CHARACTERIZATION OF THE MAMMALIAN METHYLTRANSFERASE TGS1

By

Li Chen

B.S., Wuhan University, 2010

Submitted to the graduate degree program in Molecular & Integrative Physiology and the Graduate Faculty of the University of Kansas Medical Center in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Co-Chair: Peter Baumann

_____

Co-Chair: Lane Christenson

_____

Ron Yu

_____

Ron Conaway

_____

Timothy Fields

Date Defended: 27 June 2018

The dissertation committee for Li Chen certifies that this is the approved version of the following dissertation:

CHARACTERIZATION OF THE MAMMALIAN METHYLTRANSFERASE TGS1

_____

Co-Chair: Peter Baumann

_____

Co-Chair: Lane Christenson

Date Approved: 27 June 2018

Abstract

In eukaryotes, RNAs transcribed by RNA polymerase II are capped at their 5' end. The cap is added co-transcriptionally and, in the case of mRNAs, contains one methyl group, resulting in a monomethylguanosine cap. In the case of some noncoding RNAs, two additional methyl groups are added, resulting in a 2,2,7-trimethylguanosine cap. TMG-capped RNAs include snRNAs (U1, U2, U4, U5, U7), snoRNAs (U3, U8, U13) and telomerase RNA. The enzyme responsible for the cap hypermethylation is methyltransferase Tgs1, initially identified as PIMT in human and mouse. Tgs1 deletion in yeasts results in mild cold-sensitive phenotype. However, early embryonic lethality has been reported in Tgs1-KO mice. We identified a mutant version of Tgs1 in the putative Tgs1-KO mouse cell line. Although the mutant Tgs1 retains its methyltransferase motif, it results in processing defect of U2 snRNA. To further understand the function of mammalian Tgs1, we have used genome editing to generate a human cell line deleted for endogenous Tgs1 and carrying a tetracycline-inducible version. Repression of human Tgs1 expression to undetectable level results in a severe growth phenotype in cell culture. The effects of Tgs1-depletion on RNA cap hypermethylation and gene expression were analyzed using the human conditional knockdown cell line. Our study provides new insights on a previously reported Tgs1-KO mouse model, generates new tools for characterization of mammalian Tgs1, and provides comprehensive information on the impact of Tgs1-depletion in mammalian cells.

Acknowledgements

Writing this dissertation has been an intense learning experience for me. It would not have been possible without guidance and support from many individuals. Although I cannot include everyone that has made an impact, the following people deserve special mention:

First, I must thank my mother, for her great role in my life and her love and support throughout this journey. I also want to thank my deceased father, whose encouragement set me on the path to pursuing graduate studies abroad. Your memory is always with me.

I would like to express great gratitude to my mentor Dr. Peter Baumann. Your guidance and support throughout my thesis project is much appreciated. I am grateful for your patience when things didn't turn out as expected. Much thanks to my committee members Dr. Lane Christenson, Dr. C. Ron Yu, Dr. Ron Conaway and Dr. Timothy Fields, whose invaluable advice helped shape my research. Special thanks to Dr. Susan Abmayr and Dr. Michael Wolfe, your input during the last several months has been vital for me to complete my writing.

Thanks to every past and present member in the Baumann lab. I was very fortunate to be in a great working environment created by Ram, Aracely, Lili, Wen, Maggie, Diego, Cameron, Dai, Rob, Hui-Fang, Chi-Kang, Katie, Rachel, Lisa, Aaron, Allie et al. Special thanks to Rutendo and Morgan for their bioinformatics support.

Finally, I want to thank my friends and colleagues from both Stowers Institute and University of Kansas Medical Center. Thank you for everything you have done.

Table of Contents

List of Abbreviations

MMG: Monomethylguanosine

TMG: Trimethylguanosine

CBC: Cap binding complex

AdoMet: S-adenosyl-L-methionine

AdoHcy: S-adenosyl-L-homocysteine

PIMT: PRIP-interacting protein with methyltransferase domain

PRIP: PPAR interacting protein

PPAR: Peroxisome proliferator-activated receptor

CBP: CREB binding protein

CREB: cAMP response element binding protein

PBP: PPAR binding protein

P300: Adenovirus E1A-associated 300-kDa protein

hTR: Human telomerase RNA

TGIRT: Thermostable group II intron reverse transcriptase

ZFN: Zinc finger nuclease

CRISPR: Clustered regularly interspaced short palindromic repeats

List of Figures

Chapter 1 Introduction

1.1 5’ cap of PolII transcribed RNAs in eukaryotes

As RNAs are synthesized from 5’ to 3’, the 5’ termini provide the first chance for modification. Nascent transcripts of RNA polymerase II are modified with a guanosine nucleotide via triphosphate linkage. For mRNAs, the guanosine is methylated on the N7 position, resulting in a 7-monomethylguanosine cap (Shatkin, 1976). The m7G cap is required for mRNA stability, protects the 5’ end of the RNA from exoribonuclease degradation, and facilitates translation initiation (Hallais et al., 2013; K. M. Kim et al., 2009). For snRNAs, snoRNAs, and telomerase RNAs the m7G cap is further methylated on the N2 position, forming a 2,2,7-trimethylguanosine cap (Baserga, Gilmore-Hebert, & Yang, 1992; Franke, Gehlen, & Ehrenhofer-Murray, 2008; Jacobson & Pederson, 1998; Mattaj, 1986; Terns & Dahlberg, 1994).

1.2 Trimethylguanosine cap

The hypermethylation of the RNA 5’ cap in RNA was first identified in U1, U2 snRNA and U3 snoRNA (Reddy et al., 1974; Reddy et al., 1972). A guanosine molecule is linked to the 5’ end of an RNA via a triphosphate bridge (Figure 1.2.1). The N7 position of the guanosine contains one methyl group and the N2 position contains two methyl groups, constituting a 2,2,7-trimethylguanosine (TMG) cap.

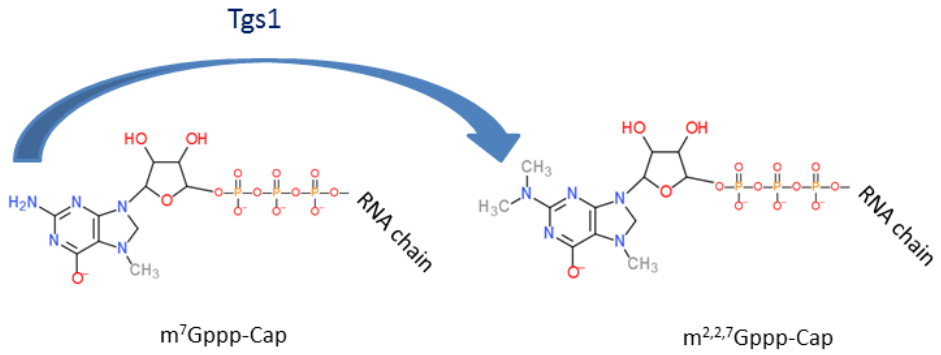


Figure 1.2.1 Hypermethylation of monomethylgaunosine cap to trimethylguanosine cap.

The TMG cap is present in a variety of small nuclear RNAs (U1, U2, U4, U5, U7, U11, and U12) and small nucleolar RNAs (the C/D and H/ACA RNAs, more than 200 in eukaryotic cells) (Matera, Terns, & Terns, 2007). Vertebrate telomerase RNA, classified as a member of box H/ACA snoRNA family, is TMG capped, too (Jady, Bertrand, & Kiss, 2004).

The enzyme that is responsible for snRNA and snoRNA TMG cap formation was identified as Tgs1 in yeast (Mouaikel, Verheggen, Bertrand, Tazi, & Bordonne, 2002). The homologs of Tgs1 in human and mouse were identified as PIMT, initially believed to be a transcriptional coactivator (Zhu et al., 2001).

1.2.1 TMG cap on snRNAs and snoRNAs

The hypermethylation of snRNAs depends on the presence of a Sm site and the binding of Sm proteins, which interact with TGS1 (Mattaj, 1986; Mouaikel et al., 2002; Raker, Plessel, & Luhrmann, 1996). Sm proteins are a set of proteins (B/B', D1, D2, D3, E, F and G) that form a

heteroheptameric ring structure around the conserved $AU_{4-6}G$ site flanked by two stem loops (Sm site) (Will & Luhrmann, 2001).

During snRNA biogenesis, the TMG cap is part of a bipartite nuclear localization signal (Hamm, Darzynkiewicz, Tahara, & Mattaj, 1990; Palacios, Hetzer, Adam, & Mattaj, 1997). The biogenesis of snRNAs has both a nuclear phase and a cytoplasmic phase. They are transcribed by RNA polymerase II (except for U6), then transported to the cytoplasm for maturation and assembly into core small nuclear ribonucleoproteins (snRNPs). The core snRNPs are transported back into the nucleus to mediate the splicing of mRNAs. TMG cap formation is required for this nuclear re-import in *Xenopus laevis* oocytes and facilitates nuclear re-import in somatic cells (Fischer, Heinrich, van Zee, Fanning, & Luhrmann, 1994; Kleinschmidt & Pederson, 1990). The TMG cap is bound by Snurportin 1, which promotes re-importation (Huber et al., 1998; Palacios et al., 1997).

For box C/D snoRNAs, the box C/D motif is necessary and sufficient for 5' hypermethylation, and the binding protein Nop58p recruits Tgs1 (Mouaikel et al., 2002; Speckmann, Terns, & Terns, 2000; Terns, Grimm, Lund, & Dahlberg, 1995). For H/ACA snoRNAs, Tgs1 is recruited by dyskerin (Girard et al., 2008). Unlike snRNAs, which are hypermethylated in the cytoplasm, snoRNAs do not have a cytoplasmic phase (Mattaj, 1986; Terns & Dahlberg, 1994; Terns et al., 1995). The TMG-capping of snoRNAs occurs in the nucleus. The TMG cap on snoRNA likely serves as a nuclear retention signal to prevent CRM1/Exportin-1 mediated nuclear exportation (Boulon et al., 2004). Instead of nuclear exportation, CRM1 helps transport TMG-capped snoRNA U3 to nucleoli.

1.2.2 TMG cap on telomerase RNA

Our interest in the trimethylguanosine cap was initiated by studies in telomerase RNA biogenesis. Telomerase is a specialized reverse transcriptase capable of extending the ends of linear chromosomes. It was first identified by Greider and Blackburn in the ciliate *Tetrahymena* in 1985 (Greider & Blackburn, 1985). Later it was shown that telomerase is a ribonucleoprotein and both the protein and RNA components are essential for its activity (Blasco, Funk, Villeponteau, & Greider, 1995; Feng et al., 1995; Greider & Blackburn, 1987; Morin, 1989; Prowse, Avilion, & Greider, 1993). The RNA component (TERC) serves as a template to add telomeric repeats as well as a scaffold for telomerase holoenzyme assembly (G. L. Yu, Bradley, Attardi, & Blackburn, 1990). The protein component (TERT) has a reverse transcriptase motif that catalyzes the addition of telomeric repeats (Lingner et al., 1997). Accessory factors like TCAB1, EST1A, dyskerin, NOP10, GAR1 and NHP2 have been identified and are believed to stabilize and regulate telomerase activity (Q. Zhang, Kim, & Feigon, 2011).

Telomerase activity is detected in human germ cells, embryonic cells, and stem cells but not in most normal somatic cells. Most cancer cells show elevated telomerase activity (N. W. Kim et al., 1994). HeLa cells, derived from a cervical cancer, transfected with antisense telomerase RNA eventually show inhibition of growth and rarely survive after 26 generations, indicating that telomerase has an important role in cell mortality (Feng et al., 1995).

The protein subunit TERT is highly conserved among eukaryotes, whereas the RNA component TERC differs greatly among species. In ciliates, the length of TERC varies from 147 to 202 nt (McCormick-Graham & Romero, 1996); vertebrate TERCs range from 312 to 559 nt (Chen, Blasco, & Greider, 2000; Xie et al., 2008); and yeast TERCs can be up to >2030 nt (Kachouri-Lafond et al., 2009).

Despite the great divergence in sequence and length, fundamental secondary structure elements of TERC are conserved (Chen et al., 2000; Lin et al., 2004). In vertebrates, there are three major structural and functional domains: the core domain, the CR4/CR5 domain, and the box H/ACA domain (Q. Zhang et al., 2011). It is worth noticing that the box H/ACA domain of vertebrate TERC has all the conserved elements of box H/ACA small nucleolar RNAs (Mitchell, Cheng, & Collins, 1999). The domain is also associated with the same four core proteins of H/ACA snoRNPs (C. Wang & Meier, 2004). Altogether, vertebrate telomerase RNA is classified into the box H/ACA snoRNA family (Kiss, Fayet-Lebaron, & Jady, 2010).

The functions of these domains and their associated protein factors have drawn most interest. In contrast, another aspect of vertebrate telomerase RNA, the hypermethylation of the 5' cap, which is conserved through vertebrate and yeasts, has generated less attention.

Vertebrate telomerase RNA is retained in the nucleus (Lukowiak, Narayanan, Li, Terns, & Terns, 2001). So far, no evidence has shown that there is a cytoplasmic phase. The biogenesis of vertebrate telomerase RNA----from transcription by Pol II to maturation and assembly to functional telomerase holoenzyme----is not well understood. Currently we do not know much about the machineries involved in 5' and 3' processing, trafficking, and assembly with other components of telomerase. However, we do know that human telomerase RNA is localized to intranuclear structures called Cajal bodies, where it is hypermethylated (Jady et al., 2004). Cajal bodies have been shown to transiently localize with telomeres during S phase. It has been proposed that telomerase RNA is hypermethylated, assembled into telomerase, and transferred to telomeres as part of Cajal body movement (Jady, Richard, Bertrand, & Kiss, 2006). Since the hypermethylation of telomerase RNA is just one step ahead of the formation of functional telomerase, it probably has an important role in the final maturation of telomerase RNA. There

are several ways in which the TMG cap may affect telomerase function, including telomerase RNA stability, telomerase holoenzyme localization, and activity and the loading of telomerase to telomeres.

Data from other members in our lab have shown that in fission yeast, the telomerase RNA level is decreased to 20% in the absence of TMG cap (Tang, Kannan, Blanchette, & Baumann, 2012). We hypothesize the TMG cap may have a similar role in stabilizing telomerase RNA in vertebrates.

Vertebrate telomerase RNA has several conserved secondary structure elements, changes in which affect either the template function or the binding of accessory factors (Ly, Blackburn, & Parslow, 2003). It is conceivable that the absence of a TMG cap causes conformational changes in telomerase RNA, leading to disruption of telomerase assembly and activity.

Certain protein factors like Snurportin-1 have been shown to recognize the TMG cap and promote the trafficking of the capped RNA. Although Snurportin-1 is primarily localized in the cytoplasm, it is possible that other nuclear factors can recognize the TMG cap of telomerase RNA and promote the trafficking of telomerase. In addition, it is possible that certain components of telomerase holoenzyme can bind to the TMG cap and initiate or facilitate the assembly of the holoenzyme.

In addition, there are many other proteins like CBC, PHAX and CRM1 that bind to a monomethylguanosine (MMG) cap, directly or indirectly (Matera et al., 2007). Should telomerase RNA fail to be hypermethylated, these proteins may bind to MMG-capped telomerase RNA and disrupt the assembly of the telomerase holoenzyme. It is also worth noting that most of these proteins are nuclear exporters. The TMG cap may serve as a Cajal body

retention signal by preventing the binding of these proteins. The Cajal body has been shown to participate in the loading of telomerase onto telomeres (Stern, Zyner, Pickett, Cohen, & Bryan, 2012). It has also been proposed that human telomerase RNA is sequestered in Cajal bodies as a regulatory mechanism to prevent improper elongation of telomeres (Cristofari et al., 2007). The localization of human telomerase RNA to Cajal bodies can be disrupted by mutation in the CAB box and the corresponding protein TCAB1. There is consequently less loading of telomerase to telomeres and a deficiency in telomere maintenance (Fu & Collins, 2003; Jady et al., 2004; Tycowski, Shu, Kukoyi, & Steitz, 2009; Venteicher et al., 2009; Zhong et al., 2011). Thus, the retention of telomerase RNA in Cajal bodies is essential for normal telomerase function.

The potential roles of the TMG cap in vertebrate telomerase RNA biogenesis make it a candidate for anti-telomerase drug design. It has been shown that telomerase reactivation enables malignant tumor progression (Ding et al., 2012). Drugs targeting telomerase biogenesis may thus serve as cancer therapies. The deletion of TGS1 in fission yeasts results in reduced telomerase RNA level and telomere shortening. If mammalian Tgs1 has a similar role in telomerase RNA biogenesis, the capping enzyme Tgs1 itself could be an etiological gene of syndromes of telomere shortening.

1.2.3 TMG cap on mRNAs

Translation initiation factor eIF4E specifically recognizes monomethylguanosine cap on mRNAs (Rutkowska-Wlodarczyk et al., 2008). It is rare that 5’ cap of mRNA undergoes hypermethylation. In *C. elegans*, some mRNAs carry TMG caps through spliced leader trans splicing. This mechanism is not present in higher eukaryotes. However, recent studies show that

mammalian selenoprotein mRNAs are TMG capped and are associated with actively translating ribosomes (Gribling-Burrer et al., 2017; Wurth et al., 2014). Selenoprotein mRNAs contain a UGA codon that is recoded as selenocysteine. A stem loop in the 3' untranslated region, known as the SelenoCysteine insertion sequence (SECIS), is essential for the recoding process (Kryukov et al., 2003). SECIS binding protein 2 (SBP2) binds selenoprotein mRNAs and recruits the SMN complex, which mediates the hypermethylation of selenoprotein mRNAs in a similar way to snRNAs (Gribling-Burrer et al., 2017).

Selenium (Se) plays important roles in human health. Selenium deficiency has been identified as a contributing factor in heart disease, neurological disorders, and inflammation (Labunskyy, Hatfield, & Gladyshev, 2014). SelenoCysteine is the major form of selenium in cells. Although the sequences and structures of selenoproteins are diverse, majority of them serve as oxidoreductases due to the presence of SelenoCysteine.

1.3 3' terminal nucleotides of PolII transcribed RNAs

The addition of 3' terminal nucleotides is another important modification in Pol II transcribed RNAs. In eukaryotes, polyadenylation is a well-known 3' end modification for mRNAs, which are transcribed by Pol II. Just as with 5' end capping, the polyadenylation occurs co-transcriptionally (Millevoi & Vagner, 2010). Nascent mRNAs are cleaved by the cleavage and polyadenylation specificity factor (CPSF) at polyadenylation site, followed by addition of adenosines by the poly (A) polymerase (PAP). The poly (A) tail is important for transportation of

mRNAs from nucleus to cytoplasm, quality control, translation efficiency, and degradation (Bresson & Conrad, 2013; Natalizio & Wente, 2013; Preiss, Muckenthaler, & Hentze, 1998).

Uridylation is a common modification at the 3' end of noncoding RNAs (Choi, Patena, Leavitt, & McManus, 2012). Well known examples include miRNA biogenesis (Heo et al., 2012; Heo et al., 2009; Newman, Mani, & Hammond, 2011). Oligo-uridylation promotes degradation of Let-7 miRNA precursors and reduces the efficiency of Dicer, while mono-uridylation of group II let-7 miRNAs facilitates Dicer processing.

It has also been reported that Tgs1 promotes certain miRNAs expression during quiescence in human primary fibroblasts (Martinez et al., 2017). A group of quiescence-induced primary miRNAs are hypermethylated by Tgs1, which is recognized by CRM1/Exportin-1 pathway. Those TMG capped primary miRNAs are transported to cytoplasm and processed by a short isoform of Drosha.

1.4 3' terminal nucleotides of U2 snRNA

snRNAs are transcribed in the nucleus and transported into cytoplasm, where they undergo TMG-capping and 3' end trimming before being transported back into the nucleus (Kiss, 2004; Matera et al., 2007; Rollenhagen & Pante, 2006). snRNA precursors have been identified in cytoplasm (Kleinschmidt & Pederson, 1987; Rohleder & Wieben, 1986; Wieben, Nenninger, & Pederson, 1985; Yuo, Ares, & Weiner, 1985). In the case of U2, approximately 10 nucleotides extend at the 3' end. After being trimmed back by one or more unknown nucleases, the snRNA is transported back into the nucleus, where splicing occurs. The untrimmed

precursors are bound by Gemin5 and facilitate snRNP assembly (Yong, Kasim, Bachorik, Wan, & Dreyfuss, 2010).

One additional non-templated adenosine is present at 3' end of mature U2, resulting in 3' terminal CCA sequence (Cho, Tomita, Suzuki, & Weiner, 2002). The CCA-adding enzyme, which synthesizes the CCA sequence of tRNAs, can also add A to U2 snRNA substrate *in vitro*. The biological function of this non-templated adenosine is unknown.

Recently, it has been reported that TOE1, an exonuclease, is involved in 3' end trimming of snRNAs (Lardelli et al., 2017). The 3' extended pre-snRNAs accumulate in the absence of TOE1. Moreover, TOE1 associates snRNAs with untemplated 3' adenosines, which suggests that 3' adenylation could serve as a mechanism to recruit TOE1 to pre-snRNAs.

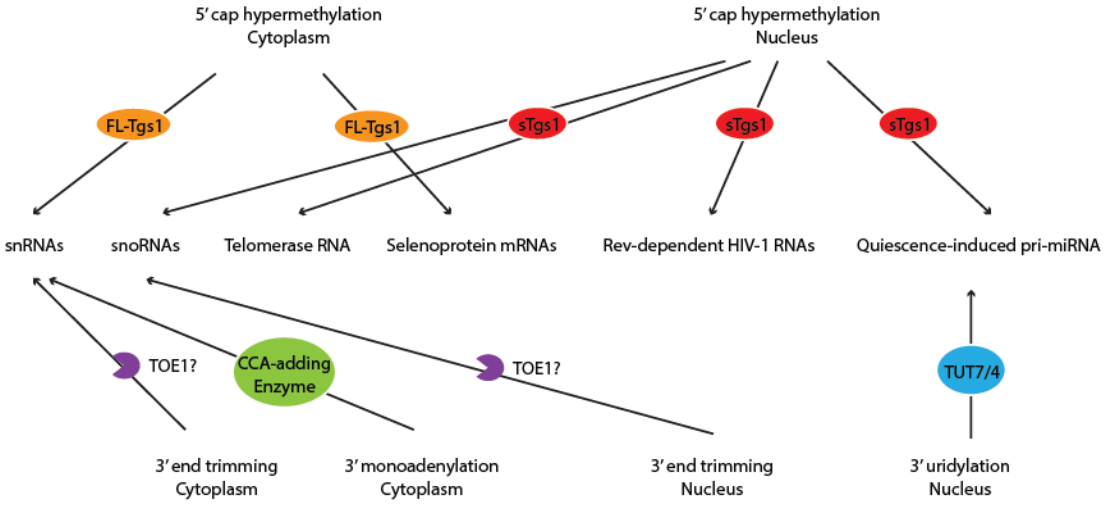


Figure 1.4.1 Summary of the 5' and 3' modifications on each relevant RNA mentioned above and the cellular location of the modification.

1.5 Trimethylguanosine synthase 1 (Tgs1)

Tgs1 is a *bona fide* methyltransferase capable of catalyzing two methylation reactions to the exocyclic nitrogen N2 of the m7G cap through a distributive mechanism (Hausmann & Shuman, 2005). Each reaction transfers one methyl group from S-adenosyl-L-methionine (AdoMet) to the m7G cap, resulting in two S-adenosyl-L-homocysteine (AdoHcy) molecules and an m2,2,7G trimethylguanosine cap.

The Tgs1 protein is evolutionarily conserved on the C-terminal domain, with a variable N-terminal domain in higher eukaryotes (Mouaikel et al., 2002). In budding yeast *Saccharomyces cerevisiae*, the Tgs1 protein consists of 331 amino acids. Fission yeast *Saccharomyces pombe* Tgs1 consists of 239 amino acids. *Caenorhabditis elegans* Tgs1 (T08G11.4) consists of 566 amino acids (Zipperlen, Fraser, Kamath, Martinez-Campos, & Ahringer, 2001). *Drosophila melanogaster* Tgs1 (DTL) consists of 491 amino acids (Komonyi et al., 2005). In mammals, both mouse and human Tgs1 consist of 853 amino acids (Jia et al., 2012; Zhu et al., 2001). Although the N-terminal domain of Tgs1 varies among different species, the mouse and human full-length Tgs1 are conserved.

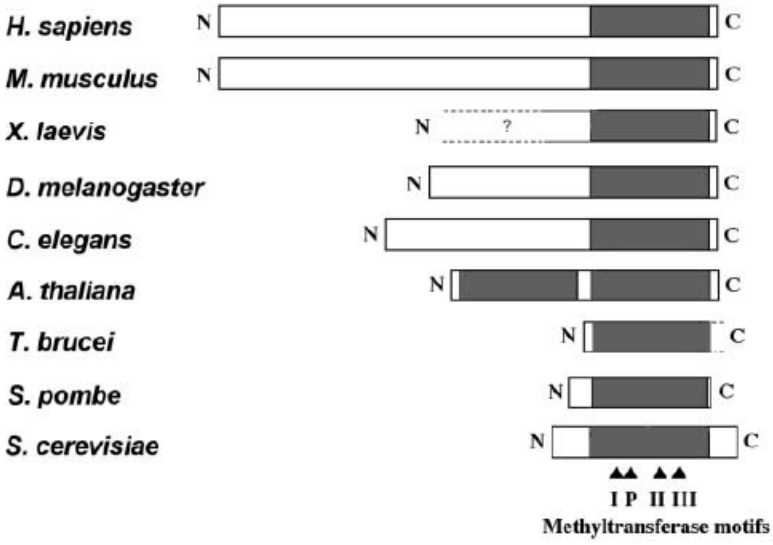


Figure 1.5.1 Evolutionary conservation of Tgs1. Adapted from (Mouaikel et al., 2002)

Two isoforms of Tgs1 exist in human cells: the full-length protein and a short isoform with only the C-terminal domain, due to limited proteolytic processing (Girard et al., 2008). The different isoforms of hTgs1 appear to be present in different cellular compartments. The full-length hTgs1 is localized in the cytoplasm, while the short isoform is localized in the nucleus.

1.5.1 C-terminal domain of Tgs1

Tgs1 contains putative methyltransferase Motifs I, II, III and Motif Post I on its C-terminal domain (Figure 1.5.1) (Mouaikel et al., 2002; Petrossian & Clarke, 2009; Zhu et al., 2001). They are conserved from yeast to human. The C-terminal domain of human Tgs1 (amino acids 576-853) is able to complement the synthetic lethality in *tgs1Δ mud2Δ* budding yeast strain (Hausmann et al., 2008). Crystal structure of hTgs1 C-terminal domain (576-853) reveals the canonical class I methyltransferase fold and an unique α-helical domain of 40 amino acids that is essential for m7G cap binding and catalysis (Monecke, Dickmanns, & Ficner, 2009). Thr-673 and Asn-704 establish water-mediated hydrogen bonds to S-adenosyl-L-homocysteine (AdoHcy) in the structure. And mutating them to alanines reduces the catalytic activity to 5%. Asp-719 makes contact to both ribose hydroxyls of AdoHcy, whose mutation to alanine reduces the catalytic activity to < 1%.

1.5.2 N-terminal domain of Tgs1

The N-terminal domain of Tgs1 varies among different species. In yeast, the N-terminal domain is dispensable for substrate binding but affects Tgs1 localization (Mouaikel, Bujnicki, Tazi, & Bordonne, 2003). The 1-40 amino acids are also required for Tgs1 self-association. (Boon, Pearson, & Kos, 2015). In human, the N-terminal domain of Tgs1 contains a consensus sequence GXXGXXI, which is present in K-homology domains of RNA-binding proteins (Lewis et al., 2000). The N-terminal domain alone (amino acids 1-384) can also enhance peroxisome proliferator-activated receptor (PPARɣ) activity (Zhu et al., 2001).

1.5.3 Phenotypes related to Tgs1 deficiency

A variety of phenotypes have been reported in Tgs1 deficient cell lines, plants and animal models. In budding yeast *Saccharomyces cerevisiae*, *tgs1Δ* strain exhibits cold-sensitive defect (Hausmann et al., 2008; Mouaikel et al., 2002). Failure of sporulation/meiosis defect is also observed (Qiu, Shuman, & Schwer, 2011). In *Arabidopsis thaliana*, Tgs1 deletion leads to growth defect under chilling conditions (J. Gao, Wallis, Jewell, & Browse, 2017). Tgs1 null mice display early embryonic lethality (Jia et al., 2012). In the absence of Tgs1, lethality in the early pupa stage is also observed in *Drosophila melanogaster* (Komonyi et al., 2005).

Tgs1 is involved in gene expression, RNA processing and transportation, either directly or indirectly through cap hypermethylation of its substrates. In human cells, Tgs1 promotes the expression of Rev-dependent HIV-1 RNAs (Yedavalli & Jeang, 2010). Specifically, Rev recruits Tgs1 to HIV-1 pre-mRNAs that contain Rev response element and Tgs1 hypermethylates them. The subsequent TMG cap affects HIV-1 protein expression by affecting the localization of the RNA. CRM1/Exportin-1 transports those TMG capped pre-mRNAs to cytoplasm, facilitating

their expression. It is worth noticing that the TMG cap seems to play contradicting roles in RNA localization. Previously it has been proposed to serve as a nuclear retention signal for U3 snoRNA by preventing CRM1/Exportin-1 mediated nuclear export (Boulon et al., 2004). What differentiates the localization of TMG-capped snoRNA and TMG-capped HIV pre-mRNA is yet to be determined.

In budding yeast, Tgs1 is required for proper U1 snRNP assembly at low temperature (18 °C) (Qiu, Schwer, & Shuman, 2015). Loss of Tgs1 leads to loss of TMG cap of U1, resulting in MMG capped U1. The MMG capped U1 is ectopically bound by cap-binding complex (CBC) at 18 °C. Mutations in Snp1 C-terminus or the cap binding site of CBC rescue the growth defect at 18 °C (Qiu et al., 2015). Tgs1 also affects splicing of certain pre-mRNAs such as PCH2 and SAE3. Both are meiotic mRNAs. In the absence of Tgs1, intron retention is observed in PCH2 and SAE3 mRNAs (Qiu et al., 2011). The SAE3 intron *per se* is a portable determinant of Tgs1-dependent splicing when transferred to other genes, indicating it is the absence of TMG cap on snRNAs, not the potential cap structure change on the SAE3 pre-mRNA itself, that leads to the splicing defect. Mutations in SEA3 intron that restore a consensus branchpoint and disrupt the stem-loop encompassing the branchpoint alleviate the dependency on Tgs1. Since Tgs1 catalytical activity is required for rescuing this dependency, the subsequent TMG cap on snRNAs is likely to be involved. Whether the TMG cap on snRNAs directly or indirectly facilitate early steps in splicing is unknown.

Absence of Tgs1 also causes disruption of cellular compartments where snRNA maturation occurs. Knockdown of Tgs1 in HeLa cells leads to loss of Cajal bodies (Lemm et al., 2006). There is some evidence that the expression level of Tgs1 is tightly controlled. Either

knockdown or overexpression of Tgs1 in *Drosophila melanogaster* is detrimental to the motor system (Borg, Fenech Salerno, Vassallo, Bordonne, & Cauchi, 2016).

1.6 Thesis overview

The primary goal of my thesis project was to examine the role of Tgs1 in mammalian RNA biogenesis, specifically telomerase RNA. As the project progressed, I realized that mouse embryonic fibroblasts derived from putative Tgs1-KO mice obtained from a collaborator contained a mutant version of Tgs1 with the intact methyltransferase motif. The snRNAs in the mutant cell line were still TMG-capped, indicating the presence of methyltransferase activity *in vivo*. Moreover, during my attempts to knock out Tgs1 in multiple human cell lines, evidence emerged that Tgs1 was essential for human cell viability. I managed to establish a human Tgs1 conditional knockdown cell line, where the cells showed severe growth defects after shutting down Tgs1. Because human telomerase RNA is a stable RNA with a long half-life (up to 3-4 weeks) (Yi, Tesmer, Savre-Train, Shay, & Wright, 1999), it is not feasible to examine the impact of Tgs1 deletion on hTR. Since then, the focus of my project was turned to two new directions. First, an unexpected U2 processing defect was observed in the mutant mouse cell line. I characterized this processing defect, gene expression profile as well as the cap hypermethylation in the mutant mouse cell line. Moreover, due to the homology of mouse and human Tgs1, I managed to create a human version of the mutant mouse Tgs1. The same U2 processing defect was recapitulated in the human mutant cell line. Second, although a severe growth defect was observed in the human Tgs1 conditional knockdown cell line, there was a brief window of time when the Tgs1 was absent and the cells were still viable. Gene expression profiling and cap

hypermethylation characterization was performed during this brief window of time. Also, potential mechanisms underlying the Tgs1-dependent growth defect were investigated. The results of this research are reported in Chapter Three of this thesis.

A comprehensive description of the materials and methods used in this thesis project is presented in Chapter Two. Chapter Four consists of a summary of information derived from the research reported in Chapter Three, as well as the future directions to further utilize the data obtained in Chapter Three.

Chapter 2 Materials and Methods

2.1 Cell culture and transfection

MEFs were kept in DMEM (Gibco, 11995) + 10% FBS, at 37 °C, 5% CO_2, 3% O_2.

HT1080, 293TRex, HT1080 Tet-off cells were kept in DMEM (Gibco, 11995) + 10% FBS, at 37 °C, 5% CO_2, atmosphere O_2.

MEFs were treated with 0.6 μM 4-OH-tamoxifen for 12h for Cre induction. Transfection of ZFN/CRISPR was performed with lipofectamine 2000 according to manufacturer's instructions.

2.2 Preparation of genomic DNA and PCR amplification

Qiagen DNeasy kit (Cat No./ID: 69504) was used to extract gDNA from MEFs and human cell lines according to manufacturer's instructions. For each PCR reaction, 100ng gDNA was used as template.

For MEF genotyping, primer P1 (CTC CTT CCT TCT GTA CCT CTG TAG C), P2 (CTG CAT GTA TGA ATC TTG GGA G) and P3 (GCA TCA AGA ATA TAC AGA ACA GAG ACT C) were used. PCR reaction with total volume 30ul contained 1ul dNTPs (10mM), 1ul each primer (10 μM), 3ul 10x thermo pol buffer (NEB) and 1ul Taq polymerase (NEB). The reaction was carried out under following condition: 3 min at 95℃; 30 cycles of 30 sec at 95℃, 30 sec at 48℃, 30 sec at 68℃; then 5 min at 68℃.

2.3 RT-PCR

To detect murine Tgs1 mRNA, 1 μg total RNA was used for each RT reaction, which contained 1 μl dNTPs (10 mM) and 1 μl reverse primer P7 (2 μM) with total volume of 11 μl. The mixture was heated at 65 ℃ for 5min, then cooled down to room temperature. 4 μl 5x first strand buffer (Invitrogen), 1 μl RNasin and 1ul DTT (0.1M) was added together with 1 μl SuperScript III reverse transcriptase (Invitrogen) as RT (+) or H2O as RT (-), incubated at 50 ℃ for 15 min, then at 55 ℃ for 45 min. 1 μl RNase H (Invitrogen) was added and incubated at 50 ℃ for 15 min. The following PCR reaction contained 2.5 μl cDNA, 1 μl dNTPs (10 mM), 0.4 μl of each Tgs1 specific primers (25 μM), 10 μl 5 x Phusion HF buffer (NEB), and 1 μl Phusion polymerase (NEB) with total volume of 50 μl. The reaction was carried out under following condition: 30 sec at 98℃; 30 cycles of 10 sec at 98℃, 15 sec at 55℃, 60 sec at 72℃; then 5 min at 72℃.

Tgs1 specific primers: P4 (gcg ACCGGT ATG TAC CCA TAC GAT GTT CCA GAT TAC GCT TCC GGA CTC ATG TGT TGC GAG AAG TGG AAC), P5 (GCC AGC TCT GGA CCA GAA GCT ATT TC), P6 (GAA ATA GCT TCT GGT CCA GAG CTG GC), P7 (CGC GGA TCC GCC TTC CTC CTG TCT CTC AC)

2.4 Immunoprecipitation

From total RNA: For each IP transfer, 50 μl Dynabeads (M-280 sheep anti-Mouse IgG, Life Technologies) were used. The beads were washed with 1 ml PBS twice. 0.5ml PBS(phosphate buffered saline, potassium dihydrogen phosphate 0.144g/L, sodium chloride 9g/L, disodium

phosphate 0.795g/L, pH=7.4) and 30 µl (3µg) of anti-TMG cap antibody (K121, Calbiochem) were added to the beads and incubated at room temperature for 30 min, then washed with 1 ml TMG (300) (10 mM Tris-HCl buffer, pH 8.0, 1 mM magnesium chloride, 10% glycerol, 300 mM sodium acetate), protease inhibitors (Promega) and 0.1mM DTT twice. 125 µg RNA was added together with 500 µl TMG (300), protease inhibitors, 0.1mM DTT, RNasin (1:1000) and 0.1% Tween 20 and incubated for 6 hours at 4℃. The beads were washed with 1 ml TMG (300), 0.1 mM DDT and 0.1% Tween 20 for 3 times, then washed with 1 ml TMG (50) (same at TMG300 but with 50 mM sodium acetate instead) and 0.1 mM DTT once. 100 µl TMG (50) and RNasin (1:500) was used to resuspend the beads after washing. To isolate RNA from the beads, 25 µl stop buffer (100mM Tris-HCl, PH 7.5, 200mM EDTA, 10mg/ml proteinase K, 2.5%SDS) was added to each resuspension and incubated for 15 min at 50 ℃. The supernatant was reserved, beads was washed by 200 µl 1 x TE (10mM Tris pH 8.0, 0.1mM EDTA pH 8.0) and the new supernatant was combined with previous one. 300 µl phenol/chloroform (equilibrated with sodium acetate, pH 5.2) was used to extract proteins from the supernatant. The remaining aqueous phase was back extracted with 50 µl 1 x TNE (10mM Tris pH7.0, 200mM NaCl, 1mM EDTA) and combined with previous aqueous phase. Another extraction was performed with 300 µl chloroform (equilibrated with sodium acetate, pH 5.2). The aqueous phase was precipitated with 1/10th volume NaOAc (3M, pH 5.2), 1 µl RNasin, 5 µg glycogen and 2.5 volume ethanol at -20℃ o/n. RNA pellets were collected via centrifugation (16000 x g, 15min, 4 ℃), washed with 70% ethanol for 3 times, air dried on ice and resuspended in 10 µl H2O plus 1 µl RNasin.

