



universität
wien

DISSERTATION

Titel der Dissertation

„Heterochromatin components are required for limiting biogenesis of siRNAs to sequences targeted for DNA elimination in *Tetrahymena*“

Verfasser

Jan Henrik Suhren

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2015

Studienkennzahl lt. Studienblatt:

A 794 685 490

Dissertationsgebiet lt. Studienblatt:

Molekulare Biologie, DK: RNA Biology

Betreuerin / Betreuer:

Dr. Kazufumi Mochizuki

Table of contents

Table of contents	1
Index of Figures	3
Abstract	5
Zusammenfassung	7
1. Introduction	9
1.1 Heterochromatin	9
1.2 RNA Interference	10
1.3 Small RNA Biogenesis	11
1.4 Silencing by RNAi	12
1.5 Nuclear RNAi Pathways and heterochromatin	13
1.5.1 Nuclear RNAi in <i>S. pombe</i>	14
1.5.2 RNA-directed DNA methylation in <i>Arabidopsis thaliana</i>	15
1.5.3 Nuclear RNAi in <i>C. elegans</i>	17
1.5.4 Nuclear piRNAs in genome defense	17
1.6 Boundary elements regulate heterochromatin spreading	19
1.7 The model organism <i>Tetrahymena thermophila</i>	22
1.8 Nuclear dimorphism and genome rearrangements in <i>Tetrahymena thermophila</i>	22
1.9 Small RNA pathways in <i>Tetrahymena</i>	25
1.10 Small RNA-directed DNA elimination	25
1.10.1 Two types of IESs	25
1.10.2 Biogenesis and scanning of Early-scnRNAs	27
1.10.3 Late-scnRNA biogenesis is heterochromatin dependent	27
2. Results	31
2.1 Coi6p is an HP1-like protein	31
2.2 Coi6p is specifically expressed during conjugation	32
2.3 Coi6p localizes to heterochromatin foci in the new macronucleus	34
2.4 Coi6p specifically binds to IESs in vivo	36
2.5 Coi6p is required for DNA elimination	38
2.6 Chromo- and chromoshadow domain are essential for Coi6p function	39
2.7 Heterochromatin components accumulate in absence of Coi6p	42
2.8 Pdd1p and H3K9/K27me3 spread outside IES boundaries in $\Delta COI6$ cells	43
2.9 Heterochromatin foci formation depends on Coi6p	47
2.10 Coi6p binding to IESs depends on scnRNAs, yet not on Pdd1p	48
2.11 Late-scnRNA production occurs outside IES boundaries in $\Delta COI6$ cells	51
2.12 Coi6p interacts with Coi7p and Lia5p	56
2.13 Conservation of Coi7p and Lia5p and co-expression with Coi6p	57
2.14 Coi7p and Lia5p are enriched on IESs	58
2.15 Generation of $\Delta COI7$ strains using CRISPR/Cas9	59
2.16 Coi7p and Lia5p are necessary for Tlr1 elimination	61
2.17 Coi6p is unstable in absence of Coi7p	62
2.18 Coi7p and Lia5p are important for precise Late-scnRNA production	63
2.19 The histone H3K27 demethylase Jmj1p is involved in maintaining precision of the Late-scnRNA production	67

2.20	Imprecise production of Late-scnRNAs causes abnormal elimination of IESs along with MDS regions	71
3.	Discussion	75
3.1	The HP1-like protein Coi6p is a new heterochromatin component in <i>Tetrahymena</i>	75
3.2	Coi7p and Lia5p are Coi6p interaction partners	76
3.3	Potential functions of Coi7p and Lia5p	76
3.4	A crucial role for Coi6p and its interaction partners in maintaining IES boundaries	77
3.5	Jmj1p restricts heterochromatin and Late-scnRNA production to IESs	78
3.6	RNAi-heterochromatin spreading-induced overelimination impacts genome integrity	79
3.7	A model for IES boundary maintenance	80
3.8	Potential primary borders for Late-scnRNA production?	82
4.	Materials and Methods	84
4.1	Growth and mating conditions for <i>Tetrahymena</i>	84
4.2	Generation of transgenic strains	84
4.3	Genomic DNA extraction	84
4.4	Polymerase Chain Reaction (PCR)	85
4.5	Overelimination PCR assay	85
4.6	TOPO TA cloning of overelimination PCR products	85
4.7	Site-directed mutagenesis	86
4.8	COI6 knock-out strains	86
4.9	CRISPR/Cas9-mediated knock-out of COI7	86
4.10	COI6 rescue strain generation	87
4.11	SDS-PAGE and Western blotting analysis	87
4.12	DMP cross-linking of antibodies to protein A dynabeads	88
4.13	Cell lysate preparation for immunoprecipitation	89
4.14	Coi6p immunoprecipitation for mass spectrometry	89
4.15	NanoLC-MS Analysis.	90
4.16	MS/MS Data Analysis	90
4.17	Small scale Coi6p, Coi7 and Lia5p immunoprecipitation	91
4.18	scnRNA analysis	91
4.19	Immunofluorescence	92
4.20	Chromatin Spreading	93
4.21	DNA fluorescence in situ hybridization	93
4.22	ChIP-sequencing	94
4.23	List of antibodies	96
4.24	List of oligonucleotides	97
	Abbreviations	99
	References	101
	Curriculum vitae	113
	Lebenslauf	115
	Acknowledgments	117

Index of Figures

Figure 1:	HP1 is an adaptor that recruits diverse activities to heterochromatin	10
Figure 2:	A comparison of RNAi pathways targeting chromatin across different organisms.....	13
Figure 3:	The nuclear RNAi and heterochromatin machinery in <i>S. pombe</i> creates a self-reinforcing positive feedback loop.....	15
Figure 4:	The RNA-directed DNA methylation (RdDM) pathway in <i>A. thaliana</i>	16
Figure 5:	piRNA pathways lead to TGS in several different organisms.....	18
Figure 6:	Heterochromatin boundaries can be established by diverse mechanisms	21
Figure 7:	Nuclear dimorphism and life cycle in <i>Tetrahymena</i>	23
Figure 8:	Outline of the lifecycle of <i>Tetrahymena thermophila</i>	23
Figure 9:	DNA rearrangements during new MAC development in <i>Tetrahymena</i>	24
Figure 10:	Early- and Late-scnRNAs target IESs for elimination.....	26
Figure 11:	Molecular details of the DNA elimination pathway	28
Figure 12:	Coi6p is a HP1-like protein with conserved functional residues.....	31
Figure 13:	Verification of α -Coi6p antibodies specificity	33
Figure 14:	Conjugation-specific expression of Coi6p	34
Figure 15:	Coi6p localizes to the new MAC and to heterochromatin foci.....	35
Figure 16:	Coi6p binds to Type-A and Type-B IESs.....	37
Figure 17:	Verification of the genotype of Δ COI6 strains.....	38
Figure 18:	Elimination of the Tlr1 IESs is inhibited in absence of Coi6p.....	39
Figure 19:	Schematic representation of the Δ COI6 rescue system.....	40
Figure 20:	Expression and localization of HA-Coi6p in Δ COI6 rescue system strains	41
Figure 21:	Tlr1 elimination in the Δ COI6 rescue system	42
Figure 22:	Lysine-methylations of histone H3K9 and K27 are established in absence of Coi6p	43
Figure 23:	Heterochromatin spreading in absence of Coi6p	44
Figure 24:	Meta-analysis of heterochromatin spreading in Δ COI6 cells.....	46
Figure 25:	Block of Pdd1p heterochromatin foci formation in Δ COI6 cells.....	47
Figure 26:	Coi6p IES binding depends on Twi1p and scnRNAs, but not on Pdd1p	48
Figure 27:	Meta-analysis - Coi6p IES binding depends on Twi1p and scnRNAs, but not on Pdd1p	50
Figure 28:	Loci unaffected and affected by scnRNA spreading in Δ COI6 cells	52
Figure 29:	Meta-analysis of WT and in Δ COI6 scnRNA-seq time courses.....	53
Figure 30:	Length and base composition of IES and MDS-mapping scnRNAs in WT and in Δ COI6 cells	55
Figure 31:	Interaction and co-expression of Coi6p, Coi7p and Lia5p.....	56
Figure 32:	Coi7p and Lia5p are enriched on IESs.....	59
Figure 33:	Generation of Δ COI7 cells using CRISPR/Cas9 technology	60
Figure 34:	Stable accumulation of Coi6p depends on Coi7p.....	62
Figure 35:	Meta-analysis of 10.5 hpm scnRNAs in WT, Δ COI6, Δ COI7 and Δ LIA5 and Δ JMJ1 cells	64
Figure 36:	Heat map analysis of 10.5 hpm scnRNAs in WT, Δ COI6, Δ COI7 and Δ LIA5 and Δ JMJ1 cells revealing genome-wide scnRNAs spreading in all KOs	67
Figure 37:	ChIP-seq analysis of heterochromatin spreading in Δ JMJ1 cells.....	68
Figure 38:	ChIP-seq meta-analysis of WT and Δ JMJ1 cells	70
Figure 39:	Spreading of Late-scnRNAs leads to elimination of MDS regions and genes.....	72
Figure 40:	Sanger-Sequencing verifies elimination of MDS regions in Δ COI6 cells	73
Figure 41:	Two models for Jmj1p activity during DNA elimination.....	79
Figure 42:	Model for initiation and IES-confinement of Late-scnRNAs and heterochromatin ...	81

Abstract

Heterochromatin is generally defined as a condensed, repressive chromatin state. Due to the self-propagating nature of heterochromatin, eukaryotic genomes must maintain heterochromatin boundaries to safeguard genome functions. This is particularly critical in the context of a process termed programmed DNA elimination, by the protist *Tetrahymena thermophila* excises thousands of transposon-related sequences from its somatic genome. The excised sequences are called internal eliminated sequences (IESs). De novo heterochromatin formation is an essential intermediate step of DNA elimination and is targeted to a large subset of IESs by a population of small RNAs, called Early-scnRNAs. The remainder of IESs requires a small RNA amplification loop that yields Late-scnRNAs for their removal. During the latter process heterochromatin probably spreads in a Late-scnRNA dependent manner and further drives production of Late-scnRNAs. Yet, how such a self-reinforcing positive feedback loop is stopped and limited to IESs in order to prevent elimination of non-IES sequences has remained obscure.

In this study I identified four proteins important for confining Late-scnRNAs to IESs: the HP1-like protein Coi6p, two Coi6p interaction partners Coi7p and Lia5p, and the histone H3 lysine 27 demethylase Jmj1p. Coi6p, Coi7p and Lia5p are heterochromatin components localizing to heterochromatic foci. Furthermore, ChIP-sequencing shows Coi6p enrichment on IESs. Knock-out (KO) of *COI6*, *COI7* or *LIA5* largely blocks elimination of IESs, and absence of Coi6p leads to spreading of core heterochromatin components H3K9/K27me3 and Pdd1p outside IESs. These results suggest that Coi6p, Coi7p and Lia5p are necessary for the precise establishment of heterochromatin on IESs. Consistent with this, sequencing of scnRNAs in *COI6*, *COI7* and *LIA5* KO cells reveals that, in the absence of either of these proteins, Late-scnRNA amplification spreads outside IESs' boundaries. Similarly, scnRNA-sequencing indicates that Jmj1p also assists in restraining Late-scnRNAs within the IESs boundaries, however this only leads to a minor block of IESs elimination. Moreover, in $\Delta COI6$, $\Delta COI7$ and $\Delta JMJ1$ cells spreading of scnRNAs causes aberrant DNA elimination. Collectively, this study demonstrates for the first time that heterochromatin components are important not only for the production of Late-scnRNAs, but are also essential for the IESs-specificity of the Late-scnRNA pool and thereby for genome integrity by preventing mistargeting of DNA elimination.

Zusammenfassung

Heterochromatin wird allgemein als repressiver Chromatinstatus definiert. Da Heterochromatin seine eigene Ausbreitung verstärkt, müssen eukaryotische Genome diese begrenzen, um die Funktionsfähigkeit des Genoms zu gewährleisten. Dies ist besonders im Kontext der programmierten DNA Eliminierung wichtig, welche der Protist *Tetrahymena thermophila* nutzt, um tausende von transposonverwandten Sequenzen aus seinem somatischen Genom herauszuschneiden. Diese Sequenzen werden interne eliminierte Sequenzen (IESs) genannt. *De novo* Heterochromatin Etablierung ist in diesem Prozess ein essentieller Zwischenschritt, der für die Mehrzahl von IESs durch eine Population von kleinen RNAs, die sich Frühe-scnRNAs nennen, eingeleitet wird. Die übrigen IESs benötigen Späte-scnRNAs für ihre Eliminierung, welche durch eine Amplifikationsschleife produziert werden, die vermutlich auf Zyklen sich gegenseitig induzierender und verstärkender Heterochromatin Entstehung und Später-scnRNA Produktion basiert. Wie ein derartiger positiver Rückkopplungsmechanismus auf IESs begrenzt werden kann, damit ausschließlich diese eliminiert werden, ist bis heute unklar. In meiner Doktorarbeit habe ich vier Proteine identifiziert, die dafür wichtig sind, um die Späte-scnRNA Produktion auf IESs zu beschränken: das HP1-verwandte Protein Coi6p, seine zwei Interaktionspartner Coi7p und Lia5p sowie die Histone H3 Lysine 27 (H3K27) Demethylase Jmj1p. Coi6p, Coi7p und Lia5p sind Heterochromatinkomponenten. In Knockouts (KOs) von *COI6*, *COI7* und *LIA5* findet IESs Elimination größtenteils nicht statt, die Akkumulation von Kernheterochromatinkomponenten wie Pdd1p wird jedoch nicht blockiert. Dies legt nahe, dass Coi6p, Coi7p und Lia5p für die anfängliche Etablierung des Heterochromatin nicht notwendig sind, sondern in einem anderen Prozess aktiv sind. In Übereinstimmung damit konnte ich mit Hilfe der Sequenzierung von scnRNAs nachweisen, dass sich in *COI6*, *COI7* und *LIA5* KO-Zellen die Produktion von Späten-scnRNAs über die Grenzen von IESs hinaus ausbreitet. Im Gegensatz zu Coi6p, Coi7p und Lia5p findet sich in *JMJ1* KO-Zellen lediglich ein schwacher Defekt bezüglich der Eliminierung von IESs. Trotzdem ist Jmj1p für die Begrenzung der Späten-scnRNA Produktion auf IESs wichtig, wie ich durch Analyse der scnRNAs in *JMJ1* KO-Zellen gezeigt habe. Darüber hinaus verursacht die unpräzise Produktion von Späten-scnRNAs in *COI6*, *COI7* und *JMJ1* KO-Zellen fehlerhafte DNA Eliminierung. Zusammenfassend kann ich in der vorliegenden Arbeit zum ersten Mal zeigen, dass Heterochromatin nicht

ZUSAMMENFASSUNG

nur für die Produktion von Späten-scnRNAs wichtig ist, sondern auch für ihre IESs-Spezifität und damit für die Genauigkeit der DNA Eliminierung.

1. Introduction

1.1 Heterochromatin

Every eukaryotic cell organizes its genome into chromatin by wrapping 146 base pairs (bp) of DNA around an octamer of histone proteins and then modifying this basic form of chromatin in ways that confer differential functions to discrete domains of the genome (Luger et al. 1997). Repressive functions are generally exerted by heterochromatin, which was first described by Heitz in 1928 by virtue of its differential staining pattern and it is generally viewed as a hypoacetylated, compacted and less accessible form of chromatin (Heitz 1928). Heterochromatin comes in many different flavors, such as constitutive heterochromatin found at centromeres and telomeres and facultative heterochromatin detected at developmentally regulated loci and the mammalian inactive-X chromosome. (reviewed in Saksouk et al. 2015; Grossniklaus and Paro 2014; Heard and Martienssen 2014).

One of the most important principles governing heterochromatin biology is that heterochromatin is often defined and propagated by protein machineries that form self-reinforcing positive feedback loops (Holoch and Moazed 2015). Furthermore, defective control of such feedback loops can cause heterochromatin to spread and invade neighboring euchromatic regions (Wang et al. 2014).

One of these feedback loops is formed by the heterochromatin protein 1 (HP1) class of proteins and SU(VAR)3-9/Clr4 type histone lysine methyltransferases (HKMTs). HP1 proteins consist of a chromodomain connected by a flexible hinge to a chromoshadow domain. The chromodomain confers the ability to bind to methylated histone H3 lysine 9 (H3K9me) (or/and in some organisms to H3K27me), whereas the chromoshadow domain allows dimerization of HP1 proteins (Jacobs and Khorasanizadeh 2002; Nielsen et al. 2002; Kaustov et al. 2011). SU(VAR)3-9 type HKMTs di- or trimethylate histone H3 at lysine 9 (H3K9me_{2/3}) (Lachner et al. 2001). In addition to HP1 SU(VAR)3-9 also contains a chromodomain, which binds to H3K9me_{2/3}, and HP1 directly binds to SU(VAR)3-9 (Schotta et al. 2002). Due to these self-reinforcing interactions, a domain of H3K9me_{2/3}, HP1 and SU(VAR)3-9 can be maintained throughout the cell-cycle and in certain conditions can spread into neighboring genomic regions by successive rounds of H3K9 methylation and HP1/SU(VAR)3-9 binding at adjacent nucleosomes. Moreover, the capacity of HP1 proteins to oligomerize along the chromatin fiber makes them ideal

adaptors for recruiting diverse activities to heterochromatin (Fig. 1) (reviewed in Eissenberg and Elgin 2014). Although, HP1 proteins are best known for their role in gene silencing, they can also be important for active gene expression (reviewed in Kwon and Workman 2011).

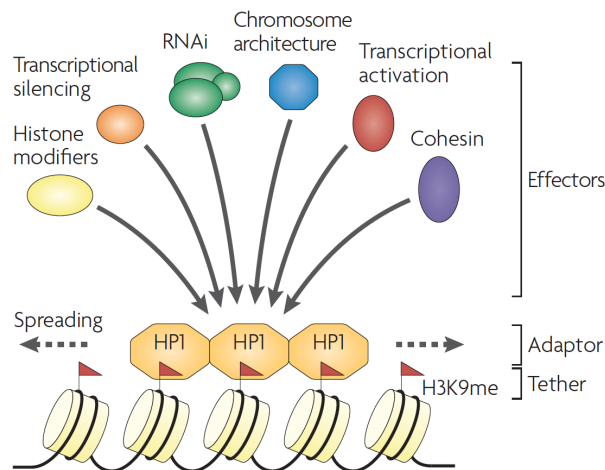


Figure 1: HP1 is an adaptor that recruits diverse activities to heterochromatin

By binding H3K9me as tether HP1 proteins have affinity to modified chromatin and act as adaptors to recruit further activities: histone modifiers, transcriptional silencing machineries, such as RNAi and histone deacetylases and factors important for chromosome segregation, e.g. Cohesin. In some cases transcriptional activation can also depend on heterochromatin. Adapted from Grewal and Jia 2007.

