THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie Santé

École doctorale Sciences Chimiques et Biologiques pour la Santé (CBS2)

Institut de Génétique Humaine - UMR 9002 CNRS - Université de Montpellier

Regulations of small RNA turnover during programmed DNA elimination in *Tetrahymena*

Présentée par Salman Shehzada
Le 28 Octobre 2022

Sous la direction de Kazufumi Mochizuki

Devant le jury composé de

Julius Brennecke, DR, IMBA, Vienna	**Rapporteur**
Mireille Betemier, DR, Université Paris-Saclay	**Rapporteur**
Severine Chambeyron, DR, Université Montpellier	**Président,Examinateur**
Clément Carré, DR, Sorbonne Université	**Examinateur**
Kazufumi Mochizuki, DR, Université Montpellier	**Directeur**

UNIVERSITÉ DE MONTPELLIER

Contents

Table of Figures 6

List of Acronyms 7

Abstract 9

Résumé 10

Acknowledgements 11

Introduction 13

Small RNA degradation 14

I. The discovery of RNA interference (RNAi) 14

II. Major component of RNA silencing pathway: Argonaute and Dicer Protein 15

a) Argonaute proteins *15*

b) Dicer protein *16*

III. sRNAs Biogenesis and mechanism of action 17

a) siRNA *17*

b) miRNA *19*

c) piRNA *22*

IV. piRNA mediated heterochromatin formation and TE silencing 25

V. Target directed micro-RNA Degradation (TDMD) 26

SUMOylation 29

I. The SUMO Protein 29

II. SUMO pathway and The Enzymology of SUMOylation 30

a) SUMO proteases *30*

b) Heterodimeric SUMO E1 enzyme *32*

c) SUMO conjugating enzyme E2/Ubc9 *32*

d) SUMO E3 ligases *33*

e) SUMO consensus motif for conjugation *35*

f) SUMO chains *37*

g) SUMO-interacting motif (SIM) *37*

Programmed genome rearrangement in *Tetrahymena thermophila* 39

I. Life cycle and nuclear dimorphism in *Tetrahymena* 40

II. Macronuclear differentiation and genome rearrangements 43

III. Small RNA-dependent heterochromatin formation in programmed genome rearrangements 45

1. Scan RNA model *45*

2. scnRNA biogenesis, TDSD and mechanism of action *47*

IV. SUMOylation in ciliates 51

Results- Research article 52

A SUMO E3 ligase is involved in target-directed small RNA degradation during DNA elimination in *Tetrahymena* 55

Abstract 55

Introduction 55

Materials and Methods 59

Strains and cell culture condition *59*

Construction of EMA2-HA strains *59*

Construction of SPT6-HA strains *59*

Construction of HA-SMT3 and His-SMT3 strains *60*

Construction of EMA2 Rescue strains *60*

Construction of wild-type and mutant HA-SPT6 strains *61*

DNA elimination assay by FISH *61*

RT-PCR analyses of lncRNA transcripts *62*

Viability test for progeny *62*

Small RNA analysis by northern hybridization *62*

Small RNA analysis by high-throughput sequencing *63*

GST pull down assay *63*

Immunoprecipitation for detecting SUMOylated proteins *64*

Immunofluorescent staining and western blotting *64*

Supplementary Table 1 *65*

Supplementary table 2 *66*

Results 68

Ema2 localizes to the parental MAC and the new MAC during conjugation 68

Ema2 is required for completing DNA elimination 68

Ema2 is important for heterochromatin formation and heterochromatin body assembly in the new MAC 69

Ema2 is required for heterochromatin formation in the parental MAC 70

Ema2 is important for Target-directed small RNA degradation (TDSD) 70

Ema2 is required for accumulation of SUMOylated proteins during conjugation 71

The conserved SP-RING domain is required for the function of Ema2 73

Ema2 is required for SUMOylation of Spt6 73

Ema2 is required for accumulation of parental MAC lncRNA transcripts 74

Ema2 is required for TDSD by promoting lncRNA transcription 76

Ema2 is required for the SUMOylation of the transcriptional regulator Spt6 77

References 79

Figures 87

Supplementary figures 97

Result- Methodology Paper 98

Discussion and prospective **115**

References 120

Table of Figures

Figure I. Mechanistic representation of Argonaute proteins 15
Figure II. *Drosophila*'s siRNA and miRNA pathway 21
Figure III. Schematic representation of piRNA. 23
Figure IV. The biogenesis pathway of piRNA in somatic and germ cells. 24
Figure V. Model of piRNA mediated heterochromatin formation and TE silencing. 26
Figure VI. Proposed model of ZSWIM8 CRL directed target-directed miRNA degradation. 28
Figure VII. SUMO maturation, deconjugation and chain editing by SENPs. 30
Figure VIII. SUMOylation cycle. 31
Figure IX. Structure/domains of PIAS protein. 34
Figure X. Phylogenetic tree of *Tetrahymena*. 39
Figure XI. *Tetrahymena thermophila* under florescence microscope. 40
Figure XII. Life cycle of *Tetrahymena*. 42
Figure XIII. Programmed genome rearrangement in *Tetrahymena*. 43
Figure XIV. Scan RNA model. 46
Figure XV. Molecular machinery of DNA elimination. 50
Figure XXV. Model for regulation of the TDSD of scnRNAs……115

Figure 1. Ema2 expression and localization. 87
Figure 2. DNA elimination and progeny viability in *EMA2* KO cells. 88
Figure 3. Heterochromatin formation in the new MACs of *EMA2* KO cells. 89
Figure 4. Heterochromatin formation in the parental MAC of *EMA2* KO cells. 90
Figure 5. TDSD in *EMA2* KO cells. 91
Figure 6. Ema2 acts as a SUMO E3 ligase. 93
Figure 7. Ema2 promotes SUMOylation of Spt6. 94
Figure 8. Ema2 is required for the accumulation of lncRNA transcripts from the parental MAC. 95
Figure 9. dsRNA localization in *EMA2* KO cells. 96

List of Acronyms

Ago	Argonaute
Aos1	Activation of Smt3 1
Aub	Aubergine
CBS	Chromosome breakage sequences
CLRC	Clr4-Rik1-Cul4 complex
CRL	Cullin-RING ubiquitin ligase
Dcr	Dicer
dsRNA	Double-stranded RNA
dsRBP	Double-stranded-RNA-binding protein
dsRBD	dsRNA binding domain
endo-siRNA	Endogenous siRNA
hpRNA	Hairpin RNA
HBs	Heterochromatin bodie
H3K9me3	Histone H3 and lysine 9 trimethylation
H3K27me3	Histone H3 and lysine 27 trimethylation
HMT	Histone methyltransferase
siRNA	Interfering RNA
IES	Internal-eliminated sequences
lncRNA	Long noncoding RNA
MDS	MAC-destined sequence
MAC	Macronucleus
MIC	Micronucleus
miRNA	microRNA
MID	Middle
NDSM	Negatively charged amino acids-dependent SUMO motif
NCS	Non-canonical sequences
sRNA	Small non-coding RNA
NMC	Non-maintained chromosomes
nt	Nucleotide
PDSM	Phosphorylation-dependent SUMO motifs

PAZ	PIWI-Argonaute-Zwille
piRNA	PIWI-interecting RNA
PTGS	Post-transcriptional gene silencing
PTM	Post-translational modification
PCNA	Proliferating cell nuclear antigen
PIAS	Protein inhibitor of activated STAT (signal transducer and activator of transcription)
RDC	Rhino (Rhi), Deadlock (Del), and Cutoff (Cuff)
RLC	RISC loading complex
RNAi	RNA interference
RdRP	RNA-dependent RNA polymerase
RDRC	RNA-dependent RNA polymerase complex
RISC	RNA-induced silencing complex
RITS	RNA-induced transcriptional silencing
scnRNA	Scan RNAs
SENP	Sentrin-specific proteases
STAT	Signal transducer and activator of transcription
SP-RING	Siz/PIAS RING
SUMO	Small ubiquitin-like modifier
SIM	SUMO-interacting motif
STUbls	SUMO-targeted ubiquitin ligases
TDMD	Target-directed miRNA degradation
TDSD	Target-directed small RNA degradation
TE	Transposable element
Uba2	Ubiquitin activating enzyme 2
Ubc9	Ubiquitin-conjugating enzyme 9/SUMO E2 ligase
UFD	Ubiquitin fold domain
Ubl	Ubiquitin-like

Abstract

Small, about 20-30 nucleotides RNAs regulate various biological processes in most eukaryotes. One of the conserved roles of small RNAs is to identify invaded selfish genetic elements and induce transcriptional silencing through heterochromatin formation. The ciliated protozoan *Tetrahymena* use such mechanism to identify transposon-derived sequences in the germline micronuclear (MIC) genome and remove them from the somatic macronuclear (MAC) genome during the sexual reproduction. During this DNA elimination process ~12,000 internal eliminated sequences (IESs), many of which are related to transposons, are removed from the new MAC. It has been revealed that IESs are targeted for heterochromatin formation followed by DNA elimination by ~29-nt small RNAs, called scnRNAs. Although scnRNAs are produced from both IESs and MAC-destined sequences (MDSs) of the germline genome at the early sexual reproduction stages, target-directed small RNA degradation (TDSD) process degrades scnRNAs complementary to MDSs in the parental MAC prior to DNA elimination. While TDSD is the key process to specifically target IESs for DNA elimination, the molecular mechanism for TDSD remains poorly understood. The aim of my PhD project is to better understand that how the TDSD process is regulated in *Tetrahymena*.

In my PhD study, I identified a novel gene, *EMA2*, which is exclusively expressed during sexual reproduction and required for completing DNA elimination. Northern blot and small RNA sequencing analysis revealed that TDSD is severely affected in the absence of *EMA2*. Ema2 protein possesses a SP-RING domain, which is shared among some SUMO E3 ligases, directly interacts with the SUMO E2 enzyme Ubc9 and is required for the accumulation of SUMOylated proteins during sexual reproduction, all indicating that Ema2 is a SUMO E3 ligase. I further found that Ema2 is necessary for the SUMOylation of the conserved transcriptional regulator Spt6 and for the genome-wide non-coding RNA transcription from the parental MAC that we previously proposed to be a prerequisite for TDSD. Altogether, I propose that Ema2p-dependent SUMOylation of Spt6p is necessary for the non-coding RNA transcription from the somatic genome and thus for TDSD.

Résumé

Les petits ARNs d'environ 20-30 nucléotides régulent divers processus biologiques chez la plupart des eucaryotes. L'un des rôles conservés des petits ARNs est d'identifier les éléments génétiques envahissseurs autonomes et d'induire leur mutisme transcriptionnel par la formation d'hétérochromatine. Le protozoaire cilié *Tetrahymena* utilise ce mécanisme pour identifier les séquences dérivées de la transposition dans le génome du micronoyau germinal (MIC) et les élimine du génome macronucléaire somatique (MAC) pendant la reproduction sexuée. Au cours de ce processus d'élimination de l'ADN, environ 12 000 séquences internes éliminées (IESs), souvent dérivées des transposons, sont ôtées du nouveau MAC. Il a été démontré que les IESs sont ciblées par des petits ARNs appelés scnRNAs (~29-nt) pour induire la formation d'hétérochromatine suivie par l'élimination de l'ADN. Cependant, les scnRNAs sont produits à partir des IESs et des séquences destinées au MAC (MDS) du génome de la lignée germinale au début de reproduction sexuée. Puis, le processus de dégradation dirigée par les petits ARNs (TDSD) élimine les scnRNAs complémentaires aux MDS dans le MAC parentale en amont de l'élimination de l'ADN. Malgré que le TDSD soit le processus clé pour cibler spécifiquement les IESs pour l'élimination de l'ADN, son mécanisme moléculaire demeure obscur. Le but de mon projet doctoral est de mieux comprendre comment le processus TDSD est régulé dans *Tetrahymena*.

Durant mon étude, j'ai identifié un nouveau gène, *EMA2*, qui est exclusivement exprimé pendant la reproduction sexuée et nécessaire pour compléter l'élimination de l'ADN. L'analyse par Northern blot et séquençage de petits ARNs a révélé que le TDSD est gravement affectée par l'absence d'*EMA2*. La protéine Ema2 possède un domaine SP-RING, partagé par une famille de ligases E3 SUMO, qui interagit directement avec l'enzyme Ubc9 E2 SUMO et qui est nécessaire pour l'accumulation de protéines SUMOylées pendant la reproduction sexuée; confirmant qu'Ema2 est un ligase E3 SUMO. J'ai également constaté qu'Ema2 est nécessaire pour la SUMOylation du régulateur transcriptionnel conservé Spt6 et pour la transcription globale du génome du MAC parentale en ARN non codant que nous avions précédemment proposé comme prérequis pour le TDSD. Au total, je propose que la SUMOylation de Spt6p dépendante d'Ema2p soit nécessaire pour la transcription du génome somatique en ARN non codant et donc pour le TDSD.

Acknowledgements

On successful completion of my PhD thesis, I would like to acknowledge the cardinal people who have always supported me to reach this phase of my life.

Firstly, I would like to extend my gratitude towards my PhD supervisor, Dr. Kazufumi Mochizuki, for his sincere guidance and utmost support in becoming an independent scientist. He have always shared his experience and made the environment so supporting that I had always craved to learn more. He is the one who furnished my basis in the sphere of RNA Biology and Epigenetics and introduced me to the exciting model organism *Tetrahymena*. I am thankful to Dr. Mochizuki for sharing his technical skills directly by himself on the bench by teaching experiments. Thus, I was a lucky PhD student.

I would also thank Dr. Clement Carre (IBPS-Sorbonne University, Paris), who being my sincere mentor, have always buffed me up to compete and ace the most competitive PhD position in the lab of Dr. Mochizuki. Clement have always spared time for me in his busy schedule whenever I handed out for his assistance. Indeed, he was the one who have introduced me to the field of RNA biology, and my journey of learning started under his mentorship.

I would also express my thanks to Dr. Severine Chambeyron and Dr. Hervé Seitz for their valuable suggestions and positive critique during our joint lab meetings.

Extending further, I would like to thank my thesis committee members, Dr. Robert Feil (IGMM, Montpellier) and Dr. Dimitris Xirodimas (CRBM, Montpellier) for their comments, appreciations, and timely guidance.

My Ph.D. would not be possible without funding. So, I would also thank "Labex EpiGenMed" for the initial three years funding of my PhD. I would also acknowledge "Fondation pour la Recherche Médicale" who have extended their support and trust in my passion and provided me the fourth year funding.

I also acknowledge the organizers of 16th Microsyposium on RNA biology 2022, 27th RNA meeting 2022 by RNA society, EMBO piRNA meeting 2022, and Ciliate Molecular Biology meeting 2021, for selecting my PhD project for oral presentation. The opportunity they provided through various conferences have positively modified me professionally and socially through numerous achievements.

I would also thank all the people of IGH and Genopolys for their support and understanding which helped me in maintain healthy work environment.

I thank all my lab fellows and friends who had worked in harmony and make this tiring journey worth remembering with loads of happy memories.
Finally, I would like to thank my family for their understanding and support. They have always pushed me to live my dream, achieve the acme despite them tolerating my absence for longer duration. Thank you all for your love and support.

Introduction

Small RNA degradation

I. The discovery of RNA interference (RNAi)

RNAi was first described in the model organism *Caenorhabditis elegans* (*C. elegans*) in 1993 (Wightman et al. 1993; Lee et al. 1993). It is a sequence-specific post-transcriptional gene silencing (PTGS) mechanism, in which a messenger RNA (mRNA) complementary to artificially-introduced double-stranded RNAs (dsRNAs) is targeted for decay (Fire et al. 1998). Later, it was shown that the introduced dsRNAs are processed to 21-22 nucleotide (nt) small interfering RNAs (siRNAs) that eventually triggers PTGS (Elbashir et al. 2001). Before the discovery of RNAi, it was known that transformation of some foreign DNA can leads to PTGS of its complementary endogenous sequence and it turned out that this was also an RNAi process triggered by the production of dsRNAs from the foreign DNA (Hamilton and Baulcombe 1999). RNAi has been a powerful gene knock-down tool in almost every model organisms (Cerutti and Casas-Mollano 2006).

Since the discovery of RNAi, number of small RNA classes have been identified, which include microRNAs (miRNAs) (Lagos-Quintana et al. 2001; Lee and Ambros 2001; Lau et al. 2001), small interfering RNAs (siRNAs) (Zamore et al. 2000; Hammond et al. 2000; Elbashir et al. 2001), and the Piwi-interacting RNAs (piRNAs) (Aravin et al. 2001; Vagin et al. 2006; Houwing et al. 2007; Aravin et al. 2006). All of them are small (~19-30 nt) non-coding regulatory RNAs (sRNA) that modulate gene expressions and/or protect the genome from internal or external threats (Ghildiyal and Zamore 2009). To execute effector functions, they follow the three key steps: (1) the biogenesis, (2) the loading and maturation, and (3) the recognition of the target. However, the different classes of sRNA use distinct molecular mechanism for these steps. In the section II, I will first discuss two key molecular machineries in the sRNA pathways: Argonaute and Dicer proteins. Then, I will overview the three steps of each sRNA class individually in the sections III.

II. Major component of RNA silencing pathway: Argonaute and Dicer Protein

a) Argonaute proteins

The all classes of sRNAs described above execute their functions by forming RNA-induced silencing complex (RISC) with Argonaute proteins. Argonaute protein was first described in *Arabidopsis thaliana* by studying the mutant *ago1*, in which leaves curled up like squid tentacles and thus the corresponding gene was named after the octopus *Argonauta argo.* (Bohmert et al. 1998). Argonaute proteins are divided into two clades: Ago and Piwi. Ago-clade Argonaute proteins are widely distributed to most eukaryotic linages whereas Piwi-clade Argonaute proteins are mainly found in animals. For instance, the fission yeast *S. pombe* has only one Argonaute protein that belongs to the Ago clade and *Arabidopsis thaliana* have 10 Argonaute proteins, all of which are in the Ago clade. On the other hand, there are five Argonaute proteins in the *Drosophila*, of which Ago-1 and Ago-2 belongs to the Ago clade and the other three, Aubergine (Aub), Argonaute-3 (Ago-3) and Piwi belong to the Piwi clade (Harris and Macdonald 2001; Cox et al. 2000; Saito et al. 2007; Vagin et al. 2006). C. elegans contains 27 genes related to Ago protein, including two related to PIWI (PRG-1 and PRG-2) (Wang and Reinke 2008).

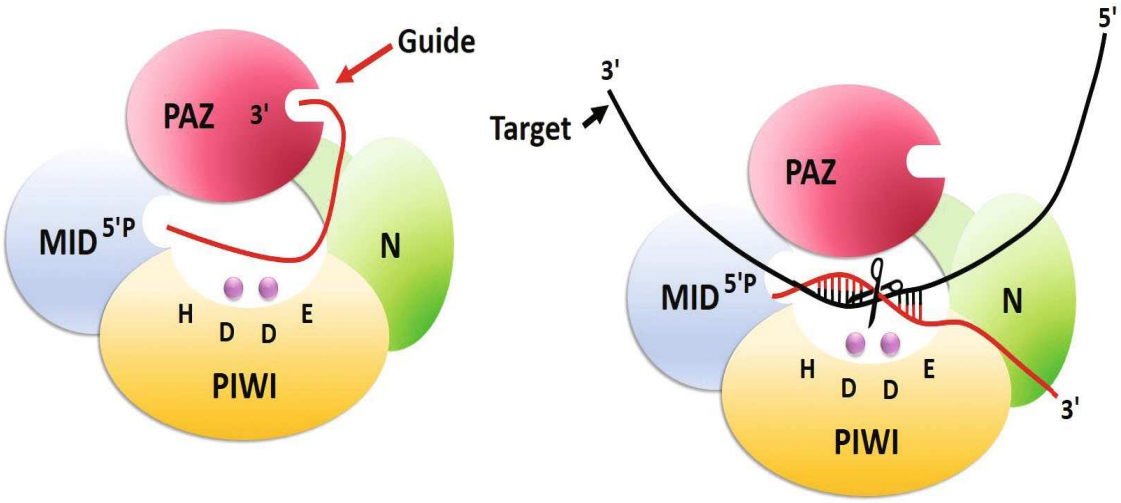


Figure I. Mechanistic representation of Argonaute proteins

Four distinct domains of Argonaute protein (PAZ, MID, N and PIWI) with the catalytic tetrad in the RNase H similar PIWI domain. In the left part, sRNA is loaded into Argonaute protein or in the right part the target RNA is bound with sRNA. The two purple dots represent two coordinated magnesium ions, which is important for the catalytic activity. Adapted from (Olina et al. 2018).

Argonaute proteins contains four characteristic domains: N-terminal, PAZ (PIWI-Argonaute-Zwille), MID (Middle) and PIWI. The channel formed by MID-PIWI and N-PAZ binds to nucleic acid (Figure. I) (Olina et al. 2018). The N-terminal domain of Ago protein have been shown to be required for assembling mature RISC by unwinding the duplex to remove the passenger strand by slicer-independent or silcer-dependent manner (Kwak and Tomari 2012). The 3'-end overhangs of sRNAs can be non-specifically anchored by the PAZ domain into the specific binding pocket of Ago protein (Höck and Meister 2008; Simon et al. 2011; Tian et al. 2011). The insertion of methylated piRNA 3'-end or siRNA/miRNA 3'-hydroxy into PAZ domain of Ago protein is essential to protect sRNAs from degradation (Tian et al. 2011; Park et al. 2017). The MID domain provides interaction site through conserved tyrosine in nucleotide binding pocket with the 5'-phosporelated terminal of guide RNA (Boland et al. 2011). PIWI domain is similar to ribonuclease (RNase H) and PIWI domains of some of the Argonaute proteins have the slicer activity (target RNA cleavage) towards complementary target RNA (Jinek and Doudna 2009).

b) Dicer protein

The biogenesis of miRNA and siRNA generally (with some exceptions) requires the dsRNA-specific RNase III family ribonuclease, the Dicer (Dcr) enzyme. In contrast, the biogenesis of piRNAs does not require Dcr (see section III "c") (Vagin et al. 2006; Brennecke et al. 2007). It was initially identified in mammals and *C. elegans* that the Dcr enzyme is required for the biogenesis of siRNAs (Bernstein et al. 2001; Knight and Bass 2001). The central role of Dcr and its associated proteins is to recognize long dsRNA substrates and process them to the sRNA sized products while translocating through the substrates (Lau et al. 2012). Dcr protein usually requires several protein cofactors to initiate the RNA silencing pathway. In *C.elegans*, more than 20 Dcr-associated proteins have been identified and they exist in at least two functional complexes which are responsible for distinct silencing pathways (Duchaine et al. 2006; Thivierge et al. 2011). Similarly, Dcr-1 in flies coordinates with R2D2 and helps siRNA loading into Ago2 (Liu et al. 2003). Canonical Dcr protein possesses the N-terminal helicase domain (DEX(D/H)), a PAZ domain, two RNase III domains, a DUF283 domain, and a C-terminal dsRNA binding domain (dsRBD). However, the domain composition varies among species (Ciechanowska et al. 2021).

The activity of the Dcr protein will be further discussed in the following section of sRNAs biogenesis.

III. sRNAs Biogenesis and mechanism of action

a) siRNA

RNAi canonical inducer is linear, long, and extensive base-paired dsRNA, taken directly from the environment or introduced into the cytoplasm (Mello and Conte 2004). This dsRNA are processed to siRNA by Dicer to direct silencing (Tomari and Zamore 2005; Meister and Tuschl 2004). Initially, RNAi was observed during transgene and virus-induced silencing in plants (Mello and Conte 2004). Transposons, centromeres, and repetitive sequences were shown to be the precursors of siRNA (Lippman and Martienssen 2004). Soon after, several other sources have been identified which include convergent mRNA transcripts, duplexes involving pseudogenes, natural sense-antisense pairs, and hairpin RNAs (hpRNAs) (Golden et al. 2008).

The loading and strand selection of siRNAs for RISC assembly is best characterized in humans and *Drosophila*. In flies, Dcr-2 forms the RISC loading complex (RLC) with R2D2, which loads the duplex of siRNA into Ago-2 (Cenik et al. 2011; Liu et al. 2003). The choice of guide strand depends on the thermodynamic stability of 5'-end. R2D2 binds to the siRNA ends with more stable ends and orients siRNAs as the strand which is less stable at 5'-end being the guide strand (Khvorova et al. 2003).

Ago-2 then cleaves the passenger strand (Matranga et al. 2005; Miyoshi et al. 2005; Rand et al. 2005) and the guide strand is 2'-*O*-methylated by the RNA methyltransferase PIMET (Hen1) (Saito et al. 2007; Pélisson et al. 2007; Horwich et al. 2007). Also, some protein components of the RNAi system can affect the choice of the strand and the loading symmetry depending on cell type, developmental stage, or various diseases (Schwarz et al. 2003; Han et al. 2004; Khvorova et al. 2003). In humans, even in the absence of the dsRNA trigger three major components, Dcr, TRBP, and Ago-2, also referred to as RISC-loading complex (RLC), associate with each other (Gregory et al. 2005; Maniataki and Mourelatos 2005). This trimer interacts with dsRNA, processes the duplex, loads siRNA into Ago protein, discards the passenger strand, and forms functional RISC (MacRae et al. 2008). Remarkably, mouse KO for Dcr can still load siRNA into

functional RISC (Murchison et al. 2005; Kanellopoulou et al. 2005), which demonstrates that Dcr is not essential in mammals for RISC loading.

In worms, secondary siRNA is formed by RNA-dependent RNA polymerase (RdRP), by synthesizing new complementary RNAs from existing RNA templates, when primary siRNA engages with the target (Mello and Conte 2004; Meister and Tuschl 2004). Some other organisms such as nematodes and plants can also produce the secondary siRNAs which can lead to systematic silencing that spreads throughout the organism. On the contrary, flies and mammals do not have such enzymes (Carthew and Sontheimer 2009).

After loading and maturation, in the canonical pathway of siRNA, the guide strand directs RISC to target complementary RNA for degradation. The PIWI domain of Ago protein has the slicer activity and cleaves the target very precisely at the phosphodiester linkage which is between the bases paired with the 10^{th} and 11^{th} residues of siRNA. This cleavage results in the generation of 3'-hydroxyl and 5'-monophosphate (Tomari and Zamore 2005). After the cleavage, the product is digested by cellular endonucleases (Orban and Izaurralde 2005). The target is dissociated from siRNA after cleavage and the free RISC can undergo another round of target search (Förstemann et al. 2007; Rivas et al. 2005).

In addition to posttranscriptional repression, siRNAs are also involved in transcriptional silencing in many eukaryotes (Lippman and Martienssen 2004). In the fission yeast *S. pombe*, the Ago-1-containing effecter is known as RNA-induced transcriptional silencing (RITS) complex, which is guided by siRNA to specific chromosomal loci such as pericentromeric regions. The existing model involves the interaction of RNA polymerase II and RITS (Djupedal et al. 2005; Kato et al. 2005), which facilitates siRNA to recognize nascent transcript (Bühler et al. 2006), this engagement of RITS to nascent transcript also activates the RNA-dependent RNA polymerase complex (RDRC), which uses Rdp1, an RdRP subunit to produce secondary siRNA precursors which are processed by Dicer protein (Dcr1) (Sugiyama et al. 2005; Motamedi et al. 2004), and load into Ago-1 to form additional RITS complexes (Volpe et al. 2002). On the heterochromatin regions, RITS and RDRC form a positive feedback loop to keep the RITS complex active. RITS also recruits CLRC (Clr4-Rik1-Cul4 complex) consisting of Su(var)3-9 homolog Clr4, which catalyzes H3K9me2/3 (Zhang et al. 2008). The HP1-like proteins Chp2 and Swi6, interact with

these marks and behave as docking platforms for chromatin remodelers that will silent siRNA-targeted sequences (Boots and Moll 2009).

TGS has also been studied in plants which shows many similarities with fission yeast. They also exhibit some differences, the key among them is the direct methylation of DNA through DNA methyltransferase in addition to histone methylation and the synthesis of siRNA precursors through DNA-dependent RNA polymerases, RNA Pol IV and Pol V (Herr et al. 2005; Onodera et al. 2005; Wierzbicki et al. 2008).

b) miRNA

In animals and plants, the miRNA precursors are encoded in their genomes (Ambros 2003). Primary miRNA transcript (pri-miRNA) is synthesized by RNA polymerase II. They are several hundred to thousand nucleotides in length and contain hairpins with about 10 nucleotides terminal loop (Cullen 2004; Saini et al. 2007). Then, pri-miRNA is processed by a microprocessor complex, which contains Drosha from the RNase III family and Pasha (in flies) or DGCR8 (in mammals) (Figure II) (Lee et al. 2002; Gregory et al. 2004). Drosha cleaved the pri-miRNA to liberate ~60-70-nt miRNA percussor (pre-miRNA) with a structure of hairpin and 2-nt single-stranded 3'-overhang, the hallmark of RNase III (Lee et al. 2006). The pre-miRNA is then transported from the nucleus to the cytoplasm by exoprtin-5 and Ran-GTP (Lund and Dahlberg 2006), where Dicer cuts out miRNA in a similar fashion to siRNA (Grishok et al. 2001; Hutvágner et al. 2001; Bernstein et al. 2001).

Double-stranded-RNA-binding proteins (dsRBPs) contribute to the loading of miRNA into the Argonaute (Chandradoss et al. 2015). In flies, both Dcr-1 and the dsRNA binding protein Loqs-PS, a protein isoform of Loquacious which have the characteristic of RNase III process the pre-miRNA to 22-23-nt miRNA:miRNA* duplex corresponding to the two sides of the base of the stem (Zamore et al. 2000; Lee et al. 2004; Bernstein et al. 2001). The duplex strands of miRNA contain the passenger (star *) strand and guide strand which follow similar thermodynamic criteria to siRNA for strand selection (Khvorova et al. 2003).

Plant's miRNA bears 2'-*O*-methylation at their 3'- extremity while flies do not have 2'-*O*-methylation for miRNA until it loads to Ago-2 (Yu et al. 2005; Horwich et al. 2007).

Typically, the mature miRNA downregulates the gene by binding to the 3' UTR of a particular target (Lai 2002; Filipowicz et al. 2008). The complementarity between miRNA and its respective target is generally imperfect, but the complementarity between the first 2-8 nt starting at the 5'-end of miRNA, so-called the seed region, and its target plays an important role in miRNA-target binding (Lewis et al. 2003), while additional pairing towered 3'-end of miRNA also plays some role (Grimson et al. 2007). On the other hand, the perfect or extensive complementarity between miRNA and its target leads to the degradation of miRNA known as Target-directed miRNA degradation (TDMD) (Shi et al. 2020; Han et al. 2020), which will be further discussed in the section (VI).

The mechanism by which miRNA regulates the target degradation or translational repression depends on the Argonaute protein into which miRNA is loaded. In flies majority of the miRNA are loaded into Ago-1, however, it has been shown that some miRNA:miRNA* duplex strands can be loaded into Ago-2 (Figure II) (Förstemann et al. 2007). The miRNA loaded into Ago-2 presents extensive complementarity like siRNA, begins with cytidine and cleaves the target. On the other hand, miRNA bound to Ago-1 typically begins with uridine and represses the translation without cleavage (Ghildiyal and Zamore 2009).

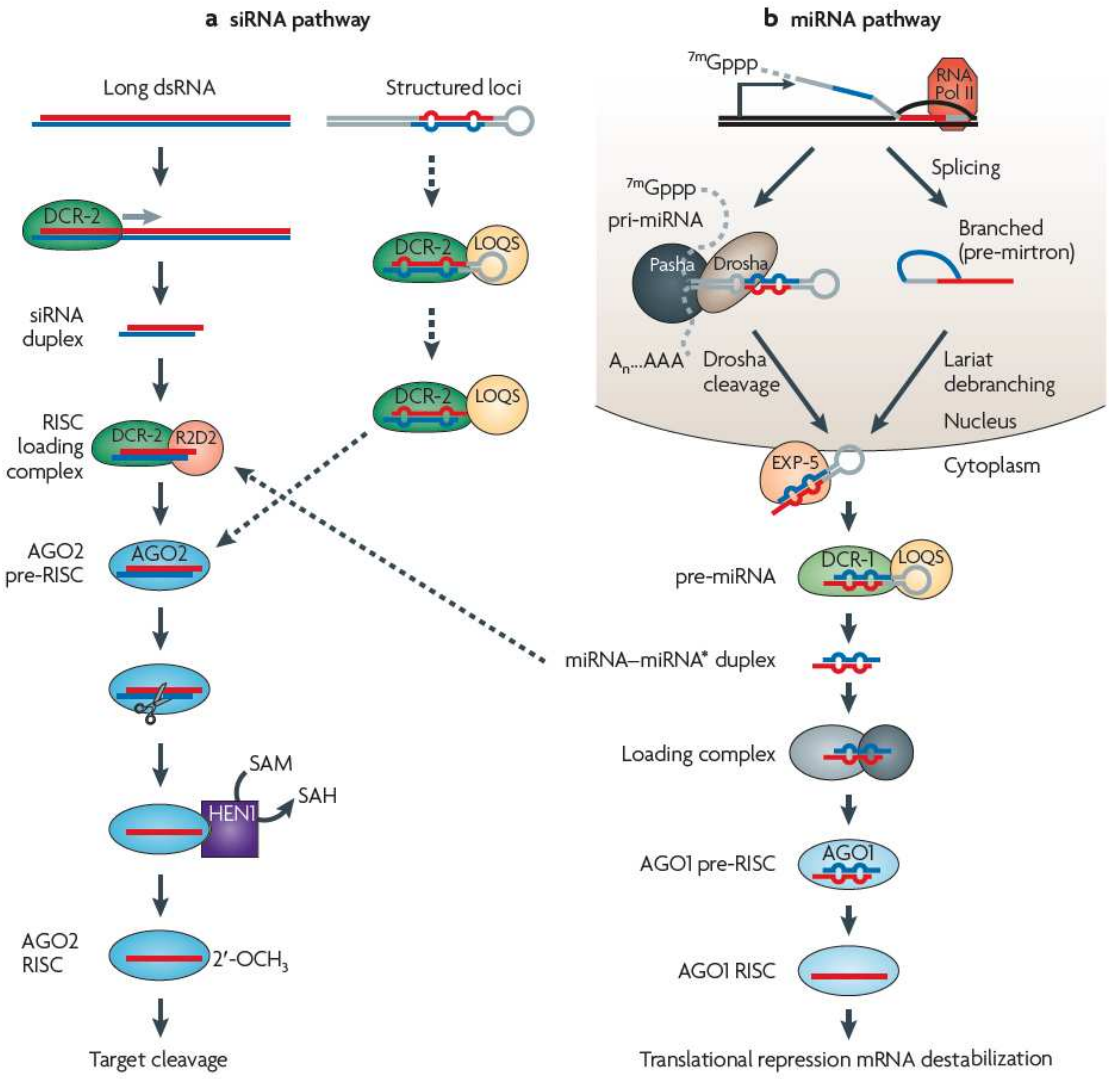


Figure II. *Drosophila*'s siRNA and miRNA pathway

a. Dcr-2 process the dsRNA precursor to produce siRNA duplex which contain the passenger strand and guide strand. The RISC-loading complex, Dcr-2 along with dsRNA-binding protein R2D2, loads siRNA into Ago-2. The endogenous siRNAs (endo-siRNAs) subset is dependent on dsRNA-binding protein Loquacious (LOQS) but not RSD2. The guide strand directs is 2'-*O*-methyalted by DmHen1 and direct Ago-2 to the target RNA while the passenger strand is destroyed. **b**. miRNAs are encoded in the genome and transcribe to the primary precursors of miRNA (pre-miRNA) transcript. Drosha cleaves the pre-miRNA to shorter pre-miRNA. Or-else, the miRNA which have introns (known as mitrons) are liberated following splicing to get authentic pre-miRNA. The Exp-5, exports pre-miRNA to the cytoplasm where Dcr-1 cleaves pre-miRNA to produce miRNA-miRNA* duplex. The passenger strand is destroyed, and the guide strand of miRNA directs Ago-2 to repress the translation of target RNA. SAM, S-adenosyl methionine; SAH, S-adenosyl homocysteine. Image adapted from (Ghildiyal and Zamore 2009).

c) piRNA

piRNA was first identified in flies as a class of repeat-associated small interfering RNAs (rasiRNAs) (~25-30 nt) (Aravin et al. 2001). Later, it was identified that they are loaded into Piwi clade Argonaute proteins and their biogenesis does not require Dicer unlike miRNA or siRNA (Pélisson et al. 2007; Vagin et al. 2006; Saito et al. 2006). Furthermore, it was found that piRNAs can be 2'-*O*-methylated at their 3'-ends in flies and mammals (Horwich et al. 2007; Saito et al. 2007).

In flies, more than 1.5 million distinct piRNAs were identified. They are collectively mapped to a few hundreds genomic clusters (Czech et al. 2008; Kawamura et al. 2008; Chung et al. 2008; Ruby et al. 2007), called piRNA clusters. piRNA clusters comprise repetitive sequences with a large portion of transposons (Brennecke et al. 2007). To protect the offspring and the gonad from further transposition, these transposon graveyards act as a memory of previous genomic invasion of transposons (Brennecke et al. 2008; Aravin et al. 2008). For example, *Flamenco* locus, the well-studied piRNA cluster, is important for suppressing *ZAM, Idefix*, and *gypsy* transposons in the genome (Prud'homme et al. 1995; Mével-Ninio et al. 2007; Pélisson et al. 2007). *Drosophila* piRNA clusters are divided into two subtypes: uni-stranded clusters and dual-stranded clusters. Uni-strand cluster is transcribed unidirectionally from a promoter located on one side of the cluster while the dual-stranded clusters do not have particular promoters at the ends but are transcribed on both strands (Brennecke et al. 2007; Malone et al. 2009).

The ovarian somatic cells are dependent on uni-stranded clusters for piRNA production (Malone et al. 2009). The uni-stranded piRNA transcript similarly undergoes RNA processing like mRNA, i.e., the addition of poly(A) and 5' capping, before being exported to the cytoplasm by specialized nuclear RNA export machinery NXF1/NXT1 to produce piRNA (Dennis et al. 2016). The transcript is then localized to "Yb bodies" in the cytoplasm of the somatic cells and cleaved by yet unidentified nuclease to get 5'- phosphate from the capped 5'-end (Olivieri et al. 2010; Yamashiro et al. 2020). The RNA helicase Armitage along with piRNA transcript and the PIWI-clade protein Piwi moves to the surface of mitochondria, later Piwi binds to the 5'-phosphate of the transcript where Zucchini (Zuc), an endonuclease processes piRNA intermediate to form 3'-end of piRNA and a new 5'-end to the intermediate (Figure 4) (Nishimasu et al. 2012). The 3'-end of piRNA is then 2'-*O*-methylated by Hen1 (Horwich et al. 2007; Saito et al. 2007; Kirino and Mourelatos 2007; Ohara et al. 2007).

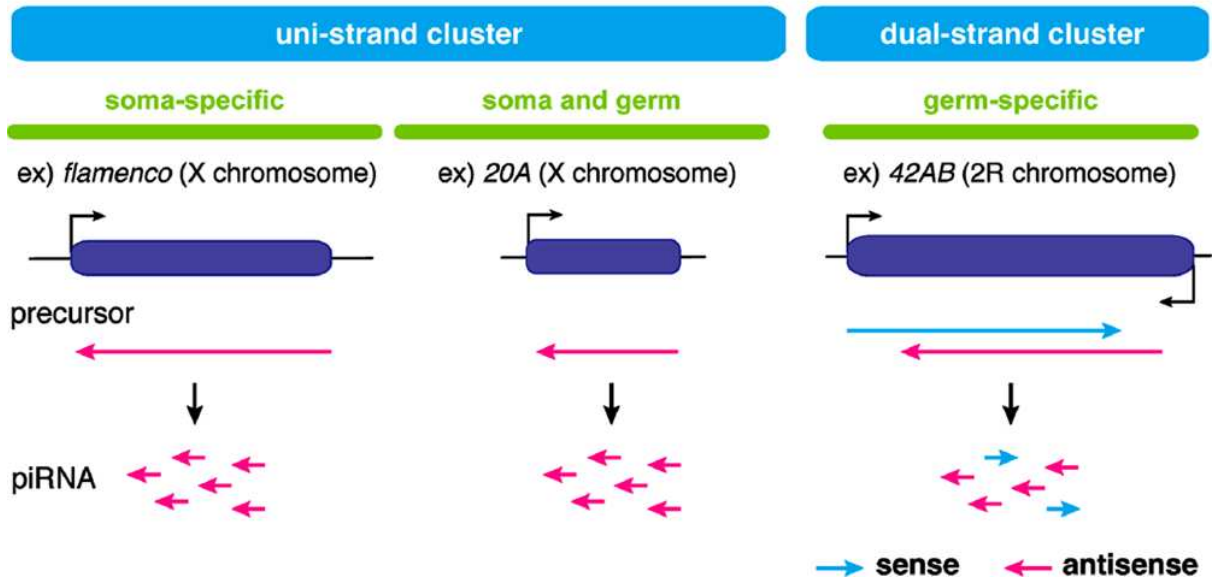


Figure III. Schematic representation of piRNA.

Uni-strand cluster produce piRNA in both somatic (e.g., *flamenco*) and germ cells (e.g., 20A). Dual-strand cluster is specific to germ cells (e.g., 42AB). Adapted from (Yamashiro and Siomi 2018).

Drosophila uses mostly dual-stranded clusters for piRNA production in germ cells (e.g., 42AB), (Figure III). In mammals, dual-strand clusters were not found (Aravin et al. 2007). Dual-strand clusters do not have defined promoters but the three nuclear piRNA factors Rhino (Rhi) , Deadlock (Del), and Cutoff (Cuff)-(RDC), play a major role in the biogenesis of piRNA from dual-stranded clusters by binding to H3K9me3 through Rhino, an HP1 homolog (Mohn et al. 2014; Klattenhoff et al. 2009). RDC recruits a transcription initiation complex that contains TFIIA (Moonshiner) and TBP (TRF2) proteins, implicated in the transcription of dual-strand piRNA cluster (Andersen et al. 2017). RDC also protects the uncapped 5'-end of the transcript (Zhang et al. 2014) and engages the transcript to the export pathway at the perinuclear structure known as "nauge" which is the counterpart of Yb bodies in germ cells (Mohn et al. 2014; Zhang et al. 2012, 2014). In the nauge, the transcript is loaded into Aub, then moved to the mitochondrial surface and eventually cleaved by Zuc (Figure IV) (Saito et al. 2010; Ge et al. 2019; Yamashiro et al. 2020).

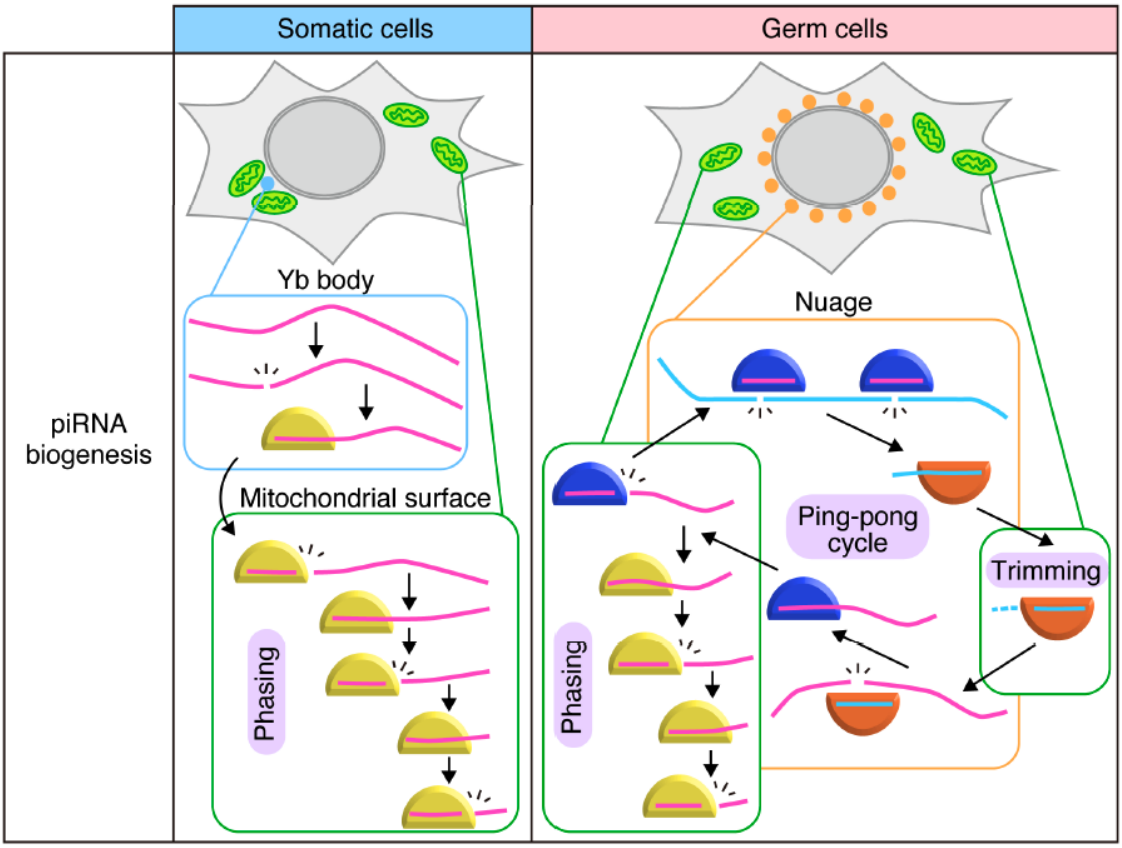


Figure IV. The biogenesis pathway of piRNA in somatic and germ cells.

