, 2018; Presnyak et al, 2015), however, attempts to model deadenylation are rare and limited to higher eukaryotes (Eisen et al, 2020). Here, we provide the first numerical model for deadenylation in budding yeast, which enabled high-precision estimation of the transcriptomic deadenylation rate of 10 A/min.

Our numerical model for deadenylation-dependent decay is based on a modified gamma distribution, used to describe experimental pA-tail length profiles. The $tanh(0.096{\cdot}N_A)$ formula inserted into the classical gamma distribution accounted for both the biological effects of the Pab1 function in protecting mRNA from decapping and in stimulating deadenylation. Specifically, stimulation of deadenylation occurs when multimeric Pab1 binds to long polyadenosine tails, whereas the protective function associated with monomeric Pab1 is lost when pA-tails are shorter than 20 As. This model is supported by previous in vitro and structural data, along with early in vivo studies performed on reporter mRNAs (Schäfer et al, 2019; Yi et al, 2018; Webster et al, 2018; Caponigro and Parker, 1995). In our study, we confirmed the essential role of Pab1 in mRNA deadenylation in vivo. The decapping onset set at a pA-tail length of 20 As is consistent with previous estimates in higher eukaryotes derived from PAL-seq2 data (Eisen et al, 2020), which we independently validated using Dcp2 or Xrn1 degron strains. Importantly, this does not negate the occurrence of deadenylation-independent decapping on specific mRNAs, particularly those that are translationally compromised (Tomecki et al, 2023). However, such aberrant transcripts are scarce compared to the functional coding transcriptome and do not constitute a mass large enough to distort the whole-transcriptome deadenylation profile. Therefore, the model does not account for those mRNA quality control events but does not exclude their existence.

Our model reveals a constant transcriptomic deadenylation rate of ten adenines per minute. This rate represents the average across the transcriptome, although individual transcripts can exhibit varying deadenylation rates. This variability is best highlighted by our in silico deadenylation simulation using a narrow pA-tail distribution (akin to a short transcription burst), which shows a gradual increase in distribution variance over time (Fig. EV5H). Although our model does not explicitly differentiate between individual mRNAs, this simulation highlights the potential for transcript-specific deadenylation rates. Notably, the precise role of Pab1 in deadenylation, particularly the function of each RNA recognition motif, cannot be fully modeled due to these complexities. The deadenylation rate holds significant implications for mRNA decapping-decay rates. For the first time we show that the in vivo time required to sufficiently shorten the pA-tail to remove the last Pab1 can explain mRNA half-life times (decapping onset) in budding yeast. Newly synthesized pA-tails typically range from 60 adenosines to a maximum of 200 (Tudek et al, 2021; Turtola et al, 2021). Decapping is likely induced on tails shorter than 20 adenosines, implying that a minimum of 50–40 adenosines must be removed before decapping, a process that likely occurs within minutes. This time frame aligns with the median mRNA half-life of 9 min, suggesting that deadenylation may be slightly faster than decay. Any differences may stem from the time required to export the newly synthesized mRNA out of the nucleus, which effectively extends its lifetime (Dargemont and Kühn, 1992; Mor et al, 2010; Bahar Halpern et al, 2015), although technical considerations cannot be ruled out.

In our study, we also present a simple method for calculating mRNA deadenylation rates based on quantiles of the pA-tail distribution. This approach allows for the estimation of deadenylation rates from small datasets (representing the pA-tail length distributions of individual transcripts) with an accuracy comparable to that of the modified gamma numerical model. This method was instrumental in unveiling several important biological observations. We established the link between codon optimality and deadenylation at the transcriptomic level, which has previously only been observed in reporter mRNAs (Presnyak et al, 2015). We also explored the longstanding hypothesis regarding a causal link between deadenylation and decapping using a model group of RPG mRNAs. We showed that under heat stress, deadenylation can accelerate, thereby shortening mRNA half-lives, which supports the classic deadenylation-decapping model. However, we also showed that deadenylation is not a prerequisite for decapping/decay, but rather serves as a major stimulatory factor. This is supported by our comprehensive analyses under stress conditions, revealing that inhibition of cytoplasmic deadenylation can lead to decapping on mRNAs with long pA-tails, suggesting the existence of additional regulatory mechanisms that can induce decapping independent of deadenylation. Specifically, for RPG mRNAs, one such mechanism appears to be linked to ongoing nuclear export. Clearly, additional research is necessary to unveil the molecular mechanisms governing the regulation of deadenylation and decapping rates. Although some mechanisms may be temperature-sensitive, where a rise in temperature boosts any enzymes' activity, others are presumably specific to transcript groups. In particular, RPG mRNAs and many other housekeeping transcripts exhibit longer lifespans compared to the transcriptomic median (Miller et al, 2011; Neymotin et al, 2014; Chan et al, 2018; Presnyak et al, 2015). Preserving the lifespan of mRNAs encoding essential cellular components, regardless of growth conditions, is likely advantageous to reduce the energy cost associated with mRNA synthesis.

Methods

Reagents and tools table

Reagent/resource	Reference or source	Identifier or catalog number
Experimental models		
Saccharomyces cerevisiae wild-type W303 and its derivatives	Appendix Table S1 page 10	Appendix Table S1 page 10
Recombinant DNA		
pUC57	Addgene	https://www.addgene.org/vector-database/4509/
pET-28 b (+)	Addgene	https://www.addgene.org/vector-database/2566/
Antibodies		
Anti-FLAG	Sigma	F1804 (RRID:AB_262044)
Anti-PGK1	Novex LifeTechnology	discontinued
Anti-Pab1	Santa Cruz	sc57953 (RRID:AB_672248)
Anti-Rpb3: 1Y26	Abcam	ab81859
Goat Anti-Mouse IgG, H&L Chain Specific Peroxidase Conjugate	Merck	401215-2 ML
Goat Anti-Rabbit IgG, H&L Chain Specific Peroxidase Conjugate	Merck	401393-2 ML
Oligonucleotides and other sequence-based reagents		
Oligonucleotide list	Appendix Table S2 page 11	Appendix Table S2 page 11
Chemicals, enzymes, and other reagents		
Auxin sodium salt	Merck	I5148-2G
Chloroform was added	Sigma	C2432
Dynabeads™ Oligo(dT)$_{25}$	LifeTechnology	61005
1-Naphthaleneacetic acid (1-NAA)	Merck	N0640-25G
Phenol solution saturated with 0.1 M citrate at pH 4.3	Sigma	P4682
Platinum SYBR™ Green qPCR SuperMix-UDG	LifeTechnology	11733046
Rapamycin	Cayman Chemicals	13346
RNase A	LifeTechnology	R1253
RNase H	New England Biolabs	M0523
RNase T1	LifeTechnology	EN0542
RT qPCR Mix SYBR	A&A Biotechnology	2008-1000
RiboLock RNase inhibitor	LifeTechnology	EO 0382
SuperScript IV reverse transcriptase	LifeTechnologies	18090050
Thiolutin	Sigma	T3450
T4 RNA Ligase	LifeTechnologies	EL0021
ULTRA-Hyb Oligo Hybridization buffer	Invitrogen	AM8663
[5′-32P]pCp (3000 Ci/mmol, 10 mCi/ml)	Hartmann Analytic	https://www.hartmann-analytic.de/gamma-p32-atp-3000-ci-mmol-10-mci-ml-fp-301.html
Software		
Code to estimate the *γ_shape* and *γ_rate* parameters of the modified gamma distribution based on experimental distributions	Mendeley data and Appendix section 5 page 27	https://doi.org/10.17632/2j3hh37zzs.1 and Appendix section 5 page 27
Guppy	Oxford Nanopore Technology (ONT)	versions listed in Dataset EV2 (4.4.1; 5.0.11; 6.0.0)
ImageJ software	https://imagej.net/	versin 1.50i version 1.54j
Inkscape	https://inkscape.org/	version 1.3.2
Minimap2	https://github.com/lh3/minimap2	version 2.17
Nanopolish polya	https://github.com/jts/nanopolish	versions listed in Dataset EV2
Origin	www.originlab.com	version 10.0
R	The R Foundation for Statistical Computing https://www.r-project.org/	version 3.6.2 (2019-12-12) version R-4.2.1 (2022-06-23) version R-4.4.1 (2024-06-14)
RStudio	https://posit.co/products/open-source/rstudio/	version 1.2.5033 for Mac, 4.2.3, RStudio Team, 2019 version 2024.042 build 764 for Windows, Posit team, 2024
Other		
Direct RNA Sequencing kit	Oxford Nanopore Technology (ONT)	SQK-RNA002
FLA 9000 scanner	N/A	N/A
Hybond-N+ membrane	GE Healthcare	RPN303B
LightCycler LC480 Roche	Roche	N/A
Nitrocellulose membrane	Bio-Rad	1620112
PhosphorImager screen	FujiFilm	N/A
Whatman paper	Bio-Rad	1703967

Yeast culture conditions

Yeast cultures were prepared in YPDA media. The Mex67-AA and thiolutin chase experiments were conducted at 25 °C. To deplete Mex67 using the Anchor-Away tag (Haruki et al, 2008) rapamycin (Cayman Chemicals cat. no. 13346) was added to a final concentration of 1 μg/ml. The heat stress chase experiment was performed by pre-culturing cells at 25 °C and adding an equal volume of media pre-heated to 51 °C, resulting in a final temperature of 37 °C, which was maintained by culture incubation in a waterbath. Treatment with thiolutin (Sigma; T3450) was performed by adding the compound to a final concentration of 4 ug/ml. Samples for all chase experiments were collected by mixing the cell culture with an equal volume of ice-cold ethanol, which inactivates cellular metabolism. To deplete Xrn1 and Dcp2 using AID tags (Nishimura et al, 2009; Morawska and Ulrich, 2013), 1-Naphthaleneacetic acid (1-NAA; N0640-25G, Merck) was added to a final concentration of 1 mM. Pab1 depletion using the AID system involved adding auxin sodium salt (I5148-2G; Merck) to a final concentration of 1–3 mM. Yeast strains used in this study are listed in Appendix Table S1.

RNA extraction

RNA extraction was performed using the hot acid phenol method. Cell pellets were resuspended in 400 µl phenol solution saturated with 0.1 M citrate at pH 4.3 (Sigma; P4682), followed by the addition of 400 µl of TES buffer (10 mM Tris pH 7.5, 5 mM EDTA, 1% SDS). Samples were vortexed for 45 min at 65 °C, then centrifuged at 4 °C for 10 min. The supernatant was transferred to a fresh tube, and 400 µl phenol solution saturated with 0.1 M citrate at pH 4.3 was added. The samples were again vortexed for 20 min at 65 °C and then centrifuged at 4 °C for 10 min. The supernatant was transferred to a fresh tube and 400 µl of chloroform was added (C2432; Sigma). The samples were briefly vortexed at room temperature to remove phenol, and centrifuged at 4 °C for 10 min. The supernatant was transferred to a fresh tube; 45 µl of 2 M LiCl was added and 1 ml of 95% ethanol. Samples were precipitated at −80 °C for at least 30 min, then centrifuged at 4 °C for 25 min, washed with 400 µl of 80% ethanol, and dried at 37 °C after removing the supernatant. Pellets were resuspended in nuclease-free water, and RNA concentration was measured using a Nanodrop apparatus.

Enrichment of the pA$^+$ fraction for sequencing library preparation and qPCR analyses

The pA$^+$ fraction was prepared using magnetic beads coupled to oligo-dT from LifeTechnologies (61005). 35 µg of total RNA was resuspended in 50 µl of nuclease-free water. The RNA was mixed with 50 µl of binding buffer (20 mM Tris-HCl, ph 7.5, 1 M LiCl, 2 mM EDTA) and denatured for 2 min at 65 °C before cooling on ice. In total, 100 µl of slurry beads per 35 µg of total RNA was pre-washed three times in 1 ml of binding buffer and resuspended in 50 µl of binding buffer per sample. The beads were added to the denatured RNA and incubated at room temperature with occasional shaking for 20 min. The supernatant was removed, and beads were washed twice with 100 µl of wash buffer (10 mM Tris-HCl pH 7.5, 150 mM LiCl, 1 mM EDTA) and after removing any remnants of wash buffer, resuspended in 12 µl of nuclease-free water. The beads were then incubated for 2 min at 80 °C, and the supernatant removed from the beads was utilized for sequencing library preparation and/or qPCR analyses as the pA$^+$ fraction. Oligonucleotides used in this study are listed in Appendix Table S2.

Nanopore sequencing

DRS libraries were prepared using a Direct RNA Sequencing kit (ONT—Oxford Nanopore Technology, SQK-RNA002) according to the manufacturer's instructions, using 50–200 ng oligo-dT$_{(25)}$-enriched mRNA from *Saccharomyces cerevisiae* yeast mixed with cap-enriched or total RNA from other organisms (human, mouse, *A. thaliana*, or *C. elegans*), as described in Bilska et al (2020). Raw data were basecalled using Guppy (ONT). The raw sequencing data files (fast5) were deposited at the European Nucleotide Archive (ENA; for a list of accession numbers, see Dataset EV2).

Bioinformatic analysis

pA-tail length determination

DRS reads were mapped to the custom *S. cerevisiae* transcriptome described in Tudek et al (2021) using Minimap2 2.17 with options -k 14 -ax map-ont –secondary = no. These alignments were processed with samtools 1.9 to filter out supplementary alignments and reads mapping to the reverse strand (samtools view -b -F 2320). Any unmapped reads were filtered out and discarded from analysis. Poly(A) tail lengths for each read were estimated using the Nanopolish 0.13.2 polya function and only length estimates with the QC tag reported as PASS were considered in subsequent analyses (see "Quality Control details" in Appendix). Since the replicates strongly correlated with one another, unless otherwise indicated, samples from the same condition were analyzed together. Changes in mean/median pA-tail length and mRNA abundance were analyzed using R. Tables with the number of counts, mean, median, geometric mean pA-tail lengths, and quantiles are deposited at Mendeley data (https://doi.org/10.17632/2j3hh37zzs.1).

Modeling of the transcriptomic pA-tail distribution using the modified gamma distribution

Modification to the standard gamma distribution

The numeric model is based on the standard two-parameter gamma distribution, which has been widely used in life-testing for decades (Gupta and Groll, 1961):

$$p(x) \sim x^a \cdot e^{-b \cdot x} \tag{1}$$

where b reflects the rate of asymptotic exponential decay of the analyzed process, and a characterizes the shape of the distribution. The mean of the distribution is $(a+1)/b$, and its maximum is located at a/b (note that x does not necessarily denote time flow).

The pre-exponent factor in the model was further modified to quantitatively describe the limited contribution of RNA protection by the polyA-binding protein Pab1. We assumed that Pab1 recognizes n succeeding adenosine residues, with the free energy of interaction being identical and additive for each residue. This assumption implies an interaction with indistinguishable independent objects. Such a model has already been developed for the analogous problem of the n-spin system in the Mean Field Theory (Albert and Barabási, 2002; Kadanoff, 2009) in the form of:

$$p(n) \sim tanh(\beta \cdot n) \tag{2}$$

where n is the number of objects, and β characterizes the strength of individual interaction with the external field (here Pab1); formally, $\beta = \Delta G/RT$, where ΔG is the free energy of interaction, T is temperature, and R is the gas constant (8.13 J·K/mol). Combining the two above equations leads to the three-parameter modified gamma distribution expressed as:

$$p(i) \sim (tanh(\beta \cdot i))^{\gamma shape} \cdot e^{-\gamma rate \cdot i} \qquad (3)$$

where i is the number of adenosine residues in the pA-tail, and β was roughly estimated at 0.096. The latter number implies that ~20 succeeding adenosine residues saturate the interaction ($tanh(20^{*}\ 0.96) = 0.96$). As such, Pab1 recognizes approximately 20 residues of the pA-tail (see Fig. 4A), consistent with previous in vitro footprint experiments (Webster et al, 2018; Baer and Kornberg, 1980; Schäfer et al, 2019). The *γ_rate* parameter should be interpreted as the apparent asymptotic proportion within the pA-tail, where $p = e^{-\gamma\_rate}$ represents the ratio of abundance between two pA-tails differing in length by a single adenosine residue. There is no direct interpretation for the *γ_shape* parameter. However, comparing the gamma distribution to the Erlang one (in which parameter a in Eq. (1) is an integer), $\gamma\_shape + 1$ in Eq. (3) roughly reflects the number of critical events leading to RNA degradation (e.g., removal of individual Pab1, a relatively slow process, and rapid decapping coupled to degradation).

Estimation of γ_shape and γ_rate parameters from experimental distributions

For each experimental pA-tail distribution from the Mex67-depletion chase, the parameters of Eq. (3) (i.e., *γ_shape* and *γ_rate*) were fitted using the Nonlinear Least Squares procedure (nls) implemented in R (version 4.2.3; www.r-project.org). The procedure code in R and the table containing the list of modified gamma parameters are deposited at Mendeley data (https://doi.org/10.17632/2j3hh37zzs.1) and also included in the publication's Appendix file (Appendix Table S3). The graphical output generated as a pdf file is exclusively deposited at Mendeley. Please refer to the instruction in the Appendix section 1.8 for further code modifications necessary to adapt the code to other datasets. The time dependence of the *γ_shape* and *γ_rate* parameters was further analyzed in R using a linear model. In addition, the time evolution of the pA-tail distributions was illustrated as a 3D graph using Origin (version 10.0; www.originlab.com), assuming a linear time dependence of the *γ_shape* and *γ_rate* parameters.

Estimation of newly made mRNA pA-tail profiles

New mRNA pA-tail profiles, depicted as blue lines in Figs. 4C and EV2B, were generated by subtracting the modified gamma distribution from the experimental pA-tail profile and subsequent log2-fitting. This procedure was applied to all Mex67-AA distributions but yielded significant estimates of new mRNA production only in control samples (compare control distributions in Fig. 4C to profiles obtained after Mex67 depletion in EV2B and refer to graphs deposited at Mendeley data: https://doi.org/10.17632/2j3hh37zzs.1). This confirms the previously observed decrease in transcript synthesis determined using 4tU-labeling (Tudek et al, 2018; Schmid et al, 2018).

Modeling the pA-tail degradation kinetics

Let us consider a population of RNAs with specific of pA-tail lengths of i equal to x_i, and assume that the microscopic deadenylation rate, k, is universal, i.e., does not substantially depend on either the length of a pA-tail or on time. The time evolution of the pA-tails' population can therefore be described by a system of N differential equations, with initial conditions $x_i(t = 0) = n(i)$.

$$dx_i/dt = -k \cdot x_i + k \cdot x_{i+1} \qquad (4)$$

In the above Eq. (4), the factor '$-k{\cdot}x_i$' describes the deadenylation of a particular pA-tail of length i. In addition, the deadenylation of the pA-tail of the length $i + 1$ contributes to the apparent recovery of the pA-tail of the length i (factor '$+k{\cdot}x_{i+1}$').

This system of differential equations can be solvable analytically, leading to the Eq. (5):

$$x_i(t) = e^{-\alpha} \cdot \sum_{j=i}^{N} \left(n(j) \cdot \alpha^{(j-i)}/(j-i)!\right) \qquad (5)$$

where the parameter α, equaling $k{\cdot}t$, is the dimensionless scaling factor, and N represents the maximal length of a pA-tail.

Equation (5) enabled the in silico evolution of any initial pA-tail distribution (Figs. EV3D and EV5H). However, according to Eq. (5), the accumulation of short-tail RNAs is expected, which is not observed experimentally due to an accelerated degradation of non-protected RNAs. The latter phenomenon is incorporated into the model as the consequence of the removal of the last Pab1, arbitrarily taken into account following Eq. (2):

$$x_i(t) = (tanh(\beta \cdot i))^{\alpha} \cdot e^{-\alpha} \cdot \sum_{j=i}^{N} \left(n_k \cdot \alpha^{(j-i)}/(j-i)!\right) \qquad (6)$$

The final model demonstrating the evolution of the pA-tail distribution is depicted in Figs. EV3D and EV5H. In Fig. EV3D, the initial distribution is the same as the one in Fig. EV1E replicate A. In Fig. EV5H, the model is tested on a theoretically sharp distribution akin to a short transcription burst. In Fig. EV3D, subsequent distributions resemble properties observed experimentally for replicate A. Comparing a series of uniformly sampled distributions (e.g., for $\alpha = 0, 1, 2, 3, 4, \ldots$) evolved from the initial experimental distribution (at $t = 0/\alpha = 0$) with subsequent experimental ones (Fig. EV3E) facilitates the estimation of the microscopic rate k, denotes by the ratio α_{best}/t, i.e., the slope of a line illustrated in Fig. EV3E, where α_{best} identifies the in silico-derived distribution closest to the experimental one.

Relation between enzymatic microscopic and apparent deadenylation rates

Since the experimental pA-tail distribution can be sufficiently described by the modified gamma distribution (Eq. (3)), for long enough tails (i.e., where $tanh(\beta{\cdot}i) \approx 1$), the ratio of populations of two "subsequent" pA-tails, n_{i+1}/n_i, remains constant, equaling $p = e^{-\gamma\_rate}$. Under these assumptions, Eq. (6) can be simplified to:

$$x_i(t) = e^{-k \cdot t} \cdot \sum_{j=i}^{N} \left(n_i \cdot (p \cdot k \cdot t)^{(j-i)}/(j-i)!\right) \approx n_i \cdot e^{-k \cdot t \cdot (1-p)} \qquad (7)$$

Therefore, the exponential decay of the RNA population carrying a specific pA-tail length is expected with the apparent decay rate, k_{app}, substantially lower than the microscopic rate of enzymatic deadenylation, k:

$$k_{app} = k \cdot (1 - p) = k \cdot (1 - e^{-\gamma rate}) \tag{8}$$

Modeling of decay and quantile deadenylation rates

Normalization of RNA abundance

Initially, sequencing data were normalized to mitigate the impact of unequal library sizes. *ENO2* transcript counts were removed from the raw datasets as in some datasets they were sourced by libraries unrelated to the project (used as artificially synthesized spike-in). The abundance of this transcript often introduced unwanted bias to RNA level quantification. For decay modeling, the transcript levels were normalized across chase datasets. First, absent raw counts (NA values) were substituted with 0.01. Second, count levels were adjusted based on library size. These data were used to calculate decay rates.

Data quality control and filtering for decay and deadenylation modeling

The Appendix file, along with Appendix Figs. S3 and S4 and Appendix Table S4, details the quality control procedures and outlines the strategy for decay and quantile deadenylation. Briefly, data were filtered to calculate decay and quantile deadenylation rates. Initially, all quantile estimates derived from less than 10 counts were discarded due to low estimation quality for quantiles and medians. Next, all values representing mRNA abundance or quantile pA-tail length greater than the reference were removed. This approach is justified for two primary reasons. First, budding yeast do not code for cytoplasmic adenylases, rendering any increase in pA-tail length biologically irrelevant and likely arising for other processes. Under Mex67-depletion conditions, newly synthesized mRNAs predominantly undergo decay prior to Pap1-mediated adenylation. However, a very small fraction of de novo produced mRNAs may escape nuclear decay and acquire a long pA-tail in a process called hyperadenylation (Jensen et al, 2001). This is due to mis-regulation of Pap1 polyA-polymerase activity (Turtola et al, 2021). Although the share of these reads is negligible in short Mex67-depletion times, this phenomenon increases over time as the cytoplasmic mRNA pool gradually degrades. Similarly, during heat stress chase, mRNA levels and pA-tail lengths decrease following transcriptional shut-down, however beyond 10 min, a large pool of newly synthesized transcripts with long pA-tails can be observed following transcriptional resumption. These processes independently lead to a bi-modal distribution of pA-tail lengths, ultimately justifying the exclusion of certain samples from modeling (Figs. 1B, 2A and 6A; see Appendix).

To calculate decay and deadenylation rates under heat stress at 37 °C or thiolutin treatment at 25 °C, only mRNAs previously identified as transcriptionally down-regulated by RNAPII ChIP were selected. To this end, a reference list of such mRNAs was directly sourced from Vinayachandran et al (2018) and is deposited at Mendeley (https://doi.org/10.17632/2j3hh37zzs.1).

Exponential modeling

Exponential modeling for decay rate estimation was previously used by other authors (Miller et al, 2011; Sun et al, 2012; Neymotin et al, 2014; Presnyak et al, 2015; Chan et al, 2018). The decay constant of the standard exponential decay formula was estimated from experimental data as the fraction degraded over a given time (*ln(L_time_x/L_time_x-1)/time*), where *L* represents the pA-tail length or mRNA abundance. Decay and quantile deadenylation constants were independently calculated for each quantile in each chase replicate and subsequently averaged. Details concerning the modeling process are provided in the Appendix, including the rationale for merging replicates. When calculating decay rates for Mex67-AA chase datasets, control times were adjusted by 2.67 min (the time estimated by the modified gamma model required for manifestation of the Mex67-depletion phenotype). Dataset EV1 presents all decay and deadenylation constants for the three chase experiments. Throughout the publication, the "mean_mean" values were used. In addition, the tables used for these calculations can be accessed in Mendeley data (https://doi.org/10.17632/2j3hh37zzs.1).

Data availability

The datasets produced in this study are available in the following databases: Basecalled nanopore sequencing data: Gene Expression Omnibus GSE272785, consult the Dataset EV2 for sample accession numbers); Raw nanopore sequencing data: European Nucleotide Archive (consult the Dataset EV2 for sample accession numbers); Source data underlying each figure panel, R code to calculate *γ_rate* and *γ_shape*, are deposited at Mendeley Data, https://doi.org/10.17632/2j3hh37zzs.1. In addition, the R code is also listed in the Appendix section 5.

The source data of this paper are collected in the following database record: biostudies:S-SCDT-10_1038-S44318-024-00258-3.

Expanded view data, supplementary information, appendices are available for this paper at https://doi.org/10.1038/s44318-024-00258-3.

Peer review information

A peer review file is available at https://doi.org/10.1038/s44318-024-00258-3

References

Adams CC, Gross DS (1991) The yeast heat shock response is induced by conversion of cells to spheroplasts and by potent transcriptional inhibitors. J Bacteriol 173(23):7429–7435. https://doi.org/10.1128/jb.173.23.7429-7435.1991

Aibara S, Gordon JM, Riesterer AS, McLaughlin SH, Stewart M (2017) Structural basis for the dimerization of Nab2 generated by RNA binding provides insight into its contribution to both poly(A) tail length determination and transcript compaction in Saccharomyces cerevisiae. Nucleic Acids Res 45(3):1529–1538

Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47–97. https://doi.org/10.1103/RevModPhys.74.47

Archer SK, Shirokikh NE, Hallwirth CV, Beilharz TH, Preiss T (2015) Probing the closed-loop model of mRNA translation in living cells. RNA Biol 12(3):248–254. https://doi.org/10.1080/15476286.2015.1017242

Baer BW, Kornberg RD (1980) Repeating structure of cytoplasmic poly(A)-ribonucleoprotein. Proc Natl Acad Sci USA 77(4):1890–1892. https://doi.org/10.1073/pnas.77.4.1890

Bahar Halpern K, Caspi I, Lemze D, Levy M, Landen S, Elinav E, Ulitsky I, Itzkovitz S (2015) Nuclear retention of mRNA in mammalian tissues. Cell Rep 13(12):2653–2662. https://doi.org/10.1016/j.celrep.2015.11.036

Bilska A, Kusio-Kobiałka M, Krawczyk PS, Gewartowska O, Tarkowski B, Kobyłecki K, Nowis D, Golab J, Gruchota J, Borsuk E, Dziembowski A, Mroczek S (2020) Immunoglobulin expression and the humoral immune response is regulated by the non-canonical poly(A) polymerase TENT5C. Nat Commun 11(1):2032

Brambilla M, Martani F, Bertacchi S, Vitangeli I, Branduardi P (2019) The *Saccharomyces cerevisiae* poly (A) binding protein (Pab1): master regulator of mRNA metabolism and cell physiology. Yeast 36(1):23–34. https://doi.org/10.1002/yea.3347

Braun JE, Truffault V, Boland A, Huntzinger E, Chang CT, Haas G, Weichenrieder O, Coles M, Izaurralde E (2012) A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5′ exonucleolytic degradation. Nat Struct Mol Biol 19(12):1324–1331. https://doi.org/10.1038/nsmb.2413

Bresson S, Tuck A, Staneva D, Tollervey D (2017) Nuclear RNA decay pathways aid rapid remodeling of gene expression in yeast. Mol Cell 65(5):787–800.e5. https://doi.org/10.1016/j.molcel.2017.01.005

Brouze A, Krawczyk PS, Dziembowski A, Mroczek S (2023) Measuring the tail: methods for poly(A) tail profiling. Wiley Interdiscip Rev RNA 14(1):e1737. https://doi.org/10.1002/wrna.1737

Caponigro G, Parker R (1995) Multiple functions for the poly(A)-binding protein in mRNA decapping and deadenylation in yeast. Genes Dev 9(19):2421–2432. https://doi.org/10.1101/gad.9.19.2421

Chan LY, Mugler CF, Heinrich S, Vallotton P, Weis K (2018) Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. eLife 7:e32536. https://doi.org/10.7554/eLife.32536

Charenton C, Gaudon-Plesse C, Fourati Z, Taverniti V, Back R, Kolesnikova O, Séraphin B, Graille M (2017) A unique surface on Pat1 C-terminal domain directly interacts with Dcp2 decapping enzyme and Xrn1 5′-3′ mRNA exonuclease in yeast. Proc Natl Acad Sci USA 114(45):E9493–E9501. https://doi.org/10.1073/pnas.1711680114

Chen CY, Shyu AB (2010) Mechanisms of deadenylation-dependent decay. Wiley Interdiscip Rev RNA 2(2):167–183. https://doi.org/10.1002/wrna.40

Dargemont C, Kühn LC (1992) Export of mRNA from microinjected nuclei of *Xenopus laevis* oocytes. J Cell Biol 118(1):1–9. https://doi.org/10.1083/jcb.118.1.1

De Magistris P (2021) The great escape: mRNA export through the nuclear pore complex. Int J Mol Sci 22(21):11767. https://doi.org/10.3390/ijms222111767

Decker CJ, Parker R (1993) A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. Genes Dev 7(8):1632–1643. https://doi.org/10.1101/gad.7.8.1632

Dower K, Kuperwasser N, Merrikh H, Rosbash M (2004) A synthetic A tail rescues yeast nuclear accumulation of a ribozyme-terminated transcript. RNA 210(12):1888–1899. https://doi.org/10.1261/rna.7166704

Eisen TJ, Eichhorn SW, Subtelny AO, Lin KS, McGeary SE, Gupta S, Bartel DP (2020) The dynamics of cytoplasmic mRNA metabolism. Mol Cell 77(4):786–799.e10. https://doi.org/10.1016/j.molcel.2019.12.005

Gupta SS, Groll PA (1961) Gamma distribution in acceptance sampling based on life tests. J Am Stat Assoc 56(296):942–970. https://doi.org/10.2307/2282006

Harigaya Y, Parker R (2016) Analysis of the association between codon optimality and mRNA stability in Schizosaccharomyces pombe. BMC Genomics 17(1):895. https://doi.org/10.1186/s12864-016-3237-6

Haruki H, Nishikawa J, Laemmli UK (2008) The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol Cell 31(6):925–932. https://doi.org/10.1016/j.molcel.2008.07.020

Ho B, Baryshnikova A, Brown GW (2018) Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* proteome. Cell Syst 6(2):192–205.e3. https://doi.org/10.1016/j.cels.2017.12.004

Jensen TH, Patricio K, McCarthy T, Rosbash M (2001) A block to mRNA nuclear export in *S. cerevisiae* leads to hyperadenylation of transcripts that accumulate at the site of transcription. Mol Cell 7(4):887–898. https://doi.org/10.1016/s1097-2765(01)00232-5

Kadanoff LP (2009) More is the Same; Phase Transitions and Mean Field Theories. J Stat Phys 137(5-6):777–797 https://doi.org/10.1007/s10955-009-9814-1

Marguerat S, Lawler K, Brazma A, Bähler J (2014) Contributions of transcription and mRNA decay to gene expression dynamics of fission yeast in response to oxidative stress. RNA Biol 11(6):702–714. https://doi.org/10.4161/rna.29196

Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dölken L, Martin DE, Tresch A, Cramer P (2011) Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Syst Biol 7:458. https://doi.org/10.1038/msb.2010.112

Mor A, Suliman S, Ben-Yishay R, Yunger S, Brody Y, Shav-Tal Y (2010) Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. Nat Cell Biol 12(6):543–552. https://doi.org/10.1038/ncb2056

Morawska M, Ulrich HD (2013) An expanded tool kit for the auxin-inducible degron system in budding yeast. Yeast 30(9):341–351. https://doi.org/10.1002/yea.2967

Neymotin B, Athanasiadou R, Gresham D (2014) Determination of in vivo RNA kinetics using RATE-seq. RNA 20(10):1645–1652. https://doi.org/10.1261/rna.045104.114

Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M (2009) An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Methods 6(12):917–922. https://doi.org/10.1038/nmeth.1401

Otero LJ, Ashe MP, Sachs AB (1999) The yeast poly(A)-binding protein Pab1p stimulates in vitro poly(A)-dependent and cap-dependent translation by distinct mechanisms. EMBO J 18(11):3153–3163. https://doi.org/10.1093/emboj/18.11.3153

Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, Coller J (2015) Codon optimality is a major determinant of mRNA stability. Cell 160(6):1111–1124. https://doi.org/10.1016/j.cell.2015.02.029

Radhakrishnan A, Chen YH, Martin S, Alhusaini N, Green R, Coller J (2016) The DEAD-Box protein Dhh1p couples mRNA decay and translation by monitoring codon optimality. Cell 167(1):122–132.e9. https://doi.org/10.1016/j.cell.2016.08.053

Réka A, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47. https://doi.org/10.1103/RevModPhys.74.47

Rodríguez-Molina JB, Turtola M (2022) Birth of a poly(A) tail: mechanisms and control of mRNA polyadenylation. FEBS Open Bio. https://doi.org/10.1002/2211-5463.13528

Sachs AB, Davis RW (1989) The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. Cell 58(5):857–867. https://doi.org/10.1016/0092-8674(89)90938-0

Schäfer IB, Yamashita M, Schuller JM, Schüssler S, Reichelt P, Strauss M, Conti E (2019) Molecular basis for poly(A) RNP architecture and recognition by the Pan2-Pan3 deadenylase. Cell 177(6):1619–1631.e21. https://doi.org/10.1016/j.cell.2019.04.013

Schmid M, Olszewski P, Pelechano V, Gupta I, Steinmetz LM, Jensen TH (2015) The nuclear polyA-binding protein Nab2p is essential for mRNA production. Cell Rep 12(1):128–139. https://doi.org/10.1016/j.celrep.2015.06.008

Schmid M, Tudek A, Jensen TH (2018) Simultaneous measurement of transcriptional and post-transcriptional parameters by 3′ end RNA-Seq. Cell Rep 24(9):2468–2478.e4. https://doi.org/10.1016/j.celrep.2018.07.104

Siwiak M, Zielenkiewicz P (2010) A comprehensive, quantitative, and genome-wide model of translation. PLoS Comput Biol 6(7):e1000865. https://doi.org/10.1371/journal.pcbi.1000865

Sun M, Schwalb B, Schulz D, Pirkl N, Etzold S, Larivière L, Maier KC, Seizl M, Tresch A, Cramer P (2012) Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res 22(7):1350–1359. https://doi.org/10.1101/gr.130161.111

Tarun Jr SZ, Wells SE, Deardorff JA, Sachs AB (1997) Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. Proc Natl Acad Sci USA 94(17):9046–9051. https://doi.org/10.1073/pnas.94.17.9046

Tomecki R, Drazkowska K, Kobylecki K, Tudek A (2023) SKI complex: a multifaceted cytoplasmic RNA exosome cofactor in mRNA metabolism with links to disease, developmental processes, and antiviral responses. Wiley Interdiscip Rev RNA. e1795. https://doi.org/10.1002/wrna.1795

Tudek A, Krawczyk PS, Mroczek S, Tomecki R, Turtola M, Matylla-Kulińska K, Jensen TH, Dziembowski A (2021) Global view on the metabolism of RNA poly(A) tails in yeast *Saccharomyces cerevisiae*. Nat Commun 12(1):4951. https://doi.org/10.1038/s41467-021-25251-w

Tudek A, Schmid M, Makaras M, Barrass JD, Beggs JD, Jensen TH (2018) A nuclear export block triggers the decay of newly synthesized polyadenylated RNA. Cell Rep 24(9):2457–2467.e7. https://doi.org/10.1016/j.celrep.2018.07.103

Turtola M, Manav MC, Kumar A, Tudek A, Mroczek S, Krawczyk PS, Dziembowski A, Schmid M, Passmore LA, Casañal A, Jensen TH (2021) Three-layered control of mRNA poly(A) tail synthesis in *Saccharomyces cerevisiae*. Genes Dev 35(17–18):1290–1303. https://doi.org/10.1101/gad.348634.121

Vinayachandran V, Reja R, Rossi MJ, Park B, Rieber L, Mittal C, Mahony S, Pugh BF (2018) Widespread and precise reprogramming of yeast protein-genome interactions in response to heat shock. Genome Res 28(3):357–366. https://doi.org/10.1101/gr.226761.117

Webster MW, Chen YH, Stowell JAW, Alhusaini N, Sweet T, Graveley BR, Coller J, Passmore LA (2018) mRNA deadenylation is coupled to translation rates by the differential activities of Ccr4-Not nucleases. Mol Cell 70(6):1089–1100.e8. https://doi.org/10.1016/j.molcel.2018.05.033

Wolf J, Valkov E, Allen MD, Meineke B, Gordiyenko Y, McLaughlin SH, Olsen TM, Robinson CV, Bycroft M, Stewart M, Passmore LA (2014) Structural basis for Pan3 binding to Pan2 and its function in mRNA recruitment and deadenylation. EMBO J 33(14):1514–1526. https://doi.org/10.15252/embj.201488373

Yi H, Park J, Ha M, Lim J, Chang H, Kim VN (2018) PABP cooperates with the CCR4-NOT complex to promote mRNA deadenylation and block precocious decay. Mol Cell 70(6):1081–1088.e5. https://doi.org/10.1016/j.molcel.2018.05.009

Acknowledgements

This work was mainly supported by the National Science Centre Poland, the SONATA grant (2020/39/D/NZ2/02174 to AT), the TEAM/2016-1/3 Foundation for Polish Science grant (to AD), and HORIZON Europa ERC AdG (101097317 to AD). Work by RT was supported by National Science Centre Poland, SONATA BIS grant (2017/26/E/NZ1/00724), and by the National Centre for Research and Development (LIDER/35/46/L-3/11/NCBR/2012). SM's work was supported by the National Science Centre Poland grant (2020/38/E/NZ2/00372 to SM). MT was supported by the Academy Research Fellow grant (nos. 349698 and 353682) awarded by the Research Council of Finland. Work by MT in the laboratory of THJ was supported by a Federation of European Biochemical Societies long-term fellowship and an EMBO long-term fellowship (ALTF 328-2019). This research was performed thanks to the IIMCB IN-MOL-CELL Infrastructure funded by the European Union, co-financed under the European Funds for Smart Economy 2021-2027 (FENG) and the European Union— NextGenerationEU.

Author contributions

Agnieszka Czarnocka-Cieciura: Data curation; Validation; Investigation; Visualization; Writing—review and editing. **Jarosław Poznański**: Conceptualization; Data curation; Software; Formal analysis; Visualization; Methodology. **Matti Turtola**: Resources; Funding acquisition; Validation; Investigation. **Rafał Tomecki**: Funding acquisition; Validation; Investigation. **Paweł S Krawczyk**: Data curation; Software; Supervision. **Seweryn Mroczek**: Supervision; Funding acquisition; Methodology. **Wiktoria Orzeł**: Investigation. **Upasana Saha**: Resources. **Torben Heick Jensen**: Supervision. **Andrzej Dziembowski**: Conceptualization; Supervision; Funding acquisition; Investigation; Methodology; Writing—review and editing. **Agnieszka Tudek**: Conceptualization; Resources; Formal analysis; Supervision; Funding acquisition; Validation; Investigation; Visualization; Methodology; Writing—original draft; Writing—review and editing.

Source data underlying figure panels in this paper may have individual authorship assigned. Where available, figure panel/source data authorship is listed in the following database record: biostudies:S-SCDT-10_1038-S44318-024-00258-3.

Disclosure and competing interests statement

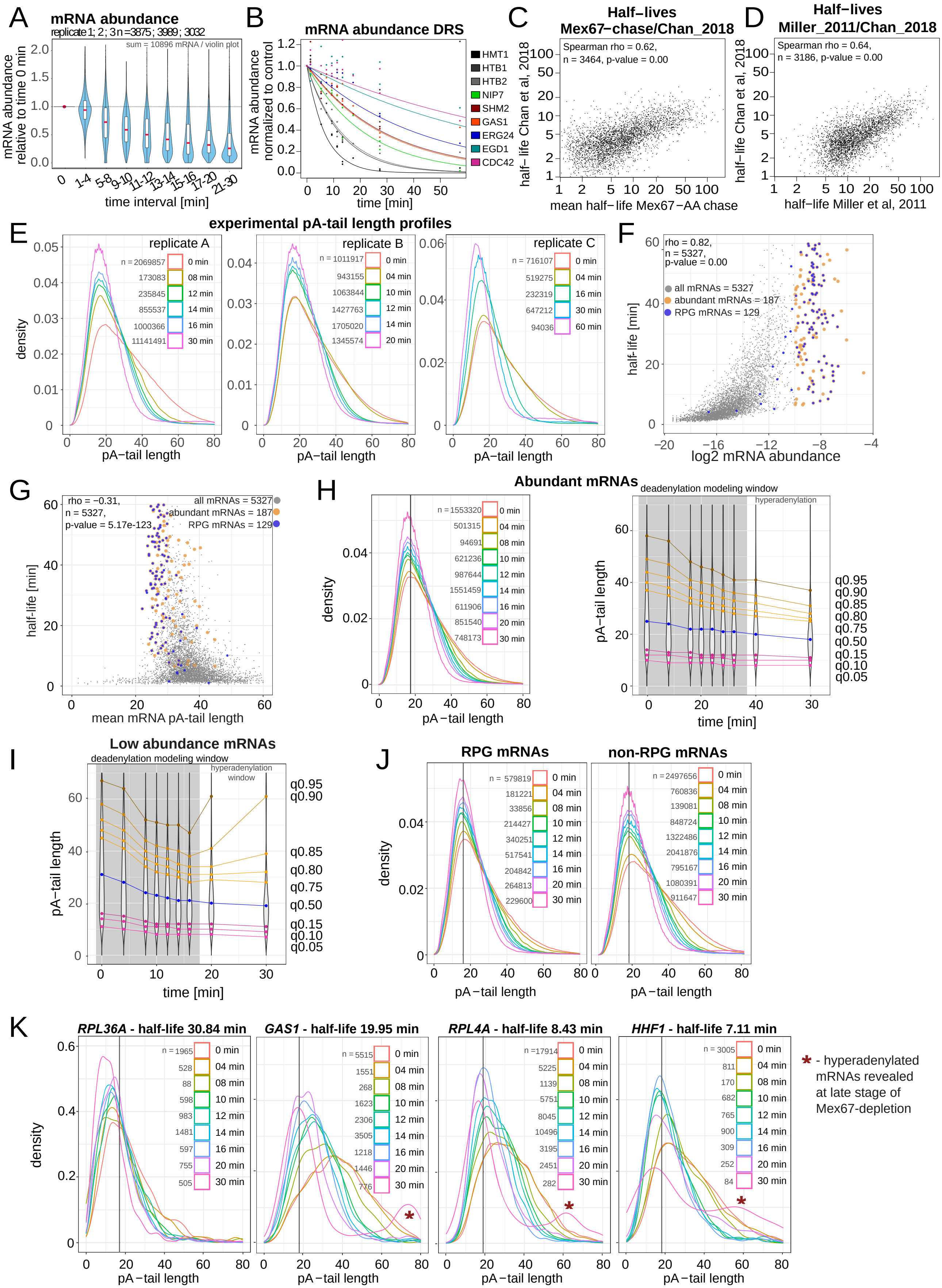
The authors declare no competing interests.