Another conceptually-related system for facultative heterochromatin formation, particularly at developmentally regulated loci, is the Polycomb system. The Polycomb repressive complex 2 (PRC2) contains the HKMT Enhancer of zeste (E(Z)), which methylates histone H3 K27 (H3K27me). The PRC2 subunit EED can bind to this histone modification and this further stimulates methylation of nearby histones (Margueron et al. 2009). Furthermore, the subunit Polycomb of the Polycomb repressive complex 1 (PRC1) contains a chromodomain, which also binds H3K27me. PRC1 harbors further enzymatic activities, which can silence transcription, once they are recruited to chromatin. Therefore, the Polycomb system also forms a self-reinforcing positive feedback loop (reviewed in Grossniklaus and Paro 2014).

In section 1.5 self-reinforcing positive feedback loops in the context of RNAi-induced heterochromatin will be discussed.

1.2 RNA Interference

The term RNA interference (RNAi) in its present definition refers broadly to gene silencing induced by a complex of a small RNA and an Argonaute (Ago) protein. Depending on its

mode of silencing this complex is termed RISC (for RNA-induced silencing complex) or RITS (for RNA-induced transcriptional silencing).

The breakthrough for small RNA research came in the late 1990s and was facilitated by seminal discoveries in the nematode *Caenorhabditis elegans* (*C. elegans*) and in plants. In 1998 Fire et al. reported efficient gene inhibition by dsRNA in *C. elegans* and in 1999 Hamilton et al. connected the phenomena of transgene and virus-induced post-transcriptional gene silencing (PTGS) to the appearance of small RNAs in tomato plants (Fire et al. 1998; Hamilton and Baulcombe 1999). Since then, there has not only been an explosion in basic knowledge about small RNAs, but RNAi has also been broadly applied as a tool in cell biology.

Given the versatility and utility of RNAi, evolution has created a large variety of RNAi pathways, however all of them converge on the common theme of silencing by a RISC or RITS complex, which occurs either as PTGS in the cytoplasm on messenger RNAs (mRNAs) or as transcriptional gene silencing (TGS) in the nucleus on chromatin. Specificity of silencing is provided by single stranded (ss) small RNAs via base-pairing to target RNAs with perfect or relaxed complementarity.

Argonaute proteins are conserved in all domains of life from bacteria to humans. Eukaryotic Argonautes can be mainly classified into two families by phylogenetic analysis: the AGO and PIWI families. In addition, there are the worm-specific WAGO-family and the trypanosoma Ago family (Swarts et al. 2014). Many Argonaute proteins not only hold the small RNA in the right orientation, but also possess slicer-endoribonuclease activity, enabling them to cut either the passenger strand of a small RNA duplex or a target RNA (Wilson and Doudna 2013).

1.3 Small RNA Biogenesis

A major distinction between different RNAi pathways is how biogenesis of the respective small RNA is achieved. One way to classify small RNA biogenesis is by analyzing which enzymes are required.

Small interfering RNA (siRNA) and microRNA (miRNA) rely on the RNase III ribonuclease Dicer to be cleaved from double stranded (ds) precursor RNAs. The dsRNA precursors can be formed by annealing of complementary transcriptional products, generated for example by bidirectional transcription, or by unidirectional transcription through inverted repeats and subsequent hairpin formation. The origin of dsRNA precursors of siRNAs can be endogenous, but also exogenous, e.g. derived of viruses or experimental manipulation

(Okamura and Lai 2008). miRNAs require additional processing, for instance by Drosha, also a RNase III ribonuclease (reviewed in Ha and Kim 2014). Another enzyme class frequently used in RNAi pathways is RNA-dependent RNA Polymerases (RdRPs) (Martienssen and Moazed 2015). Often directed by the RNAi response itself, this enzyme allows generation of dsRNA from single stranded (ss) RNA, and thus production of more dsRNA substrate for Dicer, thereby amplifying the RNAi response in a positive feedback loop. RdRPs are present in many organisms including fungi, plants and *C. elegans*, however vertebrates and insects do not contain these enzymes. In contrast to such Dicer-dependent biogenesis pathways, the animal germline specific piRNAs (PIWI-interacting RNAs), do not require Dicer, but are made of single-stranded RNA polymerase II (Pol II) transcripts in a fashion relying on an endonuclease (e.g. Zucchini in *D. melanogaster*) and/or the slicing activity of the Argonaute proteins themselves (reviewed in Iwasaki et al. 2015).

The end result of all biogenesis machineries is Argonaute proteins loaded with a single stranded small RNA.

1.4 Silencing by RNAi

Several classes of small RNAs act by PTGS. These are siRNAs, miRNAs and, in part, piRNAs. Typically, PTGS takes place in the cytoplasm and can employ two different modes of silencing: 1. Target cleavage, 2. Translational inhibition.

Target cleavage occurs when targets have perfect complementarity to the guiding small RNA and is carried out by the slicing activity of the Argonaute protein. The resulting RNA fragments are then quickly degraded. This mode of silencing is used by piRNAs against transposon transcripts (reviewed in Malone and Hannon 2009), by siRNAs to fight viruses and transposons (reviewed in Okamura and Lai 2008) and by plant miRNAs to regulate gene expression (reviewed in Jones-Rhoades et al. 2006).

The second mode of PTGS is translational repression, and it is mostly miRNAs that exert their silencing function via this route. Translational repression is imposed on mRNAs by RISC if small RNA and the target have low complementarity. This results in the recruitment of additional proteins to the mRNA, which negatively manipulate translation rates or destabilize the mRNA by decapping or deadenylation (reviewed in Fabian et al. 2010).

TGS is the second important kind of small RNA induced silencing and will be extensively discussed in the next sections.

1.5 Nuclear RNAi Pathways and heterochromatin

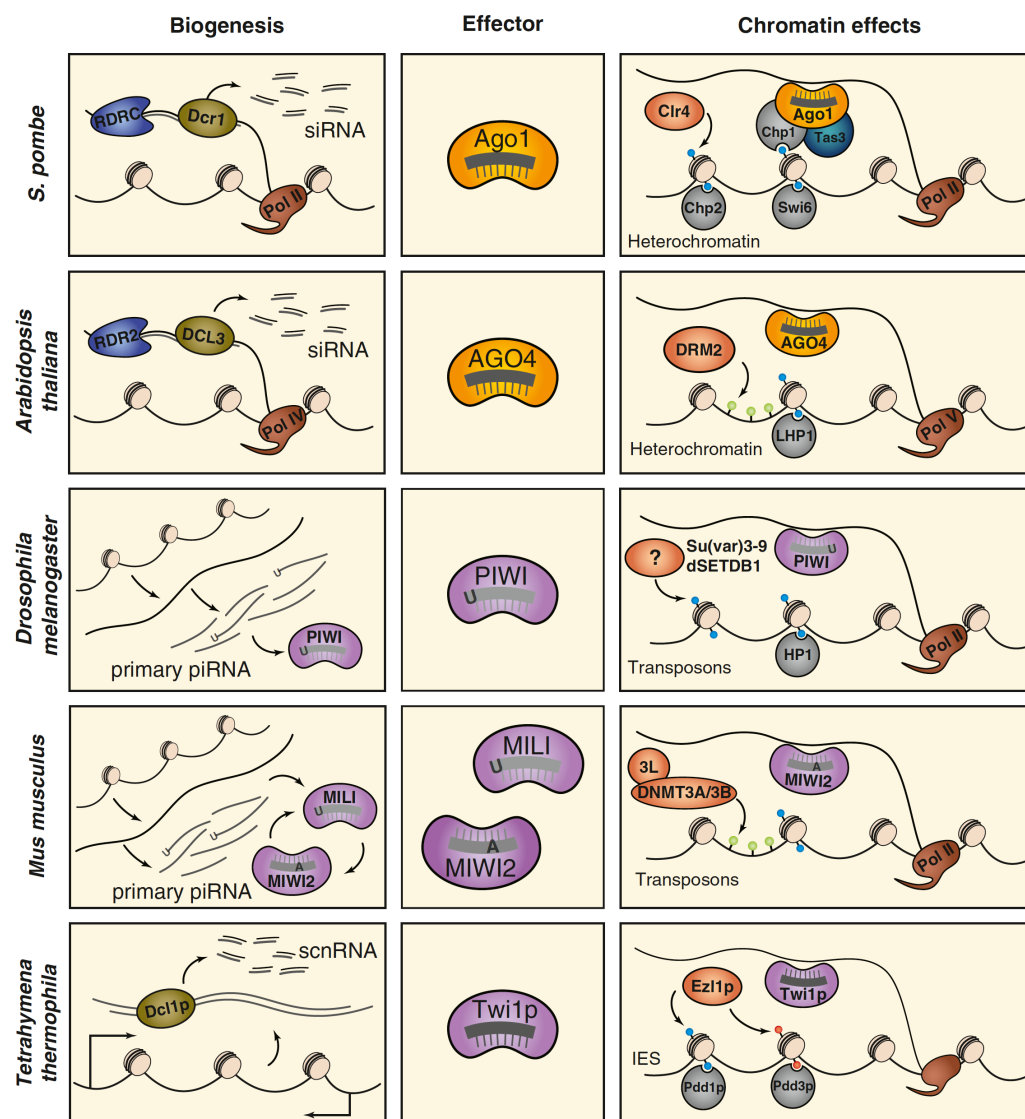


Figure 2: A comparison of RNAi pathways targeting chromatin across different organisms
 Biogenesis of small RNAs is initiated by the formation of single or double stranded RNA precursors before their transcription from the genome. Precursor RNA is processed into small RNA by Dicer-dependent or independent mechanisms. Small RNAs are incorporated into Argonaute proteins forming the effector RITS complex. This complex silences target loci (or leads to their elimination from the genome in case of *T. thermophila*). For the purpose of comparison orthologous proteins or protein modifications are indicated by the same color: Dicer (green), RNA-dependent RNA polymerase (blue), DNA-dependent RNA polymerase (brown), Ago-clade Argonaute (yellow), PIWI-clade Argonaute (purple), histone and DNA methyltransferases (orange), chromodomain-containing proteins (gray), H3K9me (blue), H3K27me (orange), DNA methylation (green). Organisms are indicated on the left side and targets of the RNAi machinery in the right panel. Specific proteins are described in the text. Adapted from Sabin et al. 2013.

RNAi and heterochromatin pathways converge in the nucleus to achieve control of repetitive and transposable elements and genome stability. In general, a complex of small RNA and Argonaute protein, termed RITS, induces or assists in maintaining transcriptional gene silencing (TGS) on genomic loci by generating a repressive and compact chromatin environment. Conceptual similarities and differences between such pathways in different organisms are highlighted in Fig. 2.

Targeting of chromatin is achieved by the base pairing of the RITS associated small RNA to nascent transcripts. This is strongly supported by several lines of evidence. However, direct small RNA – unwound DNA interaction cannot be completely ruled out (discussed in Martienssen and Moazed 2015).

1.5.1 Nuclear RNAi in *S. pombe*

In the fission yeast *S. pombe* heterochromatin is formed on two genomic locations, the centromere and the silent mating-type locus. For simplicity only the centromere will be discussed here, as additional RNAi-independent mechanisms for targeting heterochromatin are in place at the mating-type locus (Jia et al. 2004).

The *S. pombe* centromere consists of an inner central region and repeats flanking this region, called “innermost” (*imr*) and “outermost” (*otr*) repeats. Heterochromatin forms primarily on the *otr* repeats and in its absence *S. pombe* suffers from chromosome segregation defects (Volpe et al. 2003). Major RNAi factors involved are Dicer (*dcr1*⁺), Argonaute (*ago1*⁺), which incorporates ~ 22nt siRNAs, and RdRP (*rdp1*⁺), while chromatin components involved include histone deacetylases (HDACs), e.g. the SHREC complex, the histone lysine (K) methyltransferase (HKMT) Clr4, that di- or trimethylates (me2/3) histone H3K9, and three chromodomain containing proteins, Swi6, Chp2 (both HP1 proteins) and Chp1. Most of these components exist in one or several different protein complexes, which often interact with each other via additional subunits, thus constituting a tightly interwoven network of physical and functional interactions (Fig. 3).

S. pombe RITS complex consists of small RNA loaded Ago1, Tas3 and Chp1. Notably, Chp1 binds to H3K9me2/3 via its chromodomain and therefore provides a second connection between RITS and chromatin additional to small RNA interaction with nascent transcripts (Verdel et al. 2004). Importantly, RITS is capable of initiating heterochromatin formation when tethered to an ectopic locus (Buhler et al. 2006). Following its initial localization to chromatin, RITS is thought to recruit Clr4, which methylates H3K9 and stabilizes RITS chromatin association via Chp1. This directly links the RNAi and heterochromatin machineries. Clr4 resides in the Clr4-Rik1-Cul4 (CLRC) complex (with additional subunits Raf1, Raf2, Cul4 and Stc1), which interacts via its different subunits

with both RITS and the RNA-dependent RNA polymerase complex (RDRC; subunits: Rdp1, Hrr1, Cid12) (Holoch and Moazed 2015). The link to RDRC provides a positive feedback loop to the siRNA biogenesis machinery, as RDRC-generated dsRNA precursors are further processed into siRNAs by Dicer.

Moreover, Clr4-dependent H3K9me2/3 leads to binding of the HP1 proteins Swi6 and Chp2, which then establishes silenced heterochromatin by binding additional repressive activities such as the exosome complex and the SHREC HDAC complex. Moreover, Swi6 binding helps to recruit RDRC via Ers1 interaction (Hayashi et al. 2012).

All these components constitute a powerful self-reinforcing positive feedback loop that ensures heterochromatin maintenance throughout the cell cycle. However, redundancies are limited in the *S. pombe* system, as core components such as Ago1, Rdp1 and Dicer are only encoded by one gene. Thus, removal of any core RNAi component leads to a breakdown of heterochromatin and vice versa (Volpe et al. 2002; reviewed in detail in Martienssen and Moazed 2015 and Holoch and Moazed 2015).

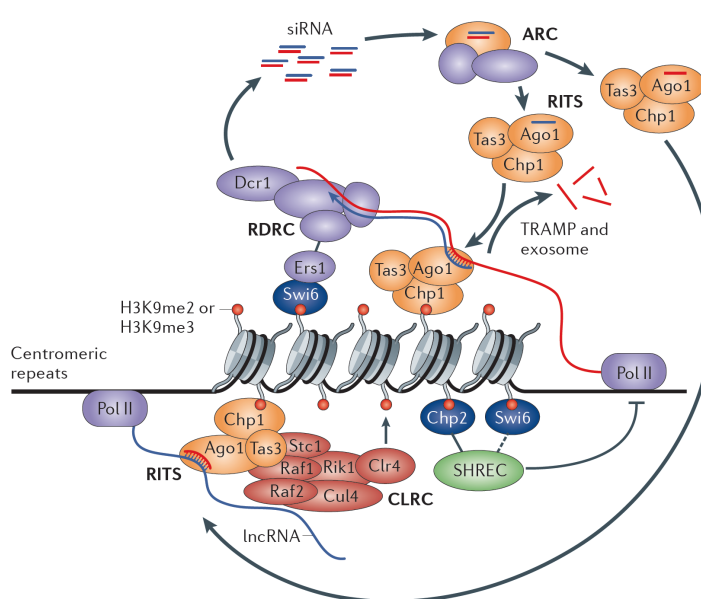


Figure 3: The nuclear RNAi and heterochromatin machinery in *S. pombe* creates a self-reinforcing positive feedback loop

On centromeric repeats in *S. pombe* heterochromatin is perpetuated throughout the cell-cycle via an RNAi-dependent feedback loop. Pol II transcripts are both substrates for siRNA biogenesis by RDRC (dsRNA synthesis by Rdp1) and Dicer 1 (Dcr1) and landing platforms for the RITS (sRNA, Ago1, Tas3, Chp1) complex. RITS recruits the CLRC complex (Clr4, Raf1, Raf2, Stc1, Rik1, Cul4), which methylates histone H3K9. H3K9me2/3 serves as a binding site for

RITS and the HP1 proteins Swi6 and Chp2, which recruit additional silencing activities to the locus, such as the Snf2-histone deacetylase repressor complex (SHREC) and TRAMP non-canonical poly(A) polymerase and the nuclear exosome complex, that can degrade RNA. Adapted from Holoch and Moazed 2015.

1.5.2 RNA-directed DNA methylation in *Arabidopsis thaliana*

The widely-used plant model organism *Arabidopsis thaliana* has been another rich source of information on nuclear RNAi in past decades. The *Arabidopsis* pathway is similar in key components to *S. pombe*, however additional factors are involved, most importantly

DNA methylation, adding additional layers of complexity to it (Fig. 4) (reviewed in Pikaard and Mittelsten Scheid 2014).

In *Arabidopsis* ~24-nt siRNAs target repetitive elements for TGS by DNA methylation, including centromeric satellite repeats, which are made up of tandem repeats of the Athila LTR retroelement. Due to the dominating role of DNA methylation this pathway is termed RNA-directed DNA methylation (RdDM). Important RNAi factors are the Argonaute protein AGO4, Dicer protein DCL3 and a RdRP, called RDR2. The complex of siRNA and AGO4 targets genomic loci by base-pairing with nascent transcripts that are produced by plant specific DNA-dependent RNA polymerase V (Pol V) (Wierzbicki et al. 2008; Wierzbicki et al. 2009). This interaction is further reinforced by direct interactions of AGO4 with the GW motifs of Pol V and its interactor KTF1, therefore stabilizing AGO4 on chromatin. Silencing is then mediated by DNA methylation via DNA methyltransferase DRM2. The protein RDM1 links AGO4 and DRM2 for this purpose ((Gao et al. 2010)).