In the periphery of nucleus, the uni-strand transcripts are translocated to the Yb bodies in the somatic cells and then move towards the surface of mitochondria to be processed by Zuc. In germ cells through ping-pong cycle in a feedback loop fashion, Aub and Ago-3 generate complementary pre-piRNAs. In the nauge, transposons-derived transcript (light blue) is cleaved and piRNA are amplified. piRNA becomes mature at the surface of mitochondria. Adapted from (Hirakata and Siomi 2019).

The secondary piRNA biogenesis process also takes place in the nauge of germ cells. There, transposable elements-derived (TE-derived) mRNA which is complementary to Aub-bound primary piRNA is cleaved. The cleaved TE transcript loads into the PIWI-clade protein Ago-3 and trim at the surface of mitochondria to be mature piRNA (Pane et al. 2007). The complementary piRNA cluster transcript which is loaded into a new Aub is sliced by this Ago-3-piRNA complex. This cycle function both as a PTGS against active transposable elements (TEs) and as a piRNA amplification loop, called a ping-pong cycle (Czech and Hannon 2016; Brennecke et al. 2007; Gunawardane et al. 2007).

IV. piRNA mediated heterochromatin formation and TE silencing

piRNAs mediated heterochromatin formation in Drosophila requires three steps: i) biogenesis of piRNA, ii) interaction of piRNA loaded Piwi protein with a nascent transcript from TE, and, iii) deposition of histone marks (H3K9me3) on TE to silence. In the germ cells, Aub and Ago-3 act in PTGS while Piwi acts predominantly in both somatic and germ cells (Le Thomas et al. 2013; Rozhkov et al. 2013; Webster et al. 2015; Izumi et al. 2020). Along with sterility and genome instability, a decrease of H3K9me3 on TEs in gonadal cells is reported in Piwi depletion (Malone and Hannon 2009; Siomi et al. 2011). Piwi-mediated silencing requires the recognition of complementary nascent transposons transcript with piRNA, which is accomplished with two Piwi interacting proteins, Asterix (Arx) and Panoramix (Panx) (Le Thomas et al. 2013; Dönertas et al. 2013; Ohtani et al. 2013; Sienski et al. 2015; Onishi et al. 2020). Panx is modified by SUMO (Andreev et al. 2022), a ubiquitin-like modifier (Gareau and Lima 2010). Panx SUMOylation occurs in a Piwi-dependent manner and this modification is essential for the interaction with Small ovary (Sov), a multi-zinc-finger protein, and hence for target silencing and heterochromatin formation (Andreev et al. 2022).

Su(var)2-10, a SUMO E3 ligase is required for transcriptional silencing, which links the piRNA complex to the silencing effector. With SUMO deposition, Su(var)2-10 recruits Eggless/SetDB1 histone methyltransferase (HMT) with the co-factor Windei complex (Ninova et al. 2020; Osumi et al. 2019). In flies, the Eggless/SetDB1 HMT is required for H3K9 methylation in piRNA-mediated silencing (Osumi et al. 2019), while Su(var)3-9 is not necessary for TGS of TEs. The current model proposes that on active transposons Piwi interacts with nascent RNA with help of Panx (Sienski et al. 2015; Yu et al. 2015), and Su(var)2-10 recruits the SetDB1/Wde complex for heterochromatin formation on TEs, and the H3K9me3 mark is bound by HP1 (Figure V) (Ninova et al. 2020; Maison and Almouzni 2004; Osumi et al. 2019; Mohn et al. 2014).

The piRNA pathway is involved in TE silencing in several other animals. In mammals, the piRNA pathway not only targets histone methylation but also DNA methylation (Aravin et al. 2008; Kuramochi-Miyagawa et al. 2008; Watanabe et al. 2018). In the model organism *Tetrahymena thermophila*, a homologous piRNA system exists, where Piwi-like protein and piRNA-like scnRNAs are involved in heterochromatin formation on TEs during sexual conjugation. I will discuss scnRNAs and piRNA-mediated heterochromatin formation of *Tetrahymena* in chapter 3.

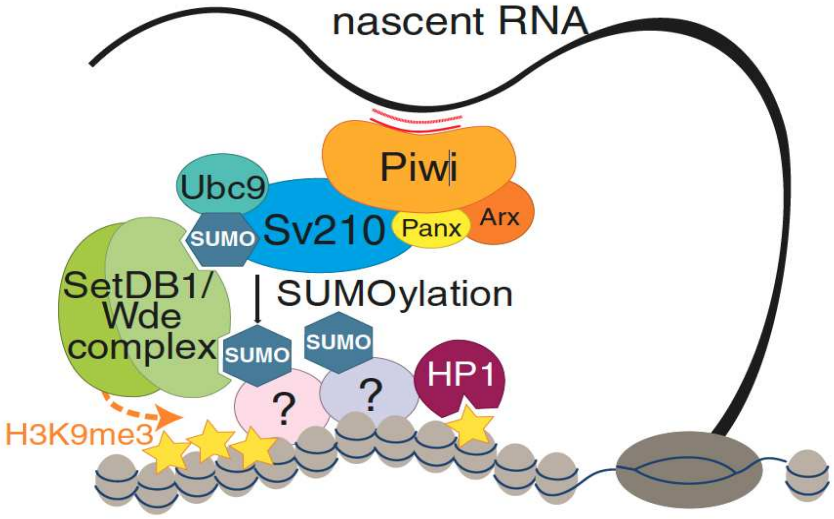


Figure V. Model of piRNA mediated heterochromatin formation and TE silencing.

Piwi-loaded piRNA interacts with the nascent RNA on active transposons with the help of Panx and Arx. Su(var)2-10 interact with Piwi-Panx-Arx complex and elevates SUMOylation of itself and yet unidentified chromatin factors. SetDB1/Wde complex is recruited by Su(var)2-10 through SIM (SUMO interaction motif) and deposit H3K9me3 mark at the target loci. Adapted from (Ninova et al. 2020a).

V. Target directed micro-RNA Degradation (TDMD)

Small regulatory RNAs, including miRNAs, are generally more stable than mRNAs, they can take days and sometimes weeks to get decay (Kingston and Bartel 2019; Reichholf et al. 2019; Eisen et al. 2020). This stability can be explained by the protection of miRNA from nucleases by the Argonaute protein (Sheu-Gruttadauria and MacRae 2017; Sheu-Gruttadauria et al. 2019). Nonetheless, the decay of some miRNAs is quick, with the range of <2 hours to 10 hours half-lives (Kingston and Bartel 2019; Reichholf et al. 2019; Rissland et al. 2011). This quick decay of some miRNAs is triggered by extensive base pairing between miRNA and its target, and the phenomenon is known as target-directed miRNA degradation (TDMD).

While TDMD was first observed by introducing artificial RNA targets with extensive pairing with miRNAs into cells (Krützfeldt et al. 2007, 2005; Ameres et al. 2010; Baccarini et al. 2011;

Xie et al. 2012), it is also identified that endogenous transcripts and viral transcripts lead to the degradation of miRNAs (Bitetti et al. 2018; Ghini et al. 2018; Lee et al. 2013; Kleaveland et al. 2018). For example, the long noncoding RNA (lncRNA) Cyrano, which has sequences extensively complementary to miR-7, induces the degradation of miR-7 (Kleaveland et al. 2018). In some cells and tissue, steady-state miR-7 level increases by 50-fold either by deletion of Cyrano or when a mutation is introduced into the cite of Cyrano that pairs extensively to miR-7 (Kleaveland et al. 2018; Kingston and Bartel 2019).

Typically, the 5' and 3'-end of miRNA are buried in Argonaute MID and PAZ domain which protects miRNA from tailing and trimming (Winter and Diederichs 2011; Wang et al. 2008; Schirle et al. 2014). The addition of non-templated nucleotide at the 3'-end of miRNA is known as tailing, which in turn recruits 3'-to-5' exoribonucleases including RNA exosome for RNA degradation (Chang et al. 2013; Ustianenko et al. 2013). While the removal of nucleotide by endonucleases from the 3'-end is known as trimming (Ameres and Zamore 2013). In the TDMD interaction, when miRNA engages with the highly complementary target then the 3'-end of miRNA is exposed for tailing and trimming (Sheu-Gruttadauria et al. 2019). In canonical interaction, miRNA interacts through their 5' seed sequences, and the 3'-end is buried in the PAZ domain and thus protected from tailing and trimming (Schirle et al. 2014). This high complementarity and thus exposure of the 3'-end of miRNA can be explained by 2'-*O*-methylation by Hen1, which protects other sRNAs (piRNA, endogenous siRNA, and plant's miRNA) from tailing and trimming (Han and Mendell 2022). In plants, miRNA exhibit extended 3'-end complementarities towards the target but are protected by 2'-*O*-methylation (Yu et al. 2005; Yang et al. 2006b), but interestingly in *Arabidopsis*, a family of 3'-to-5' exoribonucleases known as SDN1-3 can trim the 3'-end of miRNA irrespective of 2'-*O*-methylation (Ramachandran and Chen 2008; Yu et al. 2017; Chen et al. 2018).

Other conditions besides 3'-end tailing, and trimming can also cause miRNAs decay, such as TDMD mediated by ZSWIM8, a cullin-RING ubiquitin ligase (CRL) which interacts with Argonaute protein and promotes TDMD by directing miRNA-loaded complex through proteasomal degradation. Mechanistically, ZSWIM8 CRL along with possible associated proteins specifically ubiquitinylates Argonaute protein and directs towards proteasomal degradation while the miRNA-loaded complex releases miRNA to the cytoplasmic ribonucleases for degradation (Figure VI) (Shi et al. 2020; Han et al. 2020). ZSWIM8 CRL does not decay the target RNA but

subsequently releases from the miRNA-containing complex and recycling for the further round of miRNA turnover (Shi et al. 2020; Han et al. 2020).

The decay of miRNA through TDMD is well known but studies are not available for other sRNAs degradation. The ciliated protozoan *Tetrahymena* uses similar target-directed small RNA degradation (TDSD) for specifically targeting the germline-limited sequences for programmed DNA elimination by degrading Piwi-associated small RNAs complementary to the somatic genome. The aim of my PhD project is to understand how the TDSD is regulated in *Tetrahymena*, which will be further discussed in chapter 3.

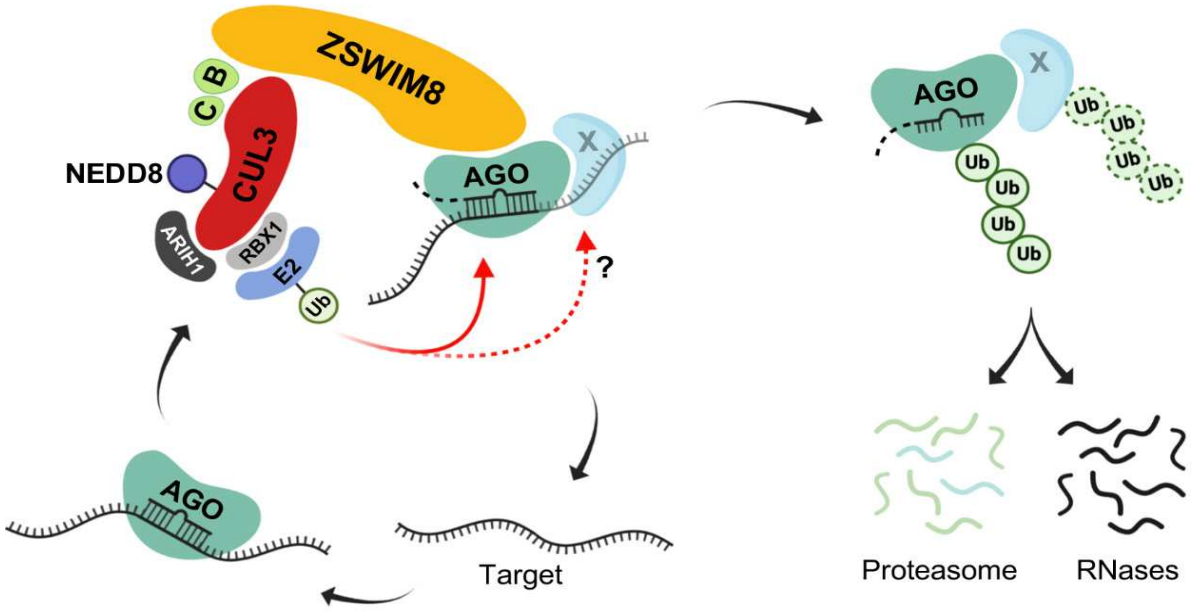


Figure VI. Proposed model of ZSWIM8 CRL directed target-directed miRNA degradation.

miRNA-containing Ago is specifically ubiquitinylate by ZSWIM8 complex along with other possible associated protein. The proteasomal degradation machinery degrades miRNA-containing Ago while miRNAs are released and possibly targeted by cytoplasmic RNases. The target transcript is released and recycled for additional round of TDMD. B, Elongin B; C, Elongin C; E2, Ubiquitin-conjugating enzyme; Ub, Ubiquitin. Adapted from (Han et al. 2020)

SUMOylation

I. The SUMO Protein

More than two decades ago, ~ 12 kDa protein which shows 18% homology with ubiquitin was co-discovered and termed Small-Ubiquitin-like Modifier (SUMO) (Wilson 2017). SUMOylation, the process by which SUMO is conjugated to a substrate, follows a similar enzymatic cascade as ubiquitylation by the sequential involvement of an activating enzyme E1, the conjugating enzyme E2, and an E3 protein ligase, although E3 ligase is not always required for SUMOylation (Kerscher et al. 2006; Capili and Lima 2007b). SUMOylation is a reversible post-translational modification (PTM) that regulates diverse cellular functions including DNA repair, transcription, replication, chromosome segregation, and heterochromatin formation (Johnson 2004; Geiss-Friedlander and Melchior 2007; Andreev et al. 2022; Ninova et al. 2020a). The substrate modified by SUMOylation can alter cellular localization, protein-protein interaction or activity of its target proteins (Gareau and Lima 2010).

The SUMO protein is highly conserved from unicellular eukaryotes to animals and plants (Chen et al. 1998). The budding yeast *Saccharomyces cerevisiae (S. cerevisiae), C. elegans* (Jones et al. 2002), and *Drosophila* (Lehembre et al. 2000) have single SUMO gene but there are eight SUMO genes in *Arabidopsis thaliana* (Kurepa et al. 2003) and humans have five SUMO paralogs, SUMO-1, -2, -3, -4 and -5 (Ryu and Hochstrasser 2021). Human SUMO-1, -2, -3, -5, and SUMO from yeast and flies are initially produced as immature precursors and processed by SUMO proteases (discussed in section II “a”) to form a mature SUMO for covalent isopeptide bonds with substrates. In contrast to other SUMO paralogues, the C-terminal tail of human SUMO-4 cannot be processed *in vivo* and therefore suggested that SUMO-4 may not be involved in covalent attachment to the substrates (Owerbach et al. 2005). SUMO-2 and 3 shares about 92% identity but only 48% identity with SUMO-1 (Kamitani et al. 1998a). Humans SUMO-1 shares 75% similarity and 48% identity with the SUMO of *S. cerevisiae* (which is called Smt3) (Huang et al. 2004). In contrast, the *Drosophila* SUMO shows more similarity to human SUMO-2 and -3 than other Human SUMOs (Smith et al. 2012) . The ciliate *Paramecium* has 3 SUMOs (Matsuda and Forney 2006), while *Tetrahymena* possesses only one SUMO which will be discussed in chapter 3 section IV.

II. SUMO pathway and The Enzymology of SUMOylation

In this section I will discuss the SUMOylation process in more detail.

a) SUMO proteases

Initially, SUMO is translated in a form of an immature precursor which is processed by sentrin-specific proteases (SENPs) to expose carboxy-terminal diglycine (Gly-Gly) motif (Figure VII and VIII) (Gareau and Lima 2010). SENPs can act both in the deconjugation and SUMO maturation process (Zilio et al. 2017). SENPs comprise Ulp1 and Ulp2 in budding yeast, Ulp1 and Ulp2 in *Drosophila*, and SENP-1, -2, -3, -5, -6, and -7 in mammals (Hickey et al. 2012; Nayak and Müller 2014) (Berdnik et al. 2012; Bhaskar et al. 2002). Different SENPs have different preferences towards SUMO paralogues and chain length (SUMO chains will be discussed in section "f") and can be involved in the process of deconjugation or maturation or both (Figure VII and VIII) (Nayak and Müller 2014).

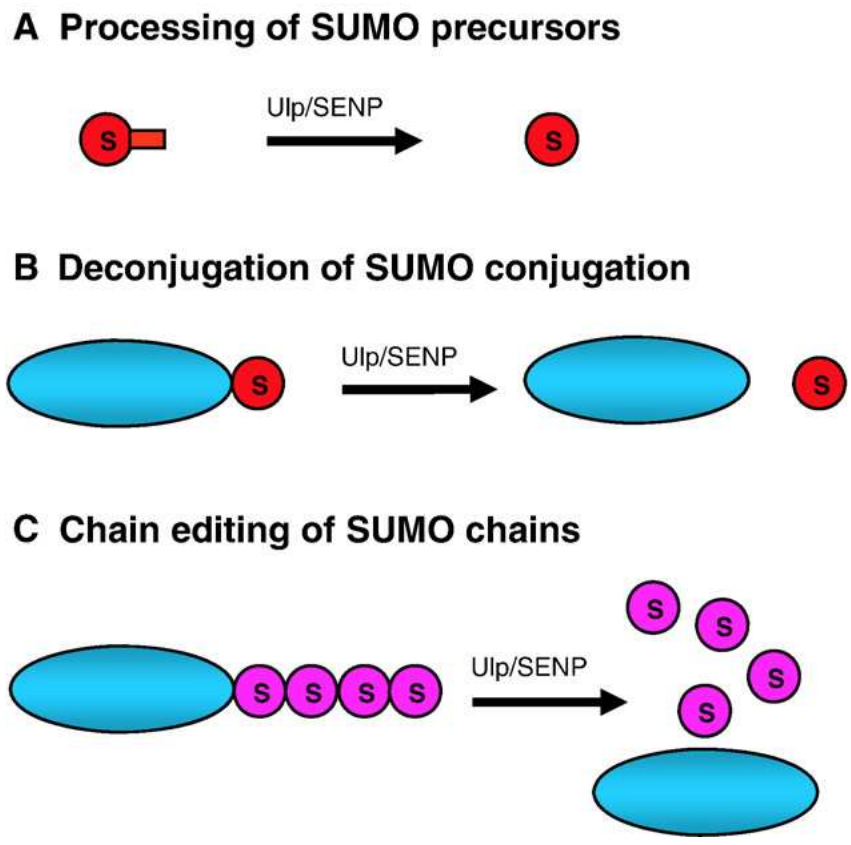


Figure VII. SUMO maturation, deconjugation and chain editing by SENPs.

(A) Ulp/SENPs process the nascent SUMO to a mature SUMO. (B) Ulp/SENPs deconjugate SUMO from the substrate. (C) Ulp/SENPs edits the SUMO chains. Adapted from (Kim and Baek 2009).

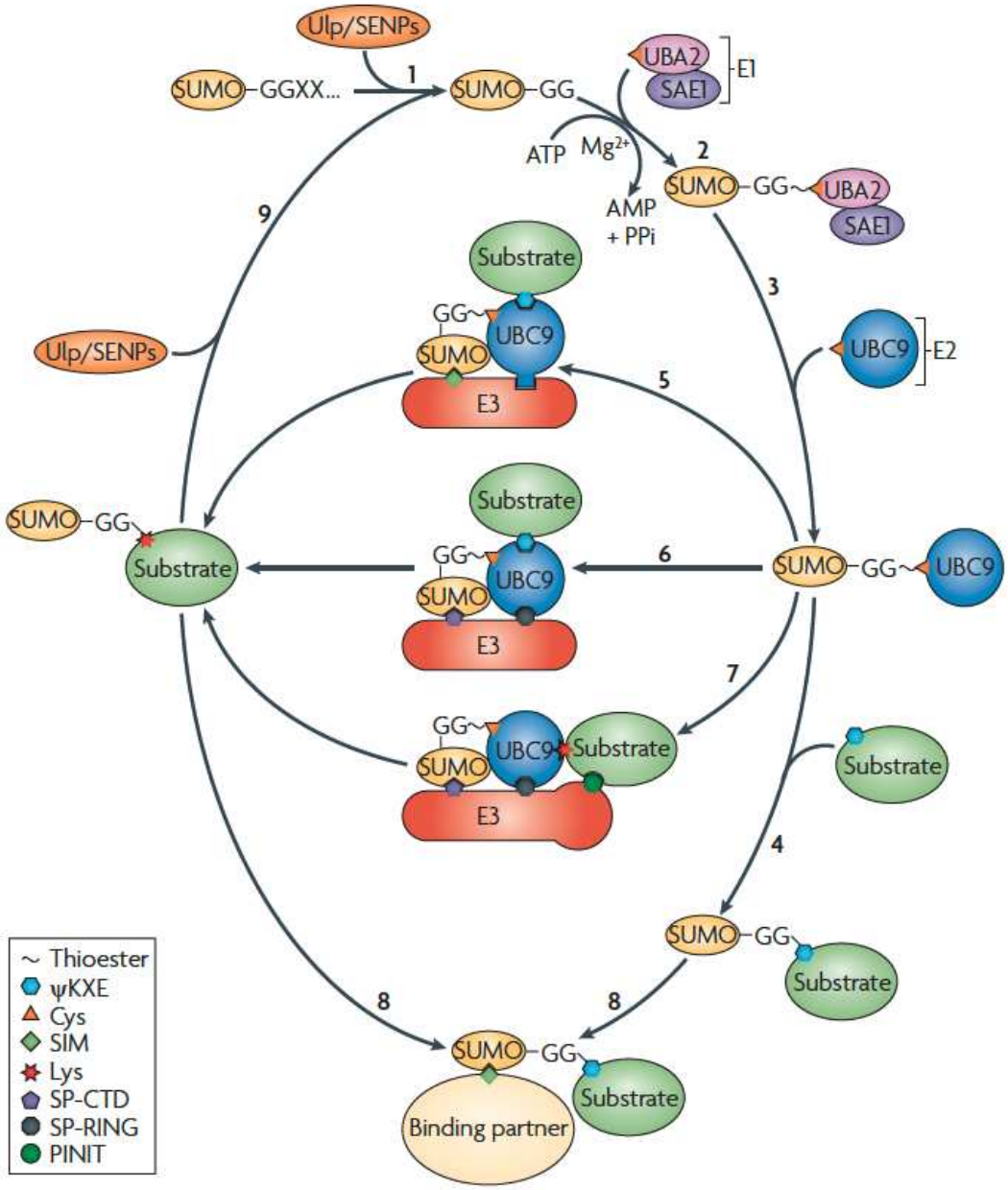


Figure VIII. SUMOylation cycle.

(1) SENPs process immature SUMO to SUMO-GG (mature SUMO), **(2)** SUMO is activated by E1 (UBA1/SAE1) in an ATP·Mg^{2+} dependent manner and transfer to the catalytic Cys of UBA2 subunit, **(3)** after the activation, SUMO is transfer to the catalytic Cys of Ubc9 (E2), **(4)** Ubc9 catalyze SUMO conjugation to substrate in an E3 independent manner, **(5-7)** in addition, SUMO E3 ligase can facilitate SUMO conjugation to substrates in different manners, **(8)** proteins modified by SUMO can interact SUMO binding proteins through their SIM, **(9)** SENPs deconjugate the SUMO from substrates and maybe recycle for another round of conjugation. Adapted from (Gareau and Lima 2010).

b) Heterodimeric SUMO E1 enzyme

Once a SUMO precursor is processed to a mature SUMO having the C-terminal diglycine, its C-terminus carboxy group is adenylated by heterodimeric the SUMO activating enzyme E1. The adenylated SUMO reacts with E1 to form a thioester bond between at a conserved cysteine (Cys) of E1 and SUMO (Figure VIII) (Gareau and Lima 2010). Then, the ubiquitin-like (Ubl)/ubiquitin fold domain (UFD) of E1 recruits SUMO conjugation enzyme E2 and transfer SUMO to catalytic cysteine of E2 (Lois and Lima 2005; Wang et al. 2010).

The SUMO E1 is a multi-domain heterodimeric protein, which consists of SUMO-activating enzyme 1 (Sae1 or Aos1) and 2 (Sae2 or Uba2) subunits. E1 possesses two catalytic activities requires for the activation of SUMO: the adenylation and the thioester bond formation (Desterro et al. 1999; Johnson et al. 1997; Okuma et al. 1999; Gong et al. 1999). The subunits of the E1 enzyme form a U-shaped heterodimeric complex with a large groove which have the ATP-binding motif at the base (Lois and Lima 2005).

c) SUMO conjugating enzyme E2/Ubc9

Although there are multiple E2 enzymes in the ubiquitin pathways, the SUMO pathway in most eukaryotes has only a single E2 conjugating enzyme (Wilson 2017; Tong et al. 1997). Furthermore, although the ubiquitin pathways require E3 ligases (Nakayama and Nakayama 2006), the SUMO E2 enzyme alone can transfer SUMO to targets for at least some substrates (Gareau and Lima 2010). SUMO E2 still requires SUMO E3 ligases for other substrates, which will be discussed in section “d”.

SUMO E2/Ubc9 has a conserved Ubc domain and is structurally shared among all E2/Ubc9 orthologs and ubiquitin E2 enzymes. Ubc domain consists of four α-helices and one anti-parallel β-sheet formed by four β-strands (αββββββ(ββ)ααα) (Tong et al. 1997; Matuschewski et al. 1996). Within around 150 amino acids domain, it contains the conserved cysteine which forms the thioester bond with the C-terminal glycine of SUMO (Olsen and Lima 2013). E2 targets lysine (K) residue(s) within a substrate for SUMOylation and form an isopeptide bond between the K and the C-terminal glycine of SUMO (Figure VIII)

d) SUMO E3 ligases

Unlike the ubiquitin pathway, SUMO E3 ligases are not required for many SUMOylation events and SUMO E1, and E2 along with ATP and maturated SUMO are sufficient *in vitro* for SUMO modification of many substrates. Also, although some SUMOylation occurs at the SUMO consensus motif, no consensus sequence has been found for ubiquitination. Based on these differences it had long been thought that the final ligation process was mechanistically different between ubiquitination and SUMOylation and E3 ligases were not required for substrate-specific SUMOylation (Hochstrasser 2001).

However, in 2001 three independent studies identified SUMO E3 ligases that greatly enhance SUMOylation both *in vitro* and *in vivo* (Kahyo et al. 2001; Takahashi et al. 2001b; Johnson and Gupta 2001). Johnson and Gupta characterized the septin-interacting protein Siz2 and its related protein Siz1 in *Saccharomyces* and found that the deletion of *SIZ1* severely decreased the SUMOylation of all three septins (Johnson and Gupta 2001). Takahashi et al and Kahyo et al found the E3 lieges with yeast two-hybrid system. Kahyo et al identified PIAS1 (Protein inhibitor of activated STAT (signal transducer and activator of transcription), which turned out to be similar to Siz1 and Siz2, as a potential E3 ligase for p53 in human cells (Kahyo et al. 2001; Hochstrasser 2001). Several other SUMO E3 ligases were later identified including RanBP2 (Pichler et al. 2002), Pc2 (polycomb 2) (Kagey et al. 2003), MMS21 (Potts and Yu 2005), and TOPORS (Weger et al. 2005; Hariharasudhan et al. 2022). Later, it became clear that SUMO E3 ligases enhance the attachments of SUMO to consensus motifs or redirect SUMO E2 to the non-consensus lysine (Lys) within a substrate. In the case of proliferating cell nuclear antigen (PCNA), where the SUMOylation of Lys 164 is strictly dependent on Siz1 but Lys 127 were found to be SUMOylated even in the absence of Siz1 both in vivo and in vitro (Yunus and Lima 2009).

The largest and most conserved family of SUMO E3 ligases is SP-RING (Siz/PIAS-really interesting new gene) domain-containing proteins, which contains, for example, PIAS-1, -2 (PIASx), -3 and -4 (PIASy) from human, Siz1, 2 and Pli1 from fungal species, Su(var)2-10 from *Drosophila* (Rytinki et al. 2009). Typical PIAS family protein contains conserved domains including N-terminal SAP ((scaffold attachment factor-A/B, acinus, and PIAS) domain, a PINT domain , an SP-RNIG domain, a SIM (discussed in section “g”) and a C-terminal serine/threonine rich (S/T) region (Figure IX) (Hochstrasser 2001; Minty et al. 2000; Duval et al. 2003; Tan et al. 2002), where most of human PIAS protein are conserved in these five domains but not the others

PIAS family protein (Rytinki et al. 2009). Among these domains, the SP-RING and SAP are more conserved while the S/T rich region is the least conserved within the PIAS family (Rytinki et al. 2009).

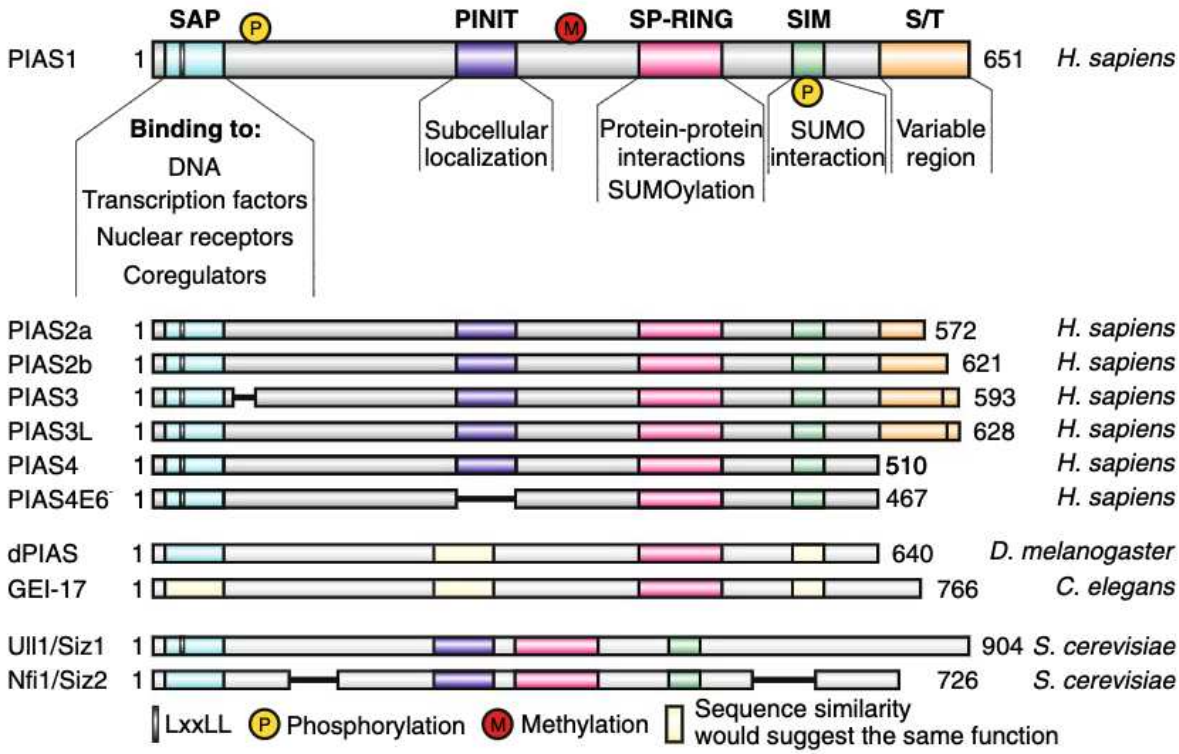


Figure IX. Structure/domains of PIAS protein.

The domains and PTMs of human PIAS protein are illustrated with the orthologs from *Drosophila*, *C. elegans* and *S. cerevisiae*. On the right side the number represents the amino acids. Adapted from (Rytinki et al. 2009).

The SAP domain exists in many proteins which are associated with chromatin and involved in structure or sequence-specific DNA binding (Aravind and Koonin 2000). Within the SAP domain, there is an LxLL motif which is found in proteins that interact with nuclear receptors (Liu et al. 2001; Heery et al. 1997). Next to the SAP domain in the C-terminus is a PINT motif, shown to be essential for the nuclear retention of PIAS3L (Duval et al. 2003). Most PIAS proteins harbor the N-terminal SIM motif, which binds non-covalently to SUMO. In the central part of the PIAS protein, there is a RING-type of zinc finger (SP-RING [Siz/PIAS RING]), which is necessary for the interaction with SUMO E2/Ubc9 for at least some PIAS proteins (Rytinki et al. 2009; Kahyo et al. 2001). It has been shown in many models that mutations in cysteine residues of the SP-RING

domain abolishes the SUMO ligase activity of PIAS proteins (Ninova et al. 2020a; Crozet et al. 2016; Kahyo et al. 2001; Munarriz et al. 2004; Takahashi et al. 2001a). For plant Siz1 as well as the mammalian PIAS3L, mutations in their SP-RING domain also disrupt their nuclear localization (Cheong et al. 2009; Duval et al. 2003).

e) SUMO consensus motif for conjugation

Typically, a SUMO substrate contains a stretch of a few amino acids known as the SUMO consensus motif. The canonical SUMO consensus motif is ΨKxE/D, in which Ψ represents hydrophobic amino acids (i.e., Phe, Val, Ile, Met or Leu with high preference for Ile and Val), K is the lysine where SUMO forms isopeptide linkage with the C-terminal glycine of SUMOs, x can be any amino acid, and the fourth is one of the acidic amino acids (Hay 2005). However, several studies found that many SUMOylated sites do not fall in the canonical ΨKxE/D motif but occur at non-canonical sequences (NCSs) (Kamitani et al. 1998b; Hoege et al. 2002; Rangasamy et al. 2000). For example, in yeast five out of ten SUMOylation events occur at NCSs (Zhou et al. 2004). These observations reveal that site selection complexity for SUMOylation is much greater than the ΨKxE/D motif. In addition to the fourth amino acid of the canonical ΨKxE/D motif, longer sequences that contains both canonical motifs and additional elements have been identified. Which includes, negatively charged amino acids-dependent SUMO motif (NDSMs)-ΨKxExxEEEE and phosphorylation-dependent SUMO motifs (PDSMs)- ΨKxExxSP while some substrates have inverted consensus motif-ExKΨ and hydrophobic cluster motif- ΨΨΨKxE (Gareau and Lima 2010; Song et al. 2005). The list of SUMO attachment motifs is increasing, currently, more than 10 SUMO motifs have been identified (Table 1) (Beauclair et al. 2015). Moreover, SUMO modification also occurs at NCS which is the case of about 50% substrates identified by Vertegaal's group (Hendriks et al. 2014).

	Name	Motif	Nb	%	References
Consensus direct	Stong consensus	$[\Psi_1]$-[K]-[x]-[α]	498	56.8%	Melchior et al., 2000 Rodriguez et al., 2001
	Consensus	$[\Psi_2]$-[K]-[x]-[α]	591	67.4%	
	Weak consensus	$[\Psi_3]$-[K]-[x]-[α]	598	68.2%	
	PDSM	$[\Psi_2]$-[K]-[x]-[α]-$[x]_2$-[S]-[P]	32	3.6%	Hietakangas et al., 2006
	NDSM	$[\Psi_2]$-[K]-[x]-[α]-[x]-$[α]_{2/6}$	231	26.3%	Yang et al., 2006
	HCSM	$[\Psi_4]_3$-[K]-[x]-[E]	105	12.0%	Matic et al., 2010
	SC-SUMO	[P/G]-$[x]_{(0-3)}$-[I/V]-[K]-[x]-[E]-$[x]_{(0-3)}$-[P/G]	110	12.5%	Benson et al., 2007
	Minimal SC-SUMO	[I/V]-[K]-[x]-[E]-$[x]_{(0-3)}$-[P]	178	20.3%	Subramanian et al., 2003
	SUMO-Acetyl switch	$[\Psi_2]$-[K]-[x]-[α]-[P]	130	14.8%	Stankovic-Valentin et al., 2007
	pSuM	$[\Psi_2]$-[K]-[x]-[$_p$S]-[P]	1	0.1%	Picard et al., 2012
Consensus inverted	Strong consensus	[α]-[x]-[K]-$[\Psi_1]$	30	3.4%	Ivanov et al., 2007 Matic et al., 2010
	Consensus	[α]-[x]-[K]-$[\Psi_2]$	77	8.8%	
	Weak consensus	[α]-[x]-[K]-$[\Psi_3]$	80	9.1%	
	Non consensus		229	26.1%	

	Name	Motif	Nb	%	References
SIM	Type 1	$[\Psi_1]$-$[\Psi_1]$-[x]-$[\Psi_1]$	63	61.8%	This study
	Type 2	$[\Psi_1]$-[x]-$[\Psi_1]$-$[\Psi_1]$	41	40.2%	
	Type 3	$[\Psi_1]$-$[\Psi_1]$-$[\Psi_1]$-$[\Psi_1]$	23	22.5%	
	Type 4	[x]-$[\Psi_1]$-$[\Psi_1]$-$[\Psi_1]$	31	30.4%	
	Type 5	$[\Psi_1]$-[α]-$[\Psi_1]$-[α]-$[\Psi_1]$	4	3.9%	This study; Ouyang et al., 2009
	Type α	[V/I]-[x]-[V/I]-[V/I]	24	23.5%	Song et al., 2004
	Type β	[V/I]-[V/I]-[x]-[V/I/L]	54	52.9%	
	Type a	[P/I/L/V/M]-[I/L/V/M]-[x]-[I/L/V/M]-$[α/S]_3$	17	16.7%	Miteva et al, 2010 Sun et al., 2012
	Type b	[P/I/L/V/M/F/Y]-[I/L/V/M]-[D]-[L]-[T]	24	23.5%	
	Type r	$[α/S]_3$-[I/L/V/M]-[x]-$[I/L/V/M/F]_2$	6	5.9%	
	Type H	[K]-$[x]_{3-5}$-[I/V]-[I/L]-[I/L]-$[x]_3$-[α/Q/N]-$[α]_2$	7	6.9%	Hannich et al., 2005
	Type M	$[\Psi_2]_2$-[x]-[S]-[x]-[S/T]-$[α]_3$	3	2.9%	Minty et al., 2000
	Non consensus		5	4.9%	

Table 1. List of SUMO consensus motif and SUMO-interacting Motif (SIM).

The upper table shows direct and inverted consensus motifs with name, residues of motifs identified with references. The lower table is the list of identifies SIMs. In the database (DB) the abundance of each motif is indicated by ((percentage (%) and number (Nb)). Ψ1 = I,L, or V; Ψ2 = V, A, I, F, L, M, P or W; Ψ3 = W, A, G, F, I, L, V, P, M or Y; Ψ4 = P, G, F, A, I, L or V; α = D or E ; pS/T = phosphorylated serine/threonine; h = hydrobobic amino acid; SIM: SUMO interacting motif. ND: not determined. Adapted from (Beauclair et al. 2015).

f) SUMO chains

Some of the SUMO paralogues, including human SUMO-2, -3, and Smt3 from *S. cerevisiae* and *S. pombe*, have the SUMO consensus motifs within the SUMO proteins. SUMO-2 and -3 possess K at position 11 which by itself can be SUMOylated and yield SUMO chains (Figure VII) (Klug et al. 2013; Gareau and Lima 2010). In contrast, human SUMO-1 and *Drosophila* Smt3 do not possess the critical K to be SUMOylated at their N-termini and are considered to be incapable to form poly SUMO chains (Matic et al. 2008; Tatham et al. 2001). However because some in vitro studies reported that SUMO-1 also can form SUMO chains (Yang et al. 2006a), the latter SUMO proteins may also form SUMO chains through atypical SUMOylation sites.

SUMO E2 enzymes are attached to SUMO not only covalently with a thioester bond but also non-covalently. It has been shown that this non-covalent Ubc9-SUMO interaction promotes SUMO chain formation (Knipscheer et al. 2007; Capili and Lima 2007a). SUMO chain formation can be enhanced also by SUMO E3 ligases (Vertegaal 2007; Pichler et al. 2002; Johnson and Gupta 2001) and counteracted by some SENPs (the process called chain-editing, see Figure VII).

g) SUMO-interacting motif (SIM)

SUMO-interacting motif (SIM) was first identified among the proteins that directly interact with SUMO-1-modified p73 (a p53 tumor suppressor) and later defined as (V/I)-x-(V/I) (V/I), which is typically flanked by acidic residues (Minty et al. 2000; Song et al. 2004; Hannich et al. 2005). For many SUMO-interacting proteins, disruption of their SIM motif abolishes the interaction with SUMO (Rytinki et al. 2009). Mechanistically, the β strand formed around SIM interacts with the β2 strand of SUMO (Sekiyama et al. 2008; Namanja et al. 2012). Similar to the SUMO consensus motif the list of SIM is expanding, and currently more than 10 different types of SIM have been reported (Table 1) (Beauclair et al. 2015). SIM-SUMO interactions play important roles in almost all types of biological processes (Andreev et al. 2022; Kerscher 2007; Song et al. 2005; Perry et al. 2008). For instance, human SUMO-targeted ubiquitin ligases (STUbls) RNF3 and RNF4 interact with multi SUMO or SUMO chains through their SIMs and induce SUMO-targeted ubiquitination (Tatham et al. 2008; Erker et al. 2013).

Like other eukaryotes, the SUMO pathway is conserved in the ciliate *Tetrahymena* (which will be discussed in chapter 3), which consists of only one SUMO (Smt3), a heterodimeric E1

enzyme (composed of Aos1/Uba2), and a single E2 (Ubc9). But until now, no study for SUMO E3 ligases and SENPs have been reported in this model. Mechanistically, very less is known about the SUMO system in the *Tetrahymena* that how it regulates programmed DNA elimination (discussed in chapter 3) and TDSD. One of the main goals of my PhD project is to study the role of SUMOylation, particularly the role of a SUMO E3 ligase in programmed DNA elimination and TDSD.

Programmed genome rearrangement in *Tetrahymena thermophila*

The ciliated protozoan *Tetrahymena thermophila* (hereafter called *Tetrahymena*) belongs to the superphylum of Alveolates within the domain Eukarya (Figure X). In wild, *Tetrahymena* live in fresh water and feed on bacteria but in the laboratory, we can culture *Tetrahymena* with inexpensive axenic media in which they grow with a doubling time of about 2-3 hours and reach high cell density (~1 x 10^6/mL). Moreover, the large cell size of *Tetrahymena* (~50 µm), together with the established genetics techniques for this organism, make *Tetrahymena* an excellent eukaryotic model for biochemical, genetics, and cell biological studies.

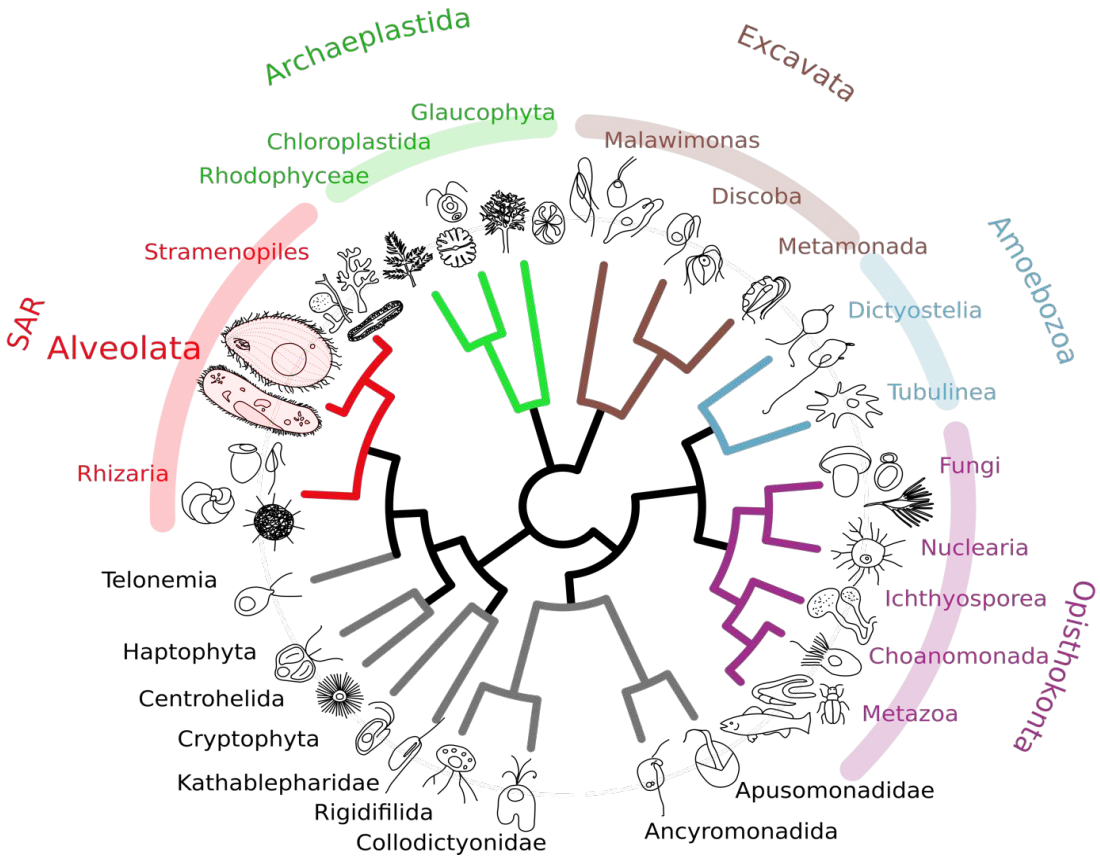


Figure X. Phylogenetic tree of *Tetrahymena*.

