TGS1 controls snRNA 3' end processing, prevents neurodegeneration and ameliorates *SMN*-dependent neurological phenotypes in vivo

Lu Chen[1,2,3*], Caitlin M. Roake[1,2,3*], Paolo Maccallini[4], Francesca Bavasso[4], Roozbeh Dehghannasiri[3], Pamela Santonicola[5], Natalia Mendoza-Ferreira[6], Livia Scatolini[4], Ludovico Rizzuti[4], Alessandro Esposito[7], Ivan Gallotta[7], Sofia Francia[8,9]; Stefano Cacchione[4], Matthias Hammerschmidt[10], Cristiano De Pittà[11], Gabriele Sales[11], Julia Salzman[1,3], Livio Pellizzoni[12,13,14], Brunhilde Wirth[6,15], Elia Di Schiavi[5,7], Maurizio Gatti[4,16], Steven E. Artandi[1,2,3] and Grazia D. Raffa[4, 17]

[1] Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94305, USA; [2] Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA; [3] Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, USA; [4] Dipartimento di Biologia e Biotecnologie, Sapienza University of Rome, Rome, IT; [5] Institute of Biosciences and BioResources, IBBR, CNR, Naples, Italy; [6] Institute of Human Genetics, Center for Molecular Medicine Cologne, Institute for Genetics, Univ. of Cologne, 50931 Cologne, Germany; [7]Institute of Genetics and Biophysics, IGB-ABT, CNR, Naples, Italy; [8]IFOM-The FIRC Institute of Molecular Oncology, Milan, Italy; [9]Istituto di Genetica Molecolare, CNR-Consiglio Nazionale delle Ricerche, Pavia, Italy; [10]Institute for Zoology, Developmental Biology, University of Cologne, 50674 Cologne, Germany; [11]Department of Biology, University of Padova, Padua, Italy; [12,]Center for Motor Neuron Biology and Disease, Columbia University, New York, NY 10032, USA; [13]Department of Pathology and Cell Biology, Columbia University, New York, NY 10032, USA; [14]Department of Neurology, Columbia University, New York, NY 10032, USA; [15]Center for Rare Diseases, University Hospital of Cologne, Univ. of Cologne, 50931 Cologne, Germany;[16]Istituto di Biologia e Patologia Molecolari (IBPM) del CNR, Rome, IT, [17]Lead Contact

*These authors contributed equally

Correspondence

Steven E. Artandi: sartandi@stanford.edu

Grazia D. Raffa: graziadaniela.raffa@uniroma1.it

ABSTRACT

Trimethylguanosine synthase 1 (TGS1) is a highly conserved enzyme that converts the 5' mono-methylguanosine cap of snRNAs to a trimethylguanosine cap. Here, we show that loss of TGS1 in *C. elegans, D. melanogaster* and *D. rerio* results in neurological phenotypes similar to those caused by Survival Motor Neuron (SMN) deficiency. Importantly, expression of human TGS1 ameliorates the *SMN*-dependent neurological phenotypes in both flies and worms, revealing that TGS1 has the ability to counteract the effects of SMN deficiency. TGS1 loss in HeLa cells leads to the accumulation of immature U2 and U4atac snRNAs with long 3' tails that are often uridylated. snRNAs with defective 3' terminations also accumulate in *Drosophila Tgs1* mutants. Consistent with defective snRNA maturation, *TGS1* and *SMN* mutant cells also exhibit partially overlapping transcriptome alterations. Together, these results identify a neuroprotective function for TGS1, reinforcing the view that defects in snRNA maturation negatively affect neuronal viability and function.

INTRODUCTION

The correct processing of small noncoding RNAs is essential to prevent severe neurodegenerative conditions (1). Small nuclear RNAs (snRNAs) are components of the spliceosomal snRNPs. Defects in the snRNP assembly pathway have been observed in Spinal Muscular Atrophy (SMA), a neurological disease caused by deficiency of the Survival Motor Neuron (SMN) protein (2). Another neurodegenerative disorder, Pontocerebellar hypoplasia type 7 (PCH7), is caused by mutations in the TOE1 deadenylase that prevents accumulation of 3'-end-extended pre-snRNAs (3).

The Sm-class snRNAs include components of both the major (U1, U2, U4, and U5) and the minor (U11, U12, U4atac) spliceosome. These snRNAs are transcribed by RNA polymerase II and the formation of their 3'- ends is co-transcriptionally controlled (4,5). Nascent U1 and U2 snRNA transcripts are subject to cleavage by the Integrator complex, resulting in precursors carrying extended 3'ends that are trimmed after export to the cytoplasm (6-8).

All snRNAs transcribed by Pol II acquire an m7 monomethylguanosine cap (m7G/MMG cap) at their 5' end. Immediately after transcription, the monomethylated snRNA precursors bind the Cap Binding Complex (CBC), and are exported to the cytoplasm (9,10), where they bind the Gemin5 subunit of the SMN complex (11) that mediates their assembly with the seven Sm proteins (Sm core) to form snRNPs (12). The snRNA precursors associated with Gemin5 have extended 3' tails that can be oligoadenylated or uridylated (11). The U2 snRNA precursors have 3' extended tails including 12 genomic-encoded nucleotides, which are removed by the TOE1 deadenylase (3).

Following interaction with SMN and the Sm core, the MMG cap of snRNAs is converted to a 2,2,7-trymethylguanosine cap (TMG) by Trimethylguanosine Synthase 1 (TGS1) (13), which physically interacts with SmB and SMN (14). Re-entry of the snRNPs into the nucleus is mediated by both the TMG cap and the Sm core (15,16). In somatic cells of *Xenopus* and mammals, the Sm signal is both necessary and sufficient for nuclear import, while the TMG cap accelerates the rate of import (17-19). Once in the nucleus, the snRNPs concentrate in the Cajal bodies (CBs) and undergo further maturation steps (20).

3' end processing of snRNAs is thought to occur after TMG cap formation (21) but whether TGS1 and cap hypermethylation have roles in snRNP processing is unknown. This is a relevant biomedical issue, as abnormal accumulations of 3' end extended snRNA precursors have been observed in PCH7 patients (3) and in cells deficient for SMN (11). To elucidate the role of *TGS1* in snRNA maturation and its functional relationships with *SMN*, here we explore the phenotypic consequences of TGS1 loss in human cells and in the *Drosophila*, C. *elegans* and *D. rerio* model organisms. Collectively our results indicate that TGS1 plays a conserved function, acting as a neuroprotective factor that can counteract the phenotypic consequences of SMN deficiency.

MATERIALS AND METHODS

C.elegans strains

Nematodes were grown and handled following standard procedures, under uncrowded conditions, at 20°C, on NGM (Nematode Growth Medium) agar plates seeded with *Escherichia coli* strain OP50(22). Wild-type animals used in this work were *C. elegans* variety Bristol, strain N2. The transgenic strains used in this work were: NA1330 *gbIs4* [GBF109 p*unc-25::cesmn-1(RNAi)*(200 ng/µL); GB301 p*chs-2::GFP* (10 ng/µL)] III; NA1524 *gbEx571* [GBF309 p*unc-25::cesmn-1(RNAi)* (10 ng/µL); GB301 p*chs-2::GFP* (10 ng/µL)]; NA1541 *gbEx568* [GBF321 p*unc-25::kal-1(RNAi)* (10 ng/µL); GB301 p*chs-2::GFP* (10 ng/µL)]; NA1252 *gbEx504* [GBF322 p*unc-119*::*dsRED* (20 ng/µL); p*elt-2*::RFP (30 ng/µL)](23); EG1285 *oxIs12* [p*unc-47*::GFP; *lin-15*(+)] X. EG2185 and N2 were provided by the *Caenorhabditis* Genetics Center (CGC), funded by NIH Office of Research Infrastructure Programs (P40 OD010440); *Is[*p*unc-47::RFP]* was kindly provided by K. Shen (Stanford University, USA).
A complete description of the transgenes obtained in this work is reported in the table below.

Strain	**Transgene name**	**Transgene description**	**Notes**
NA1539	*gbEx579*	[GBF332 p*unc-25*::*tgs-1*(RNAi)(10 ng/µL); GB301 p*chs-2*::GFP (10 ng/µL)]	*tgs-1* knock-down in D-type MNs
NA1972	*gbEx647*	[GB335 p*unc-119*::*hTGS1* (20 ng/µL); p*odr-1*::RFP (30 ng/µL)]	*hTGS1* overexpression in all neurons
	gbEx648	[GB338 p*unc-119*::*cetgs-1* (20 ng/µL); p*odr-1*::RFP (30 ng/µL)]	*ceTGS-1* overexpression in all neurons
NA1533	*gbEx576*	[GBF329 *pcetgs-1a*::GFP (50 ng/µl)]	GFP expression under *cetgs-1* isoform *a* promoter
NA1535	*gbEx577*	[GBF330 p*cetgs-1b*::GFP (50 ng/µl)]	GFP expression under *cetgs-1* isoform *b* promoter

Transgenic strains

The constructs GBF329(p*cetgs-1a*::GFP) and GBF330 (p*cetgs-1b*::GFP) for analyzing the expression of *cetgs-1* isoforms were created by PCR-fusion(24) of two DNA fragments: the presumptive promoters of *cetgs-1* and the GFP sequence. The two promoters of *cetgs-1* isoforms (*a* or *b*) have been chosen upstream of the ATG of the two isoforms, to span 360 bp (*cetgs-1a*) and 960 bp(*cetgs-1b*), respectively. The putative regulatory regions were amplified by PCR using as template wild-type genomic DNA. The GFP, followed by the 3'UTR of the *unc-54* gene to increase the stability of the constructs, was amplified from plasmid pPD95.75, kindly provided by A. Fire (Stanford University, USA). The rescue plasmids (GB335 p*unc-119*:: *hTGS1* and GB338 p*unc-119*::*cetgs-1*) for pan-neuronal expression were created by cloning *hTGS1* or *cetgs-1* CDS into the plasmid pBY103, kindly provided by O. Hobert (Columbia University, New York, USA), which contains the pan-neuronal promoter p*unc-119*. *hTGS1* and *cetgs-1* CDS were amplified by PCR from cDNA libraries and cloned in BamHI site of pBY103. D-type motoneuron-specific RNAi transgenic lines were obtained as previously described(23,25). To obtain the specific knock-down of *cetgs-1* we used a short form of *unc-25*/GAD promoter (180 bp), which is specifically expressed from embryonic to adult stages only in the 19 D-type motor neurons in the ventral cord and not in other GABAergic neurons (RME, AVL, DVB and RIS)(23,26). For gene silencing we amplified, from genomic DNA, the same exon-rich region that was used for the RNAi plasmid library prepared for RNAi by feeding experiments (27). Exon-rich regions were amplified in two separate PCR reactions to obtain the sense and antisense fragments. The *unc-25* promoter was amplified using specific primers. The promoter was subsequently fused to each orientation of the target gene by PCR fusion using internal primers. All primers are available on request. Germ line transformation was accomplished as described(28) by injecting into the gonad of adult animals a DNA mixture containing a transgenic construct together with a phenotypic marker for selection of transgenic progeny. As co-injection marker we used p*odr-1::*RFP, kindly provided by C. Bargmann, The Rockfeller University, New York, USA (RFP expression in AWC and AWB neurons) at 30 ng/µL and GB301 p*chs-2*::GFP (GFP expression in the glandular pharyngeal g1 and g2 cells, and in m3 and m4 myoepithelial cells)(29) at 10 ng/µL. At least two

independent transgenic lines were examined for each experiment. Genetic crosses were made to transfer transgenes to the appropriate genetic background. In all cases, the presence of the desired transgenes was verified by phenotyping. Two independent clones with the same genotype were examined after each cross and the mean of the two clones has been reported in the results.

Behavioral assay

Well-fed, young adult animals were used for backward movement assay (30) to test D-type motor neurons function. The assay was performed blind on NGM plates, 6 cm in diameter, seeded with bacteria. Using an eyelash the animal was touched first on the tail to induce a forward movement and then on the head to test for backward movement. A defective movement was scored when animals were unable to fully move backward. For each data set, the percentage of animals with normal locomotion among the total number of tested animals was calculated. One-way ANOVA test was used for statistical analysis.

***C.elegans* microscopy analysis**

Animals were immobilized in 0.01% tetramisole hydrochloride (Sigma-Aldrich) on 4% agar pads and visualized using Zeiss Axioskop or Leica DM6B microscopes. All microscopes were equipped with epifluorescence and DIC Nomarski optics and images were collected with Leica digital cameras Hamamatsu C11440. Confocal images have been collected using DMi8 confocal microscope.

Zebrafish experiments

Zebrafish procedures were approved by the local animal protection committee LANUV NRW; reference 84-02.04.2012.A251. Experiments for caudal primary motor neuron (CaP-MN) analysis were performed in the wild-type TL/EK (*Tupfel long fin* / *Ekkwill*) line.

Zebrafish injection

Control (non-targeting) and *tgs1* MOs were purchased from (GeneTools, LLC) using the *tgs1* XM_003197865.5 sequence as reference. MO sequences are reported in table S6. For TGS1 mRNA injection, human *TGS1* cDNA (NM_001317902) was cloned into an N-

terminal flag-pCS2+ mRNA expression vector. In vitro transcription of capped-mRNAs was performed using the mMACHINE SP6 Transcription Kit (Ambion) following manufacturer's protocol and as previously described in (31,32) . Embryos from TL/EK wild-type crossings were used to visualize the CaP-MN phenotype. Embryos were injected with the respective dose of MOs or mRNA in and aqueous solution containing 0.05% PhenolRed and 0.05% Rhodamine-Dextran (Sigma-Aldrich). 6-7hr after injection, embryos were sorted according to homogeneity of the rhodamine fluorescence signal.

Semi-quantitative RT-PCR

RNA isolation from zebrafish larvae (~34 hours post fertilization - hpf) was performed following the procedure described in (33). RNA was extracted using the RNeasy kit (Qiagen) and concentrations were determined by RiboGreen method (Life Technologies). 600ng RNA were reversely transcribed to cDNA with the Quantitect Reverse Transcription Kit (Qiagen). PCR was performed with 35 cycles to allow detection of the different *tgs1* transcripts and as multiplex together with the housekeeping gene *ef1a* (*elongation factor 1-alpha*). Primers were designed using Primer-BLAST and purchased from Integrated DNA technologies (IDT). Primers sequences (Drer-tgs1 and Drer ef1α) are reported in table S6. Amplification products from semi-quantitative RT-PCR were gel-purified using the QIAquick gel extraction kit (Qiagen) and Sanger sequenced. Densitometric analyses were performed with ImageLab 5.2.1 (BioRAD).