The control sample was proceeded following the same protocol, with anti-c-Myc antibody of same IgG subclass (SANTA CRUZ/Calbiochem/Sigma-Aldrich).

2.5 Northern blot

Total RNA and immunoprecipitation samples were separated on 4% polyacrylamide gel in 0.5 x TBE (44.5 mM Tris borate pH 8.3, 1 mM EDTA)/6.9M urea. The gel was made by mixing 30ml 6% acrylamide TBE/6.9M urea, 30µl TEMED (Tetramethylethylenediamine, Sigma-Aldrich) and 97.5µl 10% APS (Ammonium persulfate, Calbiochem). Following 1h polymerization, electrophoresis was performed in 0.5 x TBE at 15W for 1.5-2 hours. The RNA was transferred onto Biodyne membrane (Pall Corporation) and crosslinked with 12000µJ UV. The membrane was pre-hybridized in Church buffer (0.5 M Na-PO4, 1% bovine serum albumin, 1 mM EDTA, 7% SDS) for 30 min at 42 ºC. 10 million cpm probe was added and hybridized at 42 ºC o/n. One quick wash was applied with 100ml 0.1 x SSC (15 mM sodium chloride, 1.5 mM tri-sodium citrate, pH 7.0)/0.1% SDS, followed by 2 x 30 min wash with 100ml 0.1 x SSC/0.1%SDS. The membrane was wrapped and exposed to phosphor screen. 6% polyacrylamide gel was used for separation of U2 precursors. The rest of protocol remained the same. Radioactive probes were prepared by phosphorylation of target specific oligos with P32 labelled ATP.

Oligos for murine RNA included U1 (CGC AGG GGT CAG CAC ATC CGG AGT GC), U2 (CCAGGTCGATGCGTGGAGTGGACGGAGCAAGC), U3 (CTC AGA CTG TGT CCT CTC CCT CTC AAC CCT CAA G), U4 (CGC GCC TCG GAT AAA CCT CAT TGG CTA CGA TAC), U5 (GGT TTA AGA CTC AGA ATT GTT CCT CTC CAC GGA ATC TTT G), U6 (TCA TCC TTG CGC AGG GGC CAT GC) and H2B (CGGGGCAGGAGCGGACTTCGCAGGC). Human U1, U2 and U3 were detected with the same oligos due to conservation of human and murine snRNA sequences.

2.6 3’ end cloning

For each tailing reaction, 2.5 µg of total RNA was used. The RNA was brought up to total volume of 14 µl with H2O. Heat to 65°C for 3 min in PCR machine, snap cool on ice. Add 1µl RNasin, 4 µl 5x Poly(A) polymerase reaction buffer and 1 µl Poly(A) polymerase. Incubate at 30°C for 30 min. Add 2.5 µl dNTP mix (10 mM each), 5 µl oligo dT (gcgGAATTCTTTTTTTTTTTTTTTT , 25 µM), 8 µl ddH_2O, heat to 65°C for 3 min, cool down to r/t. Add to each tube 10 µl 5 x first strand buffer, 2.5 µl DTT (0.1 M), 1.0 µl RNase inhibitor, 1.0 µl superscript III RT RNase H minus (200 units/µl). Incubate at 50°C for 60 min. Add 1 µl RNase H and incubate at 37°C for 20 min. Use 2.5 µl of cDNA generated above for PCR amplification.

To amplify U2, forward primer against U2 (ATC GCT TCT CGG CCT TTT G) and reverse primer against poly(A) tail (oligo dT) were used. PCR reaction with total volume 30ul contained 1ul dNTPs (10mM), 1ul each primer (10 µM), 3ul 10x thermo pol buffer (NEB) and 1ul Taq polymerase (NEB). The reaction was carried out under following condition: 3 min at 94℃; 20 cycles of 30 sec at 94℃, 30 sec at 50℃, 30 sec at 72℃; then 7 min at 72℃.

2.7 TGIRT-U2 library

The reverse transcription of murine and human U2 was carried out by TGIRT-III Enzyme (InGex). The template switching reaction was carried out according to instructions from InGex. R1/R2 RNA (rArGrA rUrCrG rGrArA rGrArG rCrArC rArCrG rUrCrU rArGrU rUrCrU

rArCrA rGrUrC rCrGrA rCrGrA rUrC/3AmMC6T/) and R1/R2 compDNA (GATCGTCGGACTGTAGAACTAGACGTGTGCTCTTCCGATCTN) were annealed and served as the reverse primer. 2.5 µl cDNA from the template switching reaction was used for PCR amplification of U2. Adaptors and indexes for Illumina sequencing library were incorporated via forward primer with U2 specific sequence plus adaptor (AATGATACGGCGACCACCGA GATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT ATC GCT TCT CGG CCT TTT G) and reverse primer with barcodes. The PCR reaction contained 2.5 µl cDNA, 1 µl dNTPs (10 mM), 1 µl of each primer (10 µM), 4 µl DMSO (NEB), 10 µl 5 x Phusion HF buffer (NEB), and 1 µl Phusion polymerase (NEB) with total volume of 50 µl. The reaction was carried out under following condition: 30 sec at 98℃; 35 cycles of 10 sec at 98℃, 25 sec at 64.5℃, 15 sec at 72℃; then 7 min at 72℃.

The reverse primers with barcodes (lowercase) for Illumina sequencing include:

CAA GCA GAA GAC GGC ATA CGA GAT cgtgat GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

CAA GCA GAA GAC GGC ATA CGA GAT acatcg GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

CAA GCA GAA GAC GGC ATA CGA GAT gcctaa GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

CAA GCA GAA GAC GGC ATA CGA GAT tggtca GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

CAA GCA GAA GAC GGC ATA CGA GAT cactgt GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

CAA GCA GAA GAC GGC ATA CGA GAT attggc GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

CAA GCA GAA GAC GGC ATA CGA GAT gatctg GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

CAA GCA GAA GAC GGC ATA CGA GAT tcaagt GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

CAA GCA GAA GAC GGC ATA CGA GAT ctgatc GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

CAA GCA GAA GAC GGC ATA CGA GAT aagcta GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

CAA GCA GAA GAC GGC ATA CGA GAT gtagcc GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

CAA GCA GAA GAC GGC ATA CGA GAT tacaag GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T

2.8 Genome editing

2.8.1 Zinc finger nuclease

Endogenous Tgs1 alleles were knocked out by CompoZr Zinc Finger Nuclease (Sigma-Aldrich-Aldrich). The ZFN binding sequence (Uppercase) and ZFN cut site (Lowercase) are listed as below: GACCACATGCCAGTGgtactGATGGAGATGAAAGTGAG. Transfection and Surveyor assay were performed as manufacturer's instructions. Primer ZFN-F (GATGTGAGCTGGCACTTTGA) and ZFN-R (CCACTGGCTTATGCATACAAAT) were used for identifying the genome types of individual clones. The PCR reaction contained 100ng gDNA, 1 µl dNTPs (10 mM), 2.5 µl of each primer (10 µM), 10 µl 5 x Phusion HF buffer (NEB), and 1 µl Phusion polymerase (NEB) with total volume of 50 µl. The reaction was carried out under following condition: 30 sec at 98℃; 35 cycles of 10 sec at 98℃, 15 sec at 54℃, 7 sec at 72℃; then 5 min at 72℃. PCR products were purified using QIAquick PCR Purification Kit (Qiagen) according to manufacturer's instructions.

2.8.2 CRISPR

CRISPR plasmid px330 (#42230 from Feng Zhang) was obtained from Addgene. Guide RNA sequence (TACTGATGGAGATGAAAGTG) was designed using online tool

(http://crispr.mit.edu/). The cloning of gRNA sequencing into px330 was done according to instructions from the Feng Zhang lab (http://www.genome-engineering.org/crispr/).

2.9 Western blot

Human Tgs1 proteins were separated on Bis-tris 4-12% NuPage gel (Life technologies). The cell lysate was prepared with 100 µl NuPage LDS sample buffer (Life technologies) per 4 x 10^4 cells, followed by heating at 95 ℃ for 10 min, vortexing at 4 ℃ for 10 min, heating at 75 ℃ for 10 min, then snapped on ice. 10-15 µl lysate was loaded per sample. The electrophoresis was carried out in 1 x MOPS buffer (Thermo Fisher) at 200V until dye front about to enter foot of gel. Proteins were transferred to nitrocellulose membrane (Protran BA85) under 400 mA for 1 hour. The membrane was blocked in 2% milk in 1 x TTBS (20 mM Tris base, pH 7.5, 137 mM sodium chloride, 0.1% tween 20) at room temperature for 1 hour, then incubated with anti-Tgs1 antibodies (Sigma-Aldrich, 1:2000 dilution) in 1 x TTBS and 2% milk at 4 ℃ overnight. To remove the primary antibodies, 1 x brief wash and 3 x 5 min wash with TTBS was applied. The membrane was incubated with secondary antibodies (goat-anti-rabbit HRP, 1: 2000 dilution) for 1 hour at room temperature, then washed with 2 x 5min TTBS and 2 x 5min TBS. Proteins were detected with ECL plus substrate (Thermo Fisher) and scanned.

HA-tagged Tgs1 proteins were analyzed following the same protocol. Anti-HA antibodies (clone 12CA5, Roche) were used. Alpha-tubulin antibodies (clone B-5-1-2, Sigma-Aldrich) served as loading control.

Histone protein H2AX (ser139) and phosphorylated H2AX were analyzed following the same protocol, except that the MOPS buffer was replaced with MES buffer (Thermo Fisher). Anti-H2AX antibodies (EMD Millipore) and anti-phosphos-H2AX antibodies (clone JBW301, Upstate/Chemicon) were used.

2.10 Immunostaining

Coverslips (VWR) were coated with gelatin (Sigma-Aldrich) at room temperature for 1 hour, rinsed with H2O for 3 times, dried and sterilized under UV for 4 hours, then placed in a 6-well plate. 293TRex cells were digested from a confluent T75 flask with 0.25% Trypsin-EDTA solution (Sigma-Aldrich), diluted 1:4 and seeded into the 6-well plate containing coverslips. After 18-24 hours, cells were fixed with methanol (chilled at -20 ℃) for 5 min at room temperature followed by 3 x wash with ice cold PBS (Potassium dihydrogen phosphate 0.144 g/L, sodium chloride 9 g/L, disodium phosphate 0.795 g/L, pH=7.4). The coverslip was incubated with 1% BSA in PBST (PBS + 0.1% Tween 20) at room temperature for 1 hour. Anti-HA antibodies (ab9110, Abcam, 1:1000 dilution) were added and incubated at 4 ℃ overnight. 3 x wash with PBS was applied to remove the primary antibodies, followed by 1-hour incubation with secondary antibodies (Goat anti-rabbit IgG, Alexa Fluor 546, Sigma-Aldrich, 1:2000 dilution). Counterstaining was performed with 0.1 μg/ml DAPI for 10 min. The coverslip was rinsed with PBS and mounted with Fluoromount G (Southern Biotech) for examination under microscope.

2.11 Bioinformatics

TopHat and STAR were used for alignment of the reads obtained by next-generation sequencing. Gene expression were analyzed by EdgeR, and further visualized via Shiny app. Alternative splicing was analyzed by rMATS. The next-generation sequencing data were primarily analyzed by Morgan Schroeder and Rutendo Sigauke. The TMG-IP library data in 3.5.4 were analyzed by Madelaine Gal and Hua Li from the Computational Biology Core at the Stowers Institute.

Chapter 3 Results

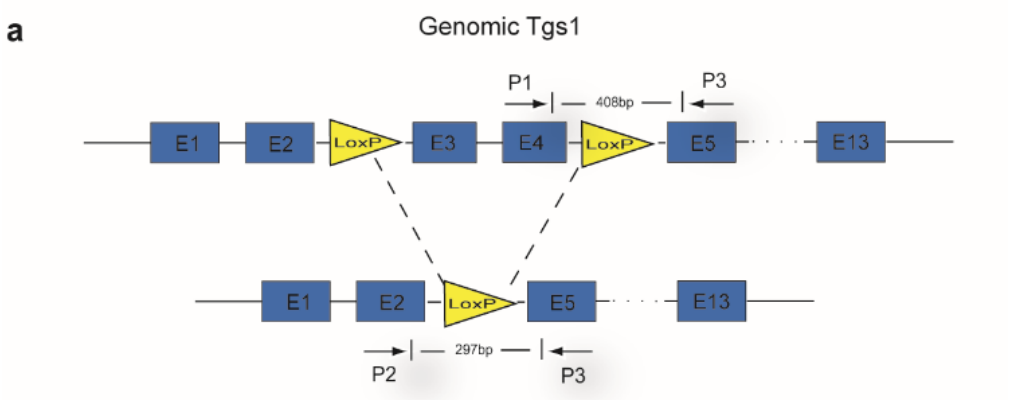
3.1 Generation of a putative conditional knockout cell line for murine Tgs1

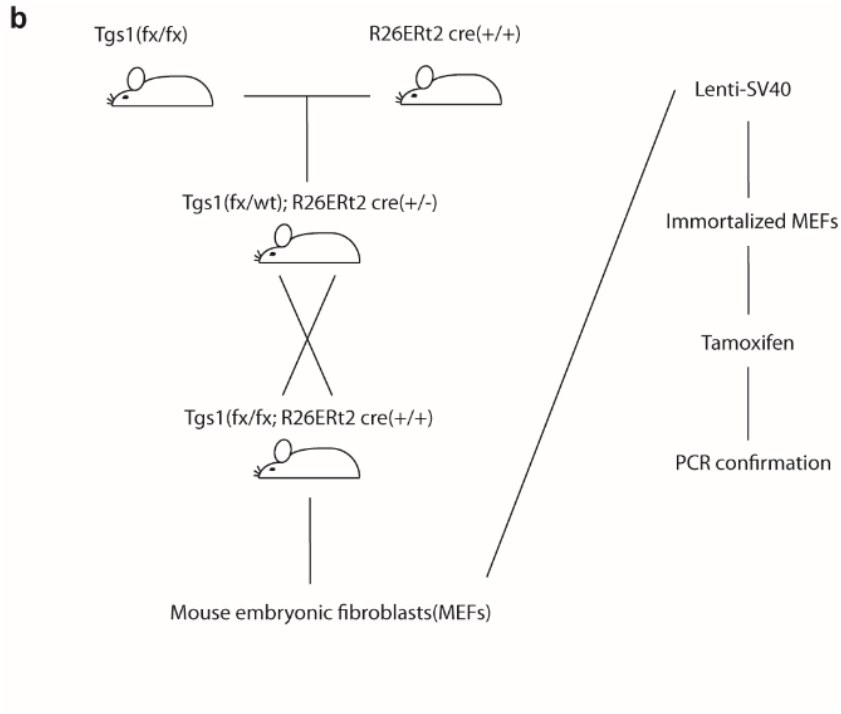
Murine Tgs1 was initially identified as PIMT (PRIP-interacting protein with methyltransferase domain) by the Reddy lab (Zhu et al., 2001). The same lab constructed a conditional knockout allele with the two-loxP, two-frt recombination system, where murine Tgs1 exon 3 and exon 4 were flanked by LoxP sites (Jia et al., 2012) (Figure 3.1.1a). We obtained the mice from the Reddy lab. The Lab Animal Services facility at Stowers Institute bred them with mice carrying tamoxifen-inducible Cre (Jackson Laboratory, stock number 008463). A cassette carrying Cre recombinase fused with estrogen receptor T2 (Cre-ERT2) is introduced into intron 1 of the gene trap ROSA26 locus. The expression is driven by ROSA26 promoter. The estrogen receptor T2 retains the Cre recombinase in the cytoplasm until tamoxifen releases the inhibition, permitting the recombination of LoxP sites. Mouse embryonic fibroblasts (MEFs) were derived from the homozygous $TGS1^{flox/flox}/CRE^{+/+}$ offspring (Figure 3.1.1b). Primary $TGS1^{flox/flox}/CRE^{+/+}$ MEFs underwent senescence after ~3 passages. It has been reported that MEFs are susceptible to oxidative stress under normal tissue culture (20%) oxygen level (Parrinello et al., 2003), which could lead to early senescence. However, after I transferred primary MEFs to physiological (3%, skin tissues) oxygen level, their lifespans were only slightly extended (senescence after ~4 passages).

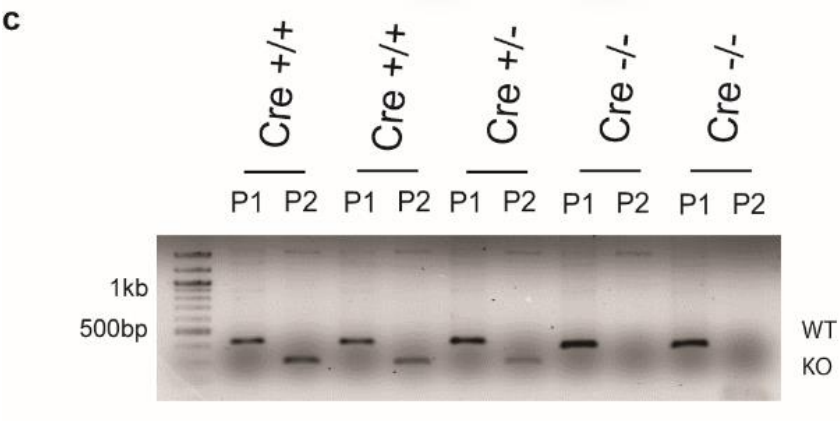
To address this problem, I immortalized the primary MEFs via infection with lentiviruses carrying SV40 large T antigen (Lenti-SV40T, Cat. G256, abm). The immortalized MEFs were treated with tamoxifen to induce the expression of Cre recombinase. Deletion of Tgs1 exon 3 and

exon 4 was examined by PCR of the genomic DNA. 72h after induction, the MEF population showed a mixture of Tgs1-KO and Tgs1-WT signals (Figure 3.1.1c). Multiple consecutive tamoxifen treatments did not change the composition of the heterogenous MEF population. Sequencing of the LoxP sites revealed that a T to C mutation was present at the 3' end of the palindrome, thereby compromising the palindrome sequence (Figure 3.1.1d). An isogenic Tgs1-KO MEF population could not be obtained, either due to the low efficiency of the compromised Cre-Lox P system or due to potential growth disadvantage of Tgs1-KO MEFs.

Subsequently, I sought to separate homozygous Tgs1-KO MEFs ($Tgs1^{-/-}$) from WT MEFs ($Tgs1^{+/+}$) and heterozygous WT/KO MEFs ($Tgs1^{+/-}$) via single cell selection. After tamoxifen treatment, the heterogeneous MEFs were diluted and seeded into 96-well plates at the density of 2 cells/well. After 7~10 days, the plates were visually inspected under a microscope. Colonies of MEFs were observed. The wells that just contained one single colony were digested and expanded gradually to 24-well plates, 6-well plates and T-75 flasks. Genomic DNA was extracted, and the genotypes were determined by PCR. Three individual clones were identified with both Tgs1 alleles deleted from the genome (Figure 3.1.1e).


a
Genomic Tgs1
P1
P3
408bp
E1
E2
LoxP
E3
E4
LoxP
E5
E13
E1
E2
LoxP
E5
E13
297bp
P2
P3


b
Tgs1(fx/fx)
R26ERt2 cre(+/+)
Tgs1(fx/wt); R26ERt2 cre(+/-)
Tgs1(fx/fx; R26ERt2 cre(+/+)
Mouse embryonic fibroblasts(MEFs)
Lenti-SV40
Immortalized MEFs
Tamoxifen
PCR confirmation


c
Cre +/+
Cre +/+
Cre +/-
Cre -/-
Cre -/-
P1 P2 P1 P2 P1 P2 P1 P2 P1 P2
1kb
500bp
WT
KO

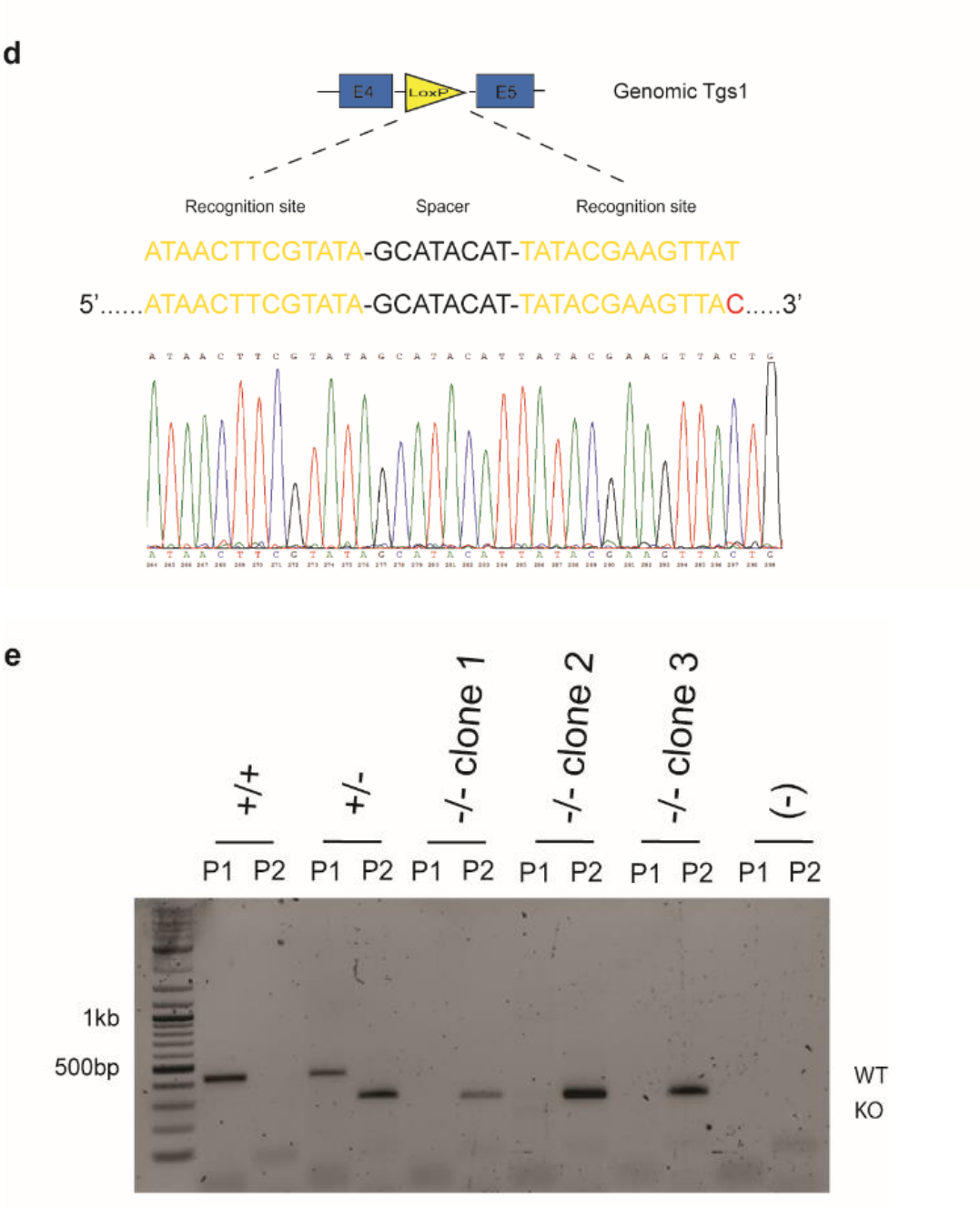


Figure 3.1.1 Generation of a putative conditional knockout cell line for murine Tgs1.

(a) Schematic of murine Tgs1 genomic locus with exons and loxP sites. **(b)** Schematic of breeding scheme to introduce inducible Cre to $Tgs1^{fx/fx}$ mice. **(c)** Partial loss of floxed exons after Cre-mediated recombination. Either $Tgs1^{fx/fx}/Cre^{+/+}$ or $Tgs1^{fx/fx}/Cre^{+/-}$ MEFs were treated with 0.6μM 4-OH-tamoxifen for 12 hours. Cells were harvested

after 72 hours and gDNA were extracted for PCR analysis. gDNA of $Tgs1^{fx/fx}/Cre^{-/-}$ MEFs served as negative control. **(d)** Sequencing of LoxP sites revealed the palindrome sequence was mutated at one position. The region carrying LoxP sites was amplified by PCR using gDNA from $Tgs1^{fx/fx}/Cre^{+/+}$ MEFs, subsequent purification and sequencing of the PCR products showed the 3' T of the palindrome sequence was mutated to C. **(e)** Successful removal of exons 3 and 4 after single cell selection. $Tgs1^{fx/fx}/Cre^{+/+}$ MEFs were treated with 0.6uM 4-OH-tamoxifen for 12 hours. Cells were counted by hemocytometer, diluted and seeded 2 cells/well in 96-well plates. The genotype of each individual clone was determined by PCR of gDNA.

It has been reported that a monoclonal antibody (K121) specifically recognizes the 2,2,7-trimethylguanosine cap (Mouaikel et al., 2002). This anti-TMG antibody precipitated snRNAs from WT yeast cells, but not in *tgs1Δ* cells (Mouaikel et al., 2002). A previous study in our lab also confirmed the specificity of this antibody against the TMG-capped telomerase RNA (Tang et al., 2012). The same antibody was used to examine the cap structure of snRNAs in Tgs1-KO MEFs. Immunoprecipitation was performed with total RNA extracted from both Tgs1-KO and WT MEFs (Figure 3.1.2 snRNAs are TMG-capped in the putative Tgs1-KO cell line.a). Surprisingly, snRNA U1, U2, U4 and U5 were precipitated equally well between Tgs1-KO and WT-derived RNA, indicating the snRNAs were still TMG-capped in Tgs1-KO MEFs (Figure 3.1.2b). These data potentially contradicted what has been observed in *tgs1Δ* yeast strain, assuming that Tgs1 is indeed inactive in the cells. A possible explanation is genetic redundancy. Unlike the case in yeast, Tgs1 may not be the only methyltransferase responsible for cap hypermethylation of murine snRNAs. However, no other protein was found with significant overall similarity with Tgs1 in mouse (Zhu et al., 2001).

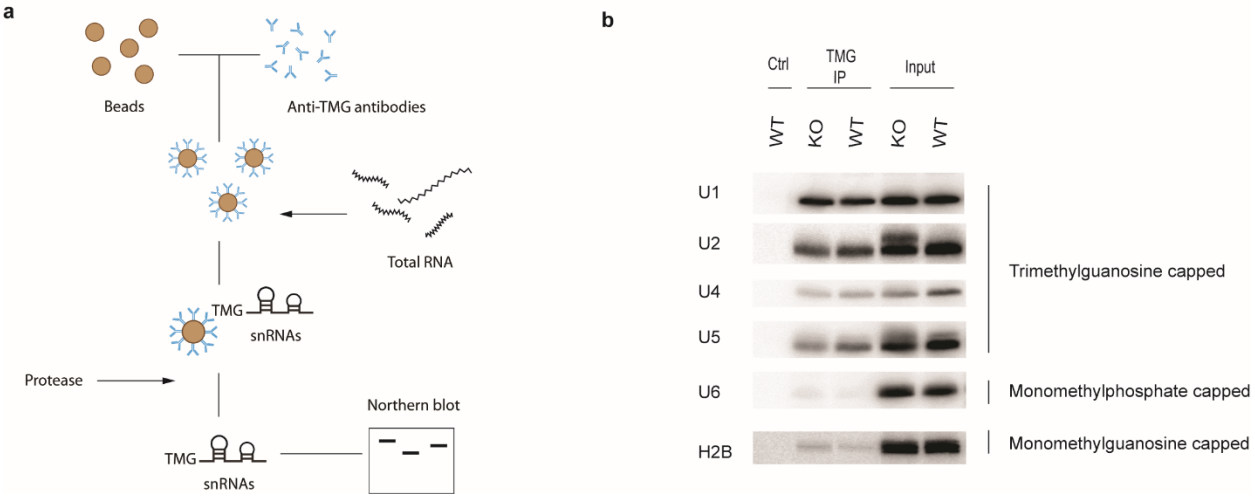


Figure 3.1.2 snRNAs are TMG-capped in the putative Tgs1-KO cell line.

(a) TMG-IP for snRNAs in Tgs1-KO MEFs. Immunoprecipitation by anti-TMG antibody (K121) was perform with total RNA from both WT and KO MEFs. The monoclonal mouse antibodies were bound to Dynabeads through pre-embedded sheep-anti-mouse IgG. Precipitated RNAs were separated from the bead via protease digestion. To rule out potential nonspecific interactions between beads/IgG and RNA, immunoprecipitation by anti-c-myc mouse antibody was performed as control (Ctrl). **(b)** snRNA U1, U2, U4 and U5 analyzed by Northern blot. 50% IP samples and 20% input samples were loaded. snRNA U6 (carrying a monomethylphosphate cap) and H2B mRNA (carrying a monomethylguanosine cap) were also analyzed as control.

As suggested above, an alternative explanation is that the putative Tgs1-KO cell line does not eliminate Tgs1 function. Deletion of exon 3 and exon 4 was confirmed at genomic loci. To determine whether stable transcripts from the remaining exons existed in cells, further examination of the putative Tgs1-KO MEFs was carried out. RT-PCR of Tgs1 mRNA was performed with primers flanking the remaining exon 10~13 (corresponding to Tgs1 C-terminus).

Tgs1-KO showed the same migrating pattern as WT, indicating the presence of stable transcripts from the remaining exons. Primers flanking the deleted exon 3~4 (corresponding to Tgs1 N-terminus) were also tested. In contrast to WT, a faster migrating band (Tgs1*) was observed in Tgs1-KO (Figure 3.1.3a). Subsequent sequencing revealed this band was a splicing product resulting from ligation of exon 2 and exon 5, after exon 3 and exon 4 were deleted. This band was not observed in WT MEFs. No annotation of such a variant was found in the mouse genomic plus transcript (Mouse G+T) database. It was a mutant version of Tgs1 specifically present in the putative Tgs1-KO MEFs. Importantly, the methyltransferase motif of this mutant was in frame (Figure 3.1.3b). Since snRNA U1, U2, U4 and U5 were still TMG-capped in Tgs1-KO MEFs, the mutant was likely translated into functional methyltransferase, which was capable of hypermethylating snRNAs.

Despite that no deficiency in TMG-capping of snRNAs was observed, it is worth noticing that mutant MEFs showed a growth defect at early stages of single cell selection. The colony in 96-well plate remained the same size for ~1 month. Then cell growth recovered, establishing a viable mutant Tgs1 cell line.

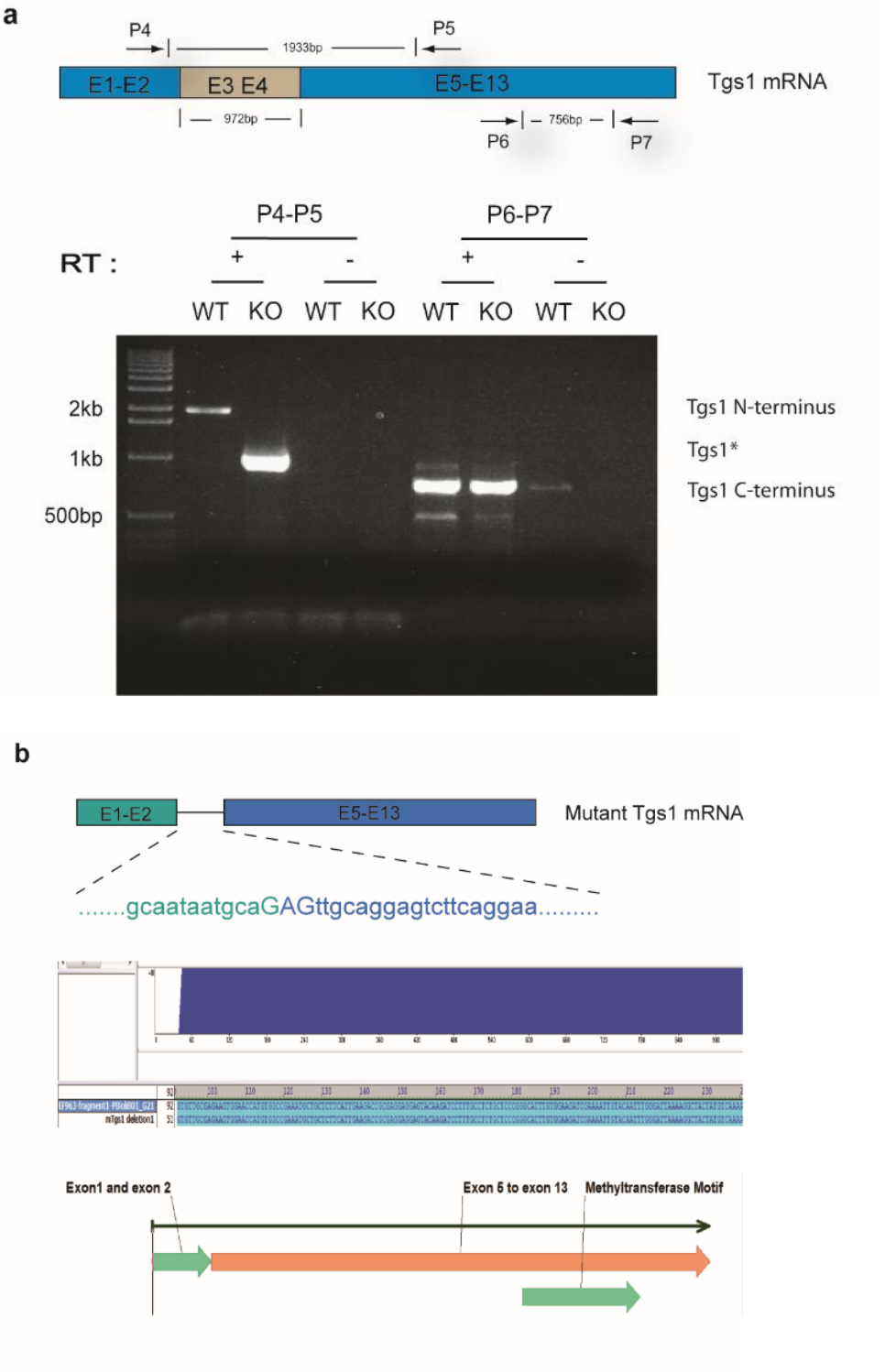


Figure 3.1.3 Identification of a mutant version of murine Tgs1 in the putative KO cell line.

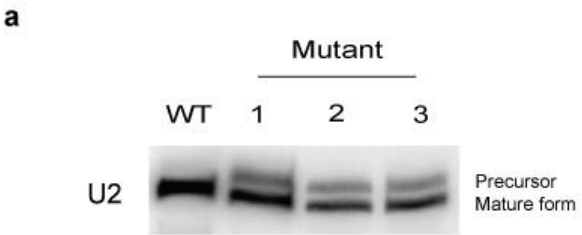
(a) RT-PCR showed a faster migrating band in putative Tgs1-KO MEFs. RNAs were extracted from both WT and KO MEFs via Trizol. RT-PCR was performed with primers flanking either the deleted exon 3~4 (Tgs1 N-terminus)

or the remaining exon 10~13 (Tgs1 C-terminus). **(b)** Sequencing result of the faster migrating band. The Tgs1* band was cut out, underwent gel purification and cloned into pCRII-TOPO vector for sequencing.

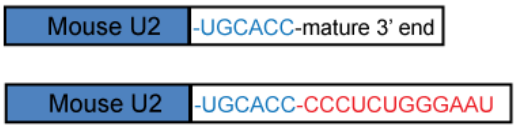
3.2 Characterization of the mutant allele of Tgs1 in MEFs

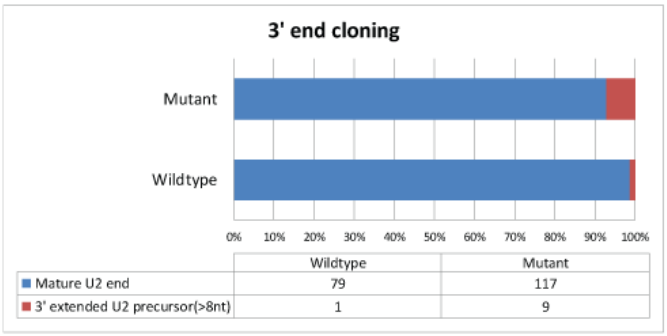
3.2.1 U2 processing defect in mutant MEFs

Although the mutant Tgs1 did not affect the TMG-cap structure of snRNAs, the migrating pattern of U2 snRNA was altered. An additional band was present above the mature form of U2. This putative U2 precursor band was present in all three mutant MEF cell lines (Figure 3.2.1a). The 3' ends of U2 were cloned via poly(A) tailing followed by RT-PCR using oligo d(T)18 and an upstream U2 specific primer. First, total RNA from mutant and WT MEF was treated with poly (A) polymerase, resulting in polyadenylation of U2 snRNA. Then oligo d(T)18 was used to synthesize cDNA from polyadenylated RNAs, including U2. Then U2 was amplified using an upstream specific primer and oligo d(T)18 by PCR. The PCR product was cloned into pCR2.1-TOPO vector and subjected to 96-well plate sequencing. Accumulation of U2 precursors with up to 10nt extension were observed in mutant MEFs (Figure 3.2.1b).