Within SAR supergroup, the ciliates *Tetrahymena* and *Paramecium* are placed in the same group of Alveolate. SAR also includes Rhizaria and Stramenopiles. Adapted from (Adl et al. 2012).

Some of the major discoveries using *Tetrahymena* as a model were recognized as Nobel prizes, which include the discoveries of the ribozyme (Cech et al. 1981) and telomerase (Greider and Blackburn 1985). Other breakthroughs made from research with *Tetrahymena* include the demonstration of the importance of histone acetylation in transcription (Brownell et al. 1996), the identification of motor proteins (dynein) (Gibbons and Rowe 1965), and the investigations of the roles of post-translational modifications (PTMs) of tubulins (Janke et al. 2005).

I. Life cycle and nuclear dimorphism in *Tetrahymena*

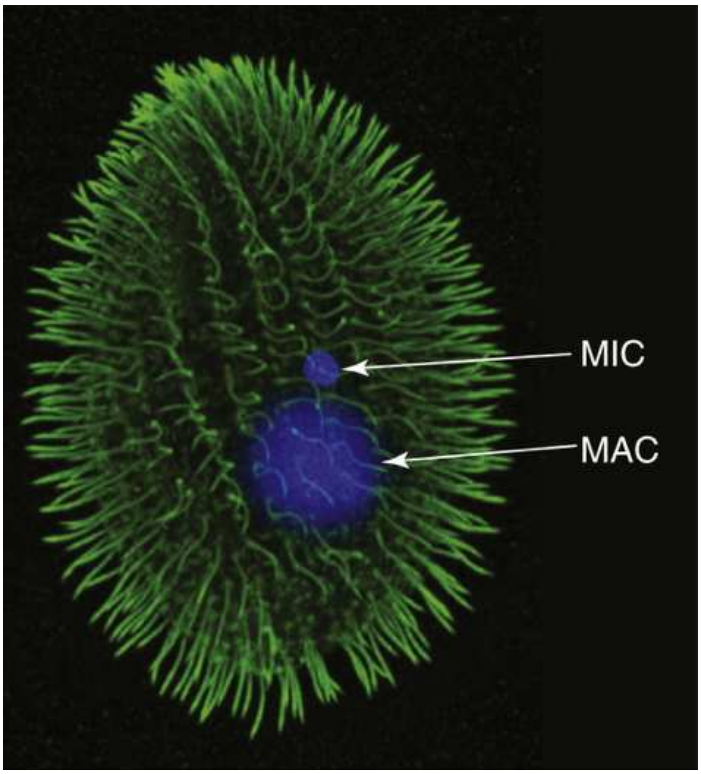


Figure XI. *Tetrahymena thermophila* under florescence microscope.

In green, anti-tubulin-stained cilia. In blue, the smaller is micronucleus (MIC) and the bigger is macronucleus (MAC) stained with DAPI. Adapted from (Goldfarb and Gorovsky 2009).

Like other ciliates, *Tetrahymena* possesses two nuclei in the same cytoplasm (Figure XI): the diploid germline micronucleus (MIC) and the polyploid somatic macronucleus (MAC) (Goldfarb and Gorovsky 2009). The MAC is the somatic nucleus that is dedicated to the maintenance of cell activities while the MIC is the germline nucleus that is responsible for heredity.

Besides their functional difference, the two *Tetrahymena* nuclei are different in several physical features. The MIC consists of 5 chromosome pairs with a genome size of ~200 Mb while the MAC is polyploid (around 90 copies) of ~250 chromosomes with a genome size of ~103 Mb (Figure XI) (Eisen et al. 2006; Hamilton et al. 2016; Sheng et al. 2020; Zhou et al. 2022). *Tetrahymena* maintain these differences through developmentally-programmed genome rearrangements and endoreplication during the sexual life cycle, where the MIC gives rise to the new MAC.

Tetrahymena acquired two life cycles. In the vegetative life cycle, the MAC divides amitotically in which replicated chromosomes are believed to be randomly segregated into two daughter nuclei. In contrast, the MIC divides mitotically and the sister chromosomes are precisely inherited to the daughters (Figure XIIA). In the sexual life cycle (termed conjugation), where cells having any two of the seven different mating types fused and exchange genetic information. In laboratory conditions, it is easy to induce conjugation by starving the cells and then mixing two strains having compatible mating types (Bruns and Brussard 1974). During conjugation (Figure XIIB), the diploid MIC undergoes meiosis and produces four haploid nuclei in each cell (Ray 1956; Tamm et al. 1975; Martindale et al. 1982), where three of the meiotic products are degraded and the selected one divide mitotically to produce two pronuclei in each cell. The two conjugated cells exchange pronuclei to produce zygotic nuclease, which mitotically divides twice to produce two new MICs and two new MACs. Simultaneously, the parental MACs are degraded in both cells and the conjugated cells are separated after ~12 hours post mixing (hpm) and give rise to two new daughter cells (Chalker et al. 2013; Davis et al. 1992).

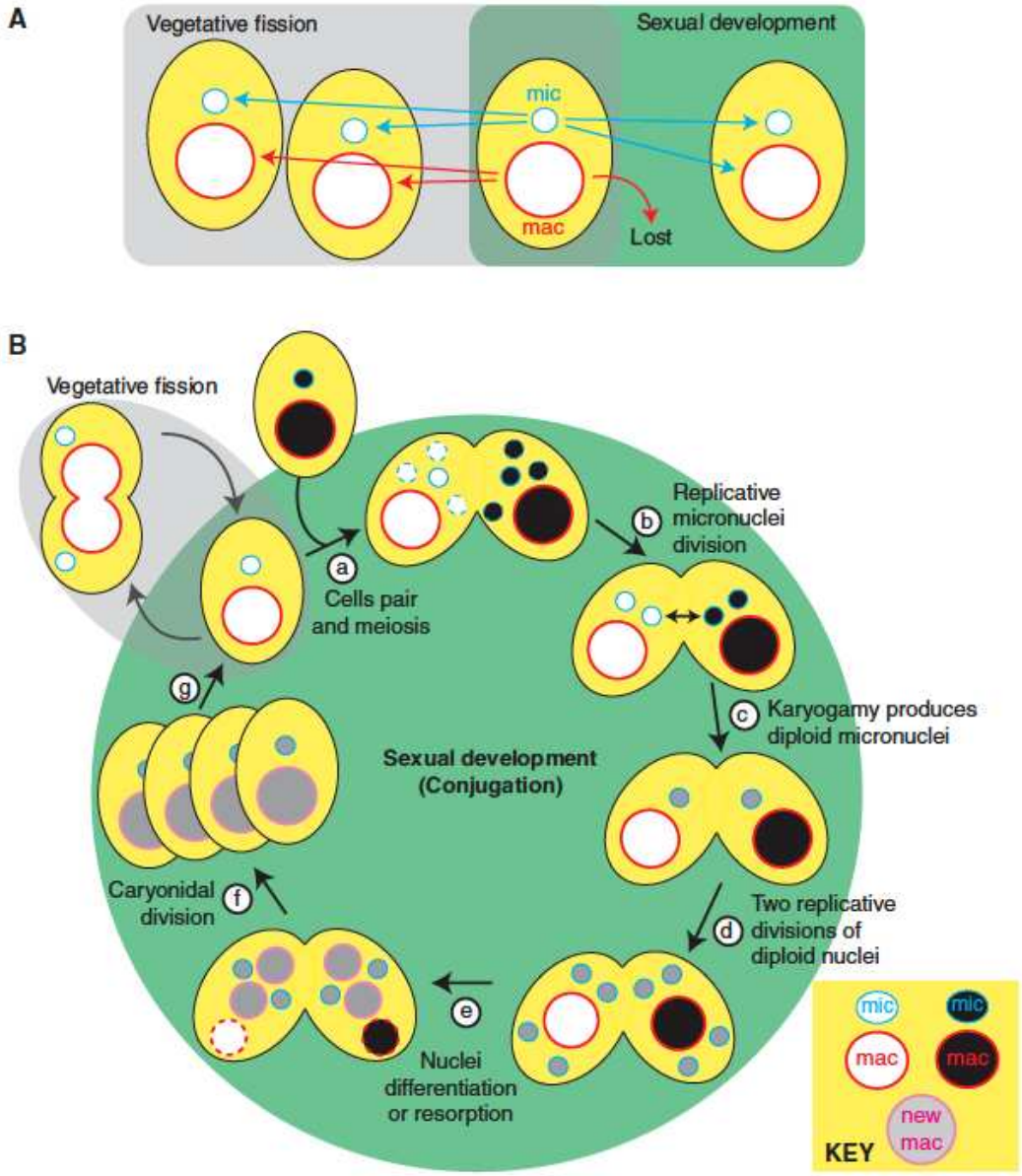


Figure XII. Life cycle of *Tetrahymena*.

A. Vegetative life cycle of *Tetrahymena*, by binary fission two cells are produce while in sexual development MIC produces new MIC and new MAC and the parental MAC degrades. **B**. during conjugation two cell pairs and MIC undergoes meiosis and produces 4 haploid nuclei (a), three of the meiotic product degrades and the selected nuclei in each cells divides mitotically to produce two pronuclei (b), which exchanges between the pairs to produce zygotic nucleus (c), zygotic diploid nucleus divide twice to produce to new MICs and two new MACs (d), the parental MACs are degraded at this stage (e), the conjugated cells separate from each other , give rise to four daughter cells (f), and resume vegetative growth (g). Adapted from (Chalker et al. 2013).

II. Macronuclear differentiation and genome rearrangements

Programmed genome rearrangement is shared among ciliates but here I will mostly focus on *Tetrahymena* and some examples from the close related ciliate *Paramecium tetraurelia*. As mentioned above, the MIC genome is ~200 Mb with 5 chromosomes and the MAC genome is ~103 Mb with ~250 chromosomes. Furthermore, it has also been demonstrated that the MIC makes the MAC, then how the genome of the MAC decreased by about 50 percent? What happens to the rest of the 50 percent genome? How from the 5 chromosomes of the MIC it become 250 in the MAC? The answers for all these questions can be given by understanding the two types of programmed genome rearrangements during conjugation.

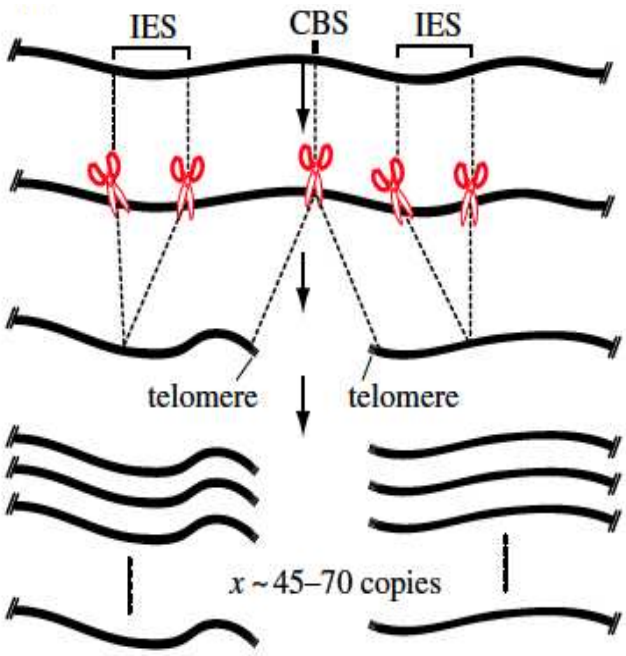


Figure XIII. Programmed genome rearrangement in *Tetrahymena.*

During the development of the new MAC, two events of programmed genome rearrangements occur. Chromosome breakage happens at CBSs and de novo telomere formation occurs. In parallel, IESs are removed by the process of programmed DNA elimination. The chromosomes in the new MACs are endo replicated to about 90 copies (According to (Zhou et al. 2022). Adapted from (Noto and Mochizuki 2017) .

The first type of genome rearrangement is chromosome breakage which occurs during the development of the new MAC at the conserved 15 bp sequences known as chromosome breakage sequences (CBSs) (Figure XIII) (Hamilton et al. 2006). The endonuclease which recognizes and

cuts the chromosomes at CBSs is yet to be identified. After the breakage, the *de novo* telomere formation happens. Pot2, a homolog of telomere binding protein, localizes to CBSs during the time of chromosome breakage and thought to be involved in *de novo* telomere formation (Cranert et al. 2014). The 5 chromosomes form the MIC is fragmented into ~250 chromosomes in the new MAC while approximately 50 of them are known as non-maintained chromosomes (NMCs), which are lost during the next ~20 rounds of the vegetative life cycle (Feng et al. 2017; Lin et al. 2016). The disappearance of NMCs is yet to be identified but some of the NMCs contain active genes (Hamilton et al. 2016; Feng et al. 2017).

The second type of genome rearrangement is programmed DNA elimination (DNA elimination) in which specific DNA fragments are eliminated from the newly formed MAC. The eliminated sequences are called internal-eliminated sequences (IESs) and the DNA retained in the MAC is known as MAC-destined sequence (MDS) (Figure XIII). After the elimination of the IESs, the MDSs are ligated. There are ~13,000 different IESs, ranging from ~0.2-40 kb, corresponding to about 97 Mb of the MIC genome, and distributed throughout the MIC genome mostly within non-coding regions (Hamilton et al. 2016; Zhou et al. 2022). Now the question is what are these IESs and why *Tetrahymena* removes them from the MAC genome?

The MIC genome contains about 19 Mb of transposable elements (TE)-related sequences. Among the classified TEs, 91% correspond to DNA transposons and only a small portion of 9% are retrotransposons. But overall, >95 % of IESs contain some TE-related sequnece and most of the TEs (>95%) are within IESs. During DNA elimination, about ~18 Mb of TEs is removed from the MAC genome but only a small portion of ~1Mb remains within the MAC genome and is thought to be a non-functional portion of TEs (Hamilton et al. 2016). Therefore, it is believed that *Tetrahymena* uses DNA elimination as a defense mechanism against the TEs to protect genome stability (Vogt and Mochizuki 2013; Coyne et al. 2012).

In *Tetrahymena* DNA elimination occurs reproducibly and precisely. Although some motifs have been discovered at the IES-boundaries they are sufficient to specify which region is to be eliminated (Lin et al. 2019; Steele et al. 1994). The next question is how DNA elimination is regulated. It is the communication between the MIC and the MAC through sRNAs known as scan-RNAs (scnRNA) of ~29 nt. How does it works? From where these scnRNAs are produced? How

scnRNAs are regulated and protect the genome from TEs will be discussed in the following section.

Note: In the literature related to ciliates or particularly *Tetrahymena*, "p" has always been used at the end of the protein, for example, Twi1p. However, I will use the short form without "p" in my thesis, i.e., Twi1.

III. Small RNA-dependent heterochromatin formation in programmed genome rearrangements

Since the discovery of Argonaute protein Twi1 (Piwi homolog in *Tetrahymena*) in 2002, a scnRNA model was proposed which explains how Twi1-loaded scnRNAs regulat programmed genome rearrangement (refers to DNA elimination) (Mochizuki et al. 2002). With time, many studies contributed to the current scnRNA model which will be discussed in this section.

1. Scan RNA model

During early conjugation of 2-3 hpm, Early-scnRNAs are produced from Type-A IESs, which consist of about 60% of ~13,000 IESs, and their surrounding regions (MDSs) of the MIC while the rest of IES, named Type-B IESs, do not produce scnRNAs (Figure XIV) (Noto et al. 2015). Type-A IESs are mostly located at the peri-centromeric and centromeric regions, known as A-regions, while Type-B IESs are in the chromosomal arms known as B-regions (Hamilton et al. 2016). The production of Early-scnRNAs is heterogeneous from the MIC genome. The production of Early-scnRNAs has two biases: the global bias is that the Type-A region preferentially produces Early-scnRNAs, and the local bias is that within the Type-A region, IESs produce Early-scnRNAs (Noto et al. 2015).

At 3-7 hpm the Early-scnRNAs are localized into the parental MAC, where the Early-scnRNAs which are complementary to the MAC genome are degraded and the others which are complementary to IESs are retained. (Mochizuki et al. 2002; Aronica et al. 2008; Schoeberl et al. 2012; Noto et al. 2015; Mochizuki and Gorovsky 2004c). This degradation process of scnRNAs was previously called scnRNA-selection but I will use the term TDSD in this manuscript.

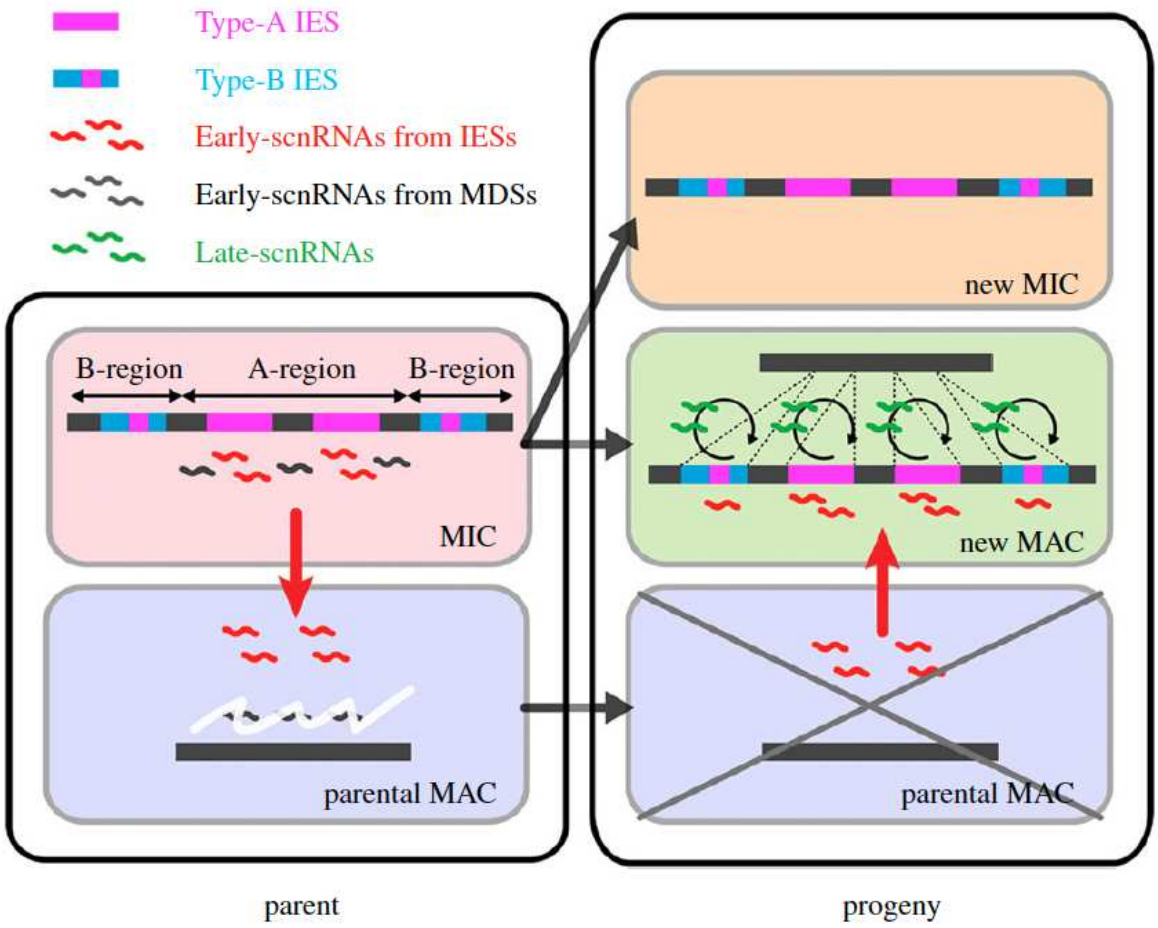


Figure XIV. Scan RNA model.

In the parental MIC, Early-scnRNAs are produce from A-region of IESs and the surrounding region (MDS). Then localizes to the parental MAC, where Early-scnRNAs derived from MDS are degraded through TDSD and those complementary to IESs retain (left). In the new MAC, the selected Early-scnRNAs recognizes not only Type-A IESs but also Type-B IESs trough the repetitive sequences common between the both types of IESs. This recognition induces the formation of Late-scnRNAs which facilitate DNA elimination. The parental MAC are degraded at the end of conjugation (right). Adapted from (Noto and Mochizuki 2017).

After TDSD, at about 7-8 hpm the Early-scnRNAs are localized into the new MAC, where the scnRNA-directed heterochromatin formation occurs on Type-A IESs (Liu et al. 2004, 2007; Mochizuki et al. 2002). Type-B IES contains repetitive sequences that are shared with Type-A IESs. Therefore, Early-scnRNAs produced from Type-A IESs can interact with Type-B IESs in *trans* to induce heterochromatin formation (Noto et al. 2015). Moreover, the Early-scnRNAs induce the production of secondary scnRNAs of ~29 nt in *cis* from both Type-A and Type-B IESs. These secondary scnRNAs are called Late-scnRNAs. Late-scnRNAs further induce heterochromatin formation and an RNAi-heterochromatin positive feedback loop is formed (Noto et al. 2015). Eventually, the heterochromatinized IESs are excised by a domesticated piggyBac transposase, Tpb2 (Vogt and Mochizuki 2013; Cheng et al. 2010). The biogenesis of scnRNA, loading, maturation, the TDSD, and scnRNA-induced heterochromatin formation will be explained in the following sections.

2. scnRNA biogenesis, TDSD and mechanism of action

The molecular mechanism of scnRNAs biogenesis and turnover is illustrated in Figure XV. Except for the meiotic prophase, the MIC is transcriptionally inactive during all stages of life (Sugai and Hiwatashi 1974; Martindale et al. 1985). In early conjugation (2-3 hpm), the MIC genome is transcribed bi-directionally (Figure XV*a*) (Chalker and Yao 2001; Aronica et al. 2008), and the dsRNA is processed to Early-scnRNAs by Dcl1, a Dicer homolog in *Tetrahymena*, within the MIC (Figure XV*b*) (Malone et al. 2005; Mochizuki and Gorovsky 2005). The two biases (local and global) in the pattern of the Early-scnRNAs production mentioned above in the scan RNA model are determined at the level of transcription (Schoeberl et al. 2012). Rpb3, an RNA Pol II subunit, localizes to the MIC during meiotic prophase, while a shared subunit of RNA Pol I and III, Rpc5 does not. Thus it is believed that the MIC is transcribed by RNA Pol II (Mochizuki and Gorovsky 2004b).

The Early-scnRNAs are then transported to the cytoplasm by a yet unknown mechanism (Figure XV*c*) and loaded into Twi1, a Piwi homolog in *Tetrahymena* (Mochizuki and Gorovsky 2004a; Noto et al. 2010). The loading process also requires Hsp90 and its co-chaperone Coi12 (Figure XV*d*) (Woehrer et al. 2015). Like some other Ago proteins, Twi1 possesses an endonuclease activity that slices out the passenger strand of the Early-scnRNAs (Figure XV*e*) (Noto et al. 2010). Then the guide strand loaded-Twi1 complex interacts with Giw1 and is

transported to the parental MAC (Figure XV*f*) (Noto et al. 2010). It is unclear how Giw1 links the Twi1-scnRNAs complex to the parental MAC-specific nuclear transport machinery.

In the parental MAC, the Early-scnRNAs that are complementary to the parental MAC are degraded in a similar manner to TDMD and the Early-scnRNAs complementary to IESs are retained (Figure XV*h*) (Mochizuki and Gorovsky 2004a; Aronica et al. 2008; Schoeberl et al. 2012). Previously the term "scnRNA selection" was used in the literature, but now we called this process Target-directed small RNA degradation (TDSD). The Early-scnRNAs are 2'-*O*-methylated by Hen1 (Figure XV*g*) (Kurth and Mochizuki 2009). This modification does not seem to be involved in the TDSD of scnRNAs process because TDSD does occur even in the absence of Hen1 (Noto and Mochizuki 2017).

The putative RNA helicase, Ema1 was shown to be required for TDSD by promoting the interaction between nascent long non-coding RNAs (lncRNA) transcript and Piwi-loaded Early-scnRNAs (Aronica et al. 2008). The requirement of nascent lncRNAs transcripts was also shown for scnRNAs selection in *Paramecium* (Lepère et al. 2008). Wag1 and Cnjb were also shown to be necessary for TDSD of scnRNAs in *Tetrahymena* (Bednenko et al. 2009). Both proteins contain GW repeats, called the AGO hook, which is characteristic of Ago-interacting proteins (Trujillo and Mosher 2017). TDSD of scnRNAs is an important process for scnRNA selection and eventually DNA elimination, but the molecular mechanism is unknown. The aim of my PhD project is to study which factors or modifications are required for the regulation of nascent lncRNA transcription and thus for the TDSD of scnRNAs in *Tetrahymena*.

In parallel to the TDSD of scnRNAs, the MIC undergoes meiosis and subsequently, the zygotic nucleus forms the new MIC and the new MAC. The Early-scnRNAs that are not degraded by TDSD move from the parental MAC to the newly formed MAC (Figure XV*i*) (Mochizuki et al. 2002; Noto et al. 2010). In the new MAC, the Twi1-loaded Early-scnRNA complex interacts with nascent lncRNA transcript with the help of Ema1 (Figure XV*j*) (Aronica et al. 2008). By a yet unknown mechanism, this interaction recruits Ezl1, a histone methyltransferase (Figure 1*k*), to methylate histone H3 lysine 9 and 27 (H3K9/H3K27me3) on Type-A IESs and A-repeats on Type-B IESs, and the HP1-like protein Pdd1 binds to these methylated histones (Figure XV*l*) (Aronica et al. 2008; Noto et al. 2015; Liu et al. 2004, 2007).

In the new MAC, the Late-scnRNAs are produced through Ezl1 and Pdd1-dependent manner (Figure XV*m*) and loaded into two Ago proteins, Twi1 and the Twi11 (Figure XV*n*) (Noto et al. 2015). The Late-scnRNA biogenesis and heterochromatin formation on IESs are facilitated by a positive feedback loop between heterochromatin and the Late-scnRNAs (Figure XV*o*) and eventually induce DNA elimination but this excision must be controlled at the border of IESs. Several boundary protection factors including Coi6 (HP1-like), Coi7, Lia5 (piggyBac transposase), and Jmj1 (H3K27 demethylase) downregulate the heterochromatin-Late-scnRNAs feedback loop (Figure XV*p*) (Vogt and Mochizuki 2013; Noto and Mochizuki 2017). The heterochromatinized IESs form heterochromatin bodies (HBs) by assembling IESs into large nuclear foci. Pdd1 phosphorylation and subsequent dephosphorylation are required for establishing HBs (Kataoka et al. 2016; Kataoka and Mochizuki 2015). RNA-Pdd1 interaction is also promoted by the dephosphorylation of Pdd1 (Kataoka and Mochizuki 2015). Therefore, it was proposed that RNA may aid to assemble multiple heterochromatinized IESs to HBs by acting as a glue. Jub6, an RNA-binding aggregation-prone protein, also promotes the formation of HBs, suggesting that RNA-mediated phase separation may play also a role in the formation of HBs (Kataoka and Mochizuki 2017).

In the final step of the DNA elimination pathway, nearly all IESs are excised by Tpb2. But the exceptional 12 IESs, some of which are found in exons, are precisely excised by Tpb1 and Tpb6. In the both cases, the non-homologous end joining machinery is required for the repair (Vogt and Mochizuki 2013; Cheng et al. 2010; Lin et al. 2012).

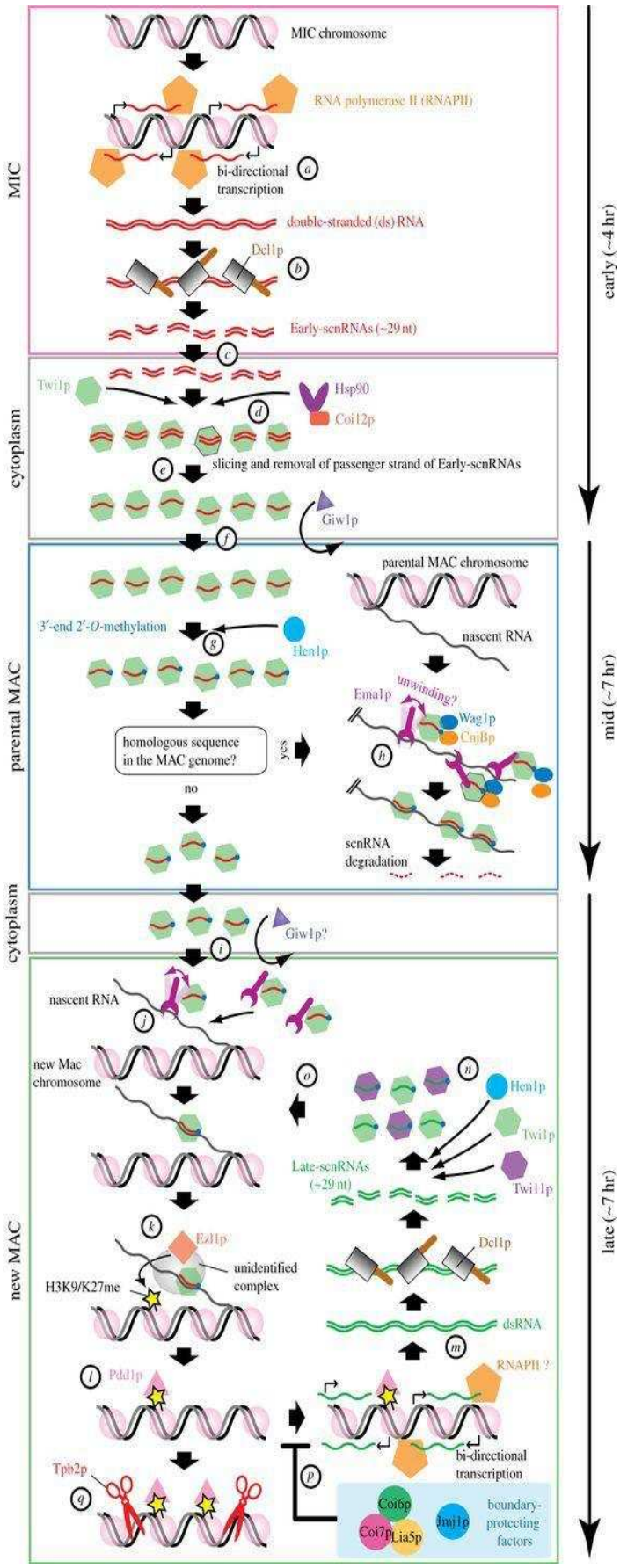


Figure XV. Molecular machinery of DNA elimination.

(a) MIC transcribe bi-directionally to dsRNA percussors of Early-scnRNAs. (b) Dcl1 then process dsRNA precursors to Early-scnRNAs, (c) and by yet unknown mechanism transported to the cytoplasm, (d) load into Twi1 with the help of Hsp90 and Coi12. (c) Twi1 cut and remove the passenger strand of scnRNA by their slicer activity, (f) and Giw1 transported the Twi1-complex to the parental MAC. (g) In the parental MAC, 3'-end of scnRNAs are 2'-*O*-methylateted by Hen1. (h) Twi1-loaded early-scan RNAs interact with the nascent transcript from the parental MAC with the help of Ema1, and TDSD of Early scnRNAs are achieved. Wag1 and CnjB are also believed to be required for the TDSD. (i) Giw1 (maybe) transported the IES-specific Early-scnRNA complex to the New MAC. (j) and Ema1 facilitate the interaction of Early-scnRNA with the nascent transcript from the IESs, (k) this interaction by unknow mechanism recruits Ezl1, which catalyzes H3K9/H3K27me3, (l) these marks are bounded by Pdd1. (m) Heterochromatin induces the production of Late-scnRNA RNAs in *cis*, Dcl1 prosses the Late-scnRNAs, (n) and loaded into Twi1 and Twi11 and are 2'-*O*-methylated by Hen1. (o) this Late-scnRNA-complex indicus further heterochromatinization by recruiting Ezl1 and form a positive feedback loop. (p) The HP1-like proteins, Coi6, Coi-7 and Lia5 downregulated the feedback. (q) Finally, the domesticated piggyBac transposase Tpb2 excised the heterochromatized IESs.

IV. SUMOylation in ciliates

Like other eukaryotes SUMO pathway in *Tetrahymena* and *Paramecium* is consisting of SUMO, a SUMO activating enzyme E1, a SUMO conjugating enzyme E2, and yet unidentified E3 ligases and SENPs. In *Paramecium*, it was shown that the SUMO E1 homolog Uba2 (Ubiquitin activating enzyme 2) and Smt3 transcript are upregulated in the sexual life cycle. RNAi knockdown of both Uba2 and Smt3 in vegetive growth showed very little effect on the cell survival or fission rate. On the other hand, the knockdown of Uba2 and Smt3 in the sexual life cycle shows severe effects, which include, failure of functional MAC formation and blockage of programmed DNA elimination (Matsuda and Forney 2006).

In *Tetrahymena*, the study of SUMOylation is not as straightforward as *Paramecium* since most of the components of the SUMO pathway are essential for vegetative growth and cell pairing during conjugation (Yang et al. 2015; Nasir et al. 2015). Therefore, the role of SUMOylation in TDSD of scnRNAs, programmed DNA elimination and heterochromatin formation, if any, has not been investigated. Nonetheless, with the currently available literature, it was shown that the SUMO E1 homolog (Uba2), Smt3, and Ubc9 (SUMO E2) are essential for *Tetrahymena*. In the vegetative growth of *Tetrahymena*, Uba2, Smt3 and Ubc9 were shown to be localized in the MAC. During conjugation, Uba2, Smt3, and Ubc9 localize to the parental MAC and the developing new MAC, consistent with the findings from flies (Nie et al. 2009; Donaghue et al. 2001), mammals (Evdokimov et al. 2008; La Salle et al. 2008), and yeasts (Dohmen et al. 1995), where components of the SUMO pathway are predominantly localizing to the nucleus since most of the SUMO target are nuclear (including chromatin-associated protein, DNA repair proteins, transcriptional factors, etc.). All the studied SUMOylation components in *Tetrahymena*, Uba2, Ubc9, and Smt3, are upregulated during conjugation and cells depleted for Uba2, Ubc9, and Smt3 cannot form mating pair. Depleted cells of Smt3, Uba, and Ubc9 were also shown to be sensitive to DNA-damaging agents (methyl methanesulfonate (MMS) and Cisplatin). Complete deletion of Ubc9 from the MIC and the MAC is lethal. In vegetative cells, Ubc9-depleted cells lose the MIC. On the other hand, the catalytically inactive mutation of Ubc9, in which the conserved Cys is mutated, resulted in multiple MICs in vegetative cells. While the same mutant arrests sexual conjugation during the developing new MAC formation (Yang et al. 2015; Nasir et al. 2015). All these findings suggest that SUMOylation plays an important role in the regulation of the sexual life cycle in *Tetrahymena*.

Results- Research article

Preface

In the following research article, I describe the main project I have been leading during my PhD under the supervision of Dr. Kazufumi Mochizuki. I investigated the Target-directed small RNA degradation (TDSD) process during the programmed DNA elimination in *Tetrahymena*. In this process, Piwi-loaded small RNAs that are complementary to the somatic genome are degraded and the remained small RNAs target transposon-related sequences in the germline genome for DNA elimination.

I found that a novel gene, *EMA2*, is required for this TDSD and for completing DNA elimination. I established *EMA2* KO strains, which lack the first half coding region of Ema2 and was described in a separate paper previously (Shehzada and Mochizuki 2022). Using the *EMA2* KO strains, I probed the genome of the DNA transposon Tlr1 by FISH and found that Ema2 is required for completing DNA elimination. By immunostaining, I further found that Ema2 is necessary for the accumulation of H3K9me3 and H3K27me3, which are required for the elimination of transposon-related sequences from the newly developed MAC.

Using the transgenic cell lines expressing HA-tagged Ema2, I found that during sexual reproduction Ema2-HA localizes to the parental MAC and the new MAC. This localization pattern was also reported in several proteins involved in DNA elimination. The localization pattern of Ema2 and its requirements in heterochromatin formation and DNA elimination are reminiscent of those of Ema1, which is also known to be involved in the TDSD of scnRNAs. Indeed, I demonstrated that Ema2 is also required for TDSD of scnRNAs. I achieved this finding by a northern hybridization for small RNAs and by deep sequencing of small RNAs.

To understand the molecular function of Ema2 in TDSD, I examined the protein sequence of Ema2 and found that Ema2 possesses an SP-RING domain which was also found in many SUMO E3 ligases. By GST-pulldown assay I demonstrated that Ema2 interacts with the SUMO E2 enzyme Ubc9. I also revealed that Ema2 enhances the accumulation of SUMOylated proteins in the SP-RING domain-dependent manner. Further, by comparing SUMOylated protein populations between wild-type and *EMA2* KO cells using immunoprecipitation of HA-tagged SUMO followed by mass spectrometry analyses, I found that Ema2 is targeting the transcriptional regulator Spt6 for SUMOylation.

The Ema2-dependent SUMOylation of Spt6 promoted me to investigate if the genome-wide occurrence of nascent long noncoding (lnc)RNAs in the parental MAC was affected in the absence of Ema2. By RT-PCR, I indeed found that Ema2 is required for the accumulation of lncRNAs at least at three loci I tested. Global disappearance of lncRNAs in the parental MAC of *EMA2* KO cells was also demonstrated by immunostaining using an anti-long double-stranded RNA antibody. Therefore, I conclude that Ema2 is required for the accumulation of lncRNA, which is a prerequisite for TDSD of scnRNAs.

Previously, the molecular mechanism of Target-directed microRNA degradation was reported. However, for the other small regulatory RNAs, the mechanism of degradation was not well studied. This work suggests that, in *Tetrahymena*, the genome-wide TDSD of Piwi-loaded small RNAs is regulated at the level of transcription of their complementary lncRNAs. While the importance of the regulation of nascent lncRNAs in small RNA-directed transposon silencing and heterochromatin formation has been emphasized in many previous publications, the regulation of lncRNAs for TDSD is conceptually new. Therefore, I believe this study is an important step toward understanding the TDSD process and the small RNA-directed transcriptional silencing mechanism in general.

A SUMO E3 ligase is involved in target-directed small RNA degradation during DNA elimination in *Tetrahymena*

Salman Shehzada[1], Tomoko Noto[1], Masatoshi Mutazono[1], Kazufumi Mochizuki[1]*

1) Institute of Human Genetics (IGH), CNRS, University of Montpellier, Montpellier, France

*Correspondence and Lead Contact: kazufumi.mochizuki@igh.cnrs.fr

Abstract

Small RNAs, about 20-30 nucleotides in length, recognize their genomic target by binding to nascent long non-coding RNA (lncRNA) and downregulate gene expression by small RNA-mediated heterochromatin formation. Piwi-associated small RNAs of the ciliated protozoan *Tetrahymena* uses a similar mechanism to induce heterochromatin formation followed by programmed DNA elimination. In this process, target-directed small RNA degradation (TDSD) of small RNAs complementary to the somatic genome enables small RNAs to specifically target the germline-limited sequences for DNA elimination. Although it has been suggested that this TDSD process is promoted by a small RNA-lncRNA interaction, the mechanisms by which lncRNA is transcribed and regulate TDSD are not well understood. We show that the SUMO E3 ligase Ema2 is required for the accumulation of lncRNAs from the somatic genome and for TDSD. Ema2 interacts with the SUMO E2 conjugating enzyme Ubc9 and enhances SUMOylation of the transcription regulator Spt6. We also show that Ema2 is required for heterochromatin formation and DNA elimination, indicating that Ema2-mediated SUMOylation also acts on the germline genome during DNA elimination. We propose that the SUMOylation of Spt6 allows genome-wide lncRNA transcription from both the somatic and the germline genome that is a prerequisite of TDSD and heterochromatin formation, respectively.

Introduction

Small RNAs, about 20-30 nucleotides in length and complexed with the Argonaute-family proteins, modulate a variety of biological processes in eukaryotes by regulating gene expressions at the posttranscriptional level through mRNA translation inhibition or degradation, or at the transcriptional level by induction of epigenetic modifications at specific genomic regions (Eulalio et al. 2008; Holoch and Moazed 2015). In plants, animals, metazoans, and yeasts, small RNA-guided deposition of the repressive chromatin mark histone 3 lysine 9 methylation (H3K9me) leads to transcriptional silencing (Volpe et al. 2002; Enke et al. 2011; Bernatavichute et al. 2008; Gu et al. 2012; Le Thomas et al. 2013; Pezic et al. 2014).

During the transcriptional repression, small RNA-Argonaute complexes recognize their genomic targets by binding to nascent long non-coding RNA (lncRNA) transcripts in many eukaryotes (Holoch and Moazed 2015; Grewal and Jia 2007). In *S. pombe*, the RNA-induced transcriptional silencing (RITS) complex establishes a physical connection between chromatin and small interfering RNAs (siRNA) by lncRNA transcript for initiation and spreading of heterochromatin (Bühler et al. 2006; Kato et al. 2005; Djupedal et al. 2005; Motamedi et al. 2004). Also, in metazoans, small RNA-mediated transcriptional repression is achieved by the binding of PIWI-interacting RNA (piRNA) to the nascent lncRNA transcript (Le Thomas et al. 2014; Pezic et al. 2014; Batki et al. 2019; Onishi et al. 2021). Similarly, in the ciliated protozoan *Tetrahymena*, the small RNA-directed mechanism recognizes lncRNA for inducing heterochromatin formation followed by programmed DNA elimination (Xu et al. 2021; Aronica et al. 2008; Mochizuki et al. 2002). In this study, we use this small RNA-directed DNA elimination process of *Tetrahymena* to understand the molecular mechanism for the lncRNA transcription.

Like other ciliated protozoans, *Tetrahymena* possesses two nuclei: the diploid germline micronucleus (MIC) and the polyploid somatic macronucleus (MAC) (Goldfarb and Gorovsky 2009). In vegetative growth, the MAC and MIC independently divide and are segregated into daughter cells. In sexual reproduction, where two cells make pair for conjugation, the diploid MIC undergoes meiosis in each cell and produces four haploid pronuclei, one of which exchange between two conjugating cells, and create a zygotic nucleus by the fusion of stationary and migratory pronuclei, which mitotically dived twice to produce the two new MICs and the two new MACs, while the parental MAC is degraded (Chalker et al. 2013; Davis et al. 1992; Martindale et al. 1982; Noto and Mochizuki 2017).

During the formation of the new MAC, ~12,000 internal elimination sequences (IESs), many of which are related to transposons, are deleted and the genome is reorganized from the remaining MAC-destined sequences (MDSs) (Noto and Mochizuki 2017). Our analysis of the latest publicly available MIC genome sequence revealed that IESs consist of around half of the 200 Mb MIC genome and the size of IESs ranges from 0.2-40 kb.

DNA elimination is epigenetically regulated by ~29-nt small RNAs, called scan RNAs (scnRNA) (Mochizuki et al. 2002; Coyne et al. 1996; Lee and Collins 2006). During early conjugation, IESs and their surrounding MDSs are bi-directionally transcribed to produce the double-stranded RNA (dsRNA) precursors of scnRNA in the MIC (Mochizuki and Gorovsky 2004b). The dsRNA precursors are processed by the Dicer-like protein Dcl1 to scnRNAs, exported to the cytoplasm (Mochizuki and Gorovsky 2005), and loaded into the Piwi-clade Aragnute protein Twi1 (Mochizuki et al. 2002), where Twi1 cut and remove the passenger strand of scnRNA by their slicer activity (Noto et al. 2010). Then Twi1-scnRNA complex is transferred to the parental MAC (Noto et al. 2010), where the scnRNAs complementary to the somatic MAC genome (= MDSs) are degraded by target-directed small RNA degradation (TDSD) (Aronica et al. 2008; Schoeberl et al. 2012; Mochizuki and Gorovsky 2004a). TDSD requires the RNA helicase Ema1, which promotes the interaction between nascent lncRNA and Twi1-scnRNA complex (Aronica et al. 2008). The remained IES-specific scnRNAs, which are not subjected to TDSD, are then transported to the new MAC, where Ema1 is also believed to facilitate the interaction of Twi1-scnRNA complex to lncRNA transcript (Aronica et al. 2008; Noto and Mochizuki 2017) to recruit the Polycomb repressive complex 2 (PRC2)-like histone methyltransferase specifically to IESs. This complex, containing Ezl1, catalyzes H3K9me3 as well as H3K27me3, resulting in IES-specific accumulation of an HP1-like protein, Pdd1, (Madireddi et al. 1996; Taverna et al. 2002), and of a piggyBac transposase, Tpb2, which excises the heterochromatinizated IESs (Cheng et al. 2010; Vogt and Mochizuki 2013).