Immunostaining of zebrafish caudal primary motor neurons

Zebrafish larvae immunostaining was performed as previously described in (31,32). Briefly, ~34 hpf zebrafish were manually dechorionated, fixed in 4% PFA-PBS, and permeabilized by Proteinase K digestion of whole larvae. To visualize CaP-MNs, fish larvae were blocked in blocking solution (PBS 0.1% Tween + 5% FCS, 2% BSA and 1% DMSO) and posteriorly incubated overnight at 4°C with mouse α- Znp1 (Synaptotagmin) antibody diluted 1:150 in blocking solution. After five washes of 1 hour with washing solution (PBS 0.1% Tween + 1%FCS + 1%BSA), the secondary mouse α-AlexaFluor488 antibody was diluted 1:250 in blocking solution and incubated at 4°C

overnight. Following 5 washes of 20 minutes each, stained fish were stored in 80% glycerol in PBS at 4°C.

For imaging of CaP-MN, fish were laterally embedded in 2% Low Melting Agarose (LMA) microslides under the binocular microscope. Slides were analysed with the fluoresce microscope AxioImagerM2 (Zeiss). The first 10 motor axons posterior to the yolk sac were considered for quantification. Based on overall appearance CaP-MN axons were classified as: normal, truncation (shortened axonal projection) or atrophy (absent axonal projection). Based on terminal branching, axons were classified as Normal, Mild (branching ventral from midline), Medium (2-3 or more branches at ventral or midline) or Severe (>3 branches ventral or dorsal from midline). Brigthfield images of zebrafish larvae were acquired with a Leica S8AP0 binocular attached to an AxioCam ERc5s camera (Zeiss).

Drosophila **strains and transgenic constructs**

The UAS-Smn RNAi construct [P{TRiP.HMC03832}attP40 (UAS-SmnRNAi)] and the *nsyb-GAL4* driver were obtained from the Bloomington Stock Center

The deficiency that removes the *Smn* gene (the Smn^{X7} allele) is gift from Dr. Artavanis-Tsakonas (34).

The *UAS-GFP-dTgs1* strain carries the pPGW-*Tgs1* construct, generated by cloning the *dTgs1* CDS into the pPGW destination vector (Stock number 1077, Drosophila Genomics Resource Center, supported by NIH grant 2P40OD010949), using the Gateway technology (Thermo Fisher Scientific); *the UAS-mst* construct (*UAS-CTRL*) encoding the unrelated Misato protein is ppGW-GFP-Mst (35).

UAS-hTGS1 construct carrying human *TGS1* carries the full-length human *TGS1* gene cloned into the pUAST-attB vector (36). Transgenic flies were obtained by injecting the constructs into either the w^{1118} or the y^1 v^1*; P{CaryP}attP40* (2L, 25C6) strains. All embryo injections were carried out by BestGene (Chino Hills, CA, USA).

The dTgs1^{R1} mutant allele were generated by CRISPR/Cas9 genome editing and described in (37).

The dTgs1^{CB0} mutant allele is the P{PTT-GB}moiCB02140 insertion(38,39).The Oregon-R strain was used as a wild type control. All flies were reared according to standard procedures at 25 °C. Lethal mutations were balanced over *either TM6B, Hu, Tb* or *CyO-TbA, Cy, Tb (40). All* genetic markers and special chromosomes are described in detail in FlyBase (http://www.flybase.org).

Cell Culture, Transfections, generation of TGS1-CRISPR HeLa cell lines and Transductions.

HeLa S3 cells were cultured in DMEM supplemented with 10% fetal bovine serum and Penicillin-Streptomycin at 37°C, in 5% CO2. Lipofectamine 2000 (Life Technologies) was used for all cDNA transfection experiments. TGS1 or SMN CRISPR cells were generated by transfection of HeLa cells with pSpCas9-2A-GFP (PX458) plasmid (41) encoding 3x FLAG Cas9-T2A-GFP and guide RNAs to the *TGS1* locus. (*TGS1-1*: AGAGAAACATTTCCGCCACG; *TGS1-2*: TGTCAGAGCGTATCTTCAGC)(42) or the *SMN* locus (SMN-1: CACCCCACTTACTATCATGC). GFP-positive cells were single-cell sorted into 96 well plates by FACS, and clones carrying mutations affecting TGS1 expression were screened by immunoblotting using anti TGS1 or anti- SMN antibody. respectively.

To generate *TGS1* mutants cells stably expressing FLAG-TGS1 (*TGS1* rescued cells), we transfected 293T cells with the pCDH-CMV-PURO-3xFLAG-TGS1 construct (FLAG-TGS1) and packaging constructs; 48 hours later, viral supernatant was collected and concentrated using Retro-X (Clontech). HeLa TGS1-mutant clones were transduced in the presence of 5 µg/ml polybrene and selected in 2 µg/mL puromycin.

For *TGS1* knockdown, UMUC 3 cells (43) were treated for up to 10 days (every 72 hrs) with 50 nM SMARTpool: siGENOME TGS1 siRNA or ON-TARGET plus Non-targeting siRNA using Dharmafect I (Dharmacon). The PAPD5 KO cell lines are described in (44)

Western blotting

For western blotting, 10–50 micrograms of protein extracts were separated by SDS-PAGE, transferred onto PVDF membrane (GE Healthcare, RPN303F), and blotted according to standard procedures. 5% milk in PBST (0.1% Tween) was used for all blocking and antibody incubation steps; PBST (0.1% Tween) was used for all washes. Detection was performed with the SuperSignal™ West Pico Chemiluminescent Substrate (Thermo); images were acquired with Chemidoc (Biorad) and analyzed using the QuantityOne image analysis software (Biorad).

Primary antibodies :
Anti-FLAG mouse monoclonal (clone M2), Sigma, F1804, 1:1000 ;
pimt/TGS1: Bethyl A300-814A (lot 1), 1:1500
Anti-β-Tubulin mouse monoclonal (clone D-10) Santa Cruz, SC-5274
Anti-SMN mouse monoclonal Antibody (clone 2B1), 05-1532 Sigma-Aldrich, 1:1000
Secondary antibodies: HRP-conjugated anti-Mouse or anti-Rabbit 1:5000 (GE Health Care).

Immunofluorescence staining (IF)

IF experiments were carried out on cells grown on coverslips. Cells were fixed with 4% paraformaldehyde and permeabilized with 0.5% Triton X-100 in PBS. Coverslips were incubated with primary antibodies in 3%BSA for 1 hr at room temperature: monoclonal anti-SMN (05-1532 Sigma-Aldrich clone 2B1, 1:1000); rabbit anti- Coilin (H300 Santa Cruz SC-32860, 200 ug/mL; rabbit anti- TCAB1 ((45), 25 ng/mL). Coverslips were washed 3x with PBS and incubated with secondary Alexa Fluor conjugated antibodies (Jackson Immunoresearch). Coverslips were washed 3x in PBS, counterstained in a 300 nM DAPI solution and mounted in ProLong Gold Anti-fade Mountant. Images were captured on a Leica wide-field fluorescence microscope and processed using Leica LAS AF and Photoshop.

RNA immunoprecipitation of TMG capped RNA

Trizol-purified RNA was subjected to IP with the R1131 anti-TMG cap specific antibody (46). 50 microliters protein G beads were washed with PBS and blocked with 20 micrograms tRNA and 20 micrograms BSA, then washed with NT2 buffer (50mM Tris/HCl pH 7.5; 150mM NaCl; 1mM MgCl2; 0.05% NP40; 1mM DTT; 100U/mL RNasin ; 400μM VRC, Vanadyl ribonucleoside complexes) and coupled O/N to either 20 μL of anti-TMG cap antibody or 1 microgram IgG in 250 μL NT2 buffer. Beads were then briefly washed with NT2 buffer and incubated with Trizol-purified RNA in 250 μL NT2 buffer for 2 hrs at 4 °C, while rotating. Beads were then washed 6 times for 2' while rotating with NT2 buffer and precipitated RNA was Trizol-extracted. Both TMG-immunopurified and input RNAs were then treated with DNase and subjected to reverse transcription and qRT PCR.

Quantitative Real Time PCR (RTqPCR)

Total RNA was extracted with Trizol reagent (Ambion), treated with Ambion™ DNase I (RNase-free) extracted with phenol/chloroform. The integrity of RNA samples was evaluated by gel electrophoresis. 1 microgram of intact RNA (with a 28S:18S rRNA ratio = 2:1) was reverse transcribed with the RevertAid H Minus Reverse Transcriptase kit (Thermo Scientific, EP0451). Real-time PCR reactions were performed with the Brilliant II SYBR® Green QPCR Master Mix (Agilent, 600828)

The relative quantification in gene expression was carried out using the 2- ΔΔCt method (47). Using this method, we obtained the fold changes in gene expression normalized to the GAPDH and Actin genes (the amplification efficiencies were not significantly different for target and reference among all samples). A total of 3 experiments were performed for three biological replicates and the significance was assessed by one-way or two-way ANOVA: For qRTPCR of immunoprecipitated RNA, samples were processed as above; Fold change was calculated by normalizing each RIP sample to the relative input.

Cell fractionation and Northern Blotting

Nuclear extracts (NE) or cytosolic extract (S100) was adopted from the Dignam protocol, as shown in (48). For northern blotting, total RNA is extracted by resuspending pelleted cells with TRIzol reagent. Nuclear and cytoplasmic RNA were extracted from 4µg NE and 12 µg S100 respectively with 5x volume of TRIzol. Precipitated RNAs were dried and resuspended sequentially in 5ul ddH2O, and then with 5ul RNA loading dye, containing 0.1X TBE, 50 mM EDTA, 0.1% Bromophenol blue, 0.1% Xylene Cyanol, 93% formamide. Boiled and rapidly chilled RNAs were fractionated in a 18x24cm 6% Urea-PAGE (19:1) gel, running at 45V for more than 20 hours until the Bromophenol Blue run-off. The gel was then transferred to Hybond-N nitrocellulose membrane under 400mA, for 2 hours in the cold room. Post-transfer gel was stained by ethidium bromide for ribosomal RNAs as loading control. UV-crosslinked membrane is blocked and hybridized in ULTRAhyb, in a 42℃ oven overnight, with P^{32} end-labeled oligonucleotide probes (sequence listed in Table S6). The washed membrane is exposed to a phosphor-imager screen, scanned by a Typhoon scanner, and quantitated with the TotalLab software.

Nascent RNAend-Seq and steady-state 3' RACE-Seq

For the isolation of nascent RNA from HeLa cells, cells were pulsed with 1 hr 4SU 250 uM or 4 hr 4SU 50 uM. For chase experiments, 4SU was removed, the cells were washed in PBS, and 2.5 mM uridine containing media was added. Cells were harvested and cell pellets resuspended in Trizol. RNA was extracted using standard Trizol protocol. 100 ug RNA was biotinylated with Biotin-HPDP (Pierce 21341) at 0.5 mg/mL in 40% DMF and 10 mM Tris pH 7.4, 1 mM EDTA, for 1.5 hours at room temperature. In vitro transcribed luciferase RNA transcribed in the presence of 4SU was spiked into the biotinylation mixture for a final concentration of 0.01 ng/uL. RNA was extracted using isopropanol/ethanol precipitation. Biotinylated RNA was immunoprecipitated from total RNA using mMacs Streptavidin Kit (Miltenyi 130-074-101). The depleted fraction was recovered using isopropanol/ethanol precipitation. Biotinylated RNA was released from beads using 100 mM DTT and cleaned using RNeasy MinElute Kit (QIAGEN 74204). Half of eluted RNA or 600 ng of depleted RNA was ligated to 5 uM of 50 adenylated, 30

blocked adaptor (Universal miRNA cloning linker, NEB S1315S) with 250 units of T4 RNA ligase truncated KQ (NEB M0373S), 25% PEG 8000, and 1 uL RnaseOUT (ThermoFisher 10777019) in a 20 uL reaction at 25 degrees for 16 hours.
Ligated RNA was cleaned up with RNA clean and concentrator columns (Clontech 740955.50) and DNase treatment, cDNA was synthesized with universal primer and SuperScript III (ThermoFisher 18080093). Amplification was carried out with Phusion (New England Biosystems M0530) and primer sets universal/specific for the RNA of interest. PCR products were directly run on an 8% PAGE gel and visualized with SYBR Gold (ThermoFisher S-11494), or subject to AMPure XP beads (Beckman Coulter A63881) for PCR cleanup and library preparation. Libraries were prepped using Kapa Hyperprep Kit (Kapa KK8504), quantified with Qubit and bioanalyzer, and run on Illumina miSeq at the Stanford Functional Genomics Facility. Reads were paired using fastq-join tool at Galaxy (http://usegalaxy. org). Reads were binned into the various forms of each specific snRNA using custom python scripts (https://cmroake.people.stanford.edu/linkspython-scripts) and the number of reads in each bin was normalized to total snRNA reads. Primer sequences can be found in table S6. Data has been deposited at SRA with BioProject ID PRJNA628085.

Transcriptome library preparation and RNA-sequencing

TruSeq Stranded mRNA kit (Illumina, San Diego, CA) has been used for library preparation following the manufacturer's instructions (library type: fr-firstrand).
RNA samples were quantified and quality tested by Agilent 2100 Bioanalyzer RNA assay (Agilent technologies, Santa Clara, CA). Final libraries were checked with both Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) and Agilent Bioanalyzer DNA assay. Libraries were then prepared for sequencing and sequenced on paired-end 150 bp mode on NovaSeq 6000 (Illumina, San Diego, CA).
The number of reads (in millions) produced for each sample is listed in the table below.