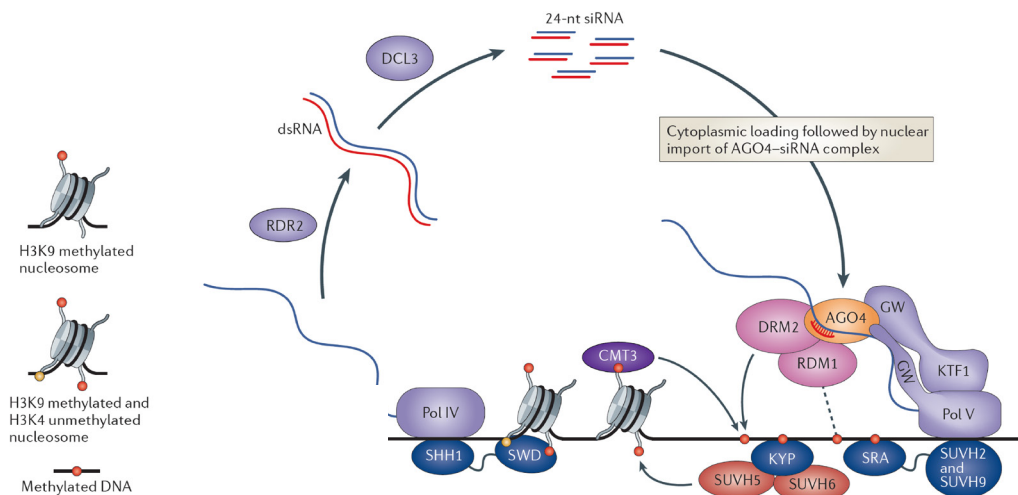


Figure 4: The RNA-directed DNA methylation (RdDM) pathway in *A. thaliana*

Repetitive elements are silenced by 24-nt siRNA directed DNA methylation in *Arabidopsis thaliana*. The plant specific DNA-dependent RNA polymerases IV and V are crucial for siRNA biogenesis and targeting, respectively. Pol IV can recognize histone modifications indicative of silenced loci by its interaction with SHH1, targeting siRNA biogenesis to such loci. Pol IV transcripts are converted to dsRNA precursors by RDR2 and processed to 24-nt siRNAs by DCL3. siRNAs incorporated into AGO4 bind loci transcribed by Pol V via siRNA-nascent transcript interaction and direct AGO4 binding to Pol V and its associated factor KTF1. DNA is methylated by DRM2. DNA methylation marks are read by HKMT KYP, which methylates H3K9. H3K9me is recognized in turn by DNA methyltransferase CMT3. Furthermore, Pol V recruiting proteins SUVH2 and SUVH9 are capable of recognizing DNA methylation, biasing Pol V recruitment to already-methylated loci. Adapted from Holoch and Moazed 2015.

Moreover, additional mechanisms for silencing are in place, for instance via H3K9 methyltransferases, called KYP, SUVH5 and SUVH6. KYP can bind to methylated DNA and contributes to the maintenance of DNA methylation at the silenced loci, thus forming

a positive feedback loop (Johnson et al. 2007). In contrast to *S. pombe* the only HP1 homologue of *Arabidopsis* LHP1 does not seem to play an important role in the RdDM pathway (Pikaard and Mittelsten Scheid 2014). Other activities are also crucial for silencing, including chromatin remodeling by SWI2/SNF2 ATPase DDM1 and additional DNA methyltransferases DRM1 and CMT3 (CMT3 contains a chromodomain, like *S. pombe* HKMT Clr4). (reviewed in Martienssen and Moazed 2015 and Pikaard and Mittelsten Scheid 2014)

Biogenesis of ~24 nt siRNAs depends on plant-specific DNA-dependent RNA polymerase IV (Pol IV). The factor SHH1 associated with Pol IV specifically binds nucleosomes with methylated H3K9 and unmethylated H3K4, therefore directing Pol IV to loci already targeted by the silencing machinery, reinforcing the silenced state (Law et al. 2013).

In summary, the *Arabidopsis* RdDM system is also shaped by self-reinforcing feedback loops. However, due to additional redundancies, the strict codependency of heterochromatin and RNAi known from the *S.pombe* RNAi system is lacking in *Arabidopsis*.

1.5.3 Nuclear RNAi in *C. elegans*

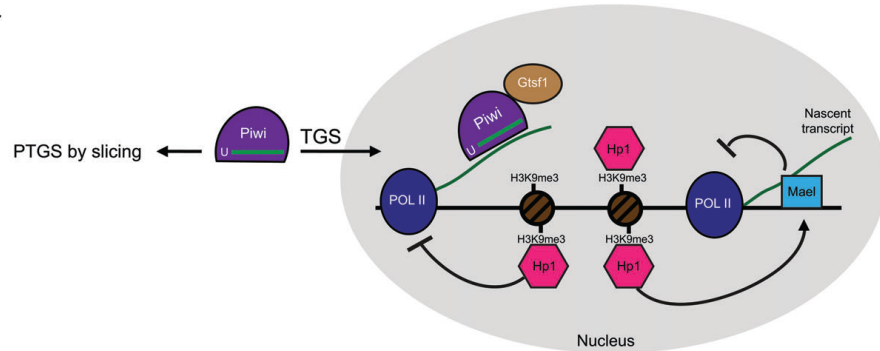
Several examples of nuclear RNAi inducing TGS have furthermore been observed in *C. elegans*. In somatic cells exogenous dsRNA triggers primary PTGS by siRNAs associated with the Argonout protein RDE-1, which in turn triggers production of secondary siRNAs via a RdRP. These are loaded into a second AGO NRDE-3 (Nuclear RNAi deficient 3). NRDE-3 can induce TGS and H3K9 methylation in the somatic nucleus together with the factors NRDE-2 and NRDE-1, although mechanistic details are not known (reviewed in Holoch and Moazed 2015).

1.5.4 Nuclear piRNAs in genome defense

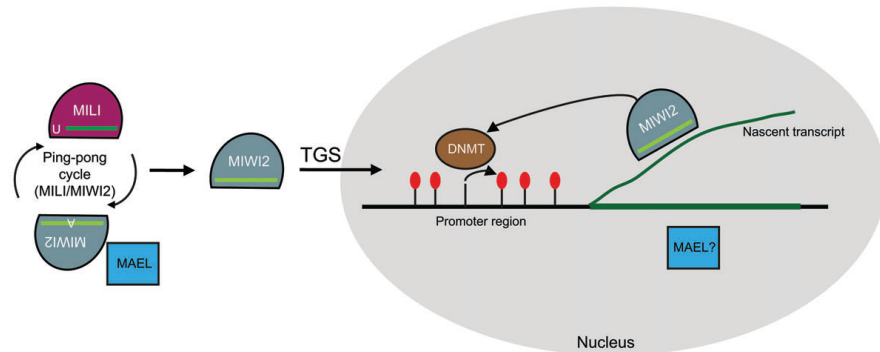
In recent years further instances of nuclear RNAi-mediated TGS have come to light, most notably piRNAs in *Drosophila*, mammals and *C. elegans*, which defend the germline against transposable and repetitive genetic elements.

piRNAs in *Drosophila* and mammals are produced from genomic clusters, which contain a collection of old transposable elements and thus can be viewed as a genomic memory of the organism's transposon encounters. Both mice and *Drosophila* use nuclear PIWI-clade Ago proteins, called MIWI2 and Piwi, respectively, to induce TGS.

A. *D. melanogaster*



B. *M. musculus*



C. *C. elegans*

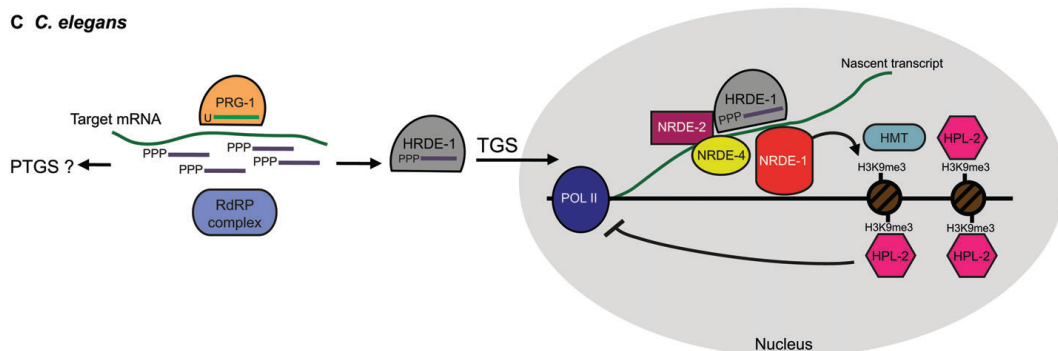


Figure 5: piRNA pathways lead to TGS in several different organisms

(A) The *Drosophila melanogaster* piRNA pathway induces TGS by piRNA-Piwi complex at DNA-dependent RNA polymerase transcribed transposon loci. H3K9 methylation, HP1 proteins and Maelstrom (Mael) are important for TGS. (B) In the *Mus Musculus* male fetal germline MIWI2 participates in a ping-pong piRNA amplification cycle in the cytoplasm. MIWI2 also localizes to the nucleus, where it induces TGS by directing DNA methyltransferases (DNMT) to transposon promoters, which initiate DNA methylation. (C) In the *C. elegans* germline the Ago protein PRG-1 is loaded with primary 21U-RNAs. This initiates an RdRP dependent amplification cycle that yields 22G-RNAs. These are incorporated into the Ago protein HRDE-1 and induce TGS in the nucleus, assisted by the factors NRDE-1, -2 and -4, HKMTs (HMTs) methylating H3K9me3 and the HP1 protein HPL-2. Adapted from Weick and Miska 2014.

In addition to PTGS by the piRNA ping-pong cycle, in *Drosophila* it was demonstrated that nuclear Piwi represses transposons by TGS, H3K9 methylation and HP1a recruitment (Huang et al. 2013; Le Thomas et al. 2013; Sienski et al. 2012), indicating

that nuclear RNAi directly targets heterochromatin formation (Fig. 5A). The protein Maelstrom might add an additional layer of silencing downstream of H3K9me-dependent heterochromatin formation (Sienski et al. 2012). Furthermore, piRNA targeted loci can be the source of piRNA production when a complex containing the HP1 protein Rhino is targeted there, which would once more constitute a self-reinforcing positive feedback loop (Zhang et al. 2014; Mohn et al. 2014).

In the mouse male fetal germline a brief time window is used by Argonaute proteins MILI and MIWI2 to generate piRNAs by a ping-pong cycle and to silence transposons by DNA methylation, yet mechanistic details for this pathway are largely lacking (Fig. 5B) (reviewed in Iwasaki et al. 2015; Weick and Miska 2014).

In the *C. elegans* germline cytoplasmic 21U-piRNAs incorporated into the PIWI-clade AGO protein PRG-1 initiate production of secondary 22G-RNA. These effector 22G small RNAs bind to the AGO protein HRDE-1, travel to the nucleus and induce stable TGS, which can be inherited across generations. Inheritance requires nuclear factors including HRDE-1 and the HP1 protein HPL-2, yet not the cytoplasmic initial stimulus: 21U-RNAs and PRG1. Interestingly PRG-1 induced silencing can be counteracted by another Ago protein called CSR-1. It was suggested that while PRG-1 samples the germline for non-self RNAs, CSR-1 incorporates small RNAs derived from all self-RNAs, thus essentially protecting the germline transcriptome from silencing. (reviewed in Holoch and Moazed 2015; Weick and Miska 2014).

In summary, the induction of TGS by RNAi is a common theme across a wide range of different organisms and is generally shaped by self-reinforcing positive feedback loops.

1.6 Boundary elements regulate heterochromatin spreading

The above-described prevalence of self-reinforcing feedback loops during heterochromatin formation and maintenance indicates that mechanisms must be in place to stop heterochromatin from invading euchromatic regions in order to assure proper gene expression. The importance of such mechanisms is well exemplified by the classic phenomenon of position effect variegation (PEV) observed in *D. melanogaster* in the 1930s, where placing of an expressed gene next to pericentric heterochromatin causes silencing of that gene by stochastic heterochromatin spreading (Muller 1930, reviewed in Elgin and Reuter 2013).

Conceptually, two kinds of mechanisms can prevent heterochromatin spreading: either a fixed boundary element, such as a *cis*-acting DNA sequence or, alternatively, the balance of euchromatic and heterochromatic activities in the border region can be the determining factor. For the latter the term negotiable border was coined (Kimura and Horikoshi 2004). *Cis*-acting DNA boundary elements that form fixed borders can recruit a large variety of activities to prevent heterochromatin spreading, however all of these aim at interfering with the self-reinforcing positive feedback loop of histone modification and modification binding that underlies most spreading processes (Fig. 6). Such activities can be represented by enzymes that directly counteract heterochromatic enzymes, for instance histone acetyltransferases, that propagate activating histone acetylation, or histone demethylases such as LSD1, which demethylates H3K9me at *S. pombe* centromeric regions and acts in direct opposition of H3K9 methyltransferase Clr4 (Fig. 6a) (Lan et al. 2007). The JmjC domain histone demethylase DDM-1 has furthermore been implicated in restricting heterochromatin spreading in the filamentous fungus *Neurospora crassa* as well as the demethylase IBM1 in *A. thaliana*. In addition, protection of activating histone marks can be used to stop heterochromatin (Fig. 6b). The protein Bdf2 confers such protection to histone H4K16 acetylation by binding to it via its double bromodomain at the *IRC* boundary element of *S. pombe* centromeres 1 and 3, where it is recruited by Epe1 (Wang et al. 2013). Epe1 has homology to JmjC histone demethylases, however, whether it is enzymatically active is unclear. Another simple concept to separate heterochromatic and euchromatic regions can be implemented by forming nucleosome-free regions, which withhold the substrate (the nucleosome) needed to sustain heterochromatin spreading by repeated cycles of histone modifications (Fig. 6c) (Bi et al. 2004). Similarly, high turnover of nucleosomes in the boundary region can be an effective measure to confine heterochromatin within its proper territory (Fig. 6d) (Nakayama et al. 2007).

Another well-documented role in preventing spreading of heterochromatin is attributed to the process of transcription itself. Generally, transcription can hinder heterochromatin propagation in two ways: the recruitment of activating histone modifiers and histone turnover by RNA-Polymerase movement through chromatin. It might be for this reason that tRNA genes have heterochromatin boundary functions in many organisms, including *S. cerevisiae*, *S. pombe* and mammals (reviewed in Wang et al. 2014). Additional to this, also the transcript can contribute to boundary activity by binding and evicting heterochromatin components, as it was shown for Swi6 in *S. pombe* (Fig. 6e) (Keller et al. 2012; Keller et al. 2013).

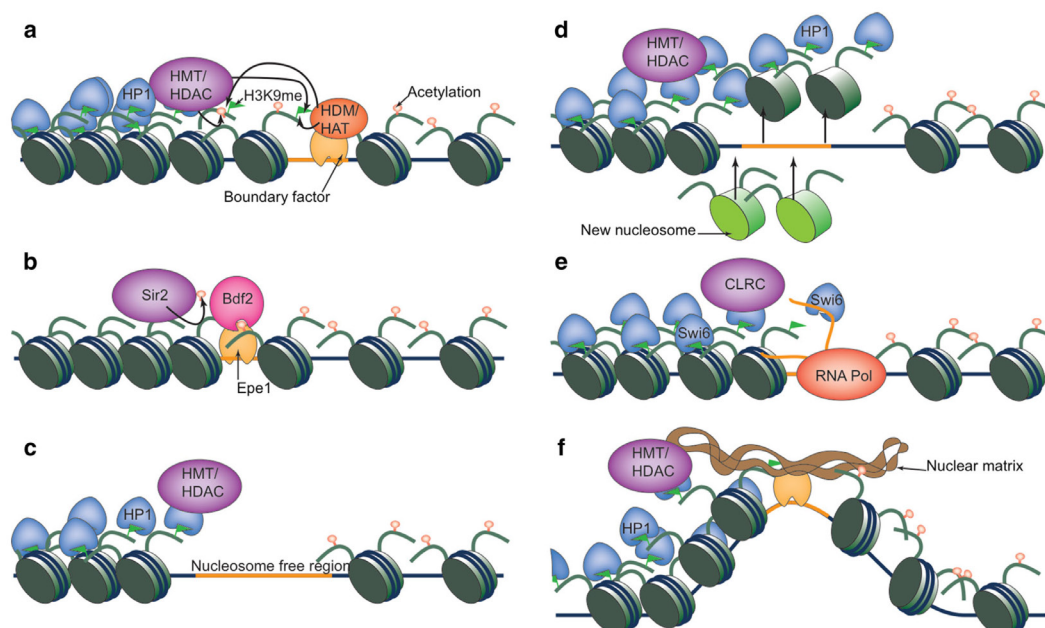


Figure 6: Heterochromatin boundaries can be established by diverse mechanisms

(a) Boundary elements can recruit enzymatic activities that oppose heterochromatic modifiers, e.g. histone acetyltransferases (HATs) or histone demethylases (HDMs). (b) Protection of euchromatic histone marks can enforce a heterochromatin boundary, e.g. by H4K16ac binding by Bdf2. (c) Nucleosome free regions can prevent spreading of heterochromatin. (d) Enhanced histone turnover at boundaries can hinder spreading of heterochromatin. (e) Transcription can implement heterochromatin boundaries in *S. pombe* by providing RNA that evicts Swi6 from chromatin. (f) Tethering of boundaries to nuclear structures blocks spreading of heterochromatin. Adapted from Wang et al. 2014.

Separation of euchromatic and heterochromatic genomic regions can furthermore be achieved by placing steric roadblocks for heterochromatin. This is often achieved by the association of boundary elements with nuclear structures (Fig. 6f). Such clusters can be TFIIIC foci in *S. pombe* or binding sites for known insulator components like CTCF and CP190, which frequently flank H3K27me3 domains in mammals and *Drosophila* (Noma et al. 2006; Li and Zhou 2013; Narendra et al. 2015). Interestingly, a recent elegant study demonstrated that CTCF in the *Drosophila* *HOX* clusters establishes euchromatin-heterochromatin boundaries and prevents euchromatin from invading heterochromatic domains, showing that in some situations euchromatin also has the capacity to spread and needs to be kept in check (Narendra et al. 2015).

Negotiable borders that do not rely on cis-acting DNA elements for recruitment of their components are formed, for instance, by opposing enzymatic activities of histone acetyltransferases and deacetylases in *S. cerevisiae* or by the presence of euchromatic proteins like transcription associated histone variant H2A.Z (Suka et al. 2002; Meneghini et al. 2003).

1.7 The model organism *Tetrahymena thermophila*

Tetrahymena thermophila is a ciliated protist, living in fresh water and pursuing a phagotrophic lifestyle in the wild. Evolutionary ciliates are eukaryotes and classified into the group of alveolates, one of the eight major branches of eukaryotic evolution (Baldauf 2003). With plants and animals residing in different major branches of the evolutionary tree, ciliates are clearly quite distinct from other model organisms, such as mice, *Drosophila melanogaster* or *Arabidopsis thaliana*. However, there is a high degree of conservation in central macromolecular cellular machineries enabling the study of these (Frankel 2000). Together with ease of handling, cultivation and manipulation, for instance rapid growth at 30 °C and easy transgene introduction by homologous recombination, *Tetrahymena* lends itself to the study of important cell biological processes. This resulted in a number of landmark discoveries made in *Tetrahymena*, for example the first description of dynein motors (Gibbons and Rowe 1965), self-splicing RNAs (Cech et al. 1981), the enzyme telomerase (Greider and Blackburn 1985) and histone acetyltransferases (Brownell et al. 1996). Notably, the Nobel Prize was awarded for two of these discoveries: for self-splicing RNAs in 1989 and for telomerase in 2009.