We previously predicted that TDSD is achieved by the interaction between Twi1-loaded scnRNA and lncRNA transcript from the MAC, which is facilitated by Ema1(Aronica et al. 2008). A similar TDSD mechanism was proposed in the closely related ciliate *Paramecium*, from the fact that RNAi knockdown of lncRNA transcripts from the parental MAC blocked DNA elimination (Lepère et al. 2008). However, the molecular mechanism for regulating the transcription of these lncRNAs is not well understood.

In this study, we show that the SUMO E3 ligase Ema2 is required for genome-wide transcription of lncRNA from the parental MAC. SUMOylation is a conserved post-translational

modification by a Small-Ubiquitin-like modifier (SUMO) and is involved in numerous biological processes, including transcription (Gareau and Lima 2010; Müller et al. 2001; Vertegaal 2022; Hendriks et al. 2017). SUMO E3 ligase acts downstream of SUMO activating enzyme E1 and conjugating enzyme E2 to conjugate SUMO to target lysine residues (Gareau and Lima 2010; Geiss-Friedlander and Melchior 2007; Kerscher et al. 2006; Capili and Lima 2007b). We show that Ema2 interacts with the *Tetrahymena* SUMO E2 enzyme Ubc9 and enhances the SUMOylation of the conserved transcriptional regulator Spt6. We further show that Ema2 is required for TDSD in the parental MAC, heterochromatin formation in both the parental and the new MAC, and required for DNA elimination. We propose that Ema2-dependent SUMOylation of Spt6 is required for the genome-wide lncRNA transcription and thus for TDSD, heterochromatin formation, and DNA elimination.

Materials and Methods

Strains and cell culture condition

The *Tetrahymena thermophila* wild-type strains, B2086, CU427 and CU428 were obtained from the *Tetrahymena* stock center. The production of the *EMA2* KO strains were described previously (Shehzada and Mochizuki 2022). The other transgenic *Tetrahymena* strains are described below. SPP medium with 1% peptone were used to grow the cells at 30 °C (Gorovsky et al. 1975). For conjugation, cells with two different mating types (~5 x10^5 cells per milliliter) were washed and resuspended in 10 mM Tris (pH 7.5), and pre-starved for overnight at 30°C.

Construction of *EMA2*-HA strains

To obtain the *EMA2*-HA construct, a part of the coding sequence of *EMA2* was amplified from the genomic DNA of B2086 using *EMA2*-C-FW and *EMA2*-HA-C-RV. Similarly, the 3' flanking sequence of *EMA2* was amplified using *EMA2*-F-FW and *EMA2*-F-RV. The two PCR products were combined by overlapping PCR using *EMA2*-C-FW and *EMA2*-F-RV and cloned into XbaI and XhoI sites of pBlueScriptSK (+) to obtain p*EMA2*. Next, the *HA-neo3* construct was excised out from pHA-neo3 (Kataoka et al. 2010), using SalI and BamHI and cloned into the same restriction sites of p*EMA2* to obtain p*EMA2-HA-neo3*. Finally, p*EMA2-HA-neo3* were digested with XbaI and XhoI and introduced into the *EMA2* MAC locus of B2086 and CU428 by homologous recombination by biolistic transformation (Bruns and Cassidy-Hanley 2000; Shehzada and Mochizuki 2022). Cells having the *EMA2-HA-neo3* construct were selected in the 1xSPP containing 100 μg/mL paromomycin (Sigma) and 1 μg/mL $CdCl_2$. Then, copies of the endogenous *EMA2* in the MAC were replaced by the construct through phenotypic assortment (Hamilton and Orias 2000) and partial replacements were confirmed by genomic PCR. The primers used are listed in Supplementary Table 1.

Construction of *SPT6*-HA strains

To obtain the *SPT6*-HA construct, a part of the coding sequence of *SPT6* was amplified from the genomic DNA of B2086 using *SPT6*-FW1-SacII and *SPT6*-RV2-overlap. Similarly, the 3' flanking sequence was amplified using *SPT6*-3F-FW2-overlap and *SPT6*-3F-RV1-KpnI. The two PCR products were combined by overlapping PCR using *SPT6*-FW1-SacII and *SPT6*-3F-RV1-KpnI

and cloned into the SacII and KpnI sites of pBlueScriptSK (+) to obtain p*SPT6*. Next, the NheI and XhoI fragment of pHA-neo3 was cloned into the same restriction sites of p*SPT6* to obtain p*SPT6-HA-neo3*. Then, *neo3* was excised from p*SPT6*-HA-neo3 with XmaI and SalI and replaced by *pur6* from pBP6MB1-TR-TUBE to obtain p*SPT6-HA-pur6*. Finally, the p*SPT6-HA-pur6* were digested by SacII and KpnI and introduced into the MAC *SPT6* locus of one of the *EMA2* KO strains by biolistic transformation. The transformed cells were selected in 300 μg/mL puromycin (Invitrogen) and 1 μg/mL $CdCl_2$ and cells were phenotypically assorted as previously described (Hamilton and Orias 2000), until cells grew in 1.2 mg/mL of puromycin. The primers used are listed in Supplementary Table 1.

Construction of HA-*SMT3* and His-*SMT3* strains

STM3 cDNA was amplified by RT-PCR from the total RNA of vegetatively growing CU428 using *SMT3*_HA_FW and *SMT3*_HisExt_RV and cloned into the BamHI and SpeI sites of pBNMB1-HA using NEBuilder HiFi DNA assembly kit to obtain pBNMB1-HA-*SMT3*. Next, *neo5* was excised with SalI and XmaI and replaced with *pur6,* which was amplified from pBP6MB1 with BTU1-LCFW and MTT1_5UTR-SeqRV, using NEBuilder HiFi DNA assembly kit to obtain pBP6MB1-HA-*SMT3*. For His-*SMT3* expression, the HA-tag encoding sequence of pBP6MB1-HA-*SMT3* was replaced by a sequence encoding 6xHis tag. Then, pBP6MB1-HA-SMT3 and pBP6MB1-His-SMT3 were digested by XhoI and introduced into the MAC *BTU1* locus of the *EMA2* KO cells. Transformation, selection, and phenotypic assortment were performed as described for the *SPT6*-HA strains. The primers used are listed in Supplementary Table 1.

Construction of *EMA2* Rescue strains

The whole coding sequence of *EMA2* was amplified with *EMA2*-PBNMB1-FW and *EMA2*-PBNMB1-RV from B2086 genomic DNA and cloned into the SpeI and BamHI sites of pBNMB1-EGFP-HA with NEBuilder HiFi DNA assembly kit to obtain pBNMB1-EGFP-HA-*EMA2*. Then, the *neo5* cassette was removed using SalI and XmaI and replaced by *pur6* to obtain pBP6MB1-EGFP-HA-*EMA2*. Next, the C72A and C75A mutations were introduced by site-directed mutagenesis using *EMA2*_mut1_S, *EMA2*_mut1_AS, and Prime STAR HS DNA polymerase (Takara) to produce pBP6MB1-EGFP-HA-*EMA2*-Mut1. Similarly, C377A and C380A mutations were introduced to pBP6MB1-EGFP-HA-*EMA2* using *EMA2*_mut2_S, *EMA2*_mut2_AS to produce pBP6MB1-EGFP-HA-*EMA2*-Mut2. All these constructs were digested with XhoI and

introduced into the *BTU1* locus of one of the *EMA2* KO strains. Transformation, selection, and phenotypic assortment were performed as described for the *SPT6*-HA strains. The primers used are listed in Supplementary Table 1.

Construction of wild-type and mutant HA-*SPT6* strains

SPT6 cDNA was amplified from total RNA of mating B2086 and CU428 at X hpm with *SPT6*-FW-Turbo and *SPT6*-RV-Turbo, cloned into SpeI and BamHI sites of pBCMB1-HA-Turbo using NEBuilder HiFi DNA assembly mix to obtain pBCMB1-HA-Turbo-*SPT6*. Then, the *MTT1* promoter and the HA-Turbo-tag were removed using AgeI and XmaI and replaced by the *SPT6* promoter and the N-terminal HA tag, using the overlapping product of two genomic PCR fragments amplified with *SPT6*-5F-FW, *SPT6*-5F-RV and with *SPT6*-N-FW-HA and *SPT6*-N-RV by NEBuilder HiFi DNA assembly mix to obtain pBCSB-HA-*SPT6*. Then, the K1401R and K1407R mutations were introduced to pBCSB-HA-*SPT6* by site-directed mutagenesis with *SPT6*_K1401R_K1407R_FW and *SPT6*_K1401R_K1407R_RV to obtained pBCSB-HA-*SPT6*-2KR. Then, K235R and K237R mutations were introduced with *SPT6*_K235R_K237R_FW and *SPT6*_K235R_K237R_RV, K1014R mutation by *SPT6*_K1014R_FW and *SPT6*_K1014R_RV, and K1166R mutation by *SPT6*_K1166R_FW and *SPT6*_K1166R_RV to finally obtain pBCSB-HA-*SPT6*-6KR. The constructs were digested with XhoI and introduced into the MAC *BTU1* locus of wild-type cells. The transformed cells were selected using 12 µg/mL cycloheximide (Sigma), and cells were phenotypically assorted until cells grew in 16 µg/m cycloheximide and replacement levels of the *BTU1* locus were examined by genomic PCR. The primers used are listed in Supplementary Table 1.

DNA elimination assay by FISH

DNA elimination assay by FISH was performed as described (Loidl and Scherthan 2004). pMBR 4C1, pMBR 2 and Tlr IntB, which contain different parts of the Tlr1-element were used to produce Cy5-labeled FISH probes by nick translation (Wuitschick et al. 2002). Images of stained cells were taken with Cy5 and DAPI filters and quantified using Fiji software. Whole areas of the new MAC (A) and MIC (I) and a part of the cytoplasm (C) were manually encircled in the images; and the intensities of DAPI and Cy5 (FISH) signals in each area were measured using Fiji software. Then, the IES retention index was calculated as follows: 0.35 × {[(mean Cy3 intensity A - mean Cy3 intensity C)]/[(mean Cy3 intensity I - mean Cy3 intensity C)]}/{[(mean DAPI intensity A - mean

DAPI intensity C)]/[(mean DAPI intensity I - mean DAPI intensity C)]}. Based on our previous observation, DNA elimination is completely blocked in the absence of *TWI1* KO therefore, the correction factor of 0.35 was obtained experimentally to make the average IES retention index of *TWI1* KO cells to 1.0.

RT-PCR analyses of lncRNA transcripts

Total RNA from mating wild-type and *EMA2* KO cells were isolated at 6 hpm using TRIzol reagent (Invitrogen). Then RNA was treated with Turbo DNase (Ambion) and cDNA was synthesized from 1 μg of RNA using SuperScript II (Invitrogen) using random 6-mer as primers. The lncRNA transcripts from the MAC loci containing the M-, L8-, and R2 IES boundaries (see also Fig. 7 A) were amplified by nested PCR (95°C for 20sec/50°C for 30sec/68°C for 1min; 30 cycles each) using the primers M5'-3 + M3'-3 (M-1st), M5'-4 + M5'-4 (M-2nd), L8_5'-1 + L8_3'-1 (L8-1st), L8_5'-2 + L8_3'-2 (L8-2nd), R2_5'-1 + R2_3'-1 (R2-1st), and R2_5'-2 + R2_3'-2 (R2-2nd) and analyzed in 1% agarose gel. As a positive control, the constitutively expressed *RPL21* mRNA was amplified using primers RPL21-FW and RPL21-RV. The primers used are listed in Supplementary Table 1.

Viability test for progeny

To test the viable progeny of *Tetrahymena*, as described by (Aronica et al. 2008; Mochizuki et al. 2002), mating pairs were isolated at 6 hpm and the growth of cells were examined at 66 h hpm using paromomycin sensitivity for *EMA2* KO cells and 6-mp resistant for wild-type cells as described previously (Mochizuki et al. 2002).

Small RNA analysis by northern hybridization

Total RNAs were isolated from conjugating wild-type and *EMA2* KO cells using TRIzol reagent, 10 μg of RNAs were separated by a 15% acrylamide-urea gel, counterstained with GelRed (GENTAUR) and analyzed by northern blot using the radio-labeled Mi-9 probe as previously described (Aronica et al. 2008). The hybridized probe was detected by Typhoon IP Phosphorimager (GE healthcare).

Small RNA analysis by high-throughput sequencing

High-throughput sequencing and analyses of small RNAs were performed as described (Mutazono et al 2019). The 2020 version of the MAC genome assembly (Sheng et al. 2020) was obtained from *Tetrahymena* Genome Database and fragmented into 10 kb pieces. Each 10 kb fragment that contains longer than 3 kb unique mappable sequence was used as a MDS tile (total 10,235 tiles). The sequences of Type-A IESs were described previously (Noto et al. 2015) and each Type-A IES was used as an IES tile (total of 4,691 tiles). Normalized numbers (RPKM [reads per kilobase of unique sequences per million]) of 26- to 32-nt small RNAs that uniquely matched the MDS and IES tiles were counted.

GST pull down assay

Synthetic *EMA2* and *UBC9* genes, both codon optimized for *E. coli*, were obtained from Integrated DNA Technologies (listed in Supplementary Table 2). The *EMA2* gene was cloned into the BamHI and XhoI sites of pGEX-4T-1 to obtain pGEX-4T-1-*EMA2*. The *UBC9* gene was cloned into NdeI and BamHI sites of pET28a (+) to obtain pET28-*UBC9*.

pET28-*UBC9* was introduced into BL21 *E. coli* strain, the expression of His-tagged Ubc9 was induced overnight (~16 hours) at 25 °C with 0.05 mM isopropylthio-β-galactoside (IPTG), the cells were lysed in 50 mM Hepes (pH 7.5), 800 mM KCL, 10% glycerol, 10 mM Imidazole, 0.2 mM PMSF, 1x cOmplete ETDA-free (Roche), incubated with Ni-NTA beads for 2 hours at 4 °C , washed the beads four times with 50 mM Hepes (pH 7.5), 800 mM KCl, 10% glycerol, 20 mM Imidazole, 0.2mM PMSF, 1x cOmplete ETDA-free, eluted with 50 mM Hepes (pH 7.5), 800 mM KCl, 10% glycerol, 250 mM Imidazole, 0.2 mM PMSF, 1x complete ETDA-free, dialyzed with 50 mM Tris pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT, 50% glycerol, 1x cOmplete ETDA-free, and finally passed through Microcon membrane (10 kDa MW cutoff, Amicon). For GST pull-down assay, 10 µg of His-Ubc9 was diluted in 20 mM Tris–HCl, pH 7.5, 100 mM NaCl, 1 %vNP-40, 5 mM $MgCl_2$, 1 mM DTT, 0.2 mM PMSF, 1x cOmplete EDTA-free.

BL21 *E. coli* cells were transformed with pGEX-4T-1 or pGEX-4T-1-*EMA2,* and expression of GST-alone or GST-Ema2 was induced overnight at 25 °C, lysed with 50 mM Tris–HCl, pH 7.5, 100 mM NaCl, 1 mM DTT, 0.2 mM PMSF,1x cOmplete EDTA-free and incubated with Glutathione Sepharose 4B for 90 min at 4°C. Then, the supernatant was discarded and incubated with 10 µg of His-Ubc9 in 20 mM Tris–HCl, pH 7.5, 100 mM NaCl, 1 %vNP-40, 5 mM $MgCl_2$, 1 mM DTT, 0.2 mM PMSF, 1x cOmplete EDTA-free for 90 min at 4°C, washed 4 times with 20 mM Tris–HCl, pH 7.5, 100 mM NaCl, 1 % NP-40, 5 mM $MgCl_2$, 1 mM DTT, 0.2mM PMSF, 1x

cOmplete EDTA-free and eluted with 2x SDS sample buffer (100mM Tris-Cl pH 6.8, 4% sodium dodecyl sulfate, 0.2 % bromophenol blue, 20% glycerol and 200 mM β-mercaptoethanol). The eluted proteins were analyzed on a western blot against an anti-GST and an anti-His antibody.

Immunoprecipitation for detecting SUMOylated proteins

Total proteins of *Tetrahymena* cells were precipitated by incubating cells in 6 % trichloroacetic (TCA) on ice for 10 min and then spined down at 13,000 rpm for 5 min at 4 °C. The precipitate was dissolved in 1x SDS sample buffer (62.5 mM Tris-HCl pH 6.8, 1% SDS, 0.01% BPB, 1% β-mercaptoethanol), neutralized with 2 M Tris, and then incubated for 5 min at 95°C. The dissolved proteins (600 µl) were then suspended in 7.4 mL with 20 mM Tris pH 7.5, 100 mM NaCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$ and incubated with Ezview Red anti-HA agarose beads (Sigma) for overnight at 4 °C. Then the beads were washed 5 times with 20 mM Tris pH 7.5, 100 mM NaCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 0.1% tween 20 and the bound proteins were eluted with by incubating the beads with 1x SDS buffer for 5 min at 95°C. The eluted proteins were analyzed by western blot against an anti-HA and an anti-Smt3 antibody.

Immunofluorescent staining and western blotting

The following antibodies were obtained from manufactuers: anti-H3K27me3 (Millipore 07-449); anti-HA (BioLegend 901513, clone 6B12); anti-H3K9me3 (Active Motif 39162,); anti-GST (BD Biosciences, 554805); anti-His (Proteintech, 66005-1-Ig,); anti-long dsRNA (SCICONS 10010200); 12G10 anti-alpha-tubulin (Developmental Studies Hybridoma Bank of the University of Iowa, AB 1157911). The anti-Smt3 antibody was a gift from Dr. James Forney and described previously (Nasir et al. 2015)). Immunofluorescent staining of long dsRNA was performed as described by Woo et al. (Woo et al. 2016). The other immunofluorescent staining experiments were performed as described (Loidl and Scherthan 2004). For immunofluorescent staining, all the primary antibodies and the respective secondary antibodies (Alexa-488-coupled anti-rabbit or mouse IgG, Invitrogen) were diluted 1:1000. For western blot, all the primary antibodies and the respective secondary antibodies (HRP-coupled anti-rabbit or mouse IgG, Jackson ImmunoResearch Lab) were diluted 1:10,000.

Supplementary Table 1

EMA2-C-FW: CCGCTCTAGATTTAATCTTTCTTATGTTAATAATCGTC

EMA2-HA-C-RV: GTCGACACTAGTGGATCCAATGCAAATATTCAAAGAGTT

EMA2-F-FW: GGATCCACTAGTGTCGACTGTTTGTCTTTATCTCGTAGTC

EMA2-F-RW: GGCGCTCGAGTTCTAGGATAACTAATAGAGTGAAC

SPT6-3F-FW2-overlap :TGCTAGCGGATCCGTCGACCTCGAGTCTCTTAGTGTTTTATATAAAACAGC

SPT6-3F-RV1-KpnI: ATTGGGTACCGGGCCCGTATCTTTAACTATCGACAAAACC

SPT6-FW1-SacII: TCCACCGCGGACTTATCACTTAAGGAAGAATATGGC

SPT6-RV2-overlap: ACTCGAGGTCGACGGATCCGCTAGCATGCTCGTTGTCATATGAACGAG

*SMT3*_HA_FW:CCTTATGATGTTCCTGATTATGCTGGTGGTTCTGGTACTGATTAAAACGC

*SMT3*_HISEXT_RV: GCTGACCGATTCAGTTCGCTCAATCAGAAAGAGCCACCAACTTGTTC

BTU1_LCFW: AATTTTAATGCGTGTATTTATTTGGGTG

MTT1_5UTR_SeqRV: AGTACTCTATTATGTTTTTTCTTATACAG

EMA2-pBNMB1-FW: ATGATGTTCCTGATTATGCTGGATCCATGAGTTCACTATTTGGAAATAAGAATAA

EMA2-pBNMB1-RV:TTAGCTGACCGATTCAGTTCGCTCAACTAGTTCAAATGCAAATATTCAAAGAGTTAAGAG

*EMA2*_mut1_S: GAATTTAGATACTTTTTGTGCTGAAAACGCCTATCTCAGTCTGCATTTTCC

*EMA2*_mut1_AS: GGAAAATGCAGACTGAGATAGGCGTTTTCAGCACAAAAAGTATCTAAATTC

*EMA2*_mut2_S: GTTAACCCTACTTGGCTGGCTCCTATTGCTAATAAGCACAAAATTTTTTTAAAAG

*EMA2*_mut2_AS: CTTTTAAAAAAATTTTGTGCTTATTAGCAATAGGAGCCAGCCAAGTAGGGTTAAC

*SPT6*_FW_Turbo: GATCTGGTGGTTCAGGAAGCGGATCCATGAGTAAACCAAGAAAAAACGAAG

*SPT6*_RV_Turbo: GCTGACCGATTCAGTTCGCTCAATCAATGCTCGTTGTCATATGAACGAG

*SPT6*_N_FW_HA: GTATCCTTATGATGTTCCTGATTATGCTGGATCCATGAGTAAACCAAGAAAAAACGAAG

*SPT6*_N_RV: ATACTTTGCGCCATAATCACCTAC

*SPT6*_5F_FW: CAAATTTTTACTGGAAAAATGCAGCTAAATTATGAGACCAAACTAAG

*SPT6*_5F_RV_HA: GCATAATCAGGAACATCATAAGGATACATTTTTCAAATATTTGCTGTGAAGACT

*SPT6*_K1401R_K1407R_FW: GGCATAAGATAAGAAGGTGGTGTCAGATATGAAAAATCAGAGAGAGGC

*SPT6*_K1401R_K1407R_RV: CATATCTGACACCACCTTCTTATCTTATGCCAATAGTTCCTTAATTAC

SPT6_K235R_K237R_FW: CGAGCGTGATGATGATAAAAGAATTAGAGAAGAGGATATACCTGAGCG

*SPT6*_K235R_K237R_RV: CGCTCAGGTATATCCTCTTCTCTAATTCTTTTATCATCATCACGCTCG

*SPT6*_K1014R_FW: GTATTAGATTTTGTCAGAGATGAGCTTATCAACCCATTTAGC

*SPT6*_K1014R_RV: GATAAGCTCATCTCTGACAAAATCTAATACAGGAGCC

*SPT6*_K1166R_FW: GAAGCTGAAGACAGGCCATTGGGTATTGATGTCACTG

*SPT6*_K1166R_RV: CAATACCCAATGGCCTGTCTTCAGCTTCAATGAACTT

M5'-3: AGGTACGATAGATCGACTGACGG

M5'-4: AATAATAAGGAACCTCTTACTGTG

M3'-3: ACTAAAATATTTATCTTCTTTTCTGC

M3'-4: ACTTTCAAAAAATTAATTTTGAGTAAAG

L8_5'-1: ATTTGCTTTTAGGAAGAGATATCAC

L8_5'-2: ACATACATTTGAAATATCACCAAGC

L8_3′-1: ACAAGAAAGGTATTCATTCATTCCTC

L8_3′-2: TGATAAAAAAGTGGTGAAAAAAGGAC

R2_3'-1: TAATTTTAGGGCGAATCACC

R2_3'-2: ACTCATAAATAAAATCATATCATAGTCTAG

R2_5'-1: TTAGCAAAGTGCATTAACCTC

R2_5'-2: TCTCTTGAAATTGGGCAAAAACTG

RPL21-FW: AAGTTGGTTATCAACTGTTGCGTT

RPL21-RV: GGGTCTTTCAAGGACGACGTA

NMC1_f53600: GTGCAATTTTACACCGTCAGG

NMC1_r55732: AGAATCTAACTCATAGCACTGG

Supplementary table 2

EMA2:

CGATCGGGATCCATGTCTAGTTTGTTCGGTAACAAGAACAAGGAAGTTATCTTCAATTTTTGCCAAAAGTCTCAGGA
AGAAGATCAAGACATGTTTGCGGACCATGAAATTCTGATTGATGAACAATTACAGTGTTTTAACTGTAACAAGTACT
TTGATCCCACGAAGTACGGGATCGATTTGTTAGACGAGATGGACTTCAACAAGAATCTGGACACTTTTTGTTGTGAG
AATTGTTATCTTTCGTTGCATTTTCCTTTTTATTATCAATTGGATTTATTGTTGAAAATGTCAAACCTGGAAGTGAA
CACGAAGTACGAATTCAACCTTGTACTGCAGGAGGAGTGCTACCAAGACAGTTGCATCCTGGGCGTGTTCTTAAGCA
AGCAAAATCTGATTAATGTCTTCAAGAAACAGACGGGGAGTCATAACTTCATCTTAAACGTGAAAATGAATGGGCAA
GACCCCTCCCAGGAAGGCAGTAACTTTTTTGTGATCGACTCGGATGACTTAACGCAGATGCAAAATACTTTGAACTT
CGAATGTCTTAACTTTTCCTTTCAGAAATCTTTCATCTCGAACATCATCTTAATGAACGACCAACAGAAAAGTCAGC
TTAAAACGGCCTCTCCGAACATCATCAGCGTAATCCTGCACTGCCGCAAAATTCGTGTCTCTGCGTATGTCGCGAAA
CTGTTTAAGAAGAAGTTCTTTAACTGTTTAAATCATGAAAGTTTATTCAAGGTTTTCGAGGAACAAGAGCAAAAATA
TAAGCAAAAAGACCAGATCCAGAAAATCAAATTACAAACCAAACAGAACTCGCTTTCAACGTCTCAAATTCATCAGT
CTAGTACCCTGAACGAAATTCAGCCCAGCCAGCTTCAAACCCATAAGGAAACTTCGAACATCTCGCAAAGTAGCCAG
CTTAACAAAGATTTCTGTGAAAAAAAAGAACTGACCGAGATCTATCTTCGCATTAAAAACTTTCACCAGATGTTCCC
GTTCATGTACAAGAATGAGATGGTGGTTAAGATGGACGATAACGCAGTTATCCAATACCCCGCGTTTACCTACCATC
ACTCGCACCCCTTTGACTTACGTGATTACTGCTACCTTAACCAGGTGAATCCTACCTGGTTGTGTCCCATCTGCAAC
AAGCACAAGATTTTCTTGAAGGAGATCCAGCTGGATTTTTATTTGTTTGCGTTGATTCAAACACAGTATCTTAAGGA
CTCCTATATTTCGGACAAAATCCAGCAGATTCTTCAGTCGAAGCCACAGTTTGATTCTGAATTTTTTAGCAATCAAC
TGAAACAGCTGAACTTGCACTCAATGAAGGTACTGGAGCAAACACAGCAACGCACCCAGTCTAACGAGATTAAGATC
CAGCAGAACTTATCCAATAAGTTGAACCAGACTAACAAGAATTTTATCCAGATCTACAAAAAGTCGAATCAGTTGGG
GCAAAAAGAGAATATTCTGAAGAACAGTACTAAAATTAAGCAATTTAAGTGTAAGTCATCTATTGATCAGAATACAA
ATTATGAACAGCGTAACCACAAGAATGGGAATCAAAATCAAAAGTCAAAAACCTCAGAGTTTTTCAATGAAAATTTG
AACTGTACTGACTTAAGCCAAGACAATTACCTGGATTTCATTCAGGATTTTGACGAATTTCAATCTTTCTTAAATCC
TTACCAGAAGCAACATTCGATGCGCAATAGCAAAGATATTTCTGTACGTCATAATATGGATCAAGCTAACCTTATCG
ACGAGATGTCTATCCTTAACATTCAGTGCAATGAAATCGCCTCCCAAAAATCTTCAACCCGTGTTTACAGCAAATTA
GTAAACAAATCTAAATCGAAGACCTCATCTTCTAACAAAGAAGAGAAGCAGAAGCAGAACGACGCATTTAAAGAAAA
CCAAATCGACATTGAGCAATCAGGCCAGATTATCGAAGAAAAAGTCCAAGAGAAACCGAGCTTGAACGATCAAATTG
ATGCTAATAAGAACATCCAGTTCTCCCAACAAATTGAGTCCTTTTCCGACCTGAGTCAGAAGTTCATCTCCCAGCAA
AAAGAAGAGCGTACTCTTGAGGTTGCAAATCAATTCTACTTGGACACTTCGTTCCTTCGCTCCCCTCTTGAAATCAA
CGTCTTGTTTGACGTATCCTATTTGCAATTCAATAAAGAAGAGCAGGAGTATCAGTCAGTTCAGCAGATCATTCAGC

TGATTGACTTCCAAGAAGACAAGCAGATGTTAGAGAAGCTTTTCAACATTGACATCCTTCAAAATCACTATTTTATT
TTCAATATTTTCAATCTGAGCTATGTAAATAATCGTCAGTATCAAGACTTGCACACAATCTTAAACCAAAACCCTTC
GACTAAAACTTTTTTTCAAGTCTTGATCAACGTTCAAGAGCAGTTCATCAACCTTGGTATGAGTTCAGAGTTTTATA
CGTTAATTTCTCAGTTCTATTATAAAGTTATGTCTAATAAGCAGCGCAATAATAAGATTTCTTACGACCTTTTGGAT
TCCGAAAGCCAAATTACGTCGCTTAATTATTCGCAAAACAAAAACTTTTTAGAGAAACTTATTGAATGCTTCCTGGG
AATGTCGATTTTTAAGCAACTGTCCAACGAAGATGTCATTGAAAATTTCCAACAAATGGTCTTTGAATACTTATTGT
TCAAGCAACAAAATATCCCGACTTTGATTAATATTCACAATATCATCATCAACTATTTTCCCAATCAATGCCTTAAC
TTAATCGAATCGATCAAGAAATCTTTGAATAGCCTGAACATTTGCATCTGACTCGAGCCCGGG

***UBC9*:**

GGCCTGGTGCCGCGCGGCAGCCATATGATGCAACAGCAGAATAAAGAAGTTAACGAACTTGTAATTACGCGCCTGAA
ACAGGAGCGCAAACAGTGGCGTCAGGATCATCCTCATCGTTTCGTCGCCAAACCCATGACTAAAGAAGACGGTACCA
TCAATATGCTTAAATGGTACTGTGAGATTCCGGGACCGGAAGGGTCACCCTGGGAGGGGGGCGTTTACATTTTATAC
ATGGACTTTTCACAAAGCTACCCTATTAAACCCCCTAAGTGCCAATTCAAACCGGTCCTTCCTCACCCCAACGTGTA
TCCAAGTGGCACCGTCTGCTTATCAATCCTGAACGAAGAAGAGGACTGGAAATCTGCAATTACGATCAAACAGATTT
TGATGGGAATCCAAAAGTTACTGAAGGATGAACCGAACATTGACAGCCCTGCCCAGTACGAGCCCTGTGCTTTATAT
CGTAGTGACAAGGAAAAATACTACAAAAAGGTACGTGAGTTTGCTGAAACTATGAAGAAGAAGGATtgagGATCCGA
ATTCGAGCTCCGTCGACA

Results

Ema2 localizes to the parental MAC and the new MAC during conjugation

Most of the known factors involved in DNA elimination are expressed exclusively during conjugation (Noto and Mochizuki 2017; Aronica et al. 2008; Noto et al. 2010; Mochizuki et al. 2002; Malone et al. 2005; Mochizuki and Gorovsky 2005; Liu et al. 2007). Therefore, to identify a novel protein involved in DNA elimination, we investigated the function of the conjugation-specific gene *EMA2 (TTHERM_00113330)* (Figure 1A).

First, to localize the Ema2 protein, Ema2 fused with HA-tag at the C-terminus (Ema2-HA) was expressed from the endogenous *EMA2* locus in the MAC and detected by immunofluorescence staining using an anti-HA antibody. Consistent with the conjugation-specific expression of *EMA2* mRNA, Ema2-HA was not detectable in the vegetative cells (Figure 1B, NM). During conjugation, Ema2-HA was first localized to the parental MAC when the MIC underwent meiosis (Figure 1B, 3 hours post-induction of mating [hpm]) and remained there until the new MAC was formed (Figure 1B, 6 hpm). Then, Ema2-HA disappeared from the parental MAC and appeared in the new MAC at the onset of its development (Figure 1B, 8 hpm) and was gradually disappeared at the later stages (Figure 1B, 12 and 14 hpm). This localization switch from the parental and to the new MAC was also reported for several other proteins involved in DNA elimination (Taverna et al. 2002; Mochizuki et al. 2002; Liu et al. 2007; Madireddi et al. 1996; Aronica et al. 2008; Noto and Mochizuki 2017).

Ema2 is required for completing DNA elimination

We next asked if Ema2 is involved in DNA elimination. The production of the *EMA2* somatic KO strains, which lack the translational start site and the first half of the *EMA2* coding sequence of all the *EMA2* copies in the polyploid MAC, were described previously (Shehzada and Mochizuki 2022). Two *EMA2* KO strains possessing complementary mating types were mated and DNA elimination in their exconjugants (progeny) was analyzed by DNA fluorescent in situ hybridization (FISH) using probes complementary to the moderately repeated transposable element Tlr1 (Wuitschick et al. 2002). Consistent with the fact that DNA elimination is normally completed by 14-16 hpm (Aronica et al. 2008), the Tlr1 element was detected only in the MICs, but not in the MACs, in the exconjugants from wild-type cells at 36 hpm (Figure 2A, WT). In contrast, the Tlr1 element was detected in both the MICs and the MACs in the exconjugants from *EMA2* KO cells

even at 36 hpm (Figure 2A, *EMA2* KO), indicating that *EMA2* is a novel gene involved in DNA elimination.

We noticed that DNA elimination was not completely blocked in the exconjugants of *EMA2* KO cells, as the intensity of the Tlr1-FISH signal in the new MACs (indicative as IES retention index in Figure 2B) was lower in the exconjugants from *EMA2* KO cells than in those from the *TWI1* KO cells, in which DNA elimination was reported to be completely blocked (Mochizuki et al. 2002). This partial block may be due to the presence of the intact *EMA2* copies in the MIC in *EMA2* KO cells and thus *EMA2* mRNA can be expressed zygotically in the new MAC. Indeed, *EMA2* mRNA shows a bi-phasic expression pattern in the wild-type cells with peaks first at 2 hpm, when the parental MAC is transcribed, and then at 8 hpm, when the parental MAC is being degenerated and the new MAC becomes transcriptionally active (Figure 1A) (Martindale et al. 1985; Mochizuki and Gorovsky 2004b; Sugai and Hiwatashi 1974).

Consistent with the previous indication that DNA elimination is required for making viable progeny (Noto and Mochizuki 2017; Aronica et al. 2008; Noto et al. 2010; Mochizuki et al. 2002; Malone et al. 2005; Mochizuki and Gorovsky 2005; Liu et al. 2007), *EMA2* KO strains did not produce viable sexual progeny (Figure 2C). Altogether, we conclude that the maternally-expressed *EMA2* is required for completing DNA elimination in the new MAC and for producing viable progeny.

Ema2 is important for heterochromatin formation and heterochromatin body assembly in the new MAC

It has been reported that the heterochromatin marks H3K9me3 and H3K27me3 are specifically deposited to IESs in the new MAC and play essential roles in DNA elimination (Taverna et al. 2002; Liu et al. 2007, 2004; Xu et al. 2021). Because *EMA2* is required for completing DNA elimination, we next examined if the heterochromatin formation was affected in *EMA2* KO cells by immunofluorescence staining.

As previously reported, in wild-type cells, both H3K9me3 and H3K27me3 localized homogeneously in the new MACs at 9 hpm and were detected in the nuclear foci at 12 hpm (Figure 3A and B, WT). These foci concomitantly occur with the excision step of DNA elimination (Kataoka and Mochizuki 2015), which starts at around 10 hpm (Yao and Yao 1994; Saveliev and Cox 2001; Mutazono et al. 2019). In *EMA2* KO cells, both H3K9me3 and H3K27me3 were greatly reduced at 9 hpm (Figure 3A and B, 9 hpm, *EMA2* KO). In contrast, at 12 hpm, these

heterochromatic marks occurred more abundantly in *EMA2* KO cells than in wild-type cells and did not constitute distinct foci (Figure 3A and B, 12 hpm, *EMA2* KO).

The delayed appearance of H3K9me3 and H3K27me3 in *EMA2* KO cells indicates that *EMA2* is required for the heterochromatin formation in the new MAC and the zygotically-expressed *EMA2* in *EMA2* KO cells can partially compensate the heterochromatin formation. Although this delay may fully explain the block of the heterochromatin foci formation and DNA elimination in *EMA2* KO cells, as described below, *EMA2* has an additional role in the parental MAC, which is also important for regulating DNA elimination in the new MAC.

Ema2 is required for heterochromatin formation in the parental MAC

It was reported that H3K27me3 not only occurs in the new MACs but also in the parental MAC during mid-conjugation stages in a manner dependent on the scnRNA-directed heterochromatin formation machineries such as Dcl1, Twi1 and Ezl1 (Liu et al. 2007). We, therefore, analyzed the accumulation of H3K27me3 in the parental MAC during conjugation in the *EMA2* KO cells. As previously reported, H3K27me3 was detected in the parental MAC at 6 hpm in wild-type cells but not in *TWI1* and *EZL1* KO cells (Figure 4). We found that H3K27me3 was also undetectable in the parental MAC in *EMA2* KO cells at the same stage (Figure 4). Therefore, although the biological role of H3K27me3 in the parental MAC during conjugation is unclear, these data further support our conclusion that Ema2 is required for the scnRNA-directed heterochromatin formation.

Ema2 is important for Target-directed small RNA degradation (TDSD)

The phenotypes of *EMA2* KO cells (the partial block of DNA elimination [Figure 2], the delayed accumulation of the heterochromatic marks in the new MAC, and the loss of the heterochromatin foci formation [Figure 3]) are reminiscent of the phenotypes of the *EMA1* somatic KO cells (Aronica et al. 2008). In addition, Ema1 is also required for the accumulation of H3K27me3 in the parental MAC (Figure 4) and first appears in the parental MAC and then in the new MAC (Aronica et al. 2008). Because *EMA1* is also required for Target-directed small RNA degradation (TDSD) of scnRNAs in the parental MAC (Aronica et al. 2008; Schoeberl et al. 2012), we next asked if *EMA2* is also involved in this process.

We previously reported that scnRNAs are predominantly produced from IESs but some are also from their surrounding MDS regions in the MIC at the early conjugation stages (~2-3 hpm). Then, scnRNAs complementary to the parental MAC genome (= MDSs) are subjected to TDSD

in the parental MAC in the mid-conjugation stages (~3-7 hpm) (Mochizuki and Gorovsky 2004a; Aronica et al. 2008; Schoeberl et al. 2012; Noto et al. 2015). As a result, only scnRNAs complementary to IESs are selectively retained (scnRNA selection) and later used for specifically targeting IESs for the heterochromatin formation and the following DNA elimination in the new MAC.

To assess TDSD, we first compared scnRNAs at different conjugation stages by northern blot using the 50-nt Mi-9 probe that is complementary to a repetitive sequence found in both IESs and MDSs (Aronica et al. 2008). Because scnRNAs detected by this probe are complementary to MDSs (in addition to IESs), they are subjected to TDSD. Indeed, in wild-type cells, the Mi-9-complementary scnRNAs were detected at 3 hpm, reduced at 4.5 hpm, and became undetectable at 6 hpm (Figure 5, top, WT). In contrast, the Mi-9-complementary scnRNAs remained detectable at 6 hpm in *EMA2* KO cells (Figure 5, top, KO), suggesting that Ema2 is required for TDSD of scnRNAs.

We also analyzed TDSD in a genome-wide scale by sequencing small RNAs at 3 hpm and 8 hpm and mapping 26- to 32-nt small RNAs, which correspond to scnRNAs (Schoeberl et al. 2012) to the genomic tiles of MDSs and IESs (see Materials and Methods). Although the normalized number of sequenced small RNAs complementary to the MDS tiles was greatly reduced from 3 to 8 hpm due to TDSD in wild-type cells (Figure 5B, WT), those in two independent crosses of *EMA2* KO cells only slightly reduced by 8 hpm (Figure 5B, *EMA2* KO). This TDSD defect in *EMA2* KO cells seems milder than that in *EMA1* KO cells, in which the amount of MDS-complementary small RNAs remained constant till 8 hpm (Figure 5B, *EMA1* KO). This difference will be discussed later. In contrast, the numbers of sequenced small RNAs complementary to the IES tiles at 3 hpm and 8 hpm were comparable in both wild-type and *EMA2* KO cells (Figure 5C), consistent with our previous observations that scnRNAs complementary to IES are not the targets of TDSD. These results suggest that *EMA2* is important for TDSD of scnRNAs at the genome-wide level.

Ema2 is required for the accumulation of SUMOylated proteins during conjugation

Ema2 possesses a zinc-finger domain at the N-terminus and an atypical SP-RING domain at the middle (Figure 6A), in the latter of which the conserved second cysteine that consists of one of the four zinc-binding residues seems to be replaced by histidine (Figure 6A, Supplementary Figure 1). Because the SP-RING domain was found in many SUMO E3 ligases and required for their interaction with the SUMO E2 enzyme (Rytinki et al. 2009; Kahyo et al. 2001), we predicted that Ema2 acts as a SUMO E3 ligase in *Tetrahymena*.

We, therefore, asked if Ema2 can directly interact with the SUMO E2 enzyme, which is encoded by *UBC9* in *Tetrahymena* (Yang et al. 2015). GST-tagged Ema2 (GST-Ema2) and GST-alone were expressed in *E. coli* and separately immobilized on beads. Then, His-tagged Ubc9 (His-Ubc9) was also prepared recombinantly from *E. coli,* incubated with the immobilized GST-Ema2 and GST-alone, and bound protein was analyzed by western blot using an anti-His antibody. As shown in Figure 6B, His-Ubc9 is specifically bound to GST-Ema2, supporting our prediction that Ema2 is a SUMO E3 ligase.

Next, we compared SUMOylated proteins during conjugation in the presence or the absence of Ema2 in *Tetrahymena*. SUMO is solely encoded by *SMT3* in *Tetrahymena* (Nasir et al. 2015). To detect SUMOylated proteins, we introduced a construct expressing HA-tagged Smt3 (HA-Smt3), which can replace the essential function of the endogenous Smt3 (unpublished), into an *EMA2* KO strain and crossed with a wild-type strain (called WT cross) or another *EMA2* KO strain (called *EMA2* KO cross) (Figure 6C). Because proteins and mRNAs are exchanged between two mating pairs through the conjugation junction (McDonald 1966), *EMA2* mRNA/Ema2 protein expressed in the wild-type partner of WT cross are expected to move into the *EMA2* KO partner cells and restores the *EMA2* KO phenotypes. We also expect that *HA-SMT3* mRNA/HA-Smt3 protein expressed in the *EMA2* KO partner contribute to SUMOylation in both sides of the pairs in both crosses.

Total proteins were harvested at 4.5 and 6 hpm and SUMOylated proteins were detected by western blot using an anti-HA antibody (Figure 6D). In both time points, high molecular weight proteins (mainly >200 kDa) were detected in WT cross (Figure 6D, WT) and they were greatly reduced in the *EMA2* KO cross (Figure 6D, *EMA2* KO), although ~70% of SUMOylated proteins still retained in this cross (Figure 6E). The Ema2-independent SUMOylations most probably occurred by other SUMO E3 ligases and/or E3-independent SUMOylations. Altogether, we conclude that Ema2 is a major, if not the sole, SUMO E3 ligase in *Tetrahymena* during the early- to mid-conjugation stages.

The conserved SP-RING domain is required for the function of Ema2

In many SUMO E3 ligases of other eukaryotes, the conserved zinc-binding residues in the SP-RING domain are essential for their activities (Crozet et al. 2016; Kahyo et al. 2001; Munarriz et al. 2004; Takahashi et al. 2001a; Ninova et al. 2020a). We found that, while ectopic expression of HA-tagged wild-type Ema2 in the *EMA2* KO cells restored the accumulation of SUMOylated proteins (Figure 6F, WT rescue) and DNA elimination (Figure 2B, WT rescue), the HA-tagged C376A/C379A mutant of Ema2, in which the two conserved cysteines (marked with * in Figure 5A) were replaced with alanines, did not rescue the defects of *EMA2* KO in SUMOylation (Figure 6F, Mut rescue) and DNA elimination (Figure 2B, Mut rescue). Therefore, the SP-RING domain is essential for the function of Ema2.

Ema2 is required for SUMOylation of Spt6

Next, to identify the SUMOylation target(s) of Ema2, we introduced a construct expressing His-tagged Smt3 into an *EMA2* KO strain and crossed the strain with a wild-type strain (WT cross) or another *EMA2* KO strain (*EMA2* KO cross). Then, SUMOylated proteins were concentrated using a nickel column and identified by mass spectrometry. Although most of the proteins were detected similarly between WT cross and *EMA2* KO cross, Spt6 was detected more abundantly from WT cross (Figure 7A).