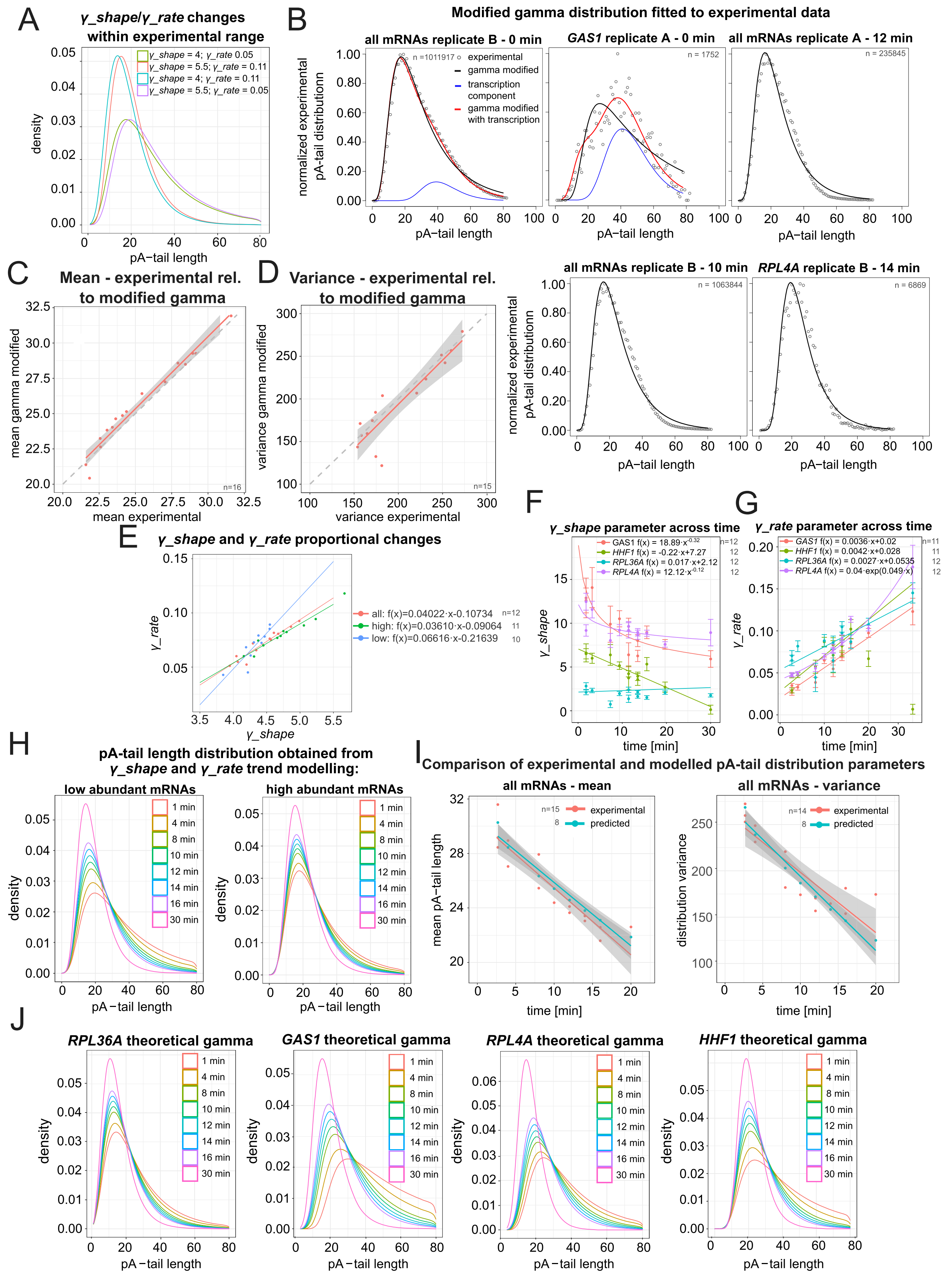
Expanded View Figures

Figure EV1. Following nuclear decay of pre-mRNAs in export-block conditions the cytoplasmic deadenylation and decay of mRNAs is revealed.

▶

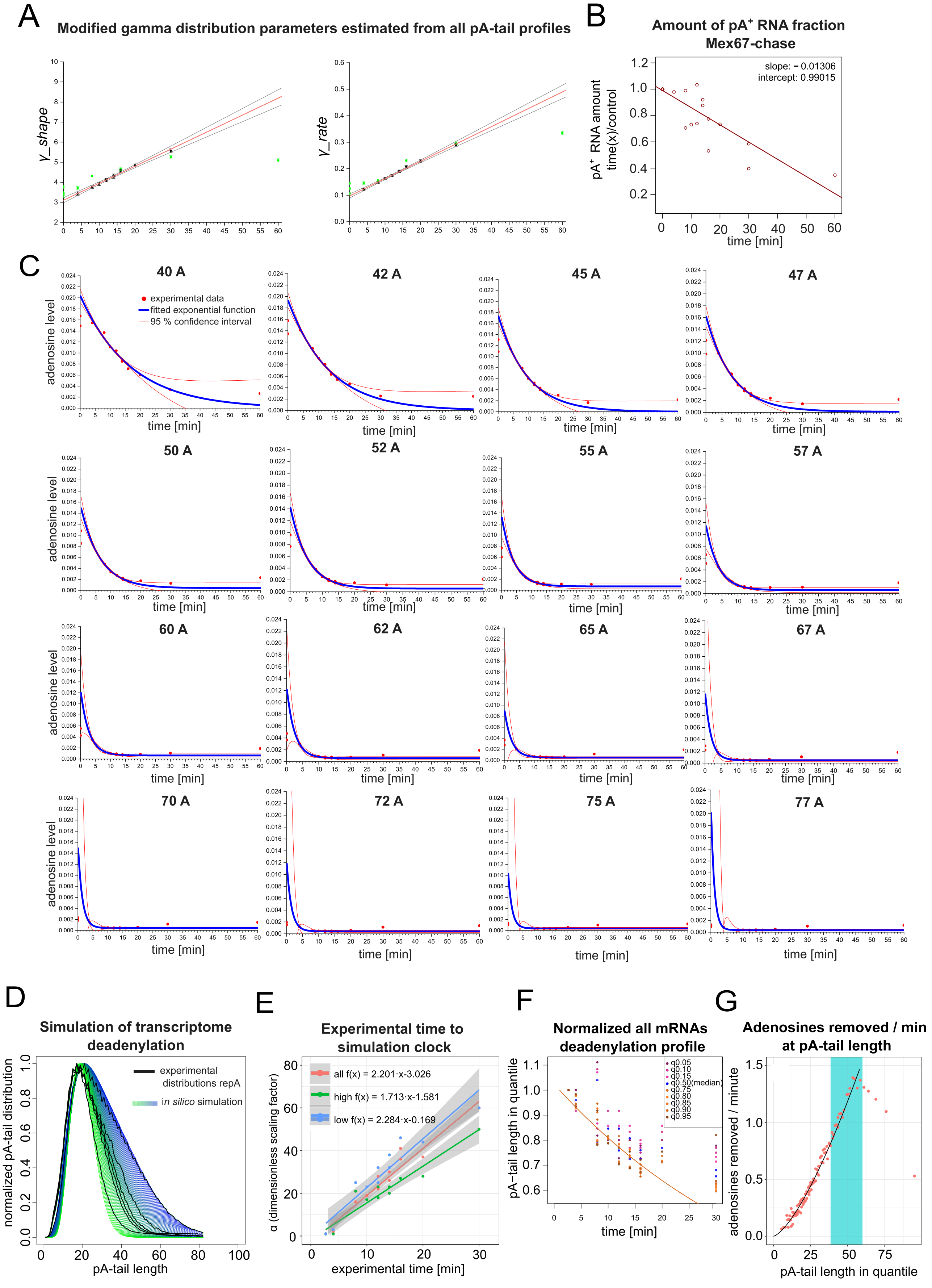
(**A**) Time-dependent change in mRNA abundance relative to control samples in Mex67-depleted cells shown in the form of a series of boxplots. The boxplot central line is the median. The box edges are the 25th and 75th percentiles. Whiskers extend to 1.5 times the IQR (Inter Quartile Range). All three replicates were combined into time point ranges. For each time-slot the number of transcripts assessed in each replicate is given (3875, 3989 and 3032). Sum of 10896 mRNAs is assessed in each time-slot. Note that those are largely overlapping). (**B**) Time-dependent abundance of selected mRNAs in DRS datasets from Mex67-depleted cells. Fitted lines demonstrate the decay factor calculated for each mRNA based on DRS data. (**C**) Scatterplot comparing half-life times in Mex67-depleted DRS datasets to those reported by Chan et al (2018). (**D**) Scatterplot comparing half-life times reported by Miller et al (2011) to those by Chan et al (2018). (**E**) Global distributions of pA-tail lengths of total mRNA in three replicates of Mex67-AA chase experiments used to model deadenylation. The number of transcripts (pA-tail estimates) in each density plot is given on the panel. (**F**) Scatterplot comparing single mRNA log2 abundance to half-life. Abundant mRNAs and RPG mRNAs are highlighted with gold and blue dots. (**G**) Scatterplot comparing mean pA-tail length to mRNA half-life. (**H**) Global distribution of pA-tail lengths of highly abundant mRNAs isolated from control samples and at various times of Mex67 depletion, presented as a density plot (left panel) and violin plot (right panel). Replicates were merged. The number of transcripts (pA-tail estimates) in each density plot is given on the panel. (**I**) Global distribution of pA-tail length of low-abundance mRNAs. Replicates were merged. (**J**) Global pA-tail distribution of RPG and non-RPG mRNAs. Replicates were merged. The number of transcripts (pA-tail estimates) in each density plot is given on the panel. (**K**) Global pA-tail distribution of individual *RPL36A*, *GAS1*, *RPL4A*, and *HHF1* mRNAs. Stars mark long pA-tailed mRNAs accumulating in later Mex67-depletion time points, indicative of hyperadenylation (Jensen et al, 2001). The number of transcripts (pA-tail estimates) in each density plot is given on the panel.


A
mRNA abundance
replicate 1; 2; 3 n = 3875; 3989; 3032
sum = 10896 mRNA / violin plot
mRNA abundance relative to time 0 min
time interval [min]
B
mRNA abundance DRS
mRNA abundance normalized to control
time [min]
HMT1
HTB1
HTB2
NIP7
SHM2
GAS1
ERG24
EGD1
CDC42
C
Half−lives Mex67−chase/Chan_2018
Spearman rho = 0.62, n = 3464, p-value = 0.00
half− life Chan et al, 2018
mean half− life Mex67−AA chase
D
Half−lives Miller_2011/Chan_2018
Spearman rho = 0.64, n = 3186, p-value = 0.00
half−life Chan et al, 2018
half−life Miller et al, 2011
E
experimental pA-tail length profiles
replicate A
replicate B
replicate C
density
pA−tail length
F
rho = 0.82, n = 5327, p-value = 0.00
all mRNAs = 5327
abundant mRNAs = 187
RPG mRNAs = 129
half-life [min]
log2 mRNA abundance
G
rho = −0.31, n = 5327, p-value = 5.17e-123
all mRNAs = 5327
abundant mRNAs = 187
RPG mRNAs = 129
half-life [min]
mean mRNA pA-tail length
H
Abundant mRNAs
deadenylation modeling window
hyperadenylation window
density
pA−tail length
pA-tail length
time [min]
q0.95 q0.90 q0.85 q0.80 q0.75 q0.50 q0.15 q0.10 q0.05
I
Low abundance mRNAs
deadenylation modeling window
hyperadenylation window
pA-tail length
time [min]
J
RPG mRNAs
non-RPG mRNAs
density
pA−tail length
K
RPL36A - half-life 30.84 min
GAS1 - half-life 19.95 min
RPL4A - half-life 8.43 min
HHF1 - half-life 7.11 min
density
pA−tail length
* - hyperadenylated mRNAs revealed at late stage of Mex67-depletion


A
γ_shape/γ_rate changes within experimental range
γ_shape = 4; γ_rate 0.05
γ_shape = 5.5; γ_rate = 0.11
γ_shape = 4; γ_rate = 0.11
γ_shape = 5.5; γ_rate = 0.05
density
pA−tail length
B
Modified gamma distribution fitted to experimental data
all mRNAs replicate B - 0 min
n =1011917
experimental
gamma modified
transcription component
gamma modified with transcription
GAS1 replicate A - 0 min
n = 1752
all mRNAs replicate A - 12 min
n = 235845
all mRNAs replicate B - 10 min
n = 1063844
RPL4A replicate B - 14 min
n = 6869
normalized experimental pA-tail distributionn
pA−tail length
C
Mean - experimental rel. to modified gamma
mean gamma modified
mean experimental
n=16
D
Variance - experimental rel. to modified gamma
variance gamma modified
variance experimental
n=15
E
γ_shape and γ_rate proportional changes
all: f(x)=0.04022·x-0.10734 n=12
high: f(x)=0.03610·x-0.09064 11
low: f(x)=0.06616·x-0.21639 10
γ_rate
γ_shape
F
γ_shape parameter across time
GAS1 f(x) = 18.89·x^-0.32 n=12
HHF1 f(x) = -0.22·x+7.27 12
RPL36A f(x) = 0.017·x+2.12 12
RPL4A f(x) = 12.12·x^-0.12 12
γ_shape
time [min]
G
γ_rate parameter across time
GAS1 f(x) = 0.0036·x+0.02 n=11
HHF1 f(x) = 0.0042·x+0.028 11
RPL36A f(x) = 0.0027·x+0.0535 12
RPL4A f(x) = 0.04·exp(0.049·x) 12
γ_rate
time [min]
H
pA-tail length distribution obtained from γ_shape and γ_rate trend modelling:
low abundant mRNAs
high abundant mRNAs
1 min
4 min
8 min
10 min
12 min
14 min
16 min
30 min
density
pA−tail length
I
Comparison of experimental and modelled pA-tail distribution parameters
all mRNAs - mean
n=15 experimental
8 predicted
mean pA-tail length
time [min]
all mRNAs - variance
n=14 experimental
8 predicted
distribution variance
time [min]
J
RPL36A theoretical gamma
GAS1 theoretical gamma
RPL4A theoretical gamma
HHF1 theoretical gamma
1 min
4 min
8 min
10 min
12 min
14 min
16 min
30 min
density
pA−tail length

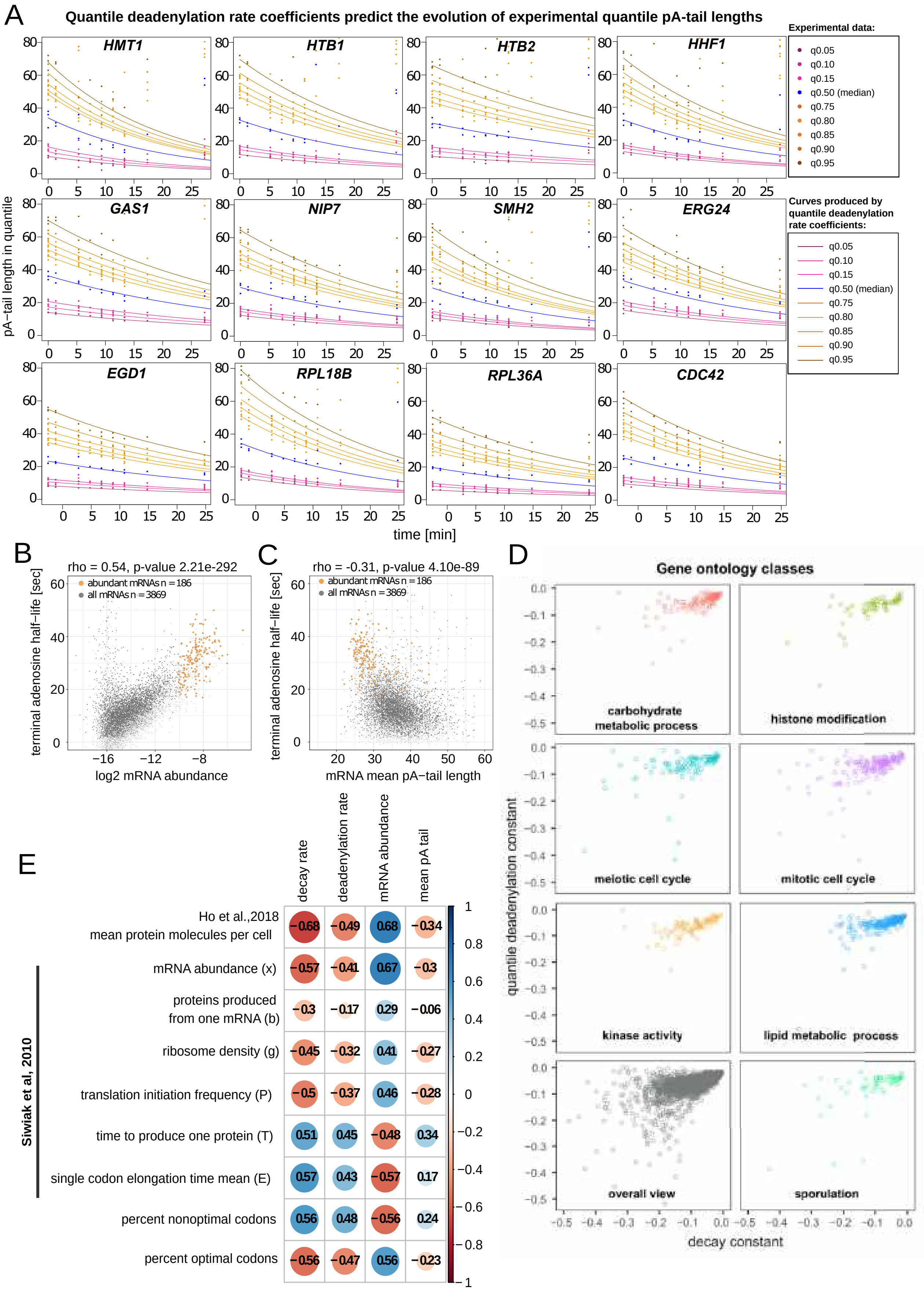
◀ **Figure EV2. A modified gamma distribution accurately describes experimental yeast pA-tail distributions.**

(**A**) Plot displaying modified gamma distributions for *γ_shape* and *γ_rate* parameters within experimental values. The green line represents a function that effectively describes the pA-tail distribution of the entire transcriptome in control cells (*γ_shape* = 4; *γ_rate* = 0.05). (**B**) Fitting of the modified gamma distribution to experimental datasets. Figure complements examples shown in Fig. 4C. (**C**) Comparison of the experimental whole-transcriptome mean pA-tail length to the mean pA-tail length of the fitted modified gamma model. A regression line with a 95% confidence interval was fitted to the data points for comparison with the diagonal. The modified gamma distribution slightly overestimates the mean due to discrepancies in estimating very long pA-tails, as seen in Fig. EV2B for pA-tails of 50 and greater. (**D**) Graph comparing the experimental whole-transcriptome pA-tail length variance length to the variance of the fitted modified gamma model. A regression line with a 95% confidence interval was fitted. (**E**) Graph comparing changes in the *γ_shape* and *γ_rate* parameters of the modified gamma distribution fitted to Mex67-depletion chase time points for all mRNAs and those of high and low abundance. The graph complements plots shown in Fig. 4D, E. (**F**, **G**) Graphs showing time-dependent changes in the value of *γ_shape* (**F**) and *γ_rate* (**G**) parameters of the modified gamma probability distribution fitted into experimental data for selected mRNAs: *GAS1*, *HHF1*, *RPL36A*, and *RPL4A*. The parameters are given as full-colored dots supplemented with vertical standard error bars. Each estimate was derived from distributions shown in Fig. EV1K (refer to the panel for the number of reads). Equations and continuous lines describe functions fitted to modified gamma parameters to predict the evolution of theoretical distributions (Fig. EV2J). (**H**) Density plots showing time-dependent evolution of modeled pA-tail length distributions for mRNAs of low and high abundance. The evolution of the median of those distributions is shown in Fig. 4G. (**I**) Graphs comparing the time-dependent evolution of the modeled modified gamma distribution mean and variance to the experimental one for the whole coding transcriptome. Lines were fitted to both datasets and are displayed with a 95% confidence interval. (**J**) Density plots showing predicted *GAS1*, *HHF1*, *RPL36A*, and *RPL4A* pA-tail length distributions obtained by modeling changes in modified gamma parameters (*γ_shape* and *γ_rate*), as shown in Fig. EV3F,G. These should be compared to experimental distributions shown in Fig. EV1K.


A
Modified gamma distribution parameters estimated from all pA-tail profiles
γ_shape
γ_rate
B
Amount of pA⁺ RNA fraction Mex67-chase
slope: − 0.01306
intercept: 0.99015
pA⁺ RNA amount time(x)/control
time [min]
C
40 A
42 A
45 A
47 A
50 A
52 A
55 A
57 A
60 A
62 A
65 A
67 A
70 A
72 A
75 A
77 A
experimental data
fitted exponential function
95 % confidence interval
adenosine level
time [min]
D
Simulation of transcriptome deadenylation
experimental distributions repA
in silico simulation
normalized pA-tail distribution
pA-tail length
E
Experimental time to simulation clock
all f(x) = 2.201·x-3.026
high f(x) = 1.713·x-1.581
low f(x) = 2.284·x-0.169
α (dimensionless scaling factor)
experimental time [min]
F
Normalized all mRNAs deadenylation profile
q0.05
q0.10
q0.15
q0.50(median)
q0.75
q0.80
q0.85
q0.90
q0.95
pA-tail length in quantile
time [min]
G
Adenosines removed / min at pA-tail length
adenosines removed / minute
pA-tail length in quantile

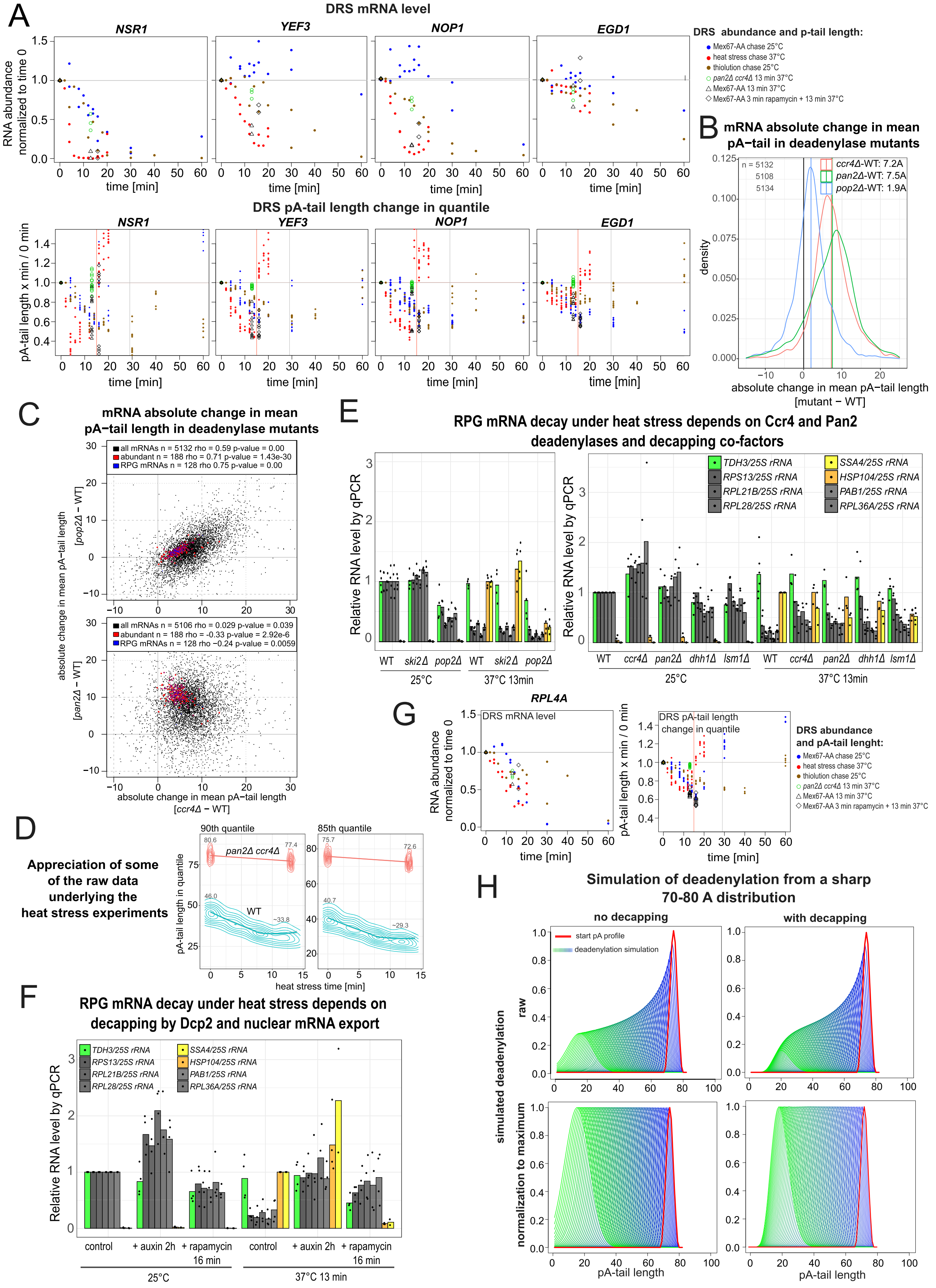
◀ **Figure EV3. The deadenylation process can be reconstituted in silico using the modified gamma distribution.**

(**A**) Graphs illustrating the time-dependent evolution of the *γ_shape* and *γ_rate* parameters of the 3-dimensional modified gamma distributions presented in Fig. 4H. (**B**) Graph depicting time-dependent change in pA^+ RNA fraction recovery from total using beads coated with oligo-$dT_{(25)}$. A linear function was fitted to predict changes in whole-transcriptome pA^+ RNA levels and was used to normalize pA-tail levels for adenosine half-life calculations shown in Fig. EV3C. (**C**) Series of graphs showing absolute change in levels of adenosines at specific pA-tail positions for the whole transcriptome. Half-lives of adenosines at each position were calculated and plotted in Fig. 4I to derive a transcriptomic apparent deadenylation rate. (**D**) Graph comparing experimental pA-tail density plots for the coding transcriptome in Mex67-depletion chase replicate A (black lines) to distributions resulting from consecutive deadenylation simulation cycles (series of blue-green lines). The constant dictating each deadenylation simulation interval was defined by the dimensionless parameter $α$ (see "Methods"). This was carried out to ensure marked differences between consecutive distributions. All distributions were normalized to the distribution peak (distribution maximum). The in silico distribution that best overlapped with the experimental distributions was selected, assigning a specific value of $α$. (**E**) Graph comparing experimental deadenylation times [min] with the ordinal number of the dimensionless $α$ parameter (dimensionless scaling factor) describing the number of in silico deadenylation steps. (**F**) Graph displaying the time-dependent decrease in pA-tail length values in each quantile normalized to control. The dots designating the upper quantiles (orange dots) cluster around a steeper slope than those designating lower quantiles (blue and maroon dots). (**G**) Graph showing the number of adenosines removed per minute as a function of pA-tail length. This graph complements the one shown in Fig. 5B.


A
Quantile deadenylation rate coefficients predict the evolution of experimental quantile pA-tail lengths
HMT1
HTB1
HTB2
HHF1
GAS1
NIP7
SMH2
ERG24
EGD1
RPL18B
RPL36A
CDC42
pA−tail length in quantile
time [min]
Experimental data:
q0.05
q0.10
q0.15
q0.50 (median)
q0.75
q0.80
q0.85
q0.90
q0.95
Curves produced by quantile deadenylation rate coefficients:
q0.05
q0.10
q0.15
q0.50 (median)
q0.75
q0.80
q0.85
q0.90
q0.95
B
rho = 0.54, p-value 2.21e-292
abundant mRNAs n = 186
all mRNAs n = 3869
terminal adenosine half−life [sec]
log2 mRNA abundance
C
rho = -0.31, p-value 4.10e-89
abundant mRNAs n = 186
all mRNAs n = 3869
terminal adenosine half−life [sec]
mRNA mean pA−tail length
D
Gene ontology classes
carbohydrate metabolic process
histone modification
meiotic cell cycle
mitotic cell cycle
kinase activity
lipid metabolic process
overall view
sporulation
quantile deadenylation constant
decay constant
E
decay rate
deadenylation rate
mRNA abundance
mean pA tail
Ho et al.,2018 mean protein molecules per cell: −0.68 −0.49 0.68 −0.34
mRNA abundance (x): −0.57 −0.41 0.67 −0.3
proteins produced from one mRNA (b): −0.3 −0.17 0.29 −0.06
ribosome density (g): −0.45 −0.32 0.41 −0.27
translation initiation frequency (P): −0.5 −0.37 0.46 −0.28
time to produce one protein (T): 0.51 0.45 −0.48 0.34
single codon elongation time mean (E): 0.57 0.43 −0.57 0.17
percent nonoptimal codons: 0.56 0.48 −0.56 0.24
percent optimal codons: −0.56 −0.47 0.56 −0.23
Siwiak et al, 2010
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

◀ **Figure EV4. A simplified method based on changes to distribution quantile values can measure deadenylation rates.**

(**A**) Series of graphs depicting as color-coded dots the changes in pA-tail length in selected quantiles (upper—75-80-85-90-95th, median—50th, and lower—15-10-5th) over time for selected single mRNA examples. The continuous lines represent the quantile deadenylation coefficients calculated from the upper quantiles. (**B**, **C**) Scatterplots comparing terminal adenosine half-life to log2 mRNA abundance (**B**) or mean mRNA pA-tail length (**C**). Spearman's rho correlations and p-values were calculated separately for each set using the rstatix package in R. (**D**) Comparison of decay to deadenylation rates for various gene ontology groups, which are also shown in Fig. 5H. (**E**) Correlation matrix comparing decay, deadenylation rates, mRNA abundance, and mean pA-tail length derived from the Mex67-depletion time course to estimates of protein abundance from Ho et al (2018), various translation rate parameters (Siwiak and Zielenkiewicz (2010)), and percent optimal or non-optimal codons.


A
DRS mRNA level
NSR1
YEF3
NOP1
EGD1
RNA abundance normalized to time 0
time [min]
DRS abundance and p-tail length:
Mex67-AA chase 25°C
heat stress chase 37°C
thiolution chase 25°C
pan2Δ ccr4Δ 13 min 37°C
Mex67-AA 13 min 37°C
Mex67-AA 3 min rapamycin + 13 min 37°C
DRS pA-tail length change in quantile
pA-tail length x min / 0 min
B
mRNA absolute change in mean pA−tail in deadenylase mutants
n = 5132
5108
5134
ccr4Δ-WT: 7.2A
pan2Δ-WT: 7.5A
pop2Δ-WT: 1.9A
density
absolute change in mean pA−tail length [mutant − WT]
C
mRNA absolute change in mean pA−tail length in deadenylase mutants
all mRNAs n = 5132 rho = 0.59 p-value = 0.00
abundant n = 188 rho = 0.71 p-value = 1.43e-30
RPG mRNAs n = 128 rho 0.75 p-value = 0.00
all mRNAs n = 5106 rho = 0.029 p-value = 0.039
abundant n = 188 rho = -0.33 p-value = 2.92e-6
RPG mRNAs n = 128 rho −0.24 p-value = 0.0059
absolute change in mean pA−tail length [pop2Δ − WT]
[pan2Δ − WT]
absolute change in mean pA−tail length [ccr4Δ − WT]
D
Appreciation of some of the raw data underlying the heat stress experiments
90th quantile
85th quantile
pA-tail length in quantile
pan2Δ ccr4Δ
WT
80.6
77.4
46.0
~33.8
75.7
72.6
40.7
~29.3
heat stress time [min]
E
RPG mRNA decay under heat stress depends on Ccr4 and Pan2 deadenylases and decapping co-factors
Relative RNA level by qPCR
WT
ski2Δ
pop2Δ
25°C
37°C 13min
TDH3/25S rRNA
RPS13/25S rRNA
RPL21B/25S rRNA
RPL28/25S rRNA
SSA4/25S rRNA
HSP104/25S rRNA
PAB1/25S rRNA
RPL36A/25S rRNA
ccr4Δ
pan2Δ
dhh1Δ
lsm1Δ
F
RPG mRNA decay under heat stress depends on decapping by Dcp2 and nuclear mRNA export
control
+ auxin 2h
+ rapamycin 16 min
25°C
37°C 13 min
G
RPL4A
DRS mRNA level
DRS pA-tail length change in quantile
RNA abundance normalized to time 0
pA-tail length x min / 0 min
time [min]
DRS abundance and pA-tail lenght:
Mex67-AA chase 25°C
heat stress chase 37°C
thiolution chase 25°C
pan2Δ ccr4Δ 13 min 37°C
Mex67-AA 13 min 37°C
Mex67-AA 3 min rapamycin + 13 min 37°C
H
Simulation of deadenylation from a sharp 70-80 A distribution
no decapping
with decapping
start pA profile
deadenylation simulation
simulated deadenylation
raw
normalization to maximum
pA-tail length

◀ **Figure EV5. RPG and non-RPG mRNAs have altered deadenylation and decay rates during response to heat stress.**

(**A**) Series of graphs showing, in rows, the time-dependent and normalized to control changes in: (top) mRNA abundance by DRS, and (bottom) pA-tail length in upper quantiles (75–95th) for single RPG mRNAs: *NSR1*, *YEF 3*, *NOP3*, and *EGD1* in Mex67 depletion, heat stress 37 °C, and thiolutin 25 °C chase sequencing data. (**B**) Mean pA-tail lengths of individual mRNAs in wild-type cells were subtracted from the mean lengths observed in deadenylase mutants (*ccr4Δ* or *pan2Δ*, or *pop2Δ*) and displayed in the form of a density plot. The mean change in each strain is marked as a vertical line and specified in the figure legend. (**C**) Scatterplots showing the absolute change in mean pA-tail length in *ccr4Δ* - WT on the x-axis compared to *pop2Δ* – WT (top graph) or *pan2Δ* – WT (bottom graph) on the y-axis. Transcripts of high abundance or RPG mRNAs are highlighted in blue and red, respectively. Spearman's rho coefficients and the number of mRNAs compared (n) are listed in each panel legend. (**D**) Plots displaying the raw pA-tail length density distribution of two example upper quantiles in the wild-type strain heat stress chase compared to the control and 13 min heat shock of a double *pan2Δ ccr4Δ* mutant strain. The plots present raw data also shown in Fig. 6E. The fitted lines are for orientation purposes only and connect the local maxima either with local or linear regression. (**E**) Abundance of selected RPG mRNAs normalized to *25S rRNA* in control or *ski2Δ*, *pop2Δ*, *ccr4Δ*, *pan2Δ*, *lsm1Δ*, and *dhh1Δ* cells at 25 °C compared to 13 min heat shock at 37 °C determined using reverse transcription coupled to qPCR. Single dots show biological replicate values used to calculate the mean. (**F**) Abundance of selected RPG mRNAs normalized to *25S rRNA* in a double Dcp2-AID and Mex67-AA strain under steady-state and after 13 min heat stress determined using reverse transcription coupled to qPCR. Prior to heat stress the strains were either treated with auxin for 2 h to deplete Dcp2 or with rapamycin for 3 min to deplete Mex67 prior (25 °C). Heat stress was conducted at 37 °C for 13 min. Single dots represent biological replicates. (**G**) Graphs showing for *RPL4A* (top) mRNA abundance by DRS, and (bottom) change in pA-tail length in upper quantiles (75–95th) normalized to control for Mex67 depletion, heat stress 37 °C, and thiolutin 25 °C chase sequencing data, along with point heat stress of *ccr4Δ pan2Δ* double mutant and cells Mex67-depleted 3 min prior to heat stress. (**H**) A theoretical pA-tail distribution of median 75As and low variance (red line) subjected to deadenylation simulation (series of blue and green lines). Artificial deadenylation was performed with and without inducing decapping, displayed with or without normalization to each distribution maximum value. Normalization to the maximum is shown to better display the systematic change in distribution variance and location of the maximum.

APPENDIX

***

Modeling of mRNA deadenylation rates reveal a complex relationship between mRNA deadenylation and decay

Agnieszka Czarnocka-Cieciura[1,#], Jarosław Poznański[2,#], Matti Turtola[3], Rafał Tomecki[2,4], Paweł S. Krawczyk[1], Seweryn Mroczek[4,1], Wiktoria Orzeł[1], Upasana Saha[5], Torben Heick Jensen[5], Andrzej Dziembowski[1,4]*, Agnieszka Tudek[2,]*

[1] International Institute of Molecular and Cell Biology, Księcia Trojdena 4, 02-109 Warsaw, Poland
[2] Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Adolfa Pawińskiego 5A, 02-106 Warsaw, Poland
[3] Department of Life Technologies, University of Turku, Biocity, Tykistökatu 6, 205240 Turku, Finland
[4] University of Warsaw, Faculty of Biology, Miecznikowa 1, 02-089 Warsaw, Poland
[5] Aarhus University, Department of Molecular Biology and Genetics – Universitetsbyen 81, 8000 Aarhus, Denmark

[#] those authors contributed equally

* correspondence should be addressed to: atudek@ibb.waw.pl or adziembowski@iimcb.gov.pl

***

TABLE OF CONTENT

1. APPENDIX METHODS

1.1. Western blot analysis

4-6 OD units of cells were resuspended in 8M urea and incubated at 80 °C. Cells were subsequently mixed with glass beads and disrupted on a vortex for 5 min. The urea cell extract was centrifuged and the supernatant was collected. The protein concentration was measured, and after adjusting to equal concentrations, extracts were resuspended in Laemmli buffer. 30 µg of protein was loaded on a 6 or 10 % denaturing SDS-PAGE gel. After gel migration, proteins were transferred to Amersham nitrocellulose Protran Western blotting membranes (GE10600001) using semi-dry transfer (semi-dry blot apparatus, Bio-Rad, cat. no.1703940). Blots were blocked in TBS-T and 5 % milk and probed with relevant antibodies overnight. Antibodies used were: anti-FLAG (F1804, Sigma), anti-PGK1 (discontinued Novex Life Technology), anti-Pab1 (Santa Cruz Cat. no.: sc57953) or anti-Rpb3 (Abcam 1Y26 cat. no.: ab81859). After washing in TBS-T blots were probed with appropriate secondary antibodies conjugated with HRP in TBS-T 5 % milk for 1 hour, washed, and developed with ECL substrate (Clarity Western ECL, BioRad).

1.2. Recombinant Xrn1 expression, purification and testing

Nucleotide sequence encoding *Thermothelomyces* (*Myceliophthora*) *thermophilus* Xrn1, with codons optimized for heterologous expression in *Escherichia coli*, was synthesized by GenScript, provided as a derivative of pUC57 plasmid, and re-cloned into NcoI/XhoI sites of the pET-28b(+) vector (Novagen) using sequence and ligation-independent cloning (SLIC) with primer pair TtXrn1For (5’-ttttgtttaactttaagaaggagatataccATGGGCGTCCCGAAGTTTTTCC-3’) - TtXrn1Rev (5’- atctcagtggtggtggtggtggtgctcgagGCTCTGCAGTGCTGCGGTCTG-3’) for insert amplification in PCR.

Resulting pET-28-TtXrn1-6xHis recombinant vector was introduced into *E. coli* BL21-CodonPlus(DE3)-RIL chemo-competent cells (Agilent; *E. coli* B F– ompT hsdS[rB– mB–] dcm+ Tetr gal λ[DE3] endA Hte [argU ileY leuW Camr]) by heat shock-based transformation. Transformants were selected in a standard solid Luria-Broth (LB) medium supplemented with 50 µg/mL kanamycin, and then used to inoculate 50-100 mL of liquid LB containing 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. Following overnight incubation at 37 °C with shaking (120 rpm), 30 mL of the starter culture was utilized to inoculate 1 L of Auto Induction Medium (AIM) Super Broth Base including Trace elements (AIMSB02, Formedium) containing 2% glycerol and both antibiotics. Bacteria were grown for 72 h at 18 °C in an orbital shaker (150 rpm) and eventually collected by centrifugation at 5000 rpm in a Sorvall H6000A/HBB6 swinging-bucket rotor for 15 min at 4 °C.