Sample name	**Reads (M)**
ID1232_1-1A-CTR1	311.29

ID1232_2-1B-CTR2	310.44
ID1232_3-2A-TGS1 M1	299.96
ID1232_4-3A-TGS1 M2	314.74
ID1232_5-4A-SMN C1	358.39
ID1232_6-5A-SMN C2	272.79

Primary bioinformatic analysis includes:

Base calling and demultiplexing. Processing raw data for both format conversion and demultiplexing by Bcl2Fastq 2.20 version of the Illumina pipeline. *https://support.illumina.com/content/dam/illuminasupport/documents/documentation/software_documentation/bcl2fastq/bcl2fastq2-v2-20-software-guide15051736-03.pdf*

Adapters masking. Adapter sequences are masked with Cutadapt v1.11 from raw fastq data using the following parameters: --anywhere (on both adapter sequences) --overlap 5 --times 2 --minimum-length 35 --mask-adapter (*http://dx.doi.org/10.14806/ej.17.1.200*).

Folder raw_reads contains files with raw reads (R1: first read sequence; R2: second read sequence) and multiqc_report.html file, which aggregates results from primary bioinformatic analysis into a single report files with parameters that give insight into overall processing and sequencing quality.

Transcripts reconstruction and analysis

Paired-ended Illumina sequencing was performed on RNA isolated from TGS1 mutant clones (M1, M2), SMN mutant clones (C1-C2), and parental HeLa cells (CTRL, two replicates) at a depth of 160 million reads/sample. Mutant samples were clustered based on both transcript and gene expression profiles (Figure S4) . We mapped reads to the human reference genome (Gencode annotation v33) using STAR v2.7.1a(49) and ran Scallop(50) on the resulting BAM files to find all novel transcripts in the data. We combined the Scallop-detected novel transcripts with the reference transcripts and used Kallisto (51) to quantify transcript expression levels. Differentially expressed novel and annotated transcripts were found using Sleuth v0.30.0 (51) (Wald test, FDR adjusted p-

value < 0.05, beta value > 2). Our analysis revealed significant changes in the expression levels of 3084 and 351 transcript isoforms in TGS1 and SMN mutant cells, respectively (Figure 7A and B; TableS1). We found changes in the expression levels of a large number of unannotated transcripts (1080/3084 for TGS1 and 158/351 for SMN mutants); 48 of these novel isoforms were found to be affected in both mutant cell types (Figure 7C; Table S1). We compared the 3' side of each novel transcript to its closest annotated transcript to determine whether the novel transcript has a shorter, extended, or the same 3' side as compared to the annotated transcript. Notably, 71% and 76% of the differentially expressed unannotated transcripts in TGS1 and SMN mutants had extended 3' sides, respectively (Figure 7A-B; Table S1).

To investigate the extent of global splicing changes induced by TGS1 and SMN knockouts, we performed differential splicing analysis. We extracted alternative splicing events from the gtf file containing all novel and annotated transcripts and then used SUPPA2 (52) to find significantly changed splicing events. We considered all 7 major categories of alternative splicing events and found the percentage of splicing-inclusion (PSI) for each extracted splicing event. A differentially spliced event between mutant and control samples should have a difference in PSI of at least 0.1 and an adjusted p-value smaller than 0.05. We observed extensive changes in the splicing behavior in TGS1 and SMN mutant cells (Figure S5A) and that the intron retention (IR) category has the largest fraction of significantly affected splicing events in both TGS1 (3.5% and 4.5% of annotated and unannotated IR events, respectively) and SMN mutants (1.2% and 2.1% of annotated and unannotated IR events, respectively) (Figure S5A-B; Table S2).

Methodological approach used for quantification of intron retention (IR) levels (Figure 7D): we performed a transcriptome-wide analysis of the expression levels for the major transcript isoforms in the intron retention events (the transcript without the intron in each event). IR (as determined by SUPPA2) is quantified by Percentage of Splicing Inclusion (PSI), which is computed as the percentage of reads mapping to the intron-inclusion isoform, relative to total reads mapping to both intron-inclusion and intron-exclusion isoforms. Differential expression of the spliced isoforms (intron exclusion) was computed as the difference in TPM values (dTPM) between mutant and control cells (plotted on the X-axis in (Figure 7E, F). The expression of the major transcripts in most of the significant

IR events does not change considerably (small difference in TPM or transcript per million) between control and mutant samples (Figure 7E, F).

As an independent pipeline for detecting differentially expressed transcripts, we also used a combination of Salmon (53) for transcript quantification and DESeq2 for differential analysis. Statistical analyses on individual transcripts were made with the test implemented in the software.

Statistics

Statistical analyses were performed in GraphPad Prism 6. Type of test, sample size, data representation and p-values are indicated in figure legends. $p \leq 0.0001$ is indicated as ****; $p > 0.05$ was considered not significant.

RESULTS

The *C. elegans* TGS1 ortholog is expressed in motoneurons and required for neuron survival and worm locomotion

To investigate the role of TGS1 in the nervous system and its relationships with SMN, we exploited the *C. elegans* model (23). We first analysed the expression of the *C.elegans* ortholog of *TGS1, T08G11.4 1,* hereafter *cetgs-1* (54), using a GFP-reporter approach (24), in which the promoter regions of the two annotated *cetgs-1* isoforms were fused to GFP (p*cetgs-1*::GFP *a* and *b*). Both isoforms are expressed in ventral cord motoneurons (MNs), 19 of which are D-type gamma-aminobutyric acid (GABA) MNs, which are specifically marked by the expression of p*unc-47*::RFP (Figure 1A) (55,56).

To selectively silence *cetgs-1* in the D-type MNs, we generated a strain carrying the [p*unc-25*::*cetgs-1(RNAi)*] transgene. This strain is viable, fertile and has a normal development. As a control we used a transgenic RNAi line that targets *kal-1*, a neurodevelopmental gene not expressed in the D-type MNs (57). We also exploited a *cesmn-1*(*RNAi*) line expressing an extrachromosomal RNAi construct (23). D-type MNs control *C. elegans* backward movement (58). 86% of wild type controls (*wt*) and 85% *kal-1*-silenced animals, showed normal backward locomotion; this proportion was reduced to 40% and 25% in *cetgs-1*(*RNAi*) and *cesmn-1*(*RNAi*) worms, respectively (Figure 1B), indicating that both *cetgs-1* and *cesmn-1* are required for proper MN function and locomotor behavior.

We next asked whether the *cetgs-1*-dependent locomotion defect is due to an alteration in MN morphology and survival. We introduced in *cetgs-1*(*RNAi*) animals the *oxIs12*[p*unc-47*::GFP] transgene that expresses GFP in D-type MNs, allowing their visualization in living animals. *wt* or *kal-1*-silenced worms invariably showed 19 D-type MNs, each extending a circumferential commissure to the dorsal side (Figure 1C-E) (58). In *cetgs-1*(*RNAi*) animals the number of visible D-type MNs was in average 7 per worm, while *cesmn-1*(*RNAi*) animals displayed only 4 neurons/worm, as previously found (23) (Figure 1C, D). The D-type MNs of *cetgs-1*(*RNAi*) animals (n = 606) showed a defect in axonal morphology, consisting of extra branching and guidance

defects (Figure 1E). This defect was present in 14% of the commissures of *cetgs-1*(*RNAi*) worms but was never observed in control (n = 586) or *kal-1*(*RNAi*) (n = 410) animals (p<0.0001; non-parametric *z*-test.). Remarkably, a similar defect was found in knockdown *cesmn-1* worms (23).

Next, we expressed in D-type MNs either human *TGS1* (*punc-119*::*hTGS1,* referred to as *hTGS1*) or the red fluorescent protein *dsRED* that served as a control (p*unc-119*::*dsRED*). *hTGS1*, but not *dsRED,* partially rescued the MN loss caused by *cetgs-1*(*RNAi*) (from 4 to 9 neurons/animal; Figure 1C,D), indicating that this phenotype is specifically caused by *cetgs-1* silencing and suggesting that the *TGS1* requirement for neuron survival is conserved from worms to humans.

Finally, we asked whether the loss of D-type MNs observed in *cetgs-1*(*RNAi*) worms is caused by cell death. Apoptotic MN death in *cesmn-1* knockdown animals is revealed by the accumulation of an endogenous auto-fluorescent marker (23). *wt* worms showed no dying MNs, while *kal-1*(*RNAi*) worms showed an average 1 dying MN per animal. In contrast, *cetgs-1*(*RNAi*) and *cesmn-1*(*RNAi*) worms displayed 7 and 5 dying MNs per worm, respectively (Figure 1F, G).Thus, both *cesmn-1* and *cetgs-1* prevent D-type MN death.

***cetgs-1* overexpression mitigates the defects elicited by *cesmn-1* depletion**

To determine whether *cetgs-1* interacts genetically with *cesmn-1*, we expressed *cetgs-1* under the control of a pan-neuronal promoter (p*unc119*::*cetgs-1*, referred to as *ceTGS-1*) in worms that also express an integrated *cesmn-1* RNAi construct (*cesmn-1(RNAi Is)* (23). 62% of the *cesmn-1*(*RNAi Is*) worms showed abnormal locomotion. Expression of *ceTGS-1* in these worms lowered the proportion of animals with abnormal locomotion to 28% (Figure 1B), while *dsRED* expression did not change the frequency of worms with locomotory defects (68%, p = 0.9) (Figure 1B). In addition, while in *cesmn-1*(*RNAi Is*) or *cesmn-1*(*RNAi Is*); dsRED worms, we observed an average of 9 dying MNs/worm, in the presence of the *ceTGS-1* or *hTGS1* transgenes, the average number of dying MNs was slightly but significantly lowered to 7 and 8, respectively (Figure 1F, G). Thus,

cetgs-1 overexpression partially suppresses the locomotion defects and the neuronal death caused by *cesmn-1* silencing.

Overexpression of dTgs1 ameliorates the Smn loss of function phenotype in Drosophila

We next assayed the effects of TGS1 overexpression in a fly SMA model. It has been reported that *Drosophila* Tgs1 (dTgs1) depletion affects motor behavior (37,59,60). We have previously shown that RNAi against *Smn* in neurons perturbs the neural circuit that controls post-eclosion events, leading to a defect in wing expansion (61). A similar phenotype has been observed in hypomorphic *dTgs1* mutants (37). We generated flies expressing a *UAS-Smn RNAi* construct in neurons, using the *nsyb-GAL4* driver. 16% of these *Smn RNAi* flies displayed unexpanded wings (Figure 2A, B), and this percentage was increased to 44% when RNAi was performed in flies carrying only one copy of *Smn*+ (Δ*Smn/+;* Figure 2B). We then examined flies co-expressing *UAS-Smn-RNAi* and *UAS-dTgs1*, both driven by *nsyb-GAL4*. As control, we used flies carrying a *UAS-CTRL* construct encoding the unrelated Mst protein (35). An increase in the dosage of dTgs1, but not of Mst, significantly lowered the frequency of flies with unexpanded wings (from 16% to 9%) (Figure 2B). A human *TGS1* transgene (*UAS-hTGS1)* which fully rescues the lethality associated with null *dTgs1* mutations (37), decreased the proportion of flies with unexpanded wings from 16% to 10% in Smn^{+}/Smn^{+} flies, and from 44% to 30% in flies bearing a single copy of *Smn* (Figure 2B). These results indicate that dTgs1 overexpression partially suppresses the defects caused by *Smn* downregulation in at least a subset of fly neurons, and that human TGS1 can substitute for the neuronal function of its fly ortholog.

TGS1 downregulation in zebrafish causes motor axon defects

To determine the consequences of TGS1 loss in a vertebrate model, we exploited zebrafish. To downregulate *tgs1*, we injected larvae with two antisense morpholino oligonucleotides (MOs): a translation blocking MO (*tgs1* ATG-MO) and a splicing-

blocking MO (*tgs1* Sp-MO) (Figure S1A, B). Each MO was injected at an optimized dosage and caused little or no effect on the overall larval morphology and development (Figure S1C).

Injection of 2 ng *tgs1* ATG-MO resulted in defects in caudal primary motoneuron (CaP-MN), including axonal truncations and increased terminal branching. 10% of the CaP-MNs exhibited truncated axonal projections and 25% showed increased terminal branching (Figure 3A, B). Negligible CaP-MN defects were observed in uninjected larvae and in larvae injected with a non-targeting MO (Figure 3A, B). 1.5 ng *tgs1* Sp-MO injection led to 15% CaP-MNs with axonal truncations and 35% with increased terminal branching (Figure 3B). Notably, these results recapitulate the motoneuron defects observed in *smn* morphants (62-64).

We also attempted to perform rescue experiments by co-injecting FLAG-tagged human *TGS1* mRNA and the *tgs1* Sp-MO. Co-injection of *tgs1* Sp-MO and 100 pg of *FLAG-hTGS1* mRNA resulted in a significant rescue of the neurological phenotype (Figure S1 D). However, injection in embryos of *FLAG-hTGS1* mRNA (100 to 400 pg) caused developmental defects in larvae (Figure S1C), preventing a firm conclusion about the rescue capacity of human *TGS1.*

Collectively, these results show that TGS1 plays a role in motoneuron development and survival in zebrafish.

TGS1 deficiency impairs 3' processing of snRNAs in human cells

Given the similarity between the phenotypes elicited by TGS1 and SMN depletion, we asked whether TGS1 loss affects snRNP biogenesis like SMN (65). To address this, we used two HeLa clones carrying CRISPR/Cas9-induced mutations in *TGS1* (*TGS1 M1*, *M2*) (42), which decrease TGS1 expression to less than 10% of the normal level but do not affect the expression of the SMN protein (Figure 4A). We have previously shown that both mutant lines exhibit a rather diffuse distribution of the Cajal body (CB) marker coilin and fail to accumulate SMN in the CBs. This defect is rescued by the expression of FLAG-TGS1 (*TGS1 M1R* cells, Figure 4A and Figure S2 A, B).

To determine the effects of *TGS1* deficiency on snRNA cap hypermethylation, we performed RNA-IP with an antibody that specifically recognizes the TMG cap (46). Quantitative RT-PCR (RT-qPCR) on RNA precipitated with the TMG antibody showed that the TMG-capped U2 snRNA molecules were 10-fold more abundant in control cells than in *TGS1* mutant cells. Actin mRNA, which has an MMG cap that is not recognized by the anti-TMG antibody was used as a control (Figure 4B).