This impressive history makes future discoveries of general relevance seem likely, an outlook that is supported by the publication of the macronuclear genome and availability of first drafts of the micronuclear genome, which will facilitate future genomic studies (Eisen et al. 2006; Tetrahymena Comparative Sequencing Project, Broad Institute of Harvard and MIT (<http://www.broadinstitute.org/>)).

1.8 Nuclear dimorphism and genome rearrangements in *Tetrahymena thermophila*

Despite being a unicellular protist, *Tetrahymena* displays somatic and germline differentiation, however not on the level of cells, as in animals and plants, but on the level of nuclei, a phenomenon termed nuclear dimorphism (reviewed in detail in Karrer 2012). A single vegetative *Tetrahymena* cell contains two kinds of nuclei, the somatic macronucleus (MAC) and the germline micronucleus (MIC) (Fig. 7). These nuclei differ in several important aspects. The small, diploid MIC, which during vegetative growth is transcriptionally silent, contains five chromosomes representing the full tetrahymena germline genome (~ 150 Mbp), including approximately 10,000 Internal Eliminated

Sequences (IESs) ranging in size from 0.2 to 40 kb (Fig. 9) (Schoeberl et al. 2012). Many of these IESs are transposon-related and repetitive sequences. In contrast to this, the MAC is transcriptionally active during vegetative growth and contains a trimmed version of the MIC genome lacking all IESs. Any genomic regions found in the MAC are by definition called macronuclear-destined sequences (MDSs). The MAC is highly polyploidy and contains approximately 45 copies of each chromosome, however these are not only devoid of IESs but also fragmented to about 180 MAC chromosomes (Fig. 9) (Karrer 2012).

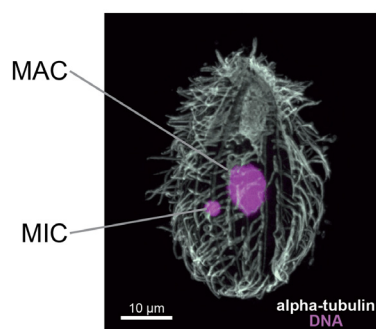


Figure 7: Nuclear dimorphism and life cycle in *Tetrahymena*

A single *Tetrahymena* cell contains two different nuclei, a germline micronucleus (MIC) and somatic macronucleus (MAC). Adapted from Kataoka and Mochizuki 2011.

During vegetative growth of *Tetrahymena* the single-celled organism simply divides its cytoplasm to duplicate. Each daughter cell receives a MIC, which duplicates by mitosis and a MAC, which divides by amitosis (random distribution of the replicated chromosomes to two new MACs) (Fig. 8).

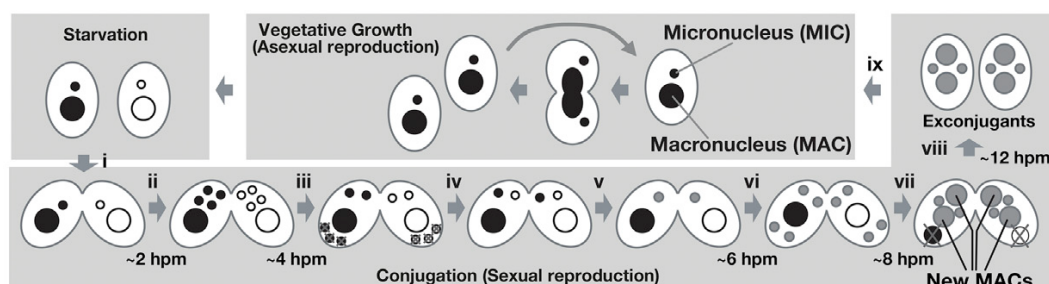


Figure 8: Outline of the lifecycle of *Tetrahymena thermophila*

During vegetative growth of *Tetrahymena*, MIC and MAC segregate independently to daughter cells. When cells of different mating types are starved and mixed conjugation is induced (i), each MIC undergoes meiosis forming four meiotic products (ii). One product is selected and divides mitotically, forming two pronuclei (iii). As cells are paired and cytoplasm fused, one pronucleus per cell can be exchanged (iv). In each cell the two pronuclei fuse (v) and divide twice mitotically (vi) in order to form two new MAC and two new MIC (vii). The parental MAC is degraded during this process. Finally, the pair resolves, yielding two separate cells, termed exconjugants (viii). As conjugation is a synchronized process the time of each event can be estimated in hours post mixing (hpm) and is indicated. Adapted from Noto et al. 2015.

More complex nuclear processes are at work during sexual reproduction (also termed conjugation) of *Tetrahymena*, which is induced by nutrition starvation and mixing of two strains possessing different mating types (Fig. 8i). Two cells with different mating types pair, their germline MICs undergo meiosis and one post-meiotic mitosis (Fig. 8ii). Thereby two zygotic pronuclei are formed and one each is exchanged between the two cells (Fig. 8iii and 8iv). In each cell the two pronuclei are fused to form one zygotic nucleus, which divides twice mitotically, yielding a total of four zygotic products per cell (Fig. 8v and 8vi). The two posterior zygotic products differentiate to become new MICs, whereas the anterior two products develop to become new MACs, while the old MAC is degraded (Fig. 8vii and 8viii).

The large difference between MIC and MAC genomes is caused by extensive genome rearrangements in the new MAC. The process of RNAi-directed DNA elimination, which removes IESs from the new MAC genome, will be discussed below. Additional to IES elimination the five MIC chromosomes, which are still intact at the onset of new MAC development, are processed into smaller MAC chromosomes at a conserved fragmentation consensus site of 15 bp, a process called chromosome breakage. After the addition of new telomeres to these approximately 180 new MAC chromosomes, they are amplified to ~45 copies by endoreplication (Karrer 2012).

The process of conjugation is highly synchronous, which facilitates its study in a time resolved manner.

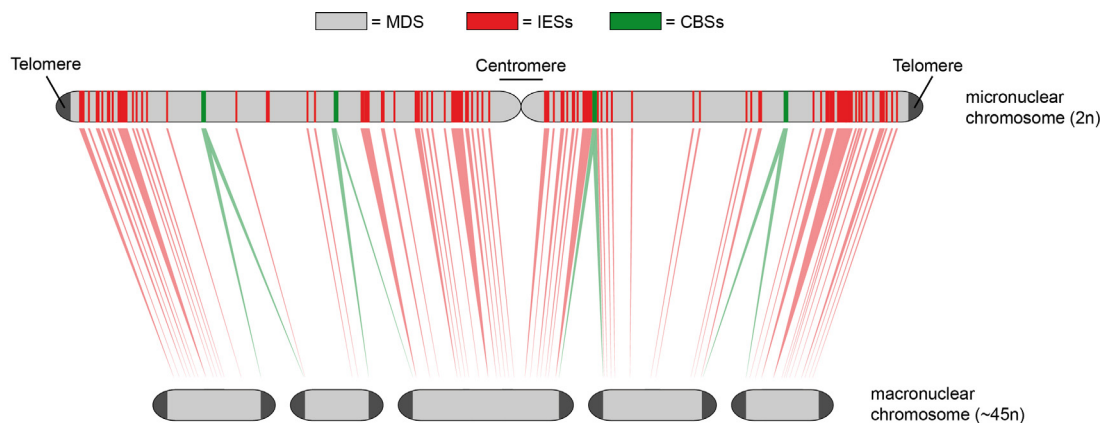


Figure 9: DNA rearrangements during new MAC development in *Tetrahymena*

During new MAC development the MIC chromosomes are fragmented into smaller MAC chromosomes by breakage at conserved chromosome breakage sequences (CBSs, green). Telomeres are attached to the new chromosome ends. Furthermore, IESs (red) are removed from the MIC chromosomes by RNAi-directed DNA elimination, leaving only macronuclear-destined sequences (MDSs). Finally, the new MAC chromosomes are replicated to approx. 45 copies. For simplicity, only the fate of one haploid MIC chromosome is depicted before endoreplication.

1.9 Small RNA pathways in *Tetrahymena*

The prevalence of a high number of different RNAi pathways in *Tetrahymena* is supported by the discovery of twelve PIWI-clade Argonaut proteins and the study of their associated small RNAs. These can be expressed either during vegetative growth or during conjugation or both (Lee and Collins 2006; Couvillion et al. 2009; Mochizuki et al. 2002; Noto et al. 2015). Furthermore, one RdRP protein Rdrp1 is present in *Tetrahymena* and three Dicer enzymes Dcr1p, Dcr2p and Dcl1p (Lee and Collins 2007; Lee et al. 2009; Mochizuki and Gorovsky 2005; Malone et al. 2005), which are important for different small RNA biogenesis pathways together with diverse additional factors (Talsky and Collins 2012). Moreover, development of RNAi as an experimental tool strongly suggests that gene silencing by small RNAs functions in *Tetrahymena* (Howard-Till and Yao 2006). In this work only a conjugation-specific population of 26-32 nt small RNAs called scnRNAs will be discussed. These scnRNAs are essential for DNA elimination during new macronuclear development.

1.10 Small RNA-directed DNA elimination

1.10.1 Two types of IESs

As described above, about 10,000 IESs comprising approximately 30% of the MIC genome are removed from the new MAC genome during conjugation in a process termed DNA elimination. Precise targeting of DNA segments for elimination is achieved by complementary small RNAs called scan-RNAs (scnRNAs) (Mochizuki et al. 2002). How the IES-specificity of this scnRNAs pool is generated was initially explained by an RNA-based genome comparison process between MIC and parental MAC, which was termed scanning model (hence the term scnRNAs) (Mochizuki and Gorovsky 2004a; Mochizuki and Gorovsky 2004c). Partially, this model still holds true, however recent findings have made clear that the scanning model needs to be refined (Schoeberl et al. 2012; Noto et al. 2015).

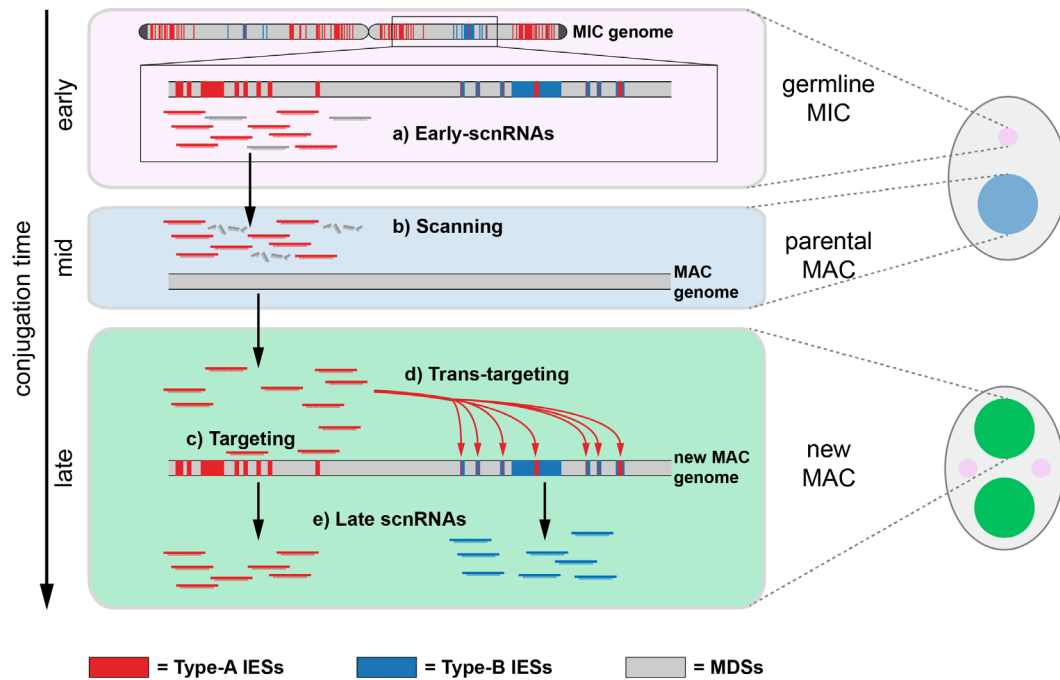


Figure 10: Early- and Late-scnRNAs target IESs for elimination

The process of DNA elimination is schematically visualized focusing on biogenesis and cellular trafficking of different scnRNAs populations. (a) Early-scnRNA biogenesis takes place in the MIC from Type-A IESs and surrounding MDS regions. (b) After import into the parental MAC MDS-specific Early-scnRNAs are degraded in a RNA-based genome comparison process called scanning. (c) Highly Type-A IESs specific Early-scnRNAs are imported in the new MAC, where they induced elimination of Type-A IESs and (d) *trans*-target complementary regions in Type-B IESs. (e) Targeting and *trans*-targeting triggers production of Late-scnRNAs, which allow subsequent elimination of Type-B IESs.

Biogenesis of scnRNAs takes place at two discrete time points during conjugation in two different nuclei. Early-scnRNAs are products of the germline MIC and are expressed at the onset of meiotic prophase in early conjugation from a subset of approximately 60% of all IESs and their neighboring MDS regions (Fig. 10a and 11a). A second subset of about 40% of IESs does not yield any scnRNAs during early conjugation, but produces scnRNAs in late conjugation in the new MAC. These are called Late-scnRNAs (Fig. 10e and 11g). Based on this behavior two types of IESs were defined: Type-A IESs expressing Early-scnRNAs and Type-B IESs expressing only Late-scnRNAs (Noto et al. 2015). Generally Type-A IESs are mostly found near telomers and presumed centromeres, while Type-B IESs are instead located in chromosomal arm regions (Fig. 10, unpublished observation Kazufumi Mochizuki). Moreover, most Type-B IESs contain short regions that have homology to Type-A IESs, called A-repeats (Noto et al. 2015).

1.10.2 Biogenesis and scanning of Early-scnRNAs

Biogenesis of Early-scnRNAs is initiated by transcriptional activation of the otherwise-silenced MIC during meiotic prophase (Sugai and Hiwatashi 1974). In this stage the MIC morphology changes dramatically by elongating to a crescent shape and DNA-dependent RNA Polymerase 2 (Pol II) and several other transcription-associated factors localize into the MIC, including TATA-binding protein and histone H2A.Z (Stargell et al. 1993; Stargell and Gorovsky 1994; Mochizuki and Gorovsky 2004b). This allows for bidirectional transcription, which is biased strongly towards Type-A IESs and, to a lesser degree, to Type-A IES neighboring MDS regions (Chalker and Yao 2001; Mochizuki and Gorovsky 2004b; Schoeberl et al. 2012). The transcription products presumably form dsRNAs, which are substrate to Dicer-like protein Dcl1p catalyzed cleavage (Fig. 10a and 11a,b) (Malone et al. 2005; Mochizuki and Gorovsky 2005). The resulting 26-32 nt long scnRNAs-duplexes are exported to the cytoplasm and assisted by the Hsp90 co-chaperone Coi12p loaded into the PIWI-clade Argonaute protein Twi1p, whose intrinsic slicing activity facilitates passenger strand removal (Fig. 11c) (Mochizuki et al. 2002; Noto et al. 2010; Woehrer et al. 2015). The Twi1p-scnRNA-RITS complex is bound by Giw1p, which senses successful removal of the passenger strand, and permits Twi1p import into the parental MAC, where scnRNAs are stabilized by 2'-O-methylation by the enzyme Hen1p and then subjected to scanning (Noto et al. 2010; Kurth and Mochizuki 2009; Aronica et al. 2008). Scanning removes scnRNAs with MDS-specificity from the initial pool of Early-scnRNAs by selectively degrading them. This process requires Twi1p interaction with non-coding (nc) RNA transcripts in the parental MAC, which is facilitated by the RNA-helicase Ema1p (Fig. 10b and 11d) (Aronica et al. 2008).

1.10.3 Late-scnRNA biogenesis is heterochromatin dependent

After scanning the highly Type-A IES specific pool of Twi1p-Early-scnRNAs complexes shuttles to the new MAC and recognizes Type-A IESs and A-repeats in Type-B IESs by base-pairing of Early-scnRNAs with nascent ncRNA, a process referred to as targeting and *trans*-targeting, respectively (Fig. 10c,d and Fig. 11e,f) (Aronica et al. 2008; Noto et al. 2015). This initial trigger then sets off two, likely co-dependent, processes: heterochromatin formation on IESs and Late-scnRNA production. These culminate in the elimination of all IESs in the new MAC, creating a new functional MAC genome. This final step of DNA elimination is catalyzed by the transposase-related protein Tpb2p, which shows cleavage preference for certain DNA motifs enriched at IES boundaries and which can interact with heterochromatin (Cheng et al. 2010; Vogt and Mochizuki 2013).