To confirm that Spt6 is SUMOylated in an Ema2-dependent manner, we introduced a construct expressing Spt6 fused with HA-tag at the C-terminus (Spt6-HA) from the endogenous *SPT6* locus into an *EMA2* KO strain and mated with a wild-type strain (WT cross) or another *EMA2* KO strain (*EMA2* KO cross). Then, total proteins harvested at 4.5 and 6 hpm were analyzed by western blot using an anti-HA antibody (Figure 7B). In WT cross, slower migrating population of Spt6-HA was detected in addition to the band corresponding to the unmodified Spt6 at the both time points (Figure 7B, WT). In contrast, such slower migrating Spt6-HA was undetectable in *EMA2* KO cross (Figure 7B, *EMA2* KO). These *EMA2*-dependent slower migrating species of Spt6-HA overlap with the major *EMA2*-dependent SUMOylated proteins detected in Figure 6D, suggesting that they are SUMOylated Spt6-HA. This was confirmed by the fact that among the immunopurified Spt6-HA by an anti-HA antibody (Figure 7C, left, WT) the slower migrating species were recognized by an anti-Smt3 antibody (Figure 7C, right, WT). The slower migrating Spt6-HA species were also detected in *EMA2* KO cross after anti-HA immunoprecipitation, although their amount was greatly lower than that in WT cross (Figure 7C). This suggests that

some level of SUMOylation still occurs on Spt6 without Ema2. All these results indicate that Ema2 is required for the majority of SUMOylation of Spt6 during conjugation.

Ema2 is required for accumulation of parental MAC lncRNA transcripts

Spt6 is a conserved transcriptional regulator reported to be involved in multiple phases of transcription from initiation, elongation, splicing, termination to RNA turnover in various eukaryotes (Nojima et al. 2018; Dronamraju et al. 2018; Clark-Adams and Winston 1987). Because we previously suggested that TDSD of scnRNAs should be preceded by the base-pairing interaction between scnRNAs and nascent long-noncoding (lnc)RNAs at the parental MAC chromosomes (Aronica et al. 2008), we speculated that the Ema2-dependent Spt6 SUMOylation is necessary for the transcription or the stability of the lncRNAs in the parental MAC.

MAC lncRNAs were first examined by RT-PCR. Because the MAC mostly shares identical sequences with the MIC and the MIC is also transcribed during conjugation to produce scnRNAs (Mochizuki and Gorovsky 2005; Malone et al. 2005; Aronica et al. 2008), we designed RT-PCR assays which amplify transcripts spanning IES-MDS borders and thus can distinguish lncRNA transcripts from the two nuclei by the lengths of PCR products (Figure 8A). Total RNA isolated from wild-type and *EMA2* KO cells at 6 hpm was used for the RT-PCR assays. For the all three loci examined, the MAC-derived lncRNA transcripts were detected in wild-type cells but not in *EMA2* KO cells, although the *RPL21* mRNA was detected similarly in the both conditions (Figure 8B). The lack of the lncRNA transcripts in *EMA2* KO cells was not due to loss of the primer binding sites by alternative DNA elimination in their prior sexual reproductions because similar sizes of genomic PCR products were detected from the parental wild-type and *EMA2* KO cells with the same primer sets (Figure 8C). We therefore conclude that *EMA2* is required for the accumulation of lncRNA transcripts in the parental MAC at least for the three tested loci.

The lncRNA accumulation in the MAC was also examined by immunofluorescent staining using the anti-long-double-stranded (ds)RNA antibody J2 (Schönborn et al. 1991). It was reported that long dsRNAs were detected by this antibody specifically during conjugation in *Tetrahymena*, first in the MIC at its pre-meiotic and meiotic stages (Woo et al. 2016; Tian et al. 2019), and then in the parental MAC at the mid stages of conjugation when TDSD of scnRNAs undergoes (Woo et al. 2016). The former is believed to be the precursors of scnRNAs (Aronica et al. 2008; Malone et al. 2005; Mochizuki and Gorovsky 2005) while the latter to be a result of the genome-wide bidirectional transcription for providing lncRNA transcripts for TDSD (Woo et al. 2016; Aronica et al. 2008). Consistent with the normal production of scnRNAs in wild-type and *EMA2* KO cells (Figure 5A, bottom), long dsRNAs were detected in the MIC at 3 hpm in the both conditions. In

contrast, although long dsRNAs were detected in the MAC at 6 hpm in the wild-type cells as previously reported, it was undetectable in *EMA2* KO cells (Figure 9). Interestingly, *EMA1* was dispensable for the accumulation of the long dsRNA in the MAC (Figure 9, *EMA*1 KO), suggesting that *EMA1* and *EMA2* act at different steps of TDSD. All these results strongly suggest that *EMA2* is required for TDSD of scnRNAs by promoting the accumulation of lncRNAs in the parental MAC.

Discussion

In this study, we identified a novel protein Ema2, which is required for TDSD of scnRNAs by promoting lncRNA transcription in *Tetrahymena*. Ema2 is a SUMO E3 ligase responsible for the accumulation of SUMOylated Spt6 during conjugation. Therefore, we propose that Ema2-dependent SUMOylation of Spt6 promotes the genome-wide lncRNA transcriptions that is the prerequisite of TDSD of scnRNAs.

Ema2 is required for TDSD by promoting lncRNA transcription

Based on our previous observations that the RNA helicase Ema1 is required for the interaction between the nascent lncRNA and the Twi1-scnRNA complex and for TDSD of scnRNAs in the parental MAC, we have proposed that scnRNAs are targeted for TDSD by their interaction with the parental MAC genome through nascent lncRNAs in *Tetrahymena* (Aronica et al. 2008; Noto and Mochizuki 2018). A similar mode of TDSD of scnRNAs was also suggested in *Paramecium* and the importance of lncRNAs for proper DNA elimination has been demonstrated by disrupting some lncRNA using RNAi in this ciliate (Lepère et al. 2008). However, there was no direct demonstration for the requirement of lncRNAs in TDSD of scnRNAs. *EMA2* KO described in this study is the first mutant that eliminate lncRNAs without obviously affecting the vegetative cell growth and the conjugation processes (except DNA elimination). Therefore, this paper provides the most direct evidence for the necessary of the parental MAC lncRNAs for TDSD of scnRNAs. We found that, in the genome-wide scale, TDSD of scnRNAs is more severely affected in *EMA 1* KO than in *EMA2* KO (Figure 5B). We speculate that TDSD of scnRNAs might still occur in the absence of Ema2 through Ema2-independent mRNA and/or ncRNAs transcriptions, while Ema1 is necessary for the all the events of TDSD of scnRNAs because it is generally required for the scnRNA-lncRNA interaction. We currently lack genome-wide view of the lncRNA transcription in the parental MAC and thus the transcriptional status of the regions complementary to scnRNAs that are subjected to TDSD in *EMA1*-dependent and *EMA2*-independent manner is unclear. Future study should establish a method to specifically investigate transcriptional status of the parental MAC at a genome-wide level without being disturbed by the lncRNA transcriptions of the MIC.
We found that Ema2 is also required for the accumulation of H3K27me3 in the parental MAC at mid-conjugation stages prior to the formation of the new MAC, indicating that the scnRNA-lncRNA interaction results in both TDSD and heterochromatin formation in the parental MAC. The fact that *EZL1,* which encodes the H3K9 and H3K27 dual-specific histone methyltransferase

(Liu et al. 2007; Xu et al. 2021), is not necessary for TDSD of scnRNAs (unpublished) indicating that H3K27me3 is not a prerequisite of the TDSD in the parental MAC. We also previously observed that scnRNAs complementary to IESs are degraded in the new MAC in parallel to the heterochromatin formation (Noto et al. 2015). Therefore, although it is unclear what is the role of TDSD of scnRNAs in the new MAC, both TDSD and heterochromatin formation are likely the common outputs of the scnRNA-lncRNA interaction in both the parental and the new MAC.

Ema2 is required for the SUMOylation of the transcriptional regulator Spt6

EMA2 encodes a SUMO E3 ligase and its disruption resulted in the loss of ~30% of the total SUMOylated proteins in the mid-conjugation stages. Therefore, we suggest that the requirement of Ema2 for TDSD of scnRNAs, heterochromatin formation and DNA elimination is due to the loss of Ema2-dependent SUMOylation in some protein(s). The requirement of protein SUMOylation in DNA elimination was demonstrated in *Paramecium* by RNAi knock-down of *UBA2* and *SUMO* (Matsuda and Forney 2006). However, because the SUMOylation machineries are essential for normal vegetative growth and for the mating pair formation in *Tetrahymena*, the significance of SUMOylation in DNA elimination could not be tested by the genetic disruptions of *UBA2* and *SMT3* in *Tetrahymena* (Nasir et al. 2015). This study showed that the SUMO pathway is also required for DNA elimination in *Tetrahymena*. Furthermore, this is the first report demonstrating that a SUMOylation machinery is necessary for both TDSD of scnRNAs and heterochromatin formation, the key processes of DNA elimination, through promoting lncRNA transcription.

We identified Spt6 as a major, if not the sole, SUMOylation target of Ema2 (Figure 7). Therefore, we predict that Ema2 regulates the lncRNA transcription in the parental MAC by promoting SUMOylation of Spt6. How the SUMOylation of Spt6 up-regulates the lncRNA transcription remains fully speculative. It was reported that the loss of *SPT6* function resulted in abnormal up-regulation of anti-sense and intergenic transcriptions in yeasts and human cells (DeGennaro et al. 2013; Kato et al. 2013; Doris et al. 2018; Narain et al. 2021). We imagine that, in *Tetrahymena*, the un-SUMOylated Spt6 also suppresses the occurrence of promiscuous lncRNA transcription by inhibiting promiscuous transcriptional initiations and/or promoting transcriptional terminations and SUMOylation of Spt6 loosens such safeguarding functions to promote genome wide lncRNA transcription and thus TDSD of scnRNAs.

Although the idea above is attractive, it remains possible that Ema2-dependent SUMOylation of a protein other than Spt6, or even a SUMOylation-independent function of Ema2, is important for

the lncRNA transcription. To test if Ema2-dependent SUMOylation of Spt6 is indeed necessary for the lncRNA transcription, future study should establish experimental systems for 1) identifying individual SUMOylated residues in *Tetrahymena* by genetically engineering *SMT3* to make proteolyzed Smt3-conjugated protein fragments detectable by mass-spectrometry, which allows us to compare SUMOylated residues of Spt6 in the presence and absence of *EMA2*; and 2) examining the roles of the identified SUMOylated residues by first establishing *SPT6* germline KO strains (in which the both copies of *SPT6* in the MIC are disrupted) and then introducing un-SUMOylatable *SPT6* mutants into the new MAC of the progeny of the *SPT6* germline KO strains. These efforts are ongoing and will clarify if and how Ema2-dependent SUMOylation of Spt6 regulates lncRNA transcription in *Tetrahymena*.

It was reported that SUMOylation occurs on Swi6, Chp2 and Clr4 and is important for RNAi-directed heterochromatin silencing in fission yeast (Shin et al. 2005). Also, SUMOylation of the histone deacetylase HAD-1 promotes small RNA-directed transcriptional silencing in *C. elegens* (Kim et al. 2021). In the piRNA-directed transposon silencing of *Drosophila*, the SUMO E3 ligase Su(var)2-10 plays an essential role for the recruitment of the histone methyltransferase complex SetDB1 that deposits H3K9me3 (Ninova et al. 2020a, 2020b). Moreover, Panoramix (Panx), which is recruited to nascent lncRNA by Piwi-piRNA complex, is SUMOylated independently of Su(var)2-10 and this SUMOylation is important for transcriptional silencing by promoting its interaction with the general heterochromatin effector Sov (Andreev et al. 2022). All these SUMOylation events seem to occur downstream of the nascent lncRNA transcription, and thus our observation that SUMOylation regulates small RNA-directed heterochromatin formation at the level of nascent lncRNA transcription adds an undescribed layer of regulatory mechanism. We expect that, because SUMOylation often occurs pervasively on many proteins act in the same biological pathways (Bhagwat et al. 2021) and SUMOylations were detected on multiple regulators of RNA polymerase II in *Drosophila* (Ninova et al. 2020b), the appearance of nascent lncRNAs is also promoted by SUMOylation of transcriptional regulators, not necessary Spt6, in different small RNA-directed pathways of various eukaryotes.

References

Andreev VI, Yu C, Wang J, Schnabl J, Tirian L, Gehre M, Handler D, Duchek P, Novatchkova M, Baumgartner L, et al. 2022. Panoramix SUMOylation on chromatin connects the piRNA pathway to the cellular heterochromatin machinery. *Nat Struct Mol Biol* **29**: 130–142.

Aronica L, Bednenko J, Noto T, DeSouza LV, Siu KWM, Loidl J, Pearlman RE, Gorovsky MA, Mochizuki K. 2008. Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in Tetrahymena. *Genes Dev* **22**: 2228–2241.

Batki J, Schnabl J, Wang J, Handler D, Andreev VI, Stieger CE, Novatchkova M, Lampersberger L, Kauneckaite K, Xie W, et al. 2019. The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. *Nat Struct Mol Biol* **26**: 720–731.

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE. 2008. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. *PLoS ONE* **3**: e3156.

Bhagwat NR, Owens SN, Ito M, Boinapalli JV, Poa P, Ditzel A, Kopparapu S, Mahalawat M, Davies OR, Collins SR, et al. 2021. SUMO is a pervasive regulator of meiosis. *eLife* **10**.

Bruns PJ, Cassidy-Hanley D. 2000. Biolistic transformation of macro- and micronuclei. *Methods Cell Biol* **62**: 501–512.

Bühler M, Verdel A, Moazed D. 2006. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**: 873–886.

Capili AD, Lima CD. 2007. Taking it step by step: mechanistic insights from structural studies of ubiquitin/ubiquitin-like protein modification pathways. *Curr Opin Struct Biol* **17**: 726–735.

Chalker DL, Meyer E, Mochizuki K. 2013. Epigenetics of ciliates. *Cold Spring Harb Perspect Biol* **5**: a017764.

Cheng C-Y, Vogt A, Mochizuki K, Yao M-C. 2010. A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila. *Mol Biol Cell* **21**: 1753–1762.

Clark-Adams CD, Winston F. 1987. The SPT6 gene is essential for growth and is required for delta-mediated transcription in Saccharomyces cerevisiae. *Mol Cell Biol* **7**: 679–686.

Coyne RS, Chalker DL, Yao MC. 1996. Genome downsizing during ciliate development: nuclear division of labor through chromosome restructuring. *Annu Rev Genet* **30**: 557–578.

Crozet P, Margalha L, Butowt R, Fernandes N, Elias CA, Orosa B, Tomanov K, Teige M, Bachmair A, Sadanandom A, et al. 2016. SUMOylation represses SnRK1 signaling in Arabidopsis. *Plant J* **85**: 120–133.

Davis MC, Ward JG, Herrick G, Allis CD. 1992. Programmed nuclear death: apoptotic-like degradation of specific nuclei in conjugating Tetrahymena. *Dev Biol* **154**: 419–432.

DeGennaro CM, Alver BH, Marguerat S, Stepanova E, Davis CP, Bähler J, Park PJ, Winston F. 2013. Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. *Mol Cell Biol* **33**: 4779–4792.

Djupedal I, Portoso M, Spåhr H, Bonilla C, Gustafsson CM, Allshire RC, Ekwall K. 2005. RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev* **19**: 2301–2306.

Doris SM, Chuang J, Viktorovskaya O, Murawska M, Spatt D, Churchman LS, Winston F. 2018. Spt6 is required for the fidelity of promoter selection. *Mol Cell* **72**: 687-699.e6.

Dronamraju R, Hepperla AJ, Shibata Y, Adams AT, Magnuson T, Davis IJ, Strahl BD. 2018. Spt6 Association with RNA Polymerase II Directs mRNA Turnover During Transcription. *Mol Cell* **70**: 1054-1066.e4.

Enke RA, Dong Z, Bender J. 2011. Small RNAs prevent transcription-coupled loss of histone H3 lysine 9 methylation in Arabidopsis thaliana. *PLoS Genet* **7**: e1002350.

Eulalio A, Huntzinger E, Izaurralde E. 2008. Getting to the root of miRNA-mediated gene silencing. *Cell* **132**: 9–14.

Gareau JR, Lima CD. 2010. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol* **11**: 861–871.

Geiss-Friedlander R, Melchior F. 2007. Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol* **8**: 947–956.

Goldfarb DS, Gorovsky MA. 2009. Nuclear dimorphism: two peas in a pod. *Curr Biol* **19**: R449-52.

Gorovsky MA, Yao MC, Keevert JB, Pleger GL. 1975. Isolation of micro- and macronuclei of Tetrahymena pyriformis. *Methods Cell Biol* **9**: 311–327.

Grewal SIS, Jia S. 2007. Heterochromatin revisited. *Nat Rev Genet* **8**: 35–46.

Gu SG, Pak J, Guang S, Maniar JM, Kennedy S, Fire A. 2012. Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nat Genet* **44**: 157–164.

Hamilton EP, Orias E. 2000. Genetic crosses: setting up crosses, testing progeny, and isolating phenotypic assortants. *Methods Cell Biol* **62**: 219–228.

Hendriks IA, Lyon D, Young C, Jensen LJ, Vertegaal ACO, Nielsen ML. 2017. Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. *Nat Struct Mol Biol* **24**: 325–336.

Holoch D, Moazed D. 2015. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet* **16**: 71–84.

Kahyo T, Nishida T, Yasuda H. 2001. Involvement of PIAS1 in the sumoylation of tumor suppressor p53. *Mol Cell* **8**: 713–718.

Kataoka K, Mochizuki K. 2015. Phosphorylation of an HP1-like Protein Regulates Heterochromatin Body Assembly for DNA Elimination. *Dev Cell* **35**: 775–788.

Kataoka K, Schoeberl UE, Mochizuki K. 2010. Modules for C-terminal epitope tagging of Tetrahymena genes. *J Microbiol Methods* **82**: 342–346.

Kato H, Goto DB, Martienssen RA, Urano T, Furukawa K, Murakami Y. 2005. RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science* **309**: 467–469.

Kato H, Okazaki K, Iida T, Nakayama J-I, Murakami Y, Urano T. 2013. Spt6 prevents transcription-coupled loss of posttranslationally modified histone H3. *Sci Rep* **3**: 2186.

Kerscher O, Felberbaum R, Hochstrasser M. 2006. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol* **22**: 159–180.

Kim H, Ding Y-H, Zhang G, Yan Y-H, Conte D, Dong M-Q, Mello CC. 2021. HDAC1 SUMOylation promotes Argonaute-directed transcriptional silencing in C. elegans. *eLife* **10**.

Lee SR, Collins K. 2006. Two classes of endogenous small RNAs in Tetrahymena thermophila. *Genes Dev* **20**: 28–33.

Le Thomas A, Marinov GK, Aravin AA. 2014. A transgenerational process defines piRNA biogenesis in Drosophila virilis. *Cell Rep* **8**: 1617–1623.

Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Tóth KF. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev* **27**: 390–399.

Lepère G, Bétermier M, Meyer E, Duharcourt S. 2008. Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in Paramecium tetraurelia. *Genes Dev* **22**: 1501–1512.

Liu Y, Mochizuki K, Gorovsky MA. 2004. Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena. *Proc Natl Acad Sci USA* **101**: 1679–1684.

Liu Y, Taverna SD, Muratore TL, Shabanowitz J, Hunt DF, Allis CD. 2007. RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. *Genes Dev* **21**: 1530–1545.

Loidl J, Scherthan H. 2004. Organization and pairing of meiotic chromosomes in the ciliate Tetrahymena thermophila. *J Cell Sci* **117**: 5791–5801.

Madireddi MT, Coyne RS, Smothers JF, Mickey KM, Yao MC, Allis CD. 1996. Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena. *Cell* **87**: 75–84.

Malone CD, Anderson AM, Motl JA, Rexer CH, Chalker DL. 2005. Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of *Tetrahymena thermophila*. *Mol Cell Biol* **25**: 9151–9164.

Martindale DW, Allis CD, Bruns PJ. 1982. Conjugation in Tetrahymena thermophila. A temporal analysis of cytological stages. *Exp Cell Res* **140**: 227–236.

Martindale DW, Allis CD, Bruns PJ. 1985. RNA and protein synthesis during meiotic prophase in Tetrahymena thermophila. *J Protozool* **32**: 644–649.

Matsuda A, Forney JD. 2006. The SUMO pathway is developmentally regulated and required for programmed DNA elimination in Paramecium tetraurelia. *Eukaryotic Cell* **5**: 806–815.

McDonald BB. 1966. The exchange of RNA and protein during conjugation in Tetrahymena. *J Protozool* **13**: 277–285.

Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA. 2002. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell* **110**: 689–699.

Mochizuki K, Gorovsky MA. 2005. A Dicer-like protein in *Tetrahymena* has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes Dev* **19**: 77–89.

Mochizuki K, Gorovsky MA. 2004a. Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev* **18**: 2068–2073.

Mochizuki K, Gorovsky MA. 2004b. RNA polymerase II localizes in Tetrahymena thermophila meiotic micronuclei when micronuclear transcription associated with genome rearrangement occurs. *Eukaryotic Cell* **3**: 1233–1240.

Motamedi MR, Verdel A, Colmenares SU, Gerber SA, Gygi SP, Moazed D. 2004. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**: 789–802.

Müller S, Hoege C, Pyrowolakis G, Jentsch S. 2001. SUMO, ubiquitin's mysterious cousin. *Nat Rev Mol Cell Biol* **2**: 202–210.

Munarriz E, Barcaroli D, Stephanou A, Townsend PA, Maisse C, Terrinoni A, Neale MH, Martin SJ, Latchman DS, Knight RA, et al. 2004. PIAS-1 is a checkpoint regulator which affects exit from G1 and G2 by sumoylation of p73. *Mol Cell Biol* **24**: 10593–10610.

Mutazono M, Noto T, Mochizuki K. 2019. Diversification of small RNA amplification mechanisms for targeting transposon-related sequences in ciliates. *Proc Natl Acad Sci USA* **116**: 14639–14644.

Narain A, Bhandare P, Adhikari B, Backes S, Eilers M, Dölken L, Schlosser A, Erhard F, Baluapuri A, Wolf E. 2021. Targeted protein degradation reveals a direct role of SPT6 in RNAPII elongation and termination. *Mol Cell* **81**: 3110-3127.e14.

Nasir AM, Yang Q, Chalker DL, Forney JD. 2015. SUMOylation is developmentally regulated and required for cell pairing during conjugation in Tetrahymena thermophila. *Eukaryotic Cell* **14**: 170–181.

Ninova M, Chen Y-CA, Godneeva B, Rogers AK, Luo Y, Fejes Tóth K, Aravin AA. 2020a. Su(var)2-10 and the SUMO Pathway Link piRNA-Guided Target Recognition to Chromatin Silencing. *Mol Cell* **77**: 556-570.e6.

Ninova M, Godneeva B, Chen Y-CA, Luo Y, Prakash SJ, Jankovics F, Erdélyi M, Aravin AA, Fejes Tóth K. 2020b. The SUMO Ligase Su(var)2-10 Controls Hetero- and Euchromatic Gene Expression via Establishing H3K9 Trimethylation and Negative Feedback Regulation. *Mol Cell* **77**: 571-585.e4.

Nojima T, Tellier M, Foxwell J, Ribeiro de Almeida C, Tan-Wong SM, Dhir S, Dujardin G, Dhir A, Murphy S, Proudfoot NJ. 2018. Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. *Mol Cell* **72**: 970-984.e7.

Noto T, Kataoka K, Suhren JH, Hayashi A, Woolcock KJ, Gorovsky MA, Mochizuki K. 2015. Small-RNA-Mediated Genome-wide trans-Recognition Network in Tetrahymena DNA Elimination. *Mol Cell* **59**: 229–242.

Noto T, Kurth HM, Kataoka K, Aronica L, DeSouza LV, Siu KWM, Pearlman RE, Gorovsky MA, Mochizuki K. 2010. The Tetrahymena argonaute-binding protein Giw1p directs a mature argonaute-siRNA complex to the nucleus. *Cell* **140**: 692–703.

Noto T, Mochizuki K. 2018. Small RNA-Mediated trans-Nuclear and trans-Element Communications in *Tetrahymena* DNA Elimination. *Curr Biol* **28**: 1938-1949.e5.

Noto T, Mochizuki K. 2017. Whats, hows and whys of programmed DNA elimination in *Tetrahymena*. *Open Biol* **7**: 170172.

Onishi R, Yamanaka S, Siomi MC. 2021. piRNA- and siRNA-mediated transcriptional repression in Drosophila, mice, and yeast: new insights and biodiversity. *EMBO Rep* **22**: e53062.

Pezic D, Manakov SA, Sachidanandam R, Aravin AA. 2014. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes Dev* **28**: 1410–1428.

Rytinki MM, Kaikkonen S, Pehkonen P, Jääskeläinen T, Palvimo JJ. 2009. PIAS proteins: pleiotropic interactors associated with SUMO. *Cell Mol Life Sci* **66**: 3029–3041.

Saveliev SV, Cox MM. 2001. Product analysis illuminates the final steps of IES deletion in Tetrahymena thermophila. *EMBO J* **20**: 3251–3261.

Schoeberl UE, Kurth HM, Noto T, Mochizuki K. 2012. Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in *Tetrahymena. Genes Dev* **26**: 1729–1742.

Schönborn J, Oberstrass J, Breyel E, Tittgen J, Schumacher J, Lukacs N. 1991. Monoclonal antibodies to double-stranded RNA as probes of RNA structure in crude nucleic acid extracts. *Nucleic Acids Res* **19**: 2993–3000.

Shehzada S, Mochizuki K. 2022. PIWI-Directed DNA Elimination for Tetrahymena Genetics. *Methods Mol Biol* **2509**: 53–68.

Shin JA, Choi ES, Kim HS, Ho JCY, Watts FZ, Park SD, Jang YK. 2005. SUMO modification is involved in the maintenance of heterochromatin stability in fission yeast. *Mol Cell* **19**: 817–828.

Sugai T, Hiwatashi K. 1974. Cytologic and autoradiographic studies of the micronucleus at meiotic prophase in Tetrahymena pyriformis. *J Protozool* **21**: 542–548.

Takahashi Y, Kahyo T, Toh-E A, Yasuda H, Kikuchi Y. 2001. Yeast Ull1/Siz1 is a novel SUMO1/Smt3 ligase for septin components and functions as an adaptor between conjugating enzyme and substrates. *J Biol Chem* **276**: 48973–48977.

Taverna SD, Coyne RS, Allis CD. 2002. Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. *Cell* **110**: 701–711.

Tian M, Mochizuki K, Loidl J. 2019. Non-coding RNA Transcription in Tetrahymena Meiotic Nuclei Requires Dedicated Mediator Complex-Associated Proteins. *Curr Biol* **29**: 2359-2370.e5.

Vertegaal ACO. 2022. Signalling mechanisms and cellular functions of SUMO. *Nat Rev Mol Cell Biol*.

Vogt A, Mochizuki K. 2013. A domesticated PiggyBac transposase interacts with heterochromatin and catalyzes reproducible DNA elimination in *Tetrahymena*. *PLoS Genet* **9**: e1004032.

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**: 1833–1837.

Woo T-T, Chao J-L, Yao M-C. 2016. Dynamic distributions of long double-stranded RNA in Tetrahymena during nuclear development and genome rearrangements. *J Cell Sci* **129**: 1046–1058.

Wuitschick JD, Gershan JA, Lochowicz AJ, Li S, Karrer KM. 2002. A novel family of mobile genetic elements is limited to the germline genome in *Tetrahymena thermophila*. *Nucleic Acids Res* **30**: 2524–2537.

Xu J, Zhao X, Mao F, Basrur V, Ueberheide B, Chait BT, Allis CD, Taverna SD, Gao S, Wang W, et al. 2021. A Polycomb repressive complex is required for RNAi-mediated heterochromatin formation and dynamic distribution of nuclear bodies. *Nucleic Acids Res* **49**: 5407–5425.

Yang Q, Nasir AM, Coyne RS, Forney JD. 2015. Depletion of UBC9 Causes Nuclear Defects during the Vegetative and Sexual Life Cycles in Tetrahymena thermophila. *Eukaryotic Cell* **14**: 1240–1252.

Yao MC, Yao CH. 1994. Detection of circular excised DNA deletion elements in Tetrahymena thermophila during development. *Nucleic Acids Res* **22**: 5702–5708.

Figures

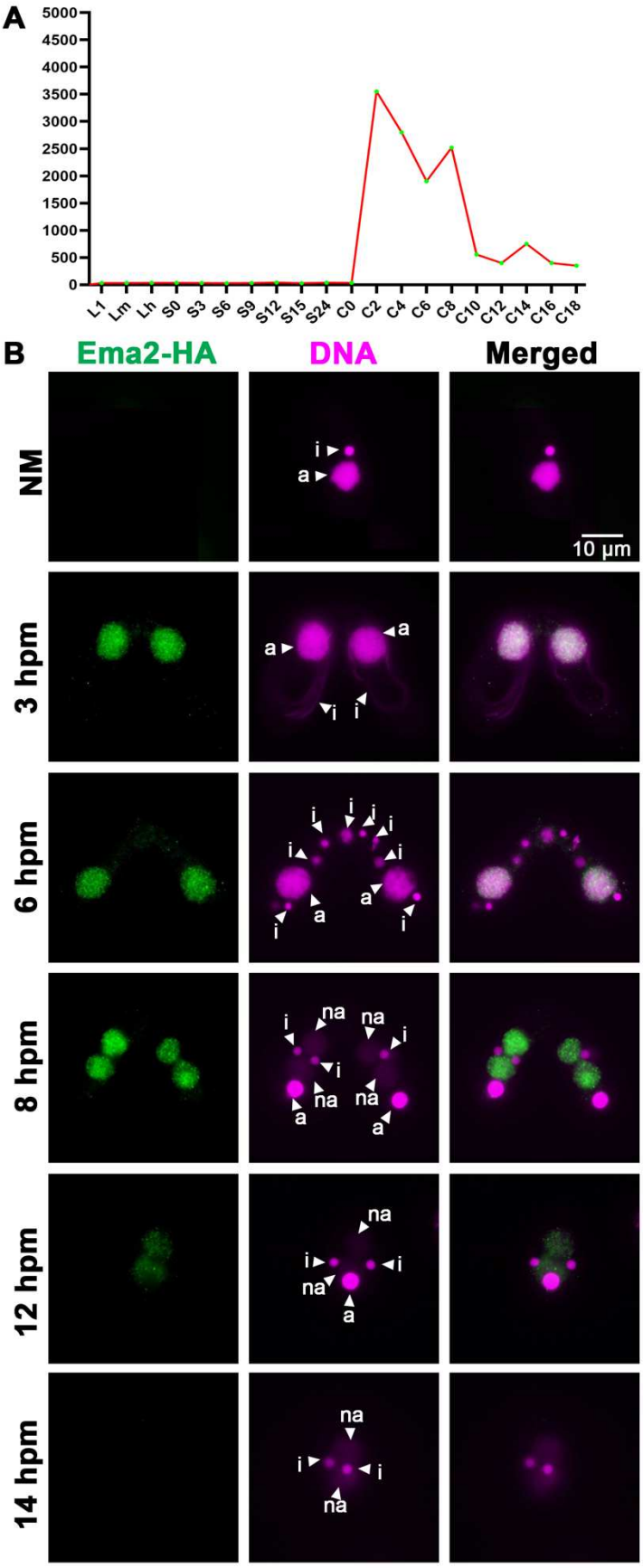
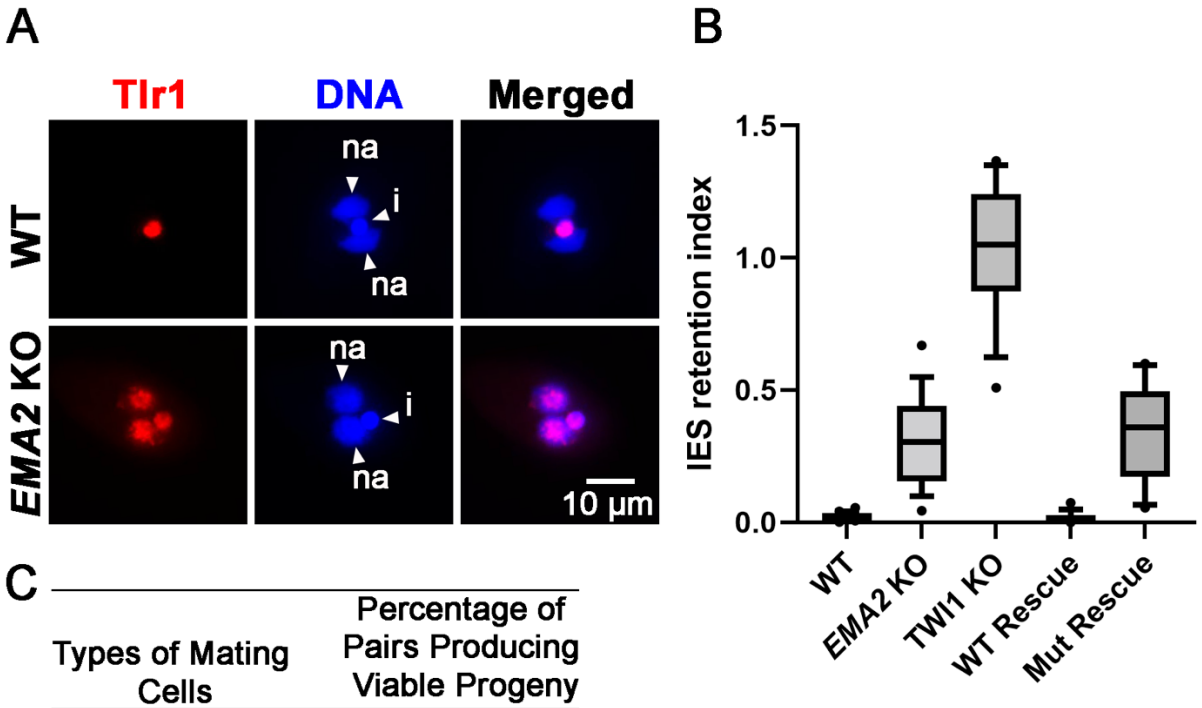


Figure 1. Ema2 expression and localization.

(A) Expression profile of *EMA2 mRNA*. On the X-axis, L1 to Lh represents growing cells, S0 to S24 represents the starvation of *Tetrahymena* cells from 0 to 24 hours. C0 to C18 represents cells undergo conjugation and post-conjugation stages from 0 to 18 hpm. On the Y-axis, the number represents the relative expression of *EMA2* mRNA. The mRNA expression data were obtained from Tetrahymena Genome Database.

(B) Ema2 localization. Two Ema2-HA strains were mixed and fixed at the timepoint indicated on the left. An anti-HA antibody were used for localizating of Ema2-HA (green) and DAPI were used to stain DNA (magenta). The MIC (i), the parental MAC (a) and the newly formed MAC (na) are marked with arrowhead. All pictures share the scalebar.



C

Types of Mating Cells	Percentage of Pairs Producing Viable Progeny
WT	76.04 (n=96)
EMA2 KO Cross 1	0 (n=96)
EMA2 KO Cross 2	0 (n=96)

Figure 2. DNA elimination and progeny viability in *EMA2 KO* cells.

(A) DNA elimination in *EMA2* KO cells. Two wild-type (WT) or two *EMA2 KO* cell lines were mixed and fixed at 36 hpm. Exconjugants were hybridized with the florescent probes complementary to the Tlr1 element (red), DNA were stained with DAPI (blue). The MIC (i) and the newly formed MAC (na) are marked with arrowhead. All pictures share the scalebar.

(B) Comparison of DNA elimination defects in different mutants. Exconjugants from wild-type (WT) cells, *EMA2* KO cells, *TWI1* KO cells, *EMA2* KO cells rescued with the wild-type *EMA2* (WT Rescue) and *EMA2* KO cells rescued with the mutant *EMA2* (Mut Rescue) were stained as in (A) and the IES retention index were calculated (see Materials and Method for details).

(C) Progeny viability of *EMA2* KO cells. Two wild-type (WT) cells and two independent crosses of *EMA2 KO* cells (Cross 1 and 2) were mated, and the conjugating pairs were isolated for the viability test. Percentage of pairs that gave rise to viable sexual progeny are shown. “n” represents the number of total pairs tested.

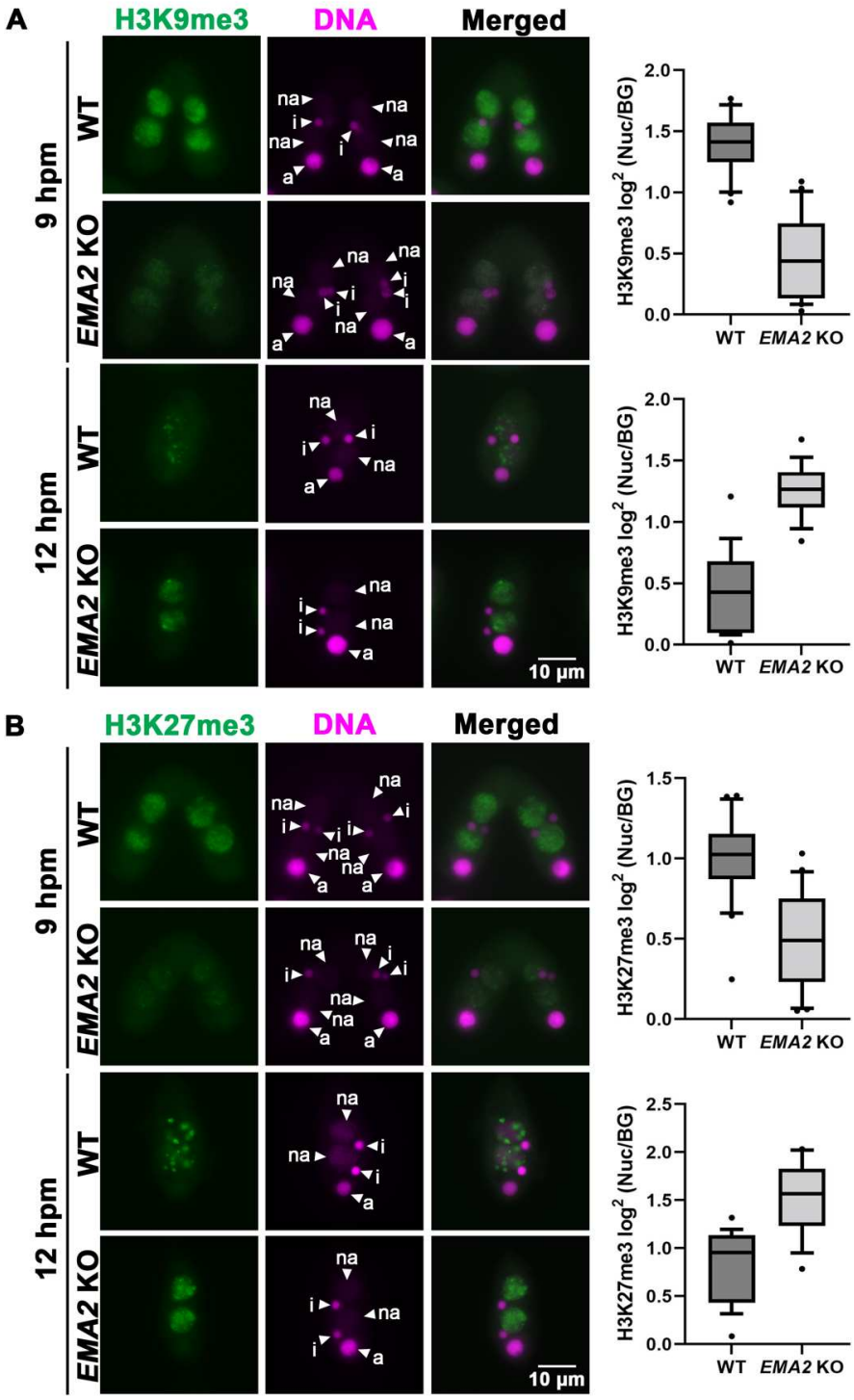


Figure 3. Heterochromatin formation in the new MACs of *EMA2* KO cells.

(Left) Two wild-type (WT) and two *EMA2* KO cells were mated, fixed at 9 and 12 hpm and stained with an anti-H3K9me3 antibody (A, green) or an anti-H3K27me3 antibody (B, green). DNA were stained with DAPI (magenta). (Right) The log2 intensity of H3K9me3 and H3K27me3 staining in the whole new MACs (Nuc) were quantified and subtracted by background cytoplasmic staining (BG). The MIC (i), the parental MAC (a) and the newly formed MAC (na) are marked with arrowhead. All pictures share the scalebar.

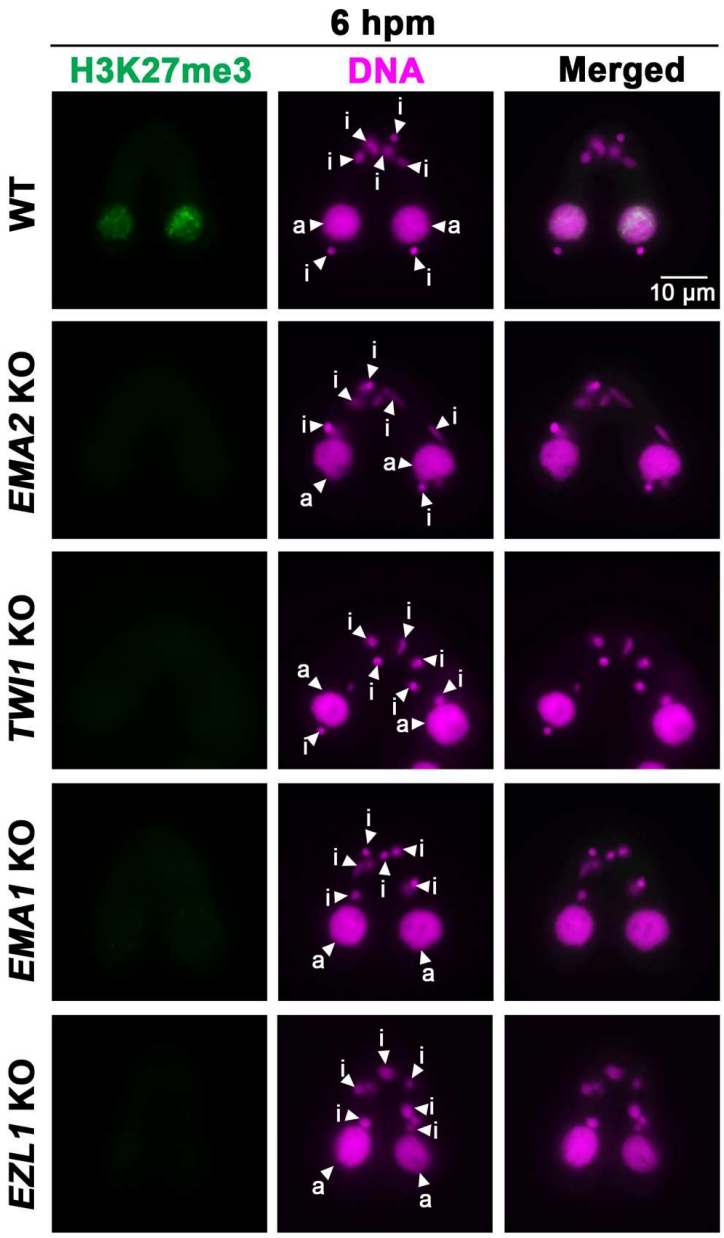


Figure 4. Heterochromatin formation in the parental MAC of *EMA2* KO cells.

Two wild-type (WT), two *EMA2* KO, two *TWI1* KO, two *EMA1* KO and two *EZL1* KO strains were mated, fixed at 6 hpm and stained with an anti-H3K27me3 antibody (green). DNA were stained with DAPI (magenta). The MIC (i) and the parental MAC (a) are marked with arrowhead. All pictures share the scalebar.