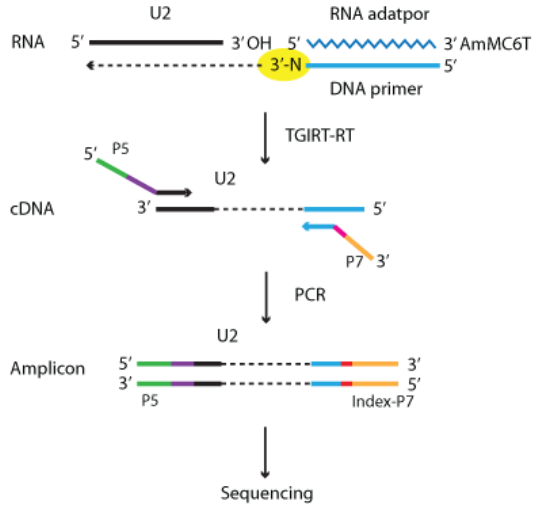
b

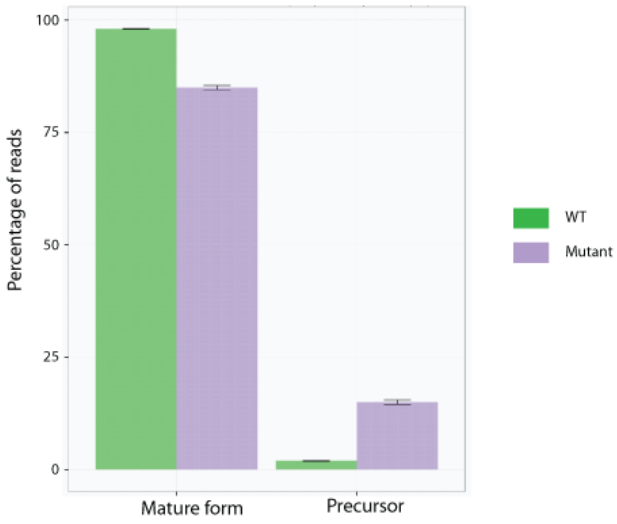


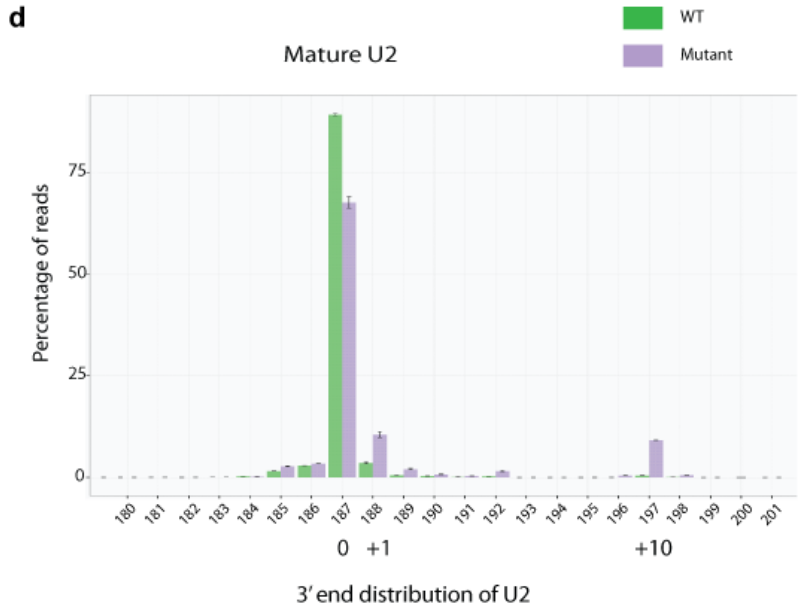


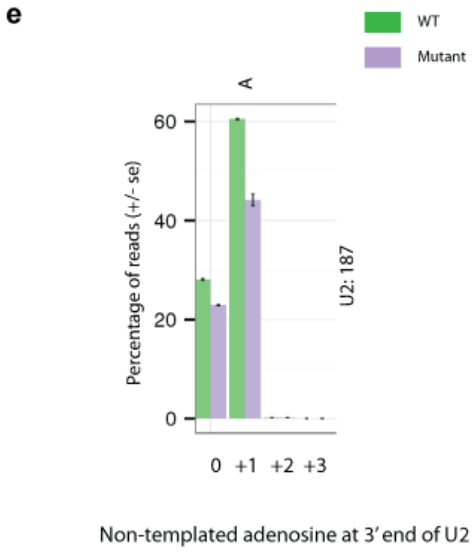
	Wildtype	Mutant
Mature U2 end	79	117
3' extended U2 precursor(>8nt)	1	9

c






d
Mature U2
WT
Mutant
Percentage of reads
75
50
25
0
180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201
0 +1
+10
3′ end distribution of U2


e
WT
Mutant
A
Percentage of reads (+/- se)
60
40
20
0
U2: 187
0 +1 +2 +3
Non-templated adenosine at 3′ end of U2

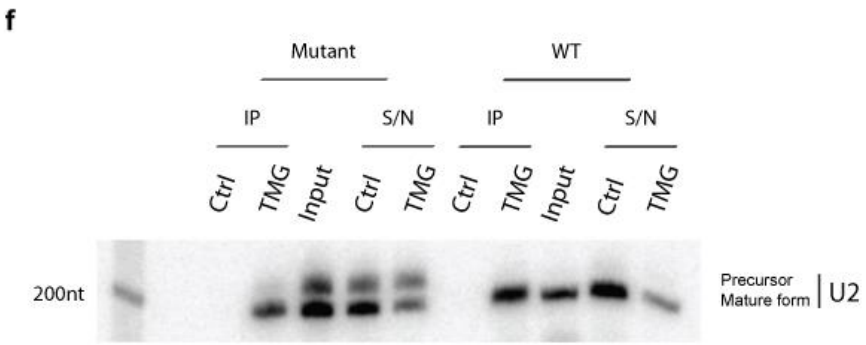


Figure 3.2.1 Processing defect of U2 snRNA in Tgs1 mutant MEFs.

(a) Northern blot for U2 snRNA. Total RNAs were extracted from both WT and three independent mutant MEF cell lines. Mature U2 and precursors were separated on 6% polyacrylamide gel. **(b)** 3' end cloning of U2 snRNA. Total RNA from WT and mutant MEFs were treated with poly(A) polymerase. cDNAs were synthesized with oligo d(T)18 primer via reverse transcription. The 3' end of U2 was amplified from cDNA by primer annealing to upstream sequence of U2 and oligo d(T)18 primer. The PCR products were cloned into pCR2.1-TOPO vector and subjected to 96-well plate sequencing. **(c)** Next-generation sequencing of U2 3' ends. Libraries of U2 3' ends were constructed using thermostable group II intron reverse transcriptase(TGIRT). The template switching reaction was initiated by a synthetic RNA/DNA duplex with a single-nucleotide 3' overhang. cDNA was synthesized and U2 was amplified by PCR with primers carrying adaptors and barcodes for next-generation sequencing. The RNA was blocked by 3' Amino Modifier C6 dT (3AmMC6T) to inhibit template switching from that end. For NGS data analysis, the reads that are mapped to mature U2 + 1nt (indistinguishable on northern blot) or shorter count as "mature", the reads that are mapped beyond full-length mature U2 +1nt count as "precursor". **(d)** 3' end distribution of U2. The mature form of U2 ends at 187nt (position 0). The reads that are mapped to each position of the genomic sequence of U2 are shown in percentages. Reads that are beyond position +10 are collapsed into position +10. **(e)** Non-templated nucleotides at the 3' end of U2. Bottom bar shows length of the non-templated nucleotides. Top bar shows the composition of the non-templated nucleotides (A: adenosine) **(f)** TMG-IP of U2 precursors. Protein A Sepharose beads were used to bind anti-TMG antibodies. Supernatant from the IP was preserved and precipitated

with EtOH. 33% of input, IP and S/N were loaded for northern blot. immunoprecipitation by anti-c-myc mouse antibody was performed as control (Ctrl).

The pre-snRNA 3' end is mapped to the genomic sequence of U2, but this templated tail is not part of the mature snRNA. However, studies show the tail functions to facilitate snRNP assembly (Yong et al., 2010). Other than the templated sequence, it has also been reported that a non-templated adenosine is present on the 3' end of U2, presumably added by the CCA-adding enzyme that also modifies the 3' end of tRNAs (Cho et al., 2002). The function of this non-templated adenosine remains unknown. It is worth examining the 3' end of U2 in greater detail to understand the implications of the processing defect in Tgs1 mutant MEFs.

There are several caveats regarding to poly(A) tailing. If adenosines are present at the 3' end of U2, they will not be distinguished from the poly(A) tails. There are two genome-coded adenosines (templated) in the U2 precursor. And a non-templated terminal adenosine has been reported as well. Those templated/non-templated terminal adenosines will not be recognized in data generated by poly(A) tailing. It has also been reported that poly(A) polymerase could be biased towards specific NTPs (Linsen et al., 2009; Raabe, Tang, Brosius, & Rozhdestvensky, 2014; Yehudai-Resheff & Schuster, 2000). Ligating an RNA adapter to the 3' end of U2 could detect the terminal adenosines, though the ligation step tended to be time-consuming, inefficient, and showed preference for RNA ends to be ligated (Lamm, Stadler, Zhang, Gent, & Fire, 2011; Levin et al., 2010).

To gain more in-depth and accurate information of terminal nucleotides of U2 in mutant MEFs, next-generation sequencing libraries for the 3' end of U2 were constructed using thermostable group II intron reverse transcriptase (TGIRT). TGIRT is capable of template

switching directly to new RNA templates that have little complementarity to the 3' end of the cDNA. It could synthesize cDNA without homopolymer tailing or RNA ligation (Mohr et al., 2013). The library construction was initiated by an RNA adapter/DNA primer duplex. The DNA primer had a single randomized nucleotide 3' overhang that can base pair with the 3' end of the target RNA (Figure 3.2.1c). The single nucleotide junction allowed TGIRT to generate a continuous cDNA that linked the adaptor sequence to 3' end of U2. The cDNA was then amplified by PCR using primers that contain barcodes and adaptors for next-generation sequencing. The 3' end distribution data obtained by Illumina sequencing showed U2 precursors accumulated in mutant MEFs (Figure 3.2.1c). In previous 96-well sequencing data, some U2 reads were one nucleotide longer than the mature form. Such difference in length would not be distinguished on northern blot. To better correspond the NGS data with the northern blot result, we used the position +1 as cutoff to separate "precursors" from mature U2 in our analysis. In WT MEFs, 1.89% reads counted as precursors. In mutant MEFs, the ratio was increased to 14.96%, consistent with the northern blot result.

Detailed 3' end distribution analysis showed that the predominant form of the precursors was 10nt extended pre-U2 (Figure 3.2.1d). 9.14% reads were mapped to position +10 in mutant MEFs, an increase from 0.48% in WT MEFs. Besides, the portion of U2 with one nucleotide longer than mature form were increased in mutant MEFs, which was not distinguished from mature U2 on northern blot. 10.59% reads were mapped to position +1 in mutant MEFs, in contrast to 3.61% in WT MEFs.

As mentioned earlier, there are studies demonstrating non-templated nucleotides on the 3' end of U2. We further examined the non-templated nucleotides. At position 0, analysis showed that one non-templated adenosine was present on 3' end of mature U2, consistent with the

previous report (Cho et al., 2002). In WT MEFs, there were 60.18% mature U2 with one 3' terminal adenosine. The ratio was reduced to 43.69% in mutant MEFs (Figure 3.2.1e). At position +1 and position +10, the non-templated nucleotides are extremely rare (<0.5%, not shown in figures).

The biogenesis of snRNAs includes transcription by a specialized RNA polymerase II, termination by the integrator complex, nuclear exportation, 3' end trimming, association with Sm proteins, 5' hypermethylation, and nuclear importation (Matera et al., 2007). It is known that the association of Sm proteins is required for recruitment of Tgs1 and 5' hypermethylation. But whether the 3' end trimming occurs before or after 5' hypermethylation is not clear.

To determine whether the 3' end trimming occurs before or after 5' hypermethylation, we asked whether the 5' cap of the longer forms of U2 are hypermethylated. Total RNA was subjected to immunoprecipitation with an antibody that specifically recognizes the trimethylated form of the 5' guanosine cap. After immunoprecipitation, RNA was isolated from the supernatant. 6% polyacrylamide gel was used to separate the ~10nt extended U2 precursor from the mature form. A probe against the shared region of mature U2 and the precursors showed that U2 precursors remained in the supernatant, while mature U2 was partially depleted by anti-TMG immunoprecipitation (Figure 3.2.1f). Unlike mature U2, the precursors did not contain trimethylguanosine caps.

3.2.2 Gene expression in mutant MEFs

Murine Tgs1 was first identified by a yeast-two-hybrid screen, using coactivator peroxisome proliferator-activated receptor-interacting protein (PRIP) as bait. In the screen, the N-terminal region of Tgs1 (1-384aa) was recovered from cDNA library. This region was found

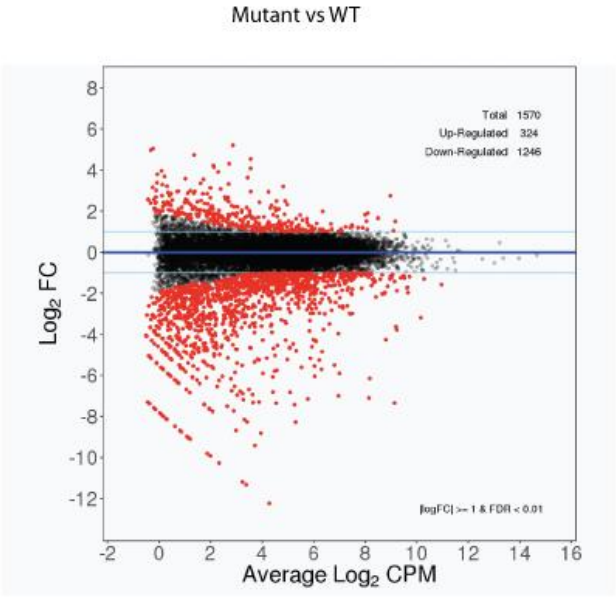
to enhance PPARɣ transcriptional activity (Zhu et al., 2001). The deleted exons 3 and 4 of Tgs1 code for amino acids 57-381. This region corresponds to transcriptional co-activator region. It is conceivable that mutant Tgs1 affects transcription of mRNAs, resulting in dysregulation of gene expression.

To test this hypothesis, we employed next-generation sequencing to access the genome-wide differential expression between mutant and WT MEFs. Total RNAs extracted from both mutant MEFs and WT MEFs were subjected to next-generation sequencing. Two biological replicates were sequenced for either mutant MEFs or WT MEFs. 30~55 million 50-base-pair (bp) single-read reads were generated for each sample and mapped by STAR. EdgeR was used to analyze differential gene expression. By the threshold of fold change (FC) of two and adjusted p-value less than 0.01, 1570 genes were categorized as dysregulated: 324 were upregulated and 1246 were downregulated (Figure 3.2.2a). Biotype analysis suggested that majority of the dysregulated genes were protein coding genes (Figure 3.2.2b, c). Among the upregulated genes, long intergenic non-coding RNA (6.19%) and anti-sense transcripts (3.1%) were the largest groups other than protein coding genes. Anti-sense transcripts (7.14%) and long intergenic non-coding RNA (3.85%) were also among the largest groups of down regulated genes.

Since U2 snRNA is one of the key components of spliceosome, the U2 processing defect could impact splicing in mutant MEFs. Since splicing mainly involved mRNAs, poly (A) enrichment was carried out in NGS library construction to enrich for mRNAs. Percent spliced out (PSO) was calculated to evaluate splicing efficiency. Surprisingly, no overall splicing defect was observed (Figure 3.2.2d). Next, different alternative splicing events were analyzed. It is conceivable that U2 processing defect may affect specific alternative splicing events without resulting in a significant difference in overall splicing. Five different alternative splicing events

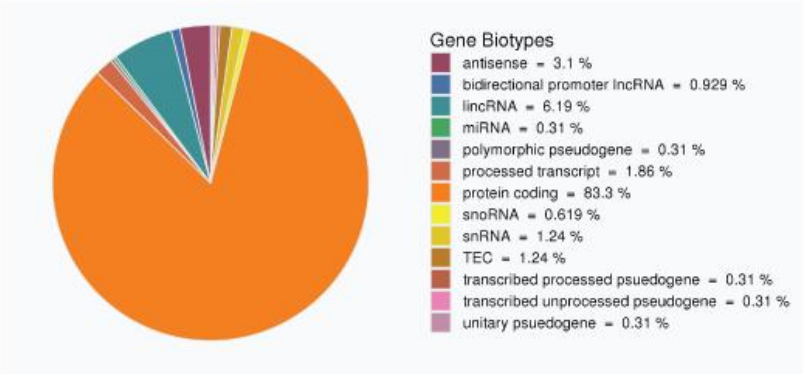
were examined with percent spliced in (PSI): skipped exon, alternative 5’ splice site, alternative 3’ end splice site, mutually exclusive exons and retained intron. No significant defect was observed in each specific alternative splicing event (Figure 3.2.2e).

a

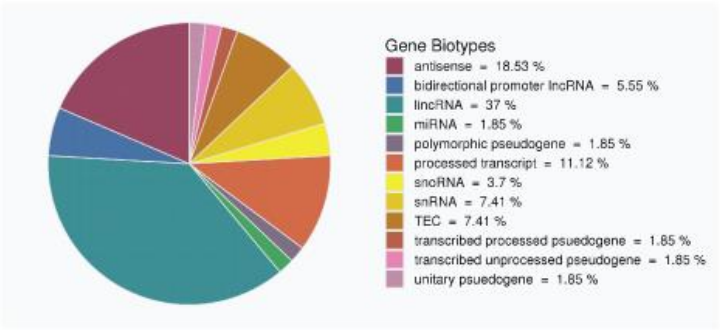


b

Up regulated

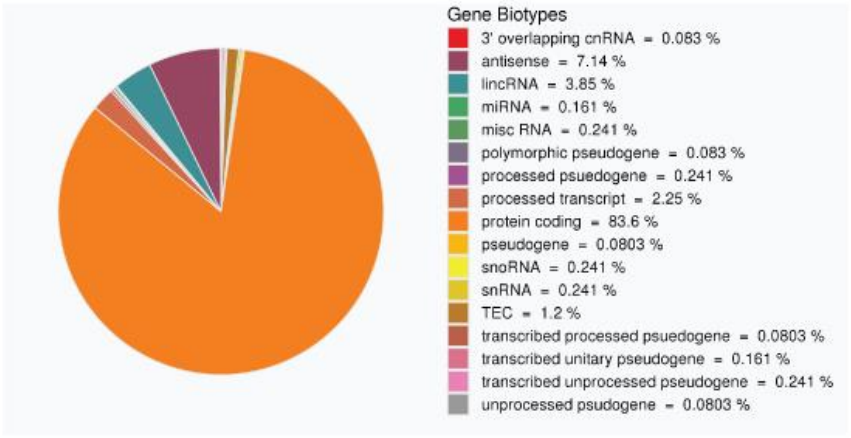


Non-protein coding transcripts only

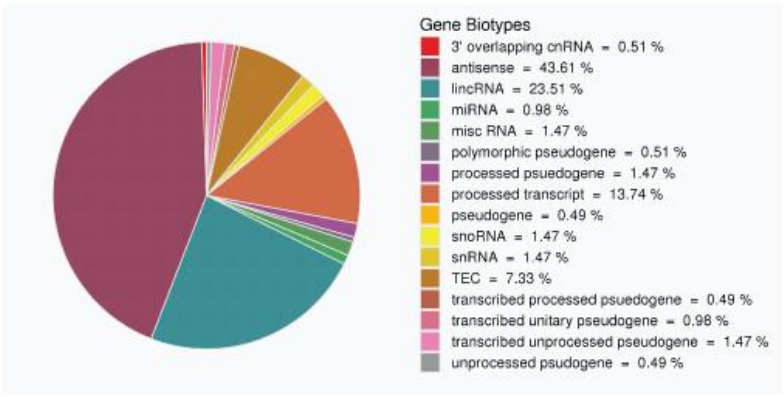


c

Down regulated

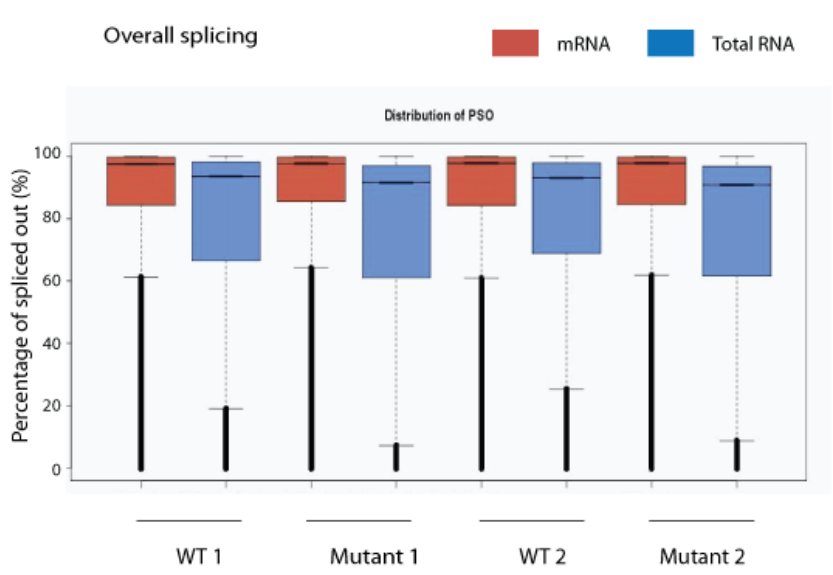


Non-protein coding transcripts only



d

Overall splicing



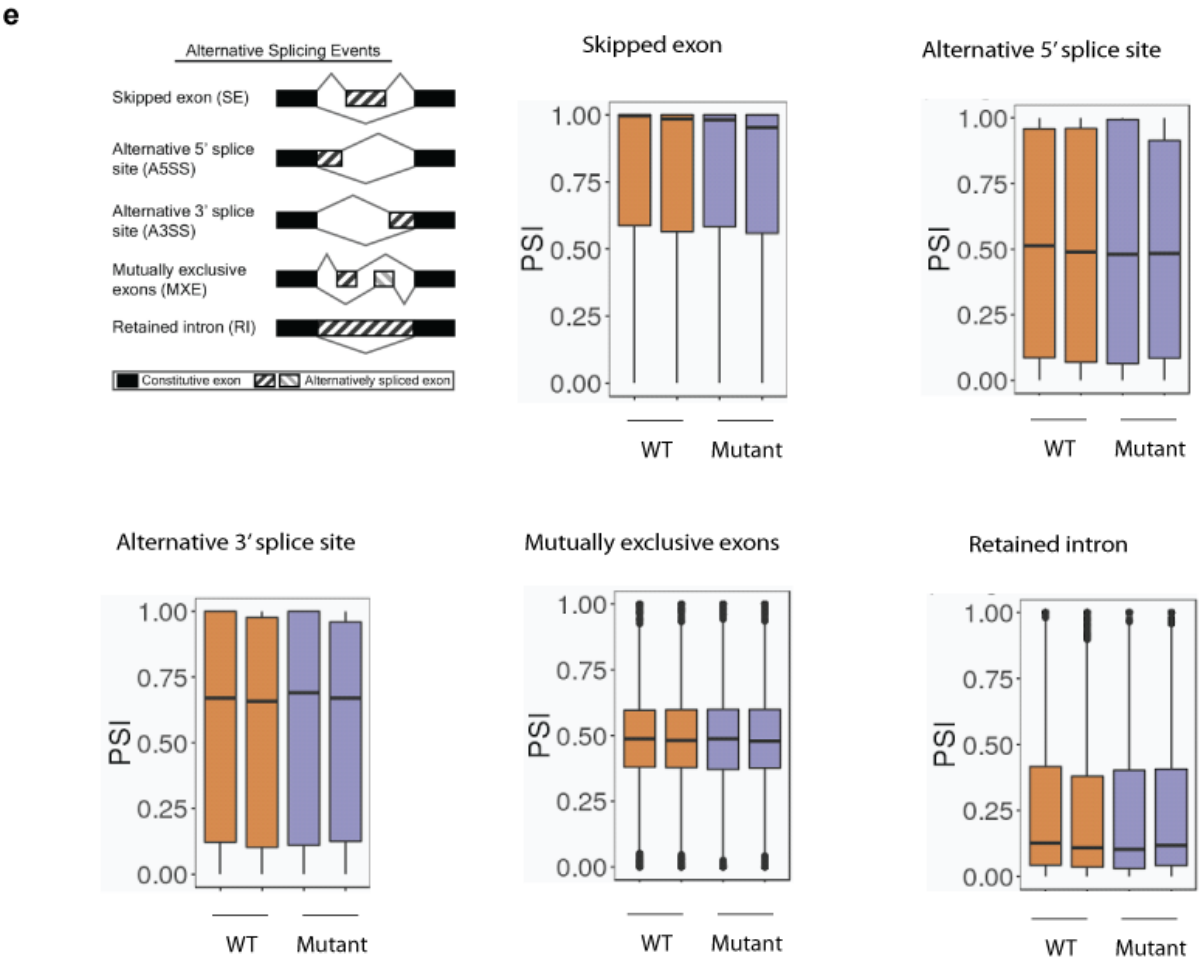


Figure 3.2.2 The effect of mutant Tgs1 on gene expression and splicing.

(a) MA plot of differentially expressed genes in mutant MEFs vs WT MEFs. FC=fold change. Red dots are genes with >2-fold differential expression (Log_2FC >1 or <-1) in mutant vs WT MEFs. **(b)** Biotypes of upregulated genes in mutant MEFs. **(c)** Biotypes of the downregulated genes in mutant MEFs. **(d)** Overall splicing efficiency in WT and mutant MEFs. Percent spliced out (PSO) was calculated to evaluate the overall splicing efficiency. Libraries were prepared using ribo-depletion + poly(A) enrichment (red bar, mRNA) or ribo-depletion only (blue bar, total RNA). **(e)** Specific splicing events in WT and mutant MEFs. Percent spliced in (PSI) was calculated to evaluate the splicing efficiency.

3.2.3 TMG-capped RNAs in mutant MEFs

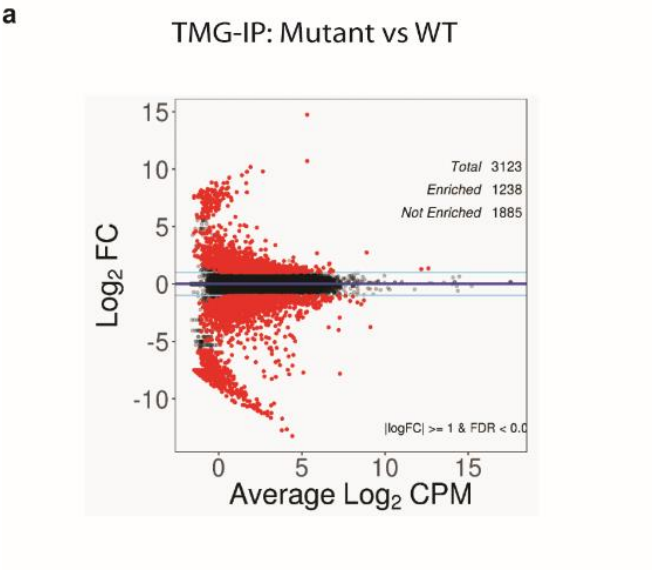
Tgs1 is recruited to its substrates via interaction with proteins such as SmB, dyskerin and Nop58 (Girard et al., 2008; Mouaikel et al., 2002; Pradet-Balade et al., 2011). The N-terminus of Tgs1 (amino acids 1-384) contains a GXXG (Gly-X-X-Gly) motif that was found in many RNA binding proteins (Burd & Dreyfuss, 1994; Lewis et al., 2000). The deletion of exon 3 and 4 does not affect methyltransferase activity but could affect the recruitment of mutant Tgs1 to certain substrates, thus affect their cap structure. TMG-IP was performed with total RNA from both mutant Tgs1 MEFs and wildtype MEFs. Libraries were made and subjected to next-generation sequencing analysis. 26-30 million 100-base-pair (bp) paired-end reads were generated. The reads were mapped to *Mus musculus* genome GRCm38.88 by STAR. The differential enrichment in TMG-IP was analyzed by EdgeR. By threshold of fold change of 2, 1238 genes showed increase in enrichment and 1885 genes showed decrease in enrichment (Figure 3.2.3a).

Genes were ranked with the magnitude of their fold change in enrichment (Figure 3.2.3b, c). Surprisingly, the majority of RNAs whose cap structures were altered were mRNAs. Canonical mRNAs, which contain monomethylguanosine caps, are not considered as substrates of Tgs1. The only TMG-capped mRNAs identified so far are selenoprotein mRNAs (Wurth et al., 2014). Further analysis was done to correlate the enrichment in TMG-IP to the gene expression level in input samples (Figure 3.2.3d, e). In category one, genes with increased TMG-IP enrichment and decreased/unchanged input level were isolated. In this category, there is high confidence that the increase in TMG-IP enrichment is not due to differences in expression level. In category two, genes with decreased TMG-IP enrichment and increased/unchanged input level were isolated. In this category are RNAs for which we are confident the decrease in TMG-IP enrichment is not due to difference in expression level. The majority of the RNAs whose cap

structures were altered were still mRNAs. It is a surprising result because in the literature, majority of known TMG-capped RNAs are noncoding RNAs. Eukaryotic translation initiation fact eIF4E has 2~3 magnitudes of higher affinity to the MMG cap than to the TMG cap (Niedzwiecka et al., 2002; Rutkowska-Wlodarczyk et al., 2008). The presence of TMG caps on mRNAs, if true, could impede translation. In this case, the TMG capping could serve as a regulatory mechanism for protein expression.

The only known factors that recruit Tgs1 to mRNAs are SECIS binding protein 2 (SBP2) and HIV-1 Rev protein, neither has been reported to associate with the mRNAs enriched in the TMG-IP. If Tgs1 is responsible for the hypermethylation of those mRNAs, novel Tgs1-recruting factors could be involved. Alternatively, those mRNAs could be methylated by a different methyltransferase (s). It is worth noticing that among the mRNAs that are enriched in TMG-IP, X-linked interleukin-1 receptor accessory protein-like 2 (Il1rapl2) has been reported to be hypermethylated on genomic DNA level (Lleo et al., 2015). DNA methyltransferase DNMTs are responsible for CpG methylation of genomic DNA on the X chromosome. DNMT2, which shares strong homology with other DNMTs, also methylates tRNAs using a DNA methyltransferase-like catalytic mechanism (Jurkowski et al., 2008). It raises the possibility that DNMT2 could methylate Il1rapl2 mRNAs due to its versatility. The third possibility is that the anti-TMG cap antibodies we used could recognize modifications other than the cap structure on those RNAs. In this case, those RNAs could still be MMG-capped and enriched in the TMG-IP.

a



b Increased enrichment in TMG-IP

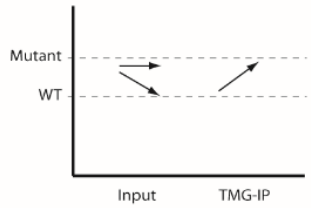
log2FC	gene_biotype	external_gene_name	description
14.7433848	snRNA	Rnu1b6	U1b6 small nuclear RNA [Source:MGI Symbol;Acc:MGI:104604]
10.7021872	lincRNA	Gm10685	predicted gene 10685 [Source:MGI Symbol;Acc:MGI:3641752]
10.1880633	protein_coding	Arfgef3	ARFGEF family member 3 [Source:MGI Symbol;Acc:MGI:106387]
9.80315171	protein_coding	Apoh	apolipoprotein H [Source:MGI Symbol;Acc:MGI:88058]
9.773438	protein_coding	Jph1	junctophilin 1 [Source:MGI Symbol;Acc:MGI:1891495]
9.47425808	antisense	Gm28376	predicted gene 28376 [Source:MGI Symbol;Acc:MGI:5579082]
9.25527803	protein_coding	Prss22	protease, serine 22 [Source:MGI Symbol;Acc:MGI:1918085]
8.83384747	protein_coding	E230025N22Rik	Riken cDNA E230025N22 gene [Source:MGI Symbol;Acc:MGI:3687212]
8.83384747	protein_coding	Agbl4	ATP/GTP binding protein-like 4 [Source:MGI Symbol;Acc:MGI:1918244]
8.77508552	protein_coding	Gm38393	predicted gene, 38393 [Source:MGI Symbol;Acc:MGI:5613898]

Figure 3.2.3 The effect of mutant Tgs1 on RNA cap structure.

c Decreased enrichment in TMG-IP

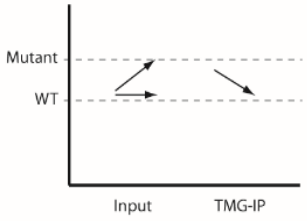
log2FC	gene_biotype	external_gene_name	description
-13.2421129	protein_coding	Efemp1	epidermal growth factor-containing fibulin-like extracellular matrix protein 1 [Source:MGI Symbol;Acc:MGI:1339998]
-12.7405378	protein_coding	Thy1	thymus cell antigen 1, theta [Source:MGI Symbol;Acc:MGI:98747]
-12.6479856	protein_coding	Lrrc17	leucine rich repeat containing 17 [Source:MGI Symbol;Acc:MGI:1921761]
-11.7194055	protein_coding	Sema3d	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3D [Source:MGI Symbol;Acc:MGI:1860118]
-11.662045	protein_coding	Dkk2	dickkopf WNT signaling pathway inhibitor 2 [Source:MGI Symbol;Acc:MGI:1890663]
-11.4471651	protein_coding	Sfrp1	secreted frizzled-related protein 1 [Source:MGI Symbol;Acc:MGI:892014]
-11.2321639	protein_coding	Lama2	laminin, alpha 2 [Source:MGI Symbol;Acc:MGI:99912]
-11.2321639	antisense	Gm16897	predicted gene, 16897 [Source:MGI Symbol;Acc:MGI:4439821]
-11.2275195	protein_coding	Prelp	proline arginine-rich end leucine-rich repeat [Source:MGI Symbol;Acc:MGI:2151110]
-11.1213661	protein_coding	AW551984	expressed sequence AW551984 [Source:MGI Symbol;Acc:MGI:2143322]

d Category 1



tmg_Log2FC	gene_biotype	external_gene_name	description
7.778375	protein_coding	Il1rapl2	interleukin 1 receptor accessory protein-like 2 [Source:MGI Symbol;Acc:MGI:1913106]
7.71726	protein_coding	Zic3	zinc finger protein of the cerebellum 3 [Source:MGI Symbol;Acc:MGI:106676]
7.443068	antisense	Gm15591	predicted gene 15591 [Source:MGI Symbol;Acc:MGI:3783038]
7.10424	protein_coding	Col17a1	collagen, type XVII, alpha 1 [Source:MGI Symbol;Acc:MGI:88450]
7.10424	protein_coding	Tmem86b	transmembrane protein 86B [Source:MGI Symbol;Acc:MGI:1915505]
6.784931	antisense	Gm11832	predicted gene 11832 [Source:MGI Symbol;Acc:MGI:3650310]
6.784931	TEC	Gm44775	predicted gene 44775 [Source:MGI Symbol;Acc:MGI:5753351]
6.660589	protein_coding	Zfp78	zinc finger protein 78 [Source:MGI Symbol;Acc:MGI:107783]
6.524511	protein_coding	Trim43c	tripartite motif-containing 43C [Source:MGI Symbol;Acc:MGI:3647365]
6.374248	antisense	Gm45187	predicted gene 45187 [Source:MGI Symbol;Acc:MGI:5753763]

e Category 2



tmg_Log2FC	gene_biotype	external_gene_name	description
-8.04373	protein_coding	Tmem74	transmembrane protein 74 [Source:MGI Symbol;Acc:MGI:2443417]
-8.00083	protein_coding	Exd1	exonuclease 3'-5' domain containing 1 [Source:MGI Symbol;Acc:MGI:3045306]
-7.81519	antisense	Med9os	mediator complex subunit 9, opposite strand [Source:MGI Symbol;Acc:MGI:1916628]
-7.7648	protein_coding	Vwc2	von Willebrand factor C domain containing 2 [Source:MGI Symbol;Acc:MGI:2442987]
-7.71258	antisense	Gm37634	predicted gene, 37634 [Source:MGI Symbol;Acc:MGI:5610862]
-7.65839	lincRNA	Gm43254	predicted gene 43254 [Source:MGI Symbol;Acc:MGI:5663391]
-7.60209	protein_coding	Rph3al	rabphilin 3A-like (without C2 domains) [Source:MGI Symbol;Acc:MGI:1923492]
-7.54351	protein_coding	Kcnip3	Kv channel interacting protein 3, calsenilin [Source:MGI Symbol;Acc:MGI:1929258]
-7.54351	antisense	Gm15563	predicted gene 15563 [Source:MGI Symbol;Acc:MGI:3783012]
-7.41868	protein_coding	Cobl	cordon-bleu WH2 repeat [Source:MGI Symbol;Acc:MGI:105056]

(a) MA plot of differential TMG-IP enrichment in mutant MEFs vs WT MEFs. FC = fold change. Red dots are genes with > 2 folds differential enrichment (Log_2FC >1 or <-1) in mutant vs WT MEFs. Adjusted p-value < 0.01. **(b)** Top 10 genes with increased TMG-IP enrichment in mutant MEFs. **(c)** Top 10 genes with decreased TMG-IP enrichment in mutant MEFs. **(d)** Increased TMG-IP enrichment adjusted by input gene expression: Category 1. Genes that were downregulated/unchanged in the input sample but showed increase in TMG-IP enrichment were included in this category. **(e)** Decreased TMG-IP enrichment adjusted by input gene expression: Category 2. Genes that were upregulated/unchanged in the input sample but showed decrease in TMG-IP enrichment were included in this category.

3.3 Tgs1 is essential in human cells

As the putative Tgs1-KO MEFs turned out to contain a mutant version of Tgs1, we sought to directly knock out Tgs1 in human cell lines via genome editing tools.