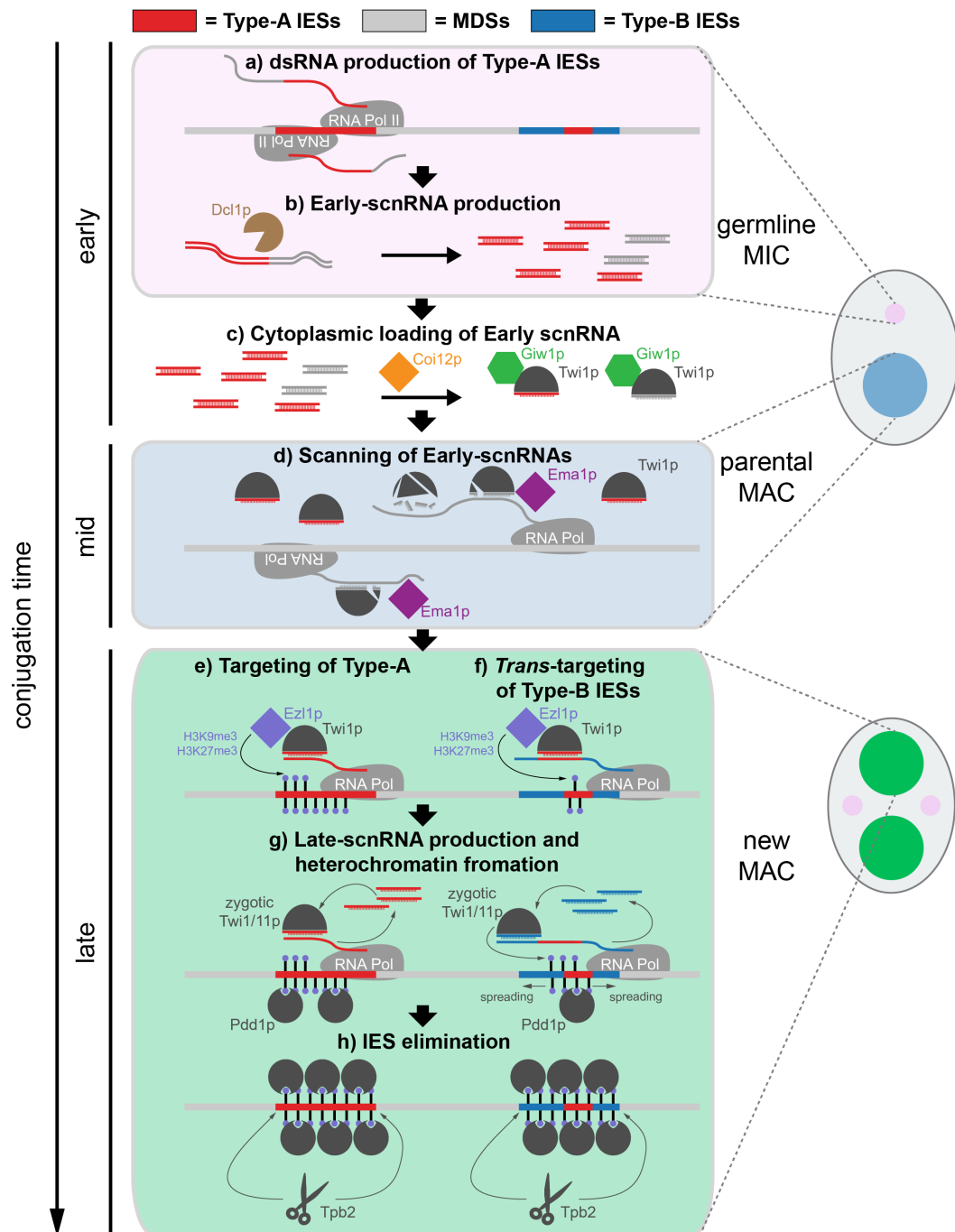


Figure 11: Molecular details of the DNA elimination pathway

(a) During early conjugation, biogenesis of Early-scRNAs takes place in the germline MIC. Pol II (grey) transcribes Type-A IESs (red) and surrounding MDS regions (grey) bidirectionally. (b) The transcripts form dsRNA, which is processed by Dcl1p (brown) into 26-32 nt long double stranded scnRNAs duplexes. (c) Early-scRNAs are exported to the cytoplasm and loaded into PIWI-clade Argonaute protein Twi1p (dark grey). This process is assisted by Hsp90 co-chaperone Col12p (yellow). Intrinsic Twi1p slicer activity removes the passenger strand of the scnRNAs duplex and presumably changes into a conformation, which is sensed by Giw1p (green), facilitating Twi1p import into the parental MAC. (d) During scanning the genome of the parental MAC is the source of

non-coding transcription, which allows MAC-matching scnRNAs (grey) to base-pair, a process facilitated by RNA helicase Ema1p (purple). Target recognition triggers degradation of the scnRNAs, and thereby removes all Early-scnRNAs with MDS-specificity. (e) Early-scnRNAs target Type-A IESs for heterochromatin formation, while (f) Type-B IESs are recognized by *trans*-targeting of short regions of Type-A IES homology. In both cases Ezi1p (blue) methylates histone H3K9 and K27 causing Pdd1p (dark grey) binding. (g) On both types of IESs Late-scnRNA production is initiated by Early-scnRNAs in a heterochromatin-dependent manner. On Type-B IESs this leads to co-dependent spreading of heterochromatin and Late-scnRNA production across the whole IES. (h) Complete heterochromatinization of Type-A and Type-B IESs allows their elimination by Tpb2p.

Due to their Type-A IESs origin, Early-scnRNAs induce heterochromatin formation on Type-A IESs in a fashion likely targeting the whole IES simultaneously (Fig. 10e,g). The E(Z)-related HKMT Ezi1p mediates methylation of histone H3K9 and K27 (Taverna et al. 2002; Liu et al. 2004; Liu et al. 2007). These marks are then bound by the HP1-like proteins Pdd1p, Pdd3p and presumably Tcd1p, establishing a heterochromatic environment on IESs in a remarkable combination of Polycomb and HP1 systems (Madireddi et al. 1996; Nikiforov et al. 2000; Taverna et al. 2002; Xu et al. 2015). Ezi1p and Pdd1p are essential for DNA elimination, whereas Tcd1p is only partially required (Coyne et al. 1999; Liu et al. 2007; Xu et al. 2015). Initial heterochromatinization is followed by formation of heterochromatin foci marked prominently by Pdd1p (Madireddi et al. 1994; Smothers et al. 1997).

Other proteins, including Lia1p, Lia3p, Lia4p, Lia5p and Drb2p, localize to these foci and have been shown to be important for IES elimination (Yao et al. 2007; Rexer and Chalker 2007; Motl and Chalker 2011; Horrell and Chalker 2014; Shieh and Chalker 2013). Yet, how these proteins contribute to elimination of IESs on a molecular level remains largely unexplored.

For Type-B IESs the heterochromatin formation process is more complex. Only the mostly short A-repeats are *trans*-targeted by Early-scnRNAs (Fig. 10d, 11f) (Noto et al. 2015). This *trans*-targeting then triggers Late-scnRNA production on Type-B IESs in the new MAC, which not only occurs on the initially targeted A-repeat, but spreads in *cis*, eventually covering the whole IES. Late-scnRNA generation strictly depends on Early-scnRNAs and furthermore on Ezi1p and Pdd1p (Noto et al. 2015). This dependency on heterochromatin components likely indicates that heterochromatin closely follows the *cis*-spreading of Late-scnRNAs on Type-B IESs, presumably by successive rounds of scnRNAs-dependent heterochromatin nucleation and heterochromatin-dependent Late-scnRNA production (Fig. 10g).

The biogenesis of Late-scnRNAs is poorly characterized beyond the requirements for Twi1p bound Early-scnRNAs, Ezi1p and Pdd1p, however they share basic properties with

Early-scnRNAs, such as length distribution, base composition and Dicer-product signature, and hence are likely also generated by Dcl1p cleavage (Mochizuki and Kurth 2013; Noto et al. 2015). In contrast to Early-scnRNAs, which are loaded into maternally, early (parental MAC) expressed Twi1p, Late-scnRNAs are loaded into zygotically, late (new MAC) expressed Twi1p and Twi11p, which is important for the successful elimination of most Type-B IESs and few Type-A IESs (Fig. 10g) (Noto et al. 2015). Association of Type-A IESs derived scnRNAs with zygotically-expressed Twi11p furthermore indicates that also Type-A IESs express Late-scnRNAs (Fig. 10g) (Noto et al. 2015). These are functionally mostly redundant with Early-scnRNAs, however presumably their production must be confined to Type-A IESs in order to prevent *cis*-spreading of heterochromatin and DNA elimination.

How such a self-reinforcing positive feedback loop could be arrested at the IES-MDS boundary has so far remained elusive, but could involve “*cis*-acting sequences”, which have been characterized for some IESs (Godiska et al. 1993; Chalker et al. 1999). The likely lack of a scanning mechanism for new MAC-expressed Late-scnRNAs makes their highly precise production even more astounding and highlights that exploring the mechanistic basis of this accuracy is a particular intriguing challenge.

2. Results

2.1 Coi6p is an HP1-like protein

Previously a screen was reported that described 22 conjugation-specifically induced (COI) genes and studied 11 of these by means of gene knock-out (KO) and DNA FISH analysis of DNA elimination (Woehrer et al. 2015). This study follows up on these findings by conducting an in-depth analysis of the conjugation-induced gene 6 (*COI6*), which had been observed to be essential for programmed DNA elimination.

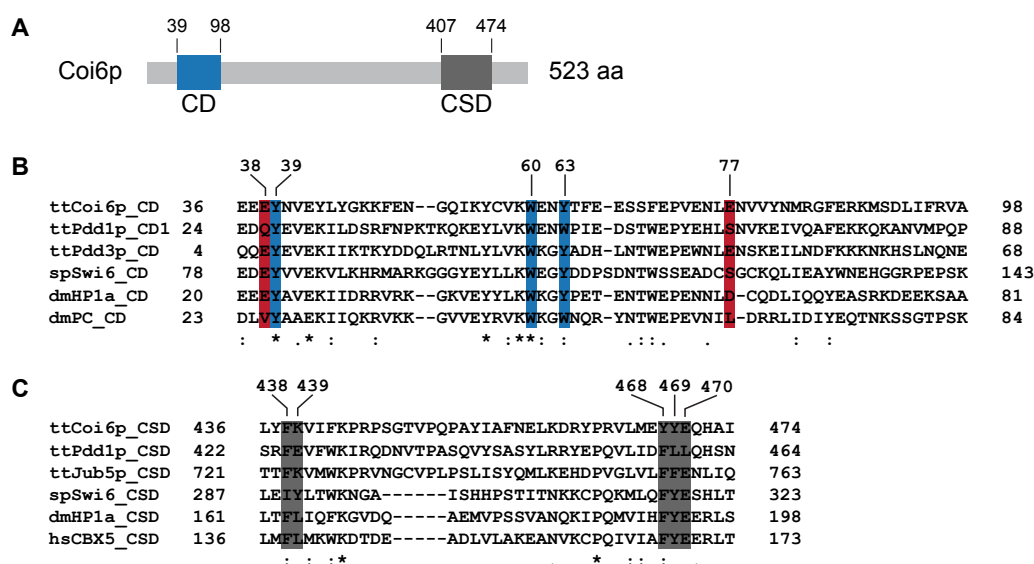


Figure 12: Coi6p is a HP1-like protein with conserved functional residues

(A) Schematic representation of the HP1-like protein Coi6p with chromodomain (CD, blue box) and chromoshadowdomain (CSD, grey box). Amino acids (aa) of domain boundaries are indicated by numbers. (B) Alignment of chromodomains of *Tetrahymena thermophila* (tt) Coi6p, Pdd1p, Pdd3p, *S. pombe* (sp) Swi6 and *D. melanogaster* (dm) HP1a and Polycomb (PC). Amino acid residues of the conserved aromatic cage are shaded in blue and residues of the polar clasp in red. Residue numbers in Coi6p are indicated above. (C) Alignment of chromoshadow domains of *Tetrahymena* Coi6p, Pdd1p, Jub5p, *S. pombe* Swi6, *D. melanogaster* HP1a and *Homo sapiens* (hs) CBX5. Conserved amino acid residues mutated in Coi6p rescue experiments are shaded in grey and their residue numbers are indicated above. (B and C) Alignments were generated using Clustal Omega. *, : and . indicate full, strong and weak conservation of the amino acid residue, respectively.

The gene *COI6* encodes a 523 amino acid (aa) long protein, called conjugation-induced protein 6 (Coi6p). Coi6p contains two conserved domains, a N-terminal chromodomain (CD, aa 39 - 98) and a C-terminal chromoshadow domain (CSD, aa 407 - 474), which are connected by a 318 aa long unstructured, low-complexity region, also called hinge region

(Fig. 12A). This domain organization resembles that of HP1 proteins, and thus Coi6p must be considered an HP1-like protein (Canzio et al. 2014).

Chromodomains are well-characterized histone post-translational modification (PTM) binding modules and they typically bind specifically to histone H3 carrying di- or trimethyl modifications at either lysine 9 (K9) or at lysine 27 (K27). Furthermore, some chromodomains bind to both these modifications (Yap and Zhou 2011). An aromatic cage made up of three conserved aromatic amino acids is crucial for the binding activity of chromodomains (Nielsen et al. 2002; Jacobs and Khorasanizadeh 2002). By aligning the chromodomain of Coi6p to other well-described chromodomains it was observed that these three crucial aromatic residues are conserved in Coi6p (Fig. 12B, Y39, W60 and Y63, blue). This indicates that Coi6p, via its chromodomain, likely binds to histone H3 methylated at either K9 or K27. Moreover, residues forming a negatively charged polar clasp in human HP1, but not Polycomb, chromodomains are conserved in Coi6p (Fig. 12B, E28, E77, red). Presence of this polar clasp confers high specificity for H3K9me3 over H3K27me3 to human HP1 chromodomains, suggesting that Coi6p might only bind H3K9me2/3 (Kaustov et al. 2011). Notably both H3K9me and H3K27me have been reported to be of crucial importance for programmed DNA elimination in *Tetrahymena* (Taverna et al. 2002; Liu et al. 2007).

Although the chromoshadow domain showed a lower degree of conservation than the chromodomain of Coi6p, several conserved residues could be identified and were used for subsequent mutagenesis (Fig. 12C, grey boxed residues).

2.2 Coi6p is specifically expressed during conjugation

Woehrer et al. described *COI6* as conjugation-specific on the level of mRNA using RT-PCR and publicly available data of microarray-based gene expression analysis (Fig. 14A) (Woehrer et al. 2015, Miao et al. 2009). In order to confirm these results, and to facilitate ease of further experiments, two α -Coi6p antibodies were obtained, one from rabbit (r) and one from guinea pig (gp), raised against recombinant MBP-tagged Coi6p (hereafter referred to as α -Coi6p(r) and α -Coi6p(gp)). In western blots α -Coi6p(r) recognized one major band present in WT cells at 8 hours post mixing (hpm), but absent from Δ *COI6* cells (Fig. 13A). Although this band migrated slightly slower than expected with approximately ~70 kDa, it still corresponds well to the predicted 60.0 kDa molecular weight of Coi6p. The α -Coi6p(gp) antibody recognized a band of comparable size in WT, but not in Δ *COI6* samples, at 14 hpm (Fig. 13B). However, this antibody also recognized ~75 kDa protein that appeared in both WT and Δ *COI6*. Nevertheless, both antibodies

stained predominantly the new MAC at 8hpm in WT, but not in $\Delta COI6$ cells (Fig. 13C). The above observations indicate that two antibodies specifically recognizing Coi6p were successfully produced. The α -Coi6p(r) antibody was then used to characterize the protein levels of Coi6p in starved cells and throughout conjugation by western blotting. This analysis confirms the conjugation-specific expression of Coi6p, as no protein was detected in starved cells (0 hpm). Moreover, it indicates that the bulk of Coi6p is expressed during the stage of heterochromatin formation and DNA elimination between 8 and 14 hpm in *Tetrahymena*, suggesting a functional link between Coi6p and these processes (Fig. 14B).

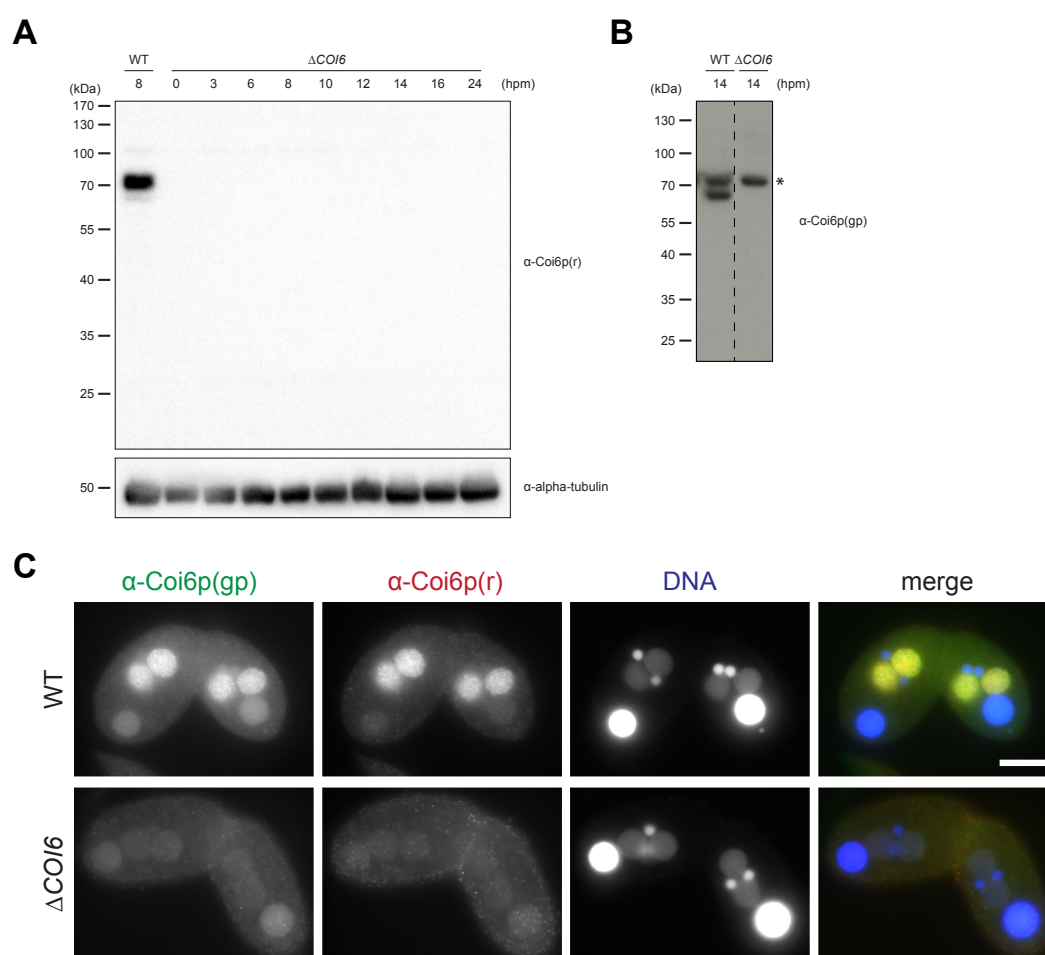


Figure 13: Verification of α -Coi6p antibodies specificity

(A) WT or $\Delta COI6$ samples of indicated time points (in hours post mixing (hpm)) were probed by western blotting with α -Coi6p rabbit (r) or α -alpha-tubulin antibodies. (B) WT or $\Delta COI6$ samples derived of cells at 14 hpm were probed with α -Coi6p guinea pig (gp) antibody; * indicates an unspecific band. (C) Immunofluorescent staining of 8hpm WT or $\Delta COI6$ cells using α -Coi6p(r) and α -Coi6p(gp).