RT-qPCR on total RNA from control cells, *TGS1* mutant cells (*TGS1 M),* or *TGS1* mutant cells bearing a FLAG-*TGS1* rescue construct *(TGS1 M1R*) showed no significant differences in snRNA abundance, with minor changes likely reflecting clonal variability (Figure 4C). However, Northern Blotting (NB) revealed that *TGS1* mutant cells exhibit slower migrating U2 snRNA species that are likely to correspond to U2 snRNAs precursors carrying genome-encoded extensions at their 3' ends (11,21). As shown in Figure 4D (lanes 1-2), the extended U2 precursors are barely detectable in total RNA from control cells but are substantially enriched in *TGS1* mutant cells.

Next, we analyzed by NB the distribution of the extended U2 species in nuclear and cytoplasmic fractions of both control and *TGS1* mutant cells. While these pre-U2 species were clearly enriched in *TGS1* mutant nuclei (Figure 4D, lanes 5, 9, 10), they were barely detectable in nuclei of control or rescued cells (*TGS1 M1R*, *TGS1 M2R;* Figure 4D, lanes 3, 4, 6, 11 and Figure 4E). The extended pre-U2 species were also detected in cytoplasmic RNA from control and *TGS1 M1R* cells (lanes 13,14,15), but their abundance was greatly increased in *TGS1* mutant cells (Figure 4D, lanes 12,16 and Figure 4E). Nuclear or cytoplasmic accumulations of extended U2 precursors were not observed in cells carrying mutations in the *PAPD5* gene that encodes a deadenylase required to remove oligo-A tails from different noncoding RNAs (44) (Figure 4D, lanes 7,15). These results indicate that TGS1 deficiency causes aberrant accumulation of unprocessed U2 precursors that might interfere with spliceosome activity, as occurs in yeast (66) .

We characterized the 3' ends of the pre-U2 molecules in total RNA by 3'RACE and Next-Gen sequencing, as described in (44,67). In control cells, 90% of the reads mapped to the mature forms of U2, most of which contained 188 nt plus a 3' adenosine added post-transcriptionally (≤189 nt), in agreement with previous results (3,68). The

remaining 10% of the U2 molecules had extended tails including genome-encoded nucleotides at their 3' ends (Figure 5A, B). Consistent with the NB results, *TGS1* mutants displayed a 4-fold higher proportion of 3' extended U2 molecules (41%) compared to control (10%), with an average tail length of 12 nt (Figure 5A, B). In *TGS1M1R* cells, the abundance of the extended U2 molecules was reduced from 41% to 20%, showing rescue of the mutant phenotype (Figure 5A). 3'RACE and sequencing performed on nuclear and cytoplasmic RNA confirmed the data obtained by NB (Figures 4E, 5C, D). In the nuclear RNA fractions, the abundance of the U2 extended species was 18% in *TGS1* mutant cells, 7% in control cells and 8% in *TGS1M1R* cells (Figure 5C). In the cytoplasmic fractions, the extended U2 molecules were 52% in *TGS1* mutant cells, 13% in control cells, and 20% in *TGS1M1R* cells (Figure 5C). *TGS1* mutant cells showed no variations in the proportion of mature U2 molecules carrying a 3' untemplated adenosine compared to controls; in both cases the 189/188 ratio was approximately 80%. Importantly, the finding that the nuclei of *TGS1* mutant cells contain more extended U2 molecules than control nuclei (Figures 4E and 5C) suggests that a fraction of these unprocessed species completes the cytoplasmic assembly pathway and is imported in the nucleus.

To assess whether TGS1 deficiency affects processing of the U2 precursors in a different cell type, we knocked down *TGS1* by RNAi in UMUC3 cells, which showed a strong reduction in the TGS1 protein and defective CBs (Figure S2C, D). 3' RACE and sequencing on total RNA revealed that in *TGS1*-RNAi cells the proportion of extended molecules is 3-fold higher than in cells treated with control siRNA (Figure 5A).

To confirm that the extended species are precursors of U2 molecules, we performed metabolic labeling of newly transcribed RNA with 4-thiouridine for 4 hours in HeLa cells, and affinity-purified this nascent RNA population with thiol-specific biotinylation and streptavidin-coated magnetic beads (44). Unlabeled RNAs (steady-state RNA) recovered after this treatment, are enriched in mature RNA species while labelled RNAs are enriched in short-lived RNA precursors. A comparison of the sequence profiles showed that extended-U2 molecules are 5-fold more abundant in nascent RNA samples (59% of total U2 reads) than in steady-state RNA (12%; Figure

5A, B), indicating that the extended U2 species are immature U2 precursors, not yet processed at their 3' ends.

In total RNA from control cells, a small fraction of the U2 reads are extended molecules carrying non-templated adenosines (0.2%) or uridines (0.6%) added post-transcriptionally at their 3' ends. The acquisition of an extra A or U is specific for the extended molecules with a mean tail length of 12 nucleotides. In *TGS1* mutant cells, the fraction of extended U2 molecules incorporating extra uridines is 6 times and 2 times higher than in control and *TGS1M1R* rescued cells, respectively (Figure 5B,E). A 5-fold increase in uridylated molecules is also observed in *TGS1* RNAi UMUC3 cells compared to cells treated with nontargeting dsRNA (Figure 5E). These data suggest that TGS1 affects not only 3' trimming but also the profile of post-transcriptional modifications on the U2 precursors. Interestingly, oligo-U extended molecules are 5-fold more abundant in nascent RNA than in mature RNA, indicating that these modifications represent an intermediate step in the 3' processing of U2 precursors (Figure 5E).

We also analyzed the 3' sequences of the U1 and the U4atac snRNAs. In both control and *TGS1* mutant cells 98% of U1 snRNA molecules were 164 nt long and less than 1% of these molecules carried untemplated adenosine or uridine residues at their ends, indicating that TGS1 loss does not affect U1 biogenesis. In contrast, *TGS1* mutant cells were enriched in U4atac extended forms showing a mean tail length of 10 nt; these forms were 19% of total reads in *TGS1* mutants, 6% in control cells, and 9% in rescued cells (Figure 5F-G). Moreover, in *TGS1* mutant cells 3% of U4atac molecules incorporated additional uridines, compared to 0.25% in control cells and 0.7% in rescued cells (Figure 5H). Collectively, these results indicate that TGS1 depletion results in extended U2 and U4atac snRNAs. However, these snRNAs differ in the type and proportion of 3' end extensions, most likely reflecting diverse dynamics of their 3' end processing.

Tgs1 controls 3' processing of snRNAs in Drosophila.

To ask whether the role of TGS1 in snRNA 3' processing is conserved in flies, we exploited *dTgs1*^{R1}, a presumably null CRISPR/Cas9-induced allele, which, when

homozygous, causes death in late second instar larvae (37). We also used *dTgs1*^{R1}/*dTgs1*^{CB0} heterozygous flies that die as third instar larvae. The lethality of both mutants is rescued by ubiquitous expression of either a *dTgs1* or a human *TGS1* trangene (37,39).

We characterized the 3' ends of U2 snRNAs by 3'RACE and RNA sequencing on total RNA from wild type larvae (*CTR*), *dTgs1*^{R1} homozygous larvae and *dTgs1*^{R1}/*dTgs1*^{CB0} larvae; as control, we also examined mutant larvae constitutively expressing *dTgs*1. *dTgs1* mutant larvae showed an increase in 3' extended U2 species compared to either wild type larvae or larvae bearing the rescue construct (Figure 6A). 95% of these extended U2 species displayed one additional uridine at the 3' end (Figure 6B). The remaining 5% of the extended molecules included species containing longer tails incorporating both genome-templated and untemplated A or U nucleotides. Thus, TGS1 is required for proper 3' processing of U2 snRNAs also in flies.

In contrast, an analysis of the U1 snRNAs showed that they are substantially unaffected by *dTgs1* mutations, consistent with the results obtained with human cells. In mutant and wild type larvae, extended species represented about 1,5% and 0.5% of total U1 RNAs, respectively.

Aberrant transcripts accumulate in both *TGS1* and *SMN* mutant cells

Studies conducted in different systems demonstrated that SMN deficiency causes profound perturbations in the transcriptome (69-74). To explore the consequences of TGS1 loss on mRNA splicing and gene expression, we performed deep sequencing on total RNA extracted from *TGS1* mutant HeLa cells, and from two independent *SMN* mutant clones derived from the same HeLa cell line used to generate *TGS1* mutant cells. These lines, designated as *SMN C1* and *SMN C2*, carry CRISPR-induced mutations in the sixth exon of the gene; *SMN C1* has a small deletion in the SMN proline rich domain, and *SMN C2* carries frameshift mutations. Both lines show reductions in the abundance of the SMN protein (Figure S3A). The protein encoded by the *SMN C2* cells has a lower MW compared to the wild type protein and fails to

concentrate in the CBs (Figure S3A, B).

Paired-ended Illumina sequencing (160 million reads/sample) was performed on RNA isolated from *TGS1* mutant clones (*M1, M2*), *SMN* mutant clones (*C1-C2*), and parental HeLa cells (*CTR*). Independent mutant clones and controls were clustered based on both transcript and gene expression profiles (Figure S4A-C). We used Kallisto (75) and Sleuth (51) for transcript quantification and differential analysis, respectively (see Methods). *TGS1* and *SMN* mutant cells showed a differential expression of both annotated and unannotated transcripts compared to controls (reconstructed by Scallop (50); Figure 7A, B; Table S1). The number of differentially expressed (DE) transcripts was higher in *TGS1* mutant cells than in *SMN* mutant cells (Figure 7A, B), likely reflecting the higher degree of functional inactivation of *TGS1* compared to *SMN* (Figures 4A and S3A). Approximately one third of both annotated and unannotated DE transcripts found in *SMN* mutant cells was also present in *TGS1* mutant cells (Figure 7C; Table S1, see Methods). Notably, many unannotated DE transcripts found in both *TGS1* and *SMN* mutant cells showed an increased number of reads for intergenic regions compared to controls (Figure 7A, B; Table S1). These 3' elongated transcripts extended beyond the 3' end of the gene and sometimes included the coding region of the next gene, likely reflecting an altered transcription termination. Most of the DE transcripts were also identified using an independent analysis workflow (50,53) (Tables S3-S5). The mutant cells also exhibited changes in splicing behavior (Figure S5A), with intron retention being the most frequent defect (Figure 7D-F; Figure S5A, B; Table S2).

To confirm the *in silico* predictions, we performed targeted validation for 4 transcripts by RT-PCR or RT-qPCR analysis on RNA samples from *CTR, TGS1 M* and *TGS1 M1R* cells. Amplification products consistent with the predicted novel transcripts isoforms were enriched in the two *TGS1* mutant clones, compared to both the parental cell line and *TGS1 M1* rescued cells (Figures S6 and S7).

DISCUSSION

We have shown that the highly conserved TGS1 hypermethylase localizes to motoneurons in *C. elegans*, and that its deficiency results in abnormalities in neuron structure and function in both *C.elegans* and *D. rerio*. Consistent with these defects, *TGS1* deficiency impairs locomotion in worms and wing expansion in flies, which is governed by bursicon-expressing neurons (76). Previous work has also shown that mutations in *Drosophila Tgs1* cause abnormal larval locomotion (37,59,60). Furthermore, we found that the phenotypes elicited by *TGS1* inhibition in flies, worm and zebrafish are highly reminiscent of those caused by loss of function of the *SMN* gene. Importantly, we showed that *TGS1* overexpression ameliorates the neurological phenotypes elicited by the impairment of the *SMN* function in both the fly and worm model systems. These findings suggest commonalities in the molecular and cellular consequences caused by *TGS1* and *SMN* loss of function. Several studies have linked neurodegeneration to defective snRNA, snRNP biogenesis and pre-mRNA splicing (3,65,73,77,78). It has been reported that SMN-depleted cells exhibit an enrichment of 3' unprocessed snRNA precursors resulting from their impaired assembly into snRNPs (11). In a mouse model of SMA, motor neurons display a much greater reduction in the snRNPs levels relative to unaffected spinal cells (79), and their selective death is the result of converging mechanisms of p53 activation driven by dysregulation of distinct mRNA splicing pathways (80-82). It has been also reported that mutations in a single gene of the U2 multicopy cluster (Rnu2-8), result in defective splicing and cause ataxia and neurodegeneration in mice (77). Unprocessed snRNAs accumulate in cells from patients affected by pontocerebellar hypoplasia (PCH7), which is also characterized by motoneuron loss. Interestingly, we show that *TGS1* mutant cells accumulate extended snRNA species with post-transcriptionally added adenosines or uridines similar to cells from PCH7 patients (3,83,84).

In both human cells and *Drosophila,* TGS1 loss results in the accumulation of untrimmed 3' extended U2 snRNA molecules but has no effect on U1 snRNAs. TGS1-deficient human cells also accumulate untrimmed U4atac snRNAs. Moreover, TGS1 deficiency results in extensive alterations in the human transcriptome, including changes in the efficiency of intron removal and accumulation of transcripts with 3' extensions spanning intergenic regions. Similar alterations are also observed in the

transcriptome of *SMN* mutant cells. Recent work has suggested that SMN directly mediates proper transcription termination by favoring polymerase II release (85), and SMN deficiency has been linked to accumulation of R-loops and DNA damage (71,86). Accumulation of aberrant read-through transcripts is consistent with these results and may contribute to the neurodegenerative defects observed in TGS1 and SMN mutant models.

Collectively, our data support a model where TGS1 and SMN play closely related functions in preventing neurodegeneration (Figure 7G-H). TGS1 plays functions in snRNA 3' maturation, pre-mRNA splicing and possibly in transcription termination. Similarly, SMN promotes assembly of snRNA into snRNPs and is essential for splicing and efficient transcription termination (73,85). Thus, loss of either protein leads to substantial transcriptome alterations. These alterations may be neurotoxic by directly affecting specific transcripts important for neuronal survival and function, a scenario supported by increasing experimental evidence (80-82,87). Importantly, we found that TGS1 overexpression can partially rescue the neurological phenotypes caused by SMN depletion. Moreover, previous work has shown that TGS1 directly interacts with the SMN protein in human cells (14), and that *Drosophila* Tgs1 physically associates with most components of the SMN complex *in vivo* (37). Thus, the physical interaction between TGS1 and SMN and the similar phenotypes elicited by their deficiency suggest a functional interaction between these proteins. We hypothesize that TGS1 overexpression in SMN-depleted cells may enhance snRNP biogenesis and possibly improve the transcription termination fidelity, compensating for the defects caused by SMN deficiency (Figure 7H).