3.3.1 Attempt to knock out Tgs1 using ZFN

Zinc-finger nucleases (ZFN) consist of two key components: the DNA binding domain of zinc finger proteins (ZFP), and the nuclease domain of *Fok*I restriction endonuclease. Each Cys2-His2 finger in ZFP recognizes 3bp of DNA. With proper combination of the fingers, ZFP can bind specific genomic sequences, where double strand breaks (DSB) are created by the *Fok*I nuclease. The requirement of dimerization for *Fok*I to cleave DNA further increases the specificity of ZFN, as two independent binding events must happen adjacently at the target site. Upon cleavage by ZFN, one of the two major pathways is involved in repairing the DSB. Non-homologous end joining (NHEJ), which is error-prone, usually leads to insertion/deletion. Alternatively, homology-directed repair (HDR) can lead to precise repair based on a template.

Both pathways can be leveraged to disrupt target genes. With NHEJ, insertion/deletion (indel) can result in shift of the open reading frame and create premature stop codons in the target gene. With HDR, an antibiotic resistance marker can be inserted and disrupt the target gene.

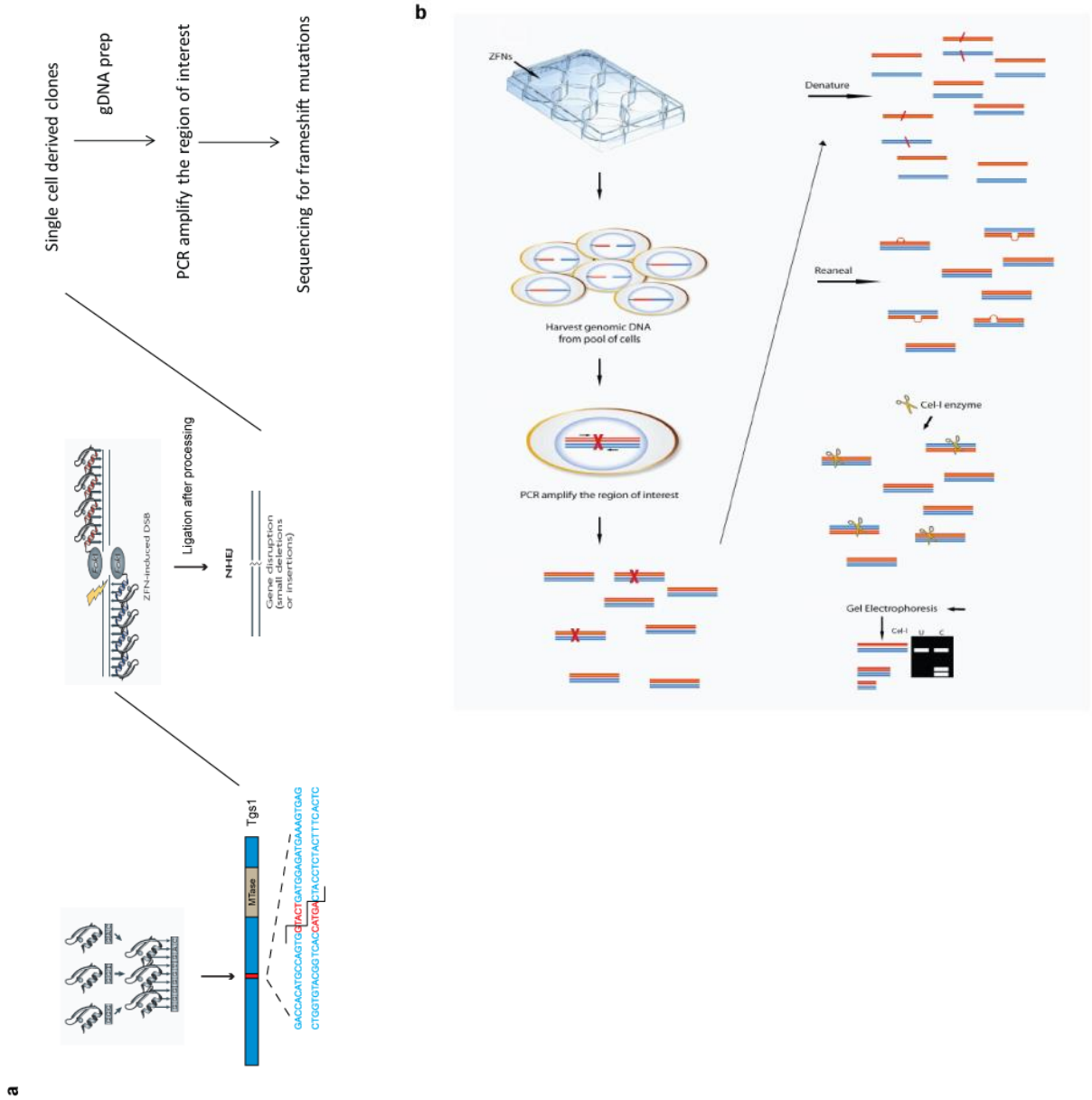
In HT1080 cells, ZFN (Sigma-Aldrich) was deployed to disrupt Tgs1 gene by creating frameshift mutations upstream of the Tgs1 methyltransferase motif via NHEJ (Figure 3.3.1a). To examine the presence of indels after ZFN treatment, SURVEYOR assay was used. In this assay, the genomic region around the ZFN cut site was amplified by PCR. The PCR products were denatured at high temperature, then reannealed. If present, indels will form imperfect matches with the wildtype strands. The hybrid DNA duplexes were digested by CEL-I, which specifically cleaved mismatched sites and resulted in altered migration patterns in gel electrophoresis (Figure 3.3.1b). ZFN activity, as defined by the frequency of indels resulted from NHEJ, was calculated based on ratio of cleaved products.

The SURVEYOR assay showed reasonably high activity of ZFN (Figure 3.3.1c). After ZFN treatment, the HT1080 population was expected to consist of $Tgs1^{wt/wt}$, $Tgs1^{wt/mutant}$ and $Tgs1^{mutant/mutant}$ cells. To isolate individual cells carrying frameshift mutations in Tgs1 gene ($Tgs1^{mutant/mutant}$), single cell selection was carried out. ZFN-treated HT1080 cells were diluted and seeded into 96-well plates at the density of 2 cells/well. After 6-7 days, the plates were visually inspected under a microscope. Colonies of HT1080 cells can be observed. The wells that just contained one single colony were digested and expanded. Genomic DNA was extracted from the single cell derived clones. ZFN target region was amplified by PCR and sequenced. The presence of ZFN-mediated mutations at the cut site resulted in mixture of signal in Sangers

sequencing (Figure 3.3.1d). 46 clones were sequenced in total, 5 of them show heterogeneous signal, indicating those were either $Tgs1^{wt/mutant}$ or $Tgs1^{mutant/mutant}$ clones. To determine the genetic identity of those clones, PCR products of their ZFN target region were cloned into pCR2.1-TOPO vector. Individual colonies were picked, and DNA isolated and sequenced in 96-well plate format. All the 5 heterozygous clones contain at least one wild type Tgs1 allele ($Tgs1^{wt/mutant}$). The mutations included 1bp, 2bp, 4bp, 14bp and 29bp deletion at the ZFN cut site (Figure 3.3.1d). Those mutations shifted open reading frame, resulting in premature stop codons, disrupting the mutant Tgs1 allele.

To obtain a Tgs1 knockout cell line, 2nd round of knockout was required to remove the remaining wildtype allele. Since the 29bp deletion removed the entire ZFN cut site, it eliminated the possibility that the mutant allele would be cut by ZFN again, which could result in another mutation restoring the open reading frame. Clone $Tgs1^{wt/\Delta 29bp}$ was treated with ZFN. The same single cell selection resulted in 64 viable clones. Since the clones were heterozygous to start with, and it was not feasible to use 96-well plate sequencing to genotype all 64 clones, we directly examined the presence of Tgs1 protein by western blot. Tgs1 protein remained in all viable cell lines after 2nd round of ZFN treatment and single cell selection (Figure 3.3.1e).

In the first round of ZFN treatment, 5 out of 46 clones carried frameshift mutations. But none of the 64 clones in the second round of ZFN treatment carried frameshift mutations. It is likely that disrupting the remaining wild type Tgs1 allele causes growth disadvantage, which will reduce the chance of viable cells carrying frameshift mutations. Either more clones need to be screened, or the efficiency of genome editing need to be further improved to knock out Tgs1.


a
Single cell derived clones
gDNA prep
PCR amplify the region of interest
Sequencing for frameshift mutations
ZFN-induced DSB
Ligation after processing
NHEJ
Gene disruption
(small deletions
or insertions)
Tgs1
MTase
GACCACATGCCAGTGGTACTGATGGAGATGAAAGTGAG
CTGGTGTACGGTCACCATGACTACCTCTACTTTCACTC
b
ZFNs
Harvest genomic DNA
from pool of cells
PCR amplify the region of interest
Denature
Reaneal
Cel-I enzyme
Gel Electrophoresis
Cel-I
U
C

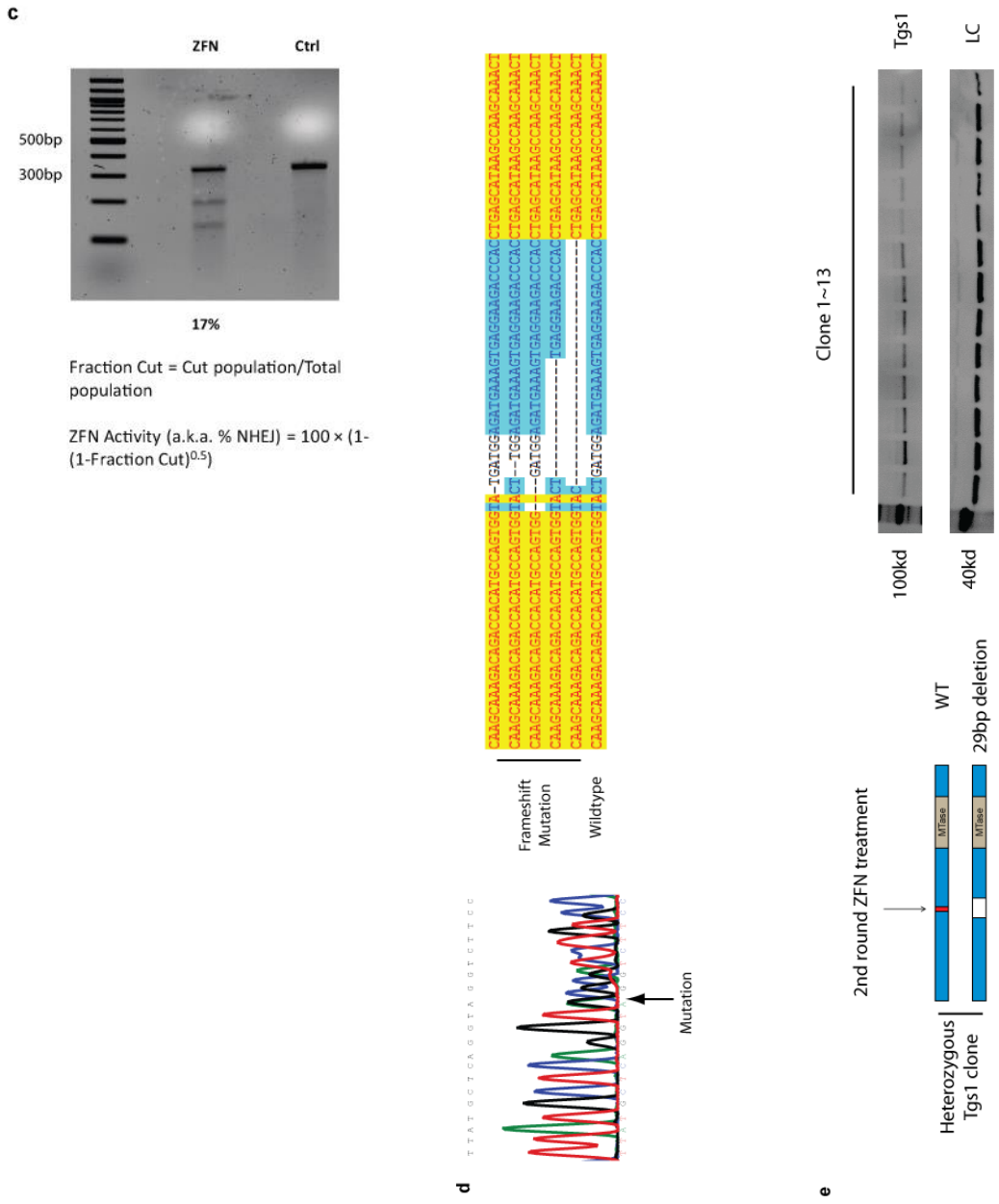


Figure 3.3.1 Attempt to knock out Tgs1 with Zinc finger nuclease (ZFN).

(a) Schematic of ZFN knockout procedure. **(b)** Schematic of SURVEYOR assay. **(c)** Activity of ZFN measured by SURVEYOR assay. **(d)** Sanger sequencing of frame shift mutations after ZFN treatment and single cell selection. **(e)** Western blot of Tgs1 from single cell derived clones after 2nd ZFN treatment. The first 13 clones were shown.

3.3.2 Attempt to knock out Tgs1 using CRISPR-Cas system

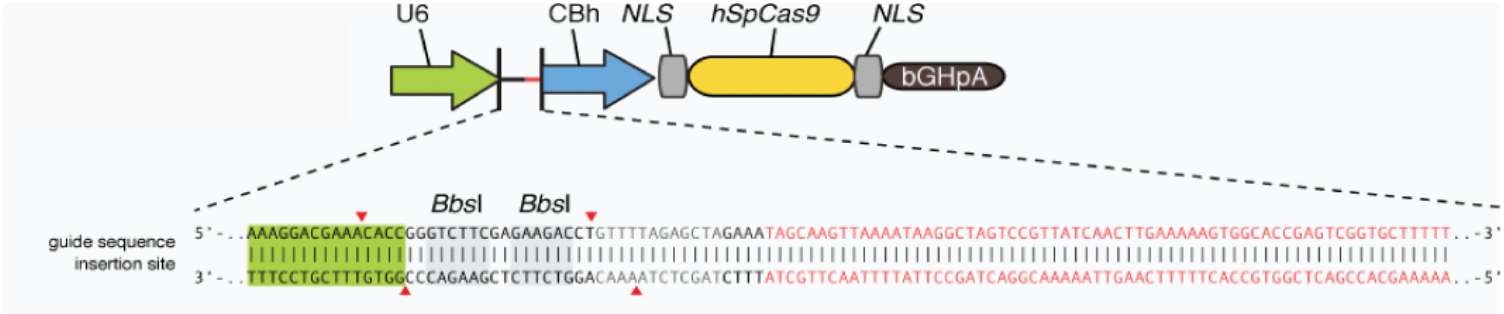
CRISPR-Cas system was identified in adaptive immune system of bacteria and archaea (Bhaya, Davison, & Barrangou, 2011). Foreign genetic elements such as viruses are integrated into a genomic region containing clustered regularly interspaced short palindromic repeats (CRISPR). Their sequences are transcribed into small RNAs, which guide Cas proteins to the invading foreign genetic elements via Watson-Crick base-pairing and cleave them.

Due to its simplicity and efficiency, CRISPR-Cas system was rapidly developed into a powerful genome editing tool. We utilized the CRISPR package developed by Feng Zhang's lab (Figure 3.3.2a) (Ran et al., 2013). This package is based on type II CRISPR system in *Streptococcus pyogenes.* It consists of the nuclease Cas9, CRISPR RNA (crRNA) array which encodes the guide RNAs and a trans-activating RNA (tracrRNA). The Cas9 from *Streptococcus pyogenes* is optimized with codon usage for expression in human cells.

The CRISPR showed slightly higher activity than ZFN (Figure 3.3.2b). Instead of relying on NHEJ, we tried a different approach to improve the chance of disrupting the remaining wild type Tgs1 allele. A homology-directed repair (HDR) donor carrying puromycin resistance marker and flanked by TGS1 genomic sequences was introduced together with ZFN, followed by puromycin selection (Figure 3.3.2c). Only cells inserted with HDR donor, which carried puromycin resistance marker, will survive. The double strand break (DSB) created by ZFN/CRISPR and the homology between the donor and TGS1 sequences strongly favored insertion at TGS1 genomic loci. Primers flanking the ZFN/CRISPR cut site were used to examine the insertion of HDR donor. In ZFN treated HT1080 cells, 9 clones were viable after puromycin selection. In CRISPR

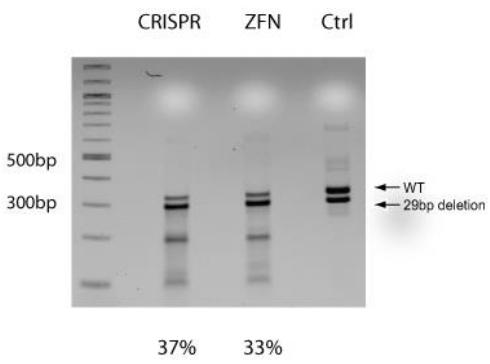
treated HT1080 cells, 16 clones were viable after puromycin selection. Unfortunately, the wild type Tgs1 allele remained in all the viable clones (Figure 3.3.2d).

a



Guide RNA sequence: 5′- CACC GtactGATGGAGATGAAAGTG - 3′
(Target Tgs1)
3′- CatgaCTACCTCTACTTTCAC CAAA - 5′

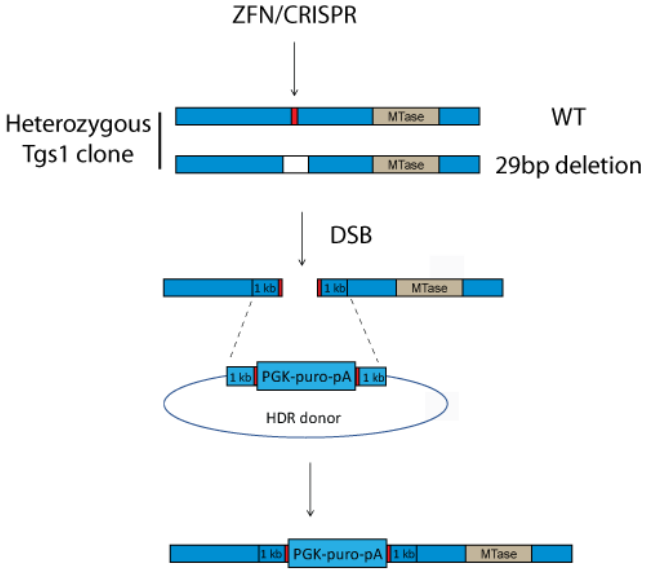
b



Fraction Cut = Cut population/Total population

ZFN Activity (a.k.a. % NHEJ) = 100 × (1-(1-Fraction Cut)$^{0.5}$)

c



d

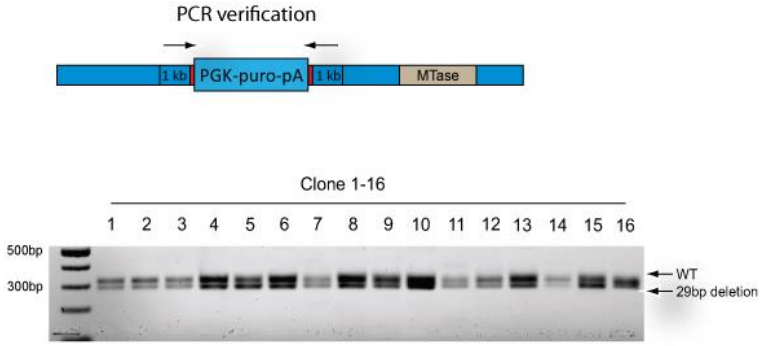


Figure 3.3.2 Attempt to knock out Tgs1 with CRISPR-Cas system.

(a) Schematic of CRISPR construct. **(b)** Schematic of homology-directed repair approach to disrupt Tgs1. **(c)** Activity of CRISPR and ZFN **(d)** PCR verification of the genome type of clones after puromycin selection. Clone 1-16 were shown.

3.3.3 Disrupt endogenous Tgs1 alleles in the presence of an integrated copy

After multiple failed attempts to knock out Tgs1 in human cell lines, we suspect that Tgs1 is essential for human cell viability. To address this possibility, a copy of CRISPR-resistant Tgs1 was integrated into the FRT site in 293 T-REx cell line (Life Technologies) (Figure 3.3.3a). In the presence of integrated Tgs1, CRISPR was employed to knock out the endogenous Tgs1 alleles. The expression of the integrated Tgs1 was turned on during the CRISPR knock out attempt and the subsequent single cell selection. Under this approach, even if Tgs1 is essential for human cell viability, disrupting the endogenous Tgs1 alleles would not grant Tgs1-KO cells disadvantage during the selection process.

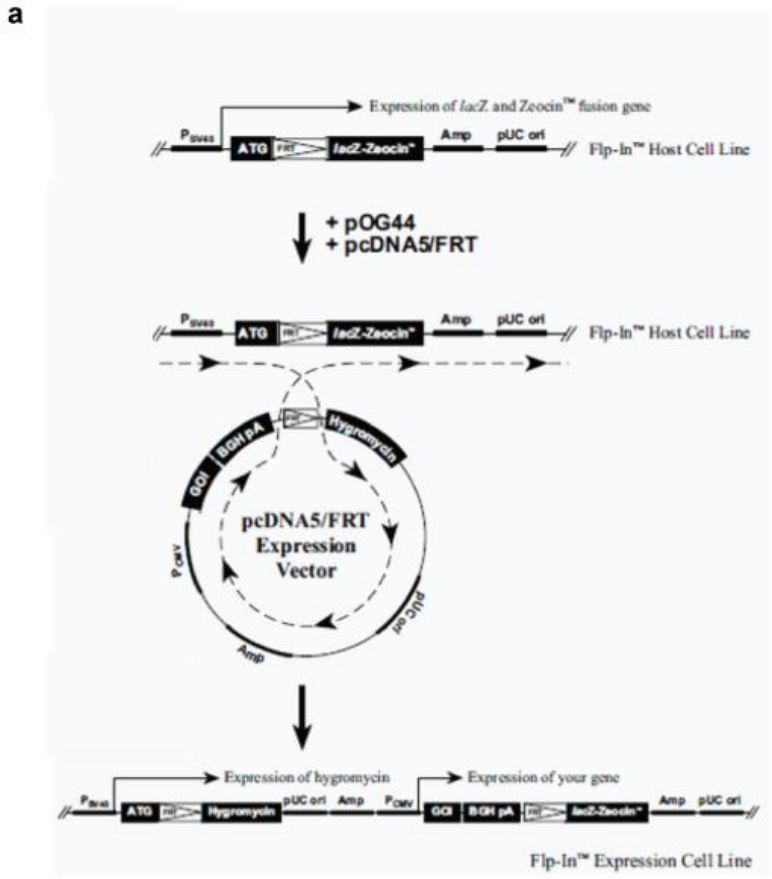
The single cell selection and genotyping were carried out as described in 3.3.1. CRISPR-treated 293 T-REx cells were diluted and seeded into 96-well plates. The plates were visually inspected under a microscope. The wells that just contained one single clone were digested and expanded. Genomic DNA was extracted from the single cell derived clones. CRISPR target region was amplified by PCR and sequenced. Heterozygous clones with mutations at the cut site will result

in mixture of signal in Sangers sequencing. The genetic identity of those heterozygous clones was further analyzed by cloning the PCR products and sequencing individual colonies.

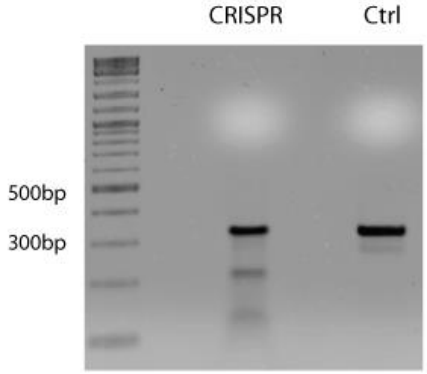
Cell lines derived from single clones with both Tgs1 alleles disrupted were identified. Even though the CRISPR activity was no higher than previous attempts (Figure 3.3.3b). Out of 72 clones that were sequenced, 8 were heterozygous. Further sequencing analysis revealed 3 clones containing only mutant alleles without the presence of wildtype Tgs1 allele. Some of the mutations were shown in Figure 3.3.3c. In clone 1, seven colonies were sequenced, four contained 3bp to 1bp substitution, resulting in premature stop codon at 1234bp in Tgs1 mRNA; the other three contained 16bp deletion, resulting in premature stop codon at 1255bp, no wildtype allele present. In clone 2, ten colonies were sequenced: four contained 2bp insertion, four contained 2bp deletion, and two contained 1bp insertion, each resulting in premature stop codon at 1273bp, 1234bp and 1237bp in Tgs1 mRNA. No wildtype allele was present. Three mutant Tgs1 alleles were present in clone 2. This is likely because 293 T-REx cell line is a hypotriploid cell line. The modal chromosome number is 64, occurring in 30% of cells.

The integrated Tgs1 is under control of an inducible promoter. If Tgs1 is indeed essential for human cell viability, turning off the integrated copy will lead to growth defect in cell lines with no endogenous wildtype Tgs1 alleles. However, the integrated Tgs1 showed high basal expression level in the absence of tetracycline induction (Figure 3.3.3d). No obvious growth defect was observed when the cells were seeded through regular procedure (2 ~ 3 x 1e5 cells per well, 6-well plate). The remaining integrated Tgs1 proteins were likely to sustain cell viability. Mild growth defect was present when cells were seeded at very low density (2 x 1e4 cells per well, 6-well plate, data not shown). Since the clones went through single cell selection, a growth

defect at low seeding density would render a disadvantage against Tgs1$^{mutant/mutant}$ clones. That was why previously we were unable to disrupt both alleles of Tgs1 without the presence of an integrated copy.



b

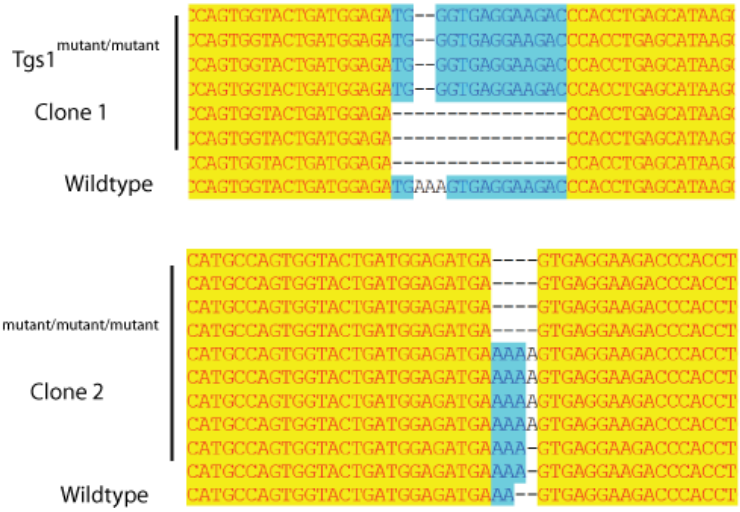


15%

Fraction Cut = Cut population/Total population

ZFN Activity (a.k.a. % NHEJ) = 100 × (1-(1-Fraction Cut)$^{0.5}$)

c



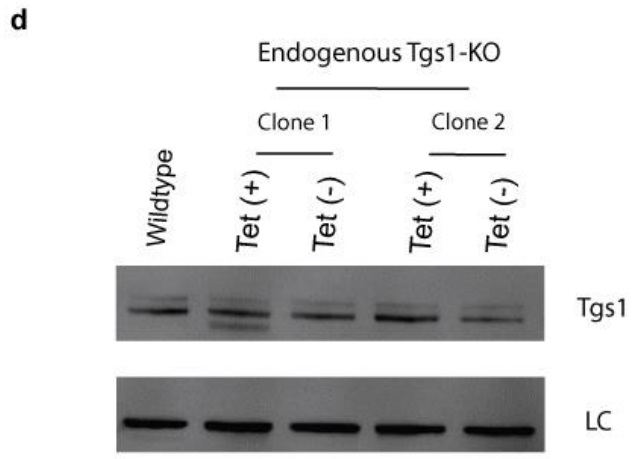


Figure 3.3.3 Disrupt endogenous Tgs1 alleles in Flp-In 293 T-REx cell line.

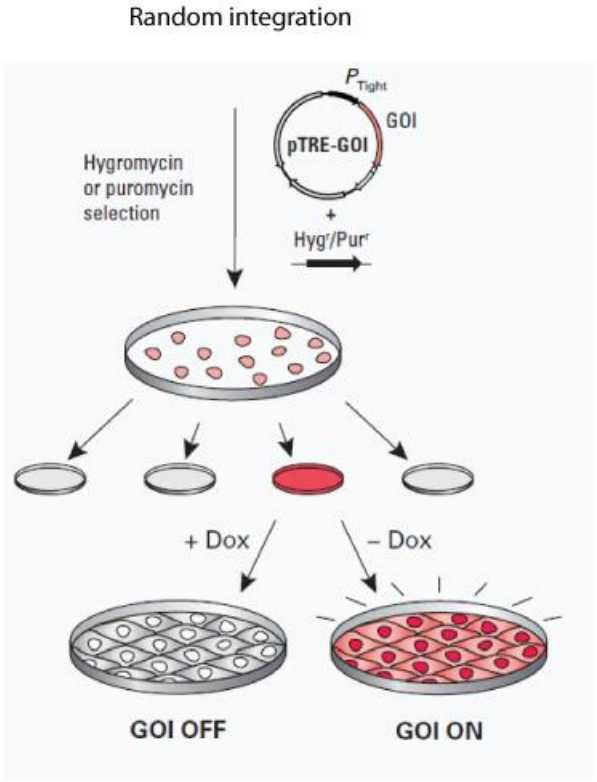
(a) Schematic of integrating full-length Tgs1 into Flp-In 293 T-REx cell line through FLP-FRT system. **(b)** SURVEYOR assay for CRISPR activity. **(c)** Frame shit mutation identified in endogenous Tgs1 alleles. **(d)** Tgs1 protein expression in the presence (+) and absence (-) of tetracycline. LC: loading control.

To reduce the basal expression level of integrated Tgs1, we switched to HT1080 Tet-off cell line (Clontech), which was supposed to have lower basal expression level. In 293 T-REx cell line, the Tet repressor binds to the Tet operator in the absence of tetracycline, resulting in inhibition of gene expression. In the presence of tetracycline, the conformation of Tet repressor is altered, releasing it from the Tet operator and leading to induction of gene expression. In HT1080 Tet-off cell line, Tet repressor is fused with the transcription domain from HSV VP16 protein. The binding of Tet repressor fusion protein to Tet operator sequence leads to induction of gene expression. In the presence of doxycycline (tetracycline derivative), the fusion protein is released, resulting in inhibition of gene expression.

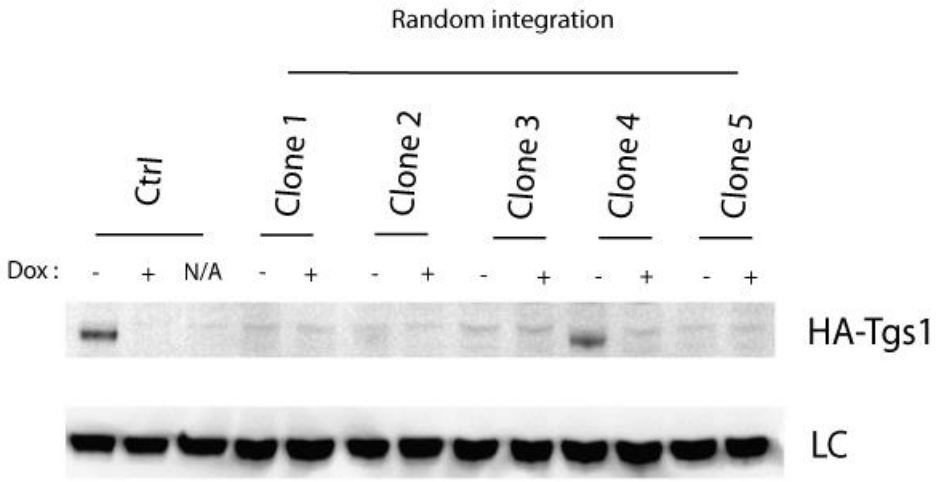
The CRISPR-insensitive copy of Tgs1 was tagged with HA tag at N-terminus (HA-Tgs1) and inserted into HT1080 Tet-off cell line through random integration (Figure 3.3.4a). Single clones integrated with HA-Tgs1 together with puromycin resistance marker were selected. The induction of integrated HA-Tgs1 was tested in the presence and absence of doxycycline (Figure 3.3.4b). The clone that carried inducible HA-Tgs1 (clone 4) was chosen for CRISPR treatment, followed by single cell selection. The disruption of endogenous Tgs1 alleles in HT1080 Tet-off cells was examined by the same procedure as described for 293 T-REx cell line. Mutations were identified in 12 cell lines (96 screened in total). 4 of them carried no wildtype Tgs1 allele. For example, the PCR product of one mutant cell line was cloned and sequenced. Eighteen colonies were sequenced in total: six contained 1bp insertion; two contained 1bp deletion; ten colonies contained 14bp deletion. No wildtype was allele present (Figure 3.3.4c). Three different mutant Tgs1 alleles were observed in the same mutant cell line. Because pseudodiploidy was frequently noted in HT1080 cells, ranging from 44 to 48 chromosomes, especially with abnormal number of group E and group C chromosomes (including chromosome 8, where TGS1 gene is located).

In the presence of doxycycline, the expression of inducible Tgs1 was shut down. A severe growth defect was observed (Figure 3.3.4d). Moreover, 3 out of the 4 mutant cell lines died after more than three passages. The remaining viable cell line did not show a growth defect in the presence of doxycycline. The Tgs1 protein is not detectable on western blot. This particular cell line may have higher residual level Tgs1 that is below detection than other cell lines, whether it is due to higher translation activity or due to higher mRNA level is unknown. Without endogenous Tgs1 allele, the randomly integrated copy likely cannot sustain long term cell growth.

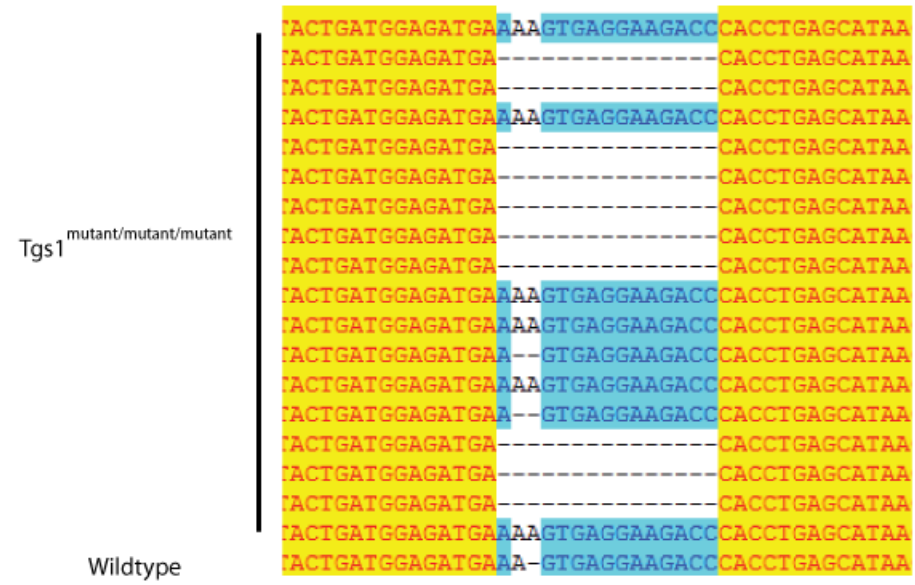
a


Random integration
P_{Tight}
GOI
pTRE-GOI
+
Hyg^r/Pur^r
Hygromycin
or puromycin
selection
+ Dox
– Dox
GOI OFF
GOI ON

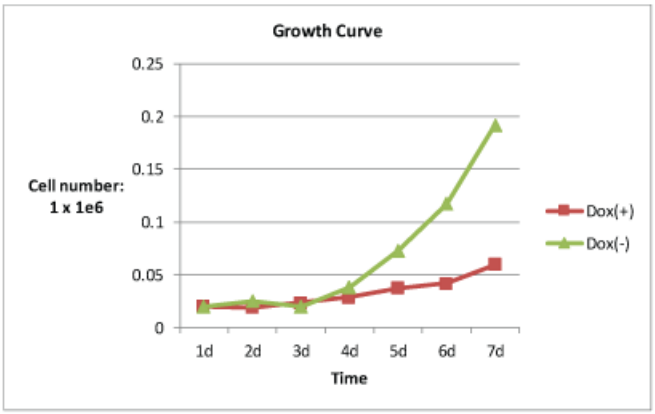
b


Random integration
Ctrl
Clone 1
Clone 2
Clone 3
Clone 4
Clone 5
Dox : - + N/A - + - + - + - + - +
HA-Tgs1
LC

c



d



	1d	2d	3d	4d	5d	6d	7d	Backman counter
Dox(+)	0.02	0.0192	0.024	0.0288	0.0372	0.042	0.06	(1 X 1e6)
Dox(-)	0.02	0.0252	0.0204	0.0384	0.0732	0.1176	0.192	

Figure 3.3.4 Disrupt endogenous Tgs1 alleles in HT1080 Tet-off cell line.

(a) Schematic of integrating full-length Tgs1 into HT1080 Tet-off cell line through random integration. **(b)** Western blot for the expression of randomly integrated Tgs1. **(c)** Frame shift mutation identified in endogenous Tgs1 alleles. **(d)** Growth curve of HT1080 Tet-off cell line without endogenous wildtype Tgs1 allele, in the presence of doxycycline (+) and absence of doxycycline (-).