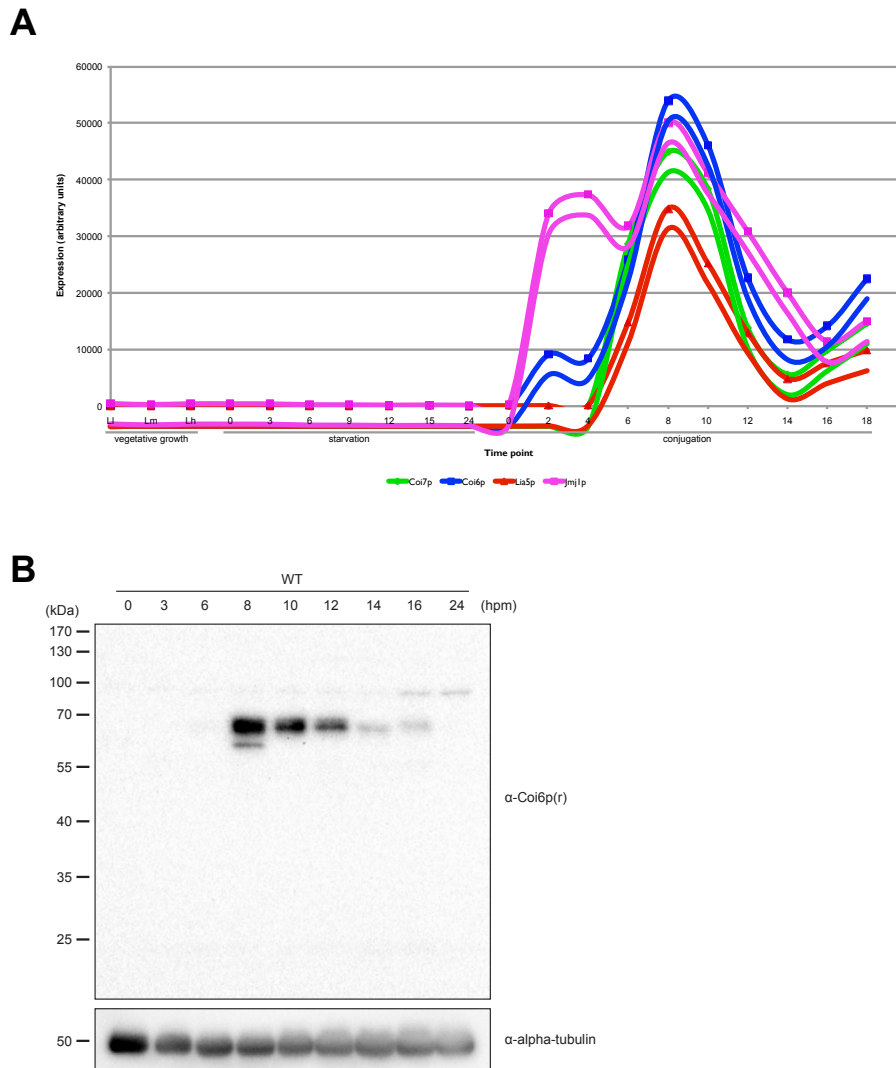


Figure 14: Conjugation-specific expression of Coi6p

(A) Graph showing mRNA expression of *COI6*, *COI7*, *LIA5* and *JMJ1*, based on publicly available data (Miao et al. 2009); numbers on the x-axis depict time points in hours. (B) Western blot showing conjugation-specific expression of Coi6p using the α -Coi6p(r) antibody and α -alpha-tubulin as loading control.

2.3 Coi6p localizes to heterochromatin foci in the new macronucleus

The cellular compartment in which heterochromatin formation and DNA elimination takes place is the new macronucleus (new MAC). Therefore, to test whether Coi6p localized to this nucleus, indirect immunofluorescent stainings using α -Coi6p(gp) and α -Coi6p(r) antibodies were performed (Fig. 13C, 15A and B). Both antibodies showed dispersed

localization of Coi6p in the new Mac at 8 hpm. Furthermore, at 14 hpm Coi6p localized in foci in the new MAC, which closely resemble heterochromatin foci (Fig. 15B). These are known nuclear structures important for DNA elimination and are characterized by an accumulation of Pdd1p (Madireddi et al. 1994; Smothers et al. 1997). Indeed, co-staining of Coi6p and Pdd1p revealed that the foci observed for Coi6p were mostly overlapping with Pdd1p-positive heterochromatin foci (Fig. 15B).

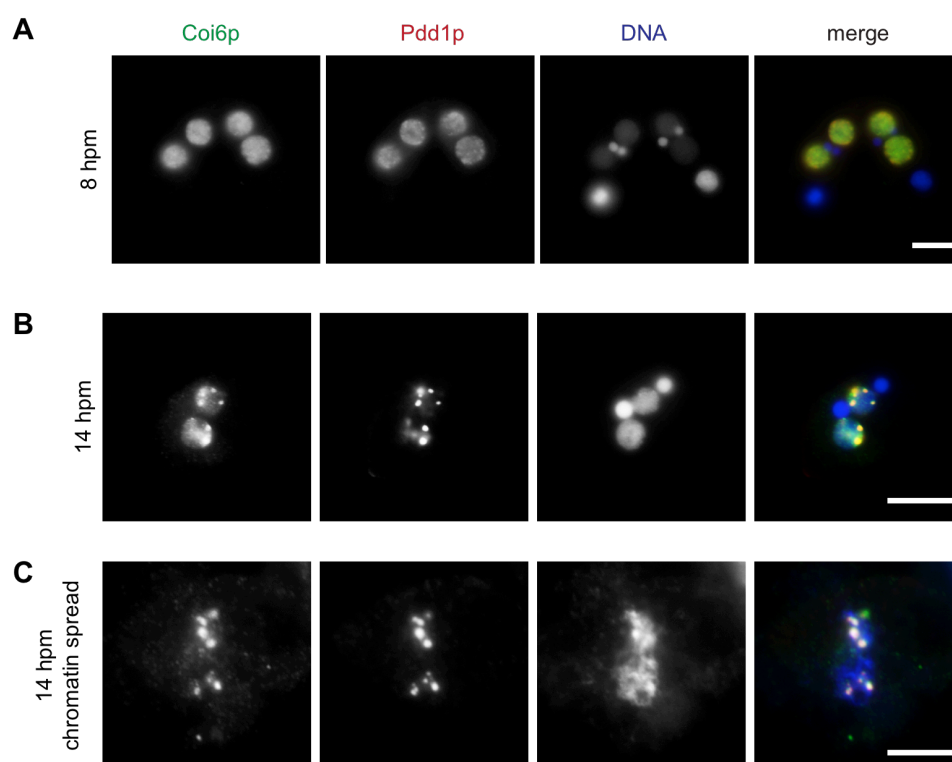


Figure 15: Coi6p localizes to the new MAC and to heterochromatin foci

(A) Co-staining of Coi6p and Pdd1p in WT at 8hpm by α -Coi6p(r) and α -Pdd1p (gp) (kindly provided by Kensuke Kataoka). (B and C) Co-staining of Coi6p and Pdd1p in WT cells at 14 hpm using normal fixation (B) or chromatin spreading (C). α -Coi6p(gp) was used to stain Coi6p and α -Pdd1p (abcam, rabbit) was used to stain Pdd1p. The scale bar indicates 10 μ m.

Despite clear co-localization of Coi6p and Pdd1p in heterochromatin foci at 14 hpm, Coi6p localization seemed to be more dispersed in the nucleus than Pdd1p. To investigate this discrepancy, chromatin spreading experiments were performed, in which the cell is exposed to a detergent-containing buffer prior to firm fixation. After this harsh treatment only proteins tightly bound to chromatin remain in the nucleus. Co-staining of Coi6p and Pdd1p after the treatment with detergent clearly showed enhanced co-localization, apparently by washing out Coi6p that did not localize to heterochromatin foci

(Fig. 15C). This hints that there might be two pools of Coi6p in the new MAC, one localizing to heterochromatin foci and tightly bound to chromatin and one loosely or not at all attached to chromatin.

2.4 Coi6p specifically binds to IESs in vivo

Both Coi6p and Pdd1p are HP1-like proteins. Coi6p co-localized with Pdd1p in heterochromatin foci and Pdd1p had previously been shown to specifically enrich on IESs in ChIP experiments (Taverna et al. 2002). These similarities prompted me to speculate that Coi6p might bind to IESs in a fashion similar to Pdd1p.

To test this, ChIP-seq (chromatin immunoprecipitation followed by high-throughput DNA sequencing) was performed on chromatin derived of cross-linked FACS-sorted nuclei that were enriched $\geq 85\%$ for new MACs. The α -Coi6p(r) antibody and a commercially available α -Pdd1p antibody were used for these experiments. Data is displayed as $\log_2(\text{ChIP}/\text{input})$, meaning that values > 0 correspond to enrichment of ChIP signal over input, while values < 0 indicate depletion. Analyzing Coi6p enrichment in two representative 100 kb genomic loci (the first locus containing mostly Type-A, the second mostly Type-B IESs) showed robust enrichment of Coi6p on all IESs (Fig. 16A). To ask in a comprehensive manner whether Coi6p shows any preference to enrich on Type-A or Type B IESs, a meta-analysis was performed distinguishing these two types of IESs. This analysis was concentrated on the two boundaries of each IES, analyzing 500 bp up- and downstream for both boundaries and then separately compiling data for 3,715 Type-A and 2,863 Type-B IESs, to which ChIP-seq reads can be reliably mapped (Fig. 16G). Coi6p clearly enriched on both Type-A and Type-B IESs (Fig. 16D, compare left and right panel). Importantly, this enrichment is lost when performing ChIP-seq with the same α -Coi6p(r) antibody in ΔCOI6 cells, proving specificity of the ChIP experiment (compare meta-analysis Fig. 16D and E). Comparing these results to the Pdd1p ChIP-seq data, it is clear that enrichment profiles of Coi6p and Pdd1p are largely overlapping on individual loci (compare Fig. 16A and C) as well as genome-wide (compare Fig. 16D and F). Thus, both HP1-like proteins localize to Type-A and Type-B IESs indiscriminately in a global manner. Strikingly, and in contrast to Coi6p ChIP, meta-analysis of Pdd1p ChIP revealed not only high enrichment of Pdd1p on IESs, but also its strong depletion from MDS regions (compare MDS regions in Fig. 16D and F). This noteworthy difference between Coi6p and Pdd1p ChIP might be explained by a higher protein concentration of Pdd1p on chromatin or a better ChIP efficiency for Pdd1p. Yet, it might reflect some difference in

biophysical properties between these two HP1-like proteins, which we do not comprehend at present.

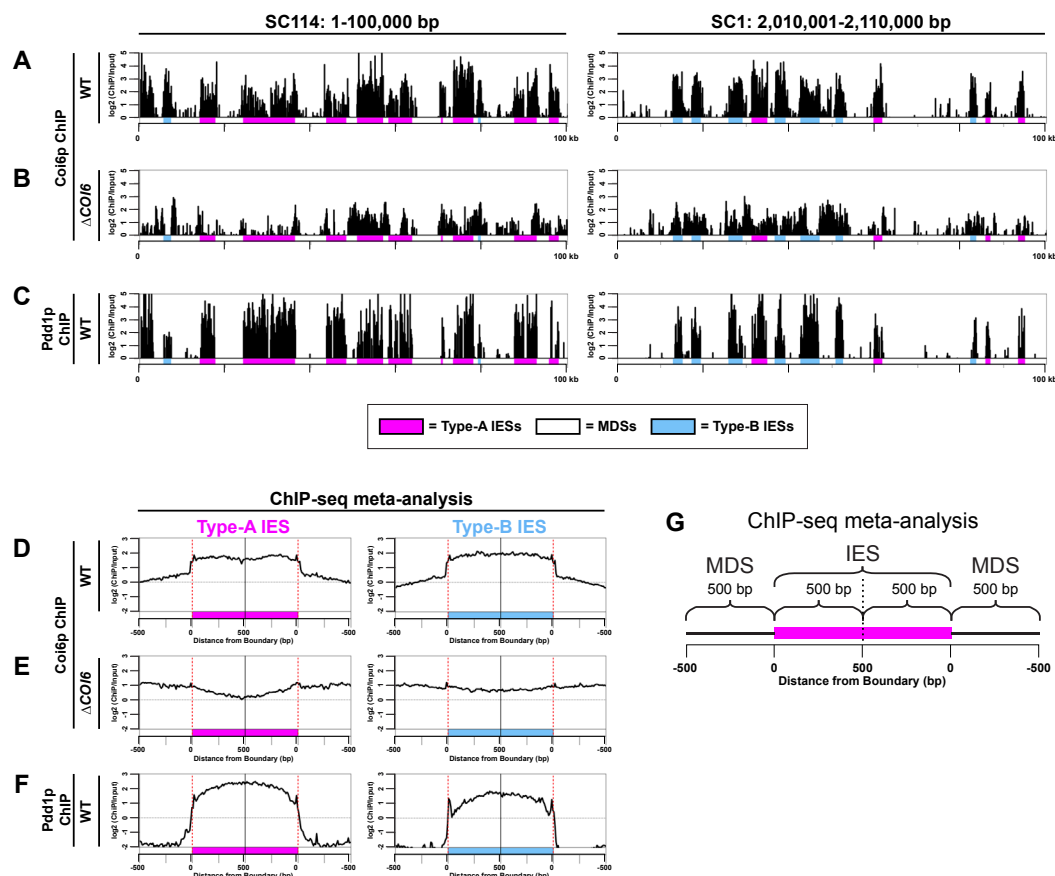


Figure 16: Coi6p binds to Type-A and Type-B IESs

(A-C) Analysis of ChIP-seq data in two 100 kb genomic regions for Coi6p in WT and $\Delta COI6$ cells and Pdd1p in WT. The left panel shows a region mostly containing Type-A IESs (indicated by purple bars beneath the x-axis), while the right panel shows a region containing mostly Type-B IESs (indicated by light blue bars beneath the x-axis). (D-F) IES boundary-focused meta-analysis (as described in the text) of the same ChIP experiments as in (A-C). Type-A (purple) and Type-B (light blue) IESs were analyzed separately. (G) Schematic representation of the boundary-focused meta-analysis.

The findings outlined above clearly point to a direct role of Coi6p in the process of DNA elimination. Coi6p is expressed specifically during the time of DNA elimination, localizes to the right cellular compartment, the new MAC, and there co-localizes with Pdd1p in heterochromatin foci, which are essential for DNA elimination. Furthermore, Coi6p enriches on IESs in ChIP-seq in a genome-wide manner.

2.5 *Coi6p* is required for DNA elimination

To investigate the role of *Coi6p* in the process of programmed DNA elimination, investigation of the *COI6* knock-out strains ($\Delta COI6$) used in our previous study were extended (Woehrer et al. 2015).

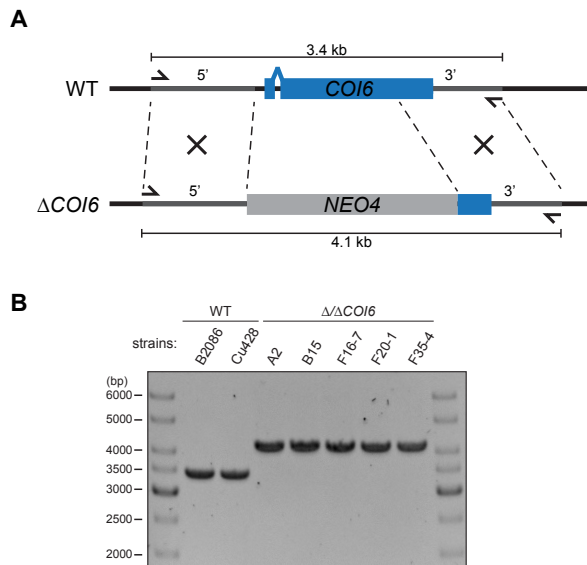


Figure 17: Verification of the genotype of $\Delta COI6$ strains

(A) Schematic depiction of the *COI6* locus in WT cells (top) and after gene knock-out by homologous recombination with the *NEO4* cassette (bottom). (B) PCR analysis of the *COI6* locus in WT and $\Delta COI6$ strains showing complete replacement of all WT *COI6* copies in the $\Delta COI6$ cells.

To confirm once more the $\Delta COI6$ strains utilized in our previous study the *COI6* locus was amplified by PCR from either wild-type (WT) strains or $\Delta COI6$ strains. While the WT locus yielded a product of approximately 3.4 kb, in $\Delta COI6$ strains a product of approximately 4.1 kb was amplified. This demonstrates clearly that all copies of the *COI6* locus in both MIC and MAC had been replaced by the *NEO4* drug resistance marker, thus verifying that the $\Delta COI6$ strains are complete KOs (Fig. 17A and B).

Subsequently, a close examination of the DNA elimination defect previously reported was undertaken, using a DNA FISH probe complementary to the repeated *Tlr1* IES and WT, $\Delta COI6$ or $\Delta PDD1$ cells at 34 hpm (Woehrer et al. 2015; Wuitschick et al. 2002). During the analysis of the FISH staining particular attention was paid to partial elimination phenotypes of the *Tlr1* element, which are demarked by areas in the new MAC that contain DAPI stained DNA, yet lack *Tlr1* signal (Fig. 18A, two green middle panels). As expected, in almost all cases (98.5%) WT cells have eliminated *Tlr1* from their new MACs and retain only *Tlr1* signal in the MIC, whereas 100% of $\Delta PDD1$ cells show a complete block of *Tlr1* elimination, yielding a homogenous *Tlr1* staining in the whole new MACs (Fig. 18A, right panel, 18B, $\Delta PDD1$). In contrast, while most $\Delta COI6$ cells (~ 84%) showed

homogenous Tlr1 staining similar to $\Delta PDD1$ cells, a significant population ($\sim 15\%$) showed only partial Tlr1 signal and a small fraction of the cells eliminated Tlr1 completely ($\sim 1\%$) (Fig. 18B, $\Delta COI6$). Thus, although Coi6p is required for DNA elimination in most cases, its absence does not completely inhibit DNA elimination.