Bacterial pellet was re-suspended in 100 mL of the lysis buffer (20 mM Tris-HCl pH 7.5, 200 mM NaCl, 10 mM imidazole, 10 mM 2-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.02 µM pepstatinA, 0.02 µg/ml chymostatin, 0.006 µM leupeptin, 20 µM benzamidine hydrochloride), incubated with lysozyme (50 µg/mL; Roth) for 30 min in a cold cabinet with head-over-tail rotation, and then broken in the EmulsiFlex-C3 High Pressure homogenizer at 1500 Bar. Homogenate was centrifuged in a Sorvall WX ULTRA SERIES ultracentrifuge (F37L rotor) at 33000 rpm for 45 min at 4°C.

The extract (supernatant after high-speed ultracentrifugation) was used for protein purification using the ÄKTA Xpress system (GE Healthcare), employing nickel affinity chromatography on the 5 mL column compatible with ÄKTA, which was manually filled with Ni-NTA Superflow resin (Qiagen). The column was equilibrated with 25 mL of low-salt (LS) buffer (20 mM Tris-HCl pH 7.4, 200 mM NaCl, 10 mM imidazole, 10 mM 2-mercaptoethanol) prior to extract loading. After protein binding, the resin was sequentially washed with 40 mL of LS buffer, 25 ml of high-salt (HS) buffer (20 mM Tris-HCl pH 7.4, 1 M NaCl, 10 mM imidazole, 10 mM 2-mercaptoethanol), and again with 20 mL of LS buffer. Bound proteins were recovered by elution with 30 mL of buffer E (50 mM Tris-HCl pH 7.4, 200 mM NaCl, 300 mM imidazole, 10 mM 2-mercaptoethanol). Further protein purification was achieved by performing size-exclusion separation of 5 mL of the eluate from the affinity chromatography step on a Hiload 16/60 Superdex S200 column (GE Healthcare), with the use of 1.2 column volumes of gel-filtration (GF) buffer (Tris-HCl pH 7.4; 150 mM NaCl), followed by ion-exchange chromatography on a Resource Q 1 mL column (GE Healthcare), using linear gradient of NaCl (150 mM-1M) in Tris-HCl pH 7.4, 1 mM DTT.

Two fractions corresponding to the maximum of A280nm absorbance were collected after ion-exchange separation, pooled together, mixed with glycerol (30% v/v), aliquoted, snap-frozen in liquid nitrogen and stored at -80 °C. Purified *Tt*Xrn1 was inspected in 10% SDS-PAGE stained with Coomassie Brilliant Blue R-250, along with commercial yeast Xrn1 (NEB; M0338) as a positive control (data not shown). PageRuler Prestained Protein Ladder, 10 to 180 kDa (ThermoScientific), was used as a molecular weight marker during electrophoresis.

Enzyme specificity towards 5'-monophosphorylated termini was tested by analyzing efficiency of degradation of synthetic 30-mer oligoribonucleotide substrates (5'-ACUCACUCACUCACCAAAAAAAAAAAAACC-3') bearing fluorescein amidite (FAM) at the 3'-end, and either 5'-monophosphate (5'P-RNA30-FAM-3'), 5'-hydroxyl (5'HO-RNA30-FAM-3') or 5'-Gppp (5'Gppp-RNA30-FAM-3') in 1x NEBuffer 3 (NEB; B7003), in 20% denaturing polyacrylamide gels, followed by fluorescence scanning in Typhoon™ FLA 9500 biomolecular imager. Furthermore, the ability to eliminate 18S and 25/28S ribosomal RNAs from total yeast/human RNA samples was examined by treating 1 µg of respective RNA with TtXrn1 for 30 and 60 minutes and running degradation products in 1% agarose gel in 1x TBE containing ethidium bromide. In all biochemical analyses, parallel reactions were carried out using equivalent amounts of commercial Xrn1 as positive control.

1.3. Xrn1 *in vitro* digestion of the pA$^+$ fraction

To remove uncapped mRNAs the 35 ng of pA$^+$ fraction was digested with home-made *Thermothelomyces (Myceliophthora) thermophilus* Xrn1 *in vitro* at 37 °C in Neb3 buffer (B7003S) for 1 hour in a total volume of 20 µl. The sample was then inactivated 10 min at 80 °C, and the RNA was extracted and precipitated as described above.

1.4. Reverse transcription and quantitative PCR analysis

cDNA synthesis was done using SuperScript IV reverse transcriptase from LifeTechnologies (18090050). The reaction mix contained random hexamers (final concentration 2.5 ng/µl), oligo-dT$_{(18)}$ (final concentration 5 µM) as primers and 0.5 mM dNTPs. 1 µg of total RNA or up to 100 ng of pA$^+$ fraction was used as a template and denatured together with oligo-dT and random hexamers for 5 min at 65 °C. After cooling the sample on ice for 2 min, the reaction mix was complemented with buffer, enzyme and RiboLock RNase inhibitor (LifeTechnology; EO 0382) and incubated for 20 min at room temperature and next for 40 min at 50 °C. The enzyme was inactivated for 10 min at 80 °C. cDNA was diluted to 400 µl. 2 µl of cDNA was used in a total of 5 µl reaction mix for qPCR analysis with commercial SYBR reaction mix [Platinum SYBR™ Green qPCR SuperMix-UDG from LifeTechnology (11733046) or RT qPCR Mix SYBR from A&A BIOTECHNOLOGY (2008-1000)] on a LightCycler LC480 Roche apparatus. Data were extracted using the 2nd derivative max method. Oligonucleotide sequences are indicated in Appendix Table S2.

1.5. RNaseH digestion of *RPS5* 3'end

20 µg of total RNA was mixed with 2 µM RNaseH-targeting oligonucleotide (5'-GGCCAAAGTTTCAGCAATGGTC-3') in annealing buffer (50 mM Tris-HCl, 50 mM KCl, pH 8.3) in a total volume of 12 µl and incubated for 2 min at 85°C to be then slowly cooled to 37°C. Where indicated, 2 µM of oligo-dT$_{(18)}$ oligonucleotide (5'-TTTTTTTTTTTTTTTTTT-3') was also included in order to target pA-tail trimming by RNaseH. Next, 8 µl of a mix preheated to 37°C and containing 2.5 U of RNase H (NEB #M0523; New England Biolabs), 2.5× RNase H reaction buffer (1× buffer: 50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl2, 10 mM DTT at pH 8.3), 25 mM DTT, and 4 U RiboLock RNase inhibitor (Thermo Scientific) was added and incubated for 30 min at 37°C. Subsequently, RNA was precipitated by the addition of 100 µl absolute ethanol and 20 µl solution containing 600 mM sodium acetate (pH 5.3), 10 mM EDTA, 5 µg tRNA, and 5 µg glycogen and incubated at −20°C. Sample pellets were washed with 70 % ethanol and resuspended in RNA loading buffer (formamide with 10mM Tris-HCl pH 8.0, 5 mM EDTA, 0.02 % xylene cyanol).

1.6. Northern blotting

RNA samples were separated on 6% urea-polyacrylamide gels by electrophoresis, transferred to Hybond-N+ membrane (GE Healthcare), and hybridized over-night at 50 ˚C in ULTRA-Hyb Oligo Hybridization buffer (Invitrogen AM8663) with 1 pmol of 5' terminally 32P-labelled DNA oligonucleotide probe for RPS5 (5'-CTTAACGGTTAGACTTGGCAACACGTTCCAATTCATCCTTCTTCTTGATAGCGTAAGAAG-3'). The membrane was washed four times with 2XSSC buffer (300 mM NaCl, 30 mM trisodium citrate, pH 7.0) containing 0.5% SDS, each time rotating for 30 min at 42 ˚C and exposed to phosphorimager screen. The upper part of the membrane was subsequently cut out, hybridized with a labeled DNA probe for SCR1 (5'-GGTCACCTTTGCTGACGCTGG-3'), and washed and exposed as above. Phosphorimager scans were processed and quantitated with ImageJ software. The indicated pA tail lengths were approximated from DNA size markers.

1.7. Bulk pA-tail length analysis

500 ng of oligo(dT)-selected poly(A)+ RNA was subjected to 3'-end labeling in 20 µl of reaction mixture containing: 2 µl of 10x Reaction Buffer for T4 RNA Ligase (Thermo Scientific), 2 µl of 10 mM ATP, 2 µl of 100 mM DTT (Invitrogen), 2 µl of 100% DMSO (Thermo Scientific), 0.5 µl of RiboLock RNase Inhibitor (40 U/µl; Thermo Scientific), 2 µl of T4 RNA Ligase 1 (EL0021; Thermo Scientific) and 1.5 µl of [5'-32P]pCp (3000Ci/mmol, 10mCi/ml; Hartmann Analytic), for 1 hour at 37 °C. Enzyme was heat-inactivated at 95 °C for 5 min. Labeled RNA was digested with 0.5 µg of RNase A and 10 U of RNase T1 (both from Thermo Scientific) for 20 min. at 37 °C. Reaction was stopped by addition of an equal amount of formamide loading dye (90% formamide in 1xTBE, 0.03% xylene cyanol, 0.03% bromophenol blue) and incubation at 95 °C for 5 min, followed by flash freezing in liquid nitrogen. 5 µl of the sample was run in denaturing 12% sequencing polyacrylamide/8 M urea/1xTBE gel. 3'-end labeled 20 nt-, 31 nt- and 51 nt-long synthetic RNA oligonucleotides (CCCCACCACCAUCACUUA$_3$, CCCCACCACCAUCACUUA$_{14}$, CCCCACCACCAUCACUUA$_{34}$; all from Future Synthesis) were run in parallel to the labeled pA-tails as molecular size markers. Following electrophoresis, the gel was transferred onto Whatman 3MM filter paper, dried at 80 °C for 2 hours under vacuum, and exposed overnight to a PhosphorImager screen (FujiFilm), which was scanned using a FLA 9000 scanner.

1.8. Consideration regarding estimation of the modified gamma parameters and specific guidelines for adaptation of the R code for other datasets.

The classical gamma distribution properties enable the estimation of the mean, variance, and mode from the gamma parameters. Likewise, those parameters can be devised from the known experimental mean and variance. The modified gamma distribution does not inherit those properties, and the estimation of its parameters is best performed using the Nonlinear Least Squares procedure, which we implemented in R (Appendix section 5 and Mendeley data: doi: 10.17632/2j3hh37zzs.1). The procedure is compatible with the Mex67-depletion dataset only (results listed in Appendix Table S3 and Mendeley data) and its further use requires adaptation to the particularities of novel datasets. The R code assumes that the left arm of the distribution (representing the shortest pA-tails) is fixed and dictated by the $tanh(\beta * N_A)$ function (where N_A represents the pA-tail lengths from 1 to X and β is a fixed value). In the case of the Mex67-AA dataset $\beta = 0.096$ and the course of the function is presented in Figure 4A. The β parameter can be adjusted manually to best represent the actual experimental datasets but should remain equal for the entire data collection. While the modified gamma parameters are not linked to the distribution mean and variance, one has to bear in mind that the *γ_rate* is nearly identical to the slope factor of the pA-tail distribution represented in the log-linear scale (Figure 4J). Therefore, having calculated *γ_rate* and after having manually adjusted β to the dataset collection, the *γ_shape* parameter can also be fitted incrementally, without the need to use the R code.

The attached R code also performs simulation of deadenylation with the devised α scaling factor, as described in the Methods section.

The general limitation of the modified and classical gamma distributions is that they cannot represent bi-modal distributions. However, such experimental distributions for budding yeast datasets are scarce, mainly due to the lack of cytoplasmic adenylases, but are quite common in other eukaryotes. Therefore, we recommend that the modified gamma modeling be used on budding yeast datasets exclusively.

1.9. Bioinformatic analyses

Data analysis in RStudio. Unless stated otherwise data analysis was performed in RStudio (version 1.2.5033 for Mac, 4.2.3 for Windows or 2024.042 build 764 for Windows; RStudio Team, 2019; Posit team, 2024 respectively). Graphs were prepared either using the standard RStudio build-in commands or using ggplot2 package.

Computation of correlation coefficients using Rstudio. Correlation coefficients presented in this study were computed using rstatix package (0.7.2) in Rstudio version 2024.042 build 764. The function used was cor_test performed using Spearman's correlation test with a 0.95 confidence interval and a two-sided correlation significance test.

Binning transcripts by pA-tail length. For each set of samples, transcripts were divided into non-overlapping bins based on their median pA-tail lengths (binwidth=5). Each bin contained transcripts represented by over 30 reads. Then, for each condition and poly(A) length bin, the number of transcripts in each bin was plotted (only transcripts detected in each dataset were included).

Binning reads by pA-tail length. For each set of samples, reads were divided into bins of multiple of 5 pA-tails length. Then for each condition and poly(A) length bin read abundances (normalized to sequencing depth) were plotted.

Gene ontology analysis. Gene ontology analysis was performed using the BioMart R package based on goslim terms. Only transcripts represented by 10 reads in each biological repeat of the chase experiment on Mex67-AA were included in the GO-term analysis. As for all other analyses the 'mean_mean' decay and deadenylation coefficients were used.

1.10. Description of dataset quality control and considerations regarding the estimation of decay and quantile deadenylation coefficients

General quality control metrics. As specified in the Methods section, the Nanopolish 0.13.2 polya function was used to estimate the quality of pA-tail length estimation, and only the reads flagged as PASS were taken for further analyses. The Expanded View Dataset 2 lists the DRS sequencing runs along with the total library read count performed for this study. The Appendix Table S4 lists the PASS counts number for libraries other than the chase experiments, whereas the bar plots in Appendix Figure S3A show the read number and quality distribution in each chase data point. Prior to further analyses, DRS replicates were compared for reproducibility of individual transcript count and mean pA-tail length by producing a Spearman rho correlation coefficient for all reads and mRNAs transcriptionally repressed during heat stress, including RPG mRNAs (Appendix Table S4). Such in-depth analysis for the chase DRS datasets is specified in the subsequent section. Overall, the mRNA abundance was reproduced with a Spearman rho factor within the range of 0.8 to 0.99. The reproducibility of mean pA-tail length for all mRNA ranged from 0.25 to 0.6, which, as expected, strongly depended on the sequencing depth. Importantly, for highly abundant RPG mRNAs, which were the main functional group of mRNAs analyzed in our study, the mean pA-tail length was replicated with a Spearman rho coefficient oscillating between 0.73 and 0.99. Those parameters indicated that the obtained data was of good quality, especially for RPG mRNAs, and could be used to model mRNA decay and deadenylation dynamics.

Chase datasets quality control and modeling strategy. Expanded View Dataset 2 lists the number of reads in each chase library, regardless of quality, while Appendix Figure S3A divides those counts into quality classes for each chase dataset sample.

The high reproducibility between the dataset biological replicates was crucial for the estimation of deadenylation and decay rates. Before commencing modeling the data was visually inspected in order to assess the reproducibility of the time-dependent phenotypes associated with each mutant (Mex67-AA) or growth condition (wild-type under heat stress at 37°C or thiolutin at 25°C). For each chase sample Appendix Figure S3B shows a scatterplot comparing the log2 mRNA abundance to mean pA-tail length. As previously observed (Tudek et al, 2021), in all control samples the mean pA tail length was inversely correlated with mRNA abundance. Specifically, mRNAs expressed at high levels, primarily the ribosomal protein-encoding mRNAs (RPG mRNAs) indicated by red dots, exhibited the shortest mean pA-tails.

During the Mex67-AA depletion at 25°C (see Appendix Figure S3B), we observed a gradual reduction in the mRNA mean pA-tail lengths and a decrease in abundance of most transcripts. As stated in the main text (in reference to Figures 1B, 2A, and EV1K), the terminal time-points of each chase experiment were excluded from subsequent deadenylation modeling. It is because for those samples, we observed an increase in mean pA-tail length, which was likely an effect of hyperadenylation occurring on a few nuclear mRNAs in the Mex67-depleted cells (Jensen et al., 2001). These hyperadenylated species are scarce compared to the cytoplasmic mRNA fraction. However, with significant loss of cytoplasmic mRNAs, they become dominant over time and thus strongly affect the estimation of the mean pA-tail length.

Heat stress at 37°C induced transcriptomic changes distinct from those observed during the Mex67-depletion chase at 25°C (refer to Appendix Figure S3B). Most mRNAs did not exhibit significant changes in abundance or mean pA-tail length. However, the group of transcripts identified by Vinayachandran et al (2018) as being transcriptionally silenced during the acute stress response underwent substantial changes in both abundance and adenylation status (blue dots for non-RPG and red dots for RPG mRNAs). These changes were indicative of deadenylation and decay. Despite this, a notable recovery in mean pA-tail length and mRNA abundance was observed at the final time points of each chase experiment (18-20 minutes), likely due to the resumption of transcription in this group of mRNAs. Consequently, these time points were excluded from deadenylation modeling, as marked in Figure 6A.

Cell treatment with thiolutin at 25°C qualitatively resembled the events occurring during heat stress response (Appendix Figure S3B). Specifically, there was a gradual decrease in both the abundance and mean pA-tail length of mRNAs transcriptionally silenced during heat stress, including RPG mRNAs (Vinayachandran et al, 2018). In contrast, the abundance and pA-tail length of other transcripts remained unchanged. Given this, we conclude that thiolutin at the low concentration used in our experiments, triggers a transcriptional response akin to reaction to heat stress, as previously suggested by Adams and Gross (1991). Notably, no recovery in pA-tail length or mRNA abundance was observed in the terminal chase data points (20 minutes and beyond). Therefore, all samples were included in the deadenylation and decay modeling.

The next quality control step involved assessing the reproducibility of mRNA abundance and pA-tail length estimations across each replicate control sample (Appendix Figure S4A). This focus on control samples was justified due to the dynamic changes in mRNA levels and mean pA-tail length observed in the later chase data points. Read counts among individual samples correlated strongly, with Spearman's rho coefficients ranging from 0.85 to 0.96, a level of reproducibility comparable to other sequencing techniques such as Illumina. The reproducibility of mean pA-tail lengths for the entire transcriptome had rho values between 0.26 and 0.49 (black dots). However, when restricting the analysis to datasets containing at least 10 reads per transcript, the rho coefficient increased to a range of 0.52 to 0.70 (grey dots). Notably, for highly expressed RPG mRNAs, this coefficient ranged from 0.69 to 0.94. These observations prompted us to conduct deadenylation modeling exclusively on datasets with at least 10 reads per transcript. This also indicated that any estimations of deadenylation rates for RPG mRNAs are highly reliable, regardless of library size, as these transcripts are typically represented by many dozens to hundreds of reads individually.

Next, to assess the reproducibility of decay and deadenylation rate estimates across replicates, we performed a trial calculation for the Mex67-AA chase dataset (Appendix Figure S4B). To this end, we estimated the decay and deadenylation coefficient for each replicate separately using the given replicate control sample as a reference. Therefore, the decay factor was estimated from five samples in each replicate, while the deadenylation factor was estimated from four (for replicate A and B) or three (for replicate H) samples (Appendix Figure S4C, calculation setting presented in the top panel). As described in the Methods section and the main text, the exponential deadenylation factor was calculated separately from the 75th, 80th, 85th, 90th , and 95th quantiles and then averaged. Next, we compared the decay and deadenylation coefficients across replicates. The decay coefficient demonstrated high reproducibility, with Spearman's rho values ranging between 0.65 and 0.79 across all mRNAs and between 0.72 and 0.84 for transcripts represented by at least 10 reads. The Spearman's rho correlation coefficients for the quantile deadenylation rate were lower than those for decay, but this outcome was anticipated due to the greater complexity of estimation of pA-tail length statistics, as previously discussed. For mRNAs represented by at least 10 reads, the rho values were 0.52, 0.47, and 0.38. The lowest correlations were observed with replicate H, which had the poorest overall read count and three instead of four time points used for modeling deadenylation. Notably, when considering only the abundant ribosomal protein-encoding mRNAs (RPG mRNAs), the correlation was substantially stronger across all replicates, with rho values ranging from 0.60 to 0.74. We also performed the same calculations on the thiolutin and heat stress chase experiments. For this analysis, we narrowed the datasets to 517-519 transcriptionally down-regulated mRNAs described by Vinayachandran et al (2018). We observed a high correlation of decay (0.66-0.81 for heat stress and 0.75 for thiolutin) and deadenylation rates (0.42-0.7 for heat stress and 0.66 for thiolutin) across all replicates. In summary, we demonstrated that the three chase datasets exhibited strong reproducibility of key phenotypic features, including mRNA abundance and mean pA-tail length. Furthermore, these datasets proved effective for calculating exponential decay and deadenylation coefficients, yielding highly satisfactory results.

The quality of deadenylation and decay rate estimation depends on the sequencing depth of the obtained datasets and the number of time-points. Therefore, to enhance the modeling power, we increased the number of samples used to estimate decay and deadenylation coefficients. This was achieved by treating each time point in the biological replicate chase dataset as a control for the consecutive time-points (Appendix Figure S4C, compare modeling strategies shown in the the top and bottom schemes). Analyzing the Mex67-AA chase dataset, we encountered a challenge. To observe cytoplasmic decay events, it was essential to fully inhibit new mRNA synthesis after efficient Mex67 depletion. This time delay could potentially set the control samples considerably apart from the other time points in the dataset, introducing bias in estimates derived from the control sample alone (Appendix Figure S4C). To address this issue, the precise onset of the Mex67-depletion phenotype was calculated during the modified gamma modeling and estimated to be 2.67 min. For the heat stress response, experimental evidence from RNA polymerase II ChIP assays demonstrated that transcriptional silencing of both RPG and non-RPG mRNAs occurs within 3 minutes (Vinayachandran et al, 2018). Therefore, we assumed the time delay between the onset of the phenotype and transcriptional shutdown to be negligible, implying that transcriptional shut-down happens almost instantaneously following a rapid temperature shift. The correction applied increased the modeling power, especially for the deadenylation rate calculations.

After establishing the uniformity of the chase replicates, we conducted a bulk estimation of decay and deadenylation coefficients from all replicates. In case of the deadenylation rate, for each quantile separately (95th, 90th, 85th, 80th 75th, 50th, 15th, 10th and 5th) the mean, median and standard deviation of the deadenylation rate was calculated. Next, using the 95th-75th quantiles the mean and medians were calculated from respective intermediate values, and listed along with a standard deviation in Expanded View Dataset 1. Throughout the main manuscript the deadenylation rate '*(...)mean_mean*' values were used, including for calculation of the terminal adenosine half-lives. Next, we compared the standard deviation of the decay and deadenylation coefficient estimate calculated for each chase experiment (Appendix Figure S4D). This showed that the value of the standard deviation was substantially lower (up to 2-fold) than the decay rate for the heat stress and thiolutin chase experiments. In the case of the Mex67-AA chase the relatively high standard deviations for the decay rate estimate encouraged further investigations. We cross-compared our mRNA half-life estimates to four published datasets (see text related to Figure 1D.-E. and Appendix Figure S1C.-D.; Miller et al, 2011; Neymotin et al, 2014; Chan et al, 2018; Presnyak et al, 2015) and obtained good correlations with independent studies based on metabolic 4tU-labelling in otherwise wild-type cells (rho = 0.74 for Miller et al, 2011; rho = 0.62 for Chan et al, 2018 and rho = 0.44 for Neymotin et al, 2014). Lower reproducibility was observed in a chase dataset obtained through heat-inactivation of a strain encoding a mutated RNA polymerase II (rho = 0.31, Presnyak et al., 2015). This suggests that the relatively high standard deviation of the decay rate estimate in the Mex67-AA chase is not concerning in terms of data quality, as our estimates are comparable to those previously published. In contrast, the standard deviations of our deadenylation rate estimates were much lower than the coefficient itself (Appendix Figure S4D). We validated our deadenylation rate estimates by comparing the decay functions drawn using the obtained estimates with the experimental quantile pA-tail length values used for the calculations. For this purpose, we selected twelve mRNAs characterized by varying quantile deadenylation rates (slow or fast deadenylation; see Figure EV4A). This analysis revealed that the quantile deadenylation rates accurately predicted the changes in experimental quantile pA-tail lengths. Coupled with the low standard deviation of our quantile deadenylation rate estimates, this indicates the reliability of the obtained coefficients.

2. APPENDIX TABLES

Appendix Table S1. Yeast strains used in this study.

Name	Identifier	Source
Wild-type W303: MAT A; *leu2-3,112; trp1-1, can1-100; ura3-1, ade2-1; leu2-3,112; his3-11,15*	T. H. Jensen collection Y159; A. Tudek collection Y13	Euroscarf Cat#BMA64-1A
Dcp2-AID-FLAG: As W303, *OsTIR1:URA3, DCP2::mAID-6flag::HygR*	T. H. Jensen collection Y3798; A. Tudek collection Y45	This study
Dcp2-AID-FLAG Mex67-AA: As W303; *tor1-1 fpr1::LEU RPL13-2xFKBP12::loxP-TRP1-loxP OsTIR::URA MEX67:FRB::kanMX DCP2-mAID-6FLAG::HygR*	T. H. Jensen collection Y3927; A. Tudek collection Y62	This study
Pab1-AID-6HA; Mex67-AA: As W303, *tor1-1 fpr1::LEU RPL13-2xFKBP12::loxP-TRP1-loxP OsTIR::URA MEX67-FRB::KANmx PAB1-mAID-6HA::HygR*	T. H. Jensen collection Y3879	This study
dhh1Δ: as W303, *DHH1::URA3ca*	A. Tudek collection Y68	This study
lsm1Δ: as W303, *LSM1::URA3ca*	A. Tudek collection Y69	This study
Mex67-AA: as W303, *tor1-1 fpr1::NAT RPL13-2xFKBP12::TRP1 MEX67-FRB::kanMX6*	T. H. Jensen collection 2618; A. Tudek collection Y39	Euroscarf from Haruki et al, 2018: Cat#HHY182
Wild-type to match Tucker et al. (2001) strains: *MATa leu2-3,112 trp1-1 ura3-52 his4-539 cup1::LEU2/PGK1pG/MFA2pG*	R. Parker collection yRP840; A. Tudek collection Y79	Tucker et al, 2001
pan2Δ: *MATa leu2-3,112 trp1-1 ura3-52 his4-539 cup1::LEU2/PGK1pG/MFA2pG pan2Δ::URA3*	R. Parker collection yRP1619; A. Tudek collection Y80	Tucker et al, 2001
ccr4Δ: *MATa leu2-3,112 trp -1 ura3-52 his4-539 cup1::LEU2/PGK1pG/MFA2pG ccr4Δ::NEO*	R. Parker collection yRP1616; A. Tudek collection Y81	Tucker et al, 2001
pan2Δ ccr4Δ: *MATa leu2-3,112 trp -1 ura3-52 his4-539 cup1::LEU2/PGK1pG/MFA2pG ccr4Δ::NEO pan2Δ::URA3*	R. Parker collection yRP1620; A. Tudek collection Y82	Tucker et al, 2001
Xrn1-AID: as W303, His+ tor1-1 fpr1::LEU RPL13-2xFKBP12::loxP-TRP1-loxP HIS+ OsTIR::URA3 XRN1::AID::KANmx	T. H. Jensen collection Y3688; A. Tudek collection Y43	Tudek et al. 2018

Appendix Table S2. Oligonucleotides used in this study

Name	**Sequence**	**Source**
25S_rRNA_FWD	ATTCCCACTGTCCCTATCTACT	Custom (LifeTechnology)
25S_rRNA_REV	CTTGGCTGTGGTTTCGCT	Custom (LifeTechnology)
HHF1_FWD	ACTGCCCGGTTTTTCTTCT	Custom (LifeTechnology)
HHF1_REV	CCTAAACCCGCTATAATACACTCAT	Custom (LifeTechnology)
Hsp104_FWD	AGCTGAAGAATGTCTGGAAGT	Custom (LifeTechnology)
Hsp104_REV	CGTCATCACCTAACGTGTCA	Custom (LifeTechnology)
Markers for bulk RNA pA-tail length	CCCCACCACCAUCACUUA(3) CCCCACCACCAUCACUUA(14) CCCCACCACCAUCACUUA(34)	Future Synthesis
oligo-dT(18)	TTTTTTTTTTTTTTTTTT	Custom (LifeTechnology)
Random hexamers		Invitrogen (48190-011)
Rpl21B_FWD	ACAGATCTCGTACACGTTACA	Custom (LifeTechnology)
Rpl21B_REV	CGACAATGTCACCAACCT	Custom (LifeTechnology)
Rpl36A_FWD	AAGGTAAGAAGGTCACTAGCA	Custom (LifeTechnology)
Rpl36A_REV	GTTGGAAGCAGCACCTTT	Custom (LifeTechnology)
RPL28_FWD	CGGTAAAGGTCGTATCGGT	Custom (LifeTechnology)
RPL28_REV	TCCATGTTAATTCTGTGGTGATGT	Custom (LifeTechnology)
RPS5 Northern blot probe	CTTAACGGTTAGACTTGGCAACACGTTCCAATTCATCCTTCTTCTTGATAGCGTAAGAAG	Custom (LifeTechnology)
RPS5 RNaseH-targeting oligonucleotide	GGCCAAAGTTTCAGCAATGGTC	Custom (LifeTechnology)
RPS13_FWD	ATTTCTTCTTCTGCTATTCCATACTCT	Custom (LifeTechnology)
RPS13_REV	CCCTTTCTCGCGTACTTGAC	Custom (LifeTechnology)
SSA4_FWD	AAATTGTACTTGTTGGTGGTTCA	Custom (LifeTechnology)
SSA4_REV	GGGTTAATCGAACGGTTTGG	Custom (LifeTechnology)
SCR1 Northern blot probe	GGTCACCTTTGCTGACGCTGG	Custom (LifeTechnology)
TDH3_FWD	CTCTCACTCTTCCATCTTCGAT	Custom (LifeTechnology)
TDH3_REV	CGTACCAGGAGACCAACTT	Custom (LifeTechnology)

Appendix Table S3. List of the *γ_shape* and *γ_rate* parameters of the modified gamma distribution*.

distribution	**replicate**	**Name**	**time**	***γ_shape***	***γ_shape* standard error**	***γ_rate***	***γ_rate* standard error**
all	A	ORFs_repA	0	4,7382963016	0,1701733983	0,0443559959	0,0010031239
all	A	ORFs_repA	12	4,5458072251	0,1709085326	0,0762571365	0,0020174607
all	A	ORFs_repA	14	4,6083475182	0,168448713	0,0815975926	0,002133893
all	A	ORFs_repA	16	4,8639966029	0,1753657642	0,087245794	0,0022913976
all	A	ORFs_repA	30	5,2342823366	0,1819466638	0,1141303491	0,0030768485
all	A	ORFs_repA	8	4,3443018564	0,1549567484	0,0655559453	0,0015840391
all	B	ORFs_repB	0	3,9477063331	0,0911411231	0,0511873469	0,0007594507
all	B	ORFs_repB	10	4,2942553682	0,1421460508	0,071834078	0,001648011
all	B	ORFs_repB	12	4,5649936518	0,1616177987	0,075070907	0,0018626591
all	B	ORFs_repB	14	4,6732992589	0,1589867501	0,0807358774	0,0019612077
all	B	ORFs_repB	20	4,9889139759	0,161404659	0,09254039	0,0022133365
all	B	ORFs_repB	4	4,198432916	0,1294216769	0,0522990157	0,0010384718
all	H	ORFs_repH	0	4,0679072475	0,0776561027	0,0548363985	0,0006835345
all	H	ORFs_repH	16	4,473338404	0,146948969	0,0982197119	0,0024128111
all	H	ORFs_repH	30	4,446216747	0,127695574	0,1256616164	0,0028422138
all	H	ORFs_repH	4	4,0473310821	0,0939395946	0,0600702287	0,000927981
all	H	ORFs_repH	60	3,6058199436	0,1301168254	0,145361458	0,0041971596
all	H	ORFs_repH	8	6,2086731339	0,3371068626	0,0634482699	0,0023046447
GAS1	A	GAS1_Mex67_repA	0	12,8878769438	2,2655815304	0,0265802005	0,0032441697
GAS1	A	GAS1_Mex67_repA	12	6,3919655705	1,3689190922	0,0547458302	0,0092779551
GAS1	A	GAS1_Mex67_repA	14	6,2650623134	0,6454890465	0,0697053807	0,0050987687
GAS1	A	GAS1_Mex67_repA	16	8,9633308236	0,7226478602	0,0852400776	0,0051341791
GAS1	A	GAS1_Mex67_repA	30	5,9157295834	0,9502770809	0,1229971594	0,0158369891
GAS1	A	GAS1_Mex67_repA	8	9,8906015085	2,0648593467	0,038581405	0,0060285048
GAS1	B	GAS1_Mex67_repB	0	12,6040841543	1,4189279791	0,0348463035	0,0025265516
GAS1	B	GAS1_Mex67_repB	10	10,5083794243	0,9226133331	0,0649782919	0,0039115551
GAS1	B	GAS1_Mex67_repB	12	9,6327202168	0,9413267313	0,0635621115	0,0041820153
GAS1	B	GAS1_Mex67_repB	14	8,0122699547	0,6846031221	0,0723561782	0,0042777458
GAS1	B	GAS1_Mex67_repB	20	7,5828969199	0,5517643955	0,0974706489	0,0053612539
GAS1	B	GAS1_Mex67_repB	4	14,067586699	2,2942563187	0,0328918473	0,003448606
HHF1	B	HHF1_Mex67_repB	0	6,5107924093	0,5834895098	0,0467976725	0,0027494249
HHF1	B	HHF1_Mex67_repB	10	5,1562236603	0,4261955779	0,0869064592	0,0054112257
HHF1	B	HHF1_Mex67_repB	12	4,2297377129	0,3584940978	0,0831228148	0,00519282
HHF1	B	HHF1_Mex67_repB	14	3,9341234102	0,4833755925	0,0819018723	0,0073712193
HHF1	B	HHF1_Mex67_repB	20	2,7006263757	0,5798381797	0,0668342626	0,0092484496

HHF1	B	HHF1_Mex67_repB	4	5,9981626059	0,579284607	0,0472671672	0,002942895
HHF1	A	HHF1_Mex67_repA	0	6,6145650838	1,0320558602	0,0291373227	0,0030225197
HHF1	A	HHF1_Mex67_repA	12	3,7479265145	1,3662826645	0,0709915031	0,0208753089
HHF1	A	HHF1_Mex67_repA	14	3,572616036	0,6737905223	0,0821874673	0,0121809386
HHF1	A	HHF1_Mex67_repA	16	5,3275541288	0,757245699	0,1021711043	0,0113029587
HHF1	A	HHF1_Mex67_repA	30	0,1297825643	0,5114497365	0,0067028791	0,0058751255
HHF1	A	HHF1_Mex67_repA	8	6,3802674481	1,1872879096	0,0573295013	0,0077858905
high	A	Mex67_high_abundance_ORFS_	0	4,9654002402	0,1430404473	0,0544502193	0,0010223591
high	A	Mex67_high_abundance_ORFS_	8	4,4507002021	0,1369398016	0,0700066827	0,00148884
high	A	Mex67_high_abundance_ORFS_	12	4,7092277065	0,1653948932	0,078703171	0,001967263
high	A	Mex67_high_abundance_ORFS_	14	4,7505973006	0,1662417261	0,0831457526	0,0020981125
high	A	Mex67_high_abundance_ORFS_	16	5,057551509	0,1753887568	0,088430846	0,002246691
high	A	Mex67_high_abundance_ORFS_	30	5,6661134788	0,1847040146	0,1179515794	0,0030100054
high	B	Mex67_high_abundance_ORFS_repB	0	4,1583258244	0,0848307042	0,0600664537	0,0008161981
high	B	Mex67_high_abundance_ORFS_repB	10	4,3737259612	0,1377950671	0,074463203	0,001645561
high	B	Mex67_high_abundance_ORFS_repB	12	4,65940655	0,1590682664	0,0769404313	0,0018574119
high	B	Mex67_high_abundance_ORFS_repB	14	4,7985933133	0,1616056564	0,0823170797	0,0019938044
high	B	Mex67_high_abundance_ORFS_repB	20	5,2350598403	0,1717914696	0,0943295448	0,0023023997
high	B	Mex67_high_abundance_ORFS_repB	4	4,3113343862	0,1094349881	0,0597594955	0,0010081333
low	A	Mex67_low_abundance_ORFS_	0	4,7934942737	0,2585077417	0,0361313324	0,0012129324
low	A	Mex67_low_abundance_ORFS_	8	4,2598497515	0,1874837742	0,0606844373	0,0017680253
low	A	Mex67_low_abundance_ORFS_	12	4,3086098793	0,1826785709	0,0728297541	0,002136068
low	A	Mex67_low_abundance_ORFS_	14	4,3769701007	0,1751196888	0,0791355528	0,0022357259
low	A	Mex67_low_abundance_ORFS_	16	4,5450343598	0,1770653555	0,0852472269	0,0023951438
low	A	Mex67_low_abundance_ORFS_	30	4,3424692286	0,1930820392	0,1056367641	0,00355708
low	B	Mex67_low_abundance_ORFS_repB	0	3,8488234017	0,1266423076	0,0427803314	0,0008794938
low	B	Mex67_low_abundance_ORFS_repB	10	4,1916359993	0,1522345815	0,0682804239	0,00168943
low	B	Mex67_low_abundance_ORFS_repB	12	4,4227910912	0,1683514236	0,0723216818	0,0019028602
low	B	Mex67_low_abundance_ORFS_repB	14	4,469556271	0,1584036705	0,0782098125	0,001955639
low	B	Mex67_low_abundance_ORFS_repB	20	4,5480351401	0,1499378985	0,0892115375	0,0021480054
low	B	Mex67_low_abundance_ORFS_repB	4	4,218966619	0,1793497749	0,0450593145	0,0012005542
non_RPG	A	Mex67_NON_RPG_ORFS_	0	4,9672842436	0,2041426011	0,0420164582	0,0010823057
non_RPG	A	Mex67_NON_RPG_ORFS_	8	4,5478286662	0,1737482453	0,0640898907	0,001648397
non_RPG	A	Mex67_NON_RPG_ORFS_	12	4,7113500016	0,1880873248	0,0745866179	0,0020824396
non_RPG	A	Mex67_NON_RPG_ORFS_	14	4,7899489509	0,1817291368	0,0800705223	0,0021630123
non_RPG	A	Mex67_NON_RPG_ORFS_	16	5,0501213504	0,1880011016	0,0859854215	0,0023234117
non_RPG	A	Mex67_NON_RPG_ORFS_	30	5,3781243432	0,1882079531	0,1125937656	0,0030494312

non_RPG	B	Mex67_NON_RPG_ORFS_repB	0	4,0641918842	0,1090397351	0,0484523351	0,0008273292
non_RPG	B	Mex67_NON_RPG_ORFS_repB	10	4,4557985875	0,1593752927	0,069907468	0,001720062
non_RPG	B	Mex67_NON_RPG_ORFS_repB	12	4,7420000963	0,1795998684	0,0732750569	0,001931857
non_RPG	B	Mex67_NON_RPG_ORFS_repB	14	4,8351224907	0,1729630996	0,0790030243	0,0020061326
non_RPG	B	Mex67_NON_RPG_ORFS_repB	20	5,1578102871	0,169904363	0,0910024898	0,0022076077
non_RPG	B	Mex67_NON_RPG_ORFS_repB	4	4,3683962591	0,1548699682	0,0498433958	0,0011278062
RPG	A	Mex67_RPG_ORFS_	0	4,3858961152	0,1058700093	0,0575355128	0,0009149546
RPG	A	Mex67_RPG_ORFS_	12	4,1553868064	0,1301502105	0,0845869745	0,0019155143
RPG	A	Mex67_RPG_ORFS_	14	4,1607554066	0,1346566299	0,0893391874	0,0021160282
RPG	A	Mex67_RPG_ORFS_	16	4,3703662408	0,1408138464	0,0938420789	0,0022411947
RPG	A	Mex67_RPG_ORFS_	30	4,8103893624	0,1690452851	0,1210655792	0,0033317169
RPG	A	Mex67_RPG_ORFS_	8	3,8644026949	0,1139602911	0,0739623327	0,0015281639
RPG	B	Mex67_RPG_ORFS_repB	0	3,8555028598	0,0582620036	0,0664051418	0,0006846856
RPG	B	Mex67_RPG_ORFS_repB	10	3,9489536256	0,0993100557	0,0818493924	0,0014775611
RPG	B	Mex67_RPG_ORFS_repB	12	4,1741445384	0,1156586365	0,0843256963	0,0016897826
RPG	B	Mex67_RPG_ORFS_repB	14	4,2886229838	0,1223996468	0,0897249208	0,0018799829
RPG	B	Mex67_RPG_ORFS_repB	20	4,531451744	0,1373959954	0,1005650055	0,0022971299
RPG	B	Mex67_RPG_ORFS_repB	4	3,9586374237	0,0731793719	0,0658968609	0,0008289751
RPL36A	A	RPL36A_Mex67_repA	0	2,8192824408	0,3875104317	0,0647922677	0,0062227307
RPL36A	A	RPL36A_Mex67_repA	12	1,3721045181	0,4193238798	0,0647006171	0,0138986434
RPL36A	A	RPL36A_Mex67_repA	14	1,6425974593	0,2100203444	0,0972620201	0,0087508354
RPL36A	A	RPL36A_Mex67_repA	16	1,5166212554	0,2149594425	0,0870663529	0,0083362171
RPL36A	A	RPL36A_Mex67_repA	30	1,7650534662	0,1940463142	0,145168502	0,0120251798
RPL36A	A	RPL36A_Mex67_repA	8	0,7462916293	0,3792584594	0,0445654237	0,0168873429
RPL36A	B	RPL36A_Mex67_repB	0	2,1049159955	0,2743369736	0,0703452894	0,0062238436
RPL36A	B	RPL36A_Mex67_repB	10	1,9638093588	0,2320665218	0,0882218348	0,0074217638
RPL36A	B	RPL36A_Mex67_repB	12	2,3951711021	0,303675421	0,0859376154	0,0078766261
RPL36A	B	RPL36A_Mex67_repB	14	2,1362480698	0,1726494221	0,104737701	0,0062705979
RPL36A	B	RPL36A_Mex67_repB	20	2,0780686213	0,2028633333	0,1074758501	0,0075829429
RPL36A	B	RPL36A_Mex67_repB	4	2,3159699391	0,2573549803	0,0817314839	0,0064354323
RPL4A	B	RPL4A_Mex67_repB	0	9,172571753	0,6554593614	0,0471791874	0,0021513514
RPL4A	B	RPL4A_Mex67_repB	10	9,2060871326	0,6338473303	0,0696616726	0,0033154902
RPL4A	B	RPL4A_Mex67_repB	12	9,6473512415	0,5408277986	0,0773711923	0,0030662237
RPL4A	B	RPL4A_Mex67_repB	14	8,9636686903	0,4260715937	0,0858034461	0,0029696257
RPL4A	B	RPL4A_Mex67_repB	20	7,5764018507	0,310144215	0,1031378864	0,0032820025
RPL4A	B	RPL4A_Mex67_repB	4	11,5571626981	0,9525045631	0,0490049967	0,0025948041
RPL4A	A	RPL4A_Mex67_repA	0	12,4593701276	0,9660117938	0,0480937052	0,0024011968
RPL4A	A	RPL4A_Mex67_repA	12	9,1933802778	0,8813587195	0,0786350812	0,0055350599

RPL4A	A	RPL4A_Mex67_repA	14	8,6831377247	0,5264877475	0,0911187785	0,0041281053
RPL4A	A	RPL4A_Mex67_repA	16	8,8569352177	0,3520108131	0,1137183769	0,00355103
RPL4A	A	RPL4A_Mex67_repA	30	8,9253033819	1,5156237934	0,1760605628	0,0258553617
RPL4A	A	RPL4A_Mex67_repA	8	8,0298469685	0,8059447231	0,055392504	0,0037789706

*a .csv file is deposited at Mendeley.