3.4 Creation and characterization of a conditional knockdown cell line for human Tgs1

3.4.1 Tgs1 conditional knockdown cell line

Although the methyltransferase domain of Tgs1 protein (C-terminus) is conserved from yeast to mammals, the full-length Tgs1 proteins show wide variation in sequence, especially in the N-terminus (Figure 3.4.1a). But human and murine full-length Tgs1 enzymes are highly conserved. They both consist of 853 amino acids, sharing 73% identity and 84% similarity (Zhu et al., 2001). The conservation enabled us to construct a mutant version of human Tgs1 (Tgs1*) that mirrored the murine mutant Tgs1 identified in section 3.2 (Figure 3.4.1b).

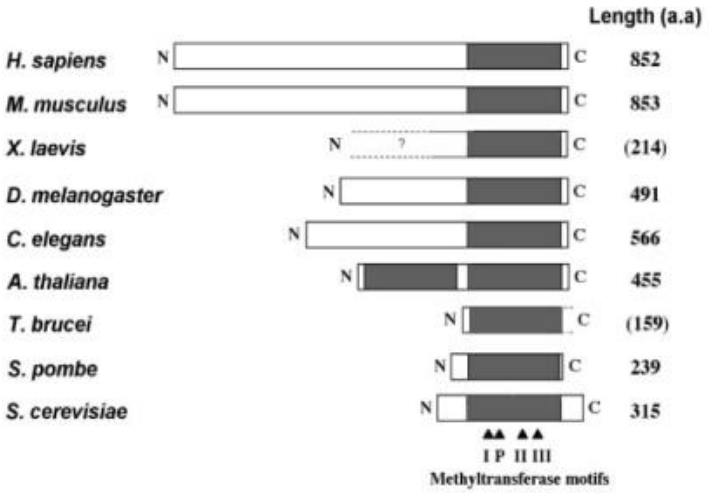
Human Tgs1* was integrated into 293 T-REx cell line via FLP-FRT system under the control of an inducible promoter. The inducible promoter contained a human cytomegalovirus (CMV) promoter followed by two Tet operator sequences ($TetO_2$). In the absence of tetracycline, Tet repressor bound to the 2 x $TetO_2$ sequences and inhibited transcription from the promoter. In the presence of tetracycline, the conformation of Tet repressor was altered and released from $TetO_2$, allowing gene expression from the promoter.

After introduction of Tgs1*, the endogenous Tgs1 alleles were disrupted by CRISPR. Single cell selection and genotyping were carried out as described in section 3.3. Forty-eight viable cell lines were obtained in total. The PCR products generated from amplifying the genomic region around CRISPR cut site were sequenced, 7 cell lines contained mutations around CRISPR cut site. PCR products from those cell lines were cloned, and individual colonies were sequenced. Further sequencing analysis revealed that one cell line contained frame shift mutations in both endogenous Tgs1 alleles. 86 *E.coli* colonies from PCR products of this

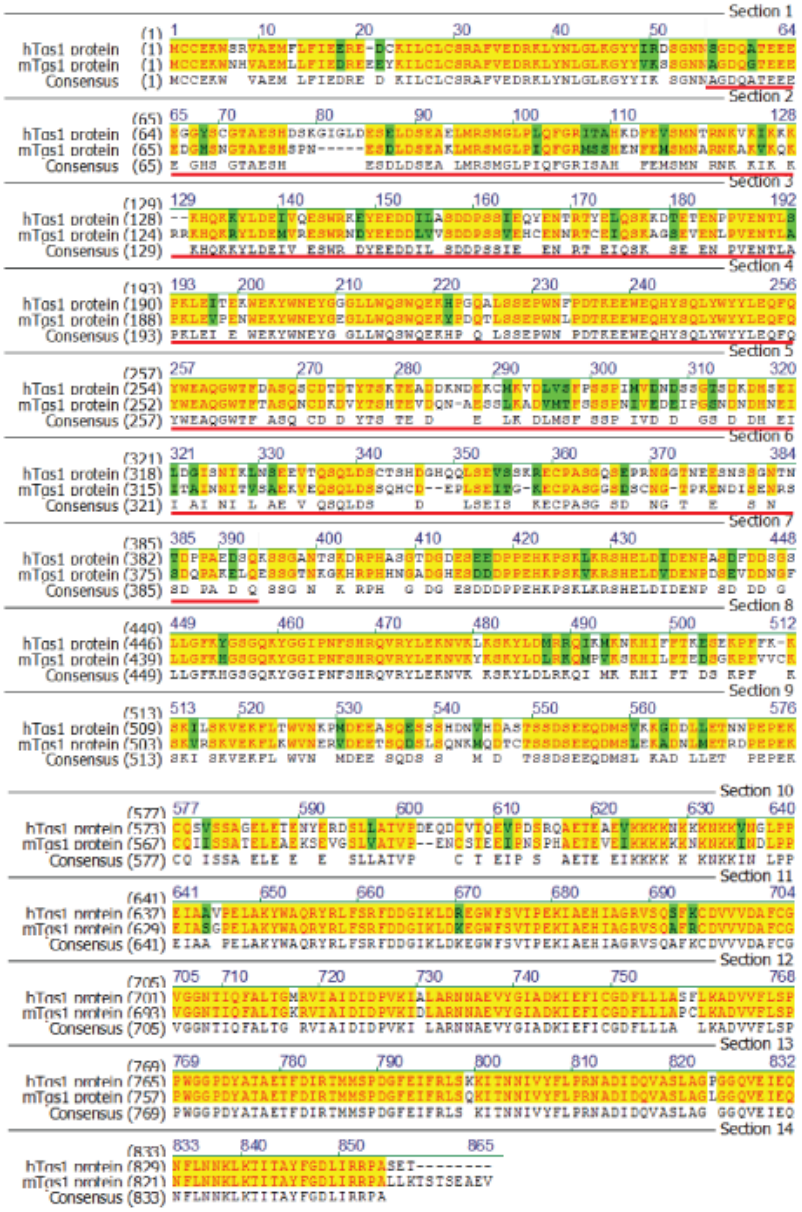
Tgs1mutant/mutant cell line contained either 1bp deletion or 2bp deletion, both resulting in premature stop codon, without presence of wildtype allele (Figure 3.4.1c). In eukaryotes, mRNA transcripts with premature stop codons are degraded through nonsense-mediated decay (NMD). RT-PCR confirmed the absence of full-length Tgs1 mRNA in the Tgs1mutant/mutant cell line (Figure 3.4.1d), indicating the disruption of all endogenous wildtype Tgs1 alleles.

In the absence of tetracycline, the integrated Tgs1* was turned off, and a severe growth defect was observed. Four days after depleting tetracycline from the media, floating cells were seen. After seven days, only a small number of viable cells were present in the absence of tetracycline. A detailed growth curve was generated. Starting at 2 x 1e4 cells per well in 6-well plate, wildtype cells (WT), mutant cells with tetracycline (Tet+), mutant cells without tetracycline (Tet-) were counted every day up to 7 days. ~30 times more cells of WT and ~8 times more cells of Mutant (Tet+) were present compared with Mutant (Tet-) (Figure 3.4.1e). It further confirmed that Tgs1 was essential for cell growth in human cells. The Tgs1* protein level eventually recovered after 10 days of shutdown (Figure 3.4.1f), likely due to the growth advantage of cells with higher expression level of integrated Tgs1*. As time went by, cells with higher basal expression level of Tgs1* replaced the ones with low basal Tgs1* expression, resulting in the resurface of Tgs1* in cell population.

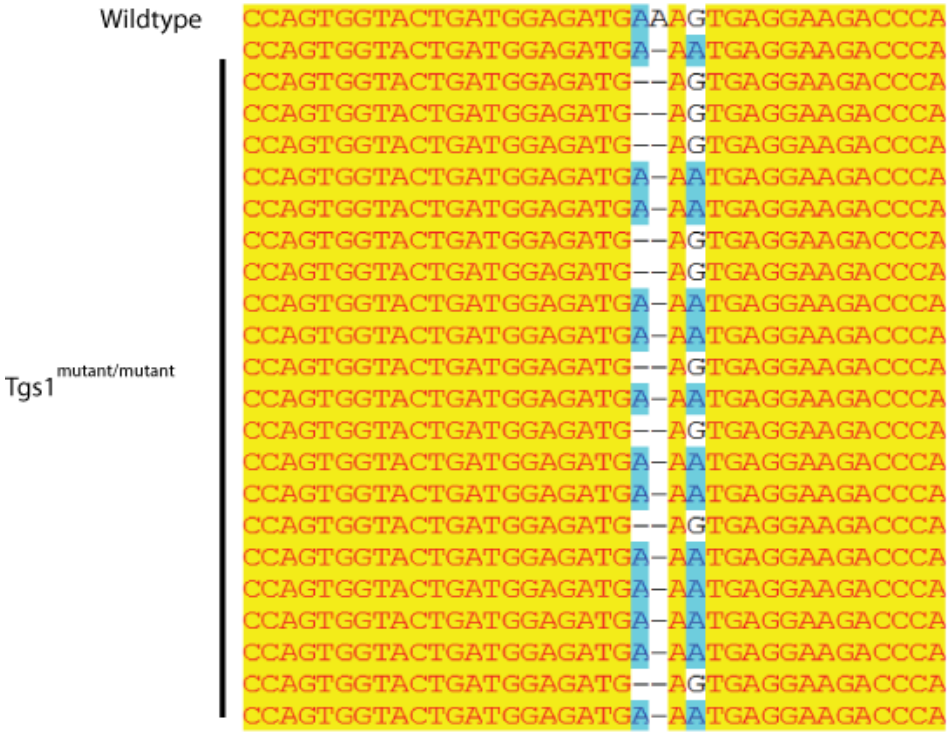
a


Length (a.a)
H. sapiens
M. musculus
X. laevis
D. melanogaster
C. elegans
A. thaliana
T. brucei
S. pombe
S. cerevisiae
N
C
?
852
853
(214)
491
566
455
(159)
239
315
I P II III
Methyltransferase motifs

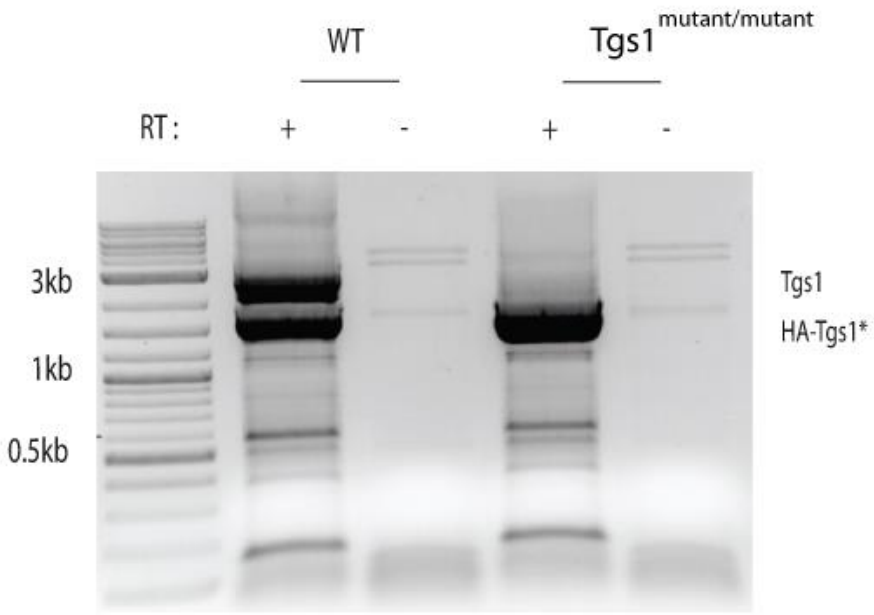
b


Section 1
Section 2
Section 3
Section 4
Section 5
Section 6
Section 7
Section 8
Section 9
Section 10
Section 11
Section 12
Section 13
Section 14
hTas1 protein
mTas1 protein
Consensus

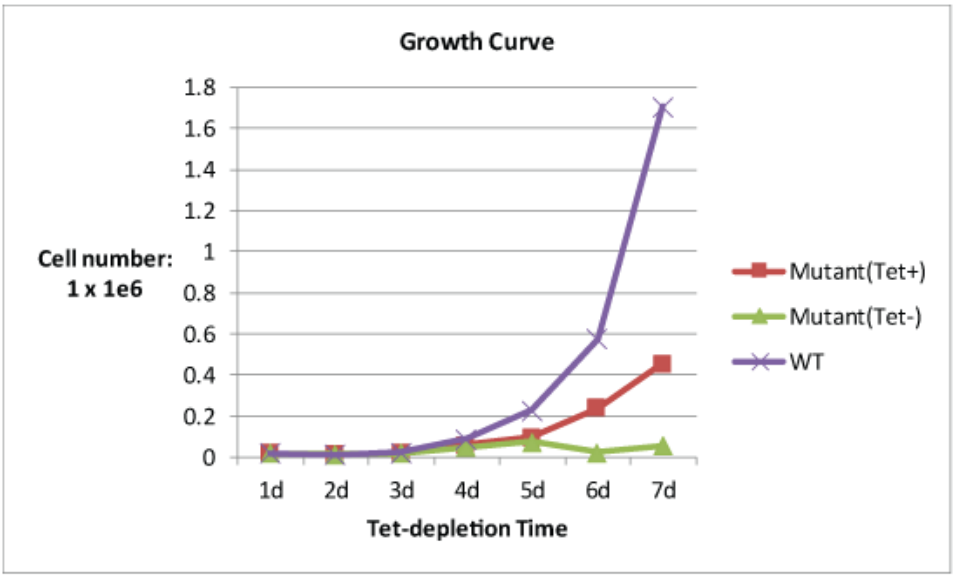
c


Wildtype
Tgs1 mutant/mutant
CCAGTGGTACTGATGGAGATGAAAGTGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA
CCAGTGGTACTGATGGAGATG--AGTGAGGAAGACCCA
CCAGTGGTACTGATGGAGATG--AGTGAGGAAGACCCA
CCAGTGGTACTGATGGAGATG--AGTGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA
CCAGTGGTACTGATGGAGATG--AGTGAGGAAGACCCA
CCAGTGGTACTGATGGAGATG--AGTGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA
CCAGTGGTACTGATGGAGATG--AGTGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA
CCAGTGGTACTGATGGAGATG--AGTGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA
CCAGTGGTACTGATGGAGATG--AGTGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA
CCAGTGGTACTGATGGAGATG--AGTGAGGAAGACCCA
CCAGTGGTACTGATGGAGATGA-AATGAGGAAGACCCA

d


WT
Tgs1 mutant/mutant
RT:
+
-
+
-
3kb
1kb
0.5kb
Tgs1
HA-Tgs1*

e



f

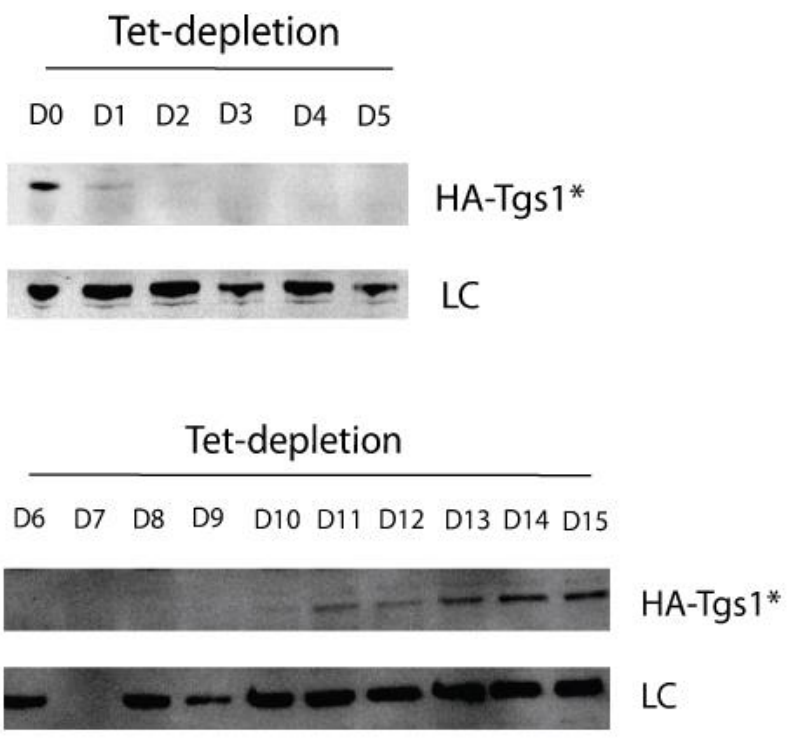


Figure 3.4.1 Creating a Tgs1* conditional knockout cell line in 293 T-REx cells.

(a) Evolutionary conservation of Tgs1 protein across different species, adapted from (Mouaikel et al., 2002). **(b)** Alignment of murine and human Tgs1 protein. Red line underscores the region that is deleted in mutant MEFs. A mutant version of human Tgs1 (Tgs1*) was created accordingly. Tgs1* was tagged with HA at the N-terminus (HA-Tgs1*). **(c)** Frame shift mutation identified in endogenous Tgs1 alleles. **(d)** RT-PCR of endogenous Tgs1 mRNA and mutant Tgs1 (HA-Tgs1*) mRNA. **(e)** Growth curve of wildtype cell line (WT) and Tgs1* conditional knockout cell line (Mutant), in the presence of tetracycline (+) and absence of tetracycline (-). **(f)** Western blot of HA-Tgs1* protein in the absence of tetracycline, from 1 day (D1) to 15 days (D15). LC: loading control.

3.4.2 U2 processing defect

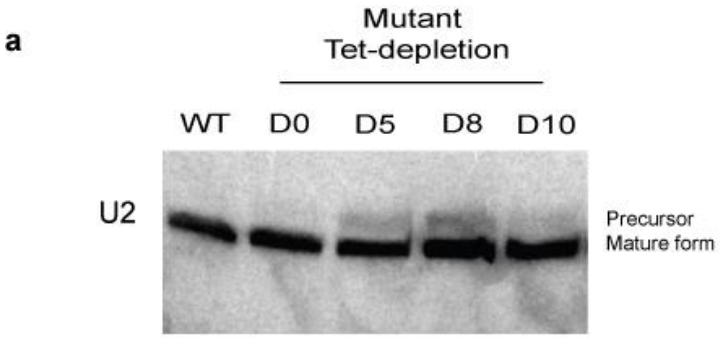
As the endogenous full-length Tgs1 was absent, the only version of Tgs1 expressed in this cell line was Tgs1*, which mirrored the murine Tgs1 mutant cell line. Since processing defect of U2 snRNA was observed in mutant MEFs, we examined U2 snRNA in the human Tgs1* 293 T-REx cell line (Figure 3.4.2a). Although the presumed precursor band above mature U2 in Tgs1* cell line was not as strong as observed in mutant MEFs, it became more obvious after shutting down Tgs1* by tetracycline depletion. The 3' ends of U2 were cloned via poly(A) tailing as described in section 3.2.1. Accumulation of human U2 precursors with ~10nt extension at 3' ends was observed (Figure 3.4.2b).

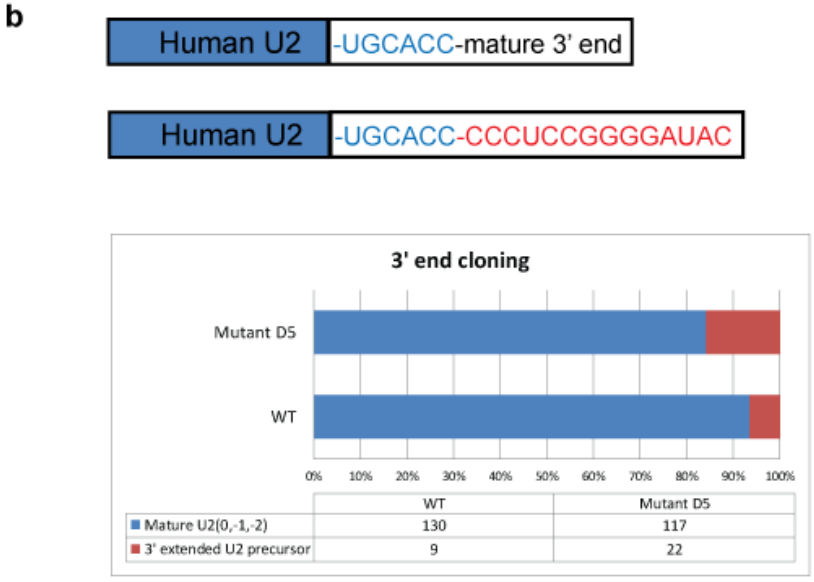
A next generation sequencing library for human U2 was also generated and analyzed as described in section 3.2.1. Accumulation of human U2 precursors, as indicated by the percentage of precursors, was slightly higher in Tgs1* 293 T-REx cell line compared with wildtype cell line. After shutting down Tgs1* by depleting tetracycline in the media for 5 days, the ratio grew higher. In wildtype cells, 6.1% U2 precursors were present compared to 93.9% mature form. In the presence of tetracycline, 7.2% U2 precursors were present. After depleting tetracycline for 5 days, the percentage of U2 precursors increased to 11.1%. As Tgs1* expression started to

recover after Tet-depletion for 10 days, the accumulation of U2 precursors was reduced correspondingly (from 11.1% to 8.8%) (Figure 3.4.2c). Detailed 3' end distribution of U2 precursors showed the accumulation of 10nt extended precursors that matched the size of the upper band in U2 northern blot. A small portion of 1nt extended precursors was also revealed. Previously, this portion was regarded as mature form due to the resolution of northern blot. In wildtype cells, among the 93.9% previously identified mature form, 6.4% were 1nt extended precursors (188nt, position +1). 79.4% were the exact mature form (187nt, position 0). The remaining 7.1% were 1~3nt shorter than mature form (position 184~186). In the presence of tetracycline, the 1nt extended precursors were 8.1%. After depleting tetracycline for 5 days, the 1nt extended precursors increased to 11.5%. As Tgs1* expression resurfaced after Tet-depletion for 10 days, the 1nt extended precursors reduced from 11.5% to 9.8%. (Figure 3.4.2 d). 3'-terminal non-templated nucleotides were further analyzed. At position 0, analysis showed that one non-templated adenosine was present on 3' end of mature U2. In wildtype cell line, there were 51.73% mature U2 with one 3' terminal adenosine. The ratio was increased to 58.87% in mutant cell line, but reduced to 48.77% after 5-day Tet-depletion, then increased again to 54.41% as Tgs1* expression resurfaced after Tet-depletion for 10 days (Figure 3.4.2 e). At position +1 and position +10, the non-templated nucleotides are extremely rare (<0.5%, not shown in figures).

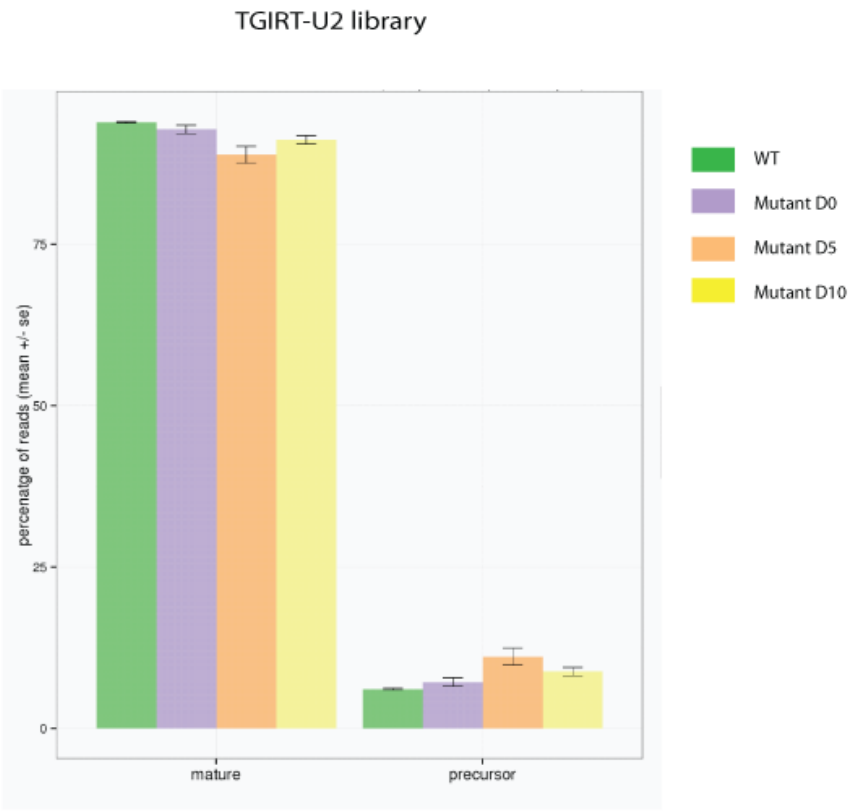
The results were largely consistent with what was observed in Tgs1 mutant MEF cell line, except that the magnitude of difference between wildtype and mutant cell lines was reduced. A possible explanation for the difference in phenotype is that mutant Tgs1 was expressed under control of Tet-inducible promoter in Tgs1* 293 T-REx cell line. In Tgs1 mutant

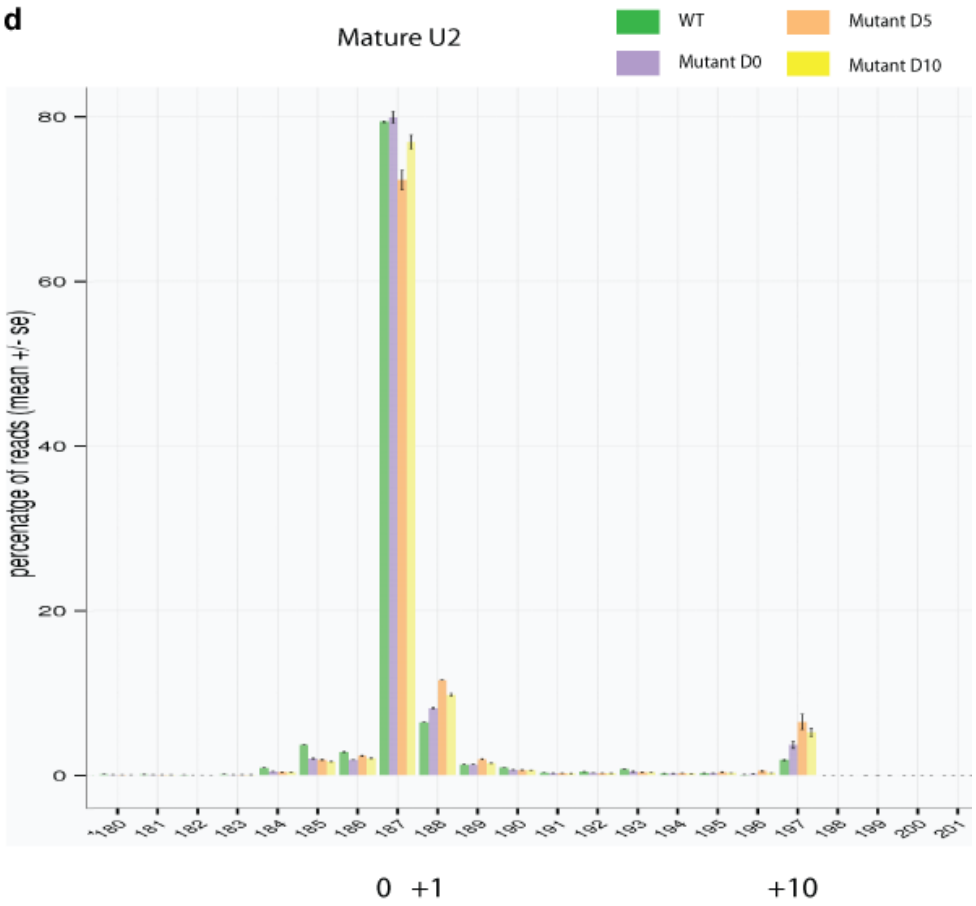
MEF cell line, mutant Tgs1 was under the control of the endogenous promoter. The overexpression of Tgs1* in the context of human cells likely weakened the U2 processing defect.





c


TGIRT-U2 library
percenatge of reads (mean +/- se)
0
25
50
75
mature
precursor
WT
Mutant D0
Mutant D5
Mutant D10


d
Mature U2
WT
Mutant D0
Mutant D5
Mutant D10
percenatge of reads (mean +/- se)
0
20
40
60
80
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
0 +1
+10
3′end distribution of U2

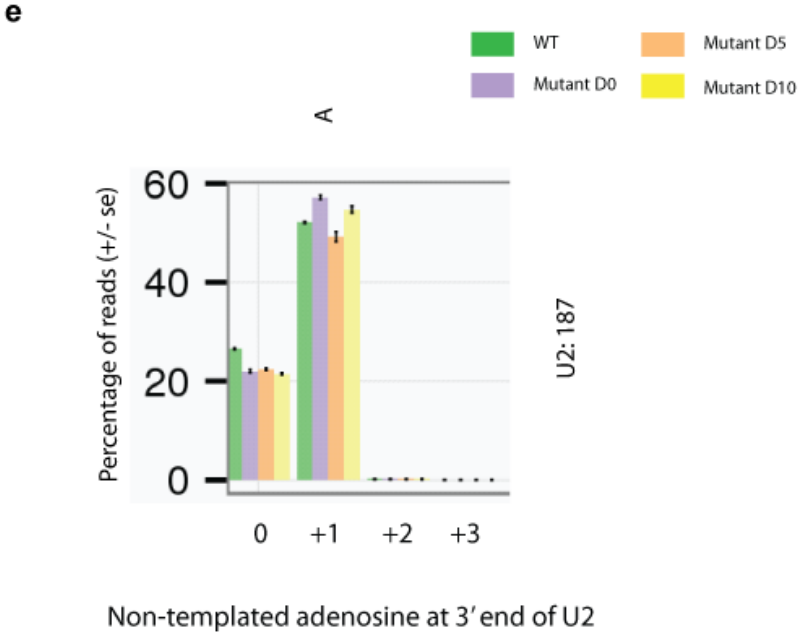

e
WT
Mutant D0
Mutant D5
Mutant D10
A
Percentage of reads (+/- se)
0
20
40
60
0
+1
+2
+3
U2: 187
Non-templated adenosine at 3′end of U2

Figure 3.4.2 Processing defect of U2 snRNA in Tgs1* conditional knockdown cell line.

(a) Northern blot for U2 snRNA in wildtype (WT) and Tgs1* conditional knockdown cell line (Mutant). After tetracycline depletion, total RNA was collected at different time points (5-day, 8-day and 10-day) and analyzed. **(b)** 3' end cloning of U2 snRNA. **(c)** Next-generation sequencing of U2 3' ends. The reads that are mapped to mature U2 + 1nt (indistinguishable on northern blot) or shorter count as "mature", the reads that are mapped beyond full-length mature U2 +1nt count as "precursor". **(d)** 3' end distribution of U2. The mature form of U2 ends at 187nt (position 0). The reads that are mapped to each position of the genomic sequence of U2 are shown in percentages. Reads that are beyond position +10 are collapsed into position +10. **(e)** Non-templated nucleotides at the 3' end of U2. Bottom bar shows length of the non-templated nucleotides. Top bar shows the composition of the non-templated nucleotides (A: adenosine)

3.4.3 Rescue of the Tgs1 conditional knockdown cell line

In both murine and human cell lines, accumulation of snRNA U2 precursors was observed after deletion of Tgs1 N-terminus (murine mutant Tgs1: 57-381aa; human Tgs1*: 55-390aa). The methyltransferase motif at C-terminus (murine Tgs1: 683-835aa; human Tgs1: 691-843aa) was intact in those cell lines. It seems that the N-terminus of Tgs1 has a role in U2 processing independent of the methyltransferase motif.

Alternatively, full-length Tgs1 may be required to recruit the methyltransferase to its substrates. The snRNAs in mouse mutant cells were still TMG-capped, it was likely that the N-terminus of Tgs1 did not affect the methyltransferase activity. But its deletion could still affect the cap hypermethylation of certain RNAs by impairing the recruitment of Tgs1 to these substrates. In this case, Tgs1's role in U2 processing is still dependent on its methyltransferase activity.

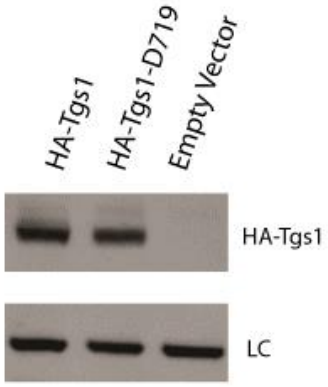
To determine whether the absence of methyltransferase activity, or the absence of Tgs1 protein itself, was responsible for the phenotypes we observed (U2 processing defect, growth defect), a catalytically inactive version of Tgs1 was constructed. By mutating the key amino acid Asp-719 into alanine, the hydrogen bond between the ribose hydroxyls of AdoHcy and the AdoHcy-binding site was disrupted. Tgs1-D719A mutant showed no *in vivo* activity in complementation of the synthetic lethality in *tgs1Δmud2Δ* yeast strain (Hausmann et al., 2008; Monecke et al., 2009). Tgs1-D719A retained the full-length Tgs1 protein sequences without the methyltransferase activity. If absence of the methyltransferase activity is responsible for the growth defect and U2 processing defect, Tgs1-D719A will not be able to rescue those phenotypes. Otherwise, Tgs1 protein itself, independent of the methyltransferase activity, will be able to rescue U2 processing defect and growth defect.

Both wildtype full-length Tgs1 and the Tgs1-D719A mutant were tagged with HA and cloned into pCMV6 vector. After tetracycline depletion, HA-Tgs1 and HA-Tgs1-D719 were transfected into Tgs1* conditional knockdown 293 T-REx cell line. The Tgs1 proteins were expressed at approximately the same level between HA-Tgs1 and HA-Tgs1-D719 (Figure 3.4.3a). Tgs1 D719 was unable to rescue the growth defect after tetracycline depletion. The same number of cells were seeded (2 x 1e4 cells per well, 6-well plate) and counted every other day after transfection and tetracycline depletion. The growth curve of cells transfected with Tgs1-D719 and empty vector largely overlapped (Figure 3.4.3b). Only cells transfected with catalytically active Tgs1 resumed cell growth. It indicated that the catalytically activity of Tgs1 was essential for human cell viability.

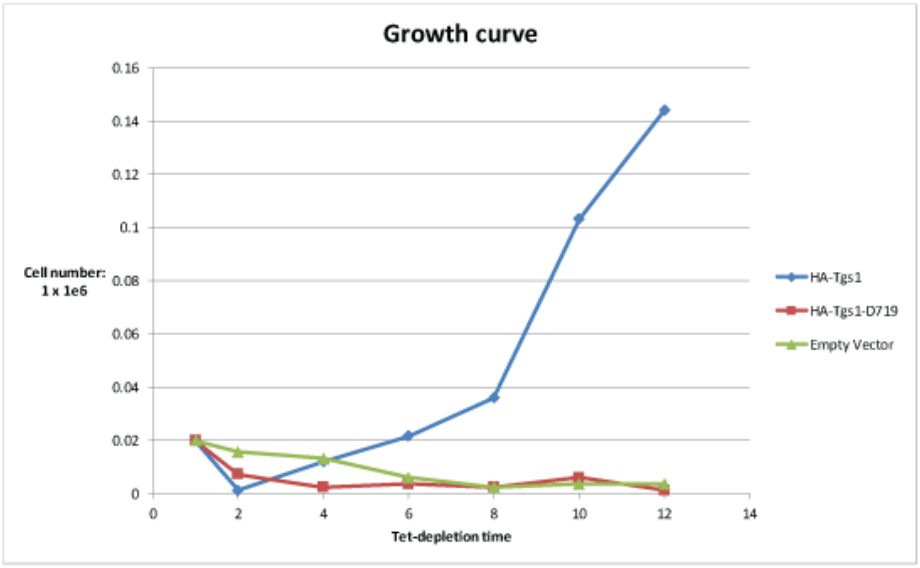
Next, we constructed a next generation sequencing library for human U2 as previously described in 3.2.1. Five days after tetracycline depletion and transfection, total RNAs were extracted from cells transfected with catalytically active Tgs1, Tgs1-D719 and empty vector. cDNA synthesis was performed using thermostable group II intron reverse transcriptase (TGIRT). 5' U2 specific primer and 3' adaptor primers were used to amplify the 3' end of U2. Three biological replicates were analyzed for each sample. Untransfected Tgs1* conditional knockdown cell line in the presence of tetracycline served as control. Control samples contained 6.2% U2 precursors compared with 93.8% mature U2. After tetracycline depletion, cells transfected with empty vector contained 18.4% U2 precursors (standard error of ± 0.65%). Cells transfected with catalytically inactive Tgs1-D719 contained 19.5% U2 precursors (standard error of ± 1.87%). Cells transfected with catalytically active Tgs1 contained 12.5% U2 precursors. Only catalytically active Tgs1 can reduce the accumulation of U2 precursors (Figure 3.4.3c). A detailed 3' end distribution of U2 revealed that the rescue effect resulted from reduced accumulation of the 10nt extended U2 precursors. The 1nt extended U2 precursors were largely unaffected by either catalytically active Tgs1 or Tgs1-D719 (Figure 3.4.3 d). 3'-terminal non-templated nucleotide analysis showed that only catalytically active Tgs1 can partially restore the reduction of the non-templated adenosine (Figure 3.4.3 e).