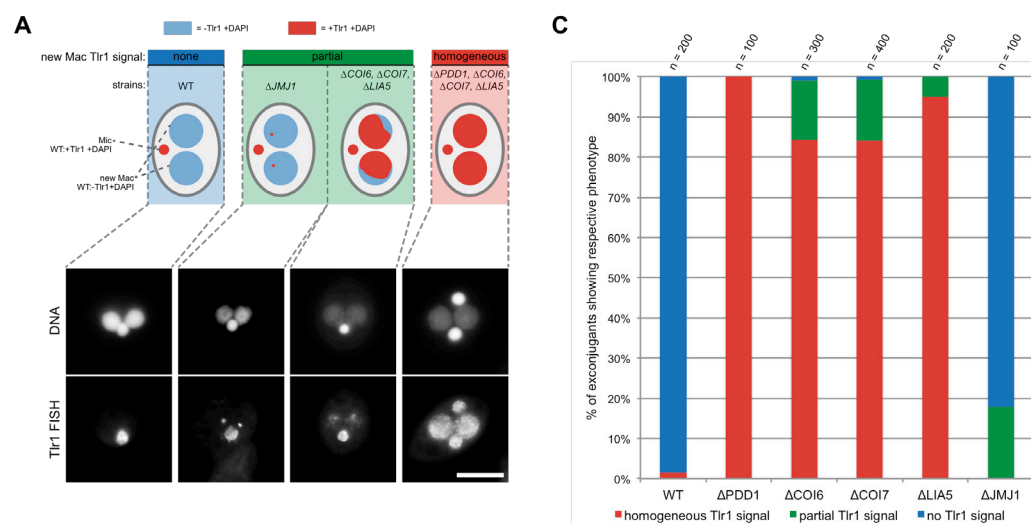


Figure 18: Elimination of the Tlr1 IESs is inhibited in absence of Coi6p

(A) Schematic representation and DNA FISH stainings showing the different Tlr1 FISH phenotypes observed in *Tetrahymena* cells at 34 hpm. All cells show Tlr1 signal (red) in the MIC. WT cells eliminate all Tlr1 from their new MACs (left). Some $\Delta JMJ1$ cells show few discrete foci of Tlr1 in each new MAC (2nd from left). In $\Delta COI6$, $\Delta COI7$ and $\Delta LIA5$ some cells can partially remove Tlr1 from new MACs leaving nuclear areas devoid of Tlr1 signal, but positive for DAPI (blue, 2nd from right). All KO cells, except $\Delta JMJ1$ cells, show a significant number of cells not eliminating any Tlr1 (right). (B) Quantification of phenotypes depicted in (A) for WT and all KO cells. The scale bar indicates 10 μ m.

2.6 Chromo- and chromoshadow domain are essential for Coi6p function

Although the DNA elimination defect of $\Delta COI6$ cells was observed in experiments with independent KO strains, final proof of the specificity of the KO phenotype may only be achieved by rescuing it by introducing WT *COI6* into $\Delta COI6$ cells. Furthermore, it was of interest to what degree the two conserved domains of Coi6p, the chromo- and the chromoshadow domain, are important for its function.

In order to establish a rescue system that allows asking such questions, a strategy was selected that relied on introduction of a transgene into the non-essential *BTU1* (*beta-tubulin 1*) locus in the parental MAC of $\Delta COI6$ cells (Fig. 19A). The transgene contained N-terminally HA-tagged WT or mutant *COI6*, which was placed under the control of the Cadmium- (Cd) inducible *MTT1* promoter. In order to assess the importance of the Coi6p

chromodomain methyl-lysine binding activity, the aromatic cage was mutated by alanine substitutions (Y39A, W60A, Y63A; designated *Coi6p_CDmut*). For testing relevance of the chromoshadow domain two mutants were created with alanine substitutions of highly conserved amino acids (F438A, K439A designated *Coi6p_CSDmut1* and Y468A, Y469A, E470A labeled *Coi6p_CSDmut2*) (Fig. 19B).

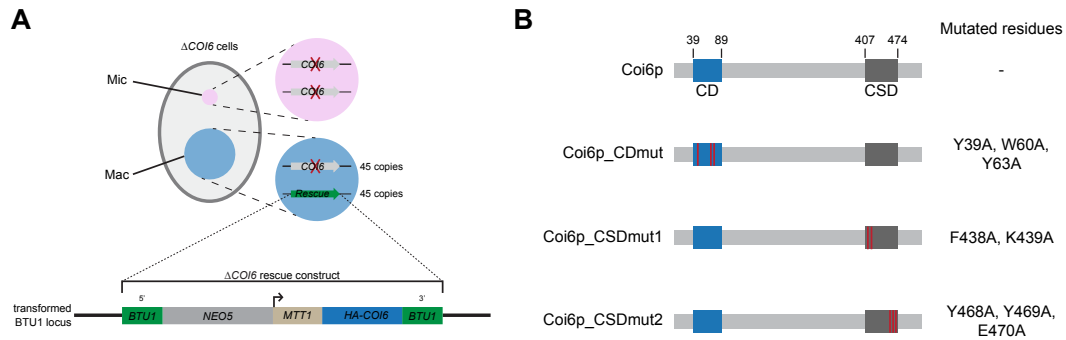


Figure 19: Schematic representation of the $\Delta COI6$ rescue system

(A) $\Delta COI6$ rescue cells are created by transforming a construct containing the *NEO5* drug resistance cassette and a *COI6* expression cassette into the non-essential parental MAC *BTU1* (Beta-Tubulin 1) locus. HA-tagged *COI6* expression was driven by the Cadmium-inducible *MTT1* (metallothionein 1) promoter. (B) Schematic depiction of *COI6* rescue constructs used in this study. Point mutations targeted either chromodomain (CD) or chromoshadow domain (CSD) of *Coi6p* and their precise positions are indicated in the right panel, while red strokes indicate their approximate position in the scheme.

The transgenes were expressed successfully and could localize to the new MAC (Fig. 20A and B). Capability of the transgenes to rescue the $\Delta COI6$ DNA elimination defect was then assessed by performing DNA FISH against the *Tlr1* IESs. For simplicity of analysis cells retaining partial or full *Tlr1* signal in their new MACs were collectively designated “not rescued” (Fig. 21A). Notably, even in non-induced cells small amounts of WT HA-*Coi6p* were expressed due to leakiness of the *MTT1* promoter and this basal *Coi6p* expression was sufficient to rescue the $\Delta COI6$ phenotype in 72% of cells as revealed by *Tlr1* FISH (Fig. 20B). Upon induction of WT HA-*Coi6p* this percentage increased to 93.5% of cells, thus DNA elimination was restored to almost WT strain levels. In stark contrast to this none of the mutant *COI6* transgenes could rescue *Tlr1* elimination efficiently (Fig. 20B; *COI6_CDmut*: 10 %, *COI6_CSDmut1*: 4.8%, *COI6_CSDmut2*: 2.8%). Indicating that both chromo- and chromoshadow domain are of crucial importance for *Coi6p* function. Moreover, as mutations in the chromodomain of *Coi6p* were designed in a way that they would interfere with presumed *Coi6p* H3K9/K27me binding activity, this implies that *Coi6p* likely requires such binding to perform its molecular function.

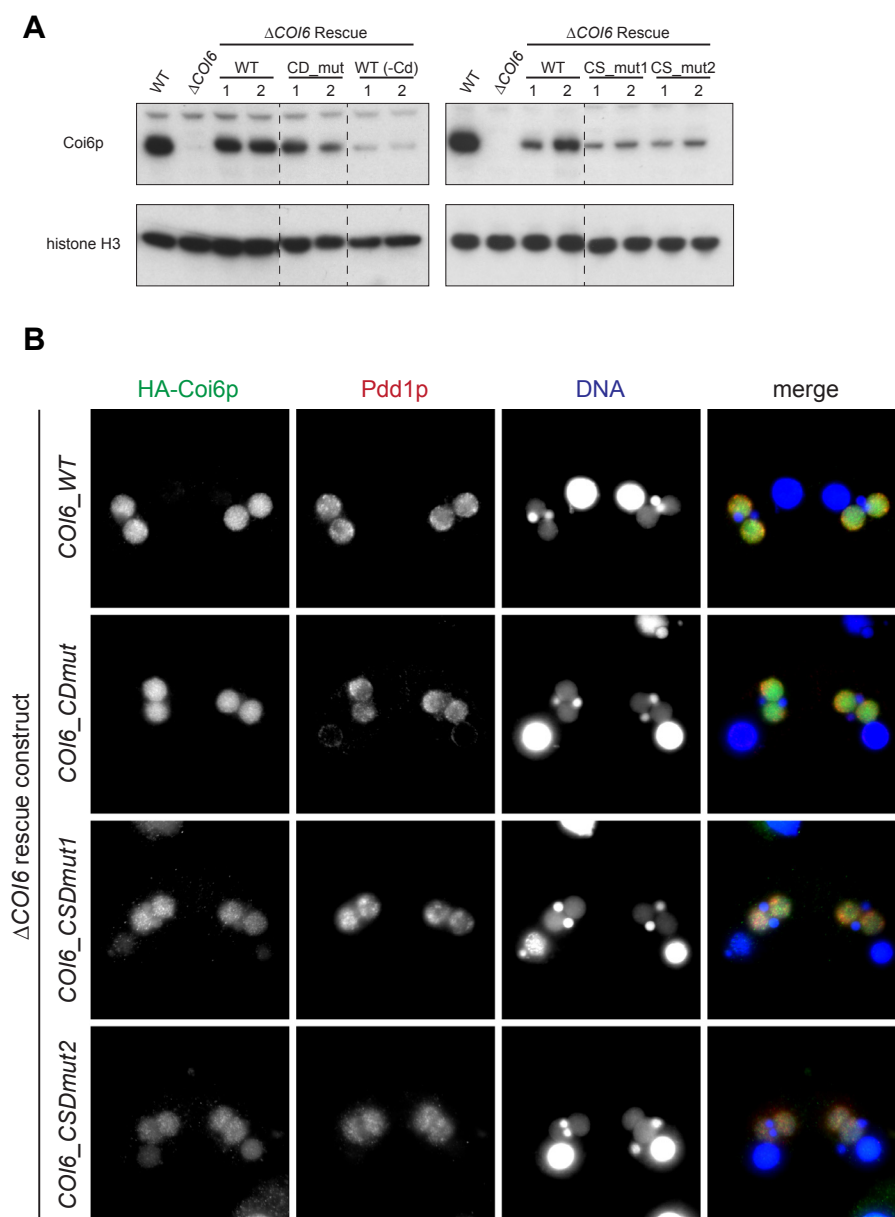


Figure 20: Expression and localization of HA-Coi6p in Δ COI6 rescue system strains

(A) Expression of HA-Coi6p in all Δ COI6 rescue system strains was tested by western blotting using α -Coi6p(r) antibody and α -histone H3 antibody (as loading control) at 8 hpm. Rescue experiments were performed in duplicates (indicated by numbers: 1 and 2). (B) Nuclear localization of HA-Coi6p was tested by IF using α -HA antibody (mouse) and α -Pdd1p antibody (abcam) as staining control at 8hpm. Strain are designated as described in the text and in Fig. 19.

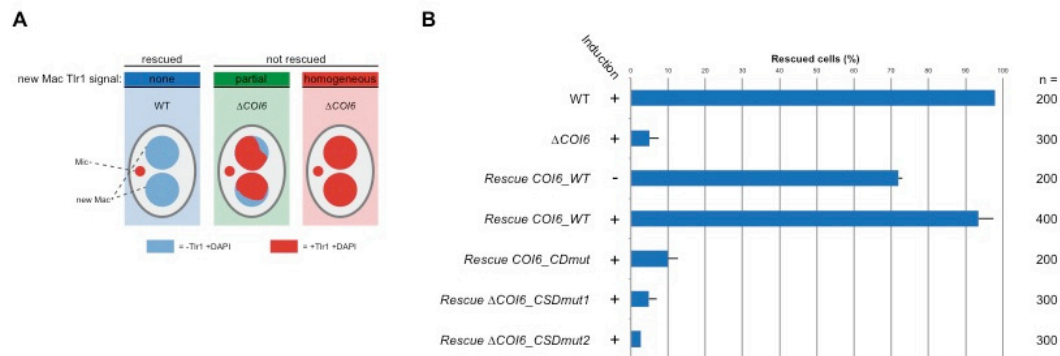


Figure 21: Tlr1 elimination in the $\Delta COI6$ rescue system

(A) Schematic depiction of the different Tlr1 DNA FISH phenotypes observed in $COI6p$ rescue strains at 34 hpm. All cells show Tlr1 signal (red) in the MIC. WT cells eliminate all Tlr1 from their new MACs (left). In $\Delta COI6$ rescue strains some cells can partially remove Tlr1 from new MACs leaving nuclear areas devoid of Tlr1 signal, but positive for DAPI (blue, middle), however most strains show a significant number of cells that do not eliminate any Tlr1 (right). Partial and homogeneous Tlr1 signal in the new MAC are summarized as “not rescued.” (B) Quantification of Tlr1 DNA FISH experiments using $COI6p$ rescue strains and WT and $\Delta COI6$ cells as control. Shown is the percentage of cells that were rescued and therefore show no Tlr1 FISH signal in their new MACs.

2.7 Heterochromatin components accumulate in absence of $COI6p$

$COI6p$ localizes to IESs and is part of the heterochromatin machinery necessary to eliminate IESs. Thus, one imminent question was whether $COI6p$ was directly or indirectly regulating other heterochromatin components that are crucial for DNA elimination, for instance Pdd1p and the histone methylations H3K9me and H3K27me. Such regulation might explain the defect observed in $\Delta COI6$ cells.

To investigate this scenario IF stainings were performed at the onset of heterochromatin formation (8 hpm) with WT or $\Delta COI6$ cells using antibodies detecting Pdd1p and the histone methylations of H3K9 and K27. No difference between WT and $\Delta COI6$ cells was detected, indicating that the crucial heterochromatin components Pdd1p, H3K9me3 and H3K27me3 accumulate normally in the new MACs in the absence of $COI6p$ (Fig. 22A and B, compare WT and $\Delta COI6$). These results indicate that the general heterochromatin machinery, including Pdd1p and the Ezl1p complex, which mediates histone methylations of H3K9/K27, localizes correctly to the new MAC in $\Delta COI6$ cells. Furthermore, enzymatic activity of the Ezl1p complex does not seem inhibited in the absence of $COI6p$, showing that $COI6p$ likely plays a role downstream or in parallel of H3K9/K27 methylation and accumulation of Pdd1p.

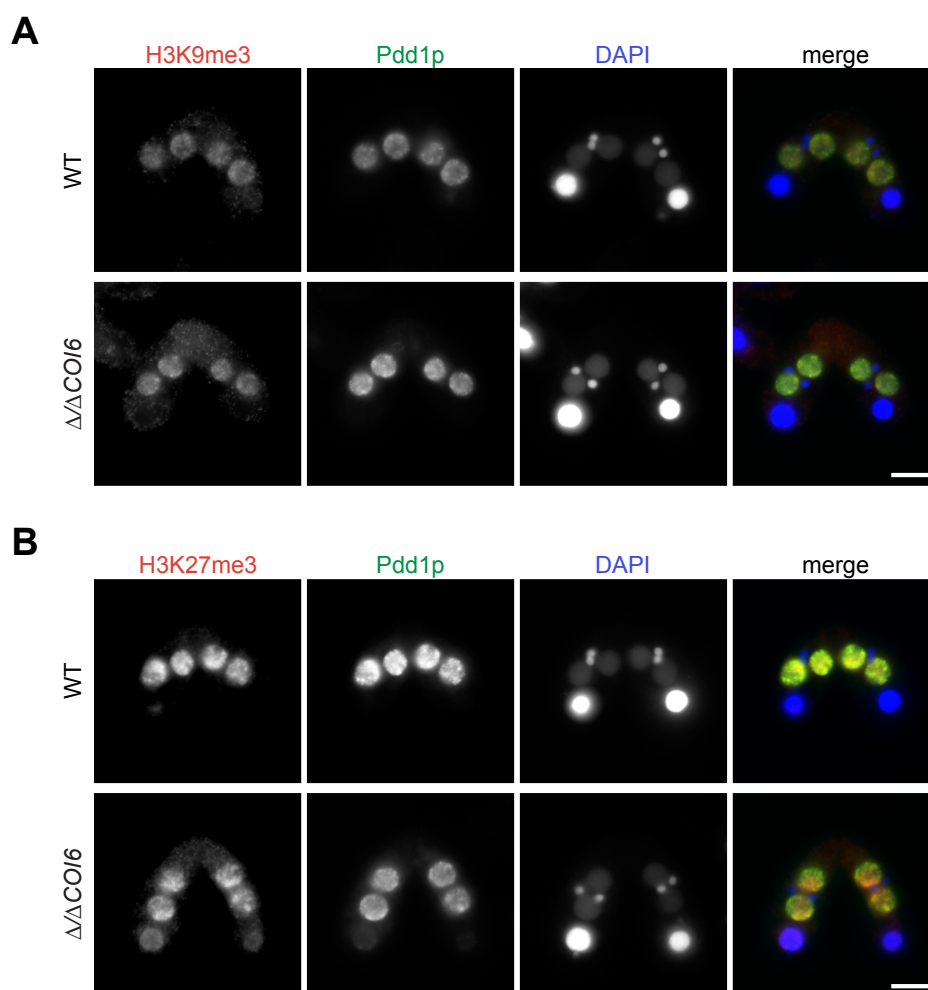


Figure 22: Lysine-methylations of histone H3K9 and K27 are established in absence of *Coi6p*
 (A) Localization and abundance of H3K9me3 and Pdd1p was tested by IF using α -H3K9me3 and α -Pdd1p (guinea pig) antibodies in WT and ΔCOI6 cells at 8 hpm. (B) same as (A) but using α -H3K27me3 antibody. The scale bar indicates 10 μm .

2.8 Pdd1p and H3K9/K27me3 spread outside IES boundaries in ΔCOI6 cells

As described above heterochromatin formation in ΔCOI6 cells seemed normal based on IF observations. However, IF stainings do not have sufficient resolution to assess heterochromatin formation on the level of single IESs and genomic loci. Therefore ChIP-seq analysis of Pdd1p, H3K9me3 and H3K27me3 was carried out in ΔCOI6 cells at 12 hpm (Fig. 23 and 24).