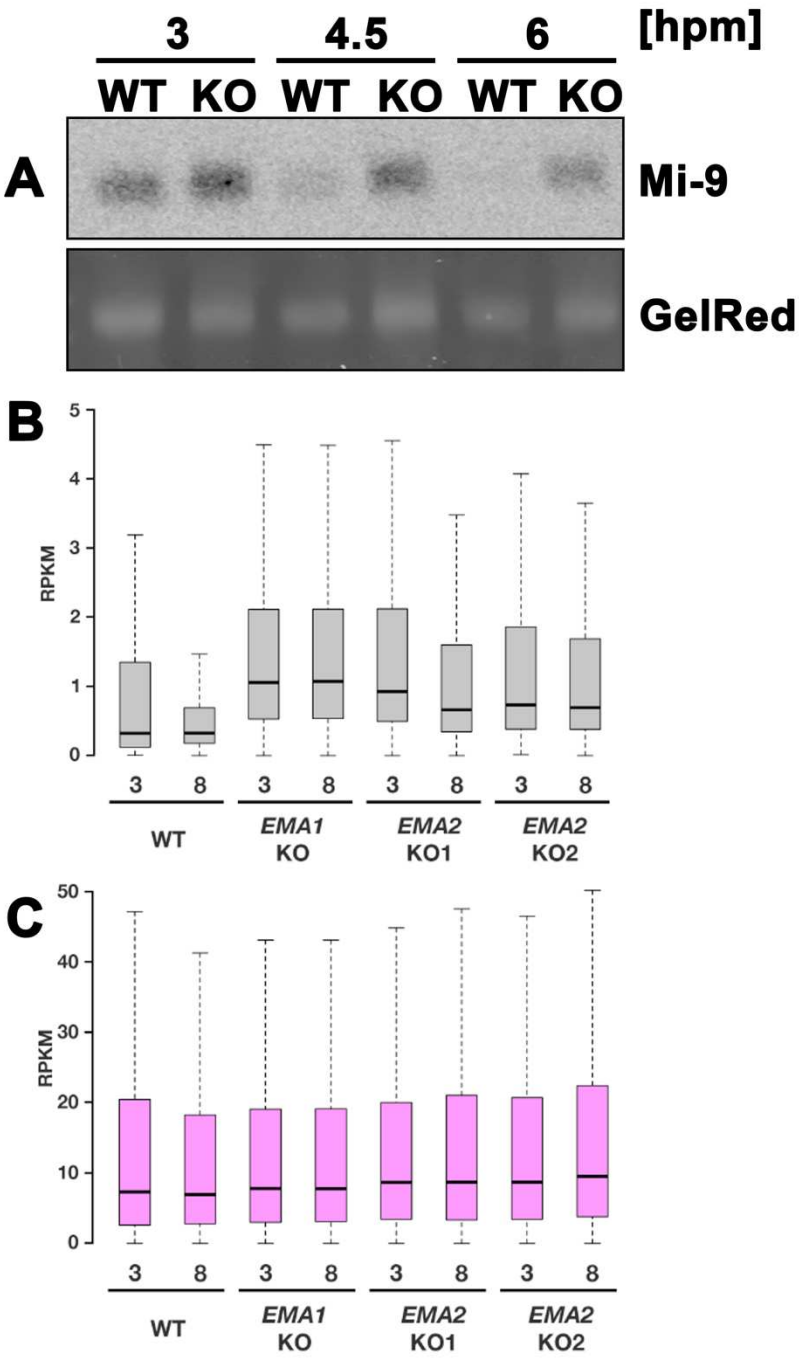


Figure 5. TDSD in *EMA2* KO cells.

(A) Total RNA was isolated from conjugating wild-type (WT) and EMA2 KO (KO) cells at 3, 4.5 and 6 hpm, separated in denaturing gel and stained with the nucleic acid dye Gel-Red (bottom). Then RNA was transferred to a membrane and hybridized with the radioactive labeled Mi-9 probe, which is complementary to a repetitive sequence in MDS (top).

(B-C) Small RNAs from conjugation wild-type (WT), *EMA1* KO, and two independent crosses of *EMA2* KO strains were isolated at 3 and 8 hpm, and analyzed ny high-throughput sequencing. Normalized numbers (RPKM [read per kilobase of unique sequence per million]) of sequenced small RNAs (26-32 nt) that uniquely matched to the MDS (B) or IES (C) genomic tiles (see Materials and Method) are shown as box plots. The median value is represented by horizontal bar in the box. The minimum and maximum value are indicated by the bar on top and bottom of the box, respectively, with 1.5x IQR.

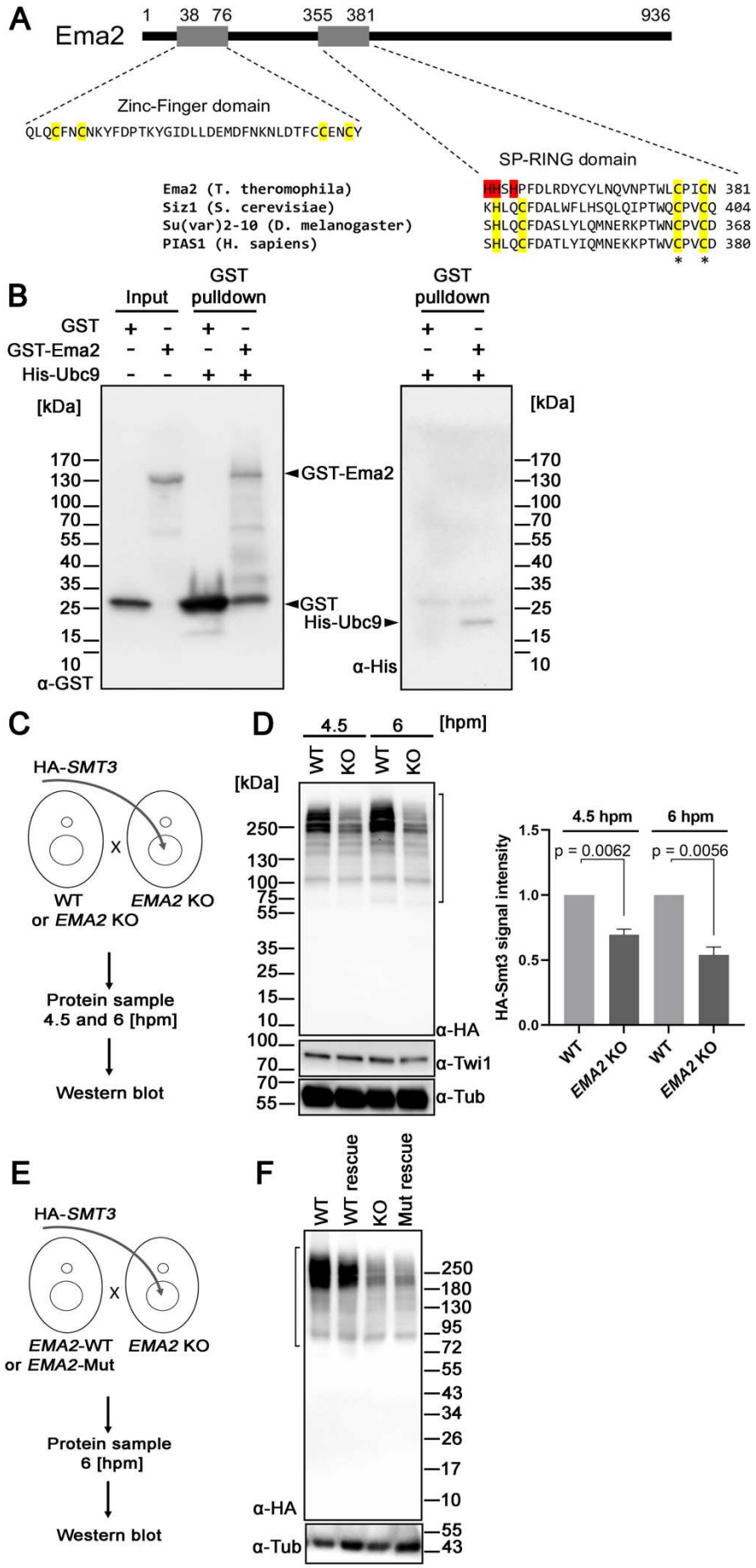

A
Ema2
1
38
76
355
381
936
Zinc-Finger domain
QLQCFNCNKYFDPTKYGIDLLDEMDFNKNLDTFCCENCY
SP-RING domain
Ema2 (T. theromophila) HHSHPFDLRDYCYLNQVNPTWLCPICN 381
Siz1 (S. cerevisiae) KHLQCFDALWFLHSQLQIPTWQCPVCQ 404
Su(var)2-10 (D. melanogaster) SHLQCFDASLYLQMNERKPTWNCPVCD 368
PIAS1 (H. sapiens) SHLQCFDATLYIQMNEKKPTWVCPVCD 380
B
Input
GST pulldown
GST pulldown
GST
GST-Ema2
His-Ubc9
[kDa]
170
130
100
70
55
40
35
25
15
10
α-GST
GST-Ema2
GST
His-Ubc9
α-His
C
HA-SMT3
WT
or EMA2 KO
EMA2 KO
Protein sample
4.5 and 6 [hpm]
Western blot
D
4.5
6
[hpm]
WT
KO
WT
KO
[kDa]
250
130
100
75
55
35
25
15
10
100
70
70
55
α-HA
α-Twi1
α-Tub
4.5 hpm
6 hpm
p = 0.0062
p = 0.0056
HA-Smt3 signal intensity
1.5
1.0
0.5
0.0
WT
EMA2 KO
WT
EMA2 KO
E
HA-SMT3
EMA2-WT
or EMA2-Mut
EMA2 KO
Protein sample
6 [hpm]
Western blot
F
WT
WT rescue
KO
Mut rescue
250
180
130
95
72
55
43
34
26
17
10
55
43
α-HA
α-Tub

Figure 6. Ema2 acts as a SUMO E3 ligase.

(A) A diagram of the Ema2 has two detectable domains, the zinc-finger domain (38-76 aa) and the SP-RING domain (355-381 aa). The SP-RING domain is compared with that of Siz1 in *S. cerevisiae*, Su(var)2-10 in *D. melanogaster* and PIAS1 in *H. sapiens*. The conserved cysteine and histidine residues are highlighted with yellow. The histidine residues of Ema2 that potentially acts as parts of SP-RING are highlighted with red (see also Supplementary Figure 1) The cystine residues of the SP-RING domain that were replaced to alanines in the Ema2 mutant are marked with asterisks.

(B) Ema2 interacts with the SUMO E2 conjugating enzyme Ubc9. GST alone (GST), GST-tagged Ema2 (GST-Ema2) and His-tagged Ubc9 that was also (His-Ubc9) were recombinantly expressed in and purified from *E. coli*. GST and GST-Ema2 were immobilized with glutathione beads and the beads were incubated with His-Ubc9. The proteins retained on the beads were eluted with SDS-sample buffer and the input and the eluted proteins were analyzed by western blot using an anti-GST (left) and an anti-His (right) antibody.

(C) Experimental design for comparing SUMOylated protein in wild-type and *EMA2* KO cells. A construct for expression of HA-tagged Smt3 (HA-*SMT3* construct) were introduced into one of the *EMA2* KO strains and crossed with wild-type (WT cross) or one of the other *EMA2 KO* (*EMA2* KO cross) strains. *EMA2* mRNA/Ema2 protein expressed in the wild-type partner of WT cross are expected to move into the *EMA2* KO partner cells and restores the *EMA2* KO phenotypes. We also expect that HA-*SMT3* mRNA/HA-Smt3 protein expressed in the *EMA2* KO partner contribute to SUMOylation in both sides of the pairs in the both crosses. Total protein was extracted from each cross at 4.5 and 6 hpm and analyzed by western blot using an anti-HA antibody to detect SUMOylated proteins.

(D) Ema2 is required for the accumulation of SUMOylated proteins *in vivo*. (top) The proteins from the WT cross (WT) or the *EMA2* KO cross (*EMA2* KO) were analyzed by western blot using an anti-HA antibody (top), an anti-Twi1 antibody (middle, to monitor mating efficiency) and an anti-alpha tubulin (bottom, loading control). The position of SUMOylated proteins is marked with a bracket. (bottom) The intensities of the HA-Smt3 signal in three independent experiments were quantified and shown as bar graphs.

(E) Experimental design for testing the activity of the SP-RING mutant of Ema2. A construct expressing wild-type (*EMA2*-WT) or C377A/C380 mutant (*EMA2*-Mut) were introduced into one of the *EMA2* KO strains and crossed with one of the other *EMA2 KO* strains that had the HA-SMT3 expressing construct (WT rescue and Mut rescue, respectively). Total protein was extracted from each cross at 6 hpm and analyzed by western blot using an anti-HA antibody.

(F) The intact SP-RING domain of Ema2 is required for the accumulation of SUMOylated proteins. The proteins from the WT rescue or the Mut rescue were compared with the WT cross and *EMA2* KO cross (shown in C and D) by western blot using an anti-HA antibody.

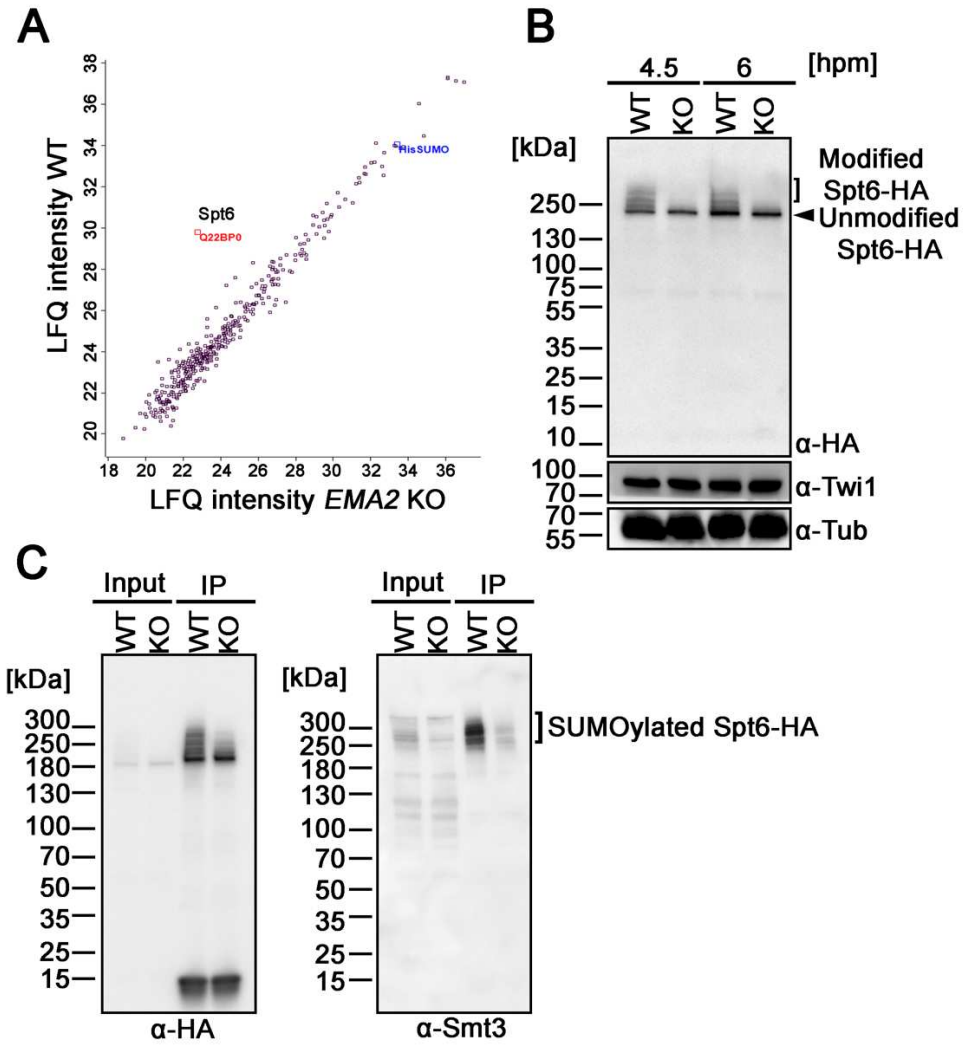


Figure 7. Ema2 promotes SUMOylation of Spt6.

(A) A construct expressing His-tagged Smt3 (His-Smt3) was introduced into a wild-type and one of the *EMA2* KO strains and crossed with another wild-type (WT cross) or one of the other *EMA2* KO strains (*EMA2* KO cross), respectively. Proteins were harvested at 6 hpm, His-Smt3-conjugated proteins were purified with Ni-NTA beads from conjugating and analyzed by mass spectrometry. Label-free quantification (LFQ) intensites of each identified protein between WT cross (WT) and *EMA2* KO cross (*EMA2* KO) were compared.

(B) A construct expressing HA-tagged Spt6 (HA-Spt6) was introduced into one of the *EMA2* KO strains and crossed with a wild-type (WT) strain or one of the other *EMA2* KO strains (KO). Total proteins were harvested at 4.5 and 6 hpm and HA-Spt6 was detected by western blotting using an anti-HA antibody. Twi1 and anti-alpha tubulin were also analyzed to monitor mating efficiency and loading, respectively.

(C) Total proteins were harvested from the WT and the KO cross shown in (B) at 6 hpm (Input) and HA-Spt6 was immunoprecipitated using an anti-HA antibody (IP). The purified proteins were analyzed by western blotting using an anti-HA (left) or an anti-Smt3 (right) antibody.

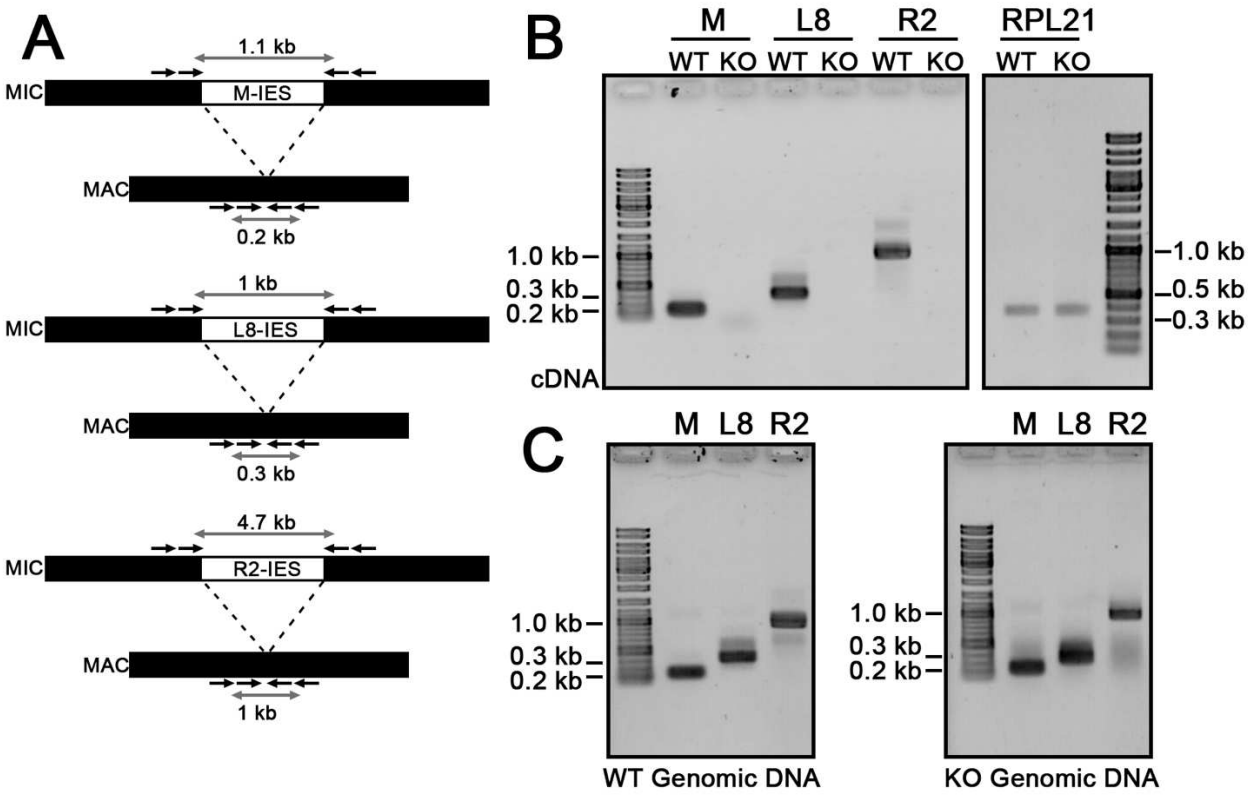


Figure 8. Ema2 is required for the accumulation of lncRNA transcripts from the parental MAC.

(A) Schematic representation of the loci used for the RT-PCR assay. The solid black box represents MDS while the open box represents M-, L8- or R2-IES. The arrows represent the primers used for nested PCR. Because the each PCR amplicon in the MIC contains an IES, transcripts from the MIC and from the MAC can be distinguished by the sizes of PCR products (double-sided arrows).

(B) Two wild-type (WT) and two *EMA2 KO* (KO) cells were mated and total RNA were isolated at 6 hpm. After reverse transcription, the cDNAs derived from the three loci shown in (A) were amplified by nested PCR. The products were separated in agarose gels. The sizes correspond to M-MDS (0.2 kb), L8-MDS (0.3 kb) and R2-MDS (1 kb) were marked on the left. *RPL21* mRNA was also analyzed as a positive control.

(C) Genomic DNAs isolated from the wild-type and the *EMA2 KO* cells were analyzed by PCR using the primer sets used for the second step of the nested PCR in (B).

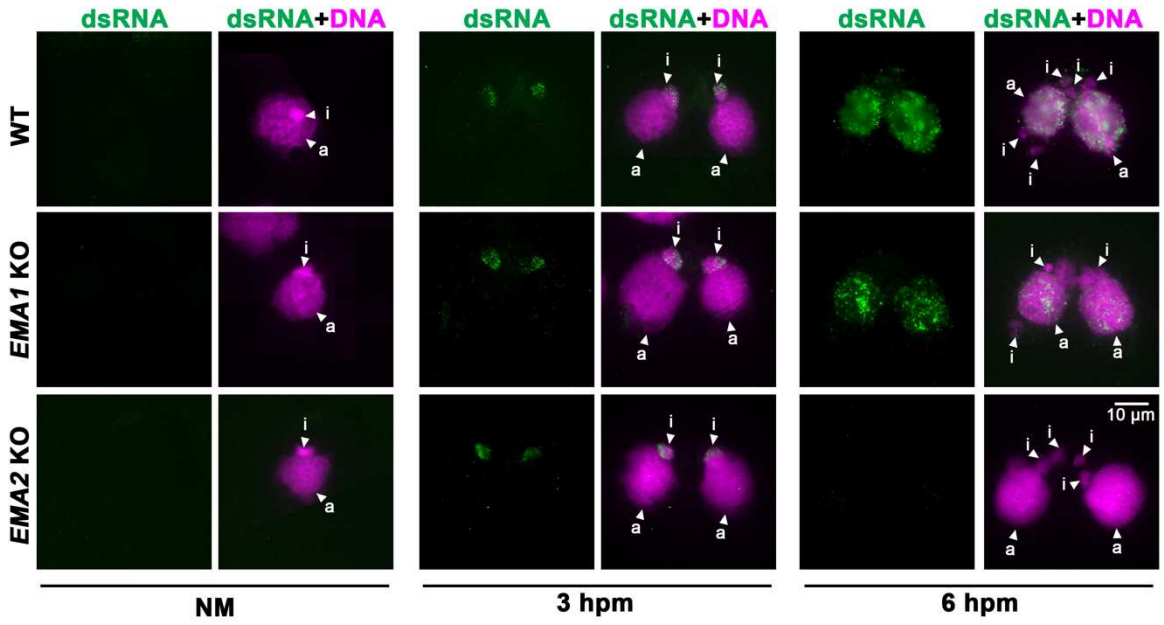
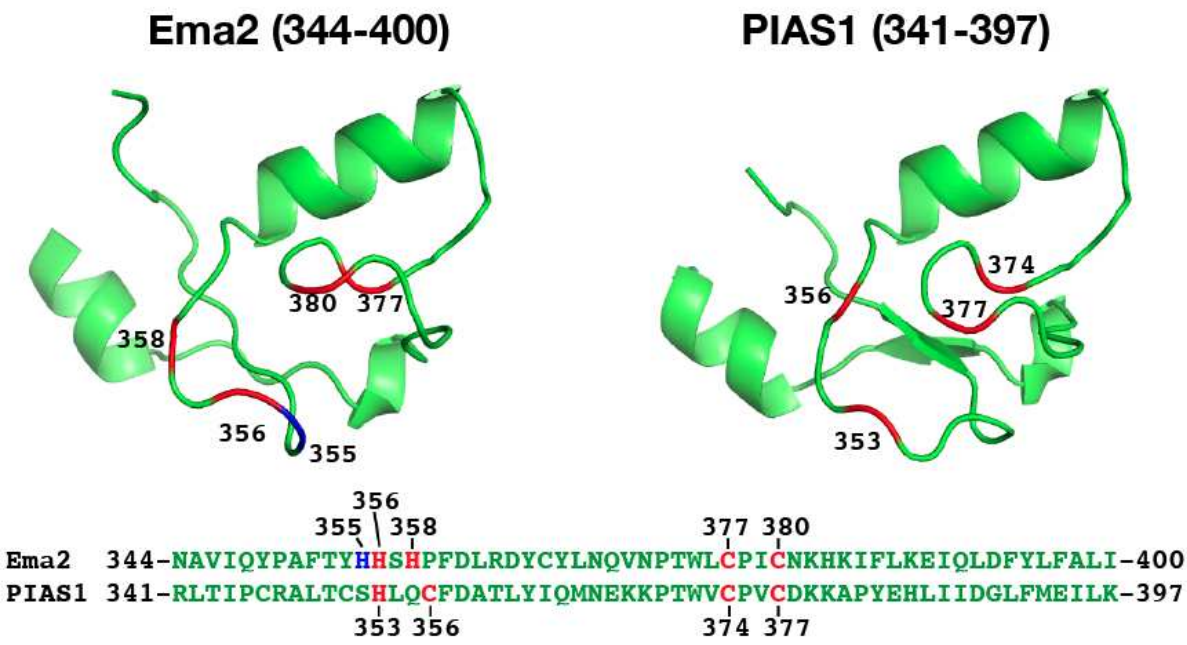


Figure 9. dsRNA localization in *EMA2* KO cells.

Vegetative (NM) or conjugating wild-type (WT), *EMA1* KO and *EMA2 KO* cells at the indicated time points were were fixed and the localization of dsRNAs was analyzed by immunostaining using the anti-long dsRNA antibody J2 (dsRNA, green). DNA was counterstained by DAPI (magenta). The MIC (i) and the parental MAC (a) are indicated with arrowhead. All pictures share the scalebar.

Supplementary figures



Supplementary Figure 1. A predicted structure of the Ema2 SP-RING domain.

The structures of the SP-RING domains of Ema2 (left) and PIAS1 (right) were predicted by AlphaFold. The conserved histidine and cysteine residues in PIAS1 and the histidine and cysteine residues of Ema2 that are closely positioned to the conserved PIAS1 residues are marked in red. The histidine residue that may be used as a metal binding site instead of the other closely located histidines is marked in blue.

Result- Methodology Paper

Preface

In the following chapter, which is published in Methods in Molecular biology, we described the production of the knockout strains for *EMA2* which is the major gene I have studied in my PhD project under the supervision of Dr. Kazufumi Mochizuki. This methodology is based on *Tetrahymena* specific knockout strategy known as coDeletion (coDel).

Tetrahymena produces small non-coding RNAs (~29-nt) known as scan RNAs(scnRNA) at two different time points during conjugation from internal eliminated sequences (IES); the Early-scnRNA are produced from Type-A IES in the MIC and Late-scnRNA production from Type-B IES is induced by early-scnRNA from Type-A IES in the new MAC, which further induced heterochromatin formation on IES to promote DNA elimination. CoDeletion can be achieved by the introduction of a desirable DNA fragment (target) into Type-A IES, and transfer this chimeric construct to a new MAC during conjugation, which will induce ectopic DNA elimination of the target by producing Late-scnRNA from the chimeric construct.

In this work, we describe the detailed procedures of coDel from the isolation of genomic DNA, construction of a targeting plasmid, introduction of the construct into *Tetrahymena*, to the selection of knockout strains. We show the procedure of the production of *EMA2* KO strains as an example of coDel experiment. We also examined in this study the minimum length of a targeting sequence and the complementarity between a targeting sequence and its target locus required for inducing coDel.

In the attached article, both SS and KM conceived the experiments, contributed to the interpretation of the results and wrote the manuscript.



Chapter 3

PIWI-Directed DNA Elimination for Tetrahymena Genetics

Salman Shehzada and Kazufumi Mochizuki

Abstract

Piwi-bound small RNAs induce programmed DNA elimination in the ciliated protozoan *Tetrahymena*. Using the phenomenon called codeletion, this process can be reprogrammed to induce ectopic DNA elimination at basically any given genomic location. Here, we describe the usage of codeletion for genetic studies in *Tetrahymena* and for investigations of the molecular mechanism of Piwi-directed programmed DNA elimination.

Key words PIWI, Small RNA, DNA elimination, Gene knockout, *Tetrahymena*

1 Introduction

Argonaute proteins are a highly conserved family found in species from humans to archaea and comprise the central protein component of the small RNA-mediated regulatory machinery. Argonaute proteins in eukaryotes can be divided into two subfamilies: the Ago and Piwi subfamilies. While the latter is found predominantly in animals, ciliated protozoans also express Piwi subfamily proteins [1–6]. Therefore, ciliates serve as unicellular models to study the molecular actions of Piwi proteins.

Ciliates are a group of eukaryotes characterized by nuclear dimorphism, in which each cell has a polyploid somatic macronucleus (MAC) and a diploid germline micronucleus (MIC). The former controls metabolism, and the latter functions in reproduction. At each sexual cycle in ciliates, the MIC produces new MICs and new MACs, and the parental MAC is degraded. During the development of the new MAC of *Tetrahymena thermophila*, a model ciliate, ~12,000 DNA segments called internal eliminated sequences (IESs) are reproducibly eliminated [7]. IESs are recognized by 26- to 32-nt small RNAs called scnRNAs [8], which are loaded into the Piwi protein Twi1p [9]. Unlike Piwi-loaded small

Nicholas F. Parrish and Yuka W. Iwasaki (eds.), *piRNA: Methods and Protocols*,
Methods in Molecular Biology, vol. 2509, https://doi.org/10.1007/978-1-0716-2380-0_3,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2022

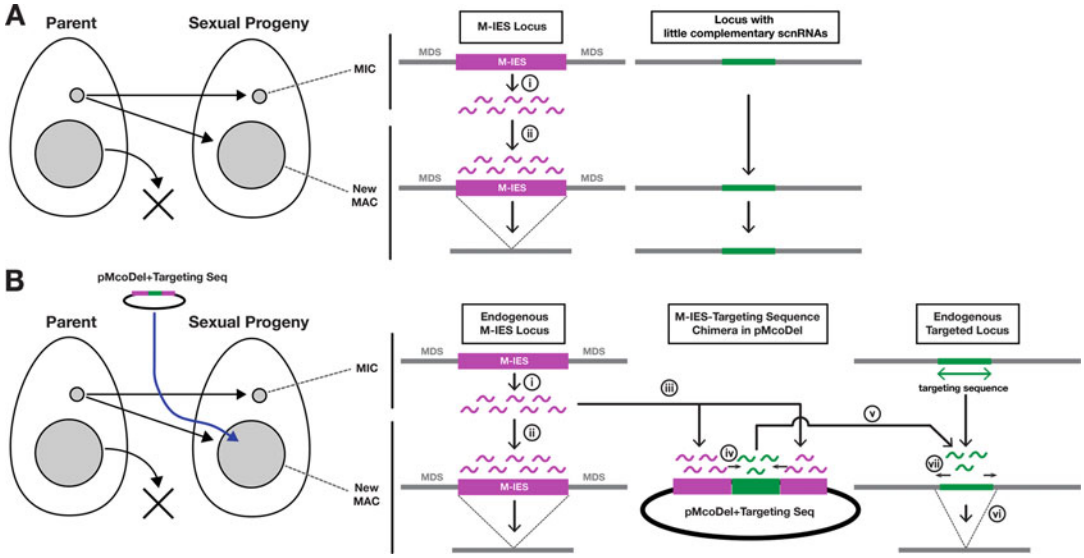


Fig. 1 The proposed actions of scnRNAs in coDel. (**a**) DNA elimination without induction of coDel. Early scnRNAs (magenta wavy lines) produced from Type-A IESs (represented by M-IES in this figure) in the MIC (i) interact with Type-A IESs in the new MAC and induce DNA elimination (ii). Non-IES loci (MAC-destined sequences [MDSs]) either do not produce early scnRNAs or homology-dependent small RNA degradation eliminates early scnRNAs derived from them [8, 12, 20, 26]; thus, DNA elimination is not induced at these loci. (**b**) Induction of coDel. When an IES (magenta box)-targeting sequence (green box) chimeric construct (pMcoDel+Targeting Seq) is introduced into the new MAC, early scnRNAs produced from M-IES in the MIC (i) interact not only with the endogenous M-IES (ii) but also with the M-IES on the introduced chimeric construct (iii). The latter interaction triggers late scnRNA (green wavy lines) production from the adjacent target sequence in cis (iv), which then interacts with the endogenous target locus in trans (v) and induces ectopic DNA elimination (vi). Late scnRNAs likely further induce the production of late scnRNAs in cis at the targeted locus (vii), making deletions longer than the targeted sequence

RNAs in animals, scnRNAs are processed from long double-stranded RNA precursors by a Dicer homolog [10, 11].

IESs in *Tetrahymena* can be classified into two types: Type A IESs, which are recognized in cis by the primary scnRNAs (early scnRNAs) that they produce, and Type B IESs, which are recognized in trans by early scnRNAs produced by Type A IESs [12]. Because secondary scnRNA (late scnRNA) biogenesis occurs concomitantly with the spreading of heterochromatin on IESs [12], the introduction of a chimeric construct of a Type A IES and a fragment of a non-IES locus (targeting sequence) into the new MAC results in the production of secondary scnRNAs from the targeting sequence that induce ectopic DNA elimination at the non-IES locus in trans (Fig. 1). This phenomenon is called codeletion (coDel).

Because coDel can be induced by the simple introduction of an IES-targeting DNA chimeric construct, this approach is suitable as a gene knockout technology in *Tetrahymena* [13–19]. Moreover,

the effects of an array of different targeting sequences can be easily examined [13], and coDel can also be used to analyze how scnRNAs are involved in the transgenerational inheritance of the pattern of DNA elimination [20] and the base-pairing requirement between small RNAs and genomic DNA to induce the downstream effect. In this chapter, we describe the usage of coDel to produce gene knockout strains in *Tetrahymena* and to investigate the length of the seed sequence of scnRNAs in DNA elimination.

2 Materials

2.1 Extraction of Tetrahymena Genomic DNA

1. *Tetrahymena thermophila* strain B2086 (or other wild-type strain of *Tetrahymena thermophila*) (*see* **Note 1**).
2. 1× Super Proteose Peptone (SPP) medium: 2% Bacto Proteose peptone (Bectone Dickinson and Company), 0.4% D-(+)-Glucose (Sigma), 0.2% Bacto Yeast extract (Bectone Dickinson and Company), 0.003% EDTA iron(III) sodium salt (Sigma) (*see* **Note 2**).
3. CHAOS buffer (stored at 4 °C): 4 M guanidine thiocyanate, 0.5% sodium lauroyl sarcosinate, 25 mM Tris (pH 8), 0.1 M 2-mercaptoethanol.
4. Phenol extraction buffer (stored at RT): 100 mM Tris (pH 8), 10 mM EDTA (pH 8), and 1% SDS.
5. 25:24:1 (v/v/v) Phenol/chloroform/isoamyl alcohol, saturated with 10 mM Tris (pH 8) and 1 mM EDTA.
6. 2-Propanol.
7. 70% ethanol.
8. Ethanol.
9. TE (Tris–EDTA) buffer, pH 8.
10. 10 mg/mL RNase A.

2.2 Construction of coDel-Targeting Plasmid

1. pMcoDel plasmid (*see* **Note 1**).
2. FW and RV primers amplifying the ~500–1000 bp region of a targeted locus (*see* **Note 3**). To assemble a construct using the NEBuilder HiFi DNA Assembly system, the primers must have >16 nt overlapping sequence with pMcoDel arms (*see* **Note 4**). For example, to induce deletion at the *TTHERM_00113330* locus (Fig. 2a), we used EMA2-coDel-FW and EMA2-coDel-RV (Table 1). To assemble a construct using T4 DNA ligase, both primers must have the NotI recognition site (*see* **Note 5**). For example, to induce deletion at a noncoding sequence on MIC chromosome 4 with the 661 bp targeting sequence (Fig. 3a), we used C121_FW1 and C121_RV1 (Table 1).
3. Taq DNA Polymerase with Standard Taq Buffer.

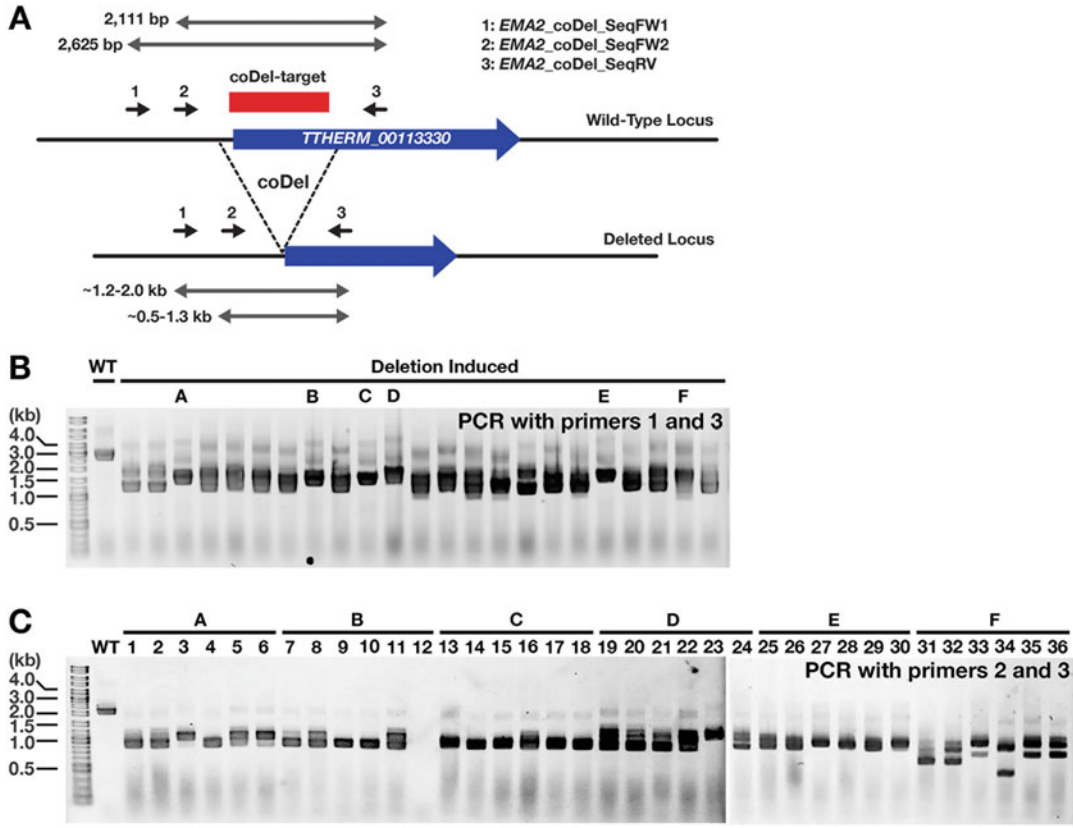


Fig. 2 Production of gene knockout strains using coDel. (**a**) A coDel was induced at the coding sequence of *TTHERM_00113330* (blue boxed arrow) in the MAC by inserting an ~1 kb sequence at the beginning of the *TTHERM_00113330* coding sequence (red box) into the pMcoDel vector. (**b**) Wild-type (WT) and 23 paromomycin-resistant transformants after introduction of the pMcoDel-*TTHERM_00113330* targeting construct (deletion induced) were examined by direct cell PCR using primers #1 and #3 shown in (**a**). The sequences of the primers are listed in Table 1. A 2625 bp product was expected to be obtained without any deletion at the targeted locus (as in WT), and all of the coDel-induced cell lines showed ~1.2–2 kb PCR products, which correspond to the locus with a deletion at the targeted region. (**c**) Six coDel lines (marked with A to F in (**b**)) were selected, six clonal cell lines each were established, and their *TTHERM_00113330* locus was analyzed by PCR using primers #2 and #3 shown in (**a**). The sequences of the primers are listed in Table 1. A 2111 bp product was expected to be obtained without any deletion at the targeted locus (as in WT), and all of the coDel-induced cell lines showed ~0.5–1.2 kb PCR products

4. 2.5 mM each dNTP mix.
5. Column-based Gel and PCR DNA Clean-up Kit.
6. *NotI* restriction enzyme.
7. NEBuilder HiFi DNA Assembly Master Mix (NEB).
8. FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific).
9. DNA Ligation Kit, Mighty Mix (Takara).

Table 1
Sequences of synthetic DNA

Primer name	Sequence (5′--3′)
EMA2_coDel_FW	ATTGTTATCATCTTATGACCGCTTTAATGTATTTGACTAGTGC
EMA2_coDel_RV	TCAAGTTGTAATGCTAAAATAAGCAAAATGAGTTCACTATTTGG
EMA2_coDel_SeqFW1	AATTGGGTACCGGGCCCCCCCTCGAGGTGGATCTGCTGGAAAAAGA TTGTATG
EMA2_coDel_SeqFW2	GCTGGCGGAGATATGCAAGATGGTC
EMA2_coDel_SeqRV	GACTAAGTCCTCTATTGGGTTTGTTC
McoDel_seqFW	ATTGAATAAGGAGACCAGCCTCTC
McoDel_seqRV	AAAACTAATAACTTGGGTCTTTAGATCAC
C121_coDel_cFW	GGTTGATGGTAGGTAGGTACCTC
C121_coDel_cRV	CAGTAGATTTCTTTGAAAGCTAACTC
C121_FW1	GCATGCGGCCGCGAGACATTTCTGGTGTGTTCGG
C121_FW2	GCATGCGGCCGCGAGAGTTAATCTTTATTAAAAGGC
C121_FW3	GCATGCGGCCGCGTATTTTAGTAACATGTATCAAG
C121_FW4	CAAAGTTGATGTTTCTGGTGGTTACAGTTGATAAAAAATTAG
C121_FW5	GCATGCGGCCGCGAGACATTTCTGGTGTGTTGG
C121_FW6	TAAGTACGTCATAATATATTGTTTTCTTGA
C121_FW7	TGACTAGTTTGTAATTATTACTAAATTTAA
C121_FW8	GCATGCGGCCGCGAGACATTTGTGGTGTGTTGG
C121_FW9	TAAGTACGTCTTAATATATTGTTTTCTTGA
C121_FW10	TGACTAGTTTGTAATTTTTACTAAATTTAA
C121_RV1	GCATGCGGCCGCGATTAATTGGCTATTCAAGCTATG
C121_RV2	GCATGCGGCCGCTTGGTGAGGAACCAAATATGTTTG
C121_RV3	GCATGCGGCCGCCTTCATGGGGAGTTAATAATTATC
C121_RV4	GCATGCGGCCGCCCCAACAATTAGATATAATTTCCC
C121_RV5	CTAATTTTTTATCAACTGTAACCACCAGAAACATCAACTTTG
C121_RV6	GCATGCGGCCGCCACCAGAAACATCAACTTTG
C121_RV7	TCAAGAAAACAATATATTATGACGTACTTA
C121_RV8	TTAAATTTAGTAATAATTACAAACTAGTCA
C121_RV9	GCATGCGGCCGCTAGGTGAGGAACCAAATATGTATG
C121_RV10	TCAAGAAAACAATATATTAAGACGTACTTA
C121_RV11	TTAAATTTAGTAAAAATTACAAACTAGTCA
C121_RV12	GCATGCGGCCGCTAGGTGAGGAAGCAAATATGTATG

(continued)

Table 1
(continued)

Primer name	Sequence (5′--3′)
C121_MUT20_1	GAGACATTTCTGGTGTGTTGGGAATTCTATTAAAACTTTCTATGTAA TCTAACTTTTATATTTTTATAAAAATTTAAAATAAAAACAAAG TTGATGTTTGTGGTGATAATTCATAAATATAAGTACGTCATAAT
C121_MUT20_2	ATATTGTTTTCTTGAATATAAGTATAAACTAATTAAGTTAAGTAAC TAAATAAGGTGAAGGATTTCGGAAATTATATCTAATTGTAGGG TAAATTATAATTTTCTACTCTTTTTTATCTTGACTAGTTTGTAAT
C121_MUT20_3	TATTACTAAATTTAATTCAATTTATTGAGAGTGTTAATCTTTA TTAAAAGGGTCAGAAAATAATAGATAATAATTAACTCCCCA TGAAGATTAGATAGAAACAAACAAACATACATATTTGGTTCC TCACCTA
C121_MUT10_1	GAGACATTTGTGGTGTGTTGGGAATTCTAATAAAACTTTCTATGTAA TCAAACTTTTATATTTTTATAATAATTTAAAATAAAAACAAAC TTGATGTTTGTGGTGATAAATCATAAATATAAGTACGTCTTAAT
C121_MUT10_2	ATATTGTTTTCTTGATTATAAGTATAAACTAATTATGTTAAGTAAC TAAATAAGGAGAAGGATTTCGGAAATTATTTCTAATTGTAGGG TAAATTTTAATTTTCTACTCTTTTTTTTCTTGACTAGTTTGTAAT
C121_MUT10_3	TTTTACTAAATTTAATTCAATATATTGAGAGTGTTAATCTTAA TTAAAAGGGTCAGAAAATTATAGATAATAATTAACTCCGCA TGAAGATTAGATAGAAAGAAACAAACATACATATTTGCTTCC TCACCTA

Underlined and double-underlined sequences are for NotI digestion and NEBuilder HiFi assembly to the arms of pMcoDel, respectively

10. NEB 5-alpha Competent *E. coli*: thaw on ice, aliquot 12.5 μL each in screw-capped 1.5-mL tubes, freeze with liquid nitrogen, and store at −80 °C.
11. Primers for colony PCR and sequencing (McoDel_seqFW and McoDel_seqRV [Table 1]).