In conclusion, the methyltransferase activity of Tgs1 is essential for both cell growth and proper U2 3' end processing. Although mammalian Tgs1 was initially identified as transcriptional coactivator and the deletion of Tgs1 N-terminus led to embryonic lethality, loss of its coactivator region in the N-terminus did not affect viability of cell lines.

a

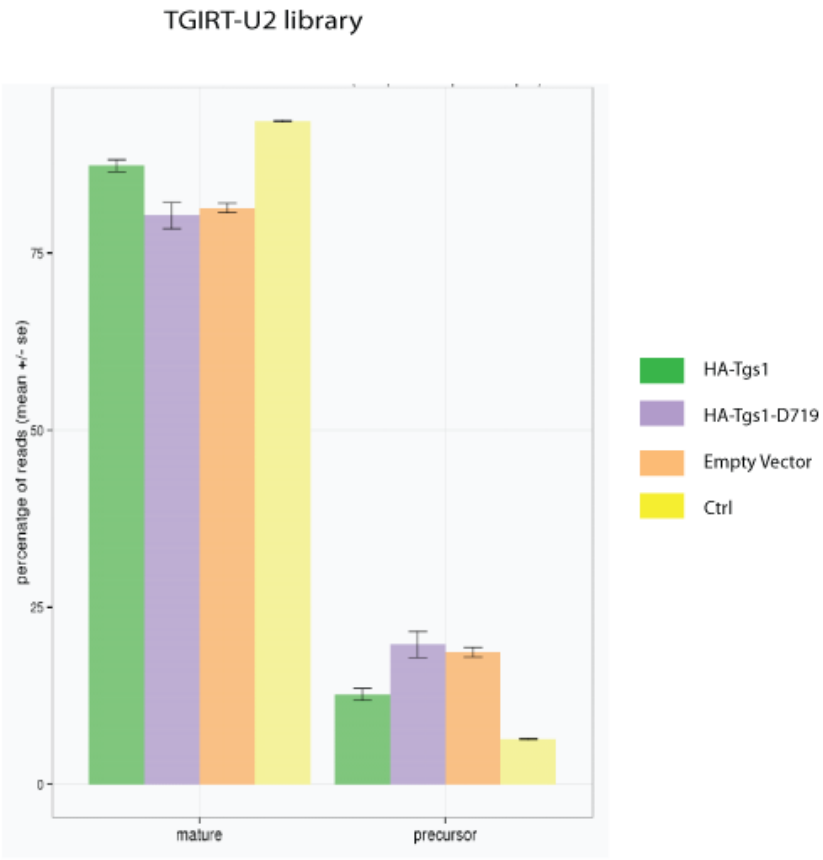


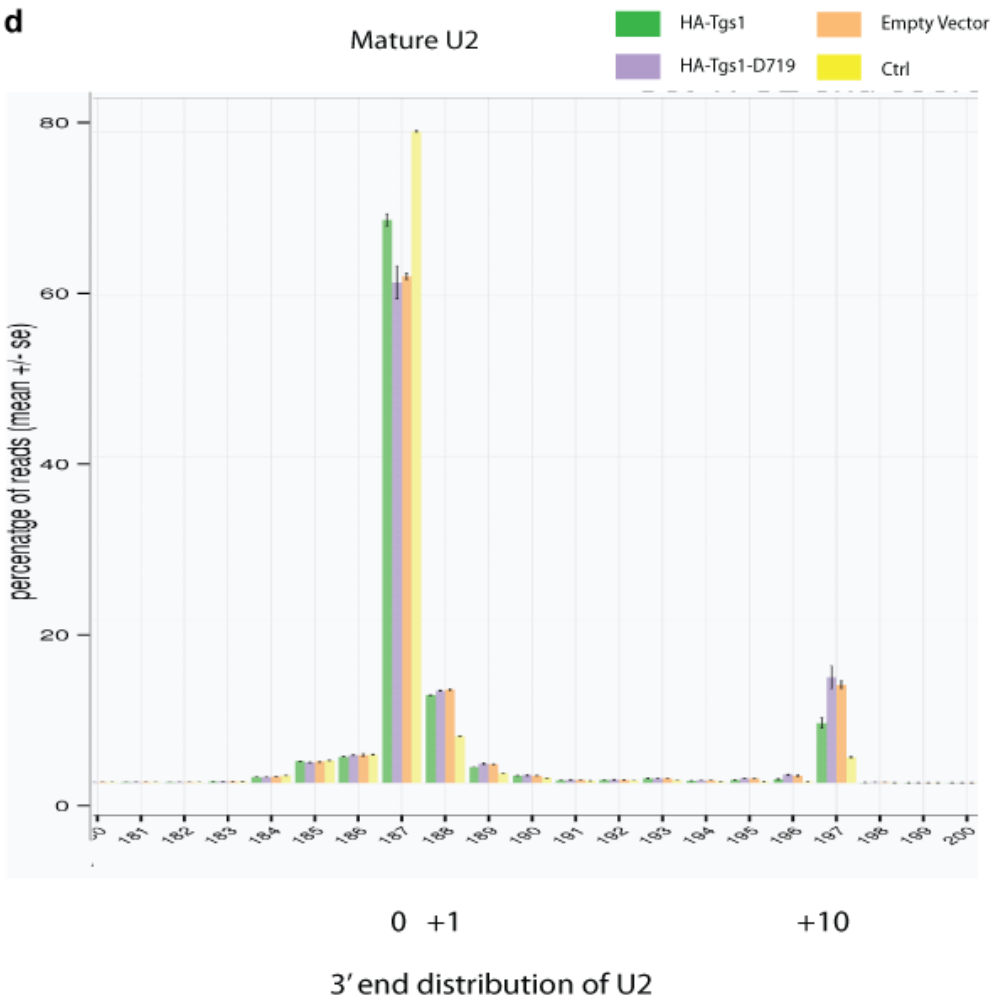
b



Day	1	2	4	6	8	10	12	Backman counter
HA-Tgs1	0.02	0.0012	0.012	0.0216	0.036	0.1032	0.144	1x1e6
HA-Tgs1-D719	0.02	0.0072	0.0024	0.0036	0.0024	0.006	0.0012	
Empty Vector	0.02	0.0156	0.0132	0.006	0.0024	0.0036	0.0036	

c


TGIRT-U2 library
percentage of reads (mean +/- se)
75
50
25
0
mature
precursor
HA-Tgs1
HA-Tgs1-D719
Empty Vector
Ctrl


d
Mature U2
HA-Tgs1
HA-Tgs1-D719
Empty Vector
Ctrl
percenatge of reads (mean +/- se)
0
20
40
60
80
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
0 +1
+10
3' end distribution of U2

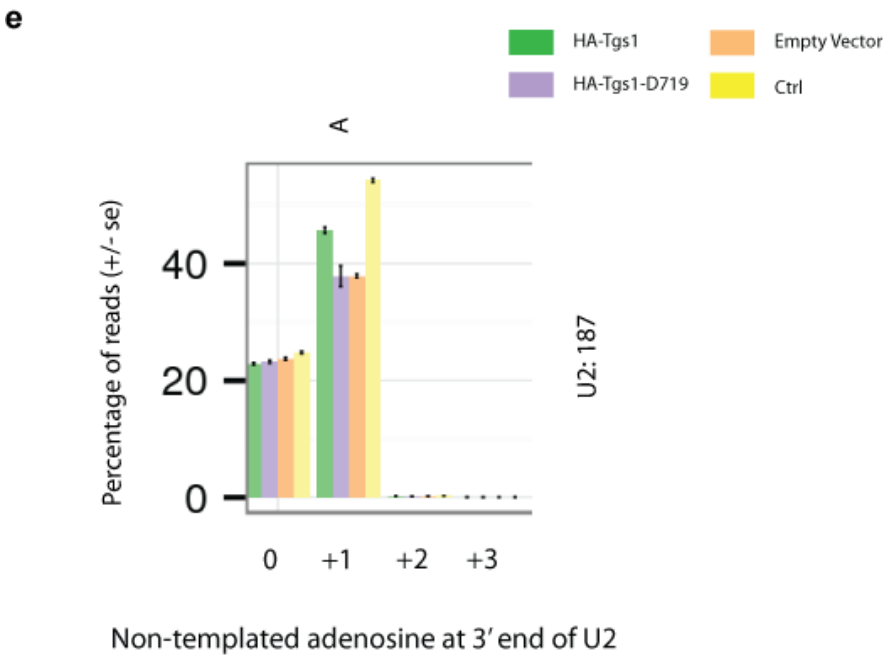


Figure 3.4.3 Rescue U2 processing defect and growth defect in Tgs1* conditional knockdown cell line.

(a) Western blot for HA-Tgs1 and HA-Tgs1-D719. After tetracycline depletion, HA-Tgs1 and HA-Tgs1-D719 were transfected into Tgs1* conditional knockdown 293 T-REx cell line. 4 days after transfection, proteins were collected and separated in 4-12% Bris-Tris NuPage gel. **(b)** Growth curve. Upon depletion of tetracycline, HA-Tgs1 and HA-Tgs1-D719 were transfected into Tgs1* conditional knockdown 293 T-REx cell line. 2 x 1e4 cells per well were seeded in 6-well plate. Each well was counted every other day. **(c)** Next-generation sequencing of U2 3' ends. HA-Tgs1, HA-Tgs1-D719 and empty vector were transfected upon the depletion of tetracycline. Cells that didn't go through tetracycline depletion served as positive control (Ctrl). The reads that are mapped to mature U2 + 1nt (indistinguishable on northern blot) or shorter count as "mature", the reads that are mapped beyond full-length mature U2 +1nt count as "precursor". **(d)** 3' end distribution of U2. The mature form of U2 ends at 187nt (position 0). The reads that are mapped to each position of the genomic sequence of U2 are shown in percentages. Reads that are beyond position +10 are collapsed into position +10. **(e)** Non-templated nucleotides at the 3' end of U2. Bottom bar shows length of the non-templated nucleotides. Top bar shows the composition of the non-templated nucleotides (A: adenosine)

3.5 Effect of shutting down Tgs1 in human cells

Since Tgs1 is essential for human cell viability, it is not feasible to shut it down for a long period of time. We utilized the limited time window when Tgs1 level was down to collect information of its impact on cell growth, RNA processing, gene expression and cap hypermethylation.

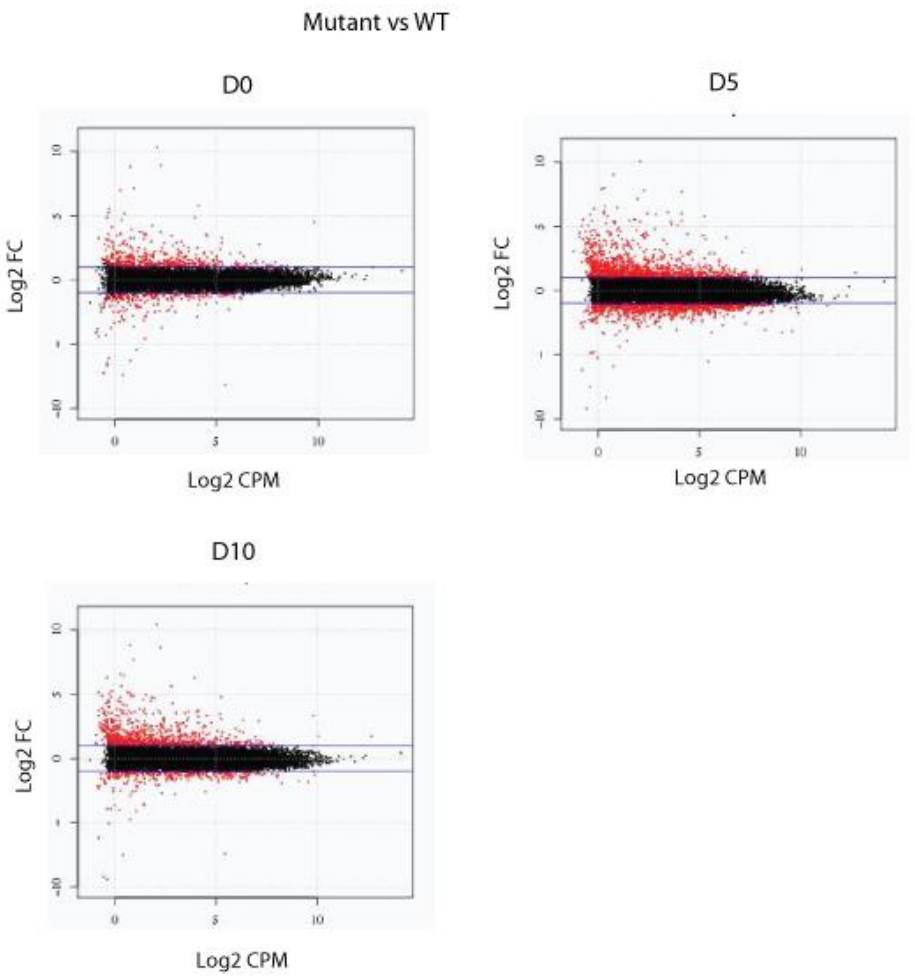
The methyltransferase activity of Tgs1 is responsible for hypermethylation of 5' end cap of RNAs, which affects RNA stability and expression. In fission yeast, TMG-cap is required for telomerase RNA stability. Deletion of Tgs1 led to loss of TMG cap on telomerase RNA, which reduced telomerase RNA level and impaired telomere maintenance. In mammals, TMG cap promotes the expression of Rev-dependent HIV-1 RNAs (Yedavalli & Jeang, 2010). Shutting down Tgs1 in Tgs1* conditional knockdown cell line will likely result in changes in steady-state levels of several RNAs.

3.5.1 Gene expression profile after shutting down Tgs1

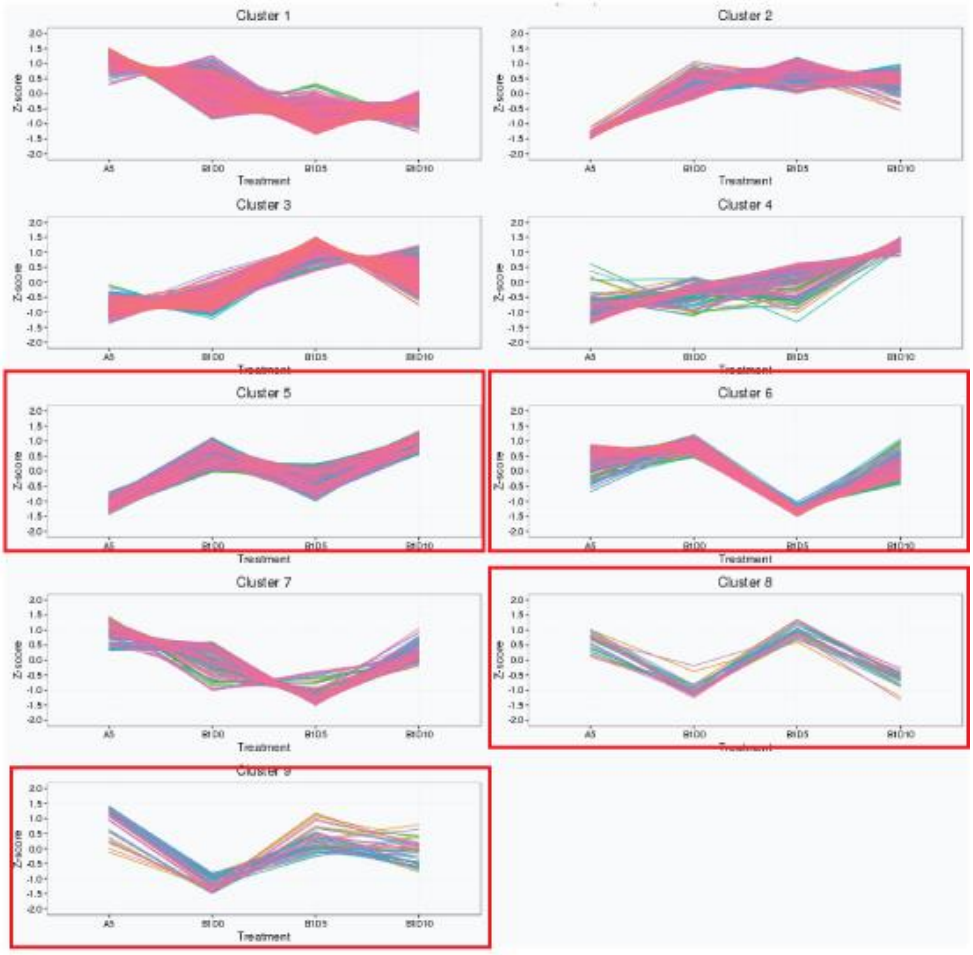
To examine the effect of shutting down Tgs1 on gene expression, total RNAs were extracted and subjected to next generation sequencing after 5 days (D5) and 10 days (D10) of tetracycline depletion. Wildtype cell line (WT) and Tgs1* conditional knockdown cell line (D0) in the presence of tetracycline served as control. Each sample contained three biological replicates. 50~66 million 50-base-pair (bp) single-read reads were generated for each sample. TopHat was used to align the reads against hg19 human genome. EdgeR was used to analyze differential gene expression. By the threshold of 2 folds change (FC) and the adjusted p-value of 0.01 or less, 679 genes were dysregulated on D0. The number increased to 1921 genes on D5 and decreased to 804 genes on D10 (Figure 3.5.1a). The patterns of the dysregulated genes during the

time course were divided into 9 clusters. Z-scores for each gene were calculated from the sample means and represented the distance between the raw score and the population mean. Euclidean distance between genes were calculated based on z-scores to create the hierarchical clustering. In cluster 5 and 6, genes that were down regulated on D5 recovered on D10. In cluster 8 and cluster 9, genes that were up regulated on D5 recovered on D10 (Figure 3.5.1b). Those 4 clusters showed that changes in expression on D5 went away on D10, corresponding to the absence and resurface of Tgs1 protein on D5 and D10 (Figure 3.4.1f). Among the most dysregulated genes after shutting down Tgs1, majority of them were protein-coding genes (Figure 3.5.1c, d, e). As shown in previous rescue experiment, the absence of methyltransferase activity was underlying the growth defect and U2 processing defect. Noncoding RNAs which contained TMG caps were expected to be directly affected. In mammals, TMG-cap structure is rarely present on mRNAs, which are transcribed from protein coding genes. The only known mRNAs that contain TMG-caps are selenoprotein mRNAs (Wurth et al., 2014). Differential expression of selenoprotein mRNAs did not meet the 2 fold threshold in our analysis. Most of them were slightly decreased by 20~30%. Multiple genes involved in tumorigenesis, cell adhesion, cytoplasmic signaling and metabolism showed changes in expression after shutting down Tgs1, but it was not clear what led to the growth defect. Nor was it clear what resulted in the U2 processing defect.

a



b



c

Mutant vs WT D0

Downregulated

name	biotype	log2FC
HTRA3	protein_coding	-9.399196
LCP1	protein_coding	-7.947139
GNAS-AS1	antisense	-7.102692
ZIC1	protein_coding	-6.439875
GUCY1B2	unitary_pseudogene	-4.681315
EPHA6	protein_coding	-4.116549
ZNF826P	transcribed_unprocessed_pseudogene	-3.990172
C2CD2	protein_coding	-3.888953
LDOC1	protein_coding	-3.77832
TSPY26P	transcribed_processed_pseudogene	-3.665434

Upregulated

name	biotype	log2FC
TNK1	protein_coding	6.609566
CEBPE	protein_coding	6.136509
SALL3	protein_coding	5.190005
KC6	lincRNA	4.925805
BNC1	protein_coding	4.301031
TGS1	protein_coding	4.193471
SPAG6	protein_coding	4.16015
CYP7B1	protein_coding	4.096729
RP11-384F7.2	lincRNA	3.917784
ASB9	protein_coding	3.785612

d

Mutant vs WT D5

Downregulated

name	biotype	log2FC
GNAS-AS1	antisense	-8.11376
HTRA3	protein_coding	-6.37287
LCP1	protein_coding	-5.73207
RP11-109M17.2	lincRNA	-5.15454
ZIC1	protein_coding	-4.10617
ETNPPL	protein_coding	-3.95052
GUCY1B2	unitary_pseudogene	-3.76584
MKX	protein_coding	-3.64063
RP6-24A23.8	TEC	-3.54331
C2CD2	protein_coding	-3.54222

Upregulated

name	biotype	log2FC
GZMB	protein_coding	8.40016
ANKRD1	protein_coding	7.560925
MMP10	protein_coding	6.822784
TNK1	protein_coding	6.638656
GS1-600G8.5	lincRNA	6.464797
PTPRN	protein_coding	6.113548
SALL3	protein_coding	6.071262
CEBPE	protein_coding	5.809776
ACTBL2	protein_coding	5.675735
KRTAP21-2	protein_coding	5.58423

e

Mutant vs WT D10

Downregulated

name	biotype	log2FC
GNAS-AS1	antisense	-8.0052
LCP1	protein_coding	-7.65776
HTRA3	protein_coding	-6.23978
ZIC1	protein_coding	-3.95895
RP11-109M17.2	lincRNA	-3.93331
ZNF826P	transcribed_unprocessed_pseudogene	-3.77713
GUCY1B2	unitary_pseudogene	-3.72178
MKX	protein_coding	-3.65866
C2CD2	protein_coding	-3.64607

Upregulated

name	biotype	log2FC
TNK1	protein_coding	6.616277
SALL3	protein_coding	5.638745
KC6	lincRNA	5.063124
LSAMP	protein_coding	4.930429
MMP10	protein_coding	4.918132
GZMB	protein_coding	4.884492
RP11-384F7.2	lincRNA	4.712388
RP11-449D8.1	lincRNA	4.578169
PAPPA	protein_coding	4.479998

Figure 3.5.1 The effect of shutting down Tgs1* on gene expression.

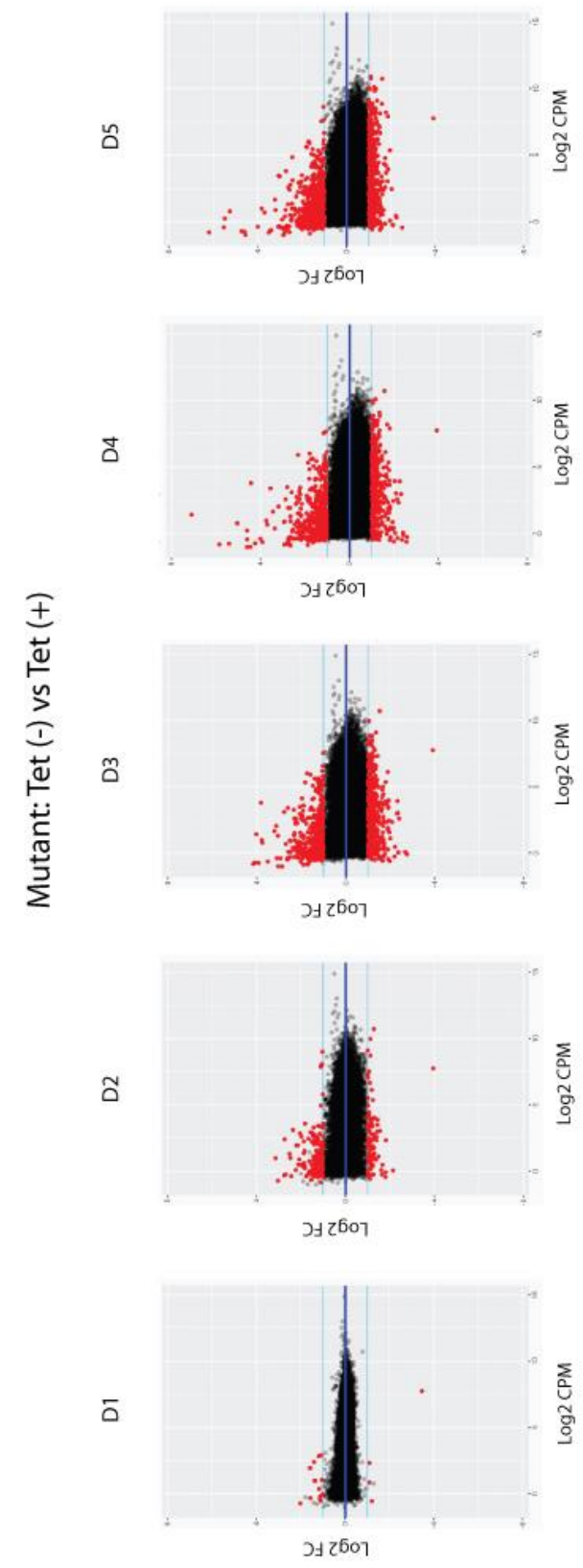
(a) MA plot of differentially expressed genes in Tgs1* 293 T-REx conditional knockdown cell line (Mutant) vs wildtype cell line (WT). FC: fold change. Red dots are genes with > 2 folds differential expression (Log_2FC >1 or <-1) in Mutant vs WT. Tetracycline was depleted for 5 days (D5) and 10 days (D10). **(b)** Cluster analysis of differentially expressed genes in Mutant vs WT. A5: WT; B1: Mutant. Z-scores for each gene calculated from the sample means. Clustering was calculated using Euclidean distance between genes. Gene clusters in which majority of gene expression changes seen on D5 went away when Tgs1* came back on D10 are underscored by red box. **(c)** Top 10 genes with differential expression in Mutant vs WT, D0. **(d)** Top 10 genes with differential expression in Mutant vs WT, D5. **(e)** Top 10 genes with differential expression in Mutant vs WT, D10.

3.5.2 Refined time course on gene expression

Other than the direct effects of the absence of methyltransferase and subsequent loss of TMG caps, secondary effects of severe growth defect could contribute to the changes in gene expression after shutting down Tgs1. It was difficult to distinguish genes directly affected by the absence of Tgs1 from genes affected by secondary effects of DNA damage response, apoptosis, and/or cell cycle arrest. Besides, the U2 processing defect could affect splicing, adding to the indirect effects of shutting down Tgs1. To address this problem, a refined time course was created to track gene expression in early stages of tetracycline depletion so that we can capture the chronological order of the changes in gene expression.

After removing tetracycline from the media, cells were collected every day for 5 consecutive days. At each time point, three biological replicates were included. Total RNAs were extracted and subjected to next generation sequencing. 52~65 million 50-base-pair (bp) single-read reads were generated for each sample. The reads were mapped to human genome GRCh38.81 reference genome by TopHat. EdgeR was used to analyze differential gene expression (Figure 3.5.2a). By the threshold of fold change (FC) of two, 22 genes were categorized as dysregulated on D1. Among the 22 genes, 18 were upregulated, 4 were downregulated. The number of dysregulated genes increased to 247 on D2, including 154 upregulated genes and 93 downregulated genes. On D3, 831 genes were dysregulated, including 416 upregulated genes and 415 down regulated genes. On D4, 880 genes were dysregulated, including 416 upregulated genes and 415 down regulated genes. On D5, the number of dysregulated gene increased to 1007, including 614 upregulated genes and 393 downregulated genes.

Among the dysregulated genes, the downregulation of Tgs1 was consistent from D1 to D5 (Figure 3.5.2b). A sharp increase of dysregulated genes was present between D1 and D2. Majority of them were protein-coding genes and unannotated noncoding RNAs. Some notable examples included TSPAN2, which is reportedly involved in cell motility and invasion during lung cancer progression (Otsubo et al., 2014); YPEL4, which was involved in adrenal cell proliferation (Oki, Plonczynski, Gomez-Sanchez, & Gomez-Sanchez, 2016); and INHBE, the inhibin beta E subunit, which inhibited growth of pancreatic exocrine cells (Hashimoto et al., 2006). The dysregulated genes through the refined time course were divided into 6 clusters (Figure 3.5.2c). Z-scores for each gene were calculated from the sample means. The clustering was based on Euclidean distance between genes calculated via z-scores. Genes in cluster 1 showed consistent decreases in expression from D1 to D5. Notable cluster 1 genes included Tgs1, snoRNA SNORD46 and SNORD83A. Genes in cluster 2 showed consistent increases in expression from D1 to D5. Surprisingly, several snRNAs and snoRNAs including U1 and U3 were present in this cluster. Given the role of TMG-cap in snRNA trafficking, the increase in snRNA expression after shutting down Tgs1 was not expected.


a
Mutant: Tet (-) vs Tet (+)
D1
D2
D3
D4
D5
Log2 FC
Log2 CPM

b

Mutant: Tet (-) vs Tet (+)

D1

Upregulated

Gene_name	D1_log2FC	gene_biotype	description
YPEL4	2.018219	protein_coding	yippee-like 4 (Drosophila) [Source:HGNC
FP236383.12	1.788435	miRNA	
FP236383.9	1.588767	miRNA	
RP11-666A8.9	1.567911	antisense	
FP671120.6	1.557266	miRNA	
MIR3687-1	1.417242	miRNA	microRNA 3687-1 [Source:HGNC Symbol;Acc:HGNC:38946]
INHBE	1.40554	protein_coding	inhibin, beta E [Source:HGNC Symbol;Acc:HGNC:24029]
RP11-282I1.1	1.383549	lincRNA	
UNC5CL	1.347275	protein_coding	unc-5 homolog C (C. elegans)-like [Source:HGNC
CLCN4	1.274072	protein_coding	chloride channel, voltage-sensitive 4 [Source:HGNC

Downregulated

Gene_name	D1_log2FC	gene_biotype	description
TGS1	-3.46397	protein_coding	Symbol;Acc:HGNC:17843]
KLHL4	-1.28509	protein_coding	Symbol;Acc:HGNC:6355]
CNN1	-1.20533	protein_coding	Symbol;Acc:HGNC:2155]
KRCC1	-1.09797	protein_coding	lysine-rich coiled-coil 1 [Source:HGNC Symbol;Acc:HGNC:28039]
PAQR8	-1.09292	protein_coding	Symbol;Acc:HGNC:15708]
TSPAN2	-0.97084	protein_coding	tetraspanin 2 [Source:HGNC Symbol;Acc:HGNC:20659]
RP5-940J5.9	-0.94328	antisense	
COL4A3	-0.85355	protein_coding	Symbol;Acc:HGNC:2204]
UBC	-0.78377	protein_coding	ubiquitin C [Source:HGNC Symbol;Acc:HGNC:12468]
MAP2	-0.76607	protein_coding	Symbol;Acc:HGNC:6839]

D2

Upregulated

Gene_name	D1_log2FC	gene_biotype	description
RP11-282I1.1	3.142226	lincRNA	
YPEL4	3.035432	protein_coding	yippee-like 4 (Drosophila) [Source:HGNC Symbol;Acc:HGNC:18328]
RP11-666A8.9	2.756089	antisense	
INHBE	2.729861	protein_coding	inhibin, beta E [Source:HGNC Symbol;Acc:HGNC:24029]
FP236383.12	2.683173	miRNA	
NPR3	2.509223	protein_coding	natriuretic peptide receptor 3 [Source:HGNC Symbol;Acc:HGNC:7945]
UNC5CL	2.475964	protein_coding	Symbol;Acc:HGNC:21203]
FGF21	2.431748	protein_coding	fibroblast growth factor 21 [Source:HGNC Symbol;Acc:HGNC:3678]
OVGP1	2.36134	protein_coding	oviductal glycoprotein 1, 120kDa [Source:HGNC Symbol;Acc:HGNC:8524]
RP11-372K14.2	2.228595	antisense	

Downregulated

Gene_name	D1_log2FC	gene_biotype	description
TGS1	-3.95187	protein_coding	Symbol;Acc:HGNC:17843]
GNG7	-2.14515	protein_coding	[Source:HGNC Symbol;Acc:HGNC:4410]
KRCC1	-1.81929	protein_coding	lysine-rich coiled-coil 1 [Source:HGNC Symbol;Acc:HGNC:28039]
C16orf74	-1.8075	protein_coding	Symbol;Acc:HGNC:23362]
PTPRZ1	-1.80237	protein_coding	[Source:HGNC Symbol;Acc:HGNC:9685]
TSPAN2	-1.67483	protein_coding	tetraspanin 2 [Source:HGNC Symbol;Acc:HGNC:20659]
ZNF837	-1.63923	protein_coding	zinc finger protein 837 [Source:HGNC Symbol;Acc:HGNC:25164]
JAKMIP1	-1.61246	protein_coding	Symbol;Acc:HGNC:26460]
ITFG3	-1.55229	protein_coding	Symbol;Acc:HGNC:14163]
COL4A3	-1.4904	protein_coding	Symbol;Acc:HGNC:2204]

D3

Upregulated

Gene_name	D1_log2FC	gene_biotype	description
GS1-600G8.5	4.165481	lincRNA	
YPEL4	4.084816	protein_coding	yippee-like 4 (Drosophila) [Source:HGNC Symbol;Acc:HGNC:18328]
KRTAP19-1	4.028963	protein_coding	keratin associated protein 19-1 [Source:HGNC Symbol;Acc:HGNC:18936]
CLCN4	3.917474	protein_coding	Symbol;Acc:HGNC:2022]
ANKRD1	3.812477	protein_coding	Symbol;Acc:HGNC:15819]
RP11-282I1.1	3.780691	lincRNA	
KLHL4	3.756	protein_coding	kelch-like family member 4 [Source:HGNC Symbol;Acc:HGNC:6355]
BMP2	3.507908	protein_coding	bone morphogenetic protein 2 [Source:HGNC Symbol;Acc:HGNC:1069]
INHBA	3.212776	protein_coding	inhibin, beta A [Source:HGNC Symbol;Acc:HGNC:6066]
CPA4	3.127768	protein_coding	carboxypeptidase A4 [Source:HGNC Symbol;Acc:HGNC:15740]

Downregulated

Gene_name	D1_log2FC	gene_biotype	description
TGS1	-3.92103	protein_coding	Symbol;Acc:HGNC:17843]
C16orf74	-2.7541	protein_coding	Symbol;Acc:HGNC:23362]
JAKMIP1	-2.67108	protein_coding	Symbol;Acc:HGNC:26460]
CTNNBIP1	-2.4056	protein_coding	Symbol;Acc:HGNC:16913]
GNG7	-2.36575	protein_coding	[Source:HGNC Symbol;Acc:HGNC:4410]
ITFG3	-2.34446	protein_coding	Symbol;Acc:HGNC:14163]
LPCAT3	-2.32651	protein_coding	Symbol;Acc:HGNC:30244]
STOX1	-2.31318	protein_coding	storkhead box 1 [Source:HGNC Symbol;Acc:HGNC:23508]
TSPAN2	-2.22043	protein_coding	tetraspanin 2 [Source:HGNC Symbol;Acc:HGNC:20659]
ZNF837	-2.19763	protein_coding	zinc finger protein 837 [Source:HGNC Symbol;Acc:HGNC:25164]

D4

Upregulated

Gene_name	D1_log2FC	gene_biotype	description
KRTAP19-1	4.028963	protein_coding	keratin associated protein 19-1 [Source:HGNC Symbol;Acc:HGNC:18936]
GS1-600G8.5	4.165481	lincRNA	
CLCN4	3.917474	protein_coding	Symbol;Acc:HGNC:2022]
KLHL4	3.756	protein_coding	kelch-like family member 4 [Source:HGNC Symbol;Acc:HGNC:6355]
NPPB	2.86258	protein_coding	natriuretic peptide B [Source:HGNC Symbol;Acc:HGNC:7940]
KRTAP19-5	1.539339	protein_coding	keratin associated protein 19-5 [Source:HGNC Symbol;Acc:HGNC:18940]
KRTAP21-2	0.831983	protein_coding	keratin associated protein 21-2 [Source:HGNC Symbol;Acc:HGNC:18946]
CDH11	3.040532	protein_coding	Symbol;Acc:HGNC:1750]
ANKRD1	3.812477	protein_coding	Symbol;Acc:HGNC:15819]
YPEL4	4.084816	protein_coding	yippee-like 4 (Drosophila) [Source:HGNC Symbol;Acc:HGNC:18328]

Downregulated

Gene_name	D1_log2FC	gene_biotype	description
TGS1	-3.93486	protein_coding	Symbol;Acc:HGNC:17843]
ZNF837	-2.59989	protein_coding	zinc finger protein 837 [Source:HGNC Symbol;Acc:HGNC:25164]
CLYBL	-2.56874	protein_coding	citrate lyase beta like [Source:HGNC Symbol;Acc:HGNC:18355]
C16orf74	-2.42753	protein_coding	Symbol;Acc:HGNC:23362]
JAKMIP1	-2.41947	protein_coding	Symbol;Acc:HGNC:26460]
PTPRZ1	-2.3671	protein_coding	[Source:HGNC Symbol;Acc:HGNC:9685]
STOX1	-2.3309	protein_coding	storkhead box 1 [Source:HGNC Symbol;Acc:HGNC:23508]
LPCAT3	-2.26891	protein_coding	Symbol;Acc:HGNC:30244]
GNAZ	-2.20356	protein_coding	polypeptide [Source:HGNC Symbol;Acc:HGNC:4395]
LRRC29	-2.1656	protein_coding	Symbol;Acc:HGNC:13605]

D5

Upregulated

Gene_name	D1_log2FC	gene_biotype	description
KRTAP19-1	8.553757	protein_coding	keratin associated protein 19-1 [Source:HGNC Symbol;Acc:HGNC:18936]
GS1-600G8.5	6.195762	lincRNA	
KRTAP21-2	5.526837	protein_coding	keratin associated protein 21-2 [Source:HGNC Symbol;Acc:HGNC:18946]
KRTAP19-5	5.486656	protein_coding	keratin associated protein 19-5 [Source:HGNC Symbol;Acc:HGNC:18940]
KLHL4	5.258377	protein_coding	kelch-like family member 4 [Source:HGNC Symbol;Acc:HGNC:6355]
YPEL4	4.729933	protein_coding	yippee-like 4 (Drosophila) [Source:HGNC Symbol;Acc:HGNC:18328]
CLCN4	4.59727	protein_coding	Symbol;Acc:HGNC:2022]
CDH11	4.551948	protein_coding	Symbol;Acc:HGNC:1750]
MMP16	4.003581	protein_coding	Symbol;Acc:HGNC:7162]
RP11-282I1.1	3.831962	lincRNA	

Downregulated

Gene_name	D1_log2FC	gene_biotype	description
TGS1	-3.92715	protein_coding	Symbol;Acc:HGNC:17843]
PTPRZ1	-2.51148	protein_coding	[Source:HGNC Symbol;Acc:HGNC:9685]
RP5-940J5.9	-2.25636	antisense	
ACTL8	-2.22963	protein_coding	actin-like 8 [Source:HGNC Symbol;Acc:HGNC:24018]
RP11-250B2.6	-2.10417	lincRNA	
ZNF837	-2.0834	protein_coding	zinc finger protein 837 [Source:HGNC Symbol;Acc:HGNC:25164]
LYPD6	-2.04028	protein_coding	Symbol;Acc:HGNC:28751]
JAKMIP1	-1.96732	protein_coding	Symbol;Acc:HGNC:26460]
FLOT2	-1.96202	protein_coding	flotillin 2 [Source:HGNC Symbol;Acc:HGNC:3758]
LPCAT3	-1.88958	protein_coding	Symbol;Acc:HGNC:30244]

c

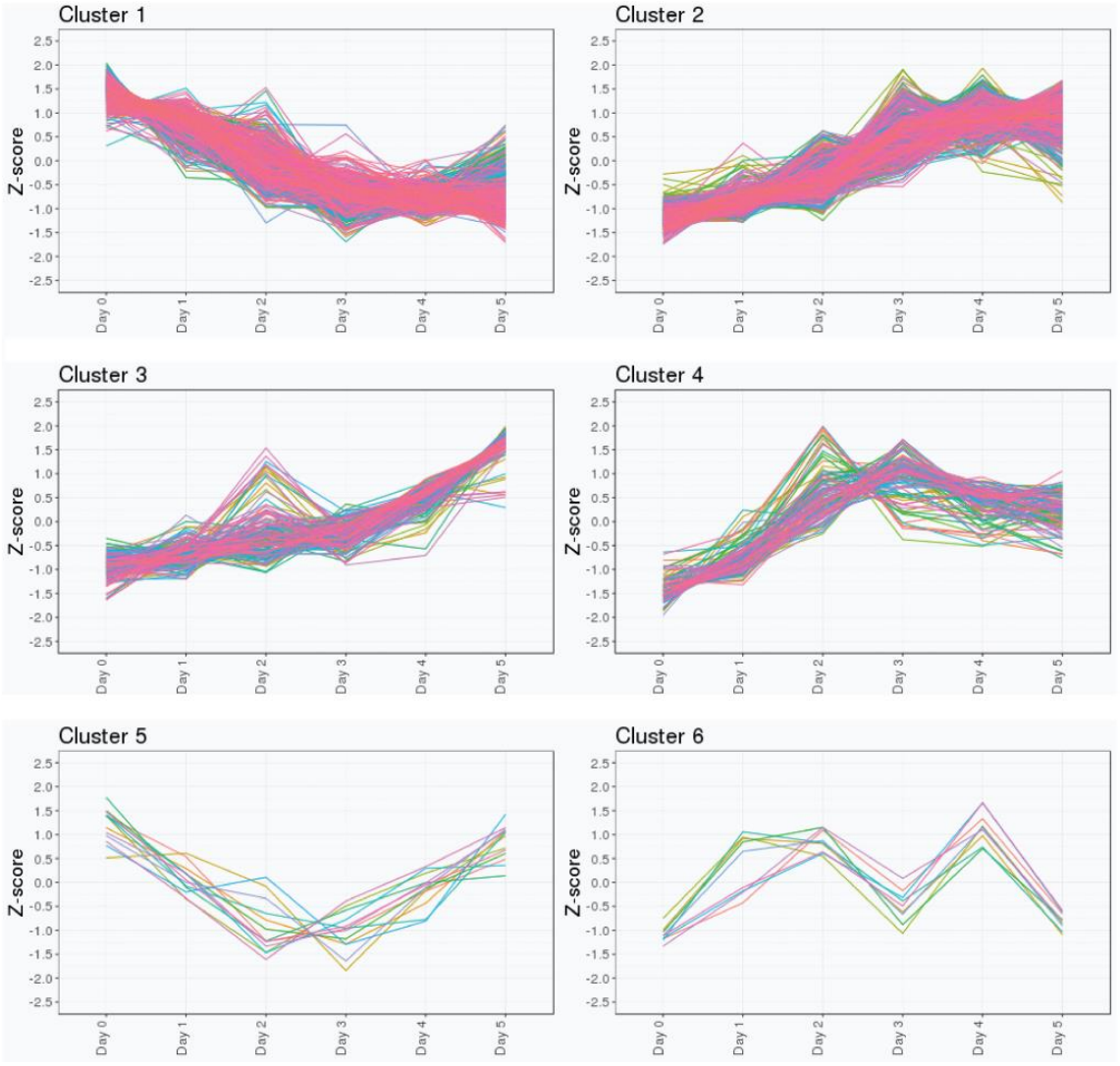


Figure 3.5.2 Refined time course for effect of shutting down Tgs1* on gene expression.