Recent remarkable advances in the field have led to the approval of disease modifying therapies for SMA patients, which include direct gene replacement via viral vector and antisense oligonucleotides designed to promote the production of full-length SMN (88-90). However, it is evident that these treatments do not represent a cure for the disease and expansion of therapeutic options for SMA patient is needed. This will likely require the development of SMN-independent strategies targeting disease-relevant downstream pathways (88-92). Candidate targets have been emerging from

studies in model organisms (63,69,81,82,93). Here, we have shown that TGS1 expression can compensate, at least in part, for SMN loss in animal models of neurodegeneration, suggesting that modulation of TGS1 activity may be a new potential therapeutic avenue for SMA.

Data Availability

Custom python scripts for 3'-RACE: https://cmroake.people.stanford.edu/links-python-scripts

RNAend-Seq data: SRA BioProject ID PRJNA628085

Transcriptome analyses: https://github.com/roozbehdn/TGS1_SMN

Acknowledgements

We thank P. Bazzicalupo for critical reading of the manuscript, G. Zampi and F. Sola for technical support, S. Schneider for support in zebrafish procedures, P. De Piante for data analysis. For strains and plasmids we thank A. Fire, K. Shen, O. Hobert, C. Bargmann and the Caenorhabditis Genetics Center (CGC), funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

Funding

This work was supported by grants from Telethon GPP13147 (to G.D.R.) and GGP16203 (to E.D.S.), NIH AG056575 and CA197563 (to S.E.A.), AIRC IG 20528 (to M.G.), from the Italian Ministry of Economy and Finance (FaReBio di Qualità) and the Italian Ministry of Health RF2009-1473235 (to E.D.S.), from Center for Molecular Medicine Cologne CMMC C16 and Deutsche Forschungsgemeinschaft Wi945/19-1, Wi945/17-1 (to B.W.). L.C. was supported by a Stanford Cancer Institute 2018 Fellowship Award, and C.M.R. by an MSTP Training Grant GM007365 and a Gerald J. Lieberman Fellowship.

Conflict of interest statement. None declared.

REFERENCES

1. Nussbacher, J.K., Tabet, R., Yeo, G.W. and Lagier-Tourenne, C. (2019) Disruption of RNA Metabolism in Neurological Diseases and Emerging Therapeutic Interventions. *Neuron*, **102**, 294-320.
2. Wirth, B., Karakaya, M., Kye, M.J. and Mendoza-Ferreira, N. (2020) Twenty-Five Years of Spinal Muscular Atrophy Research: From Phenotype to Genotype to Therapy, and What Comes Next. *Annu Rev Genomics Hum Genet*.
3. Lardelli, R.M., Schaffer, A.E., Eggens, V.R., Zaki, M.S., Grainger, S., Sathe, S., Van Nostrand, E.L., Schlachetzki, Z., Rosti, B., Akizu, N. *et al.* (2017) Biallelic mutations in the 3' exonuclease TOE1 cause pontocerebellar hypoplasia and uncover a role in snRNA processing. *Nat Genet*, **49**, 457-464.
4. de Vegvar, H.E., Lund, E. and Dahlberg, J.E. (1986) 3' end formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. *Cell*, **47**, 259-266.
5. Hernandez, N. and Weiner, A.M. (1986) Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. *Cell*, **47**, 249-258.
6. Egloff, S., Zaborowska, J., Laitem, C., Kiss, T. and Murphy, S. (2012) Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes. *Mol Cell*, **45**, 111-122.
7. O'Reilly, D., Kuznetsova, O.V., Laitem, C., Zaborowska, J., Dienstbier, M. and Murphy, S. (2014) Human snRNA genes use polyadenylation factors to promote efficient transcription termination. *Nucleic Acids Res*, **42**, 264-275.
8. Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N. and Shiekhattar, R. (2005) Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell*, **123**, 265-276.
9. Ohno, M., Segref, A., Bachi, A., Wilm, M. and Mattaj, I.W. (2000) PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell*, **101**, 187-198.
10. Kitao, S., Segref, A., Kast, J., Wilm, M., Mattaj, I.W. and Ohno, M. (2008) A compartmentalized phosphorylation/dephosphorylation system that regulates U snRNA export from the nucleus. *Mol Cell Biol*, **28**, 487-497.
11. Yong, J., Kasim, M., Bachorik, J.L., Wan, L. and Dreyfuss, G. (2010) Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. *Mol Cell*, **38**, 551-562.
12. Pellizzoni, L., Yong, J. and Dreyfuss, G. (2002) Essential role for the SMN complex in the specificity of snRNP assembly. *Science*, **298**, 1775-1779.
13. Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J. and Bordonne, R. (2002) Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol Cell*, **9**, 891-901.
14. Mouaikel, J., Narayanan, U., Verheggen, C., Matera, A.G., Bertrand, E., Tazi, J. and Bordonne, R. (2003) Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. *EMBO Rep*, **4**, 616-622.

15. Narayanan, U., Achsel, T., Luhrmann, R. and Matera, A.G. (2004) Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein. *Mol Cell*, **16**, 223-234.
16. Strasser, A., Dickmanns, A., Luhrmann, R. and Ficner, R. (2005) Structural basis for m3G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1. *EMBO J*, **24**, 2235-2243.
17. Fischer, U., Heinrich, J., van Zee, K., Fanning, E. and Luhrmann, R. (1994) Nuclear transport of U1 snRNP in somatic cells: differences in signal requirement compared with Xenopus laevis oocytes. *J Cell Biol*, **125**, 971-980.
18. Marshallsay, C. and Luhrmann, R. (1994) In vitro nuclear import of snRNPs: cytosolic factors mediate m3G-cap dependence of U1 and U2 snRNP transport. *EMBO J*, **13**, 222-231.
19. Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M. and Luhrmann, R. (1998) Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *Embo J*, **17**, 4114-4126.
20. Stanek, D. (2017) Cajal bodies and snRNPs - friends with benefits. *RNA biology*, **14**, 671-679.
21. Wieben, E.D., Nenninger, J.M. and Pederson, T. (1985) Ribonucleoprotein organization of eukaryotic RNA. XXXII. U2 small nuclear RNA precursors and their accurate 3' processing in vitro as ribonucleoprotein particles. *Journal of molecular biology*, **183**, 69-78.
22. Brenner, S. (1974) The genetics of Caenorhabditis elegans. *Genetics*, **77**, 71-94.
23. Gallotta, I., Mazzarella, N., Donato, A., Esposito, A., Chaplin, J.C., Castro, S., Zampi, G., Battaglia, G.S., Hilliard, M.A., Bazzicalupo, P. *et al.* (2016) Neuron-specific knock-down of SMN1 causes neuron degeneration and death through an apoptotic mechanism. *Hum Mol Genet*, **25**, 2564-2577.
24. Hobert, O. (2002) PCR fusion-based approach to create reporter Gene constructs for expression analysis in transgenic C. elegans. *BioTechniques*, **32**, 728-730.
25. Esposito, G., Di Schiavi, E., Bergamasco, C. and Bazzicalupo, P. (2007) Efficient and cell specific knock-down of gene function in targeted C. elegans neurons. *Gene*, **395**, 170-176.
26. Eastman, C., Horvitz, H.R. and Jin, Y. (1999) Coordinated transcriptional regulation of the unc-25 glutamic acid decarboxylase and the unc-47 GABA vesicular transporter by the Caenorhabditis elegans UNC-30 homeodomain protein. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **19**, 6225-6234.
27. Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M. *et al.* (2003) Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*, **421**, 231-237.
28. Mello, C.C., Kramer, J.M., Stinchcomb, D. and Ambros, V. (1991) Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. *The EMBO journal*, **10**, 3959-3970.
29. Veronico, P., Gray, L.J., Jones, J.T., Bazzicalupo, P., Arbucci, S., Cortese, M.R., Di Vito, M. and De Giorgi, C. (2001) Nematode chitin synthases: Gene structure,

expression and function in Caenorhabditis elegans and the plant parasitic nematode Meloidogyne artiellia. *Molecular Genetics and Genomics*, **266**, 28-34.

30. McIntire, S.L., Reimer, R.J., Schuske, K., Edwards, R.H. and Jorgensen, E.M. (1997) Identification and characterization of the vesicular GABA transporter. *Nature*, **389**, 870-876.
31. Riessland, M., Kaczmarek, A., Schneider, S., Swoboda, K.J., Lohr, H., Bradler, C., Grysko, V., Dimitriadi, M., Hosseinibarkooie, S., Torres-Benito, L. *et al.* (2017) Neurocalcin Delta Suppression Protects against Spinal Muscular Atrophy in Humans and across Species by Restoring Impaired Endocytosis. *American journal of human genetics*, **100**, 297-315.
32. Mendoza-Ferreira, N., Coutelier, M., Janzen, E., Hosseinibarkooie, S., Lohr, H., Schneider, S., Milbradt, J., Karakaya, M., Riessland, M., Pichlo, C. *et al.* (2018) Biallelic CHP1 mutation causes human autosomal recessive ataxia by impairing NHE1 function. *Neurology. Genetics*, **4**, e209.
33. Lan, C.C., Tang, R., Un San Leong, I. and Love, D.R. (2009) Quantitative real-time RT-PCR (qRT-PCR) of zebrafish transcripts: optimization of RNA extraction, quality control considerations, and data analysis. *Cold Spring Harbor protocols*, **2009**, pdb prot5314.
34. Chang, H.C., Dimlich, D.N., Yokokura, T., Mukherjee, A., Kankel, M.W., Sen, A., Sridhar, V., Fulga, T.A., Hart, A.C., Van Vactor, D. *et al.* (2008) Modeling spinal muscular atrophy in Drosophila. *PLoS One*, **3**, e3209.
35. Palumbo, V., Pellacani, C., Heesom, K.J., Rogala, K.B., Deane, C.M., Mottier-Pavie, V., Gatti, M., Bonaccorsi, S. and Wakefield, J.G. (2015) Misato Controls Mitotic Microtubule Generation by Stabilizing the TCP-1 Tubulin Chaperone Complex [corrected]. *Curr Biol*, **25**, 1777-1783.
36. Bischof, J., Maeda, R.K., Hediger, M., Karch, F. and Basler, K. (2007) An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A*, **104**, 3312-3317.
37. Maccallini, P., Bavasso, F., Scatolini, L., Bucciarelli, E., Noviello, G., Lisi, V., Palumbo, V., D'Angeli, S., Cacchione, S., Cenci, G. *et al.* (2020) Intimate functional interactions between TGS1 and the Smn complex revealed by an analysis of the Drosophila eye development. *PLoS Genet*, **16**, e1008815.
38. Buszczak, M., Paterno, S., Lighthouse, D., Bachman, J., Planck, J., Owen, S., Skora, A.D., Nystul, T.G., Ohlstein, B., Allen, A. *et al.* (2007) The carnegie protein trap library: a versatile tool for Drosophila developmental studies. *Genetics*, **175**, 1505-1531.
39. Raffa, G.D., Siriaco, G., Cugusi, S., Ciapponi, L., Cenci, G., Wojcik, E. and Gatti, M. (2009) The Drosophila modigliani (moi) gene encodes a HOAP-interacting protein required for telomere protection. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 2271-2276.
40. Lattao, R., Bonaccorsi, S., Guan, X., Wasserman, S.A. and Gatti, M. (2011) Tubby-tagged balancers for the Drosophila X and second chromosomes. *Fly*, **5**, 369-370.
41. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, **8**, 2281-2308.

42. Chen, L., Roake, C.M., Galati, A., Bavasso, F., Micheli, E., Saggio, I., Schoeftner, S., Cacchione, S., Gatti, M., Artandi, S.E. *et al.* (2020) Loss of Human TGS1 Hypermethylase Promotes Increased Telomerase RNA and Telomere Elongation. *Cell Rep*, **30**, 1358-1372 e1355.
43. Xu, L. and Blackburn, E.H. (2007) Human cancer cells harbor T-stumps, a distinct class of extremely short telomeres. *Mol Cell*, **28**, 315-327.
44. Roake, C.M., Chen, L., Chakravarthy, A.L., Ferrell, J.E., Raffa, G.D. and Artandi, S.E. (2019) Disruption of Telomerase RNA Maturation Kinetics Precipitates Disease. *Molecular Cell*, **74**, 688-700.e683.
45. Venteicher, A.S., Abreu, E.B., Meng, Z., McCann, K.E., Terns, R.M., Veenstra, T.D., Terns, M.P. and Artandi, S.E. (2009) A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science*, **323**, 644-648.
46. Bringmann, P., Reuter, R., Rinke, J., Appel, B., Bald, R. and Luhrmann, R. (1983) 5'-terminal caps of snRNAs are accessible for reaction with 2,2,7-trimethylguanosine-specific antibody in intact snRNPs. *J Biol Chem*, **258**, 2745-2747.
47. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402-408.
48. Chen, L., Ooi, S.K., Conaway, R.C. and Conaway, J.W. (2014) Generation and purification of human INO80 chromatin remodeling complexes and subcomplexes. *J Vis Exp*, e51720.
49. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.
50. Shao, M. and Kingsford, C. (2017) Accurate assembly of transcripts through phase-preserving graph decomposition. *Nature biotechnology*, **35**, 1167-1169.
51. Pimentel, H., Bray, N.L., Puente, S., Melsted, P. and Pachter, L. (2017) Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature methods*, **14**, 687-690.
52. Trincado, J.L., Entizne, J.C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D.J. and Eyras, E. (2018) SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome biology*, **19**, 40.
53. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods*, **14**, 417-419.
54. Shaye, D.D. and Greenwald, I. (2011) OrthoList: a compendium of C. elegans genes with human orthologs. *PLoS One*, **6**, e20085.
55. Spencer, W.C., Zeller, G., Watson, J.D., Henz, S.R., Watkins, K.L., McWhirter, R.D., Petersen, S., Sreedharan, V.T., Widmer, C., Jo, J. *et al.* (2011) A spatial and temporal map of C. elegans gene expression. *Genome Research*, **21**, 325-341.
56. Kaletsky, R., Lakhina, V., Arey, R., Williams, A., Landis, J., Ashraf, J. and Murphy, C.T. (2016) The C. elegans adult neuronal IIS/FOXO transcriptome reveals adult phenotype regulators. *Nature*, **529**, 92-96.