2.3 Transformation for Inducing coDel

1. Two wild-type *Tetrahymena thermophila* strains having complementary mating types. We typically use B2086 (Mating Type II) and CU428 (Mating Type IIV) (*see* **Note 1**).
2. Biolistic PDS-1000/He Particle Delivery System (Bio-Rad) (*see* **Notes 6** and **7**).
3. 10 mM Tris, pH 7.5.
4. 1× and 2× Super Proteose Peptone (SPP) medium (*see* **Note 2**).
5. 100 mg/mL paromomycin sulfate.
6. 100× Antibiotic/Antimycotic (AB/AM) mix (Millipore).
7. 96-well cell culture plates.
8. 30 °C Incubator.

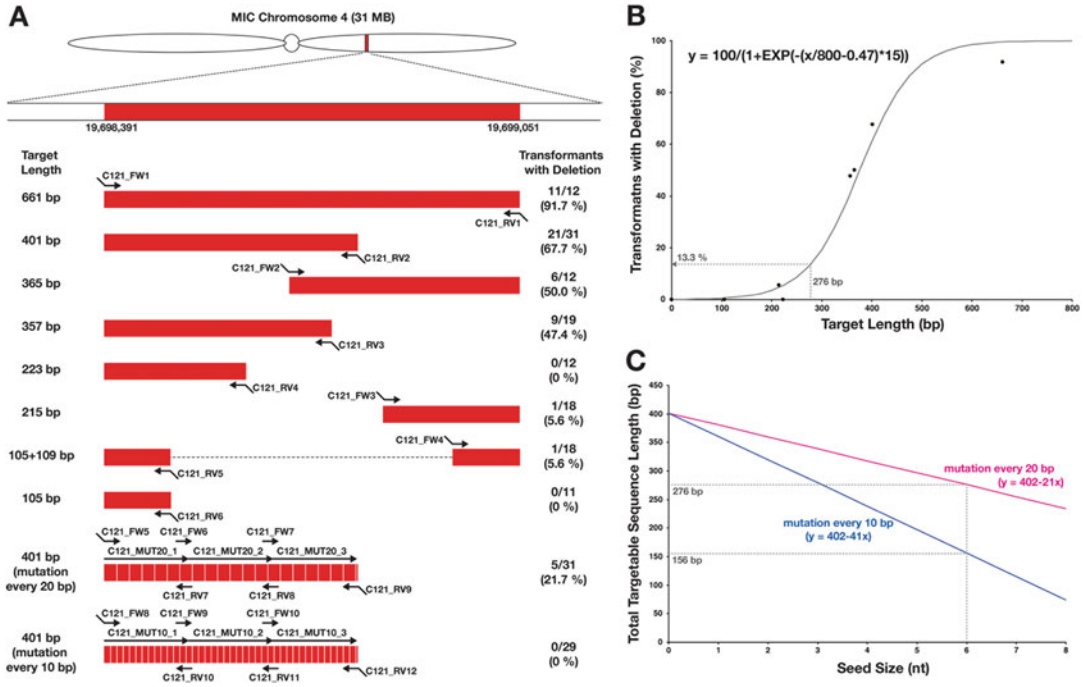


Fig. 3 Examination of the requirement of scnRNA-target interaction in DNA elimination using coDel. (**a**) Various targeting sequences that are designed to target a non-IES region at the indicated position of MIC chromosome 4 (red boxes) were introduced into the new MAC of progeny of B2086 and CU428, and the number of transformants that showed deletion at the targeted locus was counted. Arrows indicate the primers and oligo DNAs (their sequences are listed in Table 1) that were used to amplify and construct the targeting sequences. (**b**) Length of the different targeting sequences shown in (**a**) (except the last two having base replacements) and fraction of transformants that showed deletion by those targeting sequences are plotted. A sigmoidal curve that is close to the plotted experimental data is shown. Dotted lines indicate the expected fraction of deletion-positive transformants by a 276 bp targeting sequence, which corresponds to the total targetable sequence lengths if scnRNAs from the 401 bp targeting sequence having base replacement every 20 nt use 6 nt as seed (see below). (**c**) Total targetable sequence lengths within the 401 bp targeted sequence by the scnRNAs derived from the 401 bp targeting sequence that have base replacement in every 20 (magenta line) or 10 (blue line) bp were predicted (*y*-axis) for different seed sizes to initiate recognition of the targeted sequences by scnRNAs (*x*-axis). Dotted lines indicate length of predicted total targetable sequence when scnRNA uses 6 nt as seed. *See* also **Note 20** for interpretation of the results

2.4 Detecting coDel by Direct PCR

1. Primers amplifying the targeted region and its ~1 kb flanking sequences. To examine the *TTHERM_00113330* locus, we used either EMA2_coDel_SeqFW1 and EMA2_coDel_SeqRV (primers #1 and #3 in Fig. 2a) or EMA2_coDel_SeqFW2 and EMA2_coDel_SeqRV (primers #2 and #3 in Fig. 2a).
2. GoTaq Long PCR Master Mix (Promega).

2.5 Establishing Gene Knockout Cell Lines from Cells with coDel

1. DropMaker (made-to-order, *see* [21] for more explanations).
2. Bacteriological petri dish.
3. Glass capillary.
4. Aspirator tube assemblies for microcapillary pipettes.
5. Glass capillary puller.
6. Dissection Microscope.
7. 24-well and 96-well cell culture plates.
8. 13.3% DMSO in 10 mM Tris, pH 7.5 (*see* **Note 8**).
9. Cryo 1 °C "Mr. Frosty" Freezing Container (Nalgene).

2.6 Using coDel to Study DNA Elimination Mechanism

1. Primers to amplify different regions of target sequences (Fig. 3a) (Table 1).
2. Synthetic double-stranded DNA with base replacements at defined intervals (Integrated DNA Technologies) (*see* **Note 9**).
3. Synthetic single-stranded DNA corresponding to a part of the target sequence and having base replacements at defined intervals (such as Ultramer DNA oligos from Integrated DNA Technologies) (Table 1).
4. Primers amplifying the targeted region and its ~1 kb flanking sequences (Table 1).

3 Methods

3.1 Extraction of Tetrahymena Genomic DNA

1. Culture the B2086 strain of *Tetrahymena* cells in 20 mL 1× SPP at 30 °C overnight (*see* **Note 10**).
2. Place approximately 1 mL of the culture (~1×10^6 cells) in a 1.5-mL Eppendorf tube, centrifuge for ~5 s in a mini table top centrifuge, and quickly remove the supernatant using a pipette.
3. Add 300 μL CHAOS buffer and lyse cells by pipetting (*see* **Note 11**).
4. Add 300 μL of Phenol Extraction Buffer and mix vigorously by inverting the tube (*see* **Note 12**).
5. Add 600 μL of phenol/chloroform/isoamyl alcohol and mix vigorously by inverting the tube.
6. Spin down for 10 min at 14,000 rpm (or maximum speed) at RT and transfer 600 μL of the upper phase to a fresh 1.5-mL tube.
7. Add 600 μL of 2-propanol and mix well by inverting the tube.
8. Spin down for 10 min at 14,000 rpm at 4 °C and discard supernatant.

9. Add 700 μL ice-cold 70% ethanol, leave for 5 min at RT and spin down for 5 min at 14,000 rpm at 4 °C.
10. Discard the supernatant as much as possible and air dry for 5–10 min.
11. Add 200 μL TE buffer and incubate at RT for ~15 min (vortex occasionally) (*see* **Note 13**).
12. Add 1 μL of RNaseA and incubate at 37 °C for 30 min.
13. Add 200 μL TE buffer and 400 μL phenol/chloroform/isoamyl alcohol, mix by inverting tube several times, spin down for 10 min at 14,000 rpm at RT.
14. Transfer the upper phase to a fresh 1.5-mL tube, add 5% vol. of 5 M NaCl, and mix well.
15. Add 2× vol. of ethanol, mix well and incubate the tube at −20 °C for more than 1 h.
16. Spin down for 10 min at 14,000 rpm at 4 °C and discard the supernatant.
17. Add 700 μL ice-cold 70% ethanol and spin down for 5 min at 14,000 rpm at 4 °C.
18. Remove supernatant as much as possible and air dry for 5–10 min.
19. Add 100 μL TE buffer and incubate for ~15 min at RT (vortex occasionally).
20. Measure the DNA concentration, adjust to 0.5 μg/μL by adding TE and store at −20 °C.

3.2 Construction of coDel-Targeting Plasmid

1. Set up a 50 μL PCR containing 1× PCR buffer, 100 μM each dNTP, 0.2 μM FW primer, 0.2 μM RV primer, 10 ng Tetrahymena genomic DNA (Subheading 3.1, **step 20**), and 1.25 Unit Taq polymerase.
2. Perform PCR with the following conditions: 95 °C for 2 min, 35 cycles of 95 °C for 20 s/55 °C for 30 s/68 °C for 60 s, 68 °C for 3 min.
3. Check the product by running 3 μL of the reaction in a 1% agarose gel.
4. Purify the PCR product using NucleoSpin Gel and PCR Clean-up Kit (or similar glass filter base purification system). If necessary, run the PCR product in a 1% agarose gel and excise the correct PCR product.
5. Digest ~1–2 μg of pMcoDel plasmid DNA with NotI-HF and 1×CutSmart buffer in 50 μL overnight. If the NEBuilder HiFi DNA Assembly system will be used, go to **step 6**. When T4 DNA ligase will be used, go to **step 9**.

6. Extract NotI-digested pMcoDel with NucleoSpin Gel and PCR Clean-up Kit and elute with 30 μL elution buffer.
7. Mix 1 μL of NotI-digested pMcoDel (**step 6**), 1 μL PCR-amplified targeting sequence (**step 4**) and 2 μL of 2× NEBuilder HiFi DNA Assembly Mix.
8. Incubate at 50 °C for 30 min and go to **step 14**.
9. Add 1 μL of FastAP, incubate at 37 °C for 1 h and then inactivate the enzyme for 15 min at 65 °C.
10. Extract DNA with a DNA Clean-up Kit and elute with 30 μL elution buffer.
11. Digest the PCR product (**step 4**) with NotI-HF as in **step 5**; follow by purification using a DNA Clean-up Kit.
12. Mix 1 μL of NotI-digested and dephosphorylated pMcoDel (**step 10**), 1 μL NotI-digested PCR product (**step 10**), and 2 μL of 2× ligation mix.
13. Incubate at 16 °C for a few hours to overnight.
14. Transform NEB5alpha competent cell (12. 5 μL) with a 0.5 μL reaction (either from **steps 8** or **13**) and inoculate on an LB-amp plate.
15. Check the resulting plasmid construct by colony PCR using McoDel_seqFW and McoDel_seqRV.
16. Purify the plasmid construct with the correct insert size by mini-prep, and check the sequence of the insert by Sanger sequencing using McoDel_seqFW.

3.3 Transformation for Inducing coDel

1. Mix 25 mL each 5×10^5/mL prestarved B2086 and CU428 in 10 mM Tris pH 7.5 (*see* **Note 10**).
2. Introduce 5 μg of pMcoDel-Targeting Sequence plasmid into 6×10^6 cells at 7 hpm using BioRad Biolistic Particle Gun system (*see* **Note 6**).
3. Release cells into 50 mL 10 mM Tris–HCl (pH 7.5) and incubate for ~12 h at 30 °C with slow (60 rpm) rotation.
4. Add 50 mL of 2×SPP and 500 μL of 100× AB/AM mix.
5. Incubate at 30 °C for an additional ~4 h with slow (90 rpm) rotation.
6. Add 100 μL of 100 mg/mL paromomycin (final conc. 100 μg/mL).
7. Make two diluted cultures: (a) 1.5 mL of the culture + 13.5 mL of 1× SPP + 135 μL 100× AB/AM mix + 13.5 μL of 100 mg/mL paromomycin; (b) 0.3 mL of the culture + 14.7 mL of 1× SPP + 100× AB/AM mix + 14.7 μL of 100 mg/mL paromomycin.
8. Aliquot cells into 96-well plates.

Original culture: 150 μL/well, 7× 96-well plates.

1/10 diluted culture (**7a**): 150 μL/well, 1× 96-well plates.

1/50 diluted culture (**7b**): 150 μL/well, 1× 96-well plates.

9. Place the plates in a moist chamber (plastic box with wet tissue papers) and incubate for 3–4 days at 30 °C (*see* **Note 14**).
10. Select 24 wells containing paromomycin resistance cells from 96-well plate(s) in which 2/3 or less of the wells had growing cells (*see* **Note 15**).

3.4 Detecting coDel by Direct PCR

1. Mix the following in a PCR tube and keep on PCR cooler (or on ice): 10 μL GoTaq Long PCR Master Mix (2×), 0.5 μL 10 μL FW primer, 0.5 μL 10 μL RV primer, 8.5 μL water.
2. Add 0.5 μL culture (Subheading 3.2, **step 8**) to the PCR tube above (Subheading 3.3, **step 1**).
3. Perform PCR under the following conditions: 94 °C for 2 min; 35 cycles of 94 °C for 20 s, 50–55 °C for 30 s, and 67 °C for 1 min per kb; then 67 °C for 5 min.
4. Analyze 3–5 μL of reaction by agarose gel electrophoresis (Fig. 2b as an example).

3.5 Establishing Gene Knockout Cell Lines from Cells with coDel

1. Select 6–8 cell lines in which deleted copies of the target locus are dominant over the wild-type copy in Subheading 3.4, **step 4** (*see* **Note 16**).
2. Prepare 48 drops of 1× SPP on a 10 cm culture dish (1 plate for 2 lines) using a DropMaker. Alternatively, drops (~50 μL each) can be made using a pipet.
3. Pick up 24 single cells each from the corresponding culture (Subheading 3.2, **step 10**) using a microcapillary pipette under a dissection microscope and separately place into the drops above.
4. Incubate for 3 days at 30 °C in a moist chamber.
5. Choose 6–8 drops containing well-growing cells per line and transfer them individually into 1 mL fresh 1× SPP in a 24-well plate.
6. Incubate for 2 days at 30 °C.
7. Perform direct PCR as in Subheading 3.4 to find clones showing only deleted copies of the target locus (Fig. 2c as an example) (*see* **Notes 17** and **18**).
8. (If sexually mature cells are needed for the experiment) Every 2 days, inoculate 5 μL of the selected clones into 1 mL fresh 1× SPP and incubate at 30 °C (or keep cells at room temperature for a week) until passage 10. Then, perform test mating to identify combinations of clones with complementary mating types.

9. Prepare frozen cell stocks using DMSO and Cryo 1 °C Freezing Container. Refer to [21] for detailed cell storage protocol.

3.6 Using coDel to Study DNA Elimination Mechanism

1. Design modified targeting sequences (Fig. 3a as examples) (*see* **Note 19**) and prepare them by PCR from the original target or by gene synthesis.
2. Introduce the modified target sequences into pMcoDel as in Subheading 3.2, **steps 4–16**.
3. Introduce the resulting constructs into mating Tetrahymena cells as described in Subheading 3.3.
4. Analyze the target locus by direct PCR as in Subheading 3.4.
5. Measure the deletion frequency by counting the number of cell lines showing PCR products shorter than that from the wild-type target locus (Fig. 3a) (*see* **Note 20**).

4 Notes

1. The wild-type *Tetrahymena thermophila* strains and pMcoDel plasmid are available from the *Tetrahymena* Stock Center <http://tetrahymena.vet.cornell.edu/>.
2. 1× and 2× SPP is prepared from 10× SPP stock, autoclaved for 20 min at 120 °C, and stored at 4 °C (for <10 months) or at RT (for <3 months). 10× SPP is prepared by dissolving 600 g of Bacto Proteose peptone, 120 g of D-(+)-Glucose, 60 g of Bacto yeast extract, and 0.9 g of EDTA iron (III) sodium salt into 3 L (final volume) of cell culture grade water. Bacto Proteose peptone should be added in several batches. It takes around 2 h to dissolve the materials. 10× SPP stock is stored at −20 °C.
3. Our previous study suggested that coDel can be induced if an ~1 kbp segment shares ~90% identity with the targeting sequence [13]. The results shown in Fig. 3 also support this view. We therefore suggest choosing a target sequence that does not share >90% identity with any other genomic locations in all possible 200 bp windows in the genome. Nonetheless, off-target deletion may be induced in an unpredictable manner even with a carefully chosen target. Therefore, it is important to perform a genetic rescue experiment to confirm whether an observed phenotype is caused by the deletion at the targeted locus but not by some off-target deletion.
4. These designs work well for assembly by the NEBuilder HiFi assembly system: FW = 5′-CTTTATTGTTATCATCTTATGACCGC-target-3′; RV = 5′-CTCATCAAGTTGTAATGCTAAAATGC-target-3′.

5. 5′-GCATGCGGCCGC-target-3′ for both primers works fine for ligation.
6. For preparation of materials and detailed protocols for biolistic transformation, refer to [22].
7. Because pMcoDel is based on the high-copy rDNA vector pD5H8 [23], a targeting construct can also be introduced into the new MAC by electroporation [24]. However, we have not tested whether coDel could be induced by electroporation.
8. Aliquots of 1.33 mL of DMSO from freshly opened bottles were placed into 15-mL tubes and stored at −20 °C. Add 8.67 mL 10 mM Tris (pH 7.5) immediately before preparation of frozen cell stocks.
9. The AT-rich genome of *Tetrahymena* sometimes prevents efficient delivery of synthetic double-stranded DNA corresponding to some *Tetrahymena* genome loci. If an alternative region cannot be chosen, multiple shorter synthetic single-stranded DNAs can be stitched by overlapping PCR to produce the desired DNA fragment.
10. For detailed *Tetrahymena* cell culture and mating conditions, *see* [21, 25].
11. Cells lysed in CHAOS can be stored at −20 °C for more than several months.
12. Solution will be opaque. If not, add ~30 μL more phenol extraction buffer.
13. Subheadings 3.1, **steps 13–21** are not necessary to amplify up to ~1 kb DNA from the MAC by standard Taq polymerases. These steps enhance amplification of longer DNA from the MAC or any size of the MIC DNA by Taq polymerase. These steps are often necessary to amplify DNA from any genomic region by high-fidelity polymerases such as Q5 DNA polymerase (NEB) or PrimeSTAR HS DNA Polymerase (Takara).
14. pMcoDel is derived from the rDNA vector pD5H8, which contains the MIC copy of the rDNA locus containing a point mutation that causes the derived rRNA to be insensitive to paromomycin [23]. Because mature, functional rDNA can be produced from the rDNA vector only when it undergoes programmed genome rearrangement in the developing new MAC, all paromomycin-resistant cells obtained by introduction of pMcoDel should be sexual progeny of B2086 and CU428 and have pMcoDel-derived mature rDNA in the (new) MAC.
15. It is preferable to avoid plates in which all wells have growing cells because, in such a plate, the cells in each well were likely established from multiple transformation events.

16. The efficiency of coDel differs among targets. If more than six cell lines with appropriate deletions at the targeted locus are not obtained, it is necessary to screen more transformants, target different regions of the targeted gene, and/or use a longer targeting sequence. However, in our experience, extending the targeting sequence longer than 1 kb does not obviously enhance coDel.
17. If the wild-type locus is not detected in the selected line before cloning in Subheading 3.3, **step 4**, examining a few clones is probably sufficient to obtain a clone in which all the copies of the target locus in the polyploid MAC are deleted.
18. Because coDel induces DNA elimination only in the MAC, genomic PCR may detect the intact wild-type (nondeleted) copy of the targeted locus in the MIC even when all the copies in the MAC have deletions. Therefore, complete absence of the wild-type copy of the targeted gene in the MAC in selected strains may be necessary to be confirmed at the product level (such as by western blot, RT-PCR, or northern blot).
19. A series of truncated sequences and sequences having base replacements in defined intervals.
20. When the 401 bp targeting sequence has base replacement mutations every 10 bp and scnRNAs use X nucleotides as seeds to initiate recognition of the targeted sequences, the total length of the targetable sequence by such scnRNAs (Υ) can be estimated as $\Upsilon = 401-(x - 1) - 40x$. From the truncation study above, the minimum length of the targeting sequence to induce some DNA elimination at this targeted locus was ~150 bp (Fig. 3b). Because the targetable sequence length (Υ) becomes ~150 bp when $X = 6$ (Fig. 3c), we can estimate that the scnRNAs use more than 6 nucleotides for their seed sequence. This estimation of the seed length also explains the result obtained by the 401 bp targeting sequence with mutations every 20 bp, which is estimated to reduce the actual targetable sequence length to 276 bp if the seed size is 6 nt (Fig. 3c). A 276 bp targeting sequence is predicted to induce coDel in 15.3% of transformants (Fig. 3b), which was close to the actual fraction of transformants with deletion by this targeting sequence (=21.7%, Fig. 3a). The seed of scnRNA in Twi1p-scnRNA complex during *Tetrahymena* DNA elimination is most likely 6–8 bases, as in other Argonaute-mediated systems, and some mismatch(s) in the seed can be tolerated if other regions of the scnRNA are also complementary to the target.

Acknowledgments

We would like to thank Tomoko Noto and Julie Saksouk for technical support and Katrina Woolcock for the initiation of the investigation of scnRNA–target interactions. Our research was supported by an advanced grant from the "'Investissements d'avenir" program Labex EpiGenMed (ANR-10-LABX-12-01) and an "Accueil de Chercheurs de Haut Niveau" grant (ANR-16-ACHN-0017) from the French National Research Agency and Emergence de projets (2019-E10) from Cancéropôle Grand Sud-Ouest to K.M.

References

1. Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA (2002) Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell 110:689–699
2. Fang W, Wang X, Bracht JR et al (2012) Piwi-interacting RNAs protect DNA against loss during Oxytricha genome rearrangement. Cell 151:1243–1255
3. Bouhouche K, Gout J-F, Kapusta A et al (2011) Functional specialization of Piwi proteins in paramecium tetraurelia from post-transcriptional gene silencing to genome remodelling. Nucleic Acids Res 39:4249–4264
4. Couvillion MT, Sachidanandam R, Collins K (2010) A growth-essential Tetrahymena Piwi protein carries tRNA fragment cargo. Genes Dev 24:2742–2747
5. Couvillion MT, Lee SR, Hogstad B et al (2009) Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of Tetrahymena thermophila. Genes Dev 23:2016–2032
6. Furrer DI, Swart EC, Kraft MF et al (2017) Two sets of Piwi proteins are involved in distinct sRNA pathways leading to elimination of germline-specific DNA. Cell Rep 20:505–520
7. Hamilton EP, Kapusta A, Huvos PE et al (2016) Structure of the germline genome of Tetrahymena thermophila and relationship to the massively rearranged somatic genome. Elife 5:e19090. https://doi.org/10.7554/eLife.19090
8. Schoeberl UE, Kurth HM, Noto T, Mochizuki K (2012) Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in Tetrahymena. Genes Dev 26:1729–1742
9. Noto T, Kurth HM, Kataoka K et al (2010) The Tetrahymena argonaute-binding protein Giw1p directs a mature argonaute-siRNA complex to the nucleus. Cell 140:692–703
10. Mochizuki K, Gorovsky MA (2005) A dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. Genes Dev 19:77–89
11. Malone CD, Anderson AM, Motl JA et al (2005) Germ line transcripts are processed by a dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila. Mol Cell Biol 25:9151–9164
12. Noto T, Kataoka K, Suhren JH et al (2015) Small-RNA-mediated genome-wide trans-recognition network in Tetrahymena DNA elimination. Mol Cell 59:229–242
13. Hayashi A, Mochizuki K (2015) Targeted gene disruption by ectopic induction of DNA elimination in tetrahymena. Genetics 201:55–64
14. Tian M, Loidl J (2018) A chromatin-associated protein required for inducing and limiting meiotic DNA double-strand break formation. Nucleic Acids Res 46:11822–11834
15. Akematsu T, Fukuda Y, Garg J et al (2017) Post-meiotic DNA double-strand breaks occur in Tetrahymena, and require topoisomerase II and Spo11. Elife 6:e26176. https://doi.org/10.7554/eLife.26176
16. Urbanska P, Joachimiak E, Bazan R et al (2018) Ciliary proteins Fap43 and Fap44 interact with each other and are essential for proper cilia and flagella beating. Cell Mol Life Sci 75:4479–4493
17. Soh AWJ, van Dam TJP, Stemm-Wolf AJ et al (2020) Ciliary force-responsive striated fibers promote basal body connections and cortical interactions. J Cell Biol 219(1):e201904091. https://doi.org/10.1083/jcb.201904091
18. Bazan R, Schröfel A, Joachimiak E et al (2021) Ccdc113/Ccdc96 complex, a novel regulator of ciliary beating that connects radial spoke 3 to

dynein g and the nexin link. PLoS Genet 17: e1009388

19. Tian M, Agreiter C, Loidl J (2020) Spatial constraints on chromosomes are instrumental to meiotic pairing. J Cell Sci 133(22): jcs253724. https://doi.org/10.1242/jcs.253724
20. Noto T, Mochizuki K (2018) Small RNA-mediated trans-nuclear and trans-element Communications in Tetrahymena DNA elimination. Curr Biol 28:1938–1949.e5
21. Cassidy-Hanley DM (2012) Tetrahymena in the laboratory: strain resources, methods for culture, maintenance, and storage. Methods Cell Biol 109:237–276
22. Bruns PJ, Cassidy-Hanley D (2000) Biolistic transformation of macro- and micronuclei. Methods Cell Biol 62:501–512
23. Yao MC, Yao CH (1991) Transformation of Tetrahymena to cycloheximide resistance with a ribosomal protein gene through sequence replacement. Proc Natl Acad Sci U S A 88: 9493–9497
24. Gaertig J, Gorovsky MA (1992) Efficient mass transformation of Tetrahymena thermophila by electroporation of conjugants. Proc Natl Acad Sci U S A 89:9196–9200
25. Chalker DL (2012) Transformation and strain engineering of Tetrahymena. Methods Cell Biol 109:327–345
26. Noto T, Mochizuki K (2017) Whats, hows and whys of programmed DNA elimination in Tetrahymena. Open Biol 7:170172

Discussion and prospective

In my PhD study, I identified a novel gene *EMA2*, which is required for genome-wide lncRNA transcription from the parental MAC and required for heterochromatin formation in the parental MAC and the new MAC. I further show that *EMA2* is required for TDSD of scnRNAs and completing DNA elimination. *EMA2* encodes for a protein Ema2 possessing an SP-RING domain which can be found in SUMO E3 ligases of PIAS family. I show that Ema2 interact with SUMO conjugating enzyme E2, enhances global SUMOylation during conjugation and target the conserved transcriptional regulator Spt6 for SUMOylation. I propose that Ema2-dependent Spt6 SUMOylation is required for the lncRNA transcription from the parental MAC and thus for the TDSD of scnRNAs.

Role of Ema2 in the TDSD of scnRNAs by regulating lncRNA transcription

Because the RNA helicase Ema1 is required for the interaction between Twi1-scnRNA complex and lncRNAs, it was previously proposed that Twi1-scnRNA complex interact with chromatin through nascent lncRNA transcript which in turn facilitates TDSD of scnRNAs in the parental MAC and heterochromatin formation in the new MAC (Aronica et al. 2008; Noto and Mochizuki 2017). In *Paramecium* it was demonstrated that the disruption of nascent lncRNA from the parental MAC using RNAi affect the DNA elimination in this ciliate (Lepère et al. 2008). However, there was no direct evidence for the requirement of lncRNA transcription for TDSD of scnRNAs. The *EMA2* KO cell reported in this study is the first mutant which disrupts the lncRNA transcription from the parental MAC (Figure 8) and the TDSD of scnRNAs (Figure 5). Therefore, this study provides the first direct evidence for the role of lncRNA transcript form the parental MAC for the TDSD of scnRNAs.

Because *EMA1* is not required for the production of lncRNAs (Figure 9), Ema1 likely acts in downstream of Ema2. The fact that TDSD of scnRNA is more severely affected in *EMA1* KO cells than in *EMA2* KO cells (Figure 5B) may indicate that Ema2-independent mRNA or lncRNA transcripts can still induce some level of Ema1-dependent TDSD of scnRNAs. Future study should be designed to understand the genome wide pattern of lncRNAs transcription from the parental MAC in the presence and absence of Ema2.

I further found that Ema2 is required for the accumulation of H3K27me3 in the parental MAC during conjugation (Figure 4). Similarly, I found that *TWI1* and *EMA1* are required for the accumulation of H3K27me3 in the parental MAC (Figure 4), which suggests that lncRNA-scnRNA interaction results in both TDSD of scnRNAs and heterochromatin formation in the parental MAC. Ezl1, the H3K9 and H3K27 dual-specific methyltransferase (Liu et al. 2007; Xu et al. 2021), is also required for the accumulation of H3K27me3 in the parental MAC (Figure 4) (Liu et al. 2007) but our unpublished study indicated that Ezl1 is not important for the TDSD of scnRNAs. Therefore, the accumulation of H3K27me3 in the parental MAC is not a prerequisite of the TDSD of scnRNAs but the lncRNA-scnRNA interaction induces these two events in parallel. The role of H3K27me3 in the parental MAC, if any, is unclear. Future studies should clarify the distribution of H3K27me3 in the parental MAC

It was also previously observed that the scnRNAs complementary to IESs in the new MAC are downregulated in parallel to heterochromatin formation (Noto et al. 2015). Therefore, scnRNAs likely also subjected to TDSD in the new MAC and I propose that the lncRNA-scnRNA interaction results in TDSD of scnRNA and heterochromatin formation in both the parental and the new MAC. The biological importance of TDSD in the new MAC should be accessed by producing and examining germline KO strains for *EMA1* and/or *EMA2.*

The effect of Ema2 on heterochromatin formation in the new MAC

In the *EMA2* KO strains I used in this study, the heterochromatin marks (H3K9me3 and H3K27me3) are greatly reduced at 9 hpm while they are accumulated at 12 hpm (Figure 3). The delayed accumulation of H3K9me3 and H3K27me3 in the *EMA2* KO cells can be explained by the *EMA2* mRNA expression from zygotic nucleus (Figure 1A), which may drive the heterochromatin formation. The partial block of DNA elimination in the *EMA2* KO cells (Figure 2B) might be caused by the delayed heterochromatin formation. Alternatively, the loss of TDSD of scnRNAs may results in heterochromatin formation beyond IESs and thus DNA excision machineries cannot eliminate IESs properly/efficiently. Delayed heterochromatin formation and partial block of DNA elimination were also reported in the *EMA1* somatic KO cells, where the MIC copy of *EMA1* was still intact (Aronica et al. 2008). In future, the genome-wide distribution of heterochromatin in the new MAC in the presence or the absence of *EMA2* should be examined by ChIP-Seq to understand how the loss of TDSD of scnRNAs in the parental MAC affects the

heterochromatin formation in the new MAC. Also, the deletion of *EMA2* from both MAC and the MIC will further clarify the role of *EMA2* in the new MAC.

The role of Ema2-dependent SUMOylation of Spt6 in lncRNA transcription

EMA2 encodes a SUMO E3 ligase and contributes to the accumulation of about 30% of SUMOylated proteins during conjugation (Figure 6D). Therefore, I believe that Ema2 acts of lncRNA transcription, TDSD of scnRNAs, heterochromatin formation and DNA elimination through protein SUMOylation. The SUMO pathway was shown to be required for DNA elimination in the closely related ciliate *Paramecium tetraurelia* (Matsuda and Forney 2006). However, because the SUMO pathway is essential for vegetative growth in *Tetrahymena* (Nasir et al. 2015; Yang et al. 2015), and thus cells defective in the SUMOylation machinery does not undergoes sexual reproduction, it has been unclear if SUMOylation also plays some role in DNA elimination in *Tetrahymena.* This study demonstrated that SUMOylation is also important for DNA elimination in *Tetrahymena* and further showed that SUMOylation is involved in lncRNA transcription, TDSD of scnRNAs, and heterochromatin formation.

In yeasts, it was shown that SUMOylation of Swi6, Chp2 and Clr4 is important for RNAi-directed heterochromatin silencing (Shin et al. 2005). In *Drosophila*, a SUMO E3 ligase, Su(var)2-10 is important for the recruitment of SetDB1, a histone methyltransferase which deposits H3K9me3 (Ninova et al. 2020a, 2020b), although, the SUMOylation of Panx which is important to connect piRNA pathway to the cellular heterochromatin machinery is independent of Su(var)2-10 SUMOylation (Andreev et al. 2022). Similarly in *C. elegens*, the SUMOylated HAD-1, a histone deacetylase promotes sRNA-directed transposons silencing (Kim et al. 2021). All the above SUMOylated events occur downstream of nascent lncRNA transcription in the RNAi-dependent silencing pathway. In this study I show for the first time that a SUMO E3 ligase is required for the transcription of lncRNA, which add a new undescribed layer to the regulation of RNAi-dependent silencing pathway by SUMOylation.

I found that Ema2 is required for the SUMOylation of a transcriptional regulator, Spt6 (Figure 7). Because SUMOylated Spt6 (Figure 7C) overlaps to the majority of SUMOylated proteins lost in the *EMA2* KO cells (Figure 6D), Spt6 is likely the major SUMOylation target of Ema2. Spt6 is known to regulate various steps of RNA Pol II-mediated transcription, such as initiation,

elongation, and termination (Ardehali et al. 2009; Endoh et al. 2004; Doris et al. 2018; Adkins and Tyler 2006; Ivanovska et al. 2011). Therefore, Ema2-dependent SUMOylation of Spt6 likely regulates the transcription of lncRNA in the parental MAC. It was reported that loss of function mutations of *SPT6* resulted in abnormal up-regulation of intergenic and anti-sense transcriptions in human cells and yeasts (Narain et al. 2021; Doris et al. 2018; DeGennaro et al. 2013; Kato et al. 2013). Hence, I hypothesize that the un-SUMOylated Spt6 suppress the occurrence of promiscuous lncRNA transcription in vegetative growth and SUMOylation of Spt6 loosens such safeguarding ability of Spt6 to promote the genome wide lncRNA transcription and thus TDSD of scnRNAs (Figure XXV). Future study should identify Ema2-dependent SUMOylated residues of Spt6 by genetically engineering Smt3 to make it compatible for mass-spectrometry identification and then assess effects of un-SUMOylatable mutation of Spt6 in lncRNA transcription.

Model of the TSDS of scnRNAs

Based on the findings of the present study and the above discussions, I propose a model for the regulation of TDSD by Ema2-dependent SUMOylation of Spt6 (Figure XXV). I first propose that during vegetative growth (Figure XXV top), un-SUMOylated Spt6 confines transcriptions of mRNAs by RNA Pol II from the genic region and suppress the occurrence of transcription from non-coding sequences by inhibiting promiscuous transcriptional initiation and/or promoting proper termination at the ends of codings. During sexual reproduction, the transcription of the genic region may still be regulated similarly but Ema2-dependent SUMOylation of Spt6 loosens its safeguarding ability and allows RNA Pol II to transcribe nongenic regions at a genome-wide level (Figure XXV bottom). The interaction of Twi1-scnRNA complex to the nascent lncRNA is facilitated by Ema1 and induces TDSD of scnRNAs.

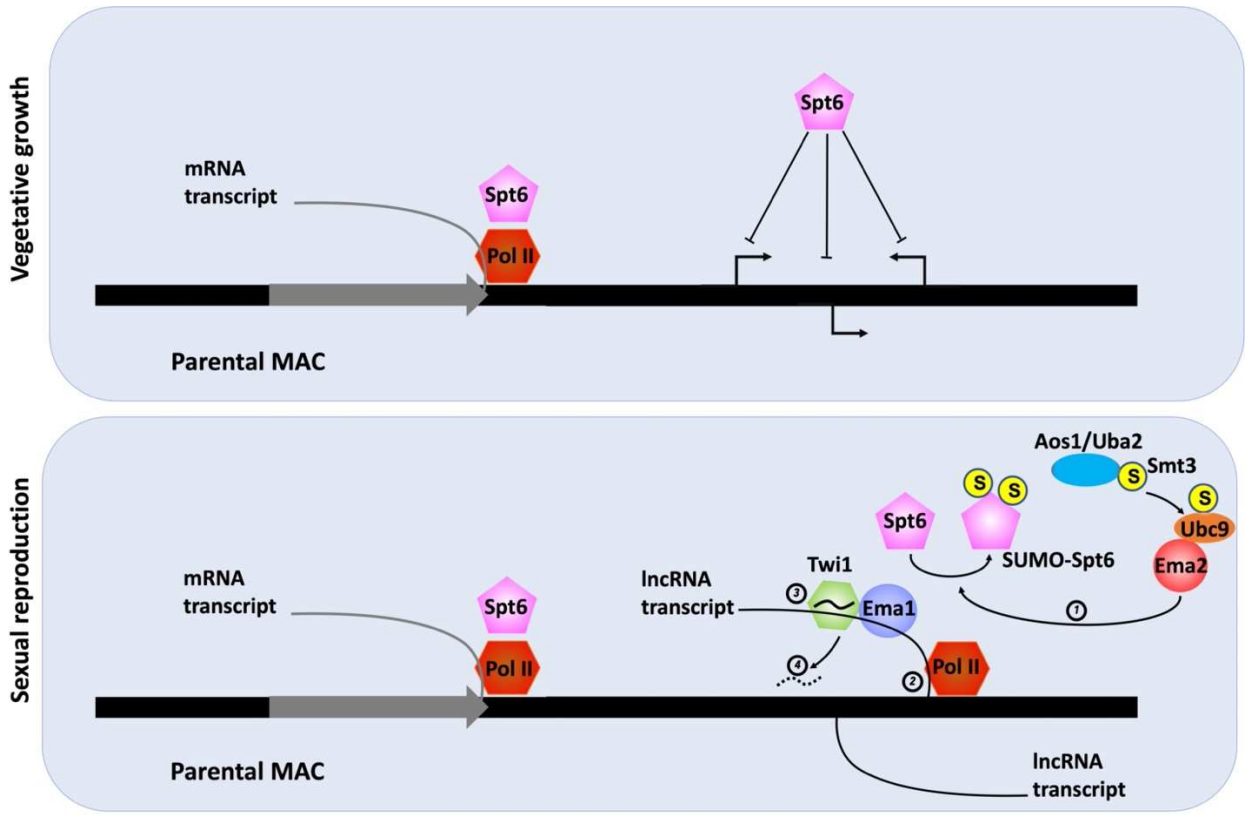


Figure XVI. Model for regulation of the TDSD of scnRNAs.

(Top) During vegetative growth, mRNA transcription is confined by Spt6-Pol II to the gene bodies and the expression of intergenic regions is suppressed by un-SUMOylated Spt6. (Bottom) During sexual reproduction, (1) Ema2-directed SUMOylation of Spt6 removes the suppression of promiscuous transcriptions from the intergenic regions and thus (2) promotes the Pol II-mediated lncRNA transcriptions in both directions, (3) the RNA helicase Ema1 facilitates the interaction between nascent lncRNA transcripts and Twi1-loaded scnRNAs, and (4) this interaction induces the TDSD of scnRNAs.

References

Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, et al. 2012. The revised classification of eukaryotes. *J Eukaryot Microbiol* **59**: 429–493.

Ambros V. 2003. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell* **113**: 673–676.

Ameres SL, Horwich MD, Hung J-H, Xu J, Ghildiyal M, Weng Z, Zamore PD. 2010. Target RNA-directed trimming and tailing of small silencing RNAs. *Science* **328**: 1534–1539.

Ameres SL, Zamore PD. 2013. Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol* **14**: 475–488.

Andersen PR, Tirian L, Vunjak M, Brennecke J. 2017. A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* **549**: 54–59.

Andreev VI, Yu C, Wang J, Schnabl J, Tirian L, Gehre M, Handler D, Duchek P, Novatchkova M, Baumgartner L, et al. 2022. Panoramix SUMOylation on chromatin connects the piRNA pathway to the cellular heterochromatin machinery. *Nat Struct Mol Biol* **29**: 130–142.

Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Curr Biol* **11**: 1017–1027.

Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**: 785–799.

Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**: 203–207.

Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Curr Biol* **11**: 1017–1027.

Aravind L, Koonin EV. 2000. SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci* **25**: 112–114.

Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**: 744–747.

Aronica L, Bednenko J, Noto T, DeSouza LV, Siu KWM, Loidl J, Pearlman RE, Gorovsky MA, Mochizuki K. 2008. Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in Tetrahymena. *Genes Dev* **22**: 2228–2241.

Baccarini A, Chauhan H, Gardner TJ, Jayaprakash AD, Sachidanandam R, Brown BD. 2011. Kinetic analysis reveals the fate of a microRNA following target regulation in mammalian cells. *Curr Biol* **21**: 369–376.

Beauclair G, Bridier-Nahmias A, Zagury J-F, Saïb A, Zamborlini A. 2015. JASSA: a comprehensive tool for prediction of SUMOylation sites and SIMs. *Bioinformatics* **31**: 3483–3491.

Bednenko J, Noto T, DeSouza LV, Siu KWM, Pearlman RE, Mochizuki K, Gorovsky MA. 2009. Two GW repeat proteins interact with Tetrahymena thermophila argonaute and promote genome rearrangement. *Mol Cell Biol* **29**: 5020–5030.

Berdnik D, Favaloro V, Luo L. 2012. The SUMO protease Verloren regulates dendrite and axon targeting in olfactory projection neurons. *J Neurosci* **32**: 8331–8340.

Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363–366.

Bhagwat NR, Owens SN, Ito M, Boinapalli JV, Poa P, Ditzel A, Kopparapu S, Mahalawat M, Davies OR, Collins SR, et al. 2021. SUMO is a pervasive regulator of meiosis. *eLife* **10**.

Bhaskar V, Smith M, Courey AJ. 2002. Conjugation of Smt3 to dorsal may potentiate the Drosophila immune response. *Mol Cell Biol* **22**: 492–504.

Bitetti A, Mallory AC, Golini E, Carrieri C, Carreño Gutiérrez H, Perlas E, Pérez-Rico YA, Tocchini-Valentini GP, Enright AJ, Norton WHJ, et al. 2018. MicroRNA degradation by a conserved target RNA regulates animal behavior. *Nat Struct Mol Biol* **25**: 244–251.

Bohmert K, Camus I, Bellini C, Bouchez D, Caboche M, Benning C. 1998. AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J* **17**: 170–180.

Boland A, Huntzinger E, Schmidt S, Izaurralde E, Weichenrieder O. 2011. Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. *Proc Natl Acad Sci USA* **108**: 10466–10471.

Boots JL, Moll I. 2009. The impact of non-coding RNAs: workshop on new functions of regulatory RNAs in pro- & eukaryotes. *EMBO Rep* **10**: 563–567.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**: 1089–1103.

Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**: 1387–1392.

Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, Allis CD. 1996. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**: 843–851.

Bruns PJ, Brussard TB. 1974. Pair formation in tetrahymena pyriformis, an inducible developmental system. *J Exp Zool* **188**: 337–344.

Bühler M, Verdel A, Moazed D. 2006. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**: 873–886.

Capili AD, Lima CD. 2007a. Structure and analysis of a complex between SUMO and Ubc9 illustrates features of a conserved E2-Ubl interaction. *J Mol Biol* **369**: 608–618.

Capili AD, Lima CD. 2007b. Taking it step by step: mechanistic insights from structural studies of ubiquitin/ubiquitin-like protein modification pathways. *Curr Opin Struct Biol* **17**: 726–735.

Carthew RW, Sontheimer EJ. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**: 642–655.

Cech TR, Zaug AJ, Grabowski PJ. 1981. In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**: 487–496.

Cenik ES, Fukunaga R, Lu G, Dutcher R, Wang Y, Tanaka Hall TM, Zamore PD. 2011. Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol Cell* **42**: 172–184.

Cerutti H, Casas-Mollano JA. 2006. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet* **50**: 81–99.

Chalker DL, Meyer E, Mochizuki K. 2013. Epigenetics of ciliates. *Cold Spring Harb Perspect Biol* **5**: a017764.

Chalker DL, Yao MC. 2001. Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila. *Genes Dev* **15**: 1287–1298.

Chandradoss SD, Schirle NT, Szczepaniak M, MacRae IJ, Joo C. 2015. A dynamic search process underlies microrna targeting. *Cell* **162**: 96–107.

Chang H-M, Triboulet R, Thornton JE, Gregory RI. 2013. A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. *Nature* **497**: 244–248.

Cheng C-Y, Vogt A, Mochizuki K, Yao M-C. 2010. A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila. *Mol Biol Cell* **21**: 1753–1762.