Appendix Table S4. Basic DRS data statistics for datasets other than the chase experiments.

Sample:	**All PASS reads**		**mRNAs reads**		**Heat stress repressed mRNA reads**		**RPG mRNA reads**	
	count number	**rho: counts/ mean pA-tail**	**count number**	**rho: counts/ mean pA-tail**	**count number**	**rho: counts/ mean pA-tail**	**count number**	**rho: counts/ mean pA-tail**
Pab1-AID control 1 Pab1-AID control 2	1461085 326593	0.93/0.42	1433042 320052	0.93/0.43	755069 167353	0.98/0.73	585883 129810	0.98/0.73
Pab1-AID 1h auxin 1 Pab1-AID 1h auxin 2	1118293 1145958	0.97/0.48	1096564 1123276	0.96/0.55	567206 540441	0.99/0.85	442538 421278	0.99/0.95
Pab1-AID 2h auxin 1 Pab1-AID 2h auxin 2	569602 1171693	0.93/0.42	554508 1148828	0.93/0.46	259303 543553	0.98/0.81	199356 415730	0.93/0.87
Control for Xrn1-AID and Dcp2AID replicate 1 replicate 2	1136743 1195900	0.97/0.44	1104241 1172287	0.97/0.54	380041 416018	0.99/0.68	243639 266540	0.98/0.93
Dcp2-AID replicate 1 Dcp2-AID replicate 2	315388 687910	0.95/0.34	302813 662182	0.95/0.39	95793 169021	0.97/0.51	58419 92399	0.96/0.75
Xrn1-AID replicate 1 Xrn1-AID replicate 2	306564 737127	0.95/0.34	296792 713751	0.95/0.41	141424 340845	0.98/0.64	101982 243608	0.99/0.89
pan2Δ ccr4Δ 25°C rep1 *pan2Δ ccr4Δ* 25°C rep2 *pan2Δ ccr4Δ* 25°C rep3	428496 541867 424573	(0.94-0.93-0.94)/(0.26-0.25-0.26)	418937 528708 413829	(0.94-0.93-0.94)/(0.28-0.28-0.27)	187184 254563 196043	(0.99-0.99-0.99)/(0.64-0.66-0.68)	127163 174407 131024	(0.99-0.98-0.99)/(0.87-0.86-0.89)
pan2Δ ccr4Δ 37°C rep1 *pan2Δ ccr4Δ* 37°C rep2 *pan2Δ ccr4Δ* 37°C rep3	592954 156325 176421	(0.89-0.89-0.87)/(0.28-0.30-0.24)	581615 152947 172571	(0.89-0.88-0.88)/(0.30-0.31-0.24)	219324 61010 66955	(0.97-0.98-0.97)/(0.53-0.66-0.52)	144836 39698 42464	(0.97-0.97-0.98)/(0.8-0.76-0.76)
Mex67-AA 25°C rep1 Mex67-AA 25°C rep2	2184459 1118897	0.96/0.47	2141570 1098396	0.96/0.49	1081952 539068	0.99/0.83	756476 374611	0.96/0.91
Mex67-AA 37°C rep1 Mex67-AA 37°C rep2	841310 1415229	0.96/0.53	827924 1392269	0.96/0.56	295373 392041	0.97/0.66	202933 243062	0.97/0.77
Mex67-AA 37°C +rapamycin rep1 Mex67-AA 37°C +rapamycin rep2	1406000 1049059	0.88/0.41	1381734 1028633	0.89/0.43	750886 549364	0.97/0.45	576594 413477	0.97/0.83
WT 30°C replicate1 WT 30°C replicate2 WT 30°C replicate3 WT 30°C replicate4 WT 30°C replicate5 WT 30°C replicate6 WT 30°C replicate7 WT 30°C replicate8	457235 960094 644548 3844655 874196 525983 774954 596447	*(0.94-0.96-0.95-0.94-0.56-0.96-0.97-0.96)/(0.54-0.65-0.61-0.55-0.76-0.71-0.75-0.70)	447134 940138 633208 3745464 867826 515572 755067 581467	*(0.93-0.96-0.95-0.92-0.54-0.95-0.97-0.96)/(0.55-0.65-0.61-0.59-0.78-0.73-0.77-0.72)	244466 528309 348321 1605245 137717 293606 415717 311665	*(0.99-0.99-0.99-0.96-0.95-0.98-0.99-0.99)/(0.80-0.91-0.87-0.87-0.64-0.84-0.80-0.80)	178995 394153 255176 1175348 92547 223015 313297 230320	*(0.95-0.97-0.94-0.92-0.83-0.91-00.96-0.97)/(0.93-0.96-0.94-0.89-0.84-0.89-0.94-0.92)
ccr4Δ 30°C replicate 1 *ccr4Δ* 30°C replicate 2	459838 285056	0.79/0.49	449734 279388	0.77/0.52	205757 154516	0.89/0.80	137171 118949	0.51/0.86
pan2Δ 30°C replicate 1 *pan2Δ* 30°C replicate 2	311266 475953	0.93/0.37	304145 464686	0.92/0.37	167681 258399	0.99/0.59	121460 191672	0.98/0.88
pop2Δ 30°C replicate 1 *pop2Δ* 30°C replicate 2	287286 525820	0.95/0.50	279923 512622	0.94/0.55	122635 217714	0.98/0.84	87657 154100	0.98/0.90
WT 30°C WT 30°C +Xrn1 *in vitro*	1291153 1090351	0.96/0.55	1251152 1042261	0.95/0.61	620428 550406	0.99/0.80	449596 402027	0.98/0.86

*due to the large number of replicates in this datasets the Spearman rho coefficients were calculated between a given replicate and the sum of all reads.

3. APPENDIX REFERENCES

Bilska, A., Kusio-Kobiałka, M., Krawczyk, P. S., Gewartowska, O., Tarkowski, B., Kobyłecki, K., Nowis, D., Golab, J., Gruchota, J., Borsuk, E., Dziembowski, A., Mroczek, S. (2020) Immunoglobulin expression and the humoral immune response is regulated by the non-canonical poly(A) polymerase TENT5C. Nat Commun. 11(1):2032. doi: 10.1038/s41467-020-15835-3.

Kadanoff, L. P. 2009. More is the Same; Phase Transitions and Mean Field Theories. Journal of Statistical Physics. 137 (5-6) , pp.777-797. DOI 10.1007/s10955-009-9814-1

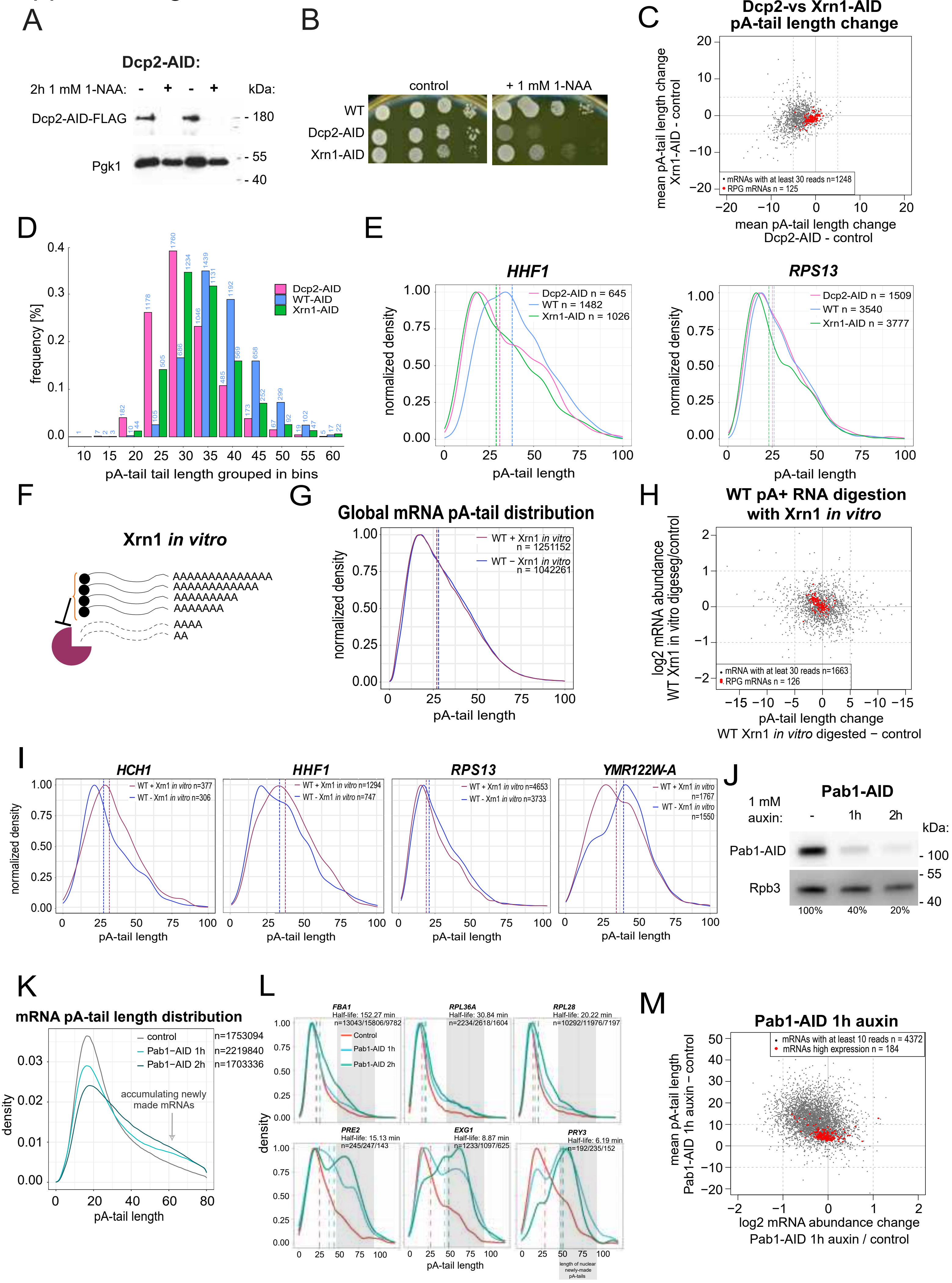
Posit team (2024). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/

RStudio Team (2019). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/

Tucker, M., Valencia-Sanchez, M. A., Staples, R. R., Chen, J., Denis, C. L., Parker, R. (2001) The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. Cell. 104(3):377-86. doi: 10.1016/s0092-8674(01)00225-2.

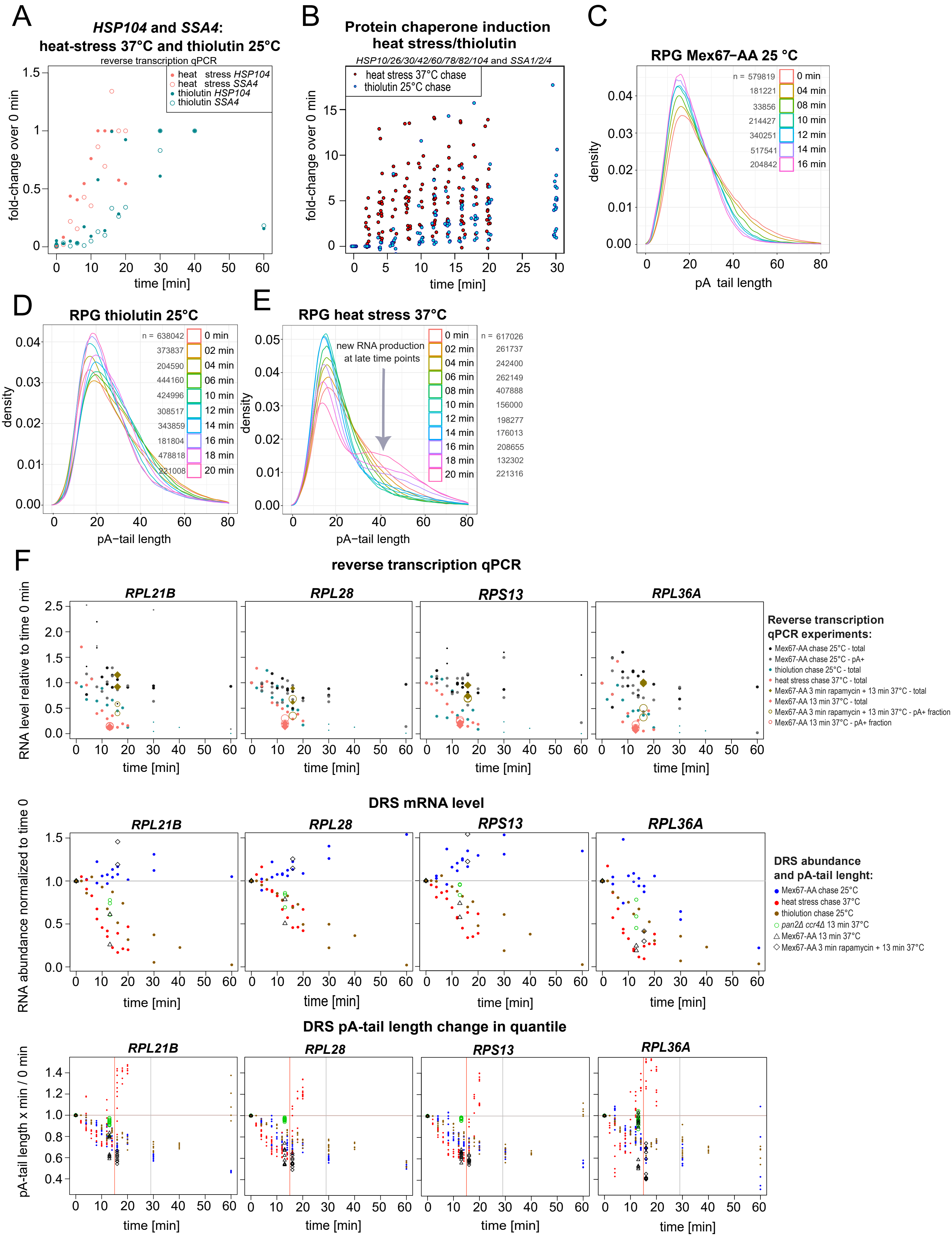
4. APPENDIX FIGURES

Appendix Figure S1



Appendix Figure S1. Assessment of the impact of Dcp2, Xrn1 and Pab1 on mRNA pA-tail lengths *in vivo*. **A.** Western blot shows the efficiency of depletion of Dcp2 using the AID system for 2 hours. PGK1 is used as a loading control. **B.** Growth test produced using 10-fold serial dilutions of Dcp2-AID and Xrn1-AID cells grown on a control or auxin-supplemented medium. **C.** Scatterplot compares the absolute change in mean pA-tail length in Dcp2- and Xrn1-depleted cells compared to control. **D.** Proportion of transcripts with specific median pA-tail lengths in designated value ranges presented for control, Dcp2- and Xrn1-depleted cells. **E.** pA-tail length distribution of *HHF1* and *RPS13* mRNAs in control and Dcp2- or Xrn1-depleted cells. **F.** Scheme shows that Xrn1 can only digest mRNAs devoid of a cap, potentially altering pA-tail distribution if transcripts are massively decapped only at a defined pA-tail length. **G.** Global pA-tail length distribution of coding transcripts in wild-type sample digested or not with Xrn1 *in vitro*. **H.** Absolute change in mean pA-tail length in relation to the log2 change in mRNA abundance in between a wild-type sample digested or not with Xrn1 *in vitro*. **I.** pA-tail length distributions of *HCH1*, *HHF1*, *RPS13* and *YMR122W-A* mRNAs in a control wild-type sample compared to the same sample digested with Xrn1 *in vitro*. **J.** Western blot shows depletion of Pab1 protein using the auxin-inducible degron for 1 and 2 hours. Rpb3 is shown as a loading control. The percentages shown below the Rpb3 loading control designate the estimated amount of Pab1-AID remaining in the sample relative to the control. **K.** Distribution of all coding transcript pA-tail lengths in control wild-type cells compared to 1 or 2 hours Pab1-depleted cells. **L.** pA-tail length distribution of *FBA1*, *RPL36A*, *RPL28*, *PRE3*, *EXG1* and *PRY3* mRNAs in control cells compared to strains depleted for Pab1 using the AID system for 1 or 2 hours. The mRNAs are ranked by half-life time. The number of reads that sum into each distribution is indicated on the panel with n='control'/'1h depletion'/'2h depletion'. **M.** Comparison of log2 change in mRNA abundance to absolute change in mean pA-tail length for cells depleted for Pab1 for 1 hour compared to wild-type control.

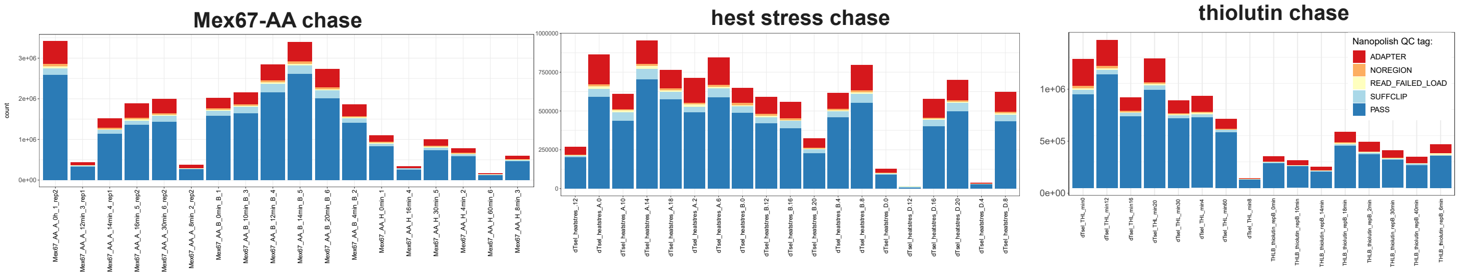
Appendix Figure S2



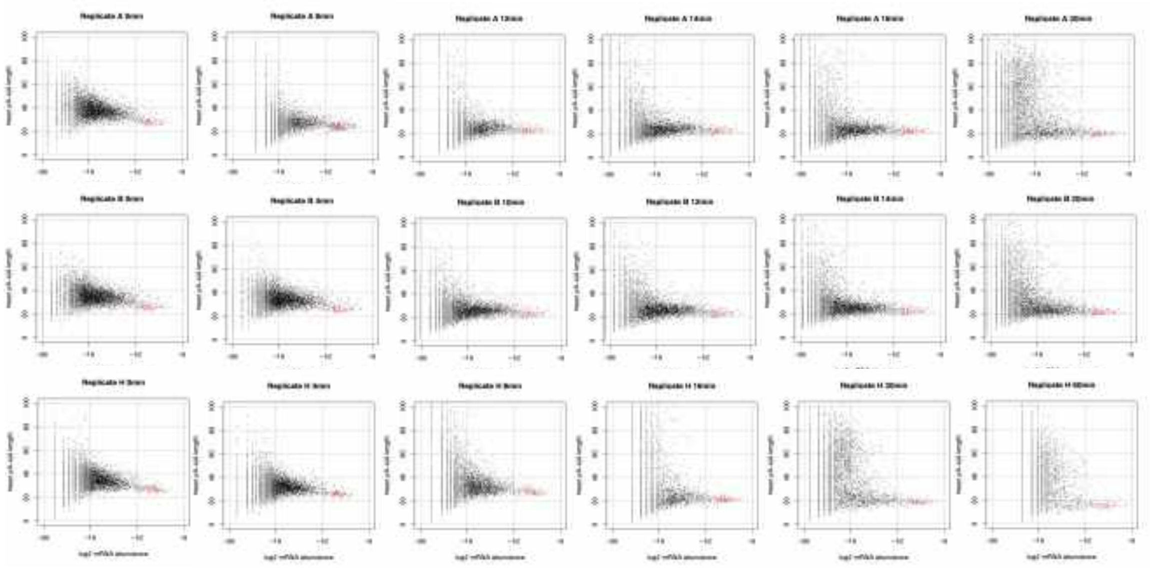
Appendix Figure S2. Transcriptomic changes upon heat stress and thiolutin treatment. **A.** Graph shows up-regulation of *HSP104* and *SSA4* mRNA levels during heat stress at 37 °C and thiolutin treatment at 25 °C by reverse transcription coupled to qPCR. **B.** Plot shows the up-regulation of a collection of chaperone mRNAs in the heat-stress at 37 °C and thiolutin treatment at 25 °C chase sequencing data. Those transcripts are known to be transcriptionally induced during heat stress (Vinayachandran et al, 2018). **C.-E.** Global distributions of RPG mRNA pA-tail lengths in chase experiments: Mex67-depletion (C.), thiolutin treatment at 25 °C (D.) and heat stress at 37 °C (E.). The arrow on the heat stress density plot points to a peak of long pA-tailed mRNAs accumulating in later heat stress chase time points. Since transcriptional down-regulation of RPGs during heat stress is only temporary (Vinayachandran et al, 2018), those can only be new transcripts. The contribution of new mRNAs in late heat stress time points was too important (they strongly distorted quantile values) and therefore those samples were not used for deadenylation modeling. **F.** Series of graphs show in rows from top to bottom time dependent: (top) mRNA abundance by reverse-transcription coupled to qPCR, mRNA, (middle) abundance by DRS and (bottom) change in the pA-tail length in upper quantiles (75-95th) normalized to control for single RPG mRNAs: *RPL21B*, *RPL36A*, *RPS13*, and *RPL28* in Mex67-depletion, heat-stress 37 °C and thiolutin 25 °C chase sequencing data. Legends for each type of experiment are given separately.

Appendix Figure S3

A **DRS library read count by category**



B **Mex67-AA chase**



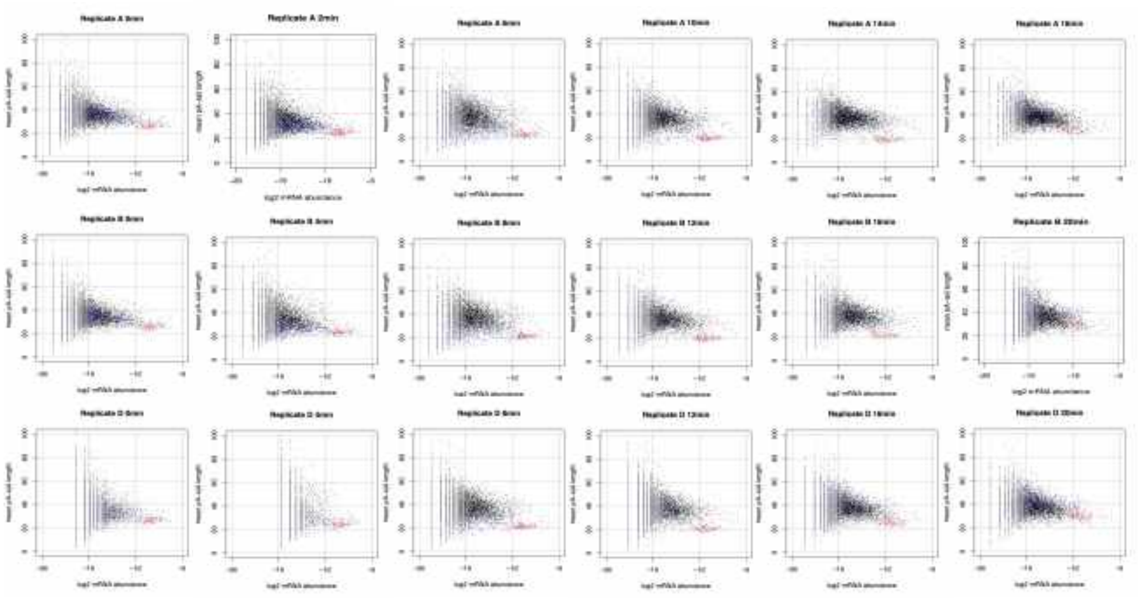
Legend:

● all mRNAs

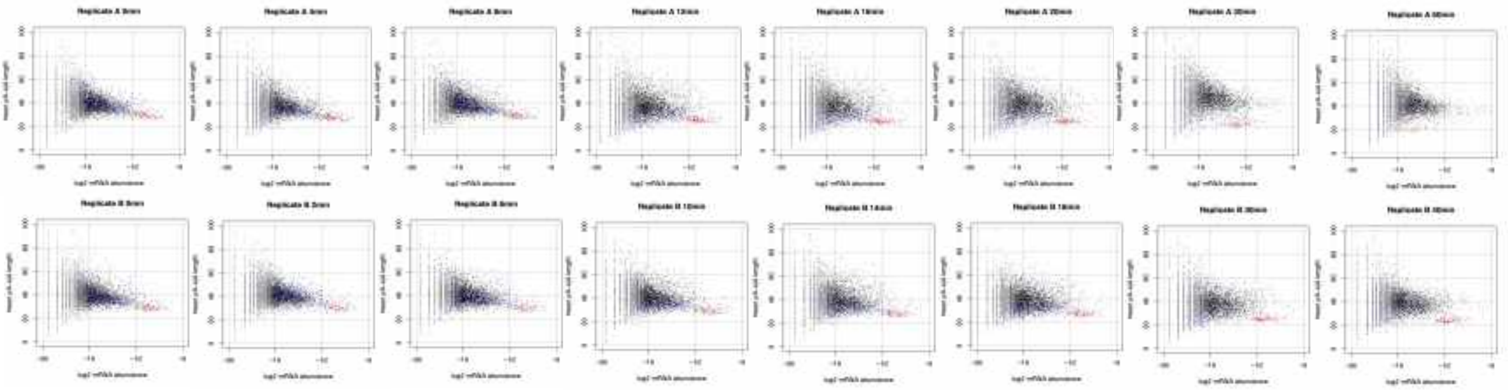
● RPG mRNAs

● non-RPG mRNAs transcriptionally silenced during heat stress

Heat stress chase

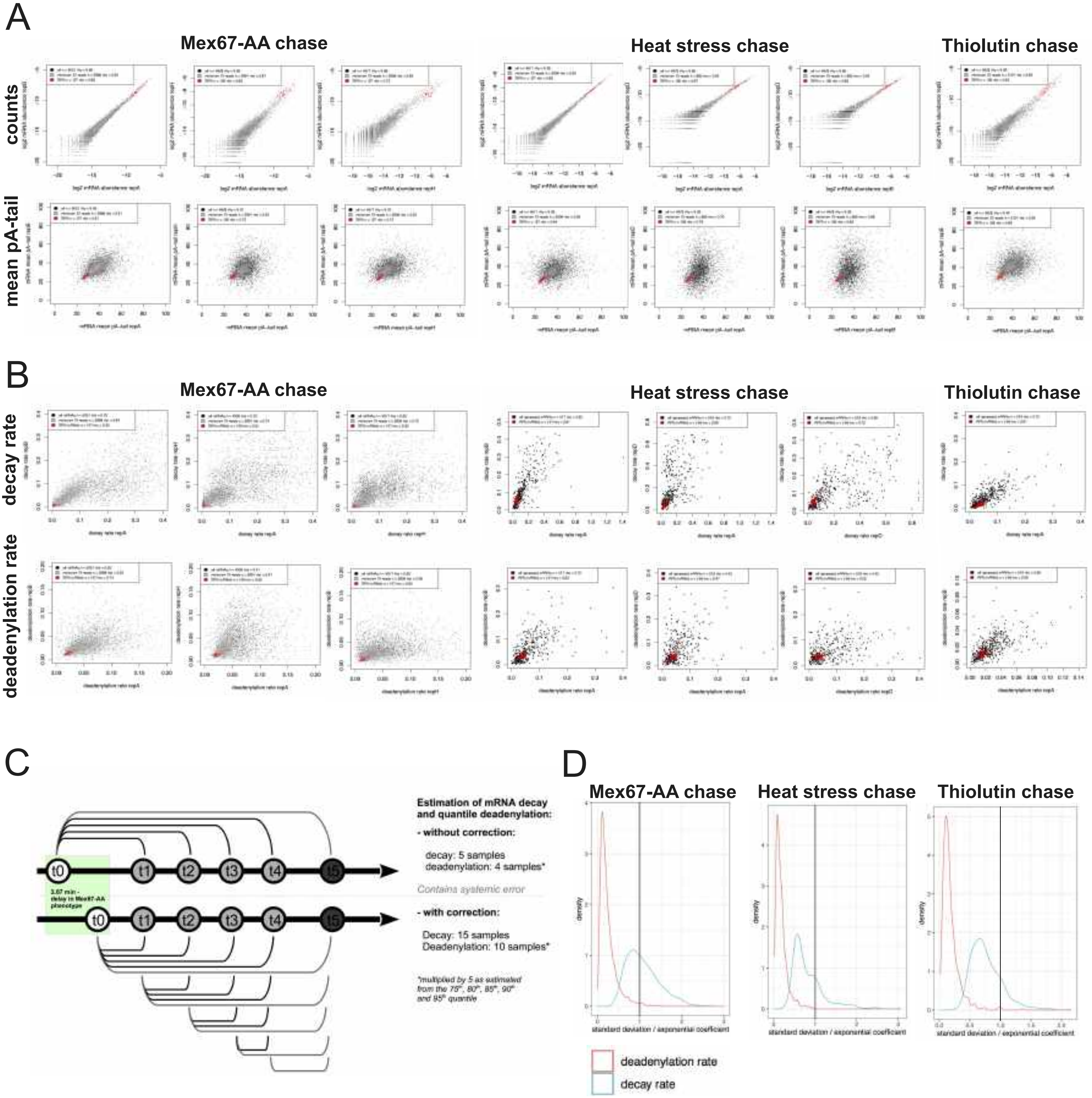


Thiolutin chase



Appendix Figure S3. General DRS chase quality control. **A.** Bar plots show each DRS chase sample read count with division into read quality control tags. Only the PASS counts were used for further analyses. **B.** Matrixes of scatterplots comparing the log2 mRNA abundance on y-scale to the transcripts' mean pA-tail length on the x-axis for the three chase experiments (Mex67-depletion, heat stress and thiolutin treatment). Where appropriate selected groups of mRNAs are highlighted as indicated in the legend. The number of mRNAs shown on each figure is indicated in each panel legend.

Appendix Figure S4



Estimation of mRNA decay and quantile deadenylation:

- without correction:

decay: 5 samples
deadenylation: 4 samples*

Contains systemic error

- with correction:

Decay: 15 samples
Deadenylation: 10 samples*

**multiplied by 5 as estimated from the 75th, 80th, 85th, 90th and 95th quantile*

Appendix Figure S4. DRS quality control and decay/deadenylation modeling strategy. **A.** Scatter plots showing the correlation between read count in log2-scale (top row) and mean pA-tail length (bottom row) in the three chase experiments. Only the control samples for each replicate were analyzed as only these samples are expected to be similar. The plots show all detected mRNAs (black dots), mRNAs detected with at least 10 reads (grey dots) and RPG mRNAs (red dots). The number of mRNAs analyzed for each group and the Spearman rho is indicated on the Figure panel. **B.** Scatter plots showing the correlation between decay (top row) and deadenylation coefficients obtained from the preliminary data analysis (see Appendix main text). The classes of mRNAs analyzed are the same as in Appendix Fig. S4A. The good correlation between replicates justified merging the replicates (Appending Fig. S4C and main text). **C.** Scheme compares two strategies for calculation of decay and quantile deadenylation rates. The top version shows the classical approach where the unique control sample is used to estimate the decay coefficient in relation to all the data-points in the replicate. In case of the Mex67-depletion this introduces a bias that is due to the time required for the phenotype onset to be fully established. This exact time was calculated when the modified gamma model was built and defined to be equal to 2.67 min. This allowed to shift the control sample on the time-line and to treat every time-point in the chase experiment as a control for the consecutive sample. This approach greatly increases the number of coefficients that can be used to estimate the decay and quantile deadenylation rates. In the latter case this number is further increased by 5-fold as each quantile deadenylation coefficient is estimated from five quantiles: 75th, 80th, 85th, 90th and 95th. The accuracy of the calculated decay and deadenylation coefficients was further tested by comparing mRNA half-lives to published datasets (Figure 1) and by evaluating the compliance of the functions drawn using the quantile deadenylation coefficients compared to actual experimental data (EV Figure 5A). **D.** Density plots compare the ratio between the standard deviation and its corresponding decay or quantile deadenylation coefficient calculated using the bulk method shown in Appendix Figure S4C (bottom version; see also main text). The number of measurments is the same as in Appendix Figure S4B.

5. R code for calculation of the modified gamma distribution parameters

```
#this code should calculate the modified gamma distribution gamma_shape and gamm_rate parameters. Remember that estimation of the modified gamma distribution parameters is most accurate on distributions containing a large number of reads.


#download dataframes and prepare a separate directory; then set the directory path in setwd(). The end result table and pdf files containing figues will be downloaded in this directory.


#note that this code works with the Mex67-chase dataset dataframes as it extracts the time point from the column header using the colnames() command. If you want to use it on other dataframes, then either adjust the column name or change the colnames() commands accordingly.
# 0.096 parameter (beta = JP_DECAP) should be adjusted to the given datasets and remain equal for all distributions.



setwd("set chosen directory")

getwd()# will check the chosen directory location



JP_DECAP<-0.096

JP_dist<-function(x,decap,shape,rate) {

  return(((tanh(x*decap))^shape)*exp(-rate*x))

}



JP_RES<-NULL

JP_LogN<-NULL

JP_Log2<-NULL


now<-Sys.time()

sname<-paste0("JP4AT-",format(Sys.time(),format="%y%m%d-%H%M"))

pdf(file=paste0(sname,".pdf"))

nn<-60  # number of curves produced for the simulation step

ll<-80  # fixed parameter describing the maximum pA-tail length; do not modify

zz<-80

pp<-0.5 # fixed parameter not to be changed

tt<-60  # maximum time in minutes

scale<-1 # curve density for the simulation step; - scale * nn is the maximum in silico evolution time

fname<-"ORFs_repB" #file name of one of the data frames



for (fname in c("name of dataframe in chosen directory")) {



  all_raw<-read.csv(paste0(fname,".csv"))

  all<-all_raw[order(all_raw[,1]),]

  all<-NULL

  ii<-0

  for(ii in 0:(ll+1)) {
```

```
  ss<-all_raw[all_raw[,1]==ii,]

  if (length(ss[,1])==0) {

    ss<-all_raw[1,]

    ss[1,1]<-ii

    ss[1,-1]<-0

  }

  ss<-gsub("NA",0,ss)

  ss<-as.numeric(ss)

  all<-rbind(all,as.numeric(ss))

}


colnames(all)<-gsub("_","",gsub("25C","",gsub("min","",gsub(fname,"",colnames(all_raw)))))

colnames(all)<-gsub("min","",gsub("Mex67AA","",colnames(all)))

#rownames(all)<-all[,1]

all<-as.data.frame(all)


#all[,1]<-as.numeric(0:(ll+1))

for(ii in 2:7) {

  #  all[,ii]<-as.numeric(all[,ii])

  all[,ii]<-all[,ii]/max(all[,ii],na.rm = TRUE)

}


dist<-all[1:(ll+2),1:2]

names(dist)<-c("n","o")

#dist$n<-0:(ll+1)

#dist$o<-0


dist$o[is.na(dist$o)]<-0

dist$o<-dist$o/max(dist$o)

#dist$o<-dist$o*0

#dist$o[70]<-1


old<-dist$o

old2<-dist$o


plot(dist$o,type="n",col="blue",log="",xlim=c(0,100),xlab = "N-Tail",

     ylab = " experimental PolyA distribution",lwd=2,main=fname)
```

```
for ( ii in 2:7) {
  cc<-as.double(colnames(all)[ii])/tt
  lines(all[,ii],type="l",col=rgb(1-cc,0,cc),lwd=2)
}

plot(dist$o,type="n",col="blue",log="xy",xlab = "N-Tail",xlim=c(1,200),
     ylab = "normalized experimental PolyA distribution",lwd=2,main=fname)
for ( ii in 2:7) {
  cc<-as.double(colnames(all)[ii])/tt
  lines(all[,ii],type="l",col=rgb(1-cc,0,cc),lwd=2)
}

plot(dist$o,type="l",col="red",log="xy",xlab = "N-Tail",ylab = "modelled PolyA distribution",lwd=2,main=fname)


wcap<-(0:(ll+1))*JP_DECAP
wcap<-(tanh(wcap))^pp


beta<-0
for (beta in 0:(nn-1)) {
  new<-dist$o*0
  for (ii in 1:(ll+1)) {
    for (jj in (ll+2-ii):1) {
      new[ii]<-(new[ii]+dist$o[ii+jj])*beta*scale/jj
    }
    new[ii]<-new[ii]+dist$o[ii]
  }
  new<-new*exp(-beta*scale)
  old<-cbind(old,new)
  colnames(old)[beta+2]<-paste0("b",beta*scale)


  wcap<-tanh((0:(ll+1))*JP_DECAP)
  wcap<-wcap^((beta*scale)^pp)
  old2<-cbind(old2,new*wcap)
  colnames(old2)[beta+2]<-paste0("b",beta*scale)


  #lines(new,type="l",col="green")
  lines(new*wcap,type="l",col=rgb(0,beta/nn,1-beta/nn))
```

```
}

write.table(old,file=paste0(fname,"_models.txt"),sep="\t")

plot(old[,1],col="blue",type="n",xlim=c(1,100),lwd=4,log="",xlab = "N-Tail",
     ylab = "modelled PolyA distribution",main=fname)
for(ii in 2:length(colnames(old))) {
  lines(old[,ii],col=rgb(0,(ii-1)/nn,1-(ii-1)/nn),type="l")
}
lines(old[,1],col="red",type="l",lwd=2)

plot(old[,1],col="blue",type="n",xlim=c(1,100),lwd=4,log="",xlab = "N-Tail",
     ylab = "normalized modelled PolyA distribution",main=fname)
for(ii in 2:length(colnames(old))) {
  #  old[,ii]<-old[,ii]/max(old[,ii])
  lines(old[,ii]/max(old[,ii]),col=rgb(0,(ii-1)/nn,1-(ii-1)/nn),type="l")
}
lines(old[,1],col="red",type="l",lwd=3)

#old2<-old*wcap
#old2[,1]<-old[,1]

plot(old2[,1],col="blue",type="n",xlim=c(1,100),lwd=4,log="",xlab = "N-Tail",
     ylab = "modelled PolyA distribution",main=fname)
for(ii in 2:length(colnames(old2))) {
  lines(old2[,ii],col=rgb(0,(ii-1)/nn,1-(ii-1)/nn),type="l")
}
lines(old2[,1],col="red",type="l",lwd=3)

plot(old2[,1],col="blue",type="n",xlim=c(1,100),lwd=4,log="",xlab = "N-Tail",
     ylab = "normalized modelled PolyA distribution",main=fname)
for(ii in 2:length(colnames(old))) {
  #  old2[,ii]<-old2[,ii]/max(old2[,ii])
  lines(old2[,ii]/max(old2[,ii]),col=rgb(0,(ii-1)/nn,1-(ii-1)/nn),type="l")
}
lines(old2[,1],col="red",type="l",lwd=3)
```

```
plot(dist$o,type="n",col="blue",log="",xlim=c(0,100),xlab = "N-Tail",
     ylab = "normalized experimental PolyA distribution",lwd=2,main=fname)
for ( ii in 2:7) {
  cc<-as.double(colnames(all)[ii])/tt
  lines(all[,ii],type="l",col=rgb(0,1-cc,cc),lwd=2)
}

plot(dist$o,type="n",col="blue",log="",xlim=c(0,100),xlab = "N-Tail",
     ylab = "normalized experimental PolyA distribution",lwd=2,main=fname)
for ( ii in 2:7) {
  cc<-as.double(colnames(all)[ii])/tt
  lines(all[,ii],type="l",col=rgb(0,1-cc,cc),lwd=2)
}
for(ii in 2:length(colnames(old))) {
  lines(old2[,ii]/max(old2[,ii]),col=rgb(0,(ii-1)/nn,1-(ii-1)/nn),type="l")
}


for ( ii in 2:7) {

  ###############################################
  #Glowne dopasowanie
  ###############################################

  cc<-as.double(colnames(all)[ii])/tt
  max<-9e99
  kk<-1
  for(jj in 2:length(colnames(old))) {
    chi2<-old2[,jj]/max(old2[,jj])-all[1:(ll+2),ii]
    chi2<-chi2*chi2
    chi2<-sum(chi2*chi2)
    if (chi2<max) {
      max<-chi2;
      kk<-jj
      #      print(c(ii,jj,chi2,"\n"))
```

```
  }

}



plot(all[,ii],type="p",pch=19,col=rgb(0,1-cc,cc),log="",xlim=c(0,100),xlab = "N-Tail",

    ylab = "normalized experimental PolyA distribution",lwd=2,main=paste0(fname," min ",cc*tt,"; in silico ",kk-1))

lines(old2[,kk]/max(old2[,kk]),col=rgb(0,(kk-1)/nn,1-(kk-1)/nn),type="l",lw=2)



dd<-all[,c(1,ii)]

dd[,1]<-as.numeric(dd[,1])

dd[,2]<-as.numeric(dd[,2])

dd<-dd[dd[,2]>0,]

dd<-dd[dd[,1]<=zz,]

x<-as.numeric(dd[,1])

y<-as.numeric(dd[,2])


rate<-0.05

shape=3



fit0<-nls(y ~ a*JP_dist(x,JP_DECAP,shape,rate),start=list(a=8e3),control=nls.control(maxiter=500))

co<-coefficients(fit0)

lines(x,predict(fit0),lw=0.5)



fit1<-nls(y ~ a*JP_dist(x,JP_DECAP,shape,rate),start=list(a=co[1],rate=rate),control=nls.control(maxiter=500))

co<-coefficients(fit1)

lines(x,predict(fit1),lw=0.5)



fit2<-nls(y ~ a*JP_dist(x,JP_DECAP,shape,rate),start=list(a=co[1],shape=shape,rate=co[2]),control=nls.control(maxiter=500))

co<-coef(fit2)

lines(x,predict(fit2),lw=1)



fit3<-nls(y ~ a*JP_dist(x,JP_DECAP,shape,rate),start=list(a=co[1],shape=co[2],rate=co[3]),control=nls.control(maxiter=500),weights = 1/(1+y))

co<-coef(fit3)

lines(x,predict(fit3),lw=3,col="black")



ssss<-summary(fit3)

JP_RES<-rbind(JP_RES,c(fname,cc*tt,co,ssss$parameters[2,],ssss$parameters[3,]))
```

```
if (as.numeric(colnames(all)[ii])<15) {

  mean<-3.75
  width<-0.31

  plot(all[,ii],type="p",pch=19,col=rgb(0,1-cc,cc),log="",xlim=c(0,100),xlab = "N-Tail",
       ylab = "normalized experimental PolyA distribution with logN",lwd=2,main=paste0(fname," min ",cc*tt,"; in silico ",kk-1))
  #    lines(old2[,kk]/max(old2[,kk]),col=rgb(0,(kk-1)/nn,1-(kk-1)/nn),type="l",lw=2)

  ww<-x
  aa<-(ww>60)+(ww<20)

  fit4<-nls(y ~ a*JP_dist(x,JP_DECAP,shape,rate),
            start=list(a=co[1],shape=co[2],rate=co[3]),control=nls.control(maxiter=500),weights = aa/(1+y))
  coss<-coef(fit4)
  lines(x,predict(fit4),lw=0.1,col="black")

  shape=coss[2]
  rate<-coss[3]

  fit5<-nls(y ~ a*JP_dist(x,JP_DECAP,shape,rate)+b*dlnorm(x,mean,width),
            start=list(a=coss[1],b=1),control=nls.control(maxiter=500),weights = 1/(1+y))
  coss<-coef(fit5)
  lines(x,predict(fit5),lw=0.1,col="black")

  fit6<-nls(y ~ a*JP_dist(x,JP_DECAP,shape,rate)+b*dlnorm(x,mean,width),
            start=list(a=coss[1],shape=shape,rate=rate,b=coss[2]),control=nls.control(maxiter=500),weights = 1/(1+y))
  coss<-coef(fit6)
  lines(x,predict(fit6),lw=.5,col="black")

  shape<-coss[2]
  rate<- coss[3]

  ok<-tryCatch({
    fit7<-nls(y ~ abs(a)*JP_dist(x,JP_DECAP,shape,rate)+abs(b)*dlnorm(x,mean,width),
              start=list(a=coss[1],b=coss[4],mean=mean,width=width),control=nls.control(maxiter=500),weights = 1/(1+y))
    coss<-coef(fit7)
    lines(x,predict(fit7),lw=1,col="black")
```

```
},error=function(cond) {
  lines(x,predict(fit6),lw=.1,col="black")
  print(c("7",as.numeric(colnames(all)[ii]),fname))
}
)

ok<-tryCatch({
  fit8<-nls(y ~ abs(a)*JP_dist(x,JP_DECAP,shape,rate)+abs(b)*dlnorm(x,mean,width),
            start=list(a=coss[1],shape=shape,rate=rate,b=coss[2],mean=coss[3],width=coss[4]),control=nls.control(maxiter=500),weights = 1/(1+y))

  coss<-coef(fit8)
  lines(x,predict(fit8),lw=3,col="red")
  lines(x,coss[4]*dlnorm(x,coss[5],coss[6]),col="blue",lw=2,lty=1)
  lines(x,coss[1]*JP_dist(x,JP_DECAP,coss[2],coss[3]),col="green",lw=2,lty=1)
  ssss<-summary(fit8)
  JP_LogN<-rbind(JP_LogN,c(fname,cc*tt,coss,ssss$parameters[1,],ssss$parameters[2,],ssss$parameters[3,],ssss$parameters[4,],ssss$parameters[5,],ssss$parameters[6,]))
},error=function(cond) {
  #	lines(x,predict(fit7),lw=.1,col="black")
  print(c("8",as.numeric(colnames(all)[ii]),fname))
}
)

ok<-tryCatch({
  mean<-3.72
  width<-0.284

  fit9<-nls(y ~ abs(a)*JP_dist(x,JP_DECAP,shape,rate)+abs(b)*dlnorm(x,mean,width),
            start=list(a=coss[1],shape=coss[2],rate=coss[3],b=coss[4]),control=nls.control(maxiter=5000),weights = 1/(1+y))

  coss<-coef(fit9)
  ssss<-summary(fit9)
  JP_Log2<-rbind(JP_Log2,c(fname,cc*tt,coss,ssss$parameters[1,],ssss$parameters[2,],ssss$parameters[3,],ssss$parameters[4,]))

  lines(x,predict(fit9),lw=3,col="red",lty=2)
  lines(x,coss[4]*dlnorm(x,mean,width),col="blue",lw=2,lty=2)
  lines(x,coss[1]*JP_dist(x,JP_DECAP,coss[2],coss[3]),col="green",lw=2,lty=2)
```

```
    },error=function(cond) {

      print(c("9",as.numeric(colnames(all)[ii]),fname))

    }

    )


    lines(x,predict(fit3),lw=3,col="black")

  }


}


#lines(old2[,22]/max(old2[,22]),type="l",col=rgb(1,0,0),lwd=3,lty=4)

#lines(old2[,18]/max(old2[,18]),type="l",col=rgb(1,0,0),lwd=3,lty=4)

#lines(old2[,15]/max(old2[,15]),type="l",col=rgb(1,0,0),lwd=3,lty=4)

#lines(old2[,4]/max(old2[,4]),type="l",col=rgb(1,0,0),lwd=3,lty=4)

#lines(old2[,2]/max(old2[,2]),type="l",col=rgb(1,0,0),lwd=3,lty=4)


dall<-all[,1]

dall<-cbind(dall,all[,2]-all[,3])

dall<-cbind(dall,all[,2]-all[,4])

dall<-cbind(dall,all[,2]-all[,5])


dall[,2]<-dall[,2]/max(dall[,2],na.rm = TRUE)

dall[,3]<-dall[,3]/max(dall[,3],na.rm = TRUE)

dall[,4]<-dall[,4]/max(dall[,4],na.rm = TRUE)


plot(dall[,2],type="n",col="blue",xlim=c(0,80),xlab = "N-Tail",

     ylab = "normalized difference in experimental PolyA distributions",lwd=2,main=fname)

lines(dall[,2],type="p",pch=19)


x<-dall[,1]

y<-dall[,2]

w<-0+(x>35)

fitex<-nls(y~a*dlnorm(x,mean,width),start=c(a=10,mean=3.8,width=0.1),weights = w)

res<-summary(fitex)

lines(predict(fitex),col="blue",lw=2)

#lines(dall[,4],type="p")

#lines(all[,1],10*dlnorm(all[,1],meanlog=3.8,.1),col="green")
```

```
  write.table(dall,file = paste0(fname,"diff.txt"),sep="\t")

  syn<-coef(fitex)


}

dev.off()

write.table(JP_RES,paste0(sname,"JP_RES.txt"),sep="\t")

write.table(JP_LogN,paste0(sname,"JP_LogN.txt"),sep="\t")

write.table(JP_Log2,paste0(sname,"JP_Log2.txt"),sep="\t")
```