57. Rugarli, E.I., Di Schiavi, E., Hilliard, M.A., Arbucci, S., Ghezzi, C., Facciolli, A., Coppola, G., Ballabio, A. and Bazzicalupo, P. (2002) The Kallmann syndrome gene homolog in C. elegans is involved in epidermal morphogenesis and neurite branching. *Development (Cambridge, England)*, **129**, 1283-1294.
58. McIntire, S.L., Jorgensen, E., Kaplan, J. and Horvitz, H.R. (1993) The GABAergic nervous system of Caenorhabditis elegans. *Nature*, **364**, 337-341.
59. Komonyi, O., Schauer, T., Papai, G., Deak, P. and Boros, I.M. (2009) A product of the bicistronic Drosophila melanogaster gene CG31241, which also encodes a trimethylguanosine synthase, plays a role in telomere protection. *J Cell Sci*, **122**, 769-774.
60. Borg, R.M., Fenech Salerno, B., Vassallo, N., Bordonne, R. and Cauchi, R.J. (2016) Disruption of snRNP biogenesis factors Tgs1 and pICln induces phenotypes that mirror aspects of SMN-Gemins complex perturbation in Drosophila, providing new insights into spinal muscular atrophy. *Neurobiology of disease*, **94**, 245-258.
61. Di Giorgio, M.L., Esposito, A., Maccallini, P., Micheli, E., Bavasso, F., Gallotta, I., Vernì, F., Feiguin, F., Cacchione, S., McCabe, B.D. *et al.* (2017) WDR79/TCAB1 plays a conserved role in the control of locomotion and ameliorates phenotypic defects in SMA models. *Neurobiology of disease*, **105**, 42-50.
62. McWhorter, M.L., Monani, U.R., Burghes, A.H. and Beattie, C.E. (2003) Knockdown of the survival motor neuron (Smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding. *J Cell Biol*, **162**, 919-931.
63. Hosseinibarkooie, S., Peters, M., Torres-Benito, L., Rastetter, R.H., Hupperich, K., Hoffmann, A., Mendoza-Ferreira, N., Kaczmarek, A., Janzen, E., Milbradt, J. *et al.* (2016) The Power of Human Protective Modifiers: PLS3 and CORO1C Unravel Impaired Endocytosis in Spinal Muscular Atrophy and Rescue SMA Phenotype. *American journal of human genetics*, **99**, 647-665.
64. Oprea, G.E., Krober, S., McWhorter, M.L., Rossoll, W., Muller, S., Krawczak, M., Bassell, G.J., Beattie, C.E. and Wirth, B. (2008) Plastin 3 is a protective modifier of autosomal recessive spinal muscular atrophy. *Science*, **320**, 524-527.
65. Tisdale, S. and Pellizzoni, L. (2015) Disease mechanisms and therapeutic approaches in spinal muscular atrophy. *J Neurosci*, **35**, 8691-8700.
66. Becker, D., Hirsch, A.G., Bender, L., Lingner, T., Salinas, G. and Krebber, H. (2019) Nuclear Pre-snRNA Export Is an Essential Quality Assurance Mechanism for Functional Spliceosomes. *Cell Rep*, **27**, 3199-3214 e3193.
67. Goldfarb, K.C. and Cech, T.R. (2013) 3' terminal diversity of MRP RNA and other human noncoding RNAs revealed by deep sequencing. *BMC molecular biology*, **14**, 23.
68. Cho, H.D., Tomita, K., Suzuki, T. and Weiner, A.M. (2002) U2 small nuclear RNA is a substrate for the CCA-adding enzyme (tRNA nucleotidyltransferase). *J Biol Chem*, **277**, 3447-3455.
69. Rizzo, F., Nizzardo, M., Vashisht, S., Molteni, E., Melzi, V., Taiana, M., Salani, S., Santonicola, P., Di Schiavi, E., Bucchia, M. *et al.* (2019) Key role of SMN/SYNCRIP and RNA-Motif 7 in spinal muscular atrophy: RNA-Seq and motif analysis of human motor neurons. *Brain : a journal of neurology*, **142**, 276-294.

70. Doktor, T.K., Hua, Y., Andersen, H.S., Broner, S., Liu, Y.H., Wieckowska, A., Dembic, M., Bruun, G.H., Krainer, A.R. and Andresen, B.S. (2017) RNA-sequencing of a mouse-model of spinal muscular atrophy reveals tissue-wide changes in splicing of U12-dependent introns. *Nucleic Acids Res*, **45**, 395-416.
71. Jangi, M., Fleet, C., Cullen, P., Gupta, S.V., Mekhoubad, S., Chiao, E., Allaire, N., Bennett, C.F., Rigo, F., Krainer, A.R. *et al.* (2017) SMN deficiency in severe models of spinal muscular atrophy causes widespread intron retention and DNA damage. *Proc Natl Acad Sci U S A*, **114**, E2347-E2356.
72. Huo, Q., Kayikci, M., Odermatt, P., Meyer, K., Michels, O., Saxena, S., Ule, J. and Schumperli, D. (2014) Splicing changes in SMA mouse motoneurons and SMN-depleted neuroblastoma cells: evidence for involvement of splicing regulatory proteins. *RNA biology*, **11**, 1430-1446.
73. Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M. and Dreyfuss, G. (2008) SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell*, **133**, 585-600.
74. Lotti, F., Imlach, W.L., Saieva, L., Beck, E.S., Hao le, T., Li, D.K., Jiao, W., Mentis, G.Z., Beattie, C.E., McCabe, B.D. *et al.* (2012) An SMN-dependent U12 splicing event essential for motor circuit function. *Cell*, **151**, 440-454.
75. Bray, N.L., Pimentel, H., Melsted, P. and Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology*, **34**, 525-527.
76. Peabody, N.C., Diao, F., Luan, H., Wang, H., Dewey, E.M., Honegger, H.W. and White, B.H. (2008) Bursicon functions within the Drosophila CNS to modulate wing expansion behavior, hormone secretion, and cell death. *J Neurosci*, **28**, 14379-14391.
77. Jia, Y., Mu, J.C. and Ackerman, S.L. (2012) Mutation of a U2 snRNA gene causes global disruption of alternative splicing and neurodegeneration. *Cell*, **148**, 296-308.
78. Tisdale, S., Lotti, F., Saieva, L., Van Meerbeke, J.P., Crawford, T.O., Sumner, C.J., Mentis, G.Z. and Pellizzoni, L. (2013) SMN is essential for the biogenesis of U7 small nuclear ribonucleoprotein and 3'-end formation of histone mRNAs. *Cell Rep*, **5**, 1187-1195.
79. Ruggiu, M., McGovern, V.L., Lotti, F., Saieva, L., Li, D.K., Kariya, S., Monani, U.R., Burghes, A.H. and Pellizzoni, L. (2012) A role for SMN exon 7 splicing in the selective vulnerability of motor neurons in spinal muscular atrophy. *Mol Cell Biol*, **32**, 126-138.
80. Simon, C.M., Dai, Y., Van Alstyne, M., Koutsioumpa, C., Pagiazitis, J.G., Chalif, J.I., Wang, X., Rabinowitz, J.E., Henderson, C.E., Pellizzoni, L. *et al.* (2017) Converging Mechanisms of p53 Activation Drive Motor Neuron Degeneration in Spinal Muscular Atrophy. *Cell Rep*, **21**, 3767-3780.
81. Simon, C.M., Van Alstyne, M., Lotti, F., Bianchetti, E., Tisdale, S., Watterson, D.M., Mentis, G.Z. and Pellizzoni, L. (2019) Stasimon Contributes to the Loss of Sensory Synapses and Motor Neuron Death in a Mouse Model of Spinal Muscular Atrophy. *Cell Rep*, **29**, 3885-3901 e3885.
82. Van Alstyne, M., Simon, C.M., Sardi, S.P., Shihabuddin, L.S., Mentis, G.Z. and Pellizzoni, L. (2018) Dysregulation of Mdm2 and Mdm4 alternative splicing

underlies motor neuron death in spinal muscular atrophy. *Genes Dev*, **32**, 1045-1059.

83. Son, A., Park, J.E. and Kim, V.N. (2018) PARN and TOE1 Constitute a 3' End Maturation Module for Nuclear Non-coding RNAs. *Cell Rep*, **23**, 888-898.
84. Lardelli, R.M. and Lykke-Andersen, J. (2020) Competition between maturation and degradation drives human snRNA 3' end quality control. *Genes Dev*, **34**, 989-1001.
85. Zhao, D.Y., Gish, G., Braunschweig, U., Li, Y., Ni, Z., Schmitges, F.W., Zhong, G., Liu, K., Li, W., Moffat, J. *et al.* (2016) SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature*, **529**, 48-53.
86. Kannan, A., Bhatia, K., Branzei, D. and Gangwani, L. (2018) Combined deficiency of Senataxin and DNA-PKcs causes DNA damage accumulation and neurodegeneration in spinal muscular atrophy. *Nucleic Acids Res*, **46**, 8326-8346.
87. Osman, E.Y., Van Alstyne, M., Yen, P.F., Lotti, F., Feng, Z., Ling, K.K., Ko, C.P., Pellizzoni, L. and Lorson, C.L. (2020) Minor snRNA gene delivery improves the loss of proprioceptive synapses on SMA motor neurons. *JCI Insight*, **5**.
88. Mendell, J.R., Al-Zaidy, S., Shell, R., Arnold, W.D., Rodino-Klapac, L.R., Prior, T.W., Lowes, L., Alfano, L., Berry, K., Church, K. *et al.* (2017) Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *The New England journal of medicine*, **377**, 1713-1722.
89. Hosseinibarkooie, S., Schneider, S. and Wirth, B. (2017) Advances in understanding the role of disease-associated proteins in spinal muscular atrophy. *Expert review of proteomics*, **14**, 581-592.
90. Finkel, R.S., Chiriboga, C.A., Vajsar, J., Day, J.W., Montes, J., De Vivo, D.C., Yamashita, M., Rigo, F., Hung, G., Schneider, E. *et al.* (2016) Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. *Lancet*, **388**, 3017-3026.
91. Sumner, C.J. and Crawford, T.O. (2018) Two breakthrough gene-targeted treatments for spinal muscular atrophy: challenges remain. *The Journal of clinical investigation*, **128**, 3219-3227.
92. Talbot, K. and Tizzano, E.F. (2017) The clinical landscape for SMA in a new therapeutic era. *Gene Ther*, **24**, 529-533.
93. Janzen, E., Mendoza-Ferreira, N., Hosseinibarkooie, S., Schneider, S., Hupperich, K., Tschanz, T., Grysko, V., Riessland, M., Hammerschmidt, M., Rigo, F. *et al.* (2018) CHP1 reduction ameliorates spinal muscular atrophy pathology by restoring calcineurin activity and endocytosis. *Brain : a journal of neurology*, **141**, 2343-2361.

FIGURE LEGENDS

Figure 1. Knockdown of *cetgs-1 or cesmn-1* in D-type MNs results in similar phenotypes that are rescued by ceTGS-1 and hTGS1 overexpression

(A) The promoters of the p*cetgs-1a* and p*cetgs-1b* isoforms drive the expression of GFP in MNs of the ventral cord. D-type (GABA) MNs express RFP under the control of the MN specific promoter (p*unc-47*). In merged images, the D-type MNs expressing both *cetgs-1 (GFP) and* p*unc-47 (RFP)* are marked by asterisks. The cells expressing only *cetgs-1(GFP)* are MNs other than D-type (arrowheads). Anterior is left and ventral is down in all images. Scale bar, 50 μM.

(B) Knockdown of *cetgs-1 or cesmn-1* in D-type MNs leads to similar locomotion defects. Wild type (Ctr) and *kal-1* silenced animals are controls. The locomotion defect elicited by *cesmn-1* silencing is rescued by *ceTGS-1* but not by *dsRED* expression. Bars represent the fraction of animals with normal backward locomotion ± SEM, from at least two independent lines/clones. Numbers within bars are the animals tested. **** p<0.0001; (one-way ANOVA).

(C) Transgenic worms expressing GFP [*oxIs12*(p*unc-47::GFP*)] show 19 D-type MNs of the ventral cord (Ctr). *cetgs-1* (*RNAi*) worms display fewer GFP expressing MNs in the ventral cord compared to controls. This phenotype is partially rescued by hTGS1 expression. MNs are indicated by arrowheads; the RNAi construct is not expressed in the tail, where one GFP positive cell is always detectable (asterisk); the heads (underlined) express the p*chs-2::GFP* injection marker in both [*cetgs-1* (*RNAi*)] and [*cetgs-1* (*RNAi*); *hTGS1*] worms; the latter were also injected with the p*odr-1::*RFP marker (arrow). Scale bar, 75 μM.

(D) Quantification of ventral cord D-type MNs. Neuron loss caused by *cetgs-1* knockdown is partially rescued by pan-neuronal expression of human *TGS1* (*hTGS1*) but not of *dsRED*. Each bar represents the mean number of visible MNs from at least two independent lines/clones ± SEM. Numbers within the bars are the animals tested. **** P<0.0001; (One-way ANOVA)

(E) In *oxIs12* [*punc-47::GFP*] transgenic animals, both MN cell bodies and axons are visible; anterior is right and ventral is down. In control (Ctr) and *kal-1(RNAi)* animals, commissures appear as single, straight axons directed to the upper side. *cetgs-1* knockdown worms exhibit commissures with extra branching and guidance defects (arrows).
(F) Apoptotic autofluorescence signals (arrowheads) in dying MNs of *cetgs-1* (*RNAi)* and *cesmn-1(RNAi Is*) animals.
(G) Quantification of dying MNs. Bars represent the average number of dying MNs in transgenic animals from at least two independent lines/clones ± SEM. Numbers within and next to bars are the animals tested. **** p<0.0001; (One-way ANOVA).

Figure 2. Tgs1 overexpression ameliorates the wing-expansion defects caused by neuronal knockdown of *Smn* in *Drosophila*

(A) Examples of flies showing wing expansion failure upon nsyb-GAL4 driven (neur>) *Smn* RNAi.