Chen A, Mannen H, Li SS. 1998. Characterization of mouse ubiquitin-like SMT3A and SMT3B cDNAs and gene/pseudogenes. *Biochem Mol Biol Int* **46**: 1161–1174.

Chen J, Liu L, You C, Gu J, Ruan W, Zhang L, Gan J, Cao C, Huang Y, Chen X, et al. 2018. Structural and biochemical insights into small RNA 3’ end trimming by Arabidopsis SDN1. *Nat Commun* **9**: 3585.

Cheong MS, Park HC, Hong MJ, Lee J, Choi W, Jin JB, Bohnert HJ, Lee SY, Bressan RA, Yun D-J. 2009. Specific domain structures control abscisic acid-, salicylic acid-, and stress-mediated SIZ1 phenotypes. *Plant Physiol* **151**: 1930–1942.

Chung W-J, Okamura K, Martin R, Lai EC. 2008. Endogenous RNA interference provides a somatic defense against Drosophila transposons. *Curr Biol* **18**: 795–802.

Ciechanowska K, Pokornowska M, Kurzyńska-Kokorniak A. 2021. Genetic Insight into the Domain Structure and Functions of Dicer-Type Ribonucleases. *Int J Mol Sci* **22**.

Cox DN, Chao A, Lin H. 2000. piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127**: 503–514.

Coyne RS, Stover NA, Miao W. 2012. Whole genome studies of Tetrahymena. *Methods Cell Biol* **109**: 53–81.

Cranert S, Heyse S, Linger BR, Lescasse R, Price C. 2014. Tetrahymena Pot2 is a developmentally regulated paralog of Pot1 that localizes to chromosome breakage sites but not to telomeres. *Eukaryotic Cell* **13**: 1519–1529.

Crozet P, Margalha L, Butowt R, Fernandes N, Elias CA, Orosa B, Tomanov K, Teige M, Bachmair A, Sadanandom A, et al. 2016. SUMOylation represses SnRK1 signaling in Arabidopsis. *Plant J* **85**: 120–133.

Cullen BR. 2004. Transcription and processing of human microRNA precursors. *Mol Cell* **16**: 861–865.

Czech B, Hannon GJ. 2016. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends Biochem Sci* **41**: 324–337.

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. 2008. An endogenous small interfering RNA pathway in Drosophila. *Nature* **453**: 798–802.

Davis MC, Ward JG, Herrick G, Allis CD. 1992. Programmed nuclear death: apoptotic-like degradation of specific nuclei in conjugating Tetrahymena. *Dev Biol* **154**: 419–432.

DeGennaro CM, Alver BH, Marguerat S, Stepanova E, Davis CP, Bähler J, Park PJ, Winston F. 2013. Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. *Mol Cell Biol* **33**: 4779–4792.

Dennis C, Brasset E, Sarkar A, Vaury C. 2016. Export of piRNA precursors by EJC triggers assembly of cytoplasmic Yb-body in Drosophila. *Nat Commun* **7**: 13739.

Desterro JM, Rodriguez MS, Kemp GD, Hay RT. 1999. Identification of the enzyme required for activation of the small ubiquitin-like protein SUMO-1. *J Biol Chem* **274**: 10618–10624.

Djupedal I, Portoso M, Spåhr H, Bonilla C, Gustafsson CM, Allshire RC, Ekwall K. 2005. RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev* **19**: 2301–2306.

Dohmen RJ, Stappen R, McGrath JP, Forrová H, Kolarov J, Goffeau A, Varshavsky A. 1995. An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. *J Biol Chem* **270**: 18099–18109.

Donaghue C, Bates H, Cotterill S. 2001. Identification and characterisation of the Drosophila homologue of the yeast Uba2 gene. *Biochim Biophys Acta* **1518**: 210–214.

Dönertas D, Sienski G, Brennecke J. 2013. Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. *Genes Dev* **27**: 1693–1705.

Doris SM, Chuang J, Viktorovskaya O, Murawska M, Spatt D, Churchman LS, Winston F. 2018. Spt6 is required for the fidelity of promoter selection. *Mol Cell* **72**: 687-699.e6.

Duchaine TF, Wohlschlegel JA, Kennedy S, Bei Y, Conte D, Pang K, Brownell DR, Harding S, Mitani S, Ruvkun G, et al. 2006. Functional proteomics reveals the biochemical niche of C. elegans DCR-1 in multiple small-RNA-mediated pathways. *Cell* **124**: 343–354.

Duval D, Duval G, Kedinger C, Poch O, Boeuf H. 2003. The “PINIT” motif, of a newly identified conserved domain of the PIAS protein family, is essential for nuclear retention of PIAS3L. *FEBS Lett* **554**: 111–118.

Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM, et al. 2006. Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. *PLoS Biol* **4**: e286.

Eisen TJ, Eichhorn SW, Subtelny AO, Lin KS, McGeary SE, Gupta S, Bartel DP. 2020. The Dynamics of Cytoplasmic mRNA Metabolism. *Mol Cell* **77**: 786-799.e10.

Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**: 494–498.

Erker Y, Neyret-Kahn H, Seeler JS, Dejean A, Atfi A, Levy L. 2013. Arkadia, a novel SUMO-targeted ubiquitin ligase involved in PML degradation. *Mol Cell Biol* **33**: 2163–2177.

Evdokimov E, Sharma P, Lockett SJ, Lualdi M, Kuehn MR. 2008. Loss of SUMO1 in mice affects RanGAP1 localization and formation of PML nuclear bodies, but is not lethal as it can be compensated by SUMO2 or SUMO3. *J Cell Sci* **121**: 4106–4113.

Feng L, Wang G, Hamilton EP, Xiong J, Yan G, Chen K, Chen X, Dui W, Plemens A, Khadr L, et al. 2017. A germline-limited piggyBac transposase gene is required for precise excision in *Tetrahymena* genome rearrangement. *Nucleic Acids Res* **45**: 9481–9502.

Filipowicz W, Bhattacharyya SN, Sonenberg N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* **9**: 102–114.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**: 806–811.

Förstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD. 2007. Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* **130**: 287–297.

Gareau JR, Lima CD. 2010. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol* **11**: 861–871.

Geiss-Friedlander R, Melchior F. 2007. Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol* **8**: 947–956.

Ge DT, Wang W, Tipping C, Gainetdinov I, Weng Z, Zamore PD. 2019. The RNA-Binding ATPase, Armitage, Couples piRNA Amplification in Nuage to Phased piRNA Production on Mitochondria. *Mol Cell* **74**: 982-995.e6.

Ghildiyal M, Zamore PD. 2009. Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10**: 94–108.

Ghini F, Rubolino C, Climent M, Simeone I, Marzi MJ, Nicassio F. 2018. Endogenous transcripts control miRNA levels and activity in mammalian cells by target-directed miRNA degradation. *Nat Commun* **9**: 3119.

Gibbons IR, Rowe AJ. 1965. Dynein: A Protein with Adenosine Triphosphatase Activity from Cilia. *Science* **149**: 424–426.

Golden DE, Gerbasi VR, Sontheimer EJ. 2008. An inside job for siRNAs. *Mol Cell* **31**: 309–312.

Goldfarb DS, Gorovsky MA. 2009. Nuclear dimorphism: two peas in a pod. *Curr Biol* **19**: R449-52.

Gong L, Li B, Millas S, Yeh ET. 1999. Molecular cloning and characterization of human AOS1 and UBA2, components of the sentrin-activating enzyme complex. *FEBS Lett* **448**: 185–189.

Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. 2005. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123**: 631–640.

Gregory RI, Yan K-P, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**: 235–240.

Greider CW, Blackburn EH. 1985. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**: 405–413.

Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**: 91–105.

Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106**: 23–34.

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science* **315**: 1587–1590.

Hamilton AJ, Baulcombe DC. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**: 950–952.

Hamilton EP, Kapusta A, Huvos PE, Bidwell SL, Zafar N, Tang H, Hadjithomas M, Krishnakumar V, Badger JH, Caler EV, et al. 2016. Structure of the germline genome of *Tetrahymena thermophila* and relationship to the massively rearranged somatic genome. *eLife* **5**.

Hamilton EP, Williamson S, Dunn S, Merriam V, Lin C, Vong L, Russell-Colantonio J, Orias E. 2006. The highly conserved family of Tetrahymena thermophila chromosome breakage elements contains an invariant 10-base-pair core. *Eukaryotic Cell* **5**: 771–780.

Hammond SM, Bernstein E, Beach D, Hannon GJ. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**: 293–296.

Hannich JT, Lewis A, Kroetz MB, Li S-J, Heide H, Emili A, Hochstrasser M. 2005. Defining the SUMO-modified proteome by multiple approaches in Saccharomyces cerevisiae. *J Biol Chem* **280**: 4102–4110.

Han J, LaVigne CA, Jones BT, Zhang H, Gillett F, Mendell JT. 2020. A ubiquitin ligase mediates target-directed microRNA decay independently of tailing and trimming. *Science* **370**.

Han J, Lee Y, Yeom K-H, Kim Y-K, Jin H, Kim VN. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**: 3016–3027.

Han J, Mendell JT. 2022. MicroRNA turnover: a tale of tailing, trimming, and targets. *Trends Biochem Sci*.

Hariharasudhan G, Jeong S-Y, Kim M-J, Jung SM, Seo G, Moon J-R, Lee S, Chang I-Y, Kee Y, You HJ, et al. 2022. TOPORS-mediated RAD51 SUMOylation facilitates homologous recombination repair. *Nucleic Acids Res* **50**: 1501–1516.

Harris AN, Macdonald PM. 2001. Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C. *Development* **128**: 2823–2832.

Hay RT. 2005. SUMO: a history of modification. *Mol Cell* **18**: 1–12.

Heery DM, Kalkhoven E, Hoare S, Parker MG. 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387**: 733–736.

Hendriks IA, D'Souza RCJ, Yang B, Verlaan-de Vries M, Mann M, Vertegaal ACO. 2014. Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat Struct Mol Biol* **21**: 927–936.

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**: 118–120.

Hickey CM, Wilson NR, Hochstrasser M. 2012. Function and regulation of SUMO proteases. *Nat Rev Mol Cell Biol* **13**: 755–766.

Hirakata S, Siomi MC. 2019. Assembly and Function of Gonad-Specific Non-Membranous Organelles in Drosophila piRNA Biogenesis. *Noncoding RNA* **5**.

Hochstrasser M. 2001. SP-RING for SUMO: new functions bloom for a ubiquitin-like protein. *Cell* **107**: 5–8.

Höck J, Meister G. 2008. The Argonaute protein family. *Genome Biol* **9**: 210.

Hoege C, Pfander B, Moldovan G-L, Pyrowolakis G, Jentsch S. 2002. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419**: 135–141.

Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, Zamore PD. 2007. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol* **17**: 1265–1272.

Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov DV, Blaser H, Raz E, Moens CB, et al. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**: 69–82.

Huang W-C, Ko T-P, Li SS-L, Wang AH-J. 2004. Crystal structures of the human SUMO-2 protein at 1.6 A and 1.2 A resolution: implication on the functional differences of SUMO proteins. *Eur J Biochem* **271**: 4114–4122.

Hutvágner G, McLachlan J, Pasquinelli AE, Bálint E, Tuschl T, Zamore PD. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**: 834–838.

Izumi N, Shoji K, Suzuki Y, Katsuma S, Tomari Y. 2020. Zucchini consensus motifs determine the mechanism of pre-piRNA production. *Nature* **578**: 311–316.

Janke C, Rogowski K, Wloga D, Regnard C, Kajava AV, Strub J-M, Temurak N, van Dijk J, Boucher D, van Dorsselaer A, et al. 2005. Tubulin polyglutamylase enzymes are members of the TTL domain protein family. *Science* **308**: 1758–1762.

Jinek M, Doudna JA. 2009. A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457**: 405–412.

Johnson ES, Gupta AA. 2001. An E3-like factor that promotes SUMO conjugation to the yeast septins. *Cell* **106**: 735–744.

Johnson ES, Schwienhorst I, Dohmen RJ, Blobel G. 1997. The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. *EMBO J* **16**: 5509–5519.

Johnson ES. 2004. Protein modification by SUMO. *Annu Rev Biochem* **73**: 355–382.

Jones D, Crowe E, Stevens TA, Candido EPM. 2002. Functional and phylogenetic analysis of the ubiquitylation system in Caenorhabditis elegans: ubiquitin-conjugating enzymes, ubiquitin-activating enzymes, and ubiquitin-like proteins. *Genome Biol* **3**: RESEARCH0002.

Kagey MH, Melhuish TA, Wotton D. 2003. The polycomb protein Pc2 is a SUMO E3. *Cell* **113**: 127–137.

Kahyo T, Nishida T, Yasuda H. 2001. Involvement of PIAS1 in the sumoylation of tumor suppressor p53. *Mol Cell* **8**: 713–718.

Kamitani T, Kito K, Nguyen HP, Fukuda-Kamitani T, Yeh ET. 1998a. Characterization of a second member of the sentrin family of ubiquitin-like proteins. *J Biol Chem* **273**: 11349–11353.

Kamitani T, Kito K, Nguyen HP, Wada H, Fukuda-Kamitani T, Yeh ET. 1998b. Identification of three major sentrinization sites in PML. *J Biol Chem* **273**: 26675–26682.

Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, Rajewsky K. 2005. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev* **19**: 489–501.

Kataoka K, Mochizuki K. 2017. Heterochromatin aggregation during DNA elimination in Tetrahymena is facilitated by a prion-like protein. *J Cell Sci* **130**: 480–489.

Kataoka K, Mochizuki K. 2015. Phosphorylation of an HP1-like Protein Regulates Heterochromatin Body Assembly for DNA Elimination. *Dev Cell* **35**: 775–788.

Kataoka K, Noto T, Mochizuki K. 2016. Phosphorylation of an HP1-like protein is a prerequisite for heterochromatin body formation in Tetrahymena DNA elimination. *Proc Natl Acad Sci USA* **113**: 9027–9032.

Kato H, Goto DB, Martienssen RA, Urano T, Furukawa K, Murakami Y. 2005. RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science* **309**: 467–469.

Kato H, Okazaki K, Iida T, Nakayama J-I, Murakami Y, Urano T. 2013. Spt6 prevents transcription-coupled loss of posttranslationally modified histone H3. *Sci Rep* **3**: 2186.

Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H. 2008. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**: 793–797.

Kerscher O, Felberbaum R, Hochstrasser M. 2006. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol* **22**: 159–180.

Kerscher O. 2007. SUMO junction-what's your function? New insights through SUMO-interacting motifs. *EMBO Rep* **8**: 550–555.

Khvorova A, Reynolds A, Jayasena SD. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**: 209–216.

Kim H, Ding Y-H, Zhang G, Yan Y-H, Conte D, Dong M-Q, Mello CC. 2021. HDAC1 SUMOylation promotes Argonaute-directed transcriptional silencing in C. elegans. *eLife* **10**.

Kim JH, Baek SH. 2009. Emerging roles of desumoylating enzymes. *Biochim Biophys Acta* **1792**: 155–162.

Kingston ER, Bartel DP. 2019. Global analyses of the dynamics of mammalian microRNA metabolism. *Genome Res* **29**: 1777–1790.

Kirino Y, Mourelatos Z. 2007. Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nat Struct Mol Biol* **14**: 347–348.

Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, Zhang F, Schultz N, Koppetsch BS, Nowosielska A, et al. 2009. The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138**: 1137–1149.

Kleaveland B, Shi CY, Stefano J, Bartel DP. 2018. A network of noncoding regulatory rnas acts in the mammalian brain. *Cell* **174**: 350-362.e17.

Klug H, Xaver M, Chaugule VK, Koidl S, Mittler G, Klein F, Pichler A. 2013. Ubc9 sumoylation controls SUMO chain formation and meiotic synapsis in Saccharomyces cerevisiae. *Mol Cell* **50**: 625–636.

Knight SW, Bass BL. 2001. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. *Science* **293**: 2269–2271.

Knipscheer P, van Dijk WJ, Olsen JV, Mann M, Sixma TK. 2007. Noncovalent interaction between Ubc9 and SUMO promotes SUMO chain formation. *EMBO J* **26**: 2797–2807.

Krützfeldt J, Kuwajima S, Braich R, Rajeev KG, Pena J, Tuschl T, Manoharan M, Stoffel M. 2007. Specificity, duplex degradation and subcellular localization of antagomirs. *Nucleic Acids Res* **35**: 2885–2892.

Krützfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, Stoffel M. 2005. Silencing of microRNAs in vivo with "antagomirs". *Nature* **438**: 685–689.

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, et al. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**: 908–917.

Kurepa J, Walker JM, Smalle J, Gosink MM, Davis SJ, Durham TL, Sung D-Y, Vierstra RD. 2003. The small ubiquitin-like modifier (SUMO) protein modification system in Arabidopsis. Accumulation of SUMO1 and -2 conjugates is increased by stress. *J Biol Chem* **278**: 6862–6872.

Kurth HM, Mochizuki K. 2009. 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena. *RNA* **15**: 675–685.

Kwak PB, Tomari Y. 2012. The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nat Struct Mol Biol* **19**: 145–151.

Lai EC. 2002. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* **30**: 363–364.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853–858.

Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**: 858–862.

Lau P-W, Guiley KZ, De N, Potter CS, Carragher B, MacRae IJ. 2012. The molecular architecture of human Dicer. *Nat Struct Mol Biol* **19**: 436–440.

La Salle S, Sun F, Zhang X-D, Matunis MJ, Handel MA. 2008. Developmental control of sumoylation pathway proteins in mouse male germ cells. *Dev Biol* **321**: 227–237.

Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Tóth KF. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev* **27**: 390–399.

Lee RC, Ambros V. 2001. An extensive class of small RNAs in Caenorhabditis elegans. *Science* **294**: 862–864.

Lee RC, Feinbaum RL, Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843–854.

Lee S, Song J, Kim S, Kim J, Hong Y, Kim Y, Kim D, Baek D, Ahn K. 2013. Selective degradation of host MicroRNAs by an intergenic HCMV noncoding RNA accelerates virus production. *Cell Host Microbe* **13**: 678–690.

Lee Y, Han J, Yeom KH, Jin H, Kim VN. 2006. Drosha in primary microRNA processing. *Cold Spring Harb Symp Quant Biol* **71**: 51–57.

Lee Y, Jeon K, Lee J-T, Kim S, Kim VN. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**: 4663–4670.

Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW. 2004. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**: 69–81.

Lehembre F, Badenhorst P, Müller S, Travers A, Schweisguth F, Dejean A. 2000. Covalent modification of the transcriptional repressor tramtrack by the ubiquitin-related protein Smt3 in Drosophila flies. *Mol Cell Biol* **20**: 1072–1082.

Lepère G, Bétermier M, Meyer E, Duharcourt S. 2008. Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in Paramecium tetraurelia. *Genes Dev* **22**: 1501–1512.

Lewis BP, Shih I, Jones-Rhoades MW, Bartel DP, Burge CB. 2003. Prediction of mammalian microRNA targets. *Cell* **115**: 787–798.

Lin C-YG, Chao J-L, Tsai H-K, Chalker D, Yao M-C. 2019. Setting boundaries for genome-wide heterochromatic DNA deletions through flanking inverted repeats in Tetrahymena thermophila. *Nucleic Acids Res* **47**: 5181–5192.

Lin C-YG, Lin I-T, Yao M-C. 2016. Programmed Minichromosome Elimination as a Mechanism for Somatic Genome Reduction in Tetrahymena thermophila. *PLoS Genet* **12**: e1006403.

Lin I-T, Chao J-L, Yao M-C. 2012. An essential role for the DNA breakage-repair protein Ku80 in programmed DNA rearrangements in Tetrahymena thermophila. *Mol Biol Cell* **23**: 2213–2225.

Lippman Z, Martienssen R. 2004. The role of RNA interference in heterochromatic silencing. *Nature* **431**: 364–370.

Liu B, Gross M, ten Hoeve J, Shuai K. 2001. A transcriptional corepressor of Stat1 with an essential LXXLL signature motif. *Proc Natl Acad Sci USA* **98**: 3203–3207.

Liu Q, Rand TA, Kalidas S, Du F, Kim H-E, Smith DP, Wang X. 2003. R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. *Science* **301**: 1921–1925.

Liu Y, Mochizuki K, Gorovsky MA. 2004. Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena. *Proc Natl Acad Sci USA* **101**: 1679–1684.

Liu Y, Taverna SD, Muratore TL, Shabanowitz J, Hunt DF, Allis CD. 2007. RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. *Genes Dev* **21**: 1530–1545.

Lois LM, Lima CD. 2005. Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *EMBO J* **24**: 439–451.

Lund E, Dahlberg JE. 2006. Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs. *Cold Spring Harb Symp Quant Biol* **71**: 59–66.

MacRae IJ, Ma E, Zhou M, Robinson CV, Doudna JA. 2008. In vitro reconstitution of the human RISC-loading complex. *Proc Natl Acad Sci USA* **105**: 512–517.

Madireddi MT, Coyne RS, Smothers JF, Mickey KM, Yao MC, Allis CD. 1996. Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena. *Cell* **87**: 75–84.

Maison C, Almouzni G. 2004. HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol* **5**: 296–304.

Malone CD, Anderson AM, Motl JA, Rexer CH, Chalker DL. 2005. Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of *Tetrahymena thermophila*. *Mol Cell Biol* **25**: 9151–9164.

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* **137**: 522–535.

Malone CD, Hannon GJ. 2009. Molecular evolution of piRNA and transposon control pathways in Drosophila. *Cold Spring Harb Symp Quant Biol* **74**: 225–234.

Maniataki E, Mourelatos Z. 2005. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev* **19**: 2979–2990.

Martindale DW, Allis CD, Bruns PJ. 1982. Conjugation in Tetrahymena thermophila. A temporal analysis of cytological stages. *Exp Cell Res* **140**: 227–236.

Martindale DW, Allis CD, Bruns PJ. 1985. RNA and protein synthesis during meiotic prophase in Tetrahymena thermophila. *J Protozool* **32**: 644–649.

Matic I, van Hagen M, Schimmel J, Macek B, Ogg SC, Tatham MH, Hay RT, Lamond AI, Mann M, Vertegaal ACO. 2008. In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. *Mol Cell Proteomics* **7**: 132–144.

Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD. 2005. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**: 607–620.

Matsuda A, Forney JD. 2006. The SUMO pathway is developmentally regulated and required for programmed DNA elimination in Paramecium tetraurelia. *Eukaryotic Cell* **5**: 806–815.

Matuschewski K, Hauser HP, Treier M, Jentsch S. 1996. Identification of a novel family of ubiquitin-conjugating enzymes with distinct amino-terminal extensions. *J Biol Chem* **271**: 2789–2794.

Meister G, Tuschl T. 2004. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**: 343–349.

Mello CC, Conte D. 2004. Revealing the world of RNA interference. *Nature* **431**: 338–342.

Mével-Ninio M, Pelisson A, Kinder J, Campos AR, Bucheton A. 2007. The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. *Genetics* **175**: 1615–1624.

Minty A, Dumont X, Kaghad M, Caput D. 2000. Covalent modification of p73alpha by SUMO-1. Two-hybrid screening with p73 identifies novel SUMO-1-interacting proteins and a SUMO-1 interaction motif. *J Biol Chem* **275**: 36316–36323.

Miyoshi K, Tsukumo H, Nagami T, Siomi H, Siomi MC. 2005. Slicer function of Drosophila Argonautes and its involvement in RISC formation. *Genes Dev* **19**: 2837–2848.

Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA. 2002. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell* **110**: 689–699.

Mochizuki K, Gorovsky MA. 2005. A Dicer-like protein in *Tetrahymena* has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes Dev* **19**: 77–89.

Mochizuki K, Gorovsky MA. 2004a. Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev* **18**: 2068–2073.

Mochizuki K, Gorovsky MA. 2004b. RNA polymerase II localizes in Tetrahymena thermophila meiotic micronuclei when micronuclear transcription associated with genome rearrangement occurs. *Eukaryotic Cell* **3**: 1233–1240.

Mochizuki K, Gorovsky MA. 2004c. Small RNAs in genome rearrangement in Tetrahymena. *Curr Opin Genet Dev* **14**: 181–187.

Mohn F, Sienski G, Handler D, Brennecke J. 2014. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. *Cell* **157**: 1364–1379.

Motamedi MR, Verdel A, Colmenares SU, Gerber SA, Gygi SP, Moazed D. 2004. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**: 789–802.

Munarriz E, Barcaroli D, Stephanou A, Townsend PA, Maisse C, Terrinoni A, Neale MH, Martin SJ, Latchman DS, Knight RA, et al. 2004. PIAS-1 is a checkpoint regulator which affects exit from G1 and G2 by sumoylation of p73. *Mol Cell Biol* **24**: 10593–10610.

Murchison EP, Partridge JF, Tam OH, Cheloufi S, Hannon GJ. 2005. Characterization of Dicer-deficient murine embryonic stem cells. *Proc Natl Acad Sci USA* **102**: 12135–12140.

Nakayama KI, Nakayama K. 2006. Ubiquitin ligases: cell-cycle control and cancer. *Nat Rev Cancer* **6**: 369–381.

Namanja AT, Li Y-J, Su Y, Wong S, Lu J, Colson LT, Wu C, Li SSC, Chen Y. 2012. Insights into high affinity small ubiquitin-like modifier (SUMO) recognition by SUMO-interacting motifs (SIMs) revealed by a combination of NMR and peptide array analysis. *J Biol Chem* **287**: 3231–3240.

Narain A, Bhandare P, Adhikari B, Backes S, Eilers M, Dölken L, Schlosser A, Erhard F, Baluapuri A, Wolf E. 2021. Targeted protein degradation reveals a direct role of SPT6 in RNAPII elongation and termination. *Mol Cell* **81**: 3110-3127.e14.

Nasir AM, Yang Q, Chalker DL, Forney JD. 2015. SUMOylation is developmentally regulated and required for cell pairing during conjugation in Tetrahymena thermophila. *Eukaryotic Cell* **14**: 170–181.

Nayak A, Müller S. 2014. SUMO-specific proteases/isopeptidases: SENPs and beyond. *Genome Biol* **15**: 422.

Nie M, Xie Y, Loo JA, Courey AJ. 2009. Genetic and proteomic evidence for roles of Drosophila SUMO in cell cycle control, Ras signaling, and early pattern formation. *PLoS ONE* **4**: e5905.

Ninova M, Chen Y-CA, Godneeva B, Rogers AK, Luo Y, Fejes Tóth K, Aravin AA. 2020a. Su(var)2-10 and the SUMO Pathway Link piRNA-Guided Target Recognition to Chromatin Silencing. *Mol Cell* **77**: 556-570.e6.

Ninova M, Godneeva B, Chen Y-CA, Luo Y, Prakash SJ, Jankovics F, Erdélyi M, Aravin AA, Fejes Tóth K. 2020b. The SUMO Ligase Su(var)2-10 Controls Hetero- and Euchromatic Gene Expression via Establishing H3K9 Trimethylation and Negative Feedback Regulation. *Mol Cell* **77**: 571-585.e4.

Nishimasu H, Ishizu H, Saito K, Fukuhara S, Kamatani MK, Bonnefond L, Matsumoto N, Nishizawa T, Nakanaga K, Aoki J, et al. 2012. Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature* **491**: 284–287.

Noto T, Kataoka K, Suhren JH, Hayashi A, Woolcock KJ, Gorovsky MA, Mochizuki K. 2015. Small-RNA-Mediated Genome-wide trans-Recognition Network in Tetrahymena DNA Elimination. *Mol Cell* **59**: 229–242.

Noto T, Kurth HM, Kataoka K, Aronica L, DeSouza LV, Siu KWM, Pearlman RE, Gorovsky MA, Mochizuki K. 2010. The Tetrahymena argonaute-binding protein Giw1p directs a mature argonaute-siRNA complex to the nucleus. *Cell* **140**: 692–703.

Noto T, Mochizuki K. 2018. Small RNA-Mediated trans-Nuclear and trans-Element Communications in *Tetrahymena* DNA Elimination. *Curr Biol* **28**: 1938-1949.e5.

Noto T, Mochizuki K. 2017. Whats, hows and whys of programmed DNA elimination in *Tetrahymena*. *Open Biol* **7**: 170172.

Ohara T, Sakaguchi Y, Suzuki T, Ueda H, Miyauchi K, Suzuki T. 2007. The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. *Nat Struct Mol Biol* **14**: 349–350.

Ohtani H, Iwasaki YW, Shibuya A, Siomi H, Siomi MC, Saito K. 2013. DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the Drosophila ovary. *Genes Dev* **27**: 1656–1661.

Okuma T, Honda R, Ichikawa G, Tsumagari N, Yasuda H. 1999. In vitro SUMO-1 modification requires two enzymatic steps, E1 and E2. *Biochem Biophys Res Commun* **254**: 693–698.

Olina AV, Kulbachinskiy AV, Aravin AA, Esyunina DM. 2018. Argonaute proteins and mechanisms of RNA interference in eukaryotes and prokaryotes. *Biochemistry Mosc* **83**: 483–497.

Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J. 2010. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *EMBO J* **29**: 3301–3317.

Olsen SK, Lima CD. 2013. Structure of a ubiquitin E1-E2 complex: insights to E1-E2 thioester transfer. *Mol Cell* **49**: 884–896.

Onishi R, Sato K, Murano K, Negishi L, Siomi H, Siomi MC. 2020. Piwi suppresses transcription of Brahma-dependent transposons via Maelstrom in ovarian somatic cells. *Sci Adv* **6**.

Onodera Y, Haag JR, Ream T, Costa Nunes P, Pontes O, Pikaard CS. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**: 613–622.

Orban TI, Izaurralde E. 2005. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* **11**: 459–469.

Osumi K, Sato K, Murano K, Siomi H, Siomi MC. 2019. Essential roles of Windei and nuclear monoubiquitination of Eggless/SETDB1 in transposon silencing. *EMBO Rep* **20**: e48296.

Owerbach D, McKay EM, Yeh ETH, Gabbay KH, Bohren KM. 2005. A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. *Biochem Biophys Res Commun* **337**: 517–520.

Pane A, Wehr K, Schüpbach T. 2007. zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. *Dev Cell* **12**: 851–862.

Park JH, Shin S-Y, Shin C. 2017. Non-canonical targets destabilize microRNAs in human Argonautes. *Nucleic Acids Res* **45**: 1569–1583.

Pélisson A, Sarot E, Payen-Groschêne G, Bucheton A. 2007. A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the Drosophila ovary. *J Virol* **81**: 1951–1960.

Perry JJP, Tainer JA, Boddy MN. 2008. A SIM-ultaneous role for SUMO and ubiquitin. *Trends Biochem Sci* **33**: 201–208.

Pichler A, Gast A, Seeler JS, Dejean A, Melchior F. 2002. The nucleoporin RanBP2 has SUMO1 E3 ligase activity. *Cell* **108**: 109–120.

Potts PR, Yu H. 2005. Human MMS21/NSE2 is a SUMO ligase required for DNA repair. *Mol Cell Biol* **25**: 7021–7032.

Prud'homme N, Gans M, Masson M, Terzian C, Bucheton A. 1995. Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. *Genetics* **139**: 697–711.

Ramachandran V, Chen X. 2008. Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. *Science* **321**: 1490–1492.

Rand TA, Petersen S, Du F, Wang X. 2005. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123**: 621–629.

Rangasamy D, Woytek K, Khan SA, Wilson VG. 2000. SUMO-1 modification of bovine papillomavirus E1 protein is required for intranuclear accumulation. *J Biol Chem* **275**: 37999–38004.

Ray C. 1956. Preparation of chromosomes of Tetrahymena pyriformis for photomicrography. *Stain Technol* **31**: 271–274.

Reichholf B, Herzog VA, Fasching N, Manzenreither RA, Sowemimo I, Ameres SL. 2019. Time-Resolved Small RNA Sequencing Unravels the Molecular Principles of MicroRNA Homeostasis. *Mol Cell* **75**: 756-768.e7.

Rissland OS, Hong S-J, Bartel DP. 2011. MicroRNA destabilization enables dynamic regulation of the miR-16 family in response to cell-cycle changes. *Mol Cell* **43**: 993–1004.

Rivas FV, Tolia NH, Song J-J, Aragon JP, Liu J, Hannon GJ, Joshua-Tor L. 2005. Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol* **12**: 340–349.

Rozhkov NV, Hammell M, Hannon GJ. 2013. Multiple roles for Piwi in silencing Drosophila transposons. *Genes Dev* **27**: 400–412.

Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Res* **17**: 1850–1864.

Rytinki MM, Kaikkonen S, Pehkonen P, Jääskeläinen T, Palvimo JJ. 2009. PIAS proteins: pleiotropic interactors associated with SUMO. *Cell Mol Life Sci* **66**: 3029–3041.

Ryu H-Y, Hochstrasser M. 2021. Histone sumoylation and chromatin dynamics. *Nucleic Acids Res* **49**: 6043–6052.

Saini HK, Griffiths-Jones S, Enright AJ. 2007. Genomic analysis of human microRNA transcripts. *Proc Natl Acad Sci USA* **104**: 17719–17724.

Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. *Genes Dev* **24**: 2493–2498.

Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev* **20**: 2214–2222.

Saito K, Sakaguchi Y, Suzuki T, Suzuki T, Siomi H, Siomi MC. 2007. Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev* **21**: 1603–1608.

Schirle NT, Sheu-Gruttadauria J, MacRae IJ. 2014. Structural basis for microRNA targeting. *Science* **346**: 608–613.

Schoeberl UE, Kurth HM, Noto T, Mochizuki K. 2012. Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in *Tetrahymena. Genes Dev* **26**: 1729–1742.

Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N, Zamore PD. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**: 199–208.

Sekiyama N, Ikegami T, Yamane T, Ikeguchi M, Uchimura Y, Baba D, Ariyoshi M, Tochio H, Saitoh H, Shirakawa M. 2008. Structure of the small ubiquitin-like modifier (SUMO)-interacting motif of MBD1-containing chromatin-associated factor 1 bound to SUMO-3. *J Biol Chem* **283**: 35966–35975.

Shehzada S, Mochizuki K. 2022. PIWI-Directed DNA Elimination for Tetrahymena Genetics. *Methods Mol Biol* **2509**: 53–68.

Sheng Y, Duan L, Cheng T, Qiao Y, Stover NA, Gao S. 2020. The completed macronuclear genome of a model ciliate *Tetrahymena thermophila* and its application in genome scrambling and copy number analyses. *Sci China Life Sci* **63**: 1534–1542.

Sheu-Gruttadauria J, MacRae IJ. 2017. Structural foundations of RNA silencing by argonaute. *J Mol Biol* **429**: 2619–2639.

Sheu-Gruttadauria J, Pawlica P, Klum SM, Wang S, Yario TA, Schirle Oakdale NT, Steitz JA, MacRae IJ. 2019. Structural Basis for Target-Directed MicroRNA Degradation. *Mol Cell* **75**: 1243-1255.e7.

Shin JA, Choi ES, Kim HS, Ho JCY, Watts FZ, Park SD, Jang YK. 2005. SUMO modification is involved in the maintenance of heterochromatin stability in fission yeast. *Mol Cell* **19**: 817–828.

Shi CY, Kingston ER, Kleaveland B, Lin DH, Stubna MW, Bartel DP. 2020. The ZSWIM8 ubiquitin ligase mediates target-directed microRNA degradation. *Science* **370**.

Sienski G, Batki J, Senti K-A, Dönertas D, Tirian L, Meixner K, Brennecke J. 2015. Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. *Genes Dev* **29**: 2258–2271.

Simon B, Kirkpatrick JP, Eckhardt S, Reuter M, Rocha EA, Andrade-Navarro MA, Sehr P, Pillai RS, Carlomagno T. 2011. Recognition of 2'-O-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein. *Structure* **19**: 172–180.

Siomi MC, Sato K, Pezic D, Aravin AA. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**: 246–258.

Smith M, Turki-Judeh W, Courey AJ. 2012. Sumoylation in drosophila development. *Biomolecules* **2**: 331–349.

Song J, Durrin LK, Wilkinson TA, Krontiris TG, Chen Y. 2004. Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. *Proc Natl Acad Sci USA* **101**: 14373–14378.

Song J, Zhang Z, Hu W, Chen Y. 2005. Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. *J Biol Chem* **280**: 40122–40129.

Steele CJ, Barkocy-Gallagher GA, Preer LB, Preer JR. 1994. Developmentally excised sequences in micronuclear DNA of Paramecium. *Proc Natl Acad Sci USA* **91**: 2255–2259.

Sugai T, Hiwatashi K. 1974. Cytologic and autoradiographic studies of the micronucleus at meiotic prophase in Tetrahymena pyriformis. *J Protozool* **21**: 542–548.

Sugiyama T, Cam H, Verdel A, Moazed D, Grewal SIS. 2005. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc Natl Acad Sci USA* **102**: 152–157.

Takahashi Y, Kahyo T, Toh-E A, Yasuda H, Kikuchi Y. 2001a. Yeast Ull1/Siz1 is a novel SUMO1/Smt3 ligase for septin components and functions as an adaptor between conjugating enzyme and substrates. *J Biol Chem* **276**: 48973–48977.

Takahashi Y, Toh-e A, Kikuchi Y. 2001b. A novel factor required for the SUMO1/Smt3 conjugation of yeast septins. *Gene* **275**: 223–231.

Tamm SL, Sonneborn TM, Dippell RV. 1975. The role of cortical orientation in the control of the direction of ciliary beat in Paramecium. *J Cell Biol* **64**: 98–112.

Tan J-A, Hall SH, Hamil KG, Grossman G, Petrusz P, French FS. 2002. Protein inhibitors of activated STAT resemble scaffold attachment factors and function as interacting nuclear receptor coregulators. *J Biol Chem* **277**: 16993–17001.

Tatham MH, Geoffroy M-C, Shen L, Plechanovova A, Hattersley N, Jaffray EG, Palvimo JJ, Hay RT. 2008. RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat Cell Biol* **10**: 538–546.

Tatham MH, Jaffray E, Vaughan OA, Desterro JM, Botting CH, Naismith JH, Hay RT. 2001. Polymeric chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9. *J Biol Chem* **276**: 35368–35374.

Thivierge C, Makil N, Flamand M, Vasale JJ, Mello CC, Wohlschlegel J, Conte D, Duchaine TF. 2011. Tudor domain ERI-5 tethers an RNA-dependent RNA polymerase to DCR-1 to potentiate endo-RNAi. *Nat Struct Mol Biol* **19**: 90–97.

Tian Y, Simanshu DK, Ma J-B, Patel DJ. 2011. Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. *Proc Natl Acad Sci USA* **108**: 903–910.

Tomari Y, Zamore PD. 2005. Perspective: machines for RNAi. *Genes Dev* **19**: 517–529.

Tong H, Hateboer G, Perrakis A, Bernards R, Sixma TK. 1997. Crystal structure of murine/human Ubc9 provides insight into the variability of the ubiquitin-conjugating system. *J Biol Chem* **272**: 21381–21387.

Trujillo JT, Mosher RA. 2017. Identification and Evolutionary Characterization of ARGONAUTE-Binding Platforms. *Methods Mol Biol* **1640**: 257–266.

Ustianenko D, Hrossova D, Potesil D, Chalupnikova K, Hrazdilova K, Pachernik J, Cetkovska K, Uldrijan S, Zdrahal Z, Vanacova S. 2013. Mammalian DIS3L2 exoribonuclease targets the uridylated precursors of let-7 miRNAs. *RNA* **19**: 1632–1638.

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**: 320–324.

Vertegaal ACO. 2007. Small ubiquitin-related modifiers in chains. *Biochem Soc Trans* **35**: 1422–1423.

Vogt A, Mochizuki K. 2013. A domesticated PiggyBac transposase interacts with heterochromatin and catalyzes reproducible DNA elimination in *Tetrahymena*. *PLoS Genet* **9**: e1004032.

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**: 1833–1837.

Wang G, Reinke V. 2008. A C. elegans Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Curr Biol* **18**: 861–867.

Wang J, Taherbhoy AM, Hunt HW, Seyedin SN, Miller DW, Miller DJ, Huang DT, Schulman BA. 2010. Crystal structure of UBA2(ufd)-Ubc9: insights into E1-E2 interactions in Sumo pathways. *PLoS ONE* **5**: e15805.

Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ. 2008. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**: 209–213.

Watanabe T, Cui X, Yuan Z, Qi H, Lin H. 2018. MIWI2 targets RNAs transcribed from piRNA-dependent regions to drive DNA methylation in mouse prospermatogonia. *EMBO J* **37**.

Webster A, Li S, Hur JK, Wachsmuth M, Bois JS, Perkins EM, Patel DJ, Aravin AA. 2015. Aub and Ago3 Are Recruited to Nuage through Two Mechanisms to Form a Ping-Pong Complex Assembled by Krimper. *Mol Cell* **59**: 564–575.

Weger S, Hammer E, Heilbronn R. 2005. Topors acts as a SUMO-1 E3 ligase for p53 in vitro and in vivo. *FEBS Lett* **579**: 5007–5012.

Wierzbicki AT, Haag JR, Pikaard CS. 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**: 635–648.

Wightman B, Ha I, Ruvkun G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**: 855–862.

Wilson VG. 2017. Introduction to Sumoylation. *Adv Exp Med Biol* **963**: 1–12.

Winter J, Diederichs S. 2011. Argonaute proteins regulate microRNA stability: Increased microRNA abundance by Argonaute proteins is due to microRNA stabilization. *RNA Biol* **8**: 1149–1157.

Woehrer SL, Aronica L, Suhren JH, Busch CJ-L, Noto T, Mochizuki K. 2015. A Tetrahymena Hsp90 co-chaperone promotes siRNA loading by ATP-dependent and ATP-independent mechanisms. *EMBO J* **34**: 559–577.

Xie J, Ameres SL, Friedline R, Hung J-H, Zhang Y, Xie Q, Zhong L, Su Q, He R, Li M, et al. 2012. Long-term, efficient inhibition of microRNA function in mice using rAAV vectors. *Nat Methods* **9**: 403–409.

Xu J, Zhao X, Mao F, Basrur V, Ueberheide B, Chait BT, Allis CD, Taverna SD, Gao S, Wang W, et al. 2021. A Polycomb repressive complex is required for RNAi-mediated heterochromatin formation and dynamic distribution of nuclear bodies. *Nucleic Acids Res* **49**: 5407–5425.

Yamashiro H, Negishi M, Kinoshita T, Ishizu H, Ohtani H, Siomi MC. 2020. Armitage determines Piwi-piRISC processing from precursor formation and quality control to inter-organelle translocation. *EMBO Rep* **21**: e48769.

Yamashiro H, Siomi MC. 2018. PIWI-Interacting RNA in Drosophila: Biogenesis, Transposon Regulation, and Beyond. *Chem Rev* **118**: 4404–4421.

Yang M, Hsu C-T, Ting C-Y, Liu LF, Hwang J. 2006a. Assembly of a polymeric chain of SUMO1 on human topoisomerase I in vitro. *J Biol Chem* **281**: 8264–8274.

Yang Q, Nasir AM, Coyne RS, Forney JD. 2015. Depletion of UBC9 Causes Nuclear Defects during the Vegetative and Sexual Life Cycles in Tetrahymena thermophila. *Eukaryotic Cell* **14**: 1240–1252.

Yang Z, Ebright YW, Yu B, Chen X. 2006b. HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res* **34**: 667–675.

Yunus AA, Lima CD. 2009. Structure of the Siz/PIAS SUMO E3 ligase Siz1 and determinants required for SUMO modification of PCNA. *Mol Cell* **35**: 669–682.

Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X. 2005. Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**: 932–935.

Yu Y, Gu J, Jin Y, Luo Y, Preall JB, Ma J, Czech B, Hannon GJ. 2015. Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science* **350**: 339–342.

Yu Y, Ji L, Le BH, Zhai J, Chen J, Luscher E, Gao L, Liu C, Cao X, Mo B, et al. 2017. ARGONAUTE10 promotes the degradation of miR165/6 through the SDN1 and SDN2 exonucleases in Arabidopsis. *PLoS Biol* **15**: e2001272.

Zamore PD, Tuschl T, Sharp PA, Bartel DP. 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**: 25–33.

Zhang F, Wang J, Xu J, Zhang Z, Koppetsch BS, Schultz N, Vreven T, Meignin C, Davis I, Zamore PD, et al. 2012. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* **151**: 871–884.

Zhang K, Mosch K, Fischle W, Grewal SIS. 2008. Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nat Struct Mol Biol* **15**: 381–388.

Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, Vreven T, Zamore PD, Weng Z, Theurkauf WE. 2014. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* **157**: 1353–1363.

Zhou W, Ryan JJ, Zhou H. 2004. Global analyses of sumoylated proteins in Saccharomyces cerevisiae. Induction of protein sumoylation by cellular stresses. *J Biol Chem* **279**: 32262–32268.

Zhou Y, Fu L, Mochizuki K, Xiong J, Miao W, Wang G. 2022. Absolute quantification of chromosome copy numbers in the polyploid macronucleus of Tetrahymena thermophila at the single-cell level. *J Eukaryot Microbiol* **69**: e12907.

Zilio N, Eifler-Olivi K, Ulrich HD. 2017. Functions of SUMO in the maintenance of genome stability. *Adv Exp Med Biol* **963**: 51–87.