3.2 Manuscript 2: TENT5-mediated polyadenylation of mRNAs encoding secreted proteins is essential for gametogenesis in mice

3.2.1 Description

It has been suggested that in mammals, TENT2 (Gld2) plays a key role in the regulation of gametogenesis. However, TENT2 knockout animals are fertile and do not display any developmental phenotypes. This study's aim was to determine the roles of other non-canonical polymerases in oogenesis and spermatogenesis. Various mouse lines with knockouts (KO) and knock-ins (KI) of non-canonical cytoplasmic poly(A) polymerases of TENT5 family were analyzed in the context of oogenesis and spermatogenesis. The analyses included the following mouse lines: *Tent5a* KO, *Tent5b* KO, *Tent5c* KO, *Tent5d* KO, *Tent5b* and *Tent5c* double KO (dKO) as well as *GFP-Tent5b*, and *Tent5b-GFP* knock-ins. Female mice with *Tent5a* KO, *Tent5b/c* dKO, (*Tent5b*$^{+/+}$/*Tent5c*$^{+/+}$) or *Tent5b-GFP* (*Tent5b*$^{wt/gfp}$ and *Tent5b*$^{gfp/gfp}$) mutations were infertile, whereas in males, this phenotype was observed for *Tent5c* KO (*Tent5c*$^{-/-}$) and *Tent5d* KO (*Tent5d*$^{-/null}$). Then, the specific stages of spermatogenesis and oogenesis where defects arise due to the absence of individual non-canonical poly(A) polymerases were determined. Subsequently, Nanopore DRS sequencing for *Tent5b/c* lines isolated from ovaries and *Tent5c* and *Tent5d* lines from testes was performed. The global distribution of poly(A) tail lengths did not reveal significant differences among the mutants (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). Consequently, differential poly(A) tail length analyses were conducted, which allowed the identification of potential targets of TENT5 proteins. These transcripts were found to be highly expressed during oogenesis. Similarly, data from male Tent5c and Tent5d mutants were processed, and for selected TENT5 targets, experiments to determine their localization in germ cells were conducted. The results, supported by GO-term analysis of potential TENT5 targets, were consistent with previous observations made for TENT5A targets (Gewartowska et al., 2021; Liudkovska et al., 2022). Proteins encoded by mRNAs undergoing cytoplasmic polyadenylation mediated by TENT5B, TENT5cC, or TENT5D were found to localize in the endoplasmic reticulum or to be secreted.

The putative role of CPEB proteins (Duran-Arqué et al., 2022) in regulating the expression of potential TENT5 targets was then investigated. An analysis of the 3′ UTR sequences of transcripts identified by DRS did not show enrichment for motifs recognized by CPEBs (mice Cytoplasmic Polyadenylation Element Binding Proteins). As no other common motifs were found in the UTRs or coding sequences (CDS) of potential TENT5 targets, we excluded the involvement of known specificity factors for cytoplasmic polyadenylation. An analysis of the protein sequences of potential TENT5 targets revealed that most of them possess an endoplasmic reticulum (ER)-targeting signal peptide (SP), confirming their localization. These findings demonstrate that TENT5-mediated polyadenylation of mRNAs encoding secreted proteins is essential for proper spermatogenesis and oogenesis in mice.

3.2.2 My contribution

In this manuscript, in which I am the second author, I was responsible for the bioinformatic analysis of Illumina RNA-seq and Nanopore DRS datasets. I conducted differential expression and differential polyadenylation analyses, identified potential TENT5 targets and examined these in terms of motif enrichment in 3′ UTRs, CPEB motif identification and signal peptide presence in protein sequences. Additionally, I performed GO-term analyses and characterized potential targets based on GC content and UTR/CDS lengths. I compiled, reviewed, and prepared figures for RNA sequencing experiments. In collaboration with Michał Brouze and Professor Andrzej Dziembowski, I contributed to writing and refining the final manuscript and supplemental materials.

3.2.3 Citation and manuscript text

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. 2024 Jun 22;15(1):5331. doi: 10.1038/s41467-024-49479-4. PMID: 38909026; PMCID: PMC11193744.

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

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*$^{-/\mathrm{null}}$), 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*$^{\mathrm{gfp/gfp}}$ 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).

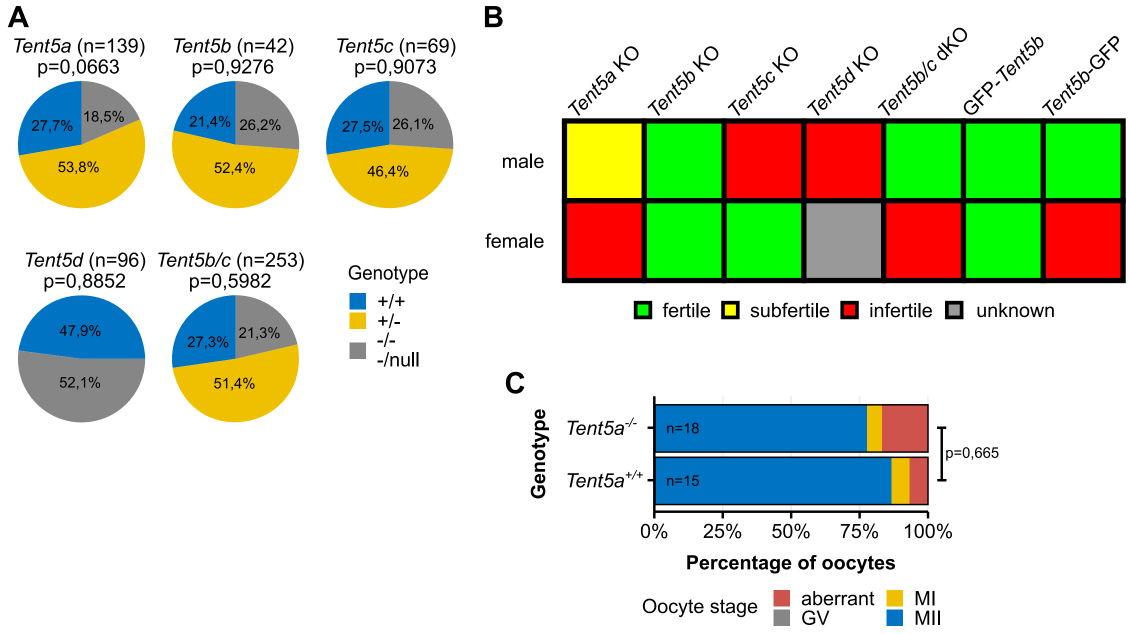


Fig. 1 | TENT5 poly(A) polymerases are involved in both spermatogenesis and oogenesis. **A** Distribution of genotypes in pups of all *Tent5* KO mouse lines from matings of heterozygous parents (only males were accounted for in the *Tent5d* KO line, in the *Tent5*b/c dKO mouse line, both parents had 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

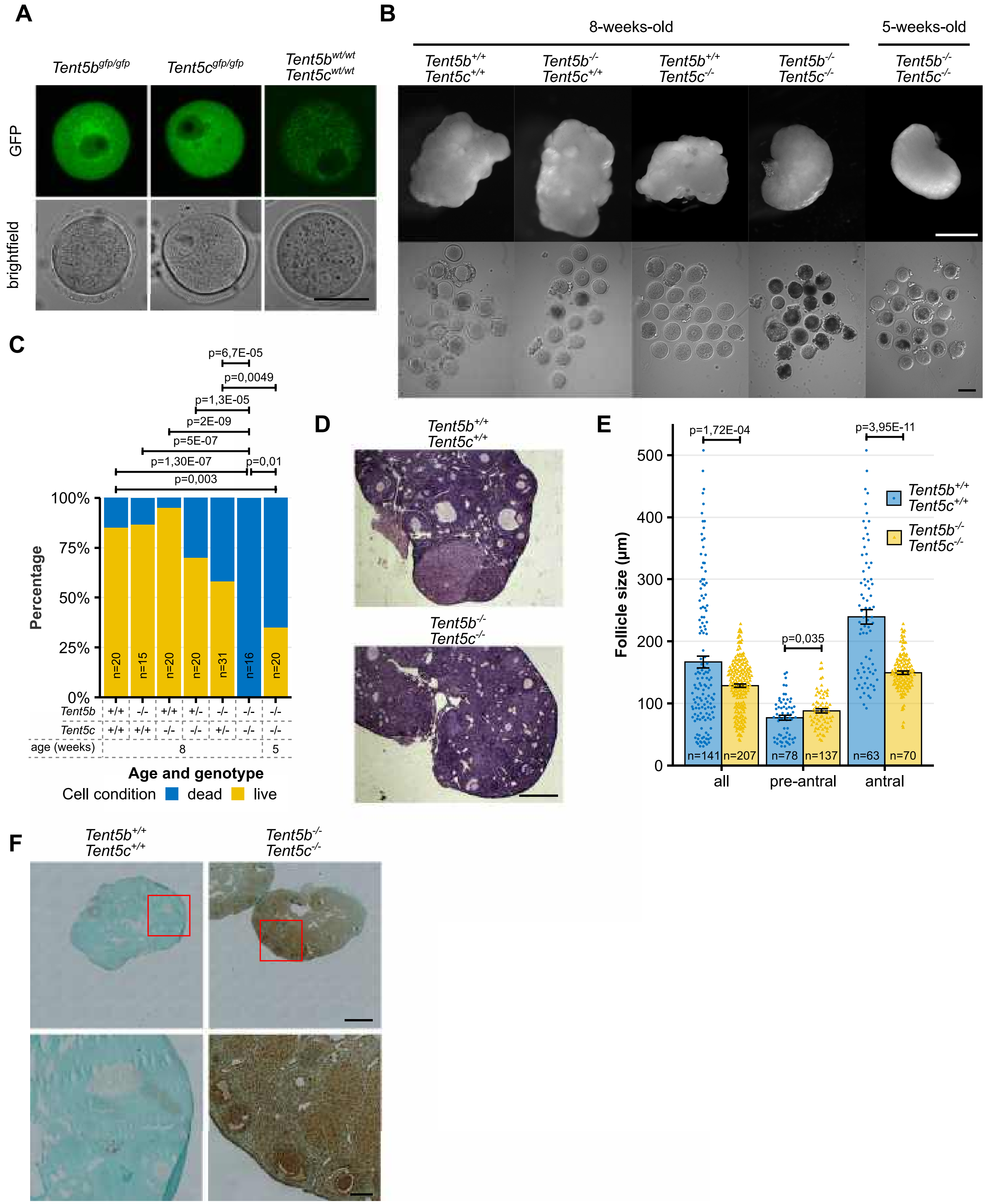
	Male	Female
Tent5a$^{-/-}$	Lowered fertility due to poor overall animal condition	Infertility due to poor overall animal condition
Tent5b$^{-/-}$	Fertile	Fertile
Tent5c$^{-/-}$	Loss of sperm head and DNA material; Fewer total germ cells produced	Fertile
Tent5d$^{-/null}$ or *Tent5d*$^{-/-}$	Abnormalities in the development of primary spermatocytes; Degeneration of testes tissue; 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; No ovulation; 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; 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



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).

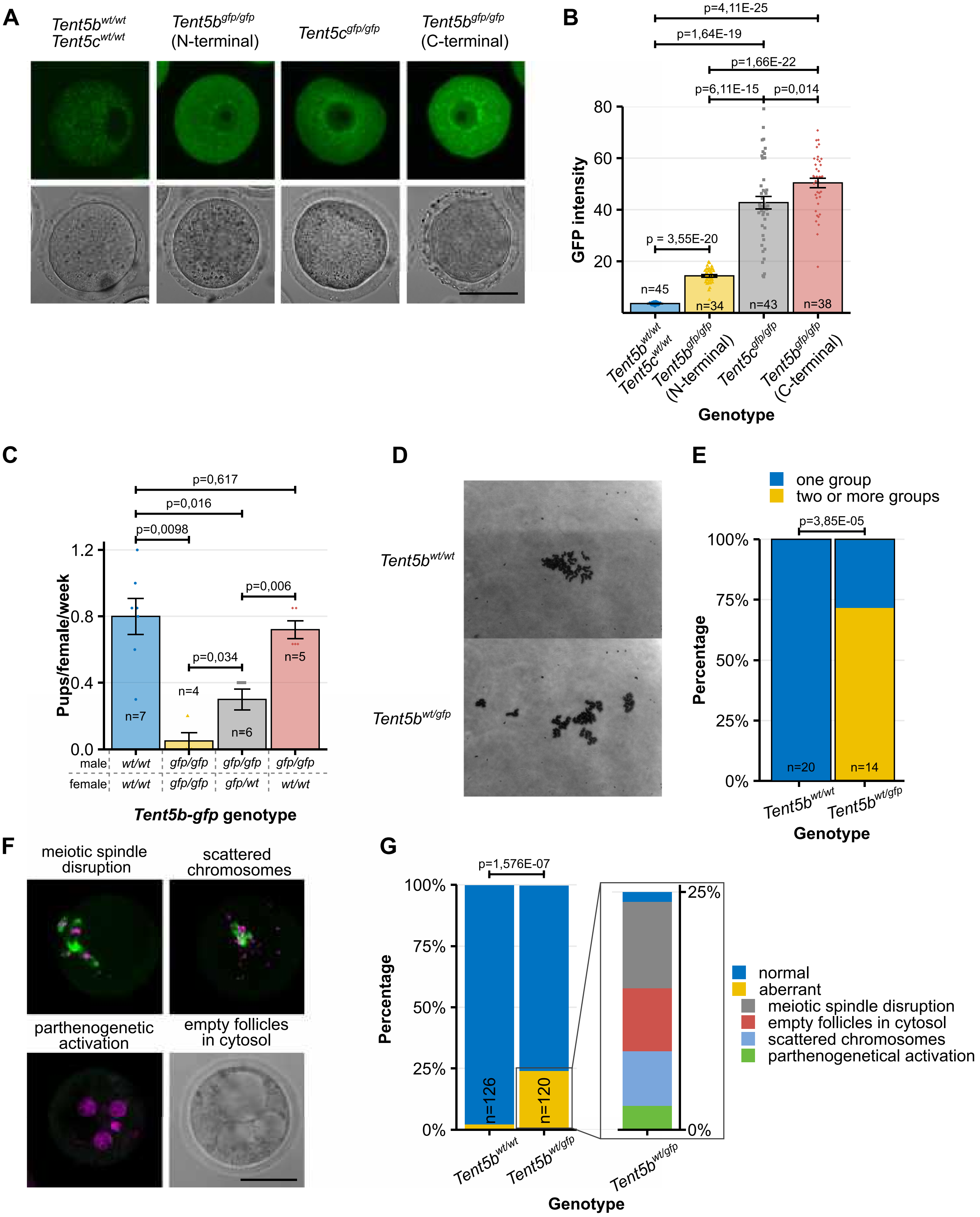
To specifically examine oocyte-enriched transcripts, we performed an RNA-seq analysis on oocytes isolated from 5-week-old WT and *Tent5b/c* dKO females using the Illumina platform (Fig. 4E). This allowed us to narrow down the number of transcripts to be analyzed to 522 highly expressed in oocytes (Supplementary 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 downregulation 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



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).

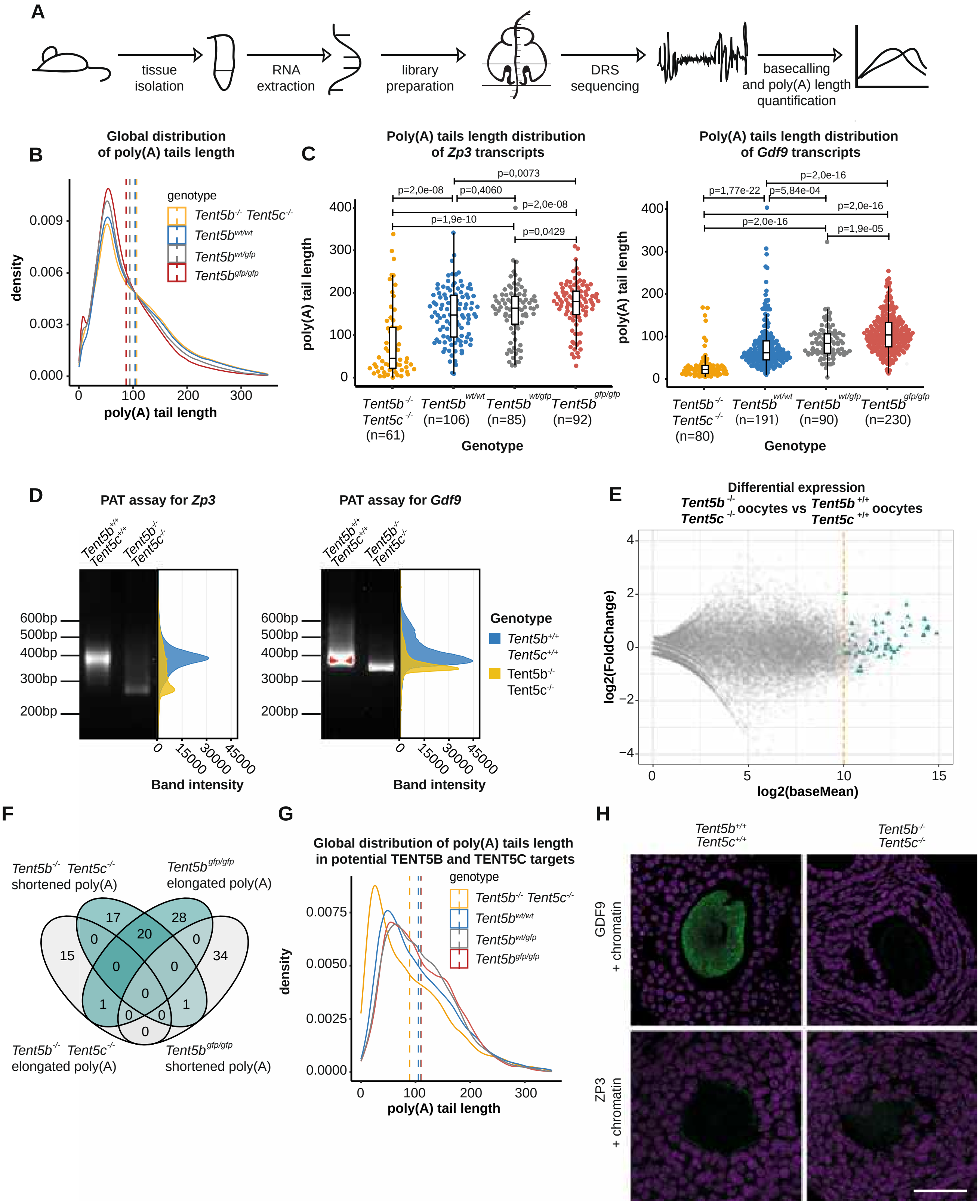
All results described above, as well as the differential localization of TENT5s, indicate that the phenotypes are mainly related to germ cells and that the substrates for TENT5C and TENT5D in testes may differ. To identify these substrates, we have performed DRS of mRNA isolated from the testes of adult *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)-



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.

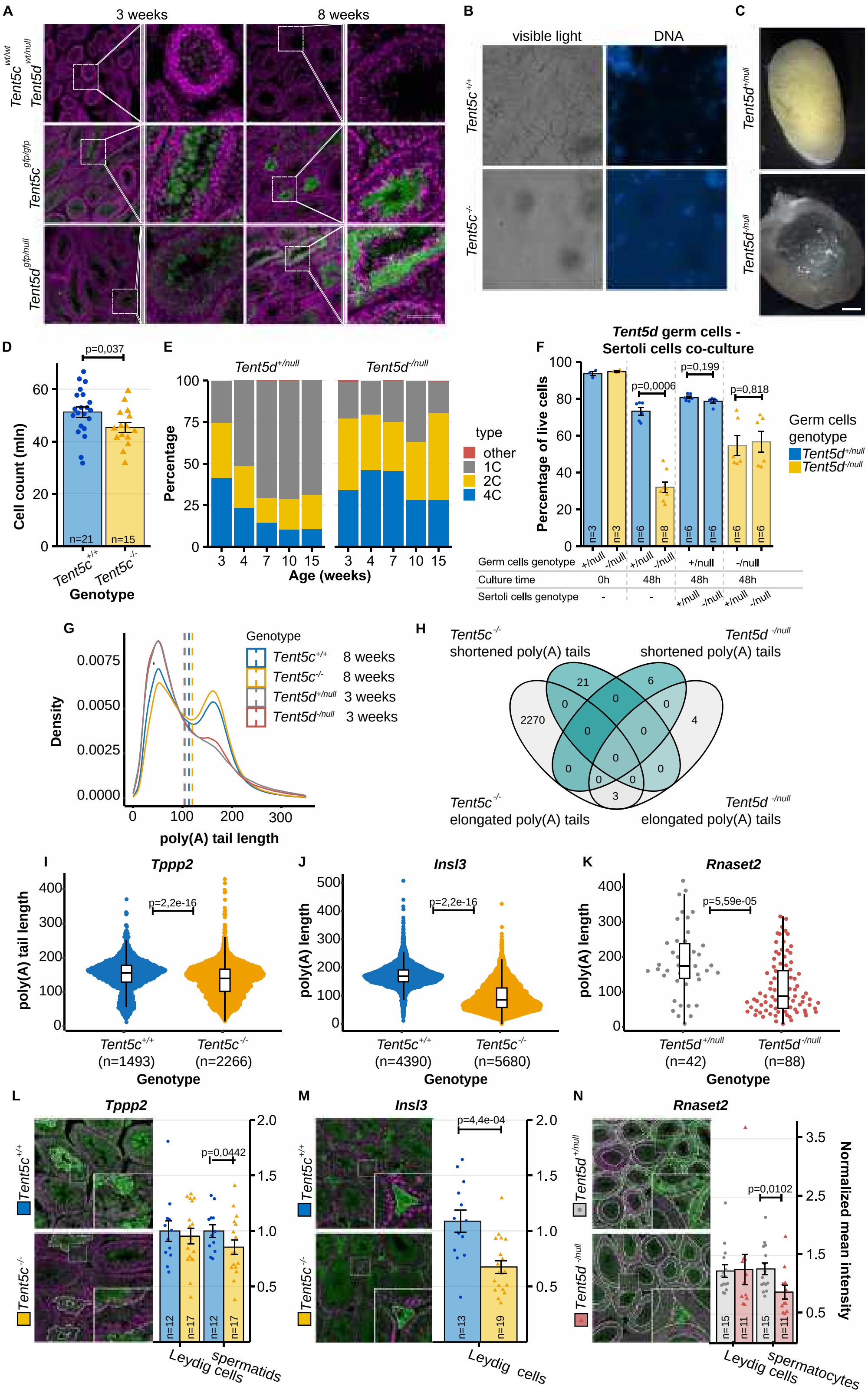
We observed that when reporter mRNAs encoded SP that recruits them to the ER, their polyadenylation rate was significantly lower in *Tent5b*$^{-/-}$ *Tent5c*$^{+/-}$ oocytes as compared to WT ones. We observed no such difference for reporter mRNAs encoding only YPET (without ER-targeting signal). Our 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].



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).

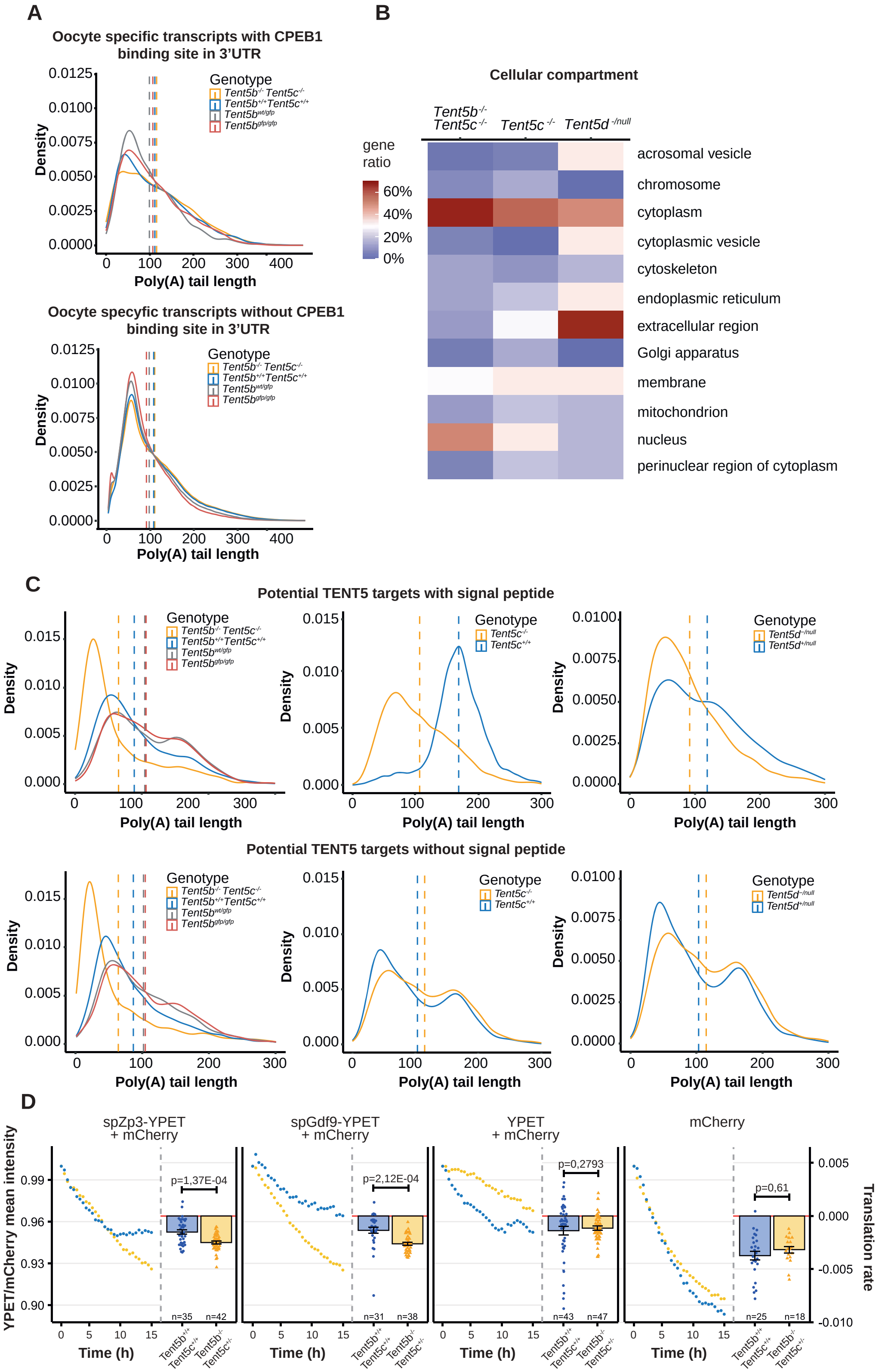
The mice were kept in conventional polystyrene cages in the animal facility of the Faculty of Biology, University of Warsaw. A 12/12 light cycle was maintained in the room, 15 air exchanges per hour. The relative humidity in the rooms was 55% ± 10%. The temperature in the rooms was 22 °C ± 2 °C. During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



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 1×10^6 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 2×10^6 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-Aldrich) 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#

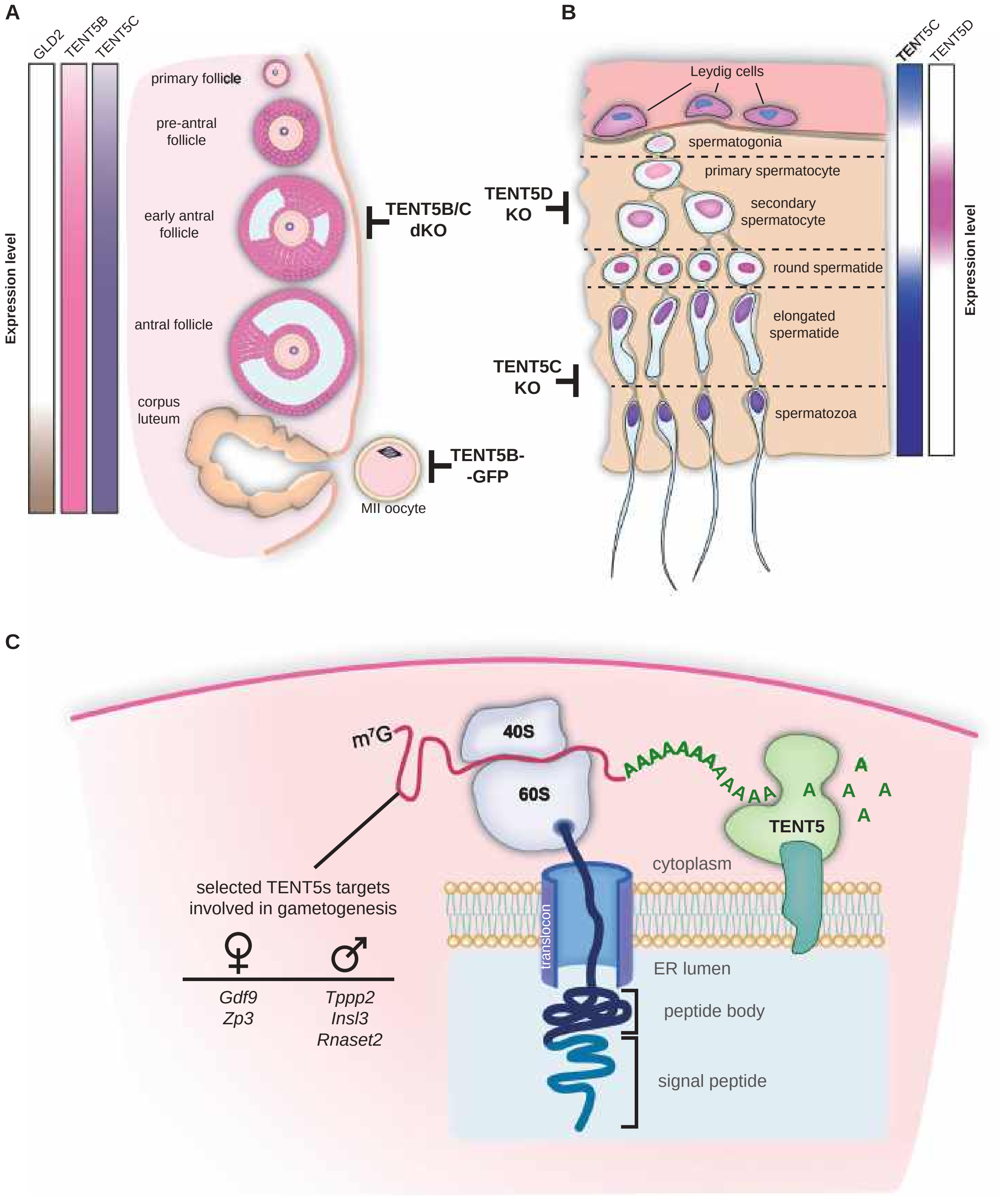


Fig. 7 | Proposed model for the activity of TENT5 proteins in gametogenesis. Simplified diagrams of oogenesis (**A**) and spermatogenesis (**B**). Black t-lines with labels indicate stages disrupted by a particular mutation. Gradient boxes indicate approximate expression of *Gld2* and *Tent5b, c*, and *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 distilled 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 Fischer 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

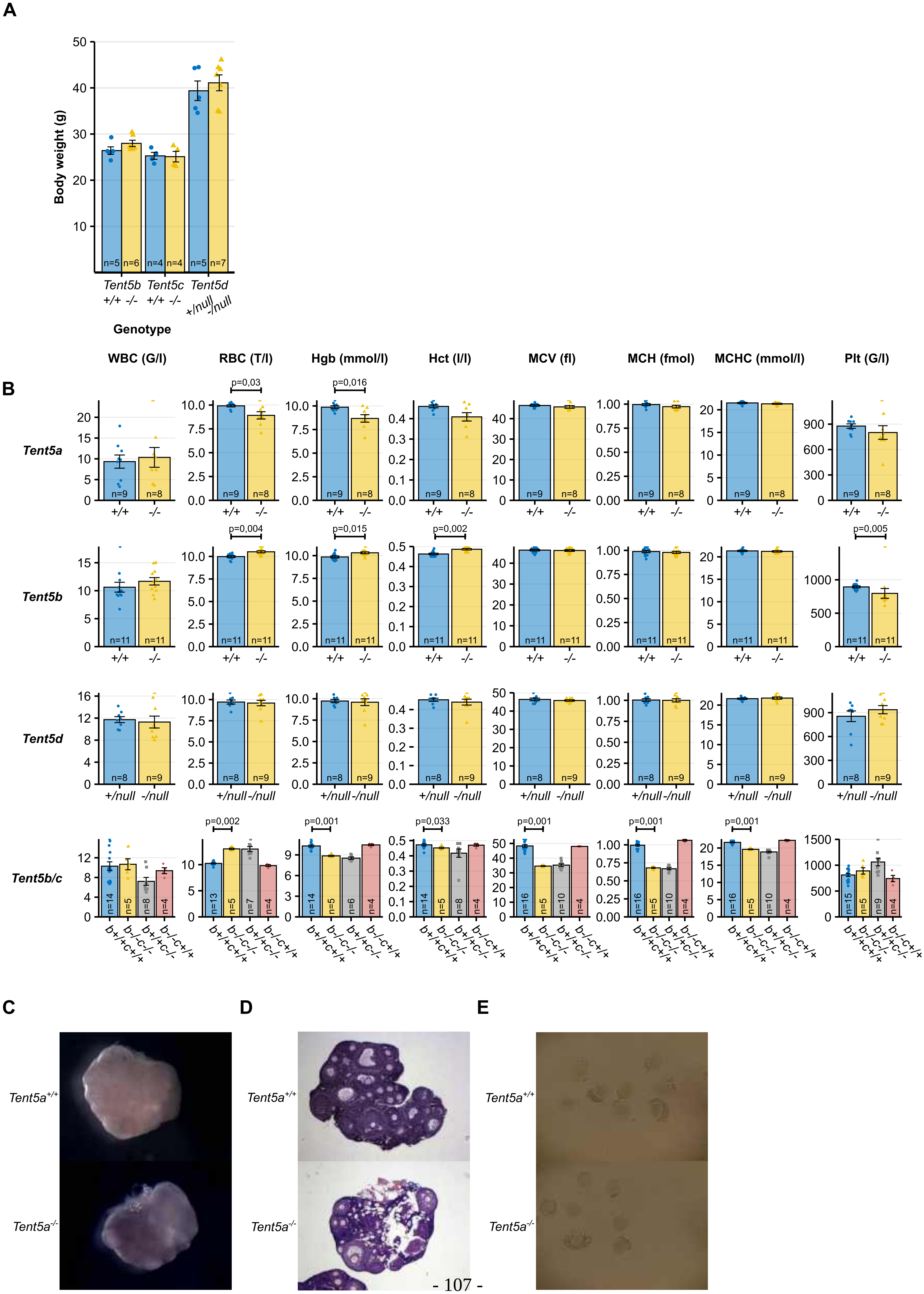
The authors declare no competing interests.