(a) MA plot of differentially expressed genes in Tgs1* 293 T-REx conditional knockdown cell line in the absence of tetracycline (Mutant: Tet (-)) vs in the presence of tetracycline (Mutant: Tet (+)). Tetracycline was depleted for 5 days. Total RNAs were collected every day through D1 to D5. FC: fold change. Red dots are genes with > 2 folds differential expression (Log_2FC >1 or <-1) in Tet (-) vs Tet (+). **(b)** Top 10 genes with differential expression in Tet (-) vs Tet (+), from D1 to D5. **(c)** Cluster analysis of differentially expressed genes in Tet (-) vs Tet (+), from D1 to

D5. Z-scores for each gene calculated from the sample means. Clustering was calculated using Euclidean distance between genes.

3.5.3 DNA damage response and cell cycle arrest

The refined time course helped us narrow down a scope of genes that may be involved in the growth defect. It was still not clear what mechanism was involved. Selenoprotein mRNAs are the only mammalian mRNAs known to carry TMG caps. It has been reported that selenoproteins are involved in protection against oxidative stress and DNA damage response (Dai, Liu, Zhou, & Huang, 2016; Jerome-Morais, Bera, Rachidi, Gann, & Diamond, 2013; Steinbrenner, Bilgic, Alili, Sies, & Brenneisen, 2006; Yant et al., 2003). Oxidative stress induces the formation of double strand breaks, which trigger phosphorylation of the histone variant H2AX, resulting in γH2AX. The level of γH2AX was examined after shutting down Tgs1. Accumulation of γH2AX was observed on D3 and the accumulation increased consistently through D4 and D5 (Figure 3.5.3a). It seemed that DNA damage response was increased as Tgs1 was depleted. However, among 25 known mammalian selenoprotein mRNAs, 12 were present in our NGS data, most of which showed no change or 10~20% decrease in expression level. It was not clear whether the mild decrease in selenoprotein mRNA expression led to the increased DNA damage response through impaired protection against oxidative stress.

It has also been reported that deletion of Tgs1 in budding yeast affects the splicing of PCH2, a meiosis regulon (Qiu et al., 2011). TRIP13, the mammalian ortholog of PCH2, is critical to mitotic exit (H. T. Ma & Poon, 2018; Marks et al., 2017; Tao et al., 2017; Vader, 2015). To test whether cell cycle was affected after shutting down Tgs1, DNA contents of the cells were detected by flow cytometry. Additional peaks of tetraploid cells in G_2/M phase were

observed on D3, D4 and D5 (Figure 3.5.3b), indicating the presence of G_2/M phase cell cycle arrest. In our NGS data, TRIP13 expression showed ~30% reduction on D3, which could lead to the G2 arrest. However, phosphorylation of H2AX also plays an important role in activation of checkpoints proteins that arrest the cell cycle. The exact cause of the cell cycle arrest is yet to be determined.

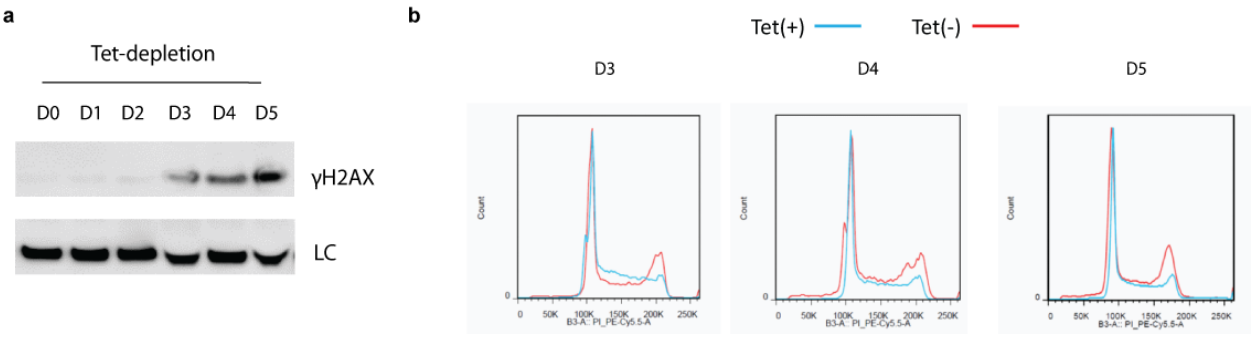


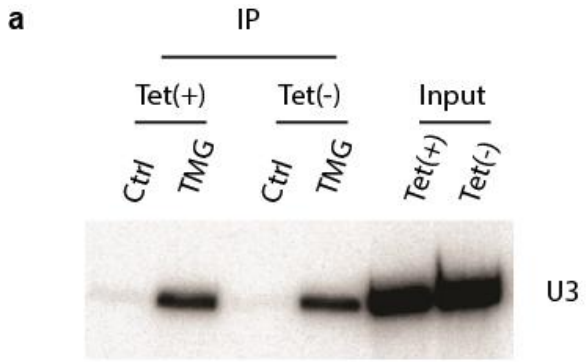
Figure 3.5.3 The effect of shutting down Tgs1* on DNA damage response and DNA content.

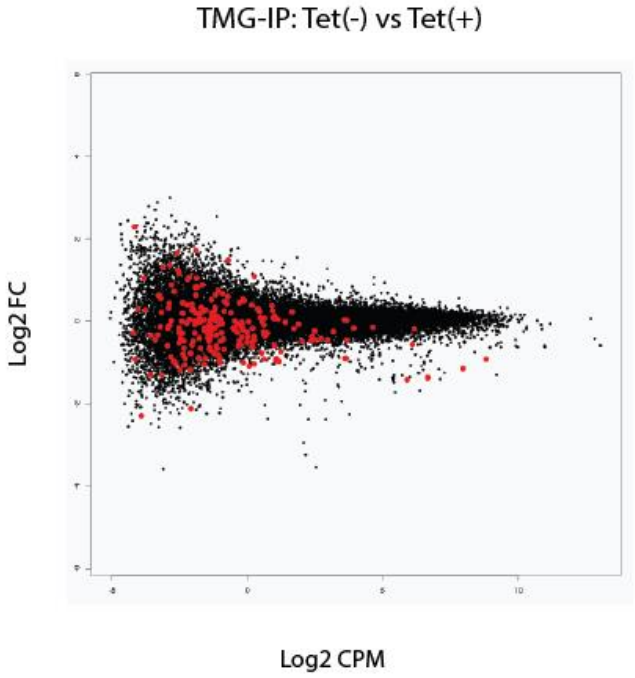
(a) Western blot of γH2AX after tetracycline depletion. Protein extracts were collected every day after tetracycline depletion from D0 to D5. γH2AX was separated on Bis-tris 4-12% NuPage gel. Alpha-tubulin served as loading control (LC). **(b)** DNA content measured by flow cytometry after tetracycline depletion. One well of cells from 6-well plate were digested on D3, D4, and D5 after tetracycline depletion. 0.5 x 1e6 cells were stained with 24µg/ml propidium iodide (PI) and ran through cytometer configured for 488nm excitation.

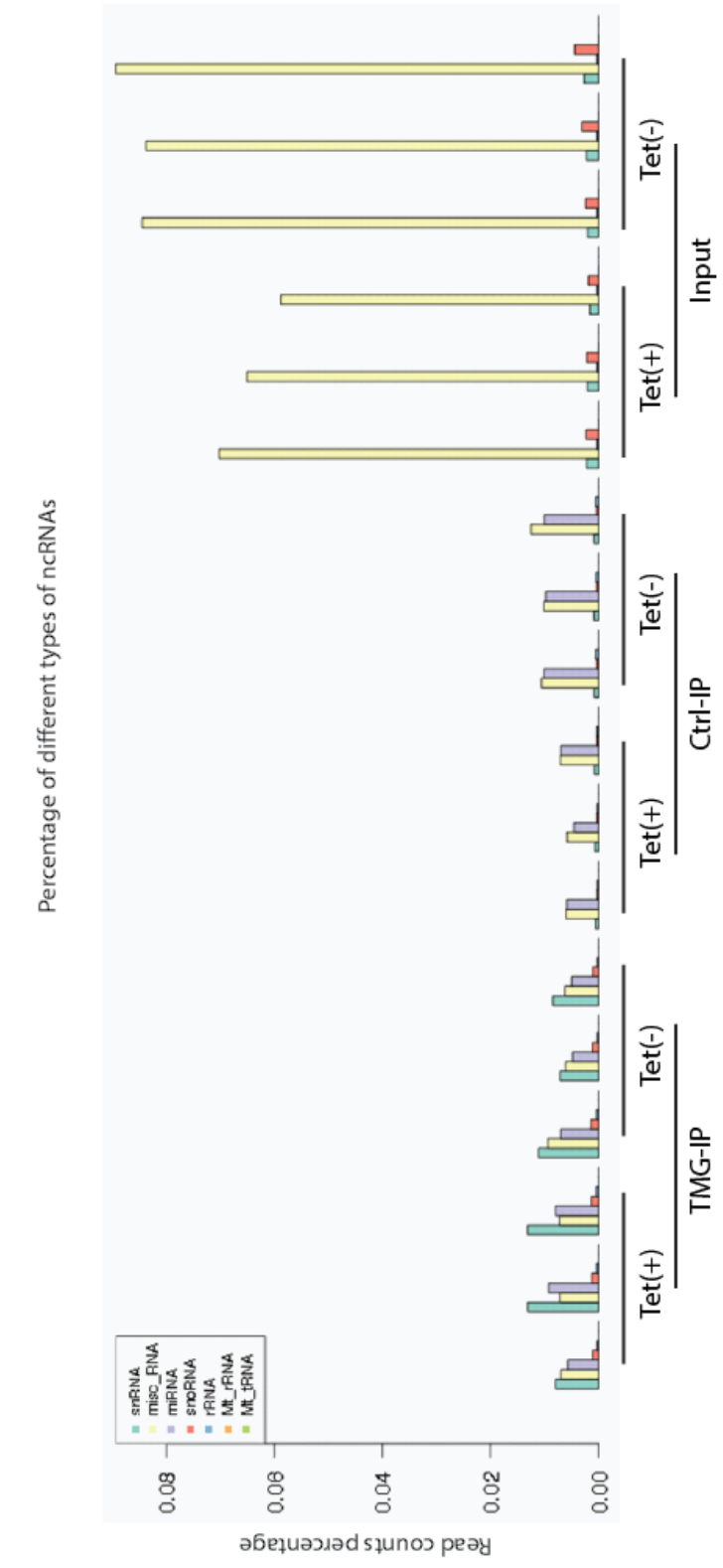
3.5.4 Cap hypermethylation

Even though we were unable to permanently deplete Tgs1, TMG-capped RNAs with short half-lives can still be impacted during the 5-day time window when Tgs1 was absent. TMG-IP was performed with RNAs from cells before (D0) and after Tgs1 shutdown (D5). At

each time point, three biological replicates were collected. snoRNA U3 was used as a marker for the TMG-IP efficiency. The K121 anti-TMG antibodies specifically pulled down U3 (Figure 3.5.4a), although the IP efficiency was similar between D0 and D5, indicating U3 didn't lose its TMG cap within 5 days. Libraries were made from IP samples and subjected to next-generation sequencing. 50-80 million 50-base-pair (bp) single read reads were generated. The reads were mapped to human genome GRCm38.78 by TopHat. The differential enrichment in TMG-IP was analyzed by EdgeR. By the threshold of fold change (FC) of two and the adjusted p-value of 0.01 or less, 538 genes showed an increase of enrichment in TMG-IP, 462 genes showed a decrease of enrichment (Figure 3.5.4b). Different biotypes of noncoding RNAs were categorized. snRNAs, snoRNAs and miRNAs were enriched in the TMG-IP samples (Figure 3.5.4c). However, no significant difference was observed between D0 and D5. Further analysis was done to single out the genes that were uniquely increased or decreased in TMG enrichment on D5 vs D0 (Figure 3.5.4d). The majority of the RNAs that were enriched in TMG-IP on D5 were mRNAs, miRNAs and long non-coding RNAs. Surprisingly, RNU6-914P, a predicted snRNA U6 pseudogene, showed dramatic decrease (down by ~89%) in TMG enrichment. The cap structure of U6 is gamma-monomethyl phosphate, which is not substrate of Tgs1.


a
IP
Tet(+)
Tet(-)
Input
Ctrl
TMG
Ctrl
TMG
Tet(+)
Tet(-)
U3


b
TMG-IP: Tet(-) vs Tet(+)
Log2 FC
Log2 CPM


c
Percentage of different types of ncRNAs
Read counts percentage
0.00
0.02
0.04
0.06
0.08
snRNA
misc_RNA
miRNA
snoRNA
rRNA
Mt_rRNA
Mt_tRNA
Tet(+)
Tet(-)
TMG-IP
Tet(+)
Tet(-)
Ctrl-IP
Tet(+)
Tet(-)
Input

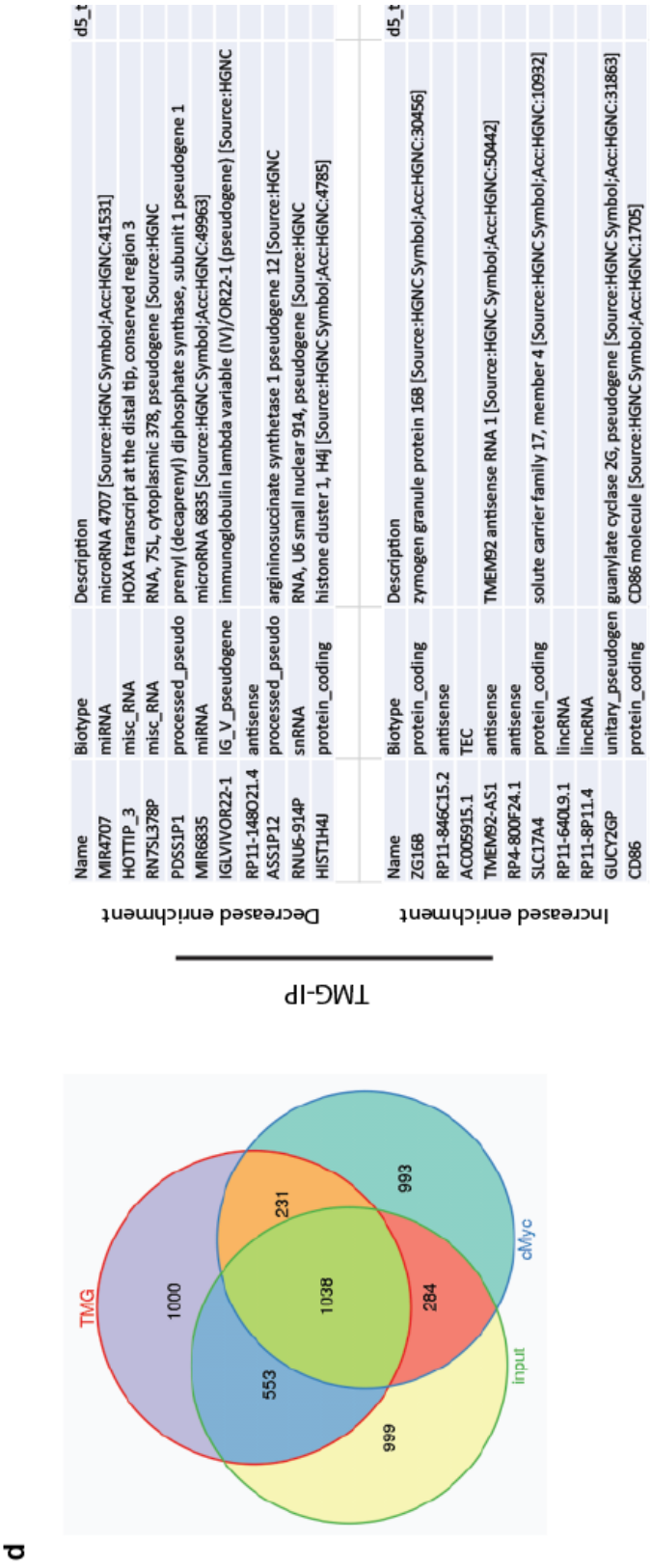


Figure 3.5.4 The effect of shutting down Tgs1 on RNA cap structure.

(a) Anti-TMG cap immunoprecipitation and northern blot of snoRNA U3. Immunoprecipitation by anti-TMG antibody (K121) was perform with total RNA from Tgs1* 293 T-REx conditional knockdown cell line in the presence or absence of tetracycline. The monoclonal mouse antibodies were bound to Dynabeads through pre-embedded sheep-anti-mouse IgG. Precipitated RNAs were separated from the bead via protease digestion. To rule out potential nonspecific interactions between beads/IgG and RNA, immunoprecipitation by anti-c-myc mouse antibody was performed as control (Ctrl). 33% of the IP sample and input samples were analyzed by northern blot. **(b)** MA plot of differential TMG-IP enrichment in the absence of tetracycline vs in the presence of tetracycline, after normalizing to the input samples. FC = fold change. Red dots are small nucleolar RNAs. **(c)** Percentage of different types of noncoding RNAs in IP libraries and input libraries. **(d)** Genes whose enrichments were specifically altered in TMG-IP. A Venn diagram was generated with genes that showed more than 2 folds increase or decrease in enrichment in TMG-IP, c-myc-IP(control) and input samples. 1000 genes were present only in the TMG-IP diagram without overlapping. Top 10 genes with increased or decreased enrichment were listed.

Chapter 4 Discussion

Trimethylguanosine synthase 1 (Tgs1) is responsible for hypermethylation of several important RNAs, including snRNAs, snoRNAs and telomerase RNA. In recent years, new substrates of Tgs1 such as selenoprotein mRNAs and miRNAs have been reported (Martinez et al., 2017; Wurth et al., 2014). The expanding role of this methyltransferase warrants a comprehensive evaluation of its impact on multiple cellular activities, including RNA processing, cell growth, splicing, DNA damage response, and cell cycle regulation. To achieve this goal, I utilized a previously reported putative Tgs1-KO mouse model, as well as created a new human Tgs1 conditional knockdown cell line. In the putative Tgs1-KO mouse embryonic fibroblasts, I identified a mutant version of Tgs1. The mutant Tgs1 retains its methyltransferase activity, as indicated by the hypermethylation of snRNAs. Although the snRNAs are still TMG-capped, U2 processing is impaired in the mutant MEFs. I also created a human Tgs1 mutant cell line, which recapitulated the U2 processing defect. Detailed mapping of the 3'-end of U2 revealed the accumulation of ~10nt extended U2 precursor as well as the reduction of the monoadenylated U2 in mutant cells. In the new human Tgs1 conditional knockdown cell line, I showed that severe growth defects were present after repression of Tgs1 expression. DNA damage response and G2/M cell cycle arrest were also observed in the absence of Tgs1. I proved that it is the methyltransferase activity, rather than the Tgs1 protein, that is required for cell growth and proper U2 processing. I also examined the gene expression profiles and RNA cap structure through next generation sequencing in both the Tgs1 mutant cell line and the Tgs1 depleted cell line. In summary, this study provided new insights on a previously reported Tgs1-KO mouse model, generated new tools for characterization of mammalian Tgs1, and provided comprehensive information on the impact of Tgs1-depletion in mammalian cells.

4.1 A mutant version of Tgs1

Mammalian Tgs1 was initially identified as PRIP-interacting protein with methyltransferase domain (PIMT) by the Reddy lab (Zhu et al., 2001). PRIP is a coactivator of peroxisome proliferator activated receptor (PPAR), which belongs to the nuclear super receptor family that regulates gene expression in lipid metabolism (Zhu et al., 2000). The Tgs1 protein contains a large N-terminal domain (1-384aa) that enhances the transcriptional activity of PPARɣ. Immunoprecipitation showed that Tgs1 forms a complex with transcription coactivators CBP, p300, PBP, and PRIP *in vivo* (Misra et al., 2002). CBP and p300 knockout mouse models were established and those models demonstrated that CBP/p300 are essential for mouse embryonic development (Kung et al., 2000; Yao et al., 1998; S. Yu & Reddy, 2007). The Reddy lab created a putative Tgs1-KO mouse model by deleting exon 3 and exon 4 of Tgs1, which removed the N-terminal amino acids 58-381. It was assumed that the theoretically remaining truncated protein would be rapidly degraded, even the exon 2 and exon 5 are still in-frame. This mouse model was regarded as Tgs1-null and showed early embryonic lethality at E5.5 (Jia et al., 2012).

In Chapter Three, I showed that a mutant version of Tgs1 is present in mouse embryonic fibroblasts (MEFs) derived from the conditional Tgs1-null mouse model. The mutant Tgs1 is a splicing variant of exon 2 and exon 5 that retains its methyltransferase activity, as the snRNAs in mutant MEF cell line still contain TMG caps. Based on the similarity between mouse and human Tgs1, I constructed a human version of mutant Tgs1 (Tgs1*). Combined with Flp-In system and CRISPR, human mutant Tgs1 was integrated 293 T-Rex cell line while the endogenous Tgs1

was knocked out, resulting in a human mutant Tgs1 cell line. The human version of mutant Tgs1 protein can be stably expressed and snRNAs in human mutant cell line are also TMG-capped. Despite the mouse embryonic lethality, both mouse and human mutant Tgs1 cell lines are viable. Those observations suggest a separate of function among different domains of mammalian Tgs1. The N-terminal domain, with its transcriptional coactivator activity, is essential in embryonic development. But this domain is not indispensable to sustain cell viability. The C-terminal domain, with the methyltransferase activity, can function independently from the N-terminal domain and sustain cell viability.

4.2 Role of Tgs1 in U2 processing

The mutant Tgs1 lacks large portion of the N-terminal domain, while the methyltransferase in the C-terminal domain is intact. Although the snRNAs are still TMG-capped in the mutant cell line, their biogenesis does not go as well as in the presence of the full-length Tgs1. Accumulation of 3’ extended pre-U2 is present in the mutant MEFs. The same processing defect can be observed in the human version of mutant Tgs1 cell line, although not as severe as in the mutant MEFs. The 3’ extended pre-snRNA U2 is not hypermethylated, indicating that TMG capping precedes the 3’ trimming. Detailed analysis of the 3’ end of U2 reveals two major forms of pre-U2 accumulate in the mutant cell line: a 10-nucleotide extended form and a 1-nucleotide extended form. The 10-nucleotide extended pre-U2 corresponds with the slow migrating band above mature U2 on northern blot. The full-length Tgs1 is able to rescue the U2 processing defect, while the catalytically inactive Tgs1 fails to do so. It is the methyltransferase activity that is required for pre-U2 3’ trimming, likely through

hypermethylation of pre-U2. Little is known about the 3' trimming of pre-snRNAs. It is not clear what the nuclease responsible for the trimming is or how the nuclease gets recruited to pre-snRNAs. One possibility is that the TMG cap itself could recruit the nuclease, either directly or through other protein factors that bind to the cap. The second possibility is that the TMG cap could affect the structure of pre-U2, making the 3' tail more accessible for trimming. Recently, TOE1, a 3'-to-5' exonuclease with a preference for adenosines *in vitro*, has been reported to be involved in snRNA 3' maturation (Lardelli et al., 2017). In their study, the effect of TOE1 on pre-U2 processing is subtle. TOE1 knockdown did not result in accumulation of U2 precursors on northern blot of total RNA like what we observed in Tgs1 mutant MEFs and human Tgs1 conditional KD cell line. Instead, TOE1 knockdown followed by immunoprecipitation with catalytically inactive TOE1 (TOE1 DE) resulted in slow migrating band of U2 that associated with TOE1 DE. To further examine whether TOE1 is the nuclease responsible for pre-U2 trimming, overexpression/ knockdown of TOE1 could be done in the background of my Tgs1 conditional KD cell line. If TOE1 overexpression alleviates pre-U2 accumulation/TOE1 KD aggravates pre-U2 accumulation in the background of Tgs1 conditional KD cell line, further question can be asked regarding to the cap structure of TOE1-associated pre-U2. The potential model is that full-length Tgs1 hypermethylates pre-U2, the TMG cap then recruits TOE1, directly or indirectly, to trim the 3' tail (Figure 4.2.1).

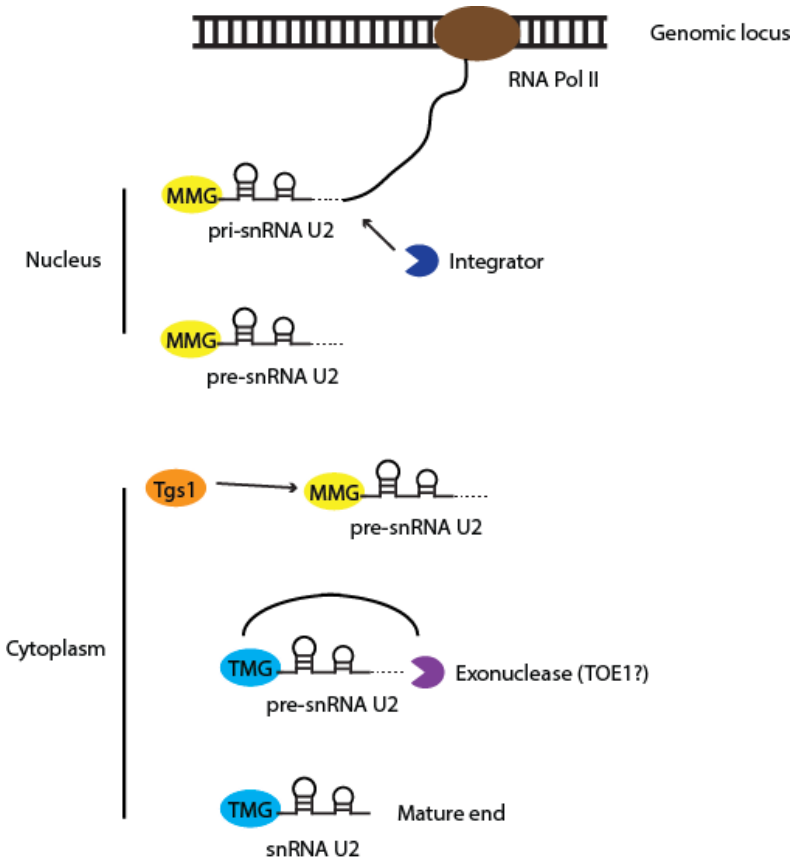


Figure 4.2.1 The role of Tgs1 in U2 processing

Primary snRNA U2 (pri-snRNA U2) is transcribed by RNA polymerase II. A monomethylgaunosine (MMG) cap is added co-transcriptionally. The Integrator complex cleaves pri-snRNA U2 in the nucleus. The 3' extended U2 precursor (pre-snRNA U2) is transported to the cytoplasm, where the MMG cap is hypermethylated by Tgs1. The trimethylguanosine (TMG) cap then recruits an unknown exonuclease(s), possibly TOE1, which trims the 3' tail.

A more complicated possibility is that instead of through the TMG cap on pre-U2, the full-length Tgs1 is affecting pre-U2 trimming indirectly through regulating its substrate genes that are required for snRNA biogenesis. Although mRNAs are generally considered to be carrying monomethylguanosine caps, a lot of mRNAs showed difference in TMG-IP enrichment between WT and mutant MEFs. And some of them are involved in snRNA biogenesis. For example, in mutant MEFs, PHAX mRNA showed ~20% decrease in TMG-IP enrichment. The CRM1/PHAX pathway is required for exportation of pre-snRNAs to the cytoplasm (Ohno,

Segref, Bachi, Wilm, & Mattaj, 2000). Currently, the details of pre-snRNA trimming are unknown. But it is known that the trimming occurs in the cytoplasm (Kleinschmidt & Pederson, 1987). The full-length Tgs1 could affect pre-U2 trimming indirectly through regulating PHAX mRNA hypermethylation and expression. In this case, the TMG cap on pre-U2 would not be required for the 3' trimming. To address this possibility, *in vitro* transcribed pre-U2 with 5'pppG cap could be tested in wildtype cell extract. If 5'pppG capped pre-U2 could still be trimmed, genes that are regulated by full-length Tgs1 need to be further examined using the expression profile and TMG-IP library data in Chapter Three.

Other than pre-U2 trimming, mutant Tgs1 also reduces the non-templated adenosine at 3' end of U2. Catalytically inactive full-length Tgs1 cannot restore the reduction, but overexpression of mutant Tgs1 or introducing catalytically active full-length Tgs1 is able to partially restore the reduction. Human CCA-adding enzyme is able to add non-templated adenosine to U2 *in vitro* (Cho et al., 2002). It is worth testing whether CCA-adding enzyme can synergize with Tgs1 on U2 monoadenylation *in vivo* by utilizing the Tgs1 conditional knockdown cell line described in Chapter Three.

It has been reported that a short isoform of human Tgs1 (sTgs1) exists in association with small nucleolar ribonucleoprotein core proteins (Girard et al., 2008). Unlike full-length Tgs1, which localizes to the cytoplasm, sTgs1 localizes to the nucleus. sTgs1 is produced by limited proteolytic processing of full-length Tgs1 N-terminus, whose length and sequence is very similar to the mutant Tgs1 we identified in the putative Tgs1-KO MEFs. Since pre-U2 is localized in cytoplasm, the localization of mutant Tgs1 likely contributes to the processing defect.

4.3 Role of Tgs1 in splicing

In Tgs1 mutant cells, despite the processing defect of U2, which is a key component of the spliceosome, no overall splicing defects were present. Further examinations for five specific splicing events (skipped exon, alternative 5' splice site, alternative 3' end splice site, mutually exclusive exons and retained intron) showed no significant difference between the mutant cells and the wildtype cells. It is possible that the processing defect of U2 is too subtle to affect splicing. After all, the mature form of U2 is largely unaffected in Tgs1 mutant cells, despite the accumulation of U2 precursors.

Alternatively, certain introns can still be affected without showing significant difference in a category of splicing events. In budding yeast *Δtgs1* strain, splicing defect is present in SAE3 and PCH2 introns. The SAE3 and PCH2 transcripts have introns that deviate from the consensus yeast branchpoint 5'-UACUAAC. SAE3 carries 5'-UAUUAAC and PCH2 carries 5'-CACUAAC, both are extremely rare among yeast genes. Mutations that restore a consensus branchpoint alleviate their dependency on Tgs1 (Qiu et al., 2011). It seems that a more degenerative branchpoint would be vulnerable. However, human intron branchpoints are already much more degenerative than yeasts (K. Gao, Masuda, Matsuura, & Ohno, 2008). Further optimization of the algorithm is needed to exploit whether certain introns are affected in mammals.

4.4 Tgs1 and cell viability

During my attempts to knock out Tgs1 in human cells, multiple measurements indicated that Tgs1-KO cell line is not viable. The human Tgs1 conditional knockdown cell line I generated shows severe growth defect upon shutting down Tgs1. Despite the essential role of Tgs1 in human cell viability, a short window of time allows us to collect data about gene expression and cap hypermethylation upon shutting down Tgs1. Notably dysregulated genes such as TSPAN2 (cell motility), YPEL4 (cell proliferation) and Inhibin (growth inhibition) emerge from early stages of Tgs1 knockdown. Also, several miRNAs and lincRNAs show alteration in cap hypermethylation. One possible link between Tgs1 depletion and the growth defect would be through altered cap hypermethylation of regulatory noncoding RNAs such as miRNAs and lincRNAs, thus affecting their mRNA targets that are involved in cell proliferation. Notably, miR-1182 is decreased in TMG-IP enrichment by ~ 60% after 5 days Tgs1 knockdown. Its expression level is also decreased by ~43%. miR-1182 is a microRNA that binds the 3' UTR of human telomerase reverse transcriptase (hTERT) mRNA (D. Zhang et al., 2015; Zhou, Dai, & Song, 2016). It suppresses hTERT expression in bladder cancer cell lines, which inhibits cell proliferation and sensitizes cancer cells to chemotherapy. One would expect for an increase of hTERT mRNA in response to the reduction of miR-1182. Surprisingly, hTERT mRNA is decreased by ~60% after Tgs1 knockdown. The decrease in hTERT mRNA is consistent with the growth defect we observed but contradicts the reported function of miR-1182 in repressing hTERT. In most cases, miRNAs that bind to the 3' UTR downregulate their target mRNAs. However, exceptions do exist when certain miRNAs bind to the 3' UTR of their targets in competition with pathways that degrade mRNAs. In those cases, binding of miRNAs to the 3' UTR upregulates their target mRNAs. For example, miRNA-466l upregulates IL-10 mRNA by competing with tristetraprolin-mediated IL-10 mRNA degradation (F. Ma et al., 2010). It would

be interesting to explore whether miR-1182 functions differently than previously reports in bladder cancer cell lines and upregulates hTERT mRNA in my Tgs1 conditional knockdown cell line, which is HEK293 cell line.

Another link between Tgs1 knockdown and the growth defect is through the TMG-capped selenoprotein mRNAs, which are involved in responding to oxidative stress by selective upregulation (Dai et al., 2016; Touat-Hamici, Legrain, Bulteau, & Chavatte, 2014). The changes in selenoprotein mRNA level after Tgs1 knockdown are mild (20%~30%). Exposing the conditional Tgs1 knockdown cell line to oxidative stress could test whether the upregulation ability of selenoprotein mRNAs is impaired in the absence of Tgs1.