(B) Frequencies of the defective wing expansion phenotype in flies co-expressing a *UAS-Smn RNAi* (*Smn RNAi*) construct and the indicated UAS-constructs driven by *nsyb-GAL4* (*neur*>). *UAS dTgs1* encodes GFP-tagged dTgs1; *UAS hTGS1* encodes GFP-tagged human TGS1; *UAS CTR* is a control construct expressing the unrelated Mst protein. *ΔSmn/+* flies are heterozygous for a deficiency of the *Smn* locus. Error bars: ± SEM. Numbers within and next to bars are the animals tested. p values: One-way ANOVA with Tukey's multiple comparisons test.

Figure 3. *Tgs1* downregulation in zebrafish leads to caudal primary motor neurons defects

(A) Lateral views of whole-mount embryos immunostained with a synaptotagmin antibody (Znp1) that labels the caudal primary motor neurons (CaP-MN). Embryos were untreated or injected with the indicated morpholinos (MO). Injection of either *tgs1* ATG-

MO or *tgs1* Sp-MO results in truncated or absent motor axons (solid arrowheads) and terminally-branched axons (open arrowheads). Scale bar, 25 µm. See also Figure S1.

(B) Based on overall appearance, CaP-MNs were classified as: normal, short (truncated axonal projection) or absent (axonal atrophy). Based on terminal branching, axons were classified as normal, mild (branching ventral from midline), intermediate (2-3 or more branches at ventral or midline) or severe (>3 branches ventral or dorsal from midline). Zebrafish larvae injected with 1.5 ng of *tgs1 Sp*-MO or 2 ng *tgs1* ATG-MO displayed CaP-MN defects compared to control (non-targeting) MO and uninjected fish. Results are percentages from 3 independent experiments (n = axons analysed; N= animals tested. ****, p<0.0001 Chi-square test); n.s. not significant. See also Figure S1D.

Figure 4. TGS1 deficiency affects maturation of U2 snRNAs in human cells

(A) Western blotting (WB) showing *TGS1* expression in two independent CRISPR clones of HeLa cells (M1, M2) . CTR, parental cell line; *M1R* and *CTR R*, M1 and parental line stably expressing FLAG-TGS1. Tubulin is a loading control. The abundance of the SMN protein is not affected by mutations in *TGS1*.

(B) Total RNA from control (CTR) and *TGS1* mutant cells (*TGS1* M1 and M2) precipitated with an anti-TMG CAP antibody or control IgGs. MMG-capped Actin mRNA is a negative control. In *TGS1* mutant cells, trimethylation of the U2 snRNAs is strongly reduced compared to controls. Data are from 3 independent experiments; p value: two-way ANOVA with Sidak's multiple comparisons test.

(C) RT-qPCR showing the abundances of the indicated snRNAs in *TGS1* mutant cells (*TGS1 M1* and *M2*) or rescued cells (*TGS1 M1R*). Data from three independent experiments are relative to parental HeLa cells (set to 1) and normalized to 5.8S rRNA. The U3 snoRNA (U3) and the uncapped U6 and U6atac snRNAs are controls. No statistically significant differences were observed between *TGS1 M* and *TGS1 M1R* cells (error bars: ± SEM; p values >0.05, calculated as in B)

(D) Northern blots showing U2 snRNA precursors (arrow) in nuclear and cytoplasmic fractions from HeLa cells of the indicated genotypes; the U3 and 7SL RNAs are loading

controls for the nuclear and cytoplasmic fractions, respectively. CTR, parental HeLa cell line; M1 and M2, *TGS1* mutant cells. M1R, *TGS1* M1 rescued cells. *PAPD5*, HeLa clone carrying mutations in *PAPD5*, used as negative control.

(**E**) Quantification of the blots in (D) by densitometric analysis. Bars: percentage of extended U2 molecules relative to total U2 (± SEM). p values: One-way ANOVA with post-hoc Tukey's test.

Figure 5. TGS1 loss leads to an accumulation of 3' extended U2 and U4atac snRNAs in human cells

(A) 3'RACE sequencing on total RNA showing that *TGS1* mutant HeLa cells accumulate more U2-extended snRNA molecules than either parental (CTR) or *TGS1* M1 rescued cells (TGS1 M1R). UMUC3 cells treated with *TGS1* siRNAs (siTGS1) accumulate more U2-extended snRNAs than cells exposed to non-targeting siRNAs (siSCR). Nascent and steady-state RNA fractions were purified from control HeLa cells (see Methods). Extended U2 molecules carry extra 3' sequences of templated and /or untemplated nucleotides . Total reads > 36k. Data are mean values of two independent clones + SEM. p values: two-sided Student's t test.

(B) Plots showing the percentage of the different sequence reads of U2 snRNAs in the same samples as in A. Numbers on the x-axis: additional nucleotides beyond the 189 nt of the mature form. Position 0 includes mature U2 snRNA species of 188 nt or 188 nt plus a post-transcriptionally added A. y-axis: percent of total reads. Blue: genomic-templated nucleotides; green: untemplated adenosine; orange: untemplated uridine.

(C) Characterization of U2 snRNA molecules from nuclear and cytoplasmic fractions from M1 and M2 mutant (*TGS1 M*), M1 rescued (*TGS1 M1R*), or control (*CTR*) cells.

(D) Tail length and composition U2 snRNAs in the C samples. y-axis: percent of total reads.

(E) Percentages of U2 molecules with tails ending with post-transcriptionally added uridines in the RNA samples described in A. Total reads per sample >75k.

(F) Percentages of 3' extended U4atac snRNAs in the RNA samples described in A. Total reads per sample >11k.

(G) Tail lengths and composition of the 3' ends of U4atac snRNAs in M1 and M2 mutant, M1 rescued (M1R), and control (CTR) cells. Numbers on the x-axis: additional nucleotides beyond the 130nt of mature U4atac (position 0). y-axis: percent of total reads.

(H) Percentages of U4atac snRNA molecules with extended tails ending with untemplated uridines. Total reads per sample >75k.

Figure 6. **3' extended U2 snRNA molecules accumulate in *Drosophila dTgs1* mutants**

(A) Percentage of extended U2 snRNA molecules determined by 3' RACE and sequencing on RNA from equally aged second or third instar mutant larvae, rescued mutant larvae (bearing a *dTgs1* construct), and wild type larvae (CTR). The most abundant population of U2 snRNAs is 192 nt long. Extended U2 snRNAs are longer than 192 nt and incorporate templated or untemplated nt at their 3' ends. Total reads > 2900.

(B) Tail length and composition of the 3' ends of U2 snRNAs for the second and third instar larvae described in A. Numbers on the x-axis: additional nucleotides beyond the mature form of 192 nt (position 0). Blue, genomic- encoded nucleotides; green, untemplated adenosine; orange, templated or untemplated uridine. y-axis: percent of total reads.

Figure 7. Mutations in *TGS1* and *SMN* cause global changes in RNA expression and splicing

(A, B) Heatmaps depicting the expression levels (TPM values estimated by Kallisto) for differentially expressed (DE) transcripts in *TGS1 M*1 and *TGS1 M2* (A), or *SMN C1* and *SMN C2* (B) mutant cells, compared to parental HeLa cells (CTR). Transcripts are classified into 5 types: annotated transcripts per GENCODE (orange); novel transcripts with extended 3' (magenta) or shortened 3' (blue); novel transcripts with annotated identical 3' (pink); and other (intergenic and opposite-strand) novel transcripts (purple).

Transcripts within each group are ranked by unsupervised clustering. Total transcript number for each group, and the DE status for each transcript (either up- or down-regulated in mutant cells) are annotated by the green/red sidebar to the right (see also Table S1).

(C) Venn diagram from data in A and B, showing the number of shared transcript isoforms between *TGS1* and *SMN* mutant cells. See also Table S1.

(D) Methodological approach used for quantification of intron retention (IR) levels (see also Methods).

(E, F) Scatter plots showing for each IR event between mutant (*TGS1* or *SMN*) and CTR cells the differential PSI (dPSI) against the differential expression level (dTPM) of the major transcript (intron exclusion). The significant IR events are color-coded in the plots, with significantly upregulated events colored in orange and significantly downregulated in teal.

(G) Roles of SMN and TGS1 in snRNA biogenesis.

(H) A model for the roles of TGS1 and SMN in prevention of neurodegeneration **.**The effects of TGS1 or SMN deficiency on snRNA biogenesis and general transcription suggest that these proteins play interconnected roles in prevention of neurodegeneration.