Additional information

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

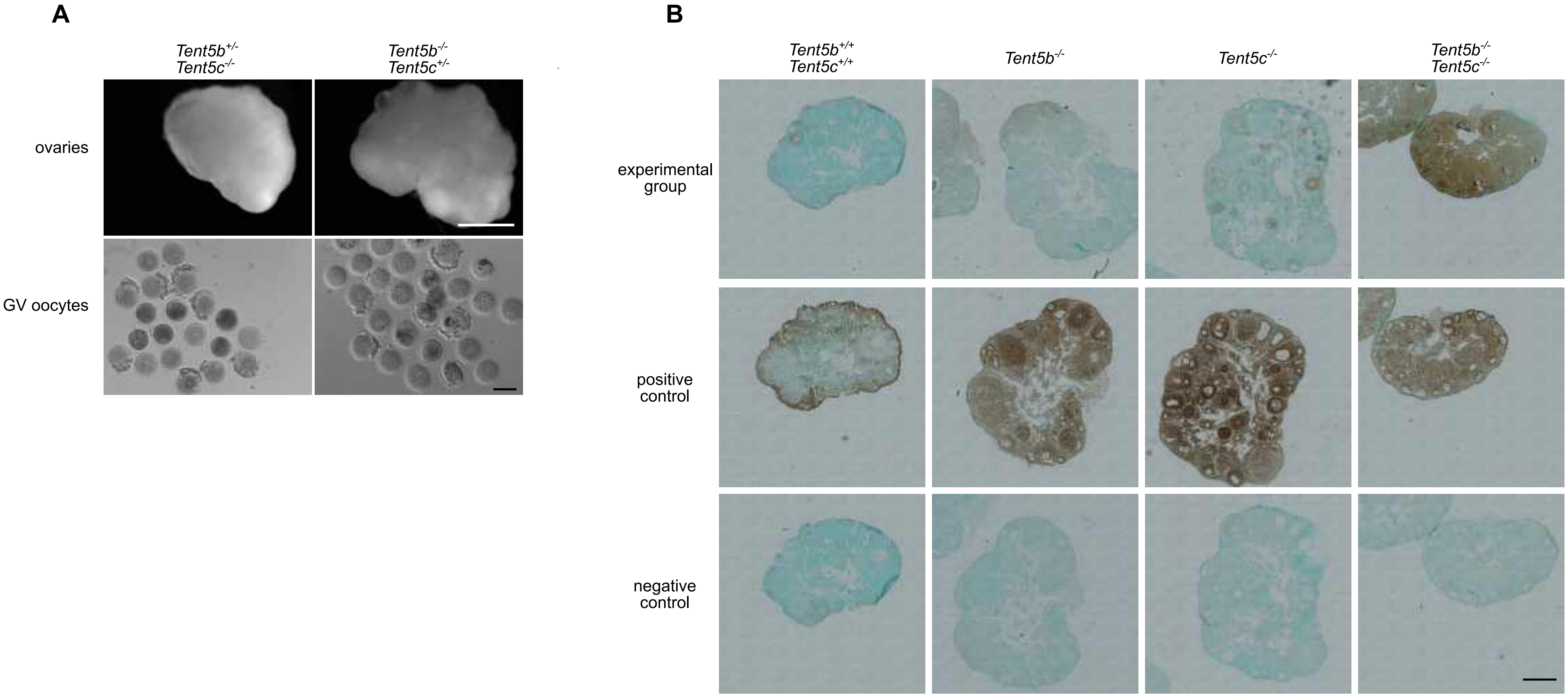
Correspondence and requests for materials should be addressed to Andrzej Dziembowski.

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.


A
Body weight (g)
50
40
30
20
10
n=5
n=6
n=4
n=4
n=5
n=7
Tent5b
+/+ -/-
Tent5c
+/+ -/-
Tent5d
+/null
-/null
Genotype
B
WBC (G/l)
RBC (T/l)
Hgb (mmol/l)
Hct (l/l)
MCV (fl)
MCH (fmol)
MCHC (mmol/l)
Plt (G/l)
Tent5a
p=0,03
p=0,016
n=9
n=8
+/+
-/-
Tent5b
p=0,004
p=0,015
p=0,002
p=0,005
n=11
n=11
+/+
-/-
Tent5d
n=8
n=9
+/null
-/null
Tent5b/c
p=0,002
p=0,001
p=0,033
p=0,001
p=0,001
p=0,001
n=14
n=5
n=8
n=4
n=13
n=7
n=6
n=16
n=10
n=15
n=9
b+/+c+/+
b-/-c-/-
b+/+c-/-
b-/-c+/+
C
Tent5a+/+
Tent5a-/-
D
Tent5a+/+
Tent5a-/-
E
Tent5a+/+
Tent5a-/-

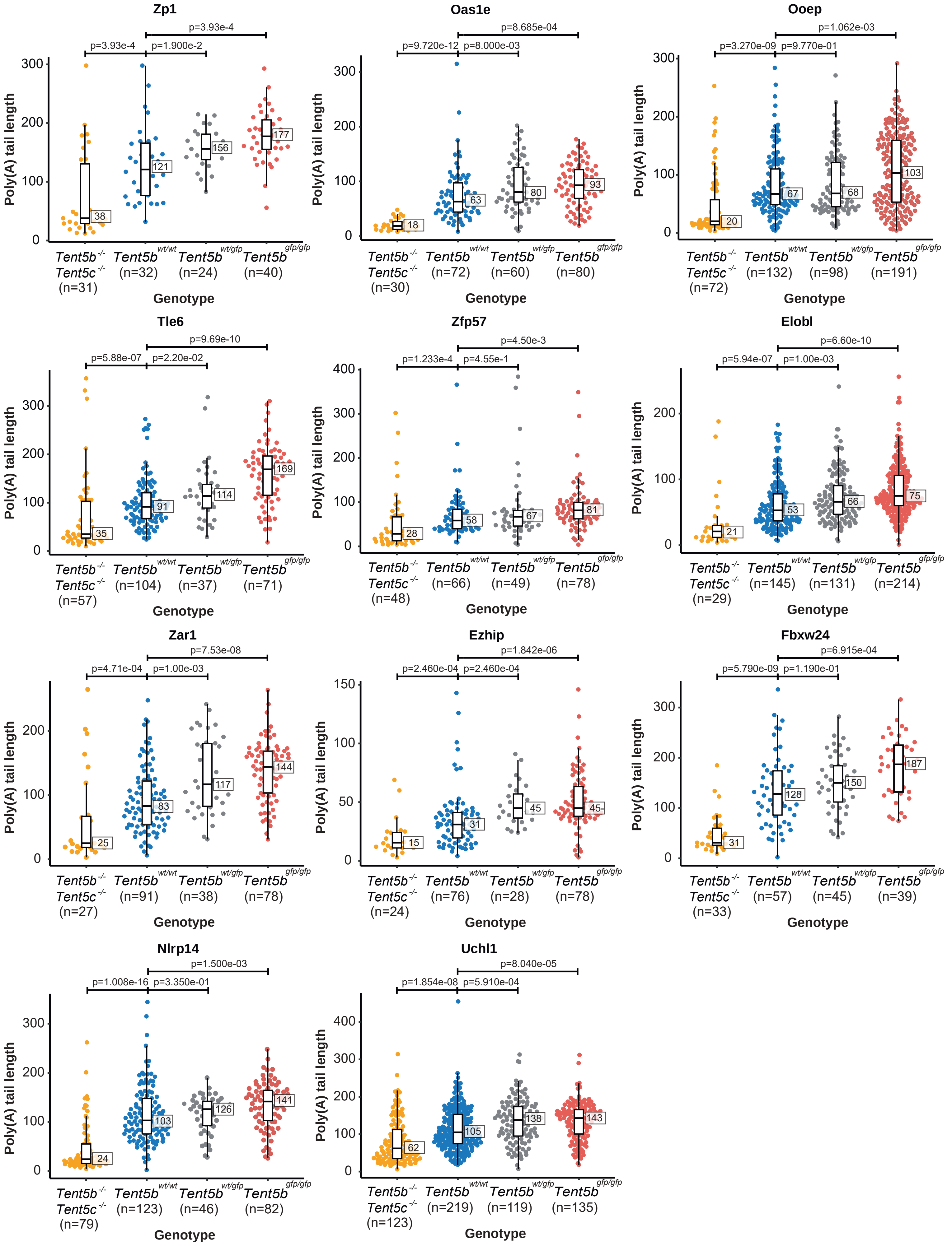
Supplementary Figure 1. Basic phenotype analysis of all *Tent5* KO mice lines.

A. Body weight of males of different *Tent5* KO mice lines. Animals of each mouse line were weighted at different age. Individual data points and n values represent individual animals weighted, bars represent mean values, error bars represent SEM, p-values reported for comparison in the Mann-Whitney-Wilcoxon test. **B.** Blood morphology parameters related to the *Tent5* mice lines genotype. Multiple differences in *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ females are related to a detrimental effect of *Tent5c*$^{-/-}$ mutation alone, reported previously by Mroczek *et al.*[27]. Individual data points and n values represent blood samples from individual animals, bars represent mean values, error bars represent SEM; p-values reported for mean parameter value comparison in Mann-Whitney-Wilcoxon test; WBC = white blood cells, RBC = red blood cells, Hgb = hemoglobin, Hct = hematocrit, MCV = mean corpuscular volume, MCH = mean cell hemoglobin, MCHC = mean corpuscular hemoglobin concentration, Plt = Platelet. **C-E.** *Tent5a*$^{-/-}$ females' ovary morphology (**C**), histology (**D**) and GV oocytes (**E**) compared to *Tent5a*$^{+/+}$ ones. No changes observed in follicle growth, ovulation, and oocyte condition at GV stage.


A
Tent5b+/- Tent5c-/-
Tent5b-/- Tent5c+/-
ovaries
GV oocytes
B
Tent5b+/+ Tent5c+/+
Tent5b-/-
Tent5c-/-
Tent5b-/- Tent5c-/-
experimental group
positive control
negative control

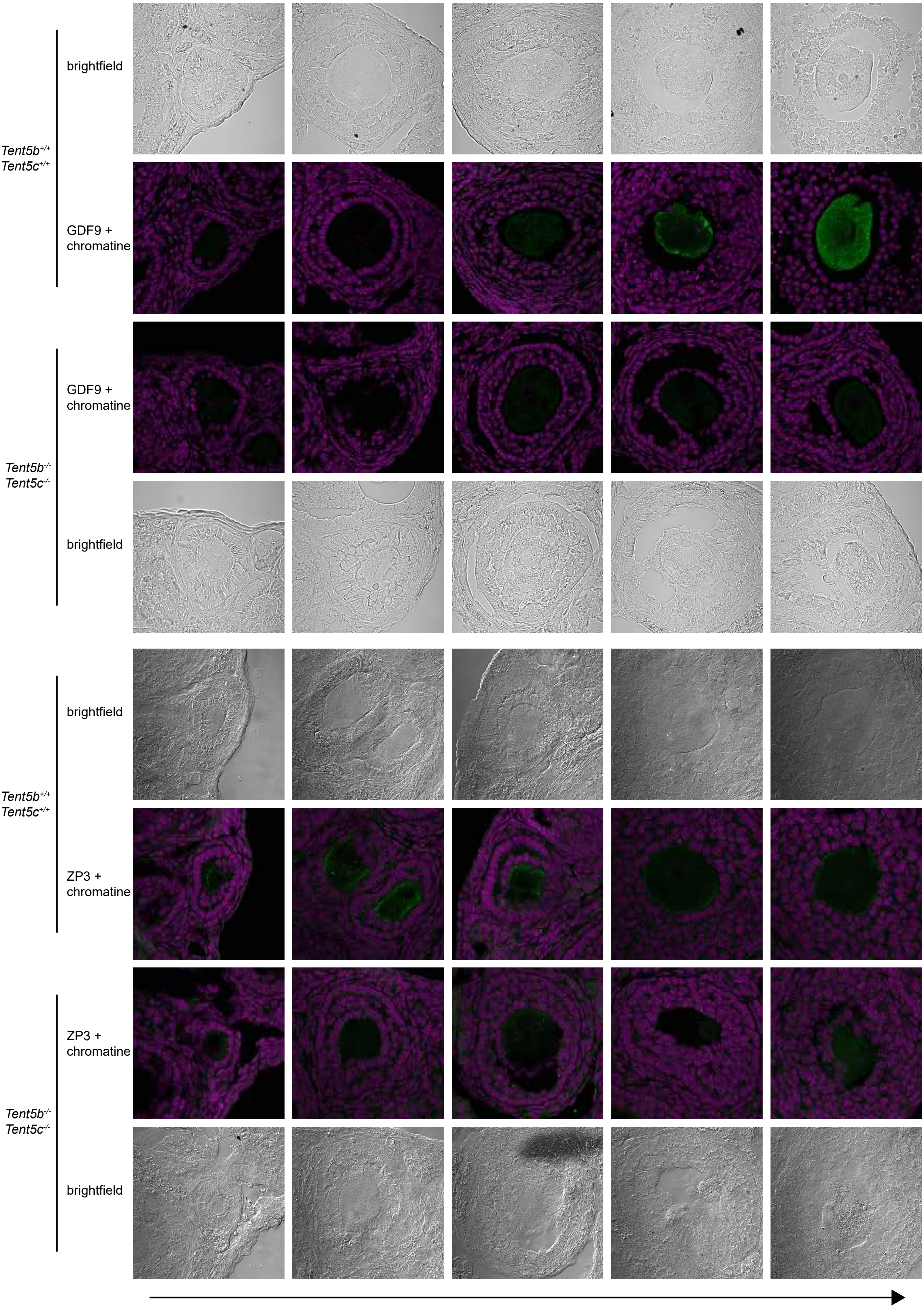
Supplementary Figure 2. *Tent5b/c* dKO ovaries and oocytes morphology and apoptosis analysis

A. Morphology of the ovaries and GV oocytes of females with one wild-type allele of either *Tent5b* or *Tent5c* gene, which is sufficient to maintain normal fertility. **B**. TUNEL assay for signs of apoptosis in cross-section of ovaries of different *Tent5b* and *Tent5c* genotypes. Only *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ mice display high levels of apoptosis in both oocytes and somatic granulosa cells surrounding them.


Zp1
p=3.93e-4
p=3.93e-4
p=1.900e-2
38
121
156
177
Oas1e
p=8.685e-04
p=9.720e-12
p=8.000e-03
18
63
80
93
Ooep
p=1.062e-03
p=3.270e-09
p=9.770e-01
20
67
68
103
Tle6
p=9.69e-10
p=5.88e-07
p=2.20e-02
35
91
114
169
Zfp57
p=4.50e-3
p=1.233e-4
p=4.55e-1
28
58
67
81
Elobl
p=6.60e-10
p=5.94e-07
p=1.00e-03
21
53
66
75
Zar1
p=7.53e-08
p=4.71e-04
p=1.00e-03
25
83
117
144
Ezhip
p=1.842e-06
p=2.460e-04
p=2.460e-04
15
31
45
45
Fbxw24
p=6.915e-04
p=5.790e-09
p=1.190e-01
31
128
150
187
Nlrp14
p=1.500e-03
p=1.008e-16
p=3.350e-01
24
103
126
141
Uchl1
p=8.040e-05
p=1.854e-08
p=5.910e-04
62
105
138
143
Poly(A) tail length
Genotype
Tent5b-/- Tent5c-/-
Tent5b wt/wt
Tent5b wt/gfp
Tent5b gfp/gfp
(n=31) (n=32) (n=24) (n=40)
(n=30) (n=72) (n=60) (n=80)
(n=72) (n=132) (n=98) (n=191)
(n=57) (n=104) (n=37) (n=71)
(n=48) (n=66) (n=49) (n=78)
(n=29) (n=145) (n=131) (n=214)
(n=27) (n=91) (n=38) (n=78)
(n=24) (n=76) (n=28) (n=78)
(n=33) (n=57) (n=45) (n=39)
(n=79) (n=123) (n=46) (n=82)
(n=123) (n=219) (n=119) (n=135)

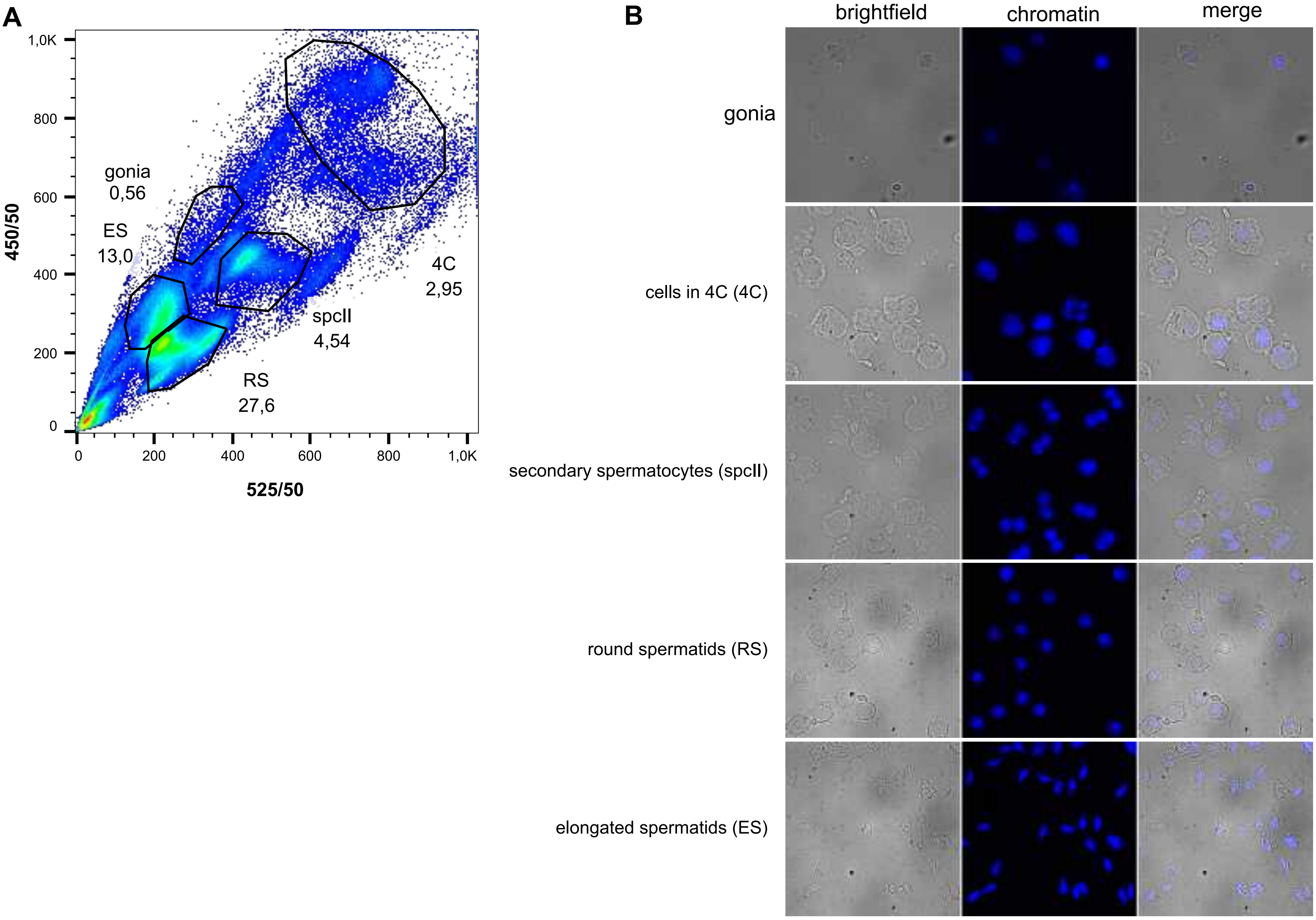
Supplementary Figure 3. TENT5s polyadenylate mRNAs in oocytes, which tight regulation is essential for oogenesis.

DRS-based poly(A) tail lengths profiling of mRNAs isolated from ovaries. Median poly(A) tail lengths are plotted in white rectangles, p-values reported for comparison in Mann-Whitney-Wilcoxon test with Bonferroni Hallberg correction.


Tent5b+/+
Tent5c+/+
brightfield
GDF9 +
chromatine
Tent5b-/-
Tent5c-/-
GDF9 +
chromatine
brightfield
Tent5b+/+
Tent5c+/+
brightfield
ZP3 +
chromatine
Tent5b-/-
Tent5c-/-
ZP3 +
chromatine
brightfield
follicle growth

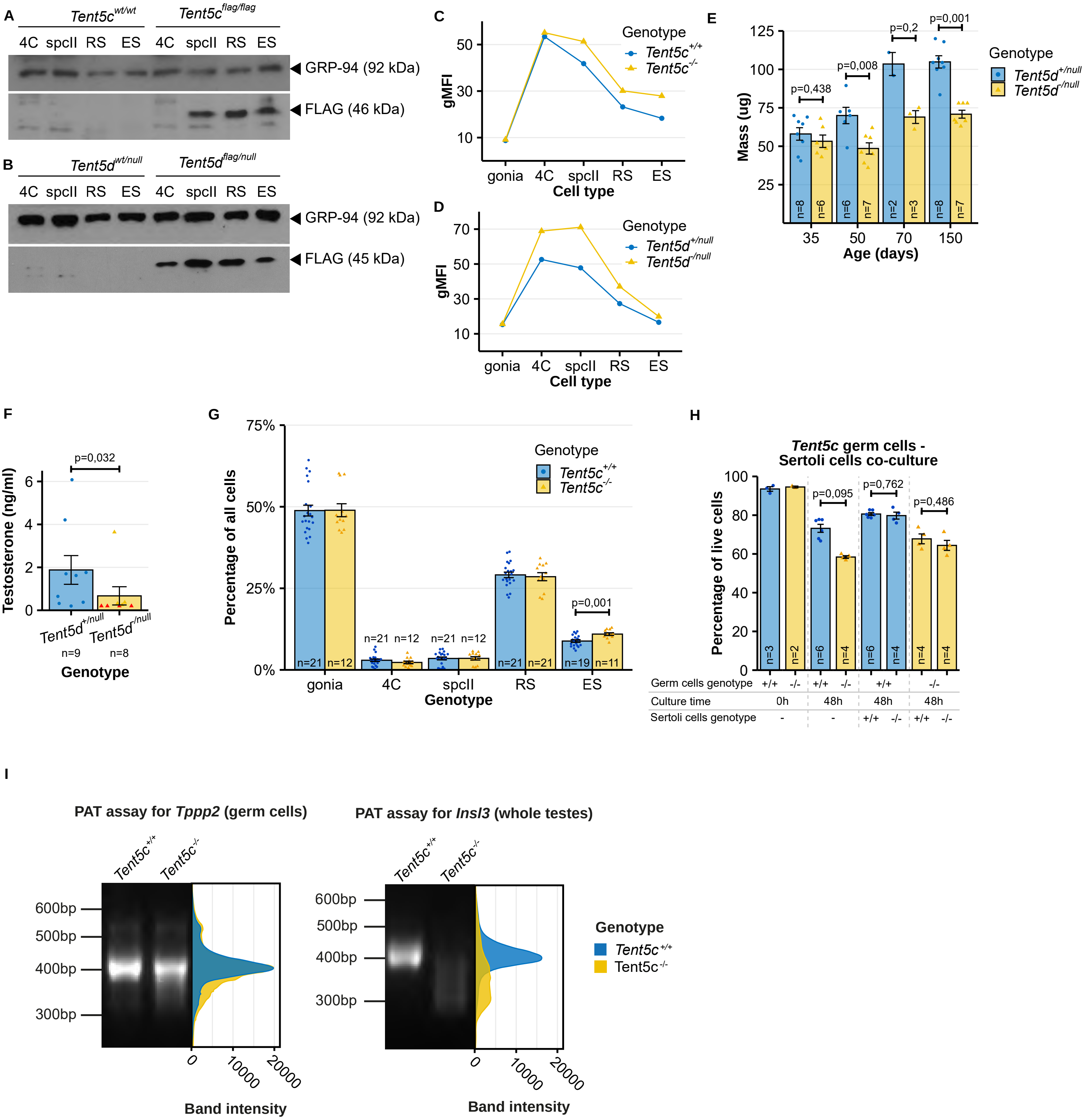
Supplementary Figure 4. Ovary immunohistochemistry staining

Immunohistochemistry staining of GDF9 (green), ZP3 (green) and chromatin (magenta) in ovaries of *Tent5b*$^{-/-}$ *Tent5c*$^{-/-}$ and *Tent5b*$^{+/+}$ *Tent5c*$^{+/+}$ females. Scale bar = 50 µm.


A
450/50
525/50
1,0K
800
600
400
200
0
gonia
0,56
ES
13,0
4C
2,95
spcII
4,54
RS
27,6
B
brightfield
chromatin
merge
gonia
cells in 4C (4C)
secondary spermatocytes (spcII)
round spermatids (RS)
elongated spermatids (ES)

Supplementary Figure 5. Male germ cells sorting

A. Gating example of cells on different stages of spermatogenesis based on cells' size, shape, and DNA content in FACS analysis. **B**. Morphology and DNA staining pictures confirming cell sorting procedures correctness. Cells at different spermatogenesis stages display different patterns of DNA staining, size and shape.


A
Tent5c wt/wt
Tent5c flag/flag
4C spcII RS ES 4C spcII RS ES
GRP-94 (92 kDa)
FLAG (46 kDa)
B
Tent5d wt/null
Tent5d flag/null
4C spcII RS ES 4C spcII RS ES
GRP-94 (92 kDa)
FLAG (45 kDa)
C
gMFI
50
30
10
gonia 4C spcII RS ES
Cell type
Genotype
Tent5c +/+
Tent5c -/-
D
gMFI
70
50
30
10
gonia 4C spcII RS ES
Cell type
Genotype
Tent5d +/null
Tent5d -/null
E
Mass (ug)
125
100
75
50
25
p=0,438
p=0,008
p=0,2
p=0,001
n=8 n=6 n=6 n=7 n=2 n=3 n=8 n=7
35 50 70 150
Age (days)
Genotype
Tent5d +/null
Tent5d -/null
F
Testosterone (ng/ml)
6
4
2
0
p=0,032
Tent5d +/null
Tent5d -/null
n=9
n=8
Genotype
G
Percentage of all cells
75%
50%
25%
0%
Genotype
Tent5c +/+
Tent5c -/-
n=21 n=12
n=21 n=12
n=21 n=12
n=21 n=21
p=0,001
n=19 n=11
gonia 4C spcII RS ES
Genotype
H
Tent5c germ cells - Sertoli cells co-culture
Percentage of live cells
100
80
60
40
20
0
p=0,095
p=0,762
p=0,486
n=3 n=2 n=6 n=4 n=6 n=4 n=4 n=4
Germ cells genotype +/+ -/- +/+ -/- +/+ -/-
Culture time 0h 48h 48h 48h
Sertoli cells genotype - - +/+ -/- +/+ -/-
I
PAT assay for Tppp2 (germ cells)
Tent5c +/+
Tent5c -/-
600bp
500bp
400bp
300bp
0
10000
20000
Band intensity
PAT assay for Insl3 (whole testes)
Tent5c +/+
Tent5c -/-
600bp
500bp
400bp
300bp
0
10000
20000
Band intensity
Genotype
Tent5c +/+
Tent5c -/-

Supplementary Figure 6. Analysis of *Tent5c* and *Tent5d* KO phenotype in males

A-B. Western blot analysis of FLAG-tagged TENT5C and TENT5D expression in different stages of spermatogenesis. Antibodies against GRP-94 were used as a loading control. Black arrowheads mark position of detected proteins. This experiment had one technical repetition. **C-D.** Cytometric analysis of GFP-tagged TENT5C and TENT5D expression in different stages of spermatogenesis, presented as changes as gMFI (geometric mean fluorescence intensity). **E.** Testes mass changes in *Tent5d*$^{+/null}$ and *Tent5d*$^{-/null}$ males in life. Individual data points and n values represent testes from individual males weighted, bars represent mean values, error bars represent SEM, p-values reported for comparison in Mann-Whitney-Wilcoxon test. **F.** Blood testosterone level in adult *Tent5d*$^{+/null}$ and *Tent5d*$^{-/null}$ males. Individual data points and n values represent samples from individual males, bars represent mean values, error bars represent SEM, p-values reported for comparison in Mann-Whitney-Wilcoxon test. **G.** Percentage of germ cells at different spermatogenesis stages among all isolated germ cells in *Tent5c*$^{+/+}$ and *Tent5c*$^{-/-}$ males. Individual data points and n values represent individual males from which germ cells were isolated, bars represent mean values, error bars represent SEM, p-value reported for comparison in t-test, two-tailed. **H.** Germ cell survival in 48h of *in vitro* culture depending on *Tent5c* genotype of germ cells and presence and genotype of Sertoli cells. Individual data points and n values represent cell culture of germ cells isolated from single male, bars represent mean values, error bars represent SEM; p-values reported for comparison in Mann-Whitney-Wilcoxon test. **I.** PAT assay visualizing poly(A) tail length distribution of *Tppp2* and *Insl3* transcripts PCR-amplified from whole mRNA isolated from germ cells (*Tppp2*) and whole testes (*Insl3*).

A

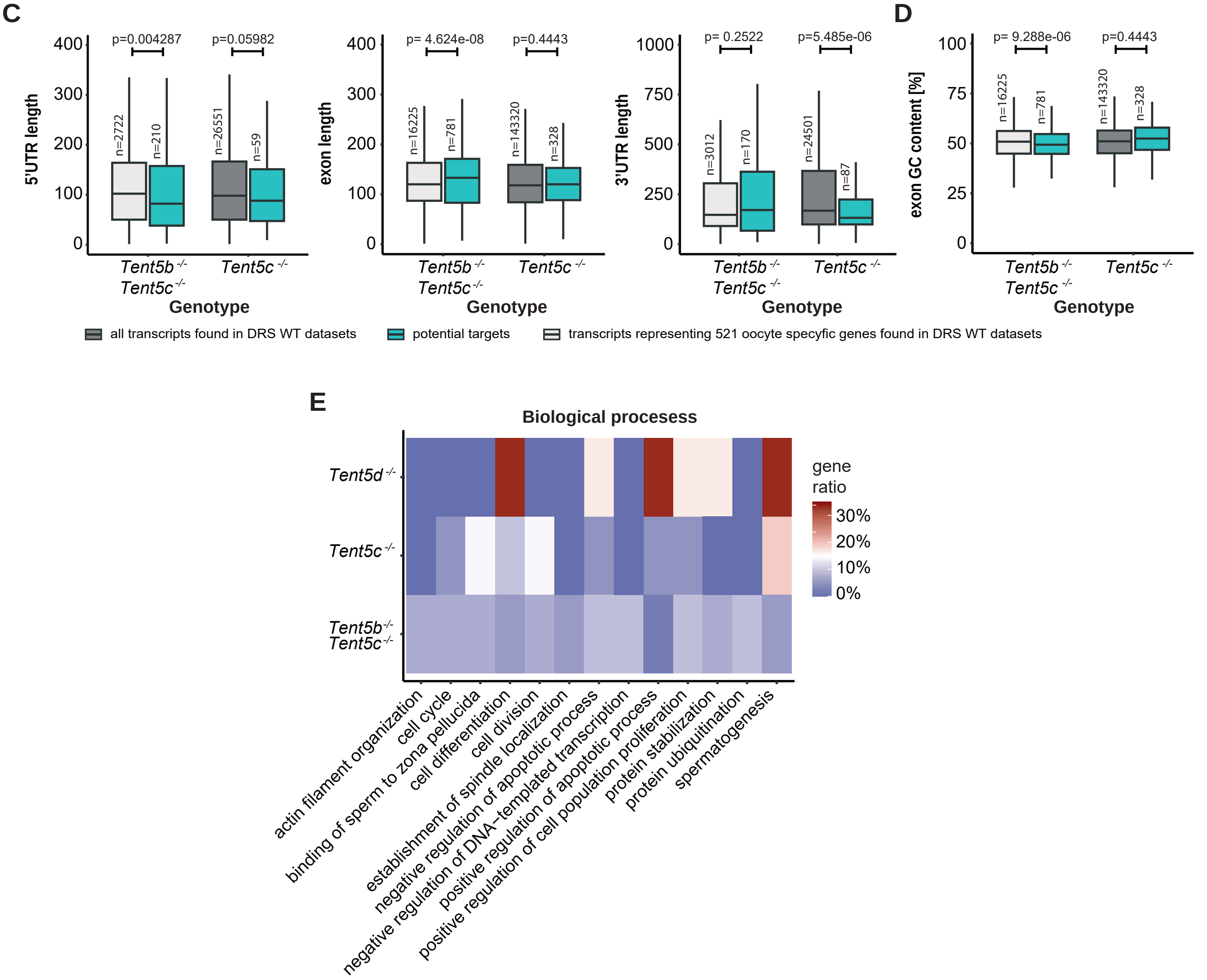
Motifs found in sequences of genes encoding potential TENT5B and TENT5C targets

region	motif	E-value	procentage of sequences maching the motive
3'UTR	AATAAAAGTTT	2.9e-002	38.2%
	TAGGAGGGA	5.6e-002	8.2%
	CAGAATACCA	1.1e-001	8.8%
	TACATCCTCCAA	2.5e-001	14.1%
5'UTR	CAGTGACACTGTG	2.8e-002	7.8%
	AGTGGCCTG	5.3e-002	10.7%
	AGTAATCCAGA	8.1e-002	4.9%

B

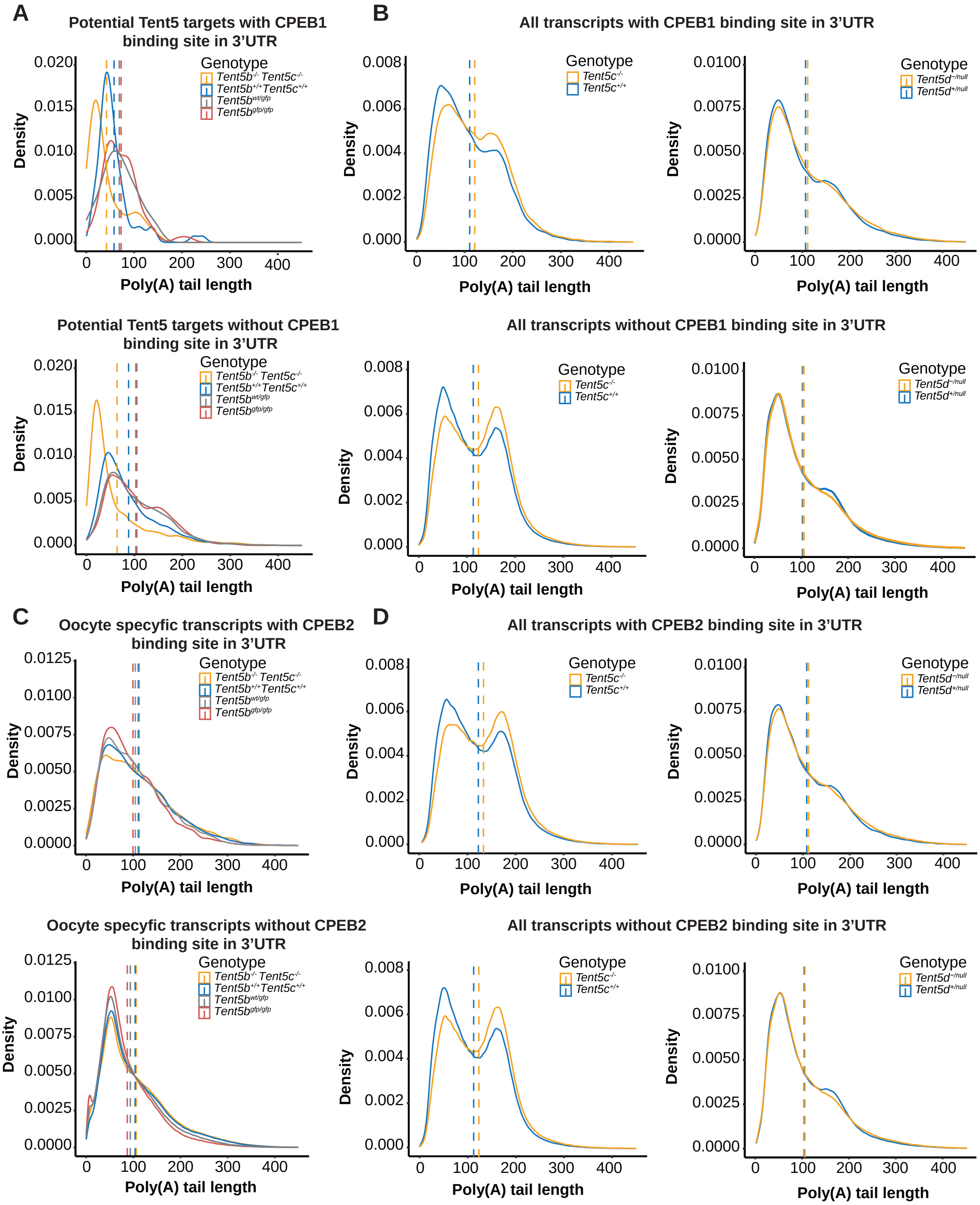
Motifs found in sequences of genes encoding potential TENT5C targets

region	motif	E-value	procentage of sequences maching the motive
3'UTR	CGTGGATGTCA	3.0e+000	10.5%
	CAAAGAAATAAAAGT	5.9e-003	7%
	ATTAGCGACA	3.0e+000	7%
5'UTR	CAGGCCCAGGCA	2.3e-002	8.5%
	ACCAGCCCC	6.4e-001	22%



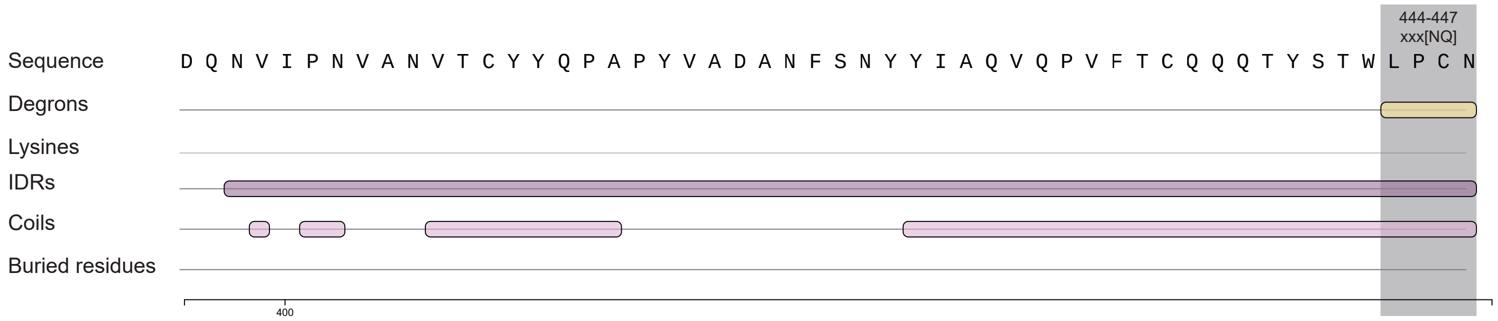
Supplementary Figure 7. Motif and Gene ontology analysis of DRS results

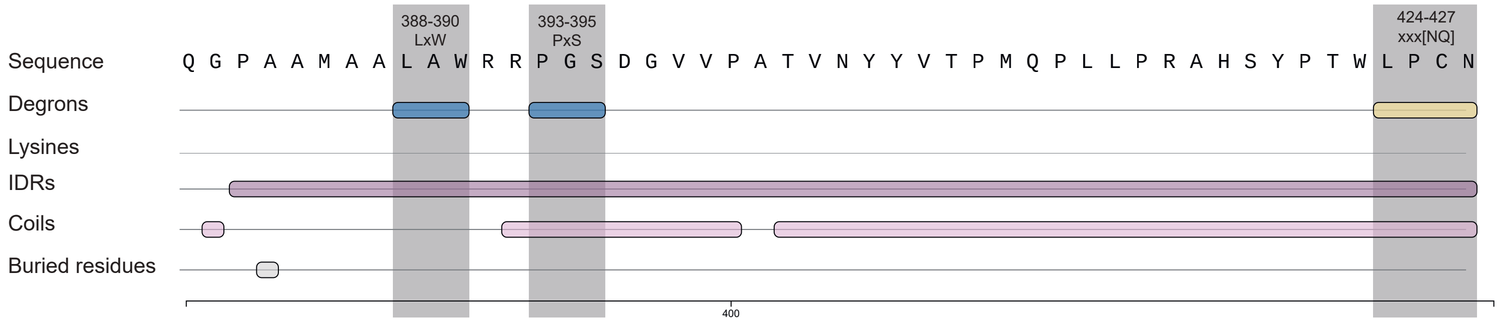
A-B. 3`UTR, 5’UTR motif analysis did not reveal any enriched motifs. In CDS only in Tent5B/C substrates, short A-rich motive is significantly enriched. **C.** Differences in 5’UTR, 3’UTR and exon length between potential targets of TENT5 proteins compared to the rest of the transcriptome detected by DRS. In the ovaries, analyses were performed for 522 oocyte-enriched mRNAs. Mann-Whitney-Wilcoxon rank sum test was calculated. **D.** Difference in GC content between potential targets of Tent5 proteins. Mann-Whitney-Wilcoxon rank sum test was calculated. **E.** Gene ontology (GO) analysis of potential TENT5 protein targets.