As mentioned before, YPEL4, which modulates adrenal cell proliferation (Oki et al., 2016), is among the consistently dysregulated genes in the early stage of Tgs1 knockdown. It has been reported that methylation of CpG island is present in the promoter region of YPEL4 (Bisarro Dos Reis et al., 2017). The DNA methyltransferases DNMTs mediate CpG methylation, which include DNMT1, DNMT2 (TRDMT1), DNMT3A, and DNMT3B. They are all decreased by 40%~60% after Tgs1 knockdown. The downregulation of DNMTs could reduce methylation of YPEL4 promoter, resulting in upregulation of YPEL4. Besides, DNMT1 itself regulates human endometrial carcinoma cell proliferation. Knockdown of DNMT1 leads to proliferation inhibition and cell cycle arrest (X. Wang & Li, 2017). In human BJ fibroblasts, DNMT2 depletion also results in decreased cell proliferation and increases sensitivity to oxidative stress (Lewinska et al., 2018). Inhibin, another consistently dysregulated gene in the early stage of Tgs1 knockdown, is regulated by DNMT1-associated long non-coding RNA (DACOR1) (Merry et al., 2015). DNMTs seem to be one of the central players in gene dysregulation upon Tgs1

knockdown. How Tgs1 knockdown results in downregulation of all four known DNMTs is yet to be determined.

4.5 Tgs1 and DNA damage response

The absence of Tgs1 protein is coupled by DNA damage response, as indicated by the increase in γH2AX. Next generation sequencing data reveals up to ~5 fold increase in all three members of the Growth Arrest and DNA Damage-inducible 45 (GADD45) proteins: GADD45A, GADD45B, and GADD45G. GADD45 is induced by DNA damage and other stress signals associated with growth arrest and apoptosis (Salvador, Brown-Clay, & Fornace, 2013). There is also a ~1.3 fold increase in DNA damage-induced apoptosis suppressor (DDIAS, human homolog of mouse Noxin) after 3 days of Tgs1 knockdown. DDIAS is the target of nuclear factor of activated T cells 1 (NFATC1), which is upregulated as well after Tgs1 knockdown. NFATC1 and DDIAS overexpression promotes lung cancer cell proliferation against cisplatin, a broad-spectrum anticancer drug (Im et al., 2016). Those observations are consistent with the conclusion that Tgs1 knockdown results in DNA damage response. The downregulation of selenoprotein mRNAs and DNMT2 in the absence of Tgs1 could sensitize cells to oxidative stress, leading to DNA damage response at atmosphere oxygen level.

4.6 Conclusions and future directions

Mammalian Tgs1 is required for proper 3’ end processing of U2. The U2 processing defect I showed in both mouse mutant Tgs1 cell line and human Tgs1 conditional knockdown

cell line is more prominent than two recent reports on the role of TOE1 in snRNA 3' end trimming (Lardelli et al., 2017; Son, Park, & Kim, 2018). Besides, TOE1 is localized to the nucleus, while the trimming of U2 occurs in the cytoplasm (Kleinschmidt & Pederson, 1987; Rohleder & Wieben, 1986). It is important to test whether TOE1 is indeed the long-sought exonuclease that trims the 3' end of snRNAs. It will also be informative to know whether mammalian Tgs1 is required for 3'end processing of a broader range of substrates. The 3' end RNA-seq libraries I generated were with a U2 specific upstream primer. Using nonspecific 5' adaptor followed by ligation would capture all the 3' ends of ncRNAs in the background of Tgs1 conditional knockdown cell line. Tgs1 is likely to facilitate U2 processing by hypermethylating pre-U2. Whether the TMG cap directly recruits the nuclease(s) or indirectly through binding of other factors is yet to be determined.

Mammalian Tgs1 methyltransferase activity is essential for cell growth. As detailed in 4.4 and 4.5, multiple pathways could be behind the growth defects followed by Tgs1 knockdown. Small molecules that target Tgs1 itself or the pathways related to Tg1-dependent cell viability can be used to inhibit cell proliferation. A variety of methyltransferases have already been screened as anticancer drug targets (Boriack-Sjodin, Ribich, & Copeland, 2018). The crystal structure of human Tgs1 methyltransferase domain has been resolved (Monecke et al., 2009). Given the role of Tgs1 in cell growth, DNA damage response, and cell cycle arrest, it could serve as one of those anticancer drug targets.

In yeast, Tgs1 protein *per se*, rather than its methyltransferase activity, is involved in ribosome synthesis (Colau, Thiry, Leduc, Bordonne, & Lafontaine, 2004). With the Tgs1 conditional knock down cell line, it is worth examining whether mammalian Tgs1 has a similar

role. Although the role of Tgs1 in ribosome synthesis is unlikely related to cell viability. Only catalytically active Tgs1 can rescue the growth defect.

References

Baserga, S. J., Gilmore-Hebert, M., & Yang, X. W. (1992). Distinct molecular signals for nuclear import of the nucleolar snRNA, U3. *Genes Dev, 6*(6), 1120-1130.

Bhaya, D., Davison, M., & Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet, 45*, 273-297. doi:10.1146/annurev-genet-110410-132430

Bisarro Dos Reis, M., Barros-Filho, M. C., Marchi, F. A., Beltrami, C. M., Kuasne, H., Pinto, C. A. L., . . . Rogatto, S. R. (2017). Prognostic Classifier Based on Genome-Wide DNA Methylation Profiling in Well-Differentiated Thyroid Tumors. *J Clin Endocrinol Metab, 102*(11), 4089-4099. doi:10.1210/jc.2017-00881

Blasco, M. A., Funk, W., Villeponteau, B., & Greider, C. W. (1995). Functional characterization and developmental regulation of mouse telomerase RNA. *Science, 269*(5228), 1267-1270.

Boon, K. L., Pearson, M. D., & Kos, M. (2015). Self-association of Trimethylguanosine Synthase Tgs1 is required for efficient snRNA/snoRNA trimethylation and pre-rRNA processing. *Sci Rep, 5*, 11282. doi:10.1038/srep11282

Borg, R. M., Fenech Salerno, B., Vassallo, N., Bordonne, R., & 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. *Neurobiol Dis, 94*, 245-258. doi:10.1016/j.nbd.2016.06.015

Boriack-Sjodin, P. A., Ribich, S., & Copeland, R. A. (2018). RNA-modifying proteins as anticancer drug targets. *Nat Rev Drug Discov, 17*(6), 435-453. doi:10.1038/nrd.2018.71

Boulon, S., Verheggen, C., Jady, B. E., Girard, C., Pescia, C., Paul, C., . . . Bertrand, E. (2004). PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. *Mol Cell, 16*(5), 777-787. doi:10.1016/j.molcel.2004.11.013

Bresson, S. M., & Conrad, N. K. (2013). The human nuclear poly(a)-binding protein promotes RNA hyperadenylation and decay. *PLoS Genet, 9*(10), e1003893. doi:10.1371/journal.pgen.1003893

Burd, C. G., & Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science, 265*(5172), 615-621.

Chen, J. L., Blasco, M. A., & Greider, C. W. (2000). Secondary structure of vertebrate telomerase RNA. *Cell, 100*(5), 503-514.

Cho, H. D., Tomita, K., Suzuki, T., & Weiner, A. M. (2002). U2 small nuclear RNA is a substrate for the CCA-adding enzyme (tRNA nucleotidyltransferase). *J Biol Chem, 277*(5), 3447-3455. doi:10.1074/jbc.M109559200

Choi, Y. S., Patena, W., Leavitt, A. D., & McManus, M. T. (2012). Widespread RNA 3'-end oligouridylation in mammals. *RNA, 18*(3), 394-401. doi:10.1261/rna.029306.111

Colau, G., Thiry, M., Leduc, V., Bordonne, R., & Lafontaine, D. L. (2004). The small nucle(ol)ar RNA cap trimethyltransferase is required for ribosome synthesis and intact nucleolar morphology. *Mol Cell Biol, 24*(18), 7976-7986. doi:10.1128/MCB.24.18.7976-7986.2004

Cristofari, G., Adolf, E., Reichenbach, P., Sikora, K., Terns, R. M., Terns, M. P., & Lingner, J. (2007). Human telomerase RNA accumulation in Cajal bodies facilitates telomerase recruitment to telomeres and telomere elongation. *Mol Cell, 27*(6), 882-889. doi:10.1016/j.molcel.2007.07.020

Dai, J., Liu, H., Zhou, J., & Huang, K. (2016). Selenoprotein R Protects Human Lens Epithelial Cells against D-Galactose-Induced Apoptosis by Regulating Oxidative Stress and Endoplasmic Reticulum Stress. *Int J Mol Sci, 17*(2), 231. doi:10.3390/ijms17020231

Ding, Z., Wu, C. J., Jaskelioff, M., Ivanova, E., Kost-Alimova, M., Protopopov, A., . . . Depinho, R. A. (2012). Telomerase Reactivation following Telomere Dysfunction Yields Murine Prostate Tumors with Bone Metastases. *Cell*. doi:10.1016/j.cell.2012.01.039

Feng, J., Funk, W. D., Wang, S. S., Weinrich, S. L., Avilion, A. A., Chiu, C. P., . . . et al. (1995). The RNA component of human telomerase. *Science, 269*(5228), 1236-1241.

Fischer, U., Heinrich, J., van Zee, K., Fanning, E., & Luhrmann, R. (1994). Nuclear transport of U1 snRNP in somatic cells: differences in signal requirement compared with Xenopus laevis oocytes. *J Cell Biol, 125*(5), 971-980.

Franke, J., Gehlen, J., & Ehrenhofer-Murray, A. E. (2008). Hypermethylation of yeast telomerase RNA by the snRNA and snoRNA methyltransferase Tgs1. *J Cell Sci, 121*(Pt 21), 3553-3560. doi:10.1242/jcs.033308

Fu, D., & Collins, K. (2003). Distinct biogenesis pathways for human telomerase RNA and H/ACA small nucleolar RNAs. *Mol Cell, 11*(5), 1361-1372.

Gao, J., Wallis, J. G., Jewell, J. B., & Browse, J. (2017). Trimethylguanosine Synthase1 (TGS1) Is Essential for Chilling Tolerance. *Plant Physiol, 174*(3), 1713-1727. doi:10.1104/pp.17.00340

Gao, K., Masuda, A., Matsuura, T., & Ohno, K. (2008). Human branch point consensus sequence is yUnAy. *Nucleic Acids Res, 36*(7), 2257-2267. doi:10.1093/nar/gkn073

Girard, C., Verheggen, C., Neel, H., Cammas, A., Vagner, S., Soret, J., . . . Bordonne, R. (2008). Characterization of a short isoform of human Tgs1 hypermethylase associating with small nucleolar ribonucleoprotein core proteins and produced by limited proteolytic processing. *J Biol Chem, 283*(4), 2060-2069. doi:10.1074/jbc.M704209200

Greider, C. W., & Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell, 43*(2 Pt 1), 405-413.

Greider, C. W., & Blackburn, E. H. (1987). The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell, 51*(6), 887-898.

Gribling-Burrer, A. S., Leichter, M., Wurth, L., Huttin, A., Schlotter, F., Troffer-Charlier, N., . . . Allmang, C. (2017). SECIS-binding protein 2 interacts with the SMN complex and the methylosome for selenoprotein mRNP assembly and translation. *Nucleic Acids Res, 45*(9), 5399-5413. doi:10.1093/nar/gkx031

Hallais, M., Pontvianne, F., Andersen, P. R., Clerici, M., Lener, D., Benbahouche Nel, H., . . . Bertrand, E. (2013). CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nat Struct Mol Biol, 20*(12), 1358-1366. doi:10.1038/nsmb.2720

Hamm, J., Darzynkiewicz, E., Tahara, S. M., & Mattaj, I. W. (1990). The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell, 62*(3), 569-577.

Hashimoto, O., Ushiro, Y., Sekiyama, K., Yamaguchi, O., Yoshioka, K., Mutoh, K., & Hasegawa, Y. (2006). Impaired growth of pancreatic exocrine cells in transgenic mice expressing human activin betaE subunit. *Biochem Biophys Res Commun, 341*(2), 416-424. doi:10.1016/j.bbrc.2005.12.205

Hausmann, S., & Shuman, S. (2005). Specificity and mechanism of RNA cap guanine-N2 methyltransferase (Tgs1). *J Biol Chem, 280*(6), 4021-4024. doi:10.1074/jbc.C400554200

Hausmann, S., Zheng, S., Costanzo, M., Brost, R. L., Garcin, D., Boone, C., . . . Schwer, B. (2008). Genetic and biochemical analysis of yeast and human cap trimethylguanosine synthase: functional overlap of 2,2,7-trimethylguanosine caps, small nuclear ribonucleoprotein components, pre-mRNA splicing factors, and RNA decay pathways. *J Biol Chem, 283*(46), 31706-31718. doi:10.1074/jbc.M806127200

Heo, I., Ha, M., Lim, J., Yoon, M. J., Park, J. E., Kwon, S. C., . . . Kim, V. N. (2012). Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell, 151*(3), 521-532. doi:10.1016/j.cell.2012.09.022

Heo, I., Joo, C., Kim, Y. K., Ha, M., Yoon, M. J., Cho, J., . . . Kim, V. N. (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell, 138*(4), 696-708. doi:10.1016/j.cell.2009.08.002

Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M., & Luhrmann, R. (1998). Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *Embo J, 17*(14), 4114-4126. doi:10.1093/emboj/17.14.4114

Im, J. Y., Lee, K. W., Won, K. J., Kim, B. K., Ban, H. S., Yoon, S. H., . . . Won, M. (2016). DNA damage-induced apoptosis suppressor (DDIAS), a novel target of NFATc1, is associated with cisplatin resistance in lung cancer. *Biochim Biophys Acta, 1863*(1), 40-49. doi:10.1016/j.bbamcr.2015.10.011

Jacobson, M. R., & Pederson, T. (1998). A 7-methylguanosine cap commits U3 and U8 small nuclear RNAs to the nucleolar localization pathway. *Nucleic Acids Res, 26*(3), 756-760.

Jady, B. E., Bertrand, E., & Kiss, T. (2004). Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *J Cell Biol, 164*(5), 647-652. doi:10.1083/jcb.200310138

Jady, B. E., Richard, P., Bertrand, E., & Kiss, T. (2006). Cell cycle-dependent recruitment of telomerase RNA and Cajal bodies to human telomeres. *Mol Biol Cell, 17*(2), 944-954. doi:10.1091/mbc.E05-09-0904

Jerome-Morais, A., Bera, S., Rachidi, W., Gann, P. H., & Diamond, A. M. (2013). The effects of selenium and the GPx-1 selenoprotein on the phosphorylation of H2AX. *Biochim Biophys Acta, 1830*(6), 3399-3406. doi:10.1016/j.bbagen.2013.03.010

Jia, Y., Viswakarma, N., Crawford, S. E., Sarkar, J., Sambasiva Rao, M., Karpus, W. J., . . . Reddy, J. K. (2012). Early embryonic lethality of mice with disrupted transcription cofactor PIMT/NCOA6IP/Tgs1 gene. *Mech Dev, 129*(9-12), 193-207. doi:10.1016/j.mod.2012.08.002

Jurkowski, T. P., Meusburger, M., Phalke, S., Helm, M., Nellen, W., Reuter, G., & Jeltsch, A. (2008). Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism. *RNA, 14*(8), 1663-1670. doi:10.1261/rna.970408

Kachouri-Lafond, R., Dujon, B., Gilson, E., Westhof, E., Fairhead, C., & Teixeira, M. T. (2009). Large telomerase RNA, telomere length heterogeneity and escape from senescence in Candida glabrata. *FEBS Lett, 583*(22), 3605-3610. doi:S0014-5793(09)00810-2 [pii]

10.1016/j.febslet.2009.10.034

Kim, K. M., Cho, H., Choi, K., Kim, J., Kim, B. W., Ko, Y. G., . . . Kim, Y. K. (2009). A new MIF4G domain-containing protein, CTIF, directs nuclear cap-binding protein CBP80/20-dependent translation. *Genes Dev, 23*(17), 2033-2045. doi:10.1101/gad.1823409

Kim, N. W., Piatyszek, M. A., Prowse, K. R., Harley, C. B., West, M. D., Ho, P. L., . . . Shay, J. W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science, 266*(5193), 2011-2015.

Kiss, T. (2004). Biogenesis of small nuclear RNPs. *J Cell Sci, 117*(Pt 25), 5949-5951. doi:10.1242/jcs.01487

Kiss, T., Fayet-Lebaron, E., & Jady, B. E. (2010). Box H/ACA small ribonucleoproteins. *Mol Cell, 37*(5), 597-606. doi:10.1016/j.molcel.2010.01.032

Kleinschmidt, A. M., & Pederson, T. (1987). Accurate and efficient 3' processing of U2 small nuclear RNA precursor in a fractionated cytoplasmic extract. *Mol Cell Biol, 7*(9), 3131-3137.

Kleinschmidt, A. M., & Pederson, T. (1990). RNA processing and ribonucleoprotein assembly studied in vivo by RNA transfection. *Proc Natl Acad Sci U S A, 87*(4), 1283-1287.

Komonyi, O., Papai, G., Enunlu, I., Muratoglu, S., Pankotai, T., Kopitova, D., . . . Boros, I. (2005). DTL, the Drosophila homolog of PIMT/Tgs1 nuclear receptor coactivator-interacting protein/RNA methyltransferase, has an essential role in development. *J Biol Chem, 280*(13), 12397-12404. doi:10.1074/jbc.M409251200

Kryukov, G. V., Castellano, S., Novoselov, S. V., Lobanov, A. V., Zehtab, O., Guigo, R., & Gladyshev, V. N. (2003). Characterization of mammalian selenoproteomes. *Science, 300*(5624), 1439-1443. doi:10.1126/science.1083516

Kung, A. L., Rebel, V. I., Bronson, R. T., Ch'ng, L. E., Sieff, C. A., Livingston, D. M., & Yao, T. P. (2000). Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP. *Genes Dev, 14*(3), 272-277.

Labunskyy, V. M., Hatfield, D. L., & Gladyshev, V. N. (2014). Selenoproteins: molecular pathways and physiological roles. *Physiol Rev, 94*(3), 739-777. doi:10.1152/physrev.00039.2013

Lamm, A. T., Stadler, M. R., Zhang, H., Gent, J. I., & Fire, A. Z. (2011). Multimodal RNA-seq using single-strand, double-strand, and CircLigase-based capture yields a refined and extended description of the C. elegans transcriptome. *Genome Res, 21*(2), 265-275. doi:10.1101/gr.108845.110

Lardelli, R. M., Schaffer, A. E., Eggens, V. R., Zaki, M. S., Grainger, S., Sathe, S., . . . Gleeson, J. G. (2017). Biallelic mutations in the 3' exonuclease TOE1 cause pontocerebellar hypoplasia and uncover a role in snRNA processing. *Nat Genet, 49*(3), 457-464. doi:10.1038/ng.3762

Lemm, I., Girard, C., Kuhn, A. N., Watkins, N. J., Schneider, M., Bordonne, R., & Luhrmann, R. (2006). Ongoing U snRNP biogenesis is required for the integrity of Cajal bodies. *Mol Biol Cell, 17*(7), 3221-3231. doi:10.1091/mbc.E06-03-0247

Levin, J. Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D. A., Friedman, N., . . . Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods, 7*(9), 709-715. doi:10.1038/nmeth.1491

Lewinska, A., Adamczyk-Grochala, J., Kwasniewicz, E., Deregowska, A., Semik, E., Zabek, T., & Wnuk, M. (2018). Reduced levels of methyltransferase DNMT2 sensitize human fibroblasts to oxidative stress and DNA damage that is accompanied by changes in proliferation-related miRNA expression. *Redox Biol, 14*, 20-34. doi:10.1016/j.redox.2017.08.012

Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., & Burley, S. K. (2000). Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell, 100*(3), 323-332.

Lin, J., Ly, H., Hussain, A., Abraham, M., Pearl, S., Tzfati, Y., . . . Blackburn, E. H. (2004). A universal telomerase RNA core structure includes structured motifs required for binding the telomerase reverse transcriptase protein. *Proc Natl Acad Sci U S A, 101*(41), 14713-14718. doi:10.1073/pnas.0405879101

0405879101 [pii]

Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V., & Cech, T. R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science, 276*(5312), 561-567.

Linsen, S. E., de Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R. K., . . . Cuppen, E. (2009). Limitations and possibilities of small RNA digital gene expression profiling. *Nat Methods, 6*(7), 474-476. doi:10.1038/nmeth0709-474

Lleo, A., Zhang, W., Zhao, M., Tan, Y., Bernuzzi, F., Zhu, B., . . . Group, P. B. C. E. S. (2015). DNA methylation profiling of the X chromosome reveals an aberrant demethylation on CXCR3 promoter in primary biliary cirrhosis. *Clin Epigenetics, 7*, 61. doi:10.1186/s13148-015-0098-9

Lukowiak, A. A., Narayanan, A., Li, Z. H., Terns, R. M., & Terns, M. P. (2001). The snoRNA domain of vertebrate telomerase RNA functions to localize the RNA within the nucleus. *RNA, 7*(12), 1833-1844.

Ly, H., Blackburn, E. H., & Parslow, T. G. (2003). Comprehensive structure-function analysis of the core domain of human telomerase RNA. *Mol Cell Biol, 23*(19), 6849-6856.

Ma, F., Liu, X., Li, D., Wang, P., Li, N., Lu, L., & Cao, X. (2010). MicroRNA-466l upregulates IL-10 expression in TLR-triggered macrophages by antagonizing RNA-binding protein tristetraprolin-mediated IL-10 mRNA degradation. *J Immunol, 184*(11), 6053-6059. doi:10.4049/jimmunol.0902308

Ma, H. T., & Poon, R. Y. C. (2018). TRIP13 Functions in the Establishment of the Spindle Assembly Checkpoint by Replenishing O-MAD2. *Cell Rep, 22*(6), 1439-1450. doi:10.1016/j.celrep.2018.01.027

Marks, D. H., Thomas, R., Chin, Y., Shah, R., Khoo, C., & Benezra, R. (2017). Mad2 Overexpression Uncovers a Critical Role for TRIP13 in Mitotic Exit. *Cell Rep, 19*(9), 1832-1845. doi:10.1016/j.celrep.2017.05.021

Martinez, I., Hayes, K. E., Barr, J. A., Harold, A. D., Xie, M., Bukhari, S. I. A., . . . DiMaio, D. (2017). An Exportin-1-dependent microRNA biogenesis pathway during human cell quiescence. *Proc Natl Acad Sci U S A, 114*(25), E4961-E4970. doi:10.1073/pnas.1618732114

Matera, A. G., Terns, R. M., & Terns, M. P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol, 8*(3), 209-220. doi:10.1038/nrm2124

Mattaj, I. W. (1986). Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell, 46*(6), 905-911.

McCormick-Graham, M., & Romero, D. P. (1996). A single telomerase RNA is sufficient for the synthesis of variable telomeric DNA repeats in ciliates of the genus Paramecium. *Mol Cell Biol, 16*(4), 1871-1879.

Merry, C. R., Forrest, M. E., Sabers, J. N., Beard, L., Gao, X. H., Hatzoglou, M., . . . Khalil, A. M. (2015). DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer. *Hum Mol Genet, 24*(21), 6240-6253. doi:10.1093/hmg/ddv343

Millevoi, S., & Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res, 38*(9), 2757-2774. doi:10.1093/nar/gkp1176

Misra, P., Qi, C., Yu, S., Shah, S. H., Cao, W. Q., Rao, M. S., . . . Reddy, J. K. (2002). Interaction of PIMT with transcriptional coactivators CBP, p300, and PBP differential role in transcriptional regulation. *J Biol Chem, 277*(22), 20011-20019. doi:10.1074/jbc.M201739200

Mitchell, J. R., Cheng, J., & Collins, K. (1999). A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol Cell Biol, 19*(1), 567-576.

Mohr, S., Ghanem, E., Smith, W., Sheeter, D., Qin, Y., King, O., . . . Lambowitz, A. M. (2013). Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA, 19*(7), 958-970. doi:10.1261/rna.039743.113

Monecke, T., Dickmanns, A., & Ficner, R. (2009). Structural basis for m7G-cap hypermethylation of small nuclear, small nucleolar and telomerase RNA by the dimethyltransferase TGS1. *Nucleic Acids Res, 37*(12), 3865-3877. doi:10.1093/nar/gkp249

Morin, G. B. (1989). The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell, 59*(3), 521-529.

Mouaikel, J., Bujnicki, J. M., Tazi, J., & Bordonne, R. (2003). Sequence-structure-function relationships of Tgs1, the yeast snRNA/snoRNA cap hypermethylase. *Nucleic Acids Res, 31*(16), 4899-4909.

Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J., & 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*(4), 891-901.

Natalizio, B. J., & Wente, S. R. (2013). Postage for the messenger: designating routes for nuclear mRNA export. *Trends Cell Biol, 23*(8), 365-373. doi:10.1016/j.tcb.2013.03.006

Newman, M. A., Mani, V., & Hammond, S. M. (2011). Deep sequencing of microRNA precursors reveals extensive 3' end modification. *RNA, 17*(10), 1795-1803. doi:10.1261/rna.2713611

Niedzwiecka, A., Marcotrigiano, J., Stepinski, J., Jankowska-Anyszka, M., Wyslouch-Cieszynska, A., Dadlez, M., . . . Stolarski, R. (2002). Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5' cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J Mol Biol, 319*(3), 615-635. doi:10.1016/S0022-2836(02)00328-5

Ohno, M., Segref, A., Bachi, A., Wilm, M., & Mattaj, I. W. (2000). PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell, 101*(2), 187-198. doi:10.1016/S0092-8674(00)80829-6

Oki, K., Plonczynski, M. W., Gomez-Sanchez, E. P., & Gomez-Sanchez, C. E. (2016). YPEL4 modulates HAC15 adrenal cell proliferation and is associated with tumor diameter. *Mol Cell Endocrinol, 434*, 93-98. doi:10.1016/j.mce.2016.06.022

Otsubo, C., Otomo, R., Miyazaki, M., Matsushima-Hibiya, Y., Kohno, T., Iwakawa, R., . . . Enari, M. (2014). TSPAN2 is involved in cell invasion and motility during lung cancer progression. *Cell Rep, 7*(2), 527-538. doi:10.1016/j.celrep.2014.03.027

Palacios, I., Hetzer, M., Adam, S. A., & Mattaj, I. W. (1997). Nuclear import of U snRNPs requires importin beta. *Embo J, 16*(22), 6783-6792. doi:10.1093/emboj/16.22.6783

Parrinello, S., Samper, E., Krtolica, A., Goldstein, J., Melov, S., & Campisi, J. (2003). Oxygen sensitivity severely limits the replicative lifespan of murine fibroblasts. *Nat Cell Biol, 5*(8), 741-747. doi:10.1038/ncb1024

Petrossian, T. C., & Clarke, S. G. (2009). Multiple Motif Scanning to identify methyltransferases from the yeast proteome. *Mol Cell Proteomics, 8*(7), 1516-1526. doi:10.1074/mcp.M900025-MCP200

Pradet-Balade, B., Girard, C., Boulon, S., Paul, C., Azzag, K., Bordonne, R., . . . Verheggen, C. (2011). CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. *Embo J, 30*(11), 2205-2218. doi:10.1038/emboj.2011.128

Preiss, T., Muckenthaler, M., & Hentze, M. W. (1998). Poly(A)-tail-promoted translation in yeast: implications for translational control. *RNA, 4*(11), 1321-1331.

Prowse, K. R., Avilion, A. A., & Greider, C. W. (1993). Identification of a nonprocessive telomerase activity from mouse cells. *Proc Natl Acad Sci U S A, 90*(4), 1493-1497.

Qiu, Z. R., Schwer, B., & Shuman, S. (2015). Two Routes to Genetic Suppression of RNA Trimethylguanosine Cap Deficiency via C-Terminal Truncation of U1 snRNP Subunit Snp1 or Overexpression of RNA Polymerase Subunit Rpo26. *G3 (Bethesda), 5*(7), 1361-1370. doi:10.1534/g3.115.016675

Qiu, Z. R., Shuman, S., & Schwer, B. (2011). An essential role for trimethylguanosine RNA caps in Saccharomyces cerevisiae meiosis and their requirement for splicing of SAE3 and PCH2 meiotic pre-mRNAs. *Nucleic Acids Res, 39*(13), 5633-5646. doi:10.1093/nar/gkr083

Raabe, C. A., Tang, T. H., Brosius, J., & Rozhdestvensky, T. S. (2014). Biases in small RNA deep sequencing data. *Nucleic Acids Res, 42*(3), 1414-1426. doi:10.1093/nar/gkt1021

Raker, V. A., Plessel, G., & Luhrmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *Embo J, 15*(9), 2256-2269.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc, 8*(11), 2281-2308. doi:10.1038/nprot.2013.143

Reddy, R., Ro-Choi, T. S., Henning, D., & Busch, H. (1974). Primary sequence of U-1 nuclear ribonucleic acid of Novikoff hepatoma ascites cells. *J Biol Chem, 249*(20), 6486-6494.

Reddy, R., Ro-Choi, T. S., Henning, D., Shibata, H., Choi, Y. C., & Busch, H. (1972). MOdified nucleosides of nuclear and nucleolar low molecular weight ribonucleic acid. *J Biol Chem, 247*(22), 7245-7250.

Rohleder, A., & Wieben, E. (1986). Solid-phase processing of U2 snRNA precursors. *Biochemistry, 25*(20), 5910-5914.

Rollenhagen, C., & Pante, N. (2006). Nuclear import of spliceosomal snRNPs. *Can J Physiol Pharmacol, 84*(3-4), 367-376. doi:10.1139/y05-101

Rutkowska-Wlodarczyk, I., Stepinski, J., Dadlez, M., Darzynkiewicz, E., Stolarski, R., & Niedzwiecka, A. (2008). Structural changes of eIF4E upon binding to the mRNA 5' monomethylguanosine and trimethylguanosine Cap. *Biochemistry, 47*(9), 2710-2720. doi:10.1021/bi701168z

Salvador, J. M., Brown-Clay, J. D., & Fornace, A. J., Jr. (2013). Gadd45 in stress signaling, cell cycle control, and apoptosis. *Adv Exp Med Biol, 793*, 1-19. doi:10.1007/978-1-4614-8289-5_1

Shatkin, A. J. (1976). Capping of eucaryotic mRNAs. *Cell, 9*(4 PT 2), 645-653.

Son, A., Park, J. E., & Kim, V. N. (2018). PARN and TOE1 Constitute a 3' End Maturation Module for Nuclear Non-coding RNAs. *Cell Rep, 23*(3), 888-898. doi:10.1016/j.celrep.2018.03.089

Speckmann, W. A., Terns, R. M., & Terns, M. P. (2000). The box C/D motif directs snoRNA 5'-cap hypermethylation. *Nucleic Acids Res, 28*(22), 4467-4473.

Steinbrenner, H., Bilgic, E., Alili, L., Sies, H., & Brenneisen, P. (2006). Selenoprotein P protects endothelial cells from oxidative damage by stimulation of glutathione peroxidase expression and activity. *Free Radic Res, 40*(9), 936-943. doi:10.1080/10715760600806248

Stern, J. L., Zyner, K. G., Pickett, H. A., Cohen, S. B., & Bryan, T. M. (2012). Telomerase recruitment requires both TCAB1 and Cajal bodies independently. *Mol Cell Biol, 32*(13), 2384-2395. doi:10.1128/MCB.00379-12

Tang, W., Kannan, R., Blanchette, M., & Baumann, P. (2012). Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes. *Nature*. doi:10.1038/nature10924

Tao, Y., Yang, G., Yang, H., Song, D., Hu, L., Xie, B., . . . Shi, J. (2017). TRIP13 impairs mitotic checkpoint surveillance and is associated with poor prognosis in multiple myeloma. *Oncotarget, 8*(16), 26718-26731. doi:10.18632/oncotarget.14957

Terns, M. P., & Dahlberg, J. E. (1994). Retention and 5' cap trimethylation of U3 snRNA in the nucleus. *Science, 264*(5161), 959-961.

Terns, M. P., Grimm, C., Lund, E., & Dahlberg, J. E. (1995). A common maturation pathway for small nucleolar RNAs. *Embo J, 14*(19), 4860-4871.

Touat-Hamici, Z., Legrain, Y., Bulteau, A. L., & Chavatte, L. (2014). Selective up-regulation of human selenoproteins in response to oxidative stress. *J Biol Chem, 289*(21), 14750-14761. doi:10.1074/jbc.M114.551994

Tycowski, K. T., Shu, M. D., Kukoyi, A., & Steitz, J. A. (2009). A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. *Mol Cell, 34*(1), 47-57. doi:10.1016/j.molcel.2009.02.020

Vader, G. (2015). Pch2(TRIP13): controlling cell division through regulation of HORMA domains. *Chromosoma, 124*(3), 333-339. doi:10.1007/s00412-015-0516-y

Venteicher, A. S., Abreu, E. B., Meng, Z., McCann, K. E., Terns, R. M., Veenstra, T. D., . . . Artandi, S. E. (2009). A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science, 323*(5914), 644-648. doi:10.1126/science.1165357

Wang, C., & Meier, U. T. (2004). Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *Embo J, 23*(8), 1857-1867. doi:10.1038/sj.emboj.7600181

Wang, X., & Li, B. (2017). DNMT1 regulates human endometrial carcinoma cell proliferation. *Onco Targets Ther, 10*, 1865-1873. doi:10.2147/OTT.S130022

Wieben, E. D., Nenninger, J. M., & Pederson, T. (1985). Ribonucleoprotein organization of eukaryotic RNA. XXXII. U2 small nuclear RNA precursors and their accurate 3' processing in vitro as ribonucleoprotein particles. *J Mol Biol, 183*(1), 69-78.

Will, C. L., & Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol, 13*(3), 290-301.

Wurth, L., Gribling-Burrer, A. S., Verheggen, C., Leichter, M., Takeuchi, A., Baudrey, S., . . . Allmang, C. (2014). Hypermethylated-capped selenoprotein mRNAs in mammals. *Nucleic Acids Res, 42*(13), 8663-8677. doi:10.1093/nar/gku580

Xie, M., Mosig, A., Qi, X., Li, Y., Stadler, P. F., & Chen, J. J. (2008). Structure and function of the smallest vertebrate telomerase RNA from teleost fish. *J Biol Chem, 283*(4), 2049-2059. doi:M708032200 [pii]

10.1074/jbc.M708032200

Yant, L. J., Ran, Q., Rao, L., Van Remmen, H., Shibatani, T., Belter, J. G., . . . Prolla, T. A. (2003). The selenoprotein GPX4 is essential for mouse development and protects from radiation and oxidative damage insults. *Free Radic Biol Med, 34*(4), 496-502.

Yao, T. P., Oh, S. P., Fuchs, M., Zhou, N. D., Ch'ng, L. E., Newsome, D., . . . Eckner, R. (1998). Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell, 93*(3), 361-372.

Yedavalli, V. S., & Jeang, K. T. (2010). Trimethylguanosine capping selectively promotes expression of Rev-dependent HIV-1 RNAs. *Proc Natl Acad Sci U S A, 107*(33), 14787-14792. doi:10.1073/pnas.1009490107

Yehudai-Resheff, S., & Schuster, G. (2000). Characterization of the E.coli poly(A) polymerase: nucleotide specificity, RNA-binding affinities and RNA structure dependence. *Nucleic Acids Res, 28*(5), 1139-1144.

Yi, X., Tesmer, V. M., Savre-Train, I., Shay, J. W., & Wright, W. E. (1999). Both transcriptional and posttranscriptional mechanisms regulate human telomerase template RNA levels. *Mol Cell Biol, 19*(6), 3989-3997.

Yong, J., Kasim, M., Bachorik, J. L., Wan, L., & Dreyfuss, G. (2010). Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. *Mol Cell, 38*(4), 551-562. doi:10.1016/j.molcel.2010.03.014

Yu, G. L., Bradley, J. D., Attardi, L. D., & Blackburn, E. H. (1990). In vivo alteration of telomere sequences and senescence caused by mutated Tetrahymena telomerase RNAs. *Nature, 344*(6262), 126-132. doi:10.1038/344126a0

Yu, S., & Reddy, J. K. (2007). Transcription coactivators for peroxisome proliferator-activated receptors. *Biochim Biophys Acta, 1771*(8), 936-951. doi:10.1016/j.bbalip.2007.01.008

Yuo, C. Y., Ares, M., Jr., & Weiner, A. M. (1985). Sequences required for 3' end formation of human U2 small nuclear RNA. *Cell, 42*(1), 193-202.

Zhang, D., Xiao, Y. F., Zhang, J. W., Xie, R., Hu, C. J., Tang, B., . . . Yang, S. M. (2015). miR-1182 attenuates gastric cancer proliferation and metastasis by targeting the open reading frame of hTERT. *Cancer Lett, 360*(2), 151-159. doi:10.1016/j.canlet.2015.01.044

Zhang, Q., Kim, N. K., & Feigon, J. (2011). Architecture of human telomerase RNA. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1100279108

Zhong, F., Savage, S. A., Shkreli, M., Giri, N., Jessop, L., Myers, T., . . . Artandi, S. E. (2011). Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita. *Genes Dev, 25*(1), 11-16. doi:10.1101/gad.2006411

Zhou, J., Dai, W., & Song, J. (2016). miR-1182 inhibits growth and mediates the chemosensitivity of bladder cancer by targeting hTERT. *Biochem Biophys Res Commun, 470*(2), 445-452. doi:10.1016/j.bbrc.2016.01.014

Zhu, Y., Kan, L., Qi, C., Kanwar, Y. S., Yeldandi, A. V., Rao, M. S., & Reddy, J. K. (2000). Isolation and characterization of peroxisome proliferator-activated receptor (PPAR) interacting protein (PRIP) as a coactivator for PPAR. *J Biol Chem, 275*(18), 13510-13516.

Zhu, Y., Qi, C., Cao, W. Q., Yeldandi, A. V., Rao, M. S., & Reddy, J. K. (2001). Cloning and characterization of PIMT, a protein with a methyltransferase domain, which interacts with and enhances nuclear receptor coactivator PRIP function. *Proc Natl Acad Sci U S A, 98*(18), 10380-10385. doi:10.1073/pnas.181347498

Zipperlen, P., Fraser, A. G., Kamath, R. S., Martinez-Campos, M., & Ahringer, J. (2001). Roles for 147 embryonic lethal genes on C.elegans chromosome I identified by RNA interference and video microscopy. *Embo J, 20*(15), 3984-3992. doi:10.1093/emboj/20.15.3984