Figure 1

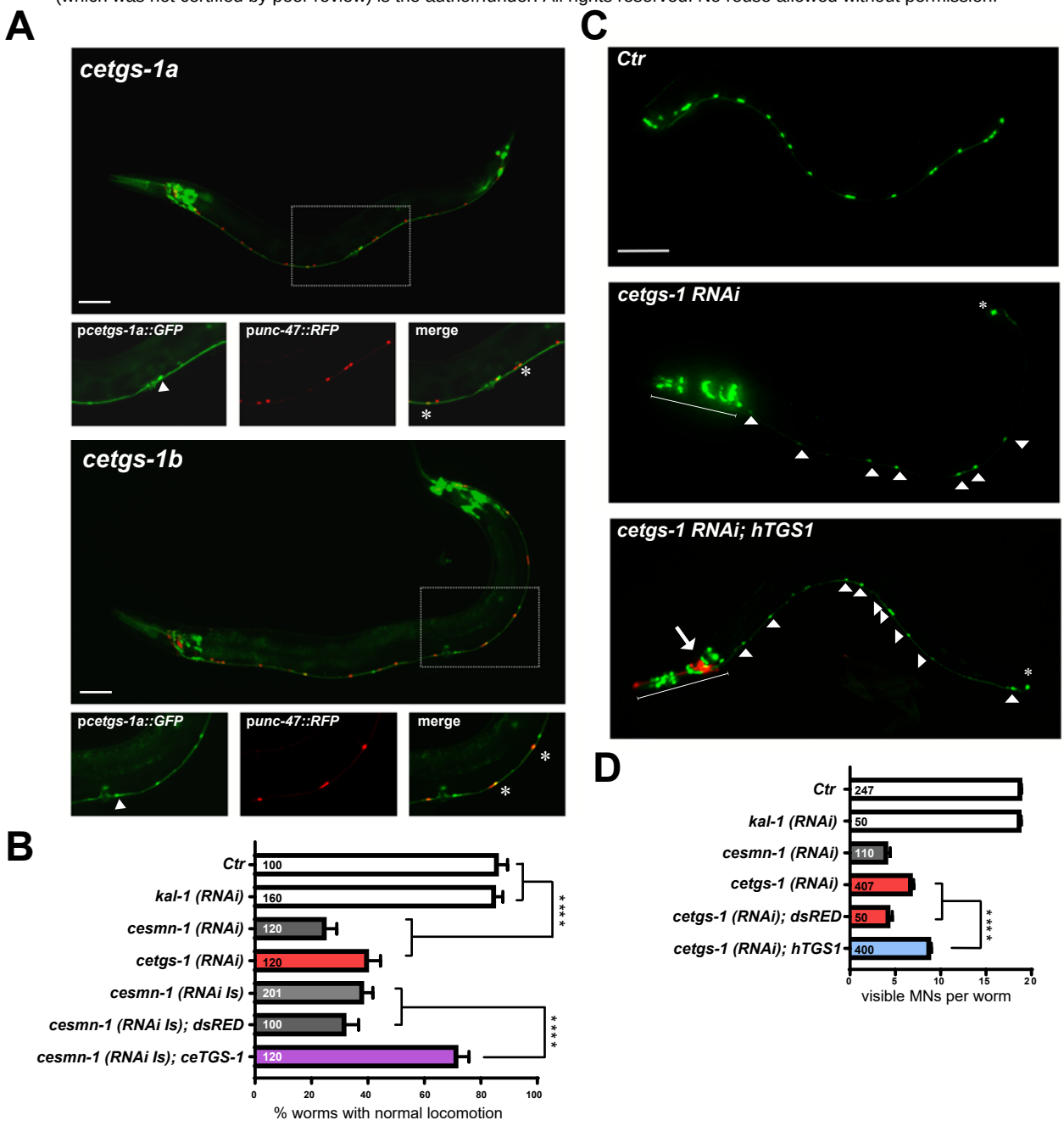

Figure 2

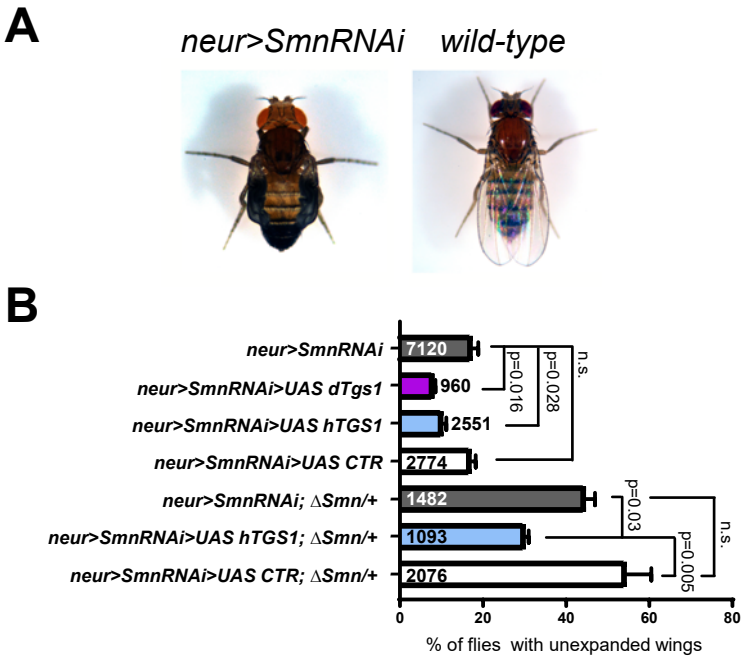

Figure 3

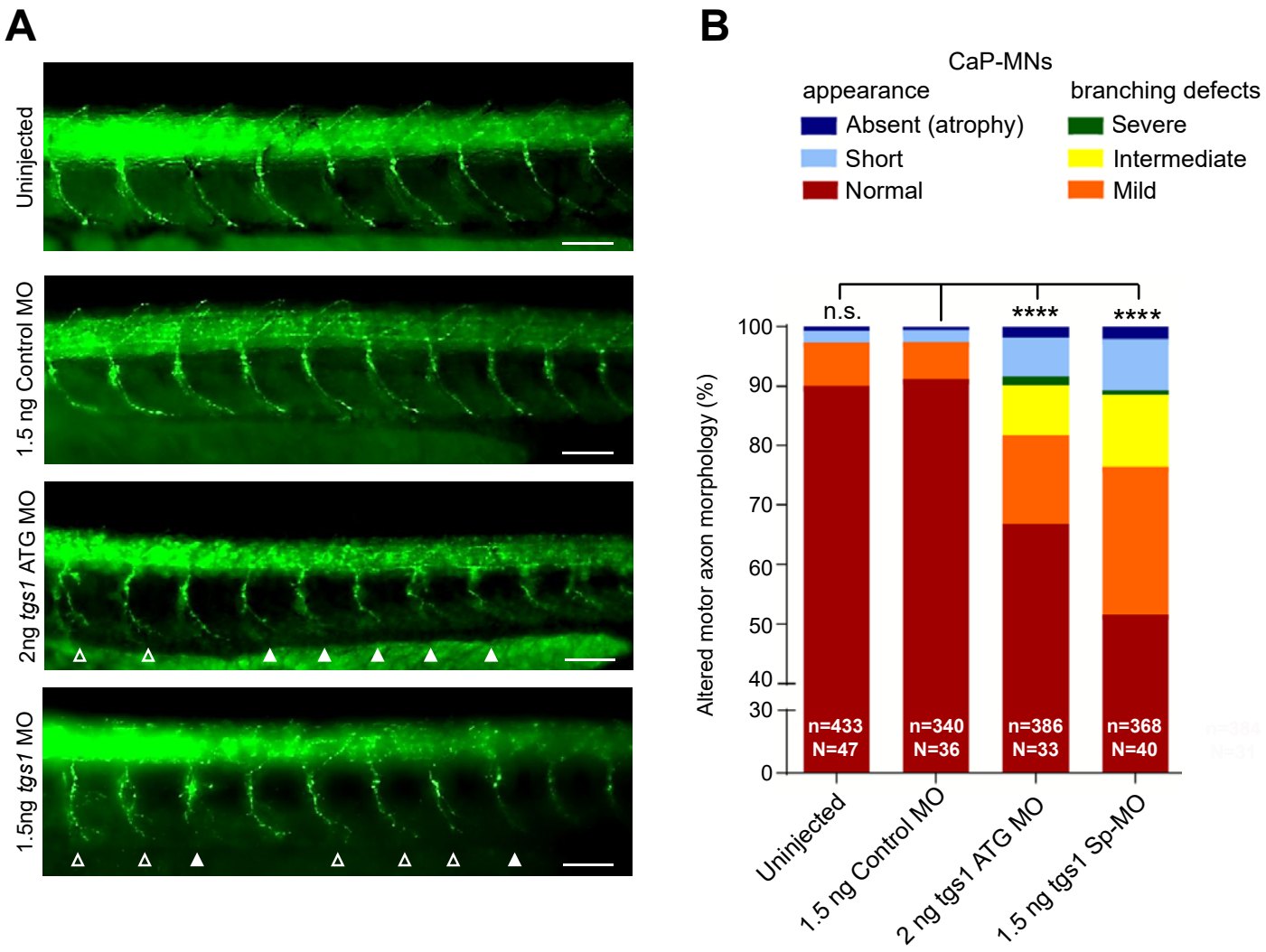

Figure 4

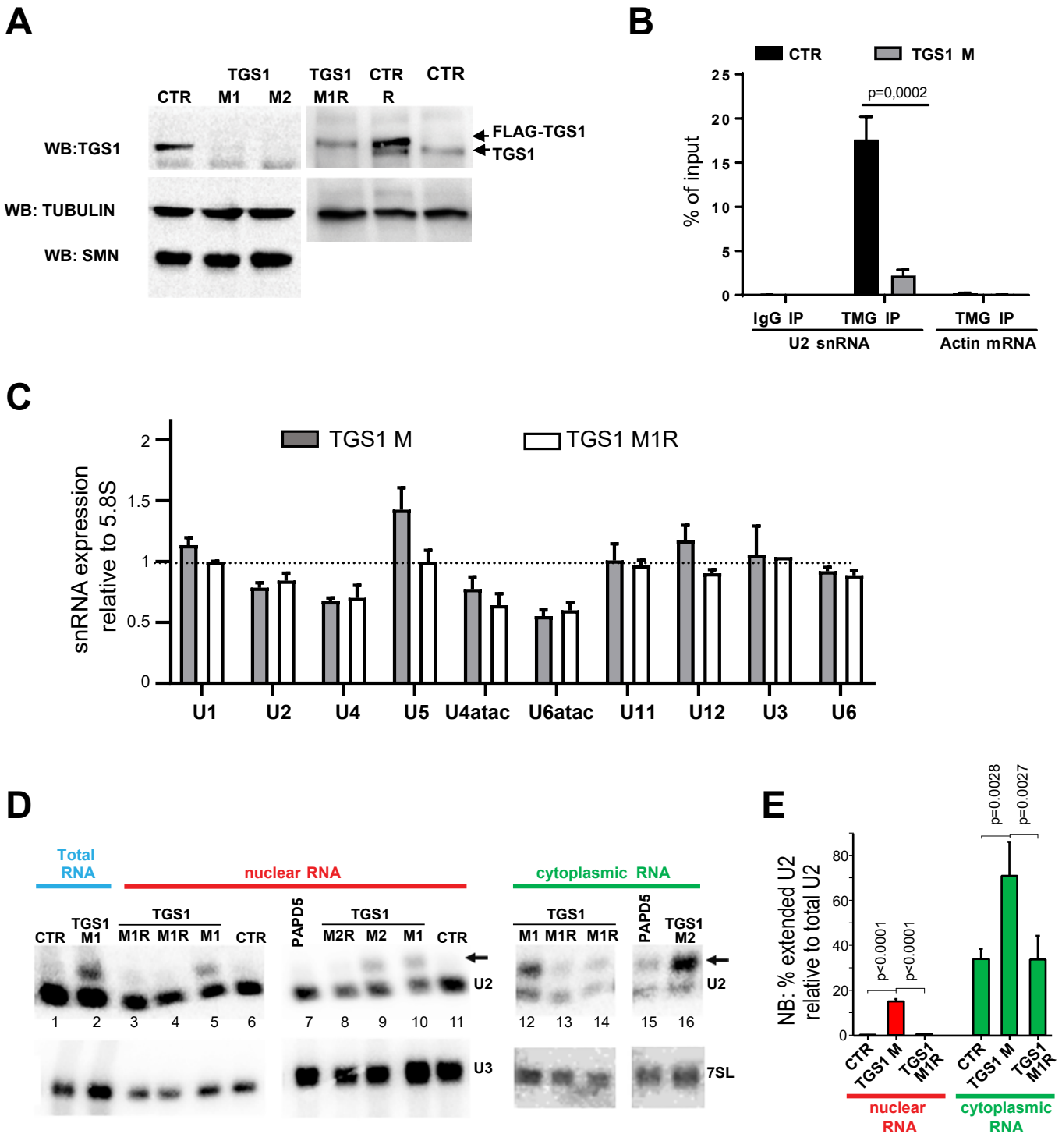

Figure 5

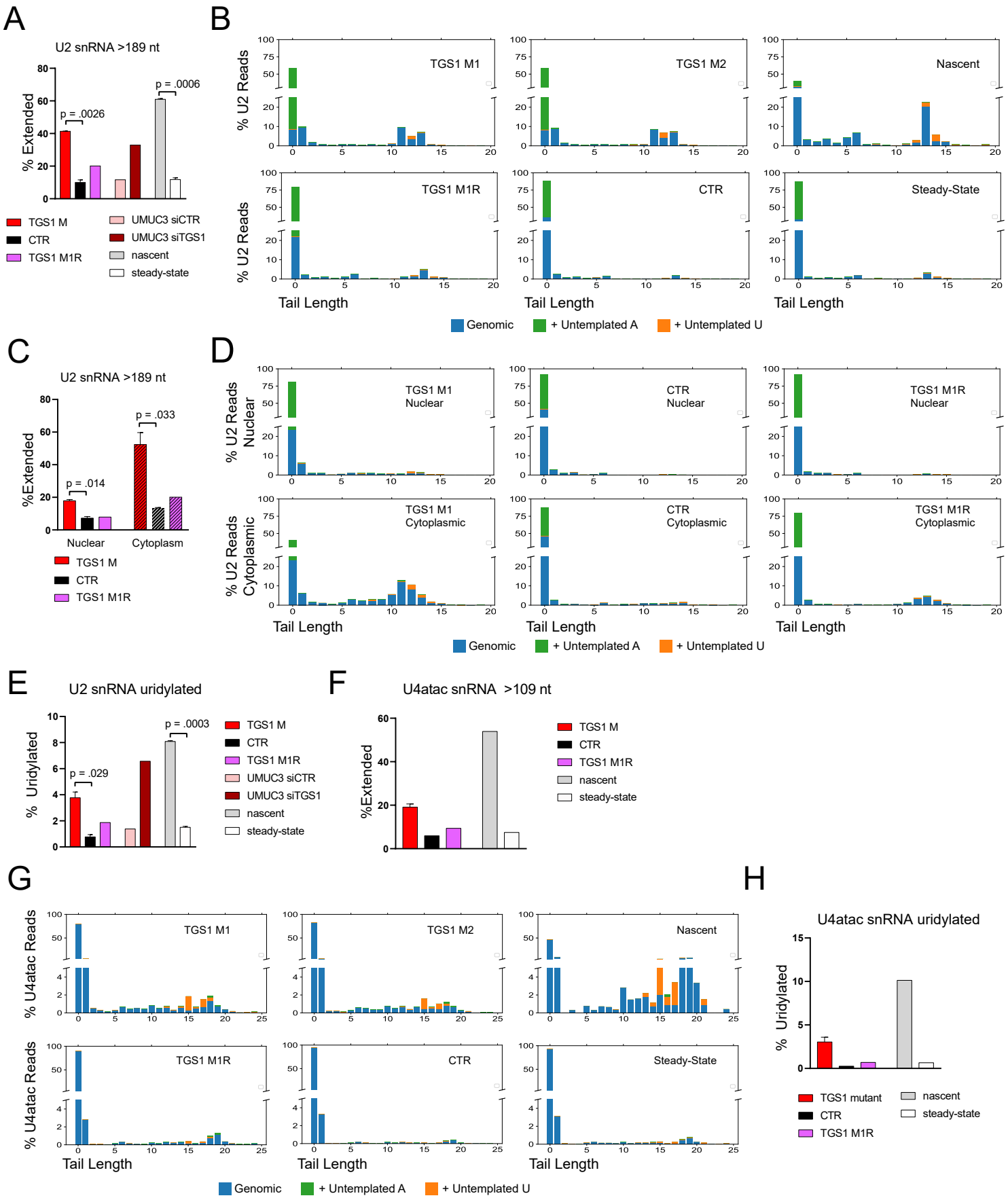

Figure 6

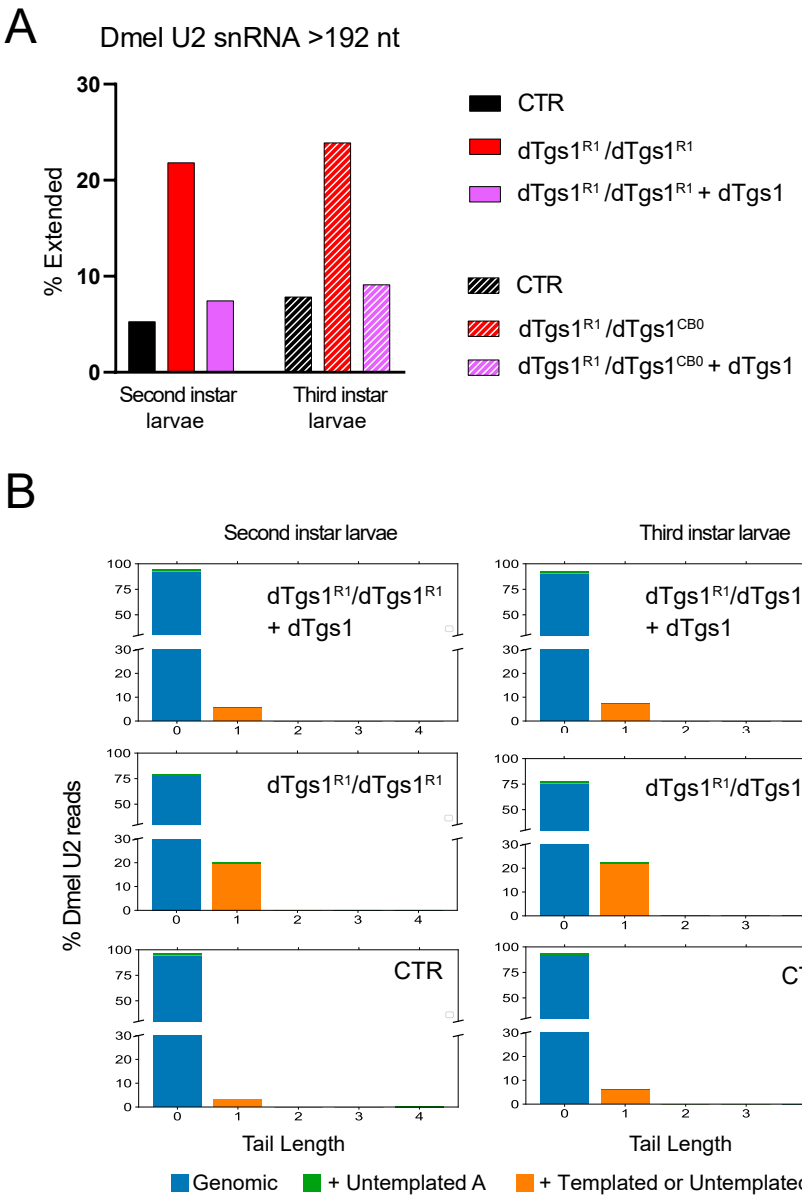

Figure 7