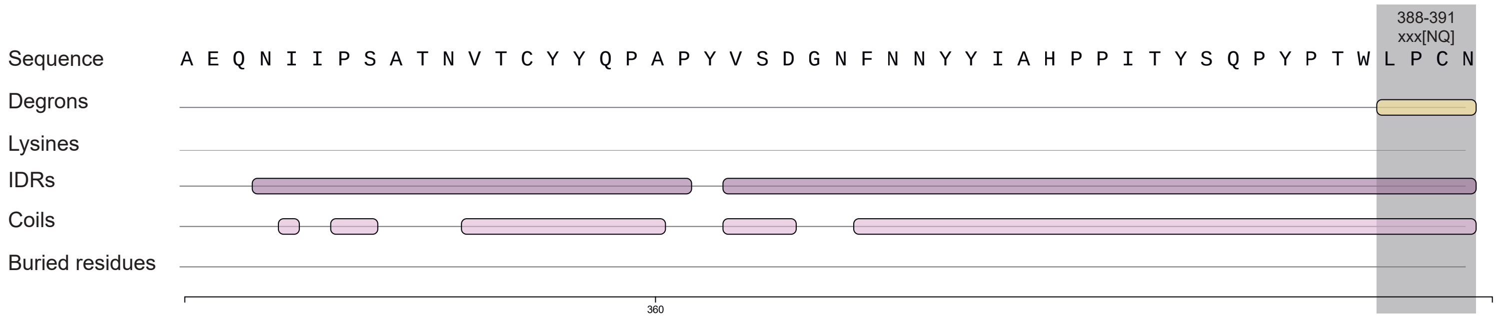

A
Potential Tent5 targets with CPEB1 binding site in 3’UTR
Potential Tent5 targets without CPEB1 binding site in 3’UTR
Genotype
Tent5b$^{-/-}$ Tent5c$^{-/-}$
Tent5b$^{+/+}$ Tent5c$^{+/+}$
Tent5b$^{wt/gfp}$
Tent5b$^{gfp/gfp}$
Density
0.000
0.005
0.010
0.015
0.020
Poly(A) tail length
0
100
200
300
400
B
All transcripts with CPEB1 binding site in 3’UTR
All transcripts without CPEB1 binding site in 3’UTR
Genotype
Tent5c$^{-/-}$
Tent5c$^{+/+}$
Tent5d$^{-/null}$
Tent5d$^{+/null}$
0.000
0.002
0.004
0.006
0.008
0.0000
0.0025
0.0050
0.0075
0.0100
C
Oocyte specyfic transcripts with CPEB2 binding site in 3’UTR
Oocyte specyfic transcripts without CPEB2 binding site in 3’UTR
0.0000
0.0025
0.0050
0.0075
0.0100
0.0125
D
All transcripts with CPEB2 binding site in 3’UTR
All transcripts without CPEB2 binding site in 3’UTR

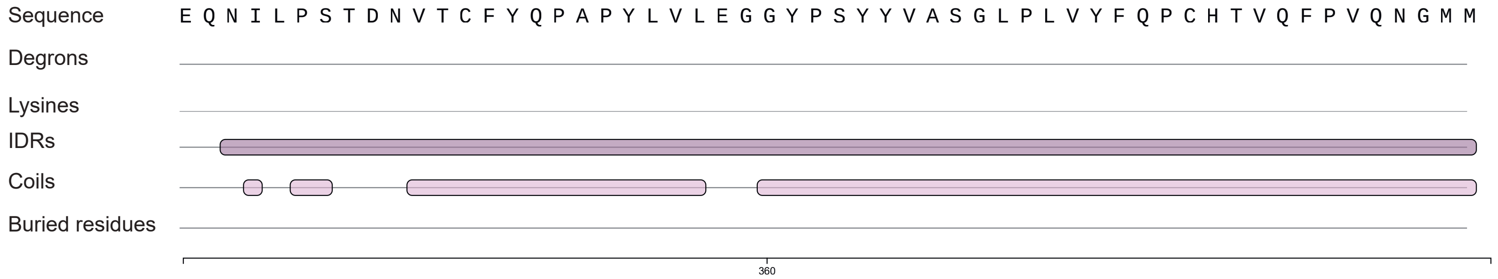
Supplementary Figure 8. CPEB1 and CPEB2 motifs analysis of DRS data

A. Changes in the global distribution of poly(A) tail lengths for group of transcripts selected as potential Tent5B/C targets. Mann-Whitney-Wilcoxon rank sum test was calculated. Dashed lines indicate the mean poly(A) tails lengths: Tent5b -/- Tent5c -/- - 42, Tent5b +/+ - 58, Tent5b wt/gfp - 69, Tent5b gfp/gfp - 73 for transcripts with CPEB1 and Tent5b-/- Tent5c-/- - 64, Tent5b+/+ - 88, Tent5b wt/gfp - 103, Tent5b gfp/gfp-105 for transcripts without CPEB1. **B.** The global distribution of polyA tails lengths of RNA isolated from Tent5c +/+, Tent5c -/-, Tent5d +/+, Tent5d -/- mice. Mann-Whitney-Wilcoxon rank sum test was calculated. Dashed lines indicate the mean poly(A) tails lengths: Tent5c -/- - 130, Tent5c +/+ - 119, Tent5d -/- - 107, Tent5d +/+ - 112 for transcripts with CPEB1 and Tent5c -/- - 124, Tent5c +/+ - 113, Tent5d -/- - 105 , Tent5d +/+ - 103 for transcripts without CPEB1. **C.** The global distribution of polyA tails lengths of RNA isolated from Tent5b -/- Tent5c -/-, Tent5b +/+, Tent5b wt/gfp and Tent5b gfp/gfp mice. Mann-Whitney-Wilcoxon rank sum test was calculated. Dashed lines indicate the mean poly(A) tails lengths: Tent5b -/- Tent5c -/- - 116, Tent5b +/+ - 111, Tent5b wt/gfp - 107, Tent5b gfp/gfp - 99 for transcripts with CPEB2 and Tent5b-/- Tent5c-/- - 106, Tent5b+/+ - 104, Tent5b wt/gfp - 94, Tent5b gfp/gfp- 87 for transcripts without CPEB2. **D.** The global distribution of polyA tails lengths of RNA isolated from Tent5c +/+110, Tent5c -/- 127, Tent5d +/+, Tent5d -/- mice. Mann-Whitney-Wilcoxon rank sum test was calculated. Dashed lines indicate the mean poly(A) tails lengths: Tent5c -/- - 129, Tent5c +/+ - 118, Tent5d -/- - 112 , Tent5d +/+ - 108 for transcripts with CPEB2 and Tent5c -/- - 123, Tent5c +/+ - 112, Tent5d -/- - 104, Tent5d +/+ - 103 for transcripts without CPEB2.


TENT5A


TENT5B


TENT5C


TENT5D

Supplementary Figure 9. Degrons in C-terminus of TENT5 proteins

Degron motifs detected in C-terminal region of TENT5 proteins using DEGRONOPEDIA. While C-terminal xxx[NQ] degron within intrinsically disordered region (IDR) is shared between TENT5A, B and C sequences, only TENT5B possesses two additional degron motifs within the same IDR.

Supplementary Table 1. Oligonucleotides, sgRNA and DNA repair templates sequences

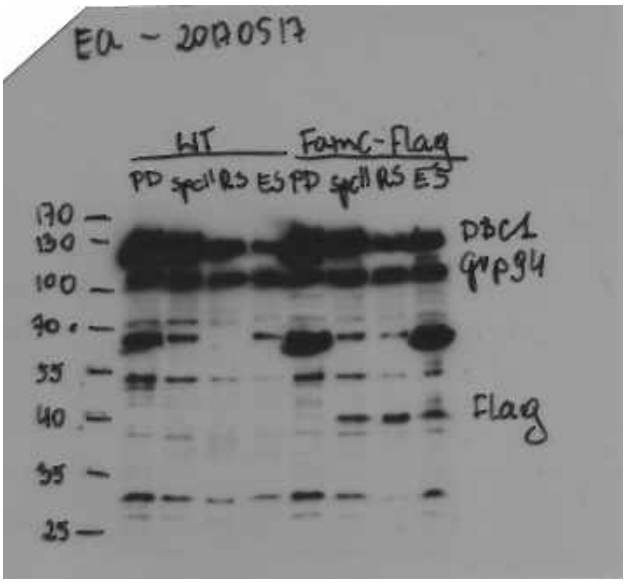
Fw primer for genotyping of Tent5a KO mouse line (#2001): CAAGCCTGATTGTGAAGGTG	Gewartowska et al.[29]
Rv primer for genotyping of *Tent5a* KO mouse line (#2002): AAGGAAGAGAAGGAAACGCA	Gewartowska et al.[29]
Fw primer for genotyping of *Tent5b* KO mouse line (large deletion) (#2007): TTAGCCTGAAGACGCTATGG	
Rv primer for genotyping of Tent5b KO mouse line (large deletion) (#2008): GCATGGGGGTCACATAGTAA	
Fw primer for genotyping of *Tent5b* KO mouse line (small deletion) (#2005): AATCGAGGCTTAGCGAGTTA	
Rv primer for genotyping of *Tent5b* KO mouse line (small deletion) (#2006)	
Fw primer for genotyping of *Tent5c* KO mouse line (#2012): AGGTCCTGACTGAGGTCGTG	Mroczek et al.[27]
Rv primer for genotyping of *Tent5c* KO mouse line (#2013): TTCCTCAAAATCCCCGTACA	Mroczek et al.[27]
Fw primer for genotyping of Tent5d KO mouse line (#2016): TCGGAGATCAGATTCAGCAAT	This paper
Rv primer for genotyping of Tent5d KO mouse line (#2017): TAGTTCTACGTTCTTCCCGTT	This paper
Fw primer for genotyping of *Tent5b*-GFP and *Tent5b*-FLAG mouse lines (#2310): GTGGTCAACGAGAGCACAGTG	This paper
Rv1 primer for genotyping of *Tent5b*-GFP and *Tent5b*-FLAG mouse lines (#2311): gaagatggtgcgctcctgg	This paper
Fw primer for genotyping of GFP-*Tent5b* mouse line (#2160): atcagaaactcctgagagcc	This paper
Rv1 primer for genotyping of GFP-*Tent5b* mouse line (#2180): GAATGGGGATCGGCTCTTTC	This paper
Rv2 primer for genotyping of GFP-*Tent5b* mouse line (#2326): gacacgctgaacttgtggc	This paper
Fw primer for genotyping of *Tent5c*-GFP and *Tent5c*-FLAG mouse lines (#2014): CTTCAGAACCACTTCTCGGA	Mroczek et al.[27]
Rv primer for genotyping of *Tent5c*-GFP and *Tent5c*-FLAG mouse lines (#2015): AGAAGTCACGCCTCCTATTG	Mroczek et al.[27]
Fw primer for genotyping of *Tent5d*-GFP and *Tent5d*-FLAG mouse lines (#2018): CCCAGTACAGACAACGTAAC	This paper
Rv primer for genotyping of *Tent5d*-GFP and *Tent5d*-FLAG mouse lines (#2019): GTGTTCTTTCATACGTTAGCC	This paper
gRNA used for generation of *Tent5b* KO mouse line: TGTAGCCTAGGCCGCTCTC	This paper
gRNA used for generation of *Tent5d* KO mouse line: ACATACTCGCAAGCCATAA	This paper
gRNA used for generation of *Tent5b*-GFP and *Tent5b*-FLAG mouse lines: AGGATCAGAGTCAGTTGCA	This paper
dsDNA used as repair template for generation of *Tent5b*-FLAG mouse line: ACCCCCATGCAGCCACTGCTGCCCCGAGCTCACTCCTATCCTACCTGGCTGCCTTGCAACGACTACAAAGACGATGACGACAAGTGACTCTGATCCTGGCCAGAAGGGAATGAGCGCCATGGGGTGGGGTGGGGTCATCAGGTA	This paper

dsDNA used as repair template for generation of *Tent5b*-GFP mouse line: ACCCCCATGCAGCCACTGCTGCCCCGAGCTCACTCCTATCCTACCTGGCTGCCTTGCAACgagaatttgtattttcagggtgatatcatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagCtggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagTGACTCTGATCCTGGCCAGAAGGGAATGAGCGCCATGGGGTGGGGTGGGGTCATCAGGTA	This paper
dsDNA used as repair template for generation of GFP-*Tent5b* mouse line: TTGGCCCGTGCACAGCCACTCTCCCTGCCCTCCTCGCCTTCACCATTCCCGGGTTTCTGCCGTCCAGGCACCGGGGCCGGGTGAATGGTttCtAAGGGaGAAGagctgttcacAggAgtggtgccTatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccaagctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggatCTggAGAAAACCTGTACTTCCAAGGAatgccATctgagagtggagctgaAagcctggagcagccagctgcgcaggtggggaccggtgcagcctcggcagtGGCCACGGCTG	This paper
dsDNA used as repair template for generation of *Tent5d*-FLAG mouse line: CCACTGGTTTATTTCCAGCCATGTCATACAGTGCAGTTCCCTGTGCAAAATGGTATGATGGACTACAAAGACGATGACGACAAGTAAGAAATACACATACCACAAGTTTTGCTTAAGCAACTCTGAAAAAGCAATTTTCCAAGT	This paper
dsDNA used as repair template for generation of *Tent5d*-GFP mouse line: CCACTGGTTTATTTCCAGCCATGTCATACAGTGCAGTTCCCTGTGCAAAATGGTATGATGgagaatttgtattttcagggtgatatcatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagTAAGAAATACACATACCACAAGTTTTGCTTAAGCAACTCTGAAAAAGCAATTTTCCAAGT	This paper
Tn5ME-A: CGTCGGCAGCGTCAGATGTGTATAAGAGACAG	Hennig et al.[64]
Tn5ME-B: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	Hennig et al.[64]
Tn5MErev: [phos]CTGTCTCTTATACACATCT	Hennig et al.[64]
Fw primer for SLIC cloning: atgaacgagctctataagagatctttcgaaaaggacgagctgtaacaaagaaagcccagctccttc	This paper
Rv primer for SLIC cloning: cacagtcgaggctgatcagcgggtttaaacaaagtgtgaaaaatacctctg	This paper

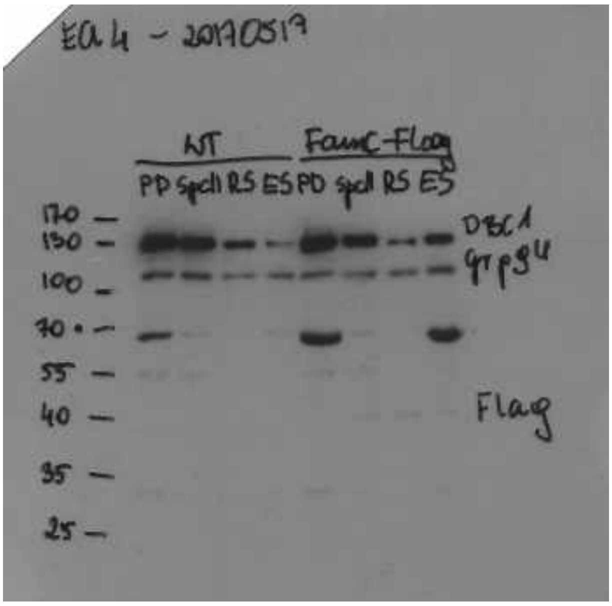
Fw oligonucleotide for Zp3 signal peptide sequence: agcttATGGCGTCAAGCTATTTCCTCTTCCTTTGTCTCCTGCTGTGTGGAGGCCCCGAGCTGTGCAATTCC	This paper
Rv oligonucleotide for Zp3 signal peptide sequence: CCGGGGAATTGCACAGCTCGGGGCCTCCACACAGCAGGAGACAAAGGAAGAGGAAATAGCTTGACGCCATa	This paper
Fw oligonucleotide for Gdf9 signal peptide sequence: agcttATGGCACTTCCCAGCAACTTCCTGTTGGGGGTTTGCTGCTTTGCCTGGCTGTGTTTTCTTAGTAGCCTTAGCTCTCAGGCTTCTACT	This paper
Rv oligonucleotide for Gdf9 signal peptide sequence: CCGGAGTAGAAGCCTGAGAGCTAAGGCTACTAAGAAAACACAGCCAGGCAAAGCAGCAAACCCCCAACAGGAAGTTGCTGGGAAGTGCCATa	This paper
Fw primer for plasmid linearization: tagagaacccactgcttactgg	This paper
Rv primer for plasmid linearization: TTTTTTTTTTTTTTTTTTTTggattgaaggagctgggcttt	This paper
Fw oligonucleotide for Zp3 PAT assay: GGGCCCCTGATATTCCTTGG	This paper
Fw oligonucleotide for Gdf9 PAT assay: CTGTGACCAGTCTCTCCGTG	This paper
Fw oligonucleotide for Tppp2 PAT assay: ACACCGGAACTCACAAGGAG	This paper
Fw oligonucleotide for Insl3 PAT assay: CGGTGCGTCCTTAATTGCTC	This paper
Rv universal oligonucleotide for PAT assay: CTTGCCTGTCGCTCTATCTT	This paper

Uncropped blots - Supplementary Figures

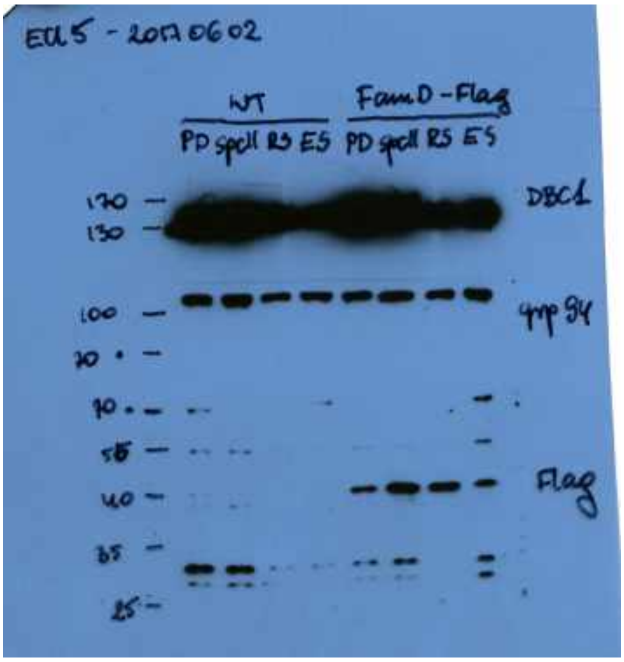
Sup. Figure 6. A - FLAG bands



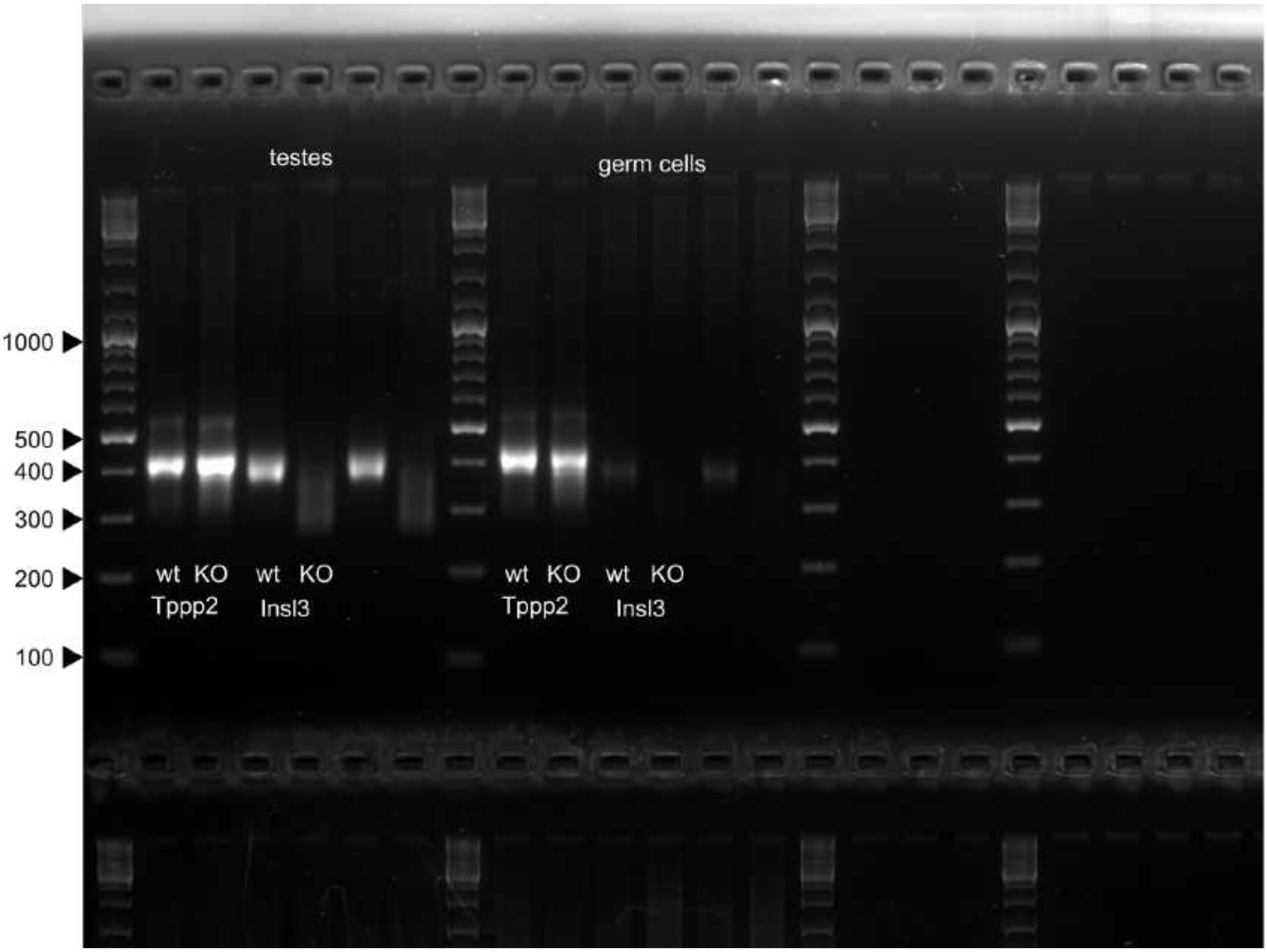
Sup. Figure 6. A - GRP-94 bands



Sup. Figure 6. B - GRP-94 and FLAG bands



Sup. Figure 6. I



3.3 Manuscript 3: Comprehensive analysis of poly(A) tails in mouse testes and ovaries using Nanopore Direct RNA Sequencing.

3.3.1 Description

During both oogenesis and spermatogenesis, diploid cells undergo meiotic division, resulting in the production of haploid gametes. Complex transcriptome remodeling involves various pathways of post-transcriptional gene regulation that are crucial at different stages of maturation. The main aim of this publication was to provide new datasets derived from TENT5 poly(A) polymerase mutant mice, shedding new light on the poly(A) tail-dependent mRNA regulation pathway. DRS data from 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 mice obtained for our previous publication (Brouze et al., 2024) were reanalyzed to examine poly(A) tail composition. Raw nanopore signals were processed using the Ninetails software to identify non-adenine nucleotides within poly(A) tails. It was discovered that mRNA molecules isolated from ovaries and testes contain a higher proportion of non-canonical nucleotides in their poly(A) tails compared to those from other mouse tissues (Krawczyk et al., 2025). To complement the DRS data, Illumina-based RNA-seq data obtained from sorted cell populations isolated from the testes of wild-type (WT) and *Tent5c* KO mice were incorporated.

Based on the poly(A) tail composition analysis, transcripts enriched in uridines at the terminal regions of their poly(A) tails were identified. mRNA uridylation is a post-transcriptional regulatory mechanism essential for proper germ cell development (Morgan et al., 2019, 2017). It has been previously demonstrated that terminal uridyltransferase 4 (TUT4) and TUT7-mediated 3′ mRNA uridylation contribute to the mRNA degradation waves observed during pachynema. Moreover, both TUT4 and TUT7 are implicated in causing embryonic growth defects (Morgan et al., 2019). Differential expression analysis performed on the RNA-seq datasets revealed that transcripts with U-rich 3′ poly(A) tails are predominantly highly expressed in spermatids. The expression levels of some of these transcripts change significantly during the transition from the round to the elongated spermatid stage. This underscores the role of uridylation in the dynamic regulation of mRNA stability and function during gametogenesis.

3.3.2 My contribution

In this manuscript, in which, together with dr Natalia Gumińska and Michał Brouze, we share a joint co-first authorship, I performed quality control analyses for RNA-seq and Nanopore DRS datasets. While dr Natalia Gumińska performed the Ninetails analyses, I handled the downstream data analysis and visualisation of poly(A) tail consistence. I was responsible for differential expression analysis and visualization of RNAseq datasets. I compiled figures to represent the sequencing data and drafted the initial version of the manuscript. In collaboration with dr Natalia Gumińska, Michał Brouze and Professor Andrzej Dziembowski, I contributed to writing and refining the final manuscript and supplemental materials.

3.3.3 Citation and manuscript text

Czarnocka-Cieciura A, Brouze M, Gumińska N, Mroczek S, Gewartowska O, Krawczyk PS, Dziembowski A. Comprehensive analysis of poly(A) tails in mouse testes and ovaries using Nanopore Direct RNA Sequencing. Sci Data. 2025 Jan 10;12(1):43. doi: 10.1038/s41597-024-04226-8. PMID: 39794363; PMCID: PMC11724052.

www.nature.com/scientificdata

scientific **data**

OPEN

DATA DESCRIPTOR

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

Agnieszka Czarnocka-Cieciura[1,5], Michał Brouze[1,5], Natalia Gumińska[1,5], Seweryn Mroczek[1,2], Olga Gewartowska[3], Paweł S. Krawczyk[1] & Andrzej Dziembowski[1,4]✉

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 (Fig. 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

---

[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. [5]These authors contributed equally: Agnieszka Czarnocka-Cieciura, Michał Brouze, Natalia Gumińska. ✉e-mail: adziembowski@iimcb.gov.pl

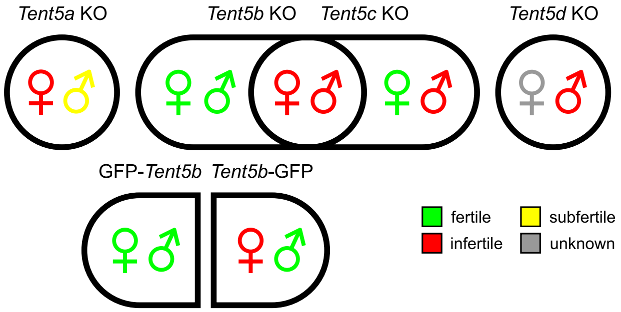


Fig. 1 Schematic representation of the Tent5 phenotypes in mouse model. Summary of fertility status of males and females with all analyzed *Tent5* genes' mutations. Subfertile category refers to significantly reduced fertility. Female *Tent5d* KO mice were unavailable for analyses due to *Tent5d* localisation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

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].

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 (Fig. 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 (Fig. 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.

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.

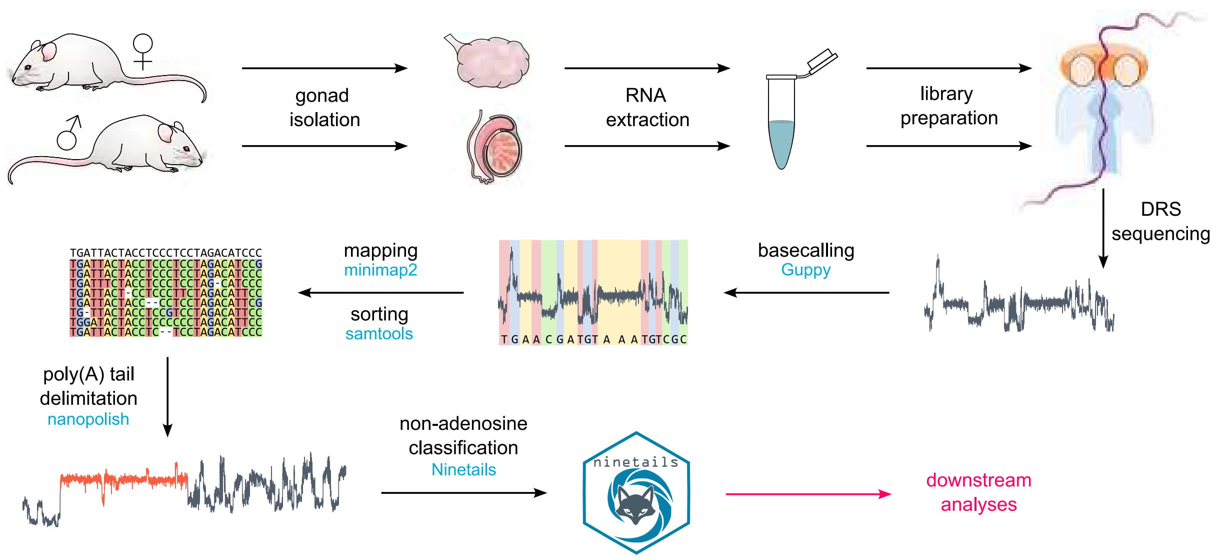


Fig. 2 Schematic representation of the non-adenosine analysis workflow. Nanopore data were basecalled with Guppy, mapped with minimap2, sorted and filtered with samtools. The coordinates of poly(A) tails within the nanopore signals were delimited with Nanopolish. Finally, Ninetails was used to analyze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

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 3 M 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].

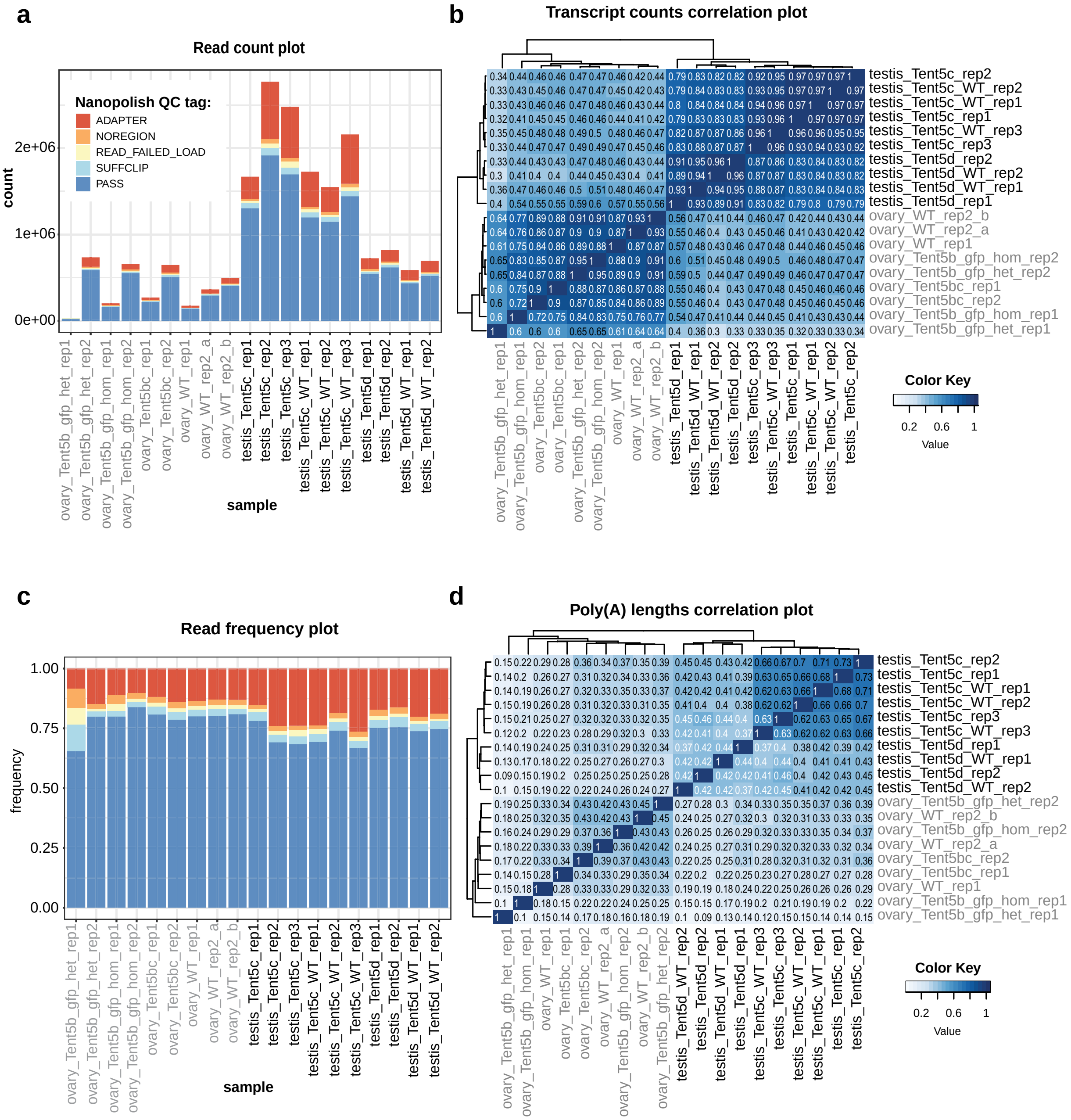


Fig. 3 Quality of underlying DRS sequencing data. (**a**) Detailed classification of reads based on poly(A) tail length prediction quality tags provided by Nanopolish. The quality of the reads was independently summarized for each sample. (**b**) Correlation matrix for all DRS datasets. Spearman correlations 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.

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]. 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.

sample_name	replicate	species	organ	platform	project	accession	seq_data_link	seq_source	nonA_data	diffexp_data
ovary_Tent5bc_rep1	1	Mus musculus	ovary	ONT (DRS)	PRJEB63526	ERS15941056	https://www.ebi.ac.uk/ena/browser/view/ERS15941056	Brouze *et al.*[1]	Y	Y
ovary_Tent5bc_rep2	2	Mus musculus	ovary	ONT (DRS)	PRJEB63526	ERS15941057	https://www.ebi.ac.uk/ena/browser/view/ERS15941057	Brouze *et al.*[1]	Y	Y
ovary_WT_rep1	1	Mus musculus	ovary	ONT (DRS)	PRJEB63526	ERS16372230	https://www.ebi.ac.uk/ena/browser/view/ERS16372230	Brouze *et al.*[1]	Y	Y
ovary_WT_rep2_a	2	Mus musculus	ovary	ONT (DRS)	PRJEB63526	ERS15941059	https://www.ebi.ac.uk/ena/browser/view/ERS15941059	Brouze *et al.*[1]	Y	Y
ovary_WT_rep2_b	2	Mus musculus	ovary	ONT (DRS)	PRJEB46685	ERS5927539	https://www.ebi.ac.uk/ena/browser/view/ERS5927539	Brouze *et al.*[1]	Y	Y
ovary_Tent5b_gfp_het_rep1	1	Mus musculus	ovary	ONT (DRS)	PRJEB63526	ERS16372228	https://www.ebi.ac.uk/ena/browser/view/ERS16372228	Brouze *et al.*[1]	Y	Y
ovary_Tent5b_gfp_het_rep2	2	Mus musculus	ovary	ONT (DRS)	PRJEB63526	ERS16372248	https://www.ebi.ac.uk/ena/browser/view/ERS16372248	Brouze *et al.*[1]	Y	Y
ovary_Tent5b_gfp_hom_rep1	1	Mus musculus	ovary	ONT (DRS)	PRJEB63526	ERS15941063	https://www.ebi.ac.uk/ena/browser/view/ERS15941063	Brouze *et al.*[1]	Y	Y
ovary_Tent5b_gfp_hom_rep2	2	Mus musculus	ovary	ONT (DRS)	PRJEB63526	ERS16372252	https://www.ebi.ac.uk/ena/browser/view/ERS16372252	Brouze *et al.*[1]	Y	Y
testis_Tent5c_rep1	1	Mus musculus	testis	ONT (DRS)	PRJEB45063	ERS5846736	https://www.ebi.ac.uk/ena/browser/view/ERS5846736	Brouze *et al.*[1]	Y	Y
testis_Tent5c_rep2	2	Mus musculus	testis	ONT (DRS)	PRJEB45063	ERS5846735	https://www.ebi.ac.uk/ena/browser/view/ERS5846735	Brouze *et al.*[1]	Y	Y
testis_Tent5c_rep3	3	Mus musculus	testis	ONT (DRS)	PRJEB45063	ERS5846739	https://www.ebi.ac.uk/ena/browser/view/ERS5846739	Brouze *et al.*[1]	Y	Y
testis_Tent5c_WT_rep1	1	Mus musculus	testis	ONT (DRS)	PRJEB45063	ERS5846734	https://www.ebi.ac.uk/ena/browser/view/ERS5846734	Brouze *et al.*[1]	Y	Y
testis_Tent5c_WT_rep2	2	Mus musculus	testis	ONT (DRS)	PRJEB45063	ERS5846733	https://www.ebi.ac.uk/ena/browser/view/ERS5846733	Brouze *et al.*[1]	Y	Y
testis_Tent5c_WT_rep3	3	Mus musculus	testis	ONT (DRS)	PRJEB45063	ERS5846741	https://www.ebi.ac.uk/ena/browser/view/ERS5846741	Brouze *et al.*[1]	Y	Y
testis_Tent5d_rep1	1	Mus musculus	testis	ONT (DRS)	PRJEB45063	ERS5846730	https://www.ebi.ac.uk/ena/browser/view/ERS5846730	Brouze *et al.*[1]	Y	Y
testis_Tent5d_rep2	2	Mus musculus	testis	ONT (DRS)	PRJEB45063	ERS5846732	https://www.ebi.ac.uk/ena/browser/view/ERS5846732	Brouze *et al.*[1]	Y	Y
testis_Tent5d_WT_rep1	1	Mus musculus	testis	ONT (DRS)	PRJEB45063	ERS5465457	https://www.ebi.ac.uk/ena/browser/view/ERS5465457	Brouze *et al.*[1]	Y	Y
testis_Tent5d_WT_rep2	2	Mus musculus	testis	ONT (DRS)	PRJEB45063	ERS5465455	https://www.ebi.ac.uk/ena/browser/view/ERS5465455	Brouze *et al.*[1]	Y	Y
Mouse_testis_spermatids_WT_16	1	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425435	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425435	this work	n	Y
Mouse_testis_spermatids_TENT5C_KO_17	1	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425436	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425436	this work	n	Y
Mouse_testis_spermatids_WT_18	2	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425437	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425437	this work	n	Y
20170221_FamC_ES	1	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425438	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425438	this work	n	Y
20170221_FamC_RS	1	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425439	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425439	this work	n	Y
20170221_WT_ES	1	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425440	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425440	this work	n	Y
20170221_WT_RS	1	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425441	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425441	this work	n	Y
20170223_FamC_ES	2	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425442	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425442	this work	n	Y
20170223_FamC_RS	2	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425443	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425443	this work	n	Y

Continued

sample_name	replicate	species	organ	platform	project	accession	seq_data_link	seq_source	nonA_data	diffexp_data
20170223_WT_ES	2	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425444	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425444	this work	n	Y
20170223_WT_RS	2	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425445	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425445	this work	n	Y
20170406_FamC_ES	3	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425446	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425446	this work	n	Y
20170406_FamC_RS	3	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425447	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425447	this work	n	Y
20170406_WT_ES	3	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425448	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425448	this work	n	Y
20170406_WT_RS	3	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425449	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425449	this work	n	Y
Mouse_testis_spermatids_TENT5C_KO_19	2	Mus musculus	testis	Illumina	PRJNA1141133	GSM8425450	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM8425450	this work	n	Y

Table 1. Metadata of deposited data record. The Ninetails analysis was not performed on Illumina RNA-seq data, because the software is incompatible with the Illumina platform (requires ONT DRS data).

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.

For both sequencing datasets differential expression analysis was performed. RNAseq data DESeq2 statistics was deposited Gene Expression Omnibus (GEO) repository[27] and for DRS dataset in Figshare[26].

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 (Fig. 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 (Fig. 3b). The distributions of read quality tags are similar across different replicates (Fig. 3c).

For the RNA-seq data, we performed correlation analysis, which revealed high homogeneity of the samples (strong positive correlation).

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

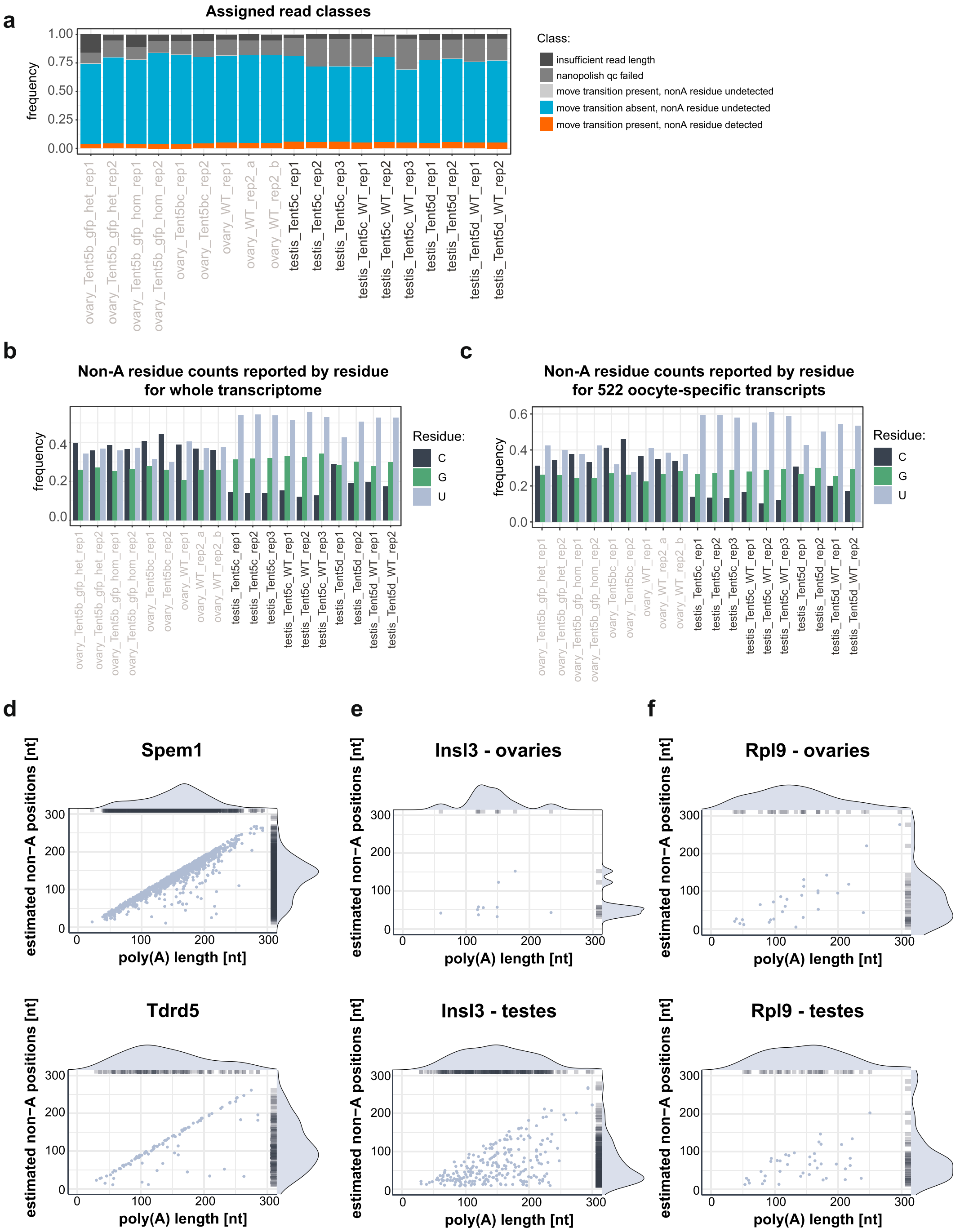


Fig. 4 Distribution of non-adenosine nucleotides in poly(A) tails. (**a**) Detailed read classification made by Ninetails for poly(A) tail composition analysis. (**b**) Frequency of occurrence of cytidine/guanosine/uridine nucleotides in poly(A) tails calculated for whole transcriptome. Only reads classified 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.*[1]. 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.

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 (Fig. 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

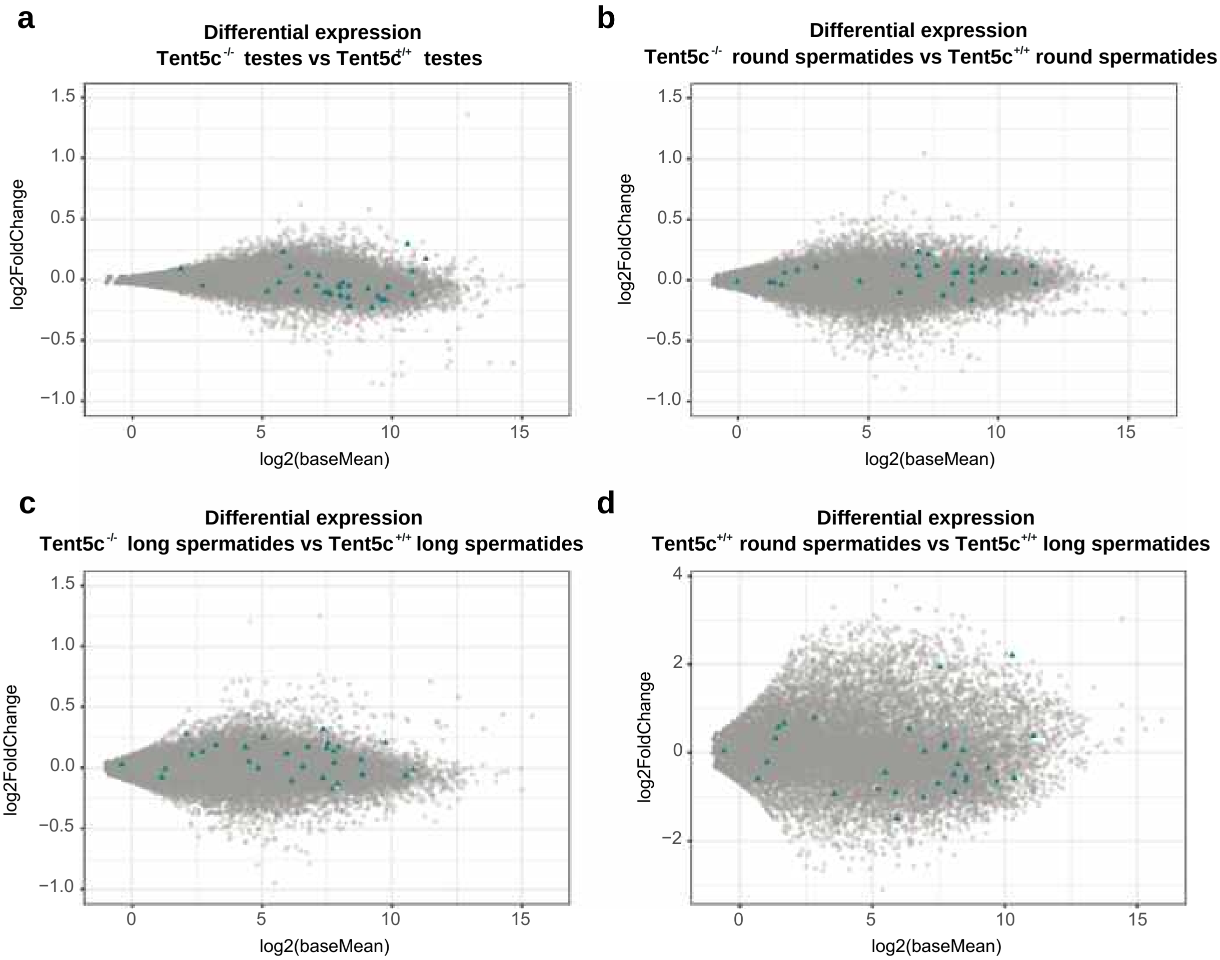


Fig. 5 Differential expression of transcripts with enrichment in uridines at the 3′- end of poly(A) tail. (**a**) RNA-seq performed on testes isolated from Tent5c$^{-/-}$ and Tent5c$^{+/+}$ mice. Most transcripts with enrichment in uridines at the 3′-end of poly(A) tails, highlighted by triangles, are 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.

each non-adenosine nucleotide in all experimental datasets (Fig. 4b) and for 522 oocyte-specific genes described in our previous work[1] (Fig. 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 (Fig. 4b,c).

In samples obtained from murine testes, we identified an overrepresentation of uridines within the heterogeneous poly(A) tails (Fig. 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 Fig. 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 Fig. 4e panel is scattered, indicating a random insertion of uridines in the poly(A) tails. A similar pattern is seen for the *Rpl9* transcript (Fig. 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.

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 (Fig. 5).

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.

Received: 29 May 2024; Accepted: 2 December 2024;

References

1. Brouze, M. *et al*. TENT5-mediated polyadenylation of mRNAs encoding secreted proteins is essential for gametogenesis in mice. *Nat Commun* **15**, 5331, https://doi.org/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*, https://doi.org/10.1111/andr.13589 (2024).
3. Sha, Y. *et al*. TENT5D disruption causes oligoasthenoteratozoospermia and male infertility. *Andrology-Us* **11**, 1121–1131, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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**, 3359, https://doi.org/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–+, https://doi.org/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, https://doi.org/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–+, https://doi.org/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, https://doi.org/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, https://doi.org/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-+, https://doi.org/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-+, https://doi.org/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, https://doi.org/10.1093/nar/gky955 (2019).
23. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574, https://doi.org/10.1093/bioinformatics/btab705 (2021).
24. Danecek, P. *et al*. Twelve years of SAMtools and BCFtools. *Gigascience* **10**. https://doi.org/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, https://doi.org/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 Direct RNA Sequencing. *Figshare* https://doi.org/10.6084/m9.figshare.27052504.v1 (2024).
27. *GEO Gene Expression Omnibus* https://identifiers.org/geo:GSE273264 (2024).
28. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21, https://doi.org/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, https://doi.org/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**, https://doi.org/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 Direct RNA Sequencing and analyzed with Ninetails software. *Figshare* https://doi.org/10.6084/m9.figshare.27051895.v1 (2024).

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://doi.org/10.6084/m9.figshare.27051787.v1 (2024).
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://doi.org/10.6084/m9.figshare.27051682.v1.
34. Zheng, H. *et al*. Lack of Spem1 causes aberrant cytoplasm removal, sperm deformation, and male infertility. *Proc Natl Acad Sci USA* **104**, 6852–6857, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1083/jcb.201009043 (2011).

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

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.D.

Reprints and permissions information is available at www.nature.com/reprints.

4. Discussion

Understanding mRNA poly(A) tail dynamics is essential for deciphering post-transcriptional gene expression regulation. This dissertation examines the landscape of poly(A) tails in two distinct biological contexts: mRNA deadenylation in yeast and cytoplasmic polyadenylation during mouse germline development. While the biological significance of these processes has been extensively discussed in the discussion sections of the respective manuscripts (manuscripts 1 and 2) or intertwined in the body of the paper (manuscript 3), this section focuses on the methodological advances brought by Direct RNA Sequencing (DRS), emphasizing how a single sequencing technique has transformed biological research.

4.1 Interplay Between Deadenylation, Decay, and Translation

Sequencing technology provided by Oxford Nanopore offers a unique advantage over conventional sequencing methods (such as Illumina or PacBio SMRT sequencing) by directly reading native RNA molecules, including full-length transcripts with their poly(A) tails intact. This ability has proven revolutionary for studying mRNA deadenylation kinetics and polyadenylation dynamics. The results of research reported in this dissertation reinforce the fundamental role of the poly(A) tail in mRNA stability and translation efficiency. The yeast study demonstrates that deadenylation is a key determinant of mRNA degradation, governing the transition between translation and decay. Using a numerical model built based on DRS data, the study establishes a constant deadenylation rate of 10 A/min, revealing that transcript half-lives are correlated with codon optimality and functional groupings. Notably, ribosomal protein-coding mRNAs (RPG mRNAs) exhibit unique degradation kinetics, showing that mRNA decay can proceed independently of deadenylation in stress conditions, emphasizing the role of nuclear export in maintaining transcript homeostasis.

The study on mammalian gametogenesis extends this understanding by demonstrating the role of TENT5 poly(A) polymerases in selective mRNA stabilization. Unlike yeast, where deadenylation primarily dictates decay, mammalian germ cells also employ cytoplasmic polyadenylation to regulate translation in a developmental context. The TENT5B and TENT5C enzymes polyadenylate mRNAs encoding secreted proteins, ensuring proper follicular and oocyte development. The loss of TENT5-mediated polyadenylation (knockout mice models) leads to infertility, highlighting the necessity of cytoplasmic poly(A) tail extension for mRNA stability and protein synthesis required for proper development. Capturing changes in the length of poly(A) tails in readenylated transcripts with sufficient resolution was only possible through the use of DRS. At the same time, analysis of this data using the Ninetails tool revealed an enrichment of germ cell transcriptomes with mRNA molecules containing non-canonical nucleotides in their poly(A) tails. Scientific literature provides insights into the role of uridylation in regulating the progression of spermatogenesis. It has also been demonstrated that the presence of non-canonical nucleotides at the 3' end of the poly(A) tail affects the processivity of the Ccr4-NOT deadenylation complex. However, whether a post-transcriptional gene regulation mechanism based on the presence of non-canonical nucleotides in poly(A) tails exists, and if so, how it functions, remains speculative due to technological limitations in the analysis of data obtained from DRS sequencing.

4.2 Advancing mRNA Research Through Nanopore Signal Analysis

A key strength of Oxford Nanopore sequencing lies in its unique signal output. Unlike fluorescence-based sequencing technologies (Illumina, PacBio), which rely on optical detection, Nanopore sequencing records electrical current perturbations. This is conceptually analogous to acoustic signal processing, allowing researchers to apply advanced computational techniques originally developed for speech and sound analysis. Unfortunately, raw electrical nanopore signals are often influenced by base drift (a gradual, systematic shift in the electrical current measurements over time during nanopore sequencing), system noise (random fluctuations and interference in the electrical current measurements that do not represent actual biological signals), and the heterogeneity of through-pore velocities. This makes the signal more challenging to analyse. This applies to both the reading of the nucleotide sequence itself and the identification of nucleic acid modifications. Also, non-canonical nucleotides in poly(A) tails using Nanopore sequencing are challenging due to background noise, which hinders software designed to detect specific signal disruptions. Techniques like signal transformation help mitigate this issue by treating Nanopore readings as overlapping waves. Tools such as BaseNet (Li et al., 2024) use transformer-based algorithms for better signal decoding, while Ninetails applies Fourier Transforms to reduce noise, improving the detection of non-canonical nucleotides in mRNA poly(A) tails.

Another challenge for basecallers in accurately interpreting raw nanopore current signals is nucleotide context. When translating nanopore signals into nucleotide sequences, difficulties arise in estimating the number of identical nucleotides within a homopolymer. In such cases, the recorded signal remains monotonous, and basecallers rely on time and movement functions to determine nucleotide count. However, as nucleic acid molecules do not pass through the pore at a constant speed, these estimates can be inaccurate. A similar issue arises in research on chemical modifications of nucleic acids, an area where Oxford Nanopore sequencing offers an interesting alternative to existing approaches. RNA molecules undergo post-transcriptional chemical modifications that affect their structure and protein interactions. By analysing the shape of the raw nanopore current signal, specialised software can identify the type and location of these modifications. Tools such as Nanocompore and DRUMMER use neural networks to detect specific signal signatures in sequencing reads (Leger et al., 2021; Abebe et al., 2022). However, the accuracy of modification detection is influenced by nucleotide context, which can alter the signal shape, leading software to discard certain sites as unreliable. This results in an underestimation of modifications, affecting structural studies. Additionally, the proximity of modified nucleotides within a sequence impacts detection. When modifications occur adjacent to one another, their combined signal may differ so significantly that neural networks fail to recognize them, further reducing detection accuracy. This limitation also affects Ninetails tool, which identifies non-canonical nucleotides in poly(A) tails. Consequently, the studies described in Manuscript 3 likely underestimate the number of heterogeneous poly(A) tails in germ cell transcriptomes.

It appears that the application of Incremental Learning (IL) may provide a solution to these problems. This machine learning technique allows models to gradually learn new information without the necessity of re-training from scratch with the entire dataset. This is crucial when analysing large datasets in which new patterns (in this case, new types of chemical modifications) may emerge over time. Incremental Learning enables models to evolve and adapt in real-time, making it essential in applications such as anomaly detection and trend analysis in raw nanopore signals (Wang et al., 2024).

The studies presented in this dissertation have extensively utilised the capabilities of Oxford Nanopore sequencing. The resulting datasets were used to model the deadenylation rate in yeast and to identify the targets of TENT5 poly(A) polymerases in mouse germ cells. The research described in the three manuscripts reinforces that mRNA poly(A) tail dynamics are not merely a passive determinant of mRNA stability but an active regulatory element in gene expression, essential for both adaptive cellular responses and developmental processes. Additionally, these studies have identified a potential new regulatory mechanism based on the presence of non-canonical nucleotides in poly(A) tails, which may play a key role in the progression of spermatogenesis. However, further progress in researching this mechanism depends on the advancement of tools for analysing raw nanopore current signals.

5. Future perspectives

The research findings described in this doctoral thesis constitute a study on the role of deadenylation, mRNA decay, and polyadenylation in the remodelling of eukaryotic cell transcriptomes. The ability of cells to adapt to changing environmental conditions is key to their survival. For many years, extensive research has been conducted on stress response pathways, both internal and external to the cell (Kitagaki and Takagi, 2014; Maragkakis et al., 2023; Święciło, 2016). These allowed for describing many mechanisms influencing gene expression changes and characterizing global changes occurring in cell transcriptomes and proteomes (Mühlhofer et al., 2019; Wu et al., 2025; Zhang et al., 2017).

Mass deadenylation and transcript decay constitute one of the key mechanisms allowing rapid reduction of mRNA molecules unnecessary from the perspective of stress response (Hilgers et al., 2006; Suzuki et al., 2022). Their removal enables saturating the pool of translating ribosomes with transcripts carrying information about proteins critical for cell survival. Based on *in vitro* experiments, it was previously assumed that deadenylation rate was constant and depended on the Ccr4-NOT complex mechanics (Hrit et al., 2014; Webster et al., 2017). The results in this dissertation illustrate differences in deadenylation rates between transcripts, introducing more degrees of freedom into the stress response process mechanics. The key questions emerging include: What determines the deadenylation rate for different transcript groups? This problem opens a broad horizon for new research. The observations described above indicate transcript expression level as one of the key factors influencing deadenylation speed. Faster deadenylation of transcripts less significant for cellular functioning is itself an adaptive mechanism. Simultaneously, the slow deadenylation of transcripts encoding ribosomal proteins enables the maintenance of high ribosome synthesis efficiency while limiting the cell's energy expenditure. This is critically important, as ribosome production requires significant resources: engaging three RNA polymerases, expressing over 100 genes, and involving splicing machinery (even in yeast cells). Ribosome mass constitutes the majority of biomass produced by the cell. Deactivating or limiting ribosome biosynthesis is one of the first stages of stress response. In this context, the results in this dissertation regarding the impact of nuclear export on mRNA decay regulation for transcripts encoding ribosomal proteins constitute an interesting starting point for new research on ribosome biosynthesis regulation during stress.

Another interesting aspect of cytoplasmic deadenylation rate regulation involves proteins interacting with deadenylase complexes and mRNA. To date, researchers have described proteins acting as Ccr4-NOT inhibitors (Poetz et al., 2021), Ccr4-NOT interactions with export machinery proteins (Kerr et al., 2011), and interactions between deadenylation complexes and ribosomes (Buschauer et al., 2020). A unique group of Ccr4-NOT interactors are RNA-binding proteins. According to the model described in this dissertation, shortening the tail to an area occupied solely by one Pab1 protein results in immediate decapping and decay. However, the group of such proteins is broader. As described in *Drosophila melanogaster* embryos, Smaug and Cup direct Ccr4-NOT towards *nanos* mRNA deadenylation. Another interesting group consists of evolutionarily conserved proteins containing pummilo domain (yeast puf 1-6, human

Pum 1-2). They recognise specific motifs in mRNA sequences and recruit the Ccr4-NOT complex to these transcripts. In my opinion, post-transcriptional mRNA level regulation through targeted deadenylation is an interesting direction for further research.

Each deadenylation model described so far assumes that detaching each subsequent nucleotide from the poly(A) tail requires the same time and effort from the deadenylating complex. However, this is an oversimplification. Poly(A) tails may contain nucleotides other than adenosine, which influence cytoplasmic deadenylation rates. The results in this dissertation regarding mRNA enrichment with non-homogeneous poly(A) tails in gamete cells illustrate another interesting aspect of deadenylation regulation: poly(A) tail composition. As shown with the COVID-19 vaccine produced by Moderna, the presence of a non-adenosine nucleotide pentamer at the 3’ end of poly(A) tail influences cytoplasmic molecule stability (Krawczyk et al. 2025). I believe that further research on the impact of non-canonical nucleotide type and their distribution in poly(A) tails may bring much benefit to mRNA-based therapeutic development.

Cell maturation exemplifies successive, tightly regulated transcriptome changes. The dissertation's results highlight the critical importance of cytoplasmic polyadenylation in regulating mammalian gametogenesis progression. Male reproductive cell differentiation represents a process aimed at reducing cell transcriptome and proteome. In contrast, female gamete maturation requires specific mechanisms maintaining a stable mRNA pool crucial for fertilisation and zygote development. While the presented results show the key role of non-canonical poly(A) polymerases from the TENT5 family in reproductive cell maturation, they also open fields for speculation about mechanisms regulating polyadenylase function and its substrate specificity. Moreover, the role of non-canonical poly(A) polymerases from the TENT5 family can also be important during fertilisation and embryonic development.

6. Bibliography

Abebe, J.S., Price, A.M., Hayer, K.E., Mohr, I., Weitzman, M.D., Wilson, A.C., Depledge, D.P., 2022. DRUMMER—rapid detection of RNA modifications through comparative nanopore sequencing. Bioinformatics 38, 3113–3115. https://doi.org/10.1093/bioinformatics/btac274

Alles, J., Legnini, I., Pacelli, M., Rajewsky, N., 2023. Rapid nuclear deadenylation of mammalian messenger RNA. iScience 26, 105878. https://doi.org/10.1016/j.isci.2022.105878

Babaian, A., Rothe, K., Girodat, D., Minia, I., Djondovic, S., Milek, M., Spencer Miko, S.E., Wieden, H.-J., Landthaler, M., Morin, G.B., Mager, D.L., 2020. Loss of m1acp3Ψ Ribosomal RNA Modification Is a Major Feature of Cancer. Cell Rep. 31, 107611. https://doi.org/10.1016/j.celrep.2020.107611

Bilska, A., Kusio-Kobiałka, M., Krawczyk, P.S., Gewartowska, O., Tarkowski, B., Kobyłecki, K., Nowis, D., Golab, J., Gruchota, J., Borsuk, E., Dziembowski, A., Mroczek, S., 2020. Immunoglobulin expression and the humoral immune response is regulated by the non-canonical poly(A) polymerase TENT5C. Nat. Commun. 11, 2032. https://doi.org/10.1038/s41467-020-15835-3

Boeck, R., Tarun, S., Rieger, M., Deardorff, J.A., Müller-Auer, S., Sachs, A.B., 1996. The Yeast Pan2 Protein Is Required for Poly(A)-binding Protein-stimulated Poly(A)-nuclease Activity. J. Biol. Chem. 271, 432–438. https://doi.org/10.1074/jbc.271.1.432

Brouze, A., Krawczyk, P.S., Dziembowski, A., Mroczek, S., 2023. Measuring the tail: Methods for poly(A) tail profiling. WIREs RNA 14, e1737. https://doi.org/10.1002/wrna.1737

Burd, C.G., Dreyfuss, G., 1994. Conserved Structures and Diversity of Functions of RNA-Binding Proteins. Science 265, 615–621. https://doi.org/10.1126/science.8036511

Chang, H., Lim, J., Ha, M., Kim, V.N., 2014. TAIL-seq: Genome-wide Determination of Poly(A) Tail Length and 3′ End Modifications. Mol. Cell 53, 1044–1052. https://doi.org/10.1016/j.molcel.2014.02.007

Chen, C.A., Ezzeddine, N., Shyu, A., 2008. Messenger RNA Half-Life Measurements in Mammalian Cells, in: Methods in Enzymology. Elsevier, pp. 335–357. https://doi.org/10.1016/S0076-6879(08)02617-7

Chen, C.A., Shyu, A., 2011. Mechanisms of deadenylation-dependent decay. WIREs RNA 2, 167–183. https://doi.org/10.1002/wrna.40

Clemente, P., Calvo-Garrido, J., Pearce, S.F., Schober, F.A., Shigematsu, M., Siira, S.J., Laine, I., Spåhr, H., Steinmetzger, C., Petzold, K., Kirino, Y., Wibom, R., Rackham, O., Filipovska, A., Rorbach, J., Freyer, C., Wredenberg, A., 2022. ANGEL2 phosphatase activity is required for non-canonical mitochondrial RNA processing. Nat. Commun. 13, 5750. https://doi.org/10.1038/s41467-022-33368-9

Drexler, H.L., Choquet, K., Churchman, L.S., 2020. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. Mol. Cell 77, 985-998.e8. https://doi.org/10.1016/j.molcel.2019.11.017

Dunn, E.F., Hammell, C.M., Hodge, C.A., Cole, C.N., 2005. Yeast poly(A)-binding protein, Pab1, and PAN, a poly(A) nuclease complex recruited by Pab1, connect mRNA biogenesis to export. Genes Dev. 19, 90–103. https://doi.org/10.1101/gad.1267005

Duran-Arqué, B., Cañete, M., Castellazzi, C.L., Bartomeu, A., Ferrer-Caelles, A., Reina, O., Caballé, A., Gay, M., Arauz-Garofalo, G., Belloc, E., Mendez, R., 2022. Comparative analyses of vertebrate CPEB proteins define two subfamilies with coordinated yet distinct functions in post-transcriptional gene regulation. Genome Biol. 23, 192. https://doi.org/10.1186/s13059-022-02759-y

Eichhorn, S.W., Subtelny, A.O., Kronja, I., Kwasnieski, J.C., Orr-Weaver, T.L., Bartel, D.P., 2016. mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in Drosophila oocytes and early embryos. eLife 5, e16955. https://doi.org/10.7554/eLife.16955

Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., Lin, K.S., McGeary, S.E., Gupta, S., Bartel, D.P., 2020. The Dynamics of Cytoplasmic mRNA Metabolism. Mol. Cell 77, 786-799.e10. https://doi.org/10.1016/j.molcel.2019.12.005

Ermisch, A.F., Wood, J.R., 2024. Regulation of Oocyte mRNA Metabolism: A Key Determinant of Oocyte Developmental Competence, in: Balboula, A.Z. (Ed.), Molecular Mechanisms Determining Mammalian Oocyte Quality, Advances in Anatomy, Embryology and Cell Biology. Springer International Publishing, Cham, pp. 23–46. https://doi.org/10.1007/978-3-031-55163-5_2

Garalde, D.R., Snell, E.A., Jachimowicz, D., Sipos, B., Lloyd, J.H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., Jordan, M., Ciccone, J., Serra, S., Keenan, J., Martin, S., McNeill, L., Wallace, E.J., Jayasinghe, L., Wright, C., Blasco, J., Young, S., Brocklebank, D., Juul, S., Clarke, J., Heron, A.J., Turner, D.J., 2018. Highly parallel direct RNA sequencing on an array of nanopores. Nat. Methods 15, 201–206. https://doi.org/10.1038/nmeth.4577

Gewartowska, O., Aranaz-Novaliches, G., Krawczyk, P.S., Mroczek, S., Kusio-Kobiałka, M., Tarkowski, B., Spoutil, F., Benada, O., Kofroňová, O., Szwedziak, P., Cysewski, D., Gruchota, J., Szpila, M., Chlebowski, A., Sedlacek, R., Prochazka, J., Dziembowski, A., 2021. Cytoplasmic polyadenylation by TENT5A is required for proper bone formation. Cell Rep. 35, 109015. https://doi.org/10.1016/j.celrep.2021.109015

Gumińska N., Matylla-Kulińska K., Krawczyk P. S., Maj M., Orzeł W., Mackiewicz Z., Brouze A., Mroczek S., Dziembowski A., 2025. Direct profiling of non-adenosines in poly(A) tails of endogenous and therapeutic mRNAs with Ninetails. Nat. Commun. (in press) https://doi.org/10.1038/s41467-025-57787-6

Hommelsheim, C.M., Frantzeskakis, L., Huang, M., Ülker, B., 2014. PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. Sci. Rep. 4, 5052. https://doi.org/10.1038/srep05052

Huang, Y.-S., Mendez, R., Fernandez, M., Richter, J.D., 2023. CPEB and translational control by cytoplasmic polyadenylation: impact on synaptic plasticity, learning, and memory. Mol. Psychiatry 28, 2728–2736. https://doi.org/10.1038/s41380-023-02088-x

Huynh, T.N., Parker, R., 2023. The PARN, TOE1, and USB1 RNA deadenylases and their roles in non-coding RNA regulation. J. Biol. Chem. 299, 105139. https://doi.org/10.1016/j.jbc.2023.105139

Jain, M., Abu-Shumays, R., Olsen, H.E., Akeson, M., 2022. Advances in nanopore direct RNA sequencing. Nat. Methods 19, 1160–1164. https://doi.org/10.1038/s41592-022-01633-w

Khanna, R., Kiledjian, M., 2004. Poly(A)-binding-protein-mediated regulation of hDcp2 decapping in vitro. EMBO J. 23, 1968–1976. https://doi.org/10.1038/sj.emboj.7600213

Kim, D., Lee, Y., Jung, S.-J., Yeo, J., Seo, J.J., Lee, Y.-Y., Lim, J., Chang, H., Song, J., Yang, J., Kim, J.-S., Jung, G., Ahn, K., Kim, V.N., 2020. Viral hijacking of the TENT4–ZCCHC14 complex protects viral RNAs via mixed tailing. Nat. Struct. Mol. Biol. 27, 581–588. https://doi.org/10.1038/s41594-020-0427-3

Krawczyk, P., Mazur M., Orzeł W., Gewartowska O., Jeleń S., Antczak W., Kasztelan K., Brouze A., Matylla-Kulińska K., Gumińska N., Tarkowski B., Owczarek E., Affek K., Turowski P., Tudek A., Sroka M., Śpiewla T., Kusio-Kobiałka M., Wesołowska A., Nowis D., Gołąb J., Kowalska J., Jemielity J., Mroczek S., Dziembowski, A., 2025. Re-adenylation by TENT5A enhances efficacy of SARS-CoV-2 mRNA vaccines. Nature (in press), DOI: 10.1038/s41586-025-08842-1

Kwiatek, L., Landry-Voyer, A.-M., Latour, M., Yague-Sanz, C., Bachand, F., 2023. PABPN1 prevents the nuclear export of an unspliced RNA with a constitutive transport element and controls human gene expression via intron retention. RNA 29, 644–662. https://doi.org/10.1261/rna.079294.122

Łabno, A., Tomecki, R., Dziembowski, A., 2016. Cytoplasmic RNA decay pathways - Enzymes and mechanisms. Biochim. Biophys. Acta BBA - Mol. Cell Res. 1863, 3125–3147. https://doi.org/10.1016/j.bbamcr.2016.09.023

Lapointe, C.P., Preston, M.A., Wilinski, D., Saunders, H.A.J., Campbell, Z.T., Wickens, M., 2017. Architecture and dynamics of overlapped RNA regulatory networks. RNA 23, 1636–1647. https://doi.org/10.1261/rna.062687.117Lee, Y., Levdansky, Y., Jung, Y., Kim, V.N., Valkov, E., 2024. Deadenylation kinetics of mixed poly(A) tails at single-nucleotide resolution. Nat. Struct. Mol. Biol. 31, 826–834. https://doi.org/10.1038/s41594-023-01187-1

Leger, A., Amaral, P.P., Pandolfini, L., Capitanchik, C., Capraro, F., Miano, V., Migliori, V., Toolan-Kerr, P., Sideri, T., Enright, A.J., Tzelepis, K., Van Werven, F.J., Luscombe, N.M., Barbieri, I., Ule, J., Fitzgerald, T., Birney, E., Leonardi, T., Kouzarides, T., 2021. RNA modifications detection by comparative Nanopore direct RNA sequencing. Nat. Commun. 12, 7198. https://doi.org/10.1038/s41467-021-27393-3

Li, Y., Misumi, I., Shiota, T., Sun, L., Lenarcic, E.M., Kim, H., Shirasaki, T., Hertel-Wulff, A., Tibbs, T., Mitchell, J.E., McKnight, K.L., Cameron, C.E., Moorman, N.J., McGivern, D.R., Cullen, J.M., Whitmire, J.K., Lemon, S.M., 2022. The ZCCHC14/TENT4 complex is required for hepatitis A virus RNA synthesis. Proc. Natl. Acad. Sci. 119, e2204511119. https://doi.org/10.1073/pnas.2204511119

Li, Q., Sun, C., Wang, D., Lou, J., 2024. BaseNet: A transformer-based toolkit for nanopore sequencing signal decoding. Comput. Struct. Biotechnol. J. 23, 3430–3444. https://doi.org/10.1016/j.csbj.2024.09.016

Lim, J., Ha, M., Chang, H., Kwon, S.C., Simanshu, D.K., Patel, D.J., Kim, V.N., 2014. Uridylation by TUT4 and TUT7 Marks mRNA for Degradation. Cell 159, 1365–1376. https://doi.org/10.1016/j.cell.2014.10.055

Lim, J., Kim, D., Lee, Y., Ha, M., Lee, M., Yeo, J., Chang, H., Song, J., Ahn, K., Kim, V.N., 2018. Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. Science 361, 701–704. https://doi.org/10.1126/science.aam5794

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

Liudkovska, V., Krawczyk, P.S., Brouze, A., Gumińska, N., Wegierski, T., Cysewski, D., Mackiewicz, Z., Ewbank, J.J., Drabikowski, K., Mroczek, S., Dziembowski, A., 2022. TENT5 cytoplasmic noncanonical poly(A) polymerases regulate the innate immune response in animals. Sci. Adv. 8, eadd9468. https://doi.org/10.1126/sciadv.add9468

Mangus, D.A., Evans, M.C., Jacobson, A., 2003. Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. Genome Biol. 4, 223. https://doi.org/10.1186/gb-2003-4-7-223

Mansur, F., Alarcon, J.M., Stackpole, E.E., Wang, R., Richter, J.D., 2021. Noncanonical cytoplasmic poly(A) polymerases regulate RNA levels, alternative RNA processing, and synaptic plasticity but not hippocampal-dependent behaviours. RNA Biol. 18, 962–971. https://doi.org/10.1080/15476286.2020.1824061

Menezes, M.R., Balzeau, J., Hagan, J.P., 2018. 3′ RNA Uridylation in Epitranscriptomics, Gene Regulation, and Disease. Front. Mol. Biosci. 5, 61. https://doi.org/10.3389/fmolb.2018.00061

Miller, C., Schwalb, B., Maier, K., Schulz, D., Dümcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dölken, L., Martin, D.E., Tresch, A., Cramer, P., 2011. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol. Syst. Biol. 7, 458. https://doi.org/10.1038/msb.2010.112

Mitchell, P., Tollervey, D., 2000. mRNA stability in eukaryotes. Curr. Opin. Genet. Dev. 10, 193–198. https://doi.org/10.1016/S0959-437X(00)00063-0

Morgan, M., Kabayama, Y., Much, C., Ivanova, I., Di Giacomo, M., Auchynnikava, T., Monahan, J.M., Vitsios, D.M., Vasiliauskaitė, L., Comazzetto, S., Rappsilber, J., Allshire, R.C., Porse, B.T., Enright, A.J., O’Carroll, D., 2019. A programmed wave of uridylation-primed mRNA degradation is essential for meiotic progression and mammalian spermatogenesis. Cell Res. 29, 221–232. https://doi.org/10.1038/s41422-018-0128-1

Morgan, M., Much, C., DiGiacomo, M., Azzi, C., Ivanova, I., Vitsios, D.M., Pistolic, J., Collier, P., Moreira, P.N., Benes, V., Enright, A.J., O’Carroll, D., 2017. mRNA 3′ uridylation and poly(A) tail length sculpt the mammalian maternal transcriptome. Nature 548, 347–351. https://doi.org/10.1038/nature23318

Mulroney, L., Wulf, M.G., Schildkraut, I., Tzertzinis, G., Buswell, J., Jain, M., Olsen, H., Diekhans, M., Corrêa, I.R., Akeson, M., Ettwiller, L., 2022. Identification of high-confidence human poly(A) RNA isoform scaffolds using nanopore sequencing. RNA 28, 162–176. https://doi.org/10.1261/rna.078703.121

Neymotin, B., Athanasiadou, R., Gresham, D., 2014. Determination of in vivo RNA kinetics using RATE-seq. RNA 20, 1645–1652. https://doi.org/10.1261/rna.045104.114

Nicholson-Shaw, A.L., Kofman, E.R., Yeo, G.W., Pasquinelli, A.E., 2022. Nuclear and cytoplasmic poly(A) binding proteins (PABPs) favor distinct transcripts and isoforms. Nucleic Acids Res. 50, 4685–4702. https://doi.org/10.1093/nar/gkac263

Preiss, T., Hentze, M.W., 1998. Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast. Nature 392, 516–520. https://doi.org/10.1038/33192

Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., Coller, J., 2015. Codon Optimality Is a Major Determinant of mRNA Stability. Cell 160, 1111–1124. https://doi.org/10.1016/j.cell.2015.02.029

Proweller, A., Butler, S., 1994. Efficient translation of poly(A)-deficient mRNAs in Saccharomyces cerevisiae. Genes Dev. 8, 2629–2640. https://doi.org/10.1101/gad.8.21.2629

Qiu, C., Dutcher, R.C., Porter, D.F., Arava, Y., Wickens, M., Hall, T.M.T., 2019. Distinct RNA-binding modules in a single PUF protein cooperate to determine RNA specificity. Nucleic Acids Res. gkz583. https://doi.org/10.1093/nar/gkz583

Read, R.L., Norbury, C.J., 2002. Roles for cytoplasmic polyadenylation in cell cycle regulation. J. Cell. Biochem. 87, 258–265. https://doi.org/10.1002/jcb.10300

Rissland, O.S., Mikulasova, A., Norbury, C.J., 2007. Efficient RNA Polyuridylation by Noncanonical Poly(A) Polymerases. Mol. Cell. Biol. 27, 3612–3624. https://doi.org/10.1128/MCB.02209-06

Roach, N.P., Sadowski, N., Alessi, A.F., Timp, W., Taylor, J., Kim, J.K., 2020. The full-length transcriptome of *C. elegans* using direct RNA sequencing. Genome Res. 30, 299–312. https://doi.org/10.1101/gr.251314.119

Rodríguez-Molina, J.B., Turtola, M., 2023. Birth of a poly(A) tail: mechanisms and control of MRNA polyadenylation. FEBS Open Bio 13, 1140–1153. https://doi.org/10.1002/2211-5463.13528

Rouhana, L., Edgar, A., Hugosson, F., Dountcheva, V., Martindale, M.Q., Ryan, J.F., 2023. Cytoplasmic Polyadenylation Is an Ancestral Hallmark of Early Development in Animals. Mol. Biol. Evol. 40, msad137. https://doi.org/10.1093/molbev/msad137

Sachs, A.B., Bond, M.W., Kornberg, R.D., 1986. A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: Domain structure and expression. Cell 45, 827–835. https://doi.org/10.1016/0092-8674(86)90557-X

Searfoss, A.M., Wickner, R.B., 2000. 3′ poly(A) is dispensable for translation. Proc. Natl. Acad. Sci. 97, 9133–9137. https://doi.org/10.1073/pnas.97.16.9133

Sehn, J.K., 2015. Insertions and Deletions (Indels), in: Clinical Genomics. Elsevier, pp. 129–150. https://doi.org/10.1016/B978-0-12-404748-8.00009-5

Soheilypour, M., Mofrad, M.R.K., 2016. Regulation of RNA-binding proteins affinity to export receptors enables the nuclear basket proteins to distinguish and retain aberrant mRNAs. Sci. Rep. 6, 35380. https://doi.org/10.1038/srep35380

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., Bartel, D.P., 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature 508, 66–71. https://doi.org/10.1038/nature13007

Tucker, M., Valencia-Sanchez, M.A., Staples, R.R., Chen, J., Denis, C.L., Parker, R., 2001. The Transcription Factor Associated Ccr4 and Caf1 Proteins Are Components of the Major Cytoplasmic mRNA Deadenylase in Saccharomyces cerevisiae. Cell 104, 377–386. https://doi.org/10.1016/S0092-8674(01)00225-2

Tudek, A., Krawczyk, P.S., Mroczek, S., Tomecki, R., Turtola, M., Matylla-Kulińska, K., Jensen, T.H., Dziembowski, A., 2021. Global view on the metabolism of RNA poly(A) tails in yeast Saccharomyces cerevisiae. Nat. Commun. 12, 4951. https://doi.org/10.1038/s41467-021-25251-w

Wahle, E., 1999. 3′-End processing of pre-mRNA in eukaryotes. FEMS Microbiol. Rev. 23, 277–295. https://doi.org/10.1016/S0168-6445(99)00008-X

Wang, M., Ogé, L., Perez-Garcia, M.-D., Hamama, L., Sakr, S., 2018. The PUF Protein Family: Overview on PUF RNA Targets, Biological Functions, and Post Transcriptional Regulation. Int. J. Mol. Sci. 19, 410. https://doi.org/10.3390/ijms19020410

Wang, Z., Fang, Y., Liu, Z., Hao, N., Zhang, H.H., Sun, X., Que, J., Ding, H., 2024. Adapting nanopore sequencing basecalling models for modification detection via incremental learning and anomaly detection. Nat. Commun. 15, 7148. https://doi.org/10.1038/s41467-024-51639-5

Wang, Z., Kiledjian, M., 2000. The Poly(A)-Binding Protein and an mRNA Stability Protein Jointly Regulate an Endoribonuclease Activity. Mol. Cell. Biol. 20, 6334–6341. https://doi.org/10.1128/MCB.20.17.6334-6341.2000

Warkocki, Z., Liudkovska, V., Gewartowska, O., Mroczek, S., Dziembowski, A., 2018. Terminal nucleotidyl transferases (TENTs) in mammalian RNA metabolism. Philos. Trans. R. Soc. B Biol. Sci. 373, 20180162. https://doi.org/10.1098/rstb.2018.0162

Webster, M.W., Chen, Y.-H., Stowell, J.A.W., Alhusaini, N., Sweet, T., Graveley, B.R., Coller, J., Passmore, L.A., 2018. mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. Mol. Cell 70, 1089-1100.e8. https://doi.org/10.1016/j.molcel.2018.05.033

Webster, M.W., Stowell, J.A.W., Tang, T.T.L., Passmore, L.A., 2017. Analysis of mRNA deadenylation by multi-protein complexes. Methods 126, 95–104. https://doi.org/10.1016/j.ymeth.2017.06.009

Wilinski, D., Qiu, C., Lapointe, C.P., Nevil, M., Campbell, Z.T., Tanaka Hall, T.M., Wickens, M., 2015. RNA regulatory networks diversified through curvature of the PUF protein scaffold. Nat. Commun. 6, 8213. https://doi.org/10.1038/ncomms9213

Wolf, J., Valkov, E., Allen, M.D., Meineke, B., Gordiyenko, Y., McLaughlin, S.H., Olsen, T.M., Robinson, C.V., Bycroft, M., Stewart, M., Passmore, L.A., 2014. Structural basis for Pan3 binding to Pan2 and its function in mRNA recruitment and deadenylation. EMBO J. 33, 1514–1526. https://doi.org/10.15252/embj.201488373

Workman, R.E., Tang, A.D., Tang, P.S., Jain, M., Tyson, J.R., Razaghi, R., Zuzarte, P.C., Gilpatrick, T., Payne, A., Quick, J., Sadowski, N., Holmes, N., De Jesus, J.G., Jones, K.L., Soulette, C.M., Snutch, T.P., Loman, N., Paten, B., Loose, M., Simpson, J.T., Olsen, H.E., Brooks, A.N., Akeson, M., Timp, W., 2019. Nanopore native RNA sequencing of a human poly(A) transcriptome. Nat. Methods 16, 1297–1305. https://doi.org/10.1038/s41592-019-0617-2

Wu, C., Roy, B., He, F., Yan, K., Jacobson, A., 2020. Poly(A)-Binding Protein Regulates the Efficiency of Translation Termination. Cell Rep. 33, 108399. https://doi.org/10.1016/j.celrep.2020.108399

Zhao, J., Hyman, L., Moore, C., 1999. Formation of mRNA 3′ Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis. Microbiol. Mol. Biol. Rev. 63, 405–445. https://doi.org/10.1128/MMBR.63.2.405-445.1999