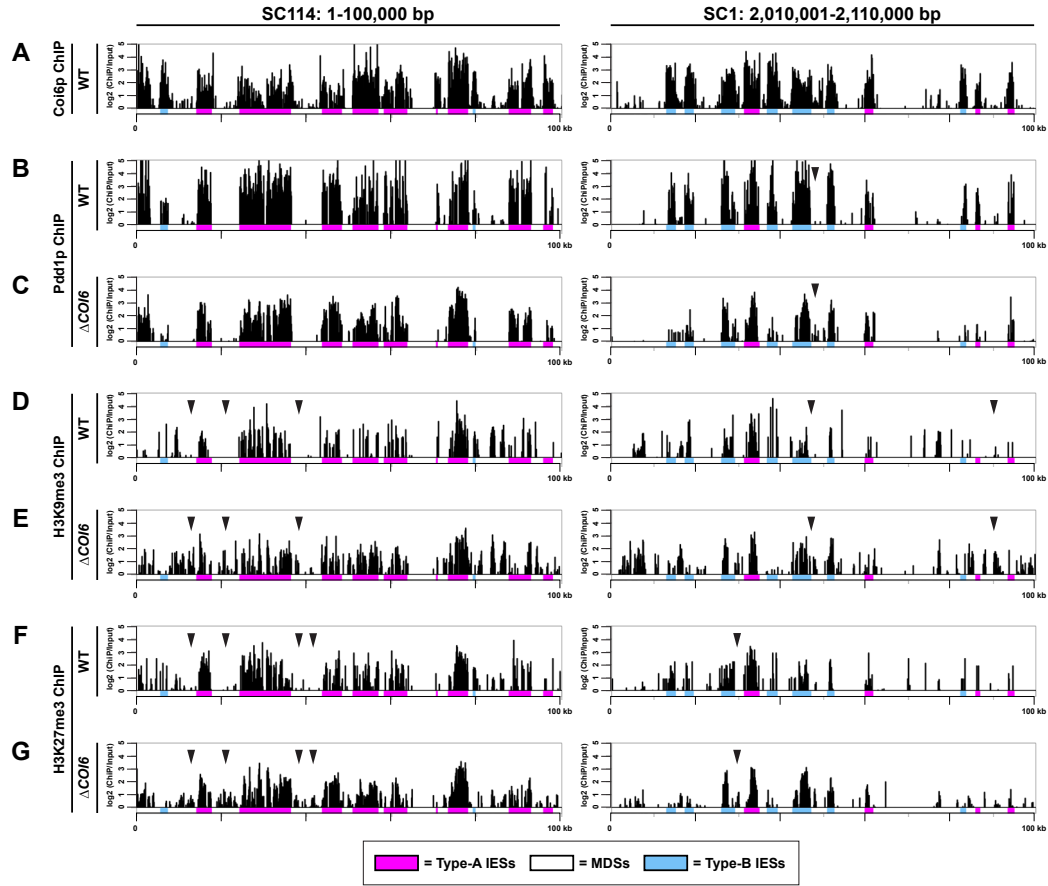


Figure 23: Heterochromatin spreading in absence of Coi6p

ChIP-seq analysis of Pdd1p, H3K9me3 and H3K27me3 in WT and $\Delta COI6$ cells. (A) For comparison purposes Coi6p ChIP-seq in WT cells is shown. (B and C) ChIP-seq analysis of Pdd1p in WT and $\Delta COI6$ cells, respectively. (D and E) ChIP-seq analysis of H3K9me3 in WT and $\Delta COI6$ cells, respectively. (F and G) ChIP-seq analysis of H3K27me3 in WT and $\Delta COI6$ cells, respectively. For each experiment two genomic 100 kb loci are shown containing either predominantly Type-A IESs (left) or Type-B IESs (right). The Y-axis depicts ChIP results as $\log_2(\text{ChIP}/\text{Input})$. Arrowheads indicate positions of heterochromatin spreading.

These experiments indicate that Pdd1p was enriched on Type-A and Type-B IESs in absence of Coi6p, albeit to slightly lesser degree when compared to WT (Fig. 23, compare IESs signal in B and C). Upon close inspection of the Pdd1p ChIP-seq data, it became apparent that in $\Delta COI6$ cells Pdd1p bound to IESs in a broader fashion when compared to WT profiles in the meta-analysis, although the overall level of binding was slightly reduced (Fig. 24B and C). This trend was clearly observed for Type-A IESs in $\Delta COI6$ cells, where Pdd1p binding continued an approximate 250 bp outside the IES boundaries compared to about 125 bp in WT cells (Fig. 24B and C, compare MDS region enrichment). For Type-B IESs the same tendency could not be detected in meta-analysis,

maybe due to sensitivity of the ChIP-seq experiment and generally lower signal intensity on Type-B IESs. Yet, on individual loci we could identify instances of Pdd1p ChIP-seq signals outside IES boundaries in $\Delta COI6$ cells, but not WT cells, for both types of IESs (Fig. 23B and C arrowheads and Fig. 28, lowest two rows in affected loci). As Pdd1p binds to both H3K9me3 and H3K27me3, I assumed that Pdd1p would merely follow the pattern of these modifications. However, ChIP-seq analysis of H3K9me3 and H3K27me3 patterns in WT or $\Delta COI6$ cells revealed that the histone modifications were much more strongly affected than anticipated. While in WT cells both modifications enrich specifically on Type-A and Type-B IESs, in $\Delta COI6$ cells their levels on IESs are generally reduced, concomitant with a strong increase of levels in MDS regions (Fig. 24 compare D and E; F and G). Furthermore, this effect is not only observed in meta-analysis, but also clearly in individual genomic locations (Fig. 23, compare D and E, F and G, see regions marked with arrowheads). All together, this indicates that Coi6p is not required for heterochromatin formation *per se*, but is required for the precise accumulation of the heterochromatin components on IESs. The fact that the precise IES localization of the heterochromatic histone modifications H3K9/K27me3 is more sensitive to loss of Coi6p than Pdd1p might indicate that mechanisms other than histone modifications also facilitate Pdd1p binding to IESs.

As IESs are the primary target of heterochromatin formation, I interpret the lack of precision of this process as expansion of heterochromatin from IESs to MDS regions, and will henceforth refer to this phenomenon as heterochromatin spreading.

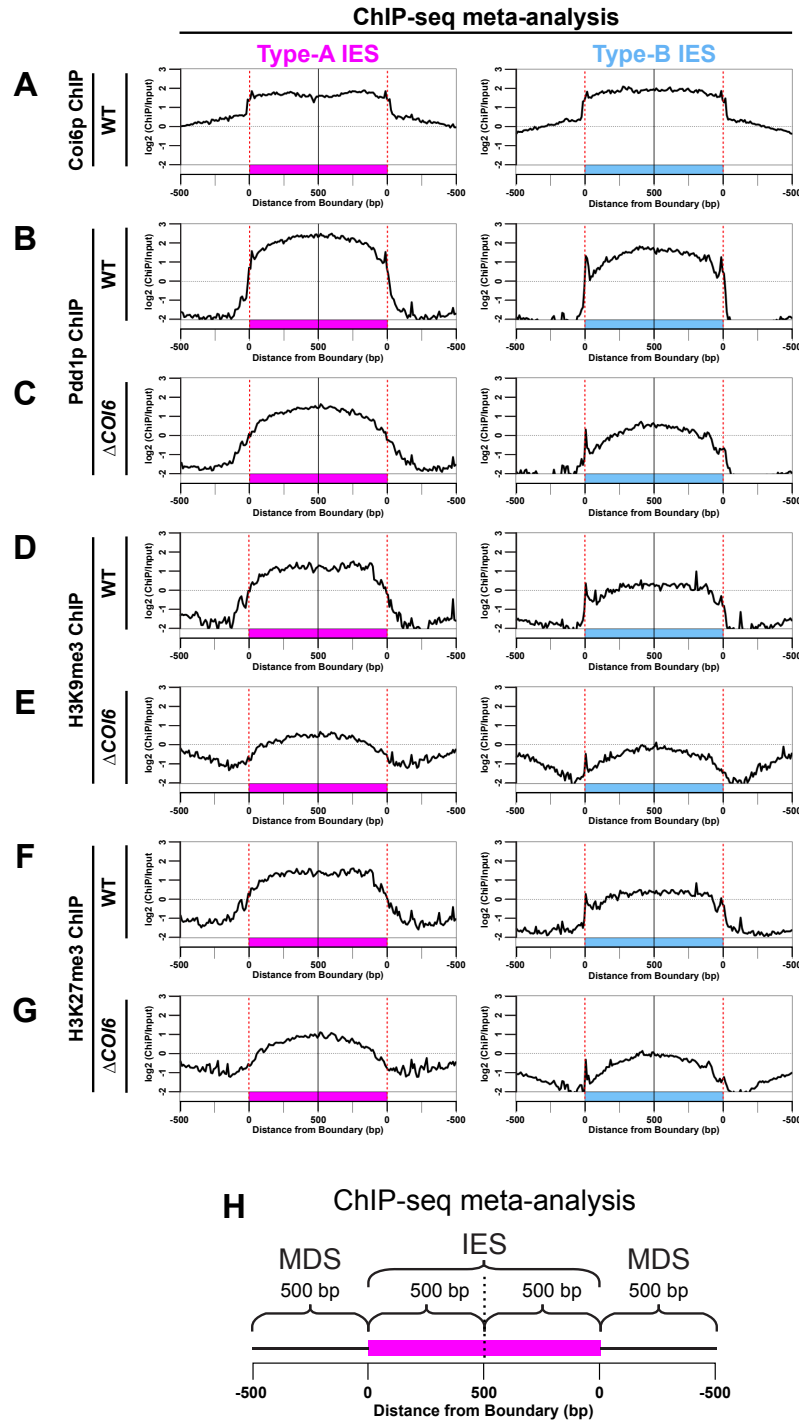


Figure 24: Meta-analysis of heterochromatin spreading in Δ COI6 cells

IES boundary-focused meta-analysis of Pdd1p, H3K9me3 and H3K27me3 ChIP-seq in WT and Δ COI6 cells as depicted in (H). (A) For comparison purposes Col6p ChIP-seq in WT cells is shown. (B and C) ChIP-seq analysis of Pdd1p in WT and Δ COI6 cells, respectively. (D and E) ChIP-seq analysis of H3K9me3 in WT and Δ COI6 cells, respectively. (F and G). ChIP-seq analysis of H3K27me3 in WT and Δ COI6 cells, respectively. Right and left panel show meta-analysis for either Type-A or Type-B IESs, respectively). All Y-axes show $\log_2(\text{ChIP}/\text{Input})$ values.

2.9 Heterochromatin foci formation depends on *Coi6p*

During new MAC development Pdd1p first localizes homogeneously in the nucleus at 8 - 10 hpm, yet then forms large foci at 12 - 14 hpm, which disappear by 16 hpm (Fig. 15A and B) (Madireddi et al. 1994; Smothers et al. 1997). The heterochromatin foci are thought to be important for DNA elimination, although their precise function has so far remained elusive.

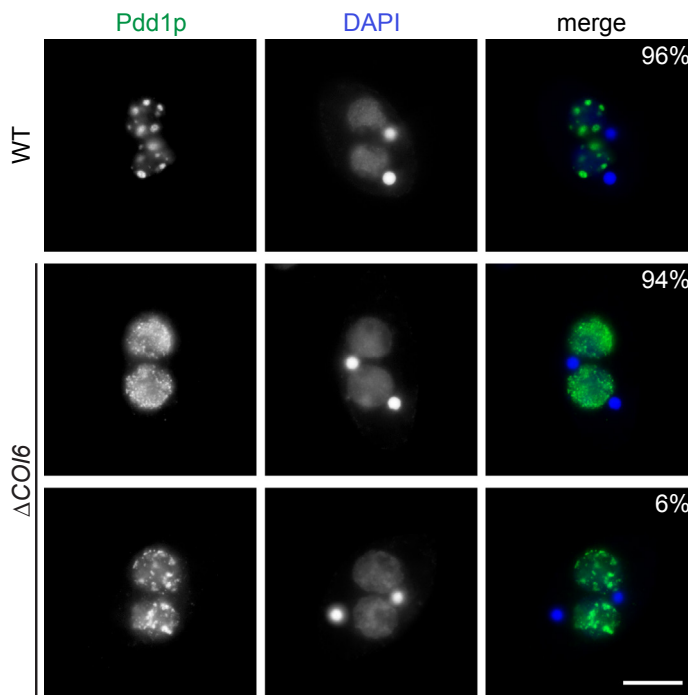


Figure 25: Block of Pdd1p heterochromatin foci formation in $\Delta COI6$ cells

Pdd1p foci formation was assessed by IF staining using commercial α -Pdd1p antibody in WT or $\Delta COI6$ cells at 14 hpm. The top two panels show the typical phenotype observed in WT or $\Delta COI6$, respectively. The lowest panel shows atypical foci rarely detected in $\Delta COI6$ cells. Percentages indicated the frequency of the respective phenotype. The scale bar indicates 10 μ m.

The formation of heterochromatin foci was tested in $\Delta COI6$ cells by localizing Pdd1p in immune-fluorescent stainings at 14 hpm. At this time point 96% of all wild-type cells have heterochromatin foci (Fig. 25, upper row), yet in $\Delta COI6$ cells 94% of cells showed homogeneous Pdd1p staining at 14 hpm (Fig. 25, middle row) and 6% of cells showed aberrant foci with unusual morphology (Fig. 25, lower row). So despite relatively moderate spreading of Pdd1p outside IESs in absence of *Coi6p* downstream Pdd1p foci formation is severely affected.

Although it is tempting to speculate that this block of Pdd1p foci formation is directly caused by heterochromatin spreading of H3K9/27me₃, Pdd1p and maybe other components, observation of similar phenotypes in KO of other heterochromatin localizing proteins, e.g. Lia5p, Lia4p, Drb2p and Lia1p, might indicate that this is rather a

general downstream phenotype detected when homeostasis of heterochromatin is disturbed (Horrell and Chalker 2014; Motl and Chalker 2011; Rexer and Chalker 2007).

2.10 *Coi6p* binding to IESs depends on *scnRNAs*, yet not on *Pdd1p*

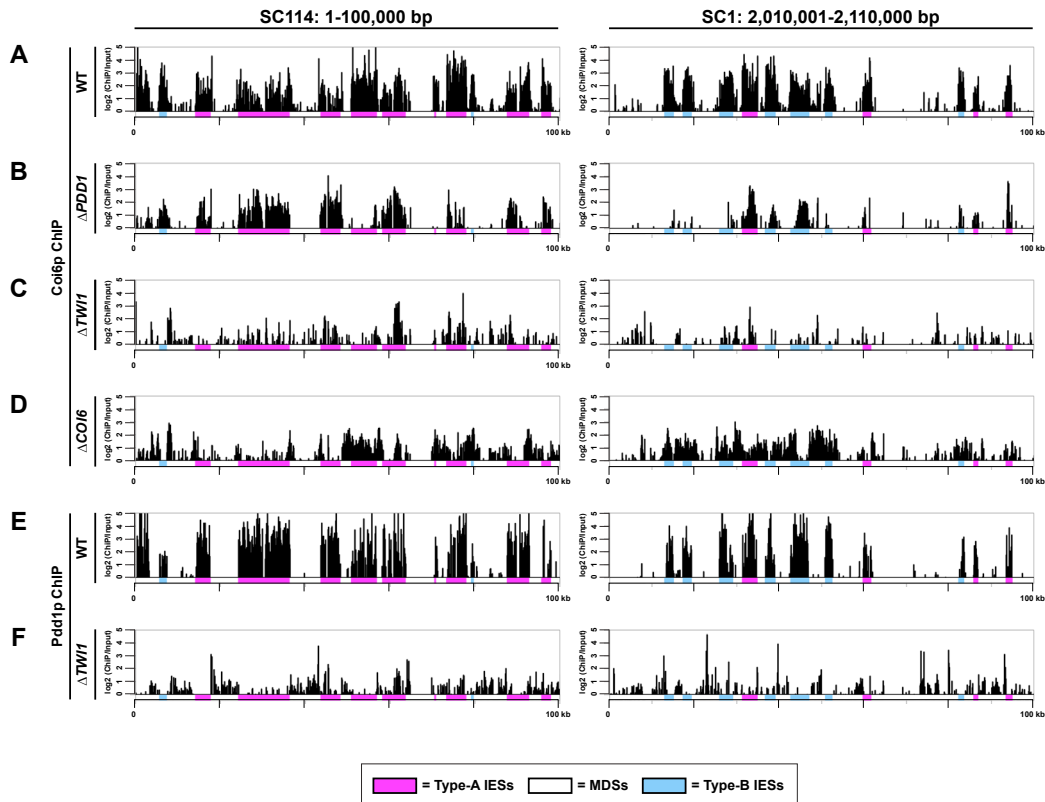


Figure 26: *Coi6p* IES binding depends on *Tw1p* and *scnRNAs*, but not on *Pdd1p*

(A, B, C and D) *Coi6p* ChIP-seq in WT, $\Delta PDD1$, $\Delta TWI1$ and $\Delta COI6$ cells. (E and F) *Pdd1p* ChIP-seq in WT and $\Delta TWI1$ cells. ChIP signal is depicted as $\log_2(\text{ChIP}/\text{Input})$. For each experiment two genomic 100 kb windows are shown containing either majorly Type-A (purple) or Type-B (light blue) IESs.

In absence of *Coi6p* heterochromatin is established, yet spreads outside IES boundaries (Fig. 23 and 24), indicating that *Coi6p* must act in a parallel or downstream pathway to *Pdd1p*. To shed further light on the relationship of these two HP1-like proteins, firstly it was tested whether *Coi6p* binding to IESs depends on the presence of *scnRNAs*. To that end ChIP-seq analysis of *Coi6p* and *Pdd1p* was performed in ΔTWI cells. These cells lack all *scnRNAs*, as Early-*scnRNAs* are unstable when not loaded into the Argonaute protein *Tw1p* and Late-*scnRNAs* require Early-*scnRNAs* for their induction (Mochizuki et al. 2002, Noto et al. 2015). Furthermore, methylations of H3K9 and K27 are not

established in the new MAC of ΔTWI cells, presumably because the HKMT Ezi1p is recruited to IESs in a scnRNA-dependent manner (Liu et al. 2004; Liu et al. 2007). Consistent with the loss of H3K9/K27me3 in the absence of Twi1p, ChIP-seq for Pdd1p fails to enrich on IESs, regardless of whether they are Type-A or Type-B (compare IESs regions in Fig. 26E and F and 27E and F). Moreover, Coi6p is not strongly enriched on Type-A and Type-B IESs in ΔTWI cells, although a minute amount of Coi6p might still bind to Type-A IESs for unknown reasons (compare IESs regions in Fig. 26A and C, 27A and C). These results indicate that both Coi6p and Pdd1p depend on Twi1p-incorporated Early-scnRNAs in order to bind to IESs.

Late-scnRNAs require heterochromatin for their production, in particular the histone H3K9/K27 methyltransferase Ezi1p and Pdd1p (Noto et al. 2015). Therefore, $\Delta PDD1$ cells can be used as a tool to study the dependency of Coi6p IES binding on Late-scnRNAs. ChIP-seq analysis of Coi6p in $\Delta PDD1$ cells showed that Coi6p was still enriched on Type-A IESs, which are targeted by Pdd1p-independent Early-scnRNAs (compare Type-A IESs in Fig. 26A and B; 27A and B). Importantly, this demonstrates that the presence of Pdd1p itself is dispensable for the ability of Coi6p to bind to IESs. However, contrary to the situation on Type-A IESs, Coi6p enrichment on Type-B IESs was markedly reduced in $\Delta PDD1$ cells, demonstrating that Coi6p requires Pdd1p-dependent Late-scnRNAs for targeting to Type-B IESs (compare Type-B IESs in Fig. 26A and B; 27A and B). The residual enrichment of Coi6p on Type-B IESs observed in $\Delta PDD1$ cells might be explained by the previously-reported *trans*-targeting of Type-B IESs by Type-A derived scnRNAs (Noto et al. 2015). Thus in general targeting of Coi6p to IESs depends on scnRNAs, but does not require Pdd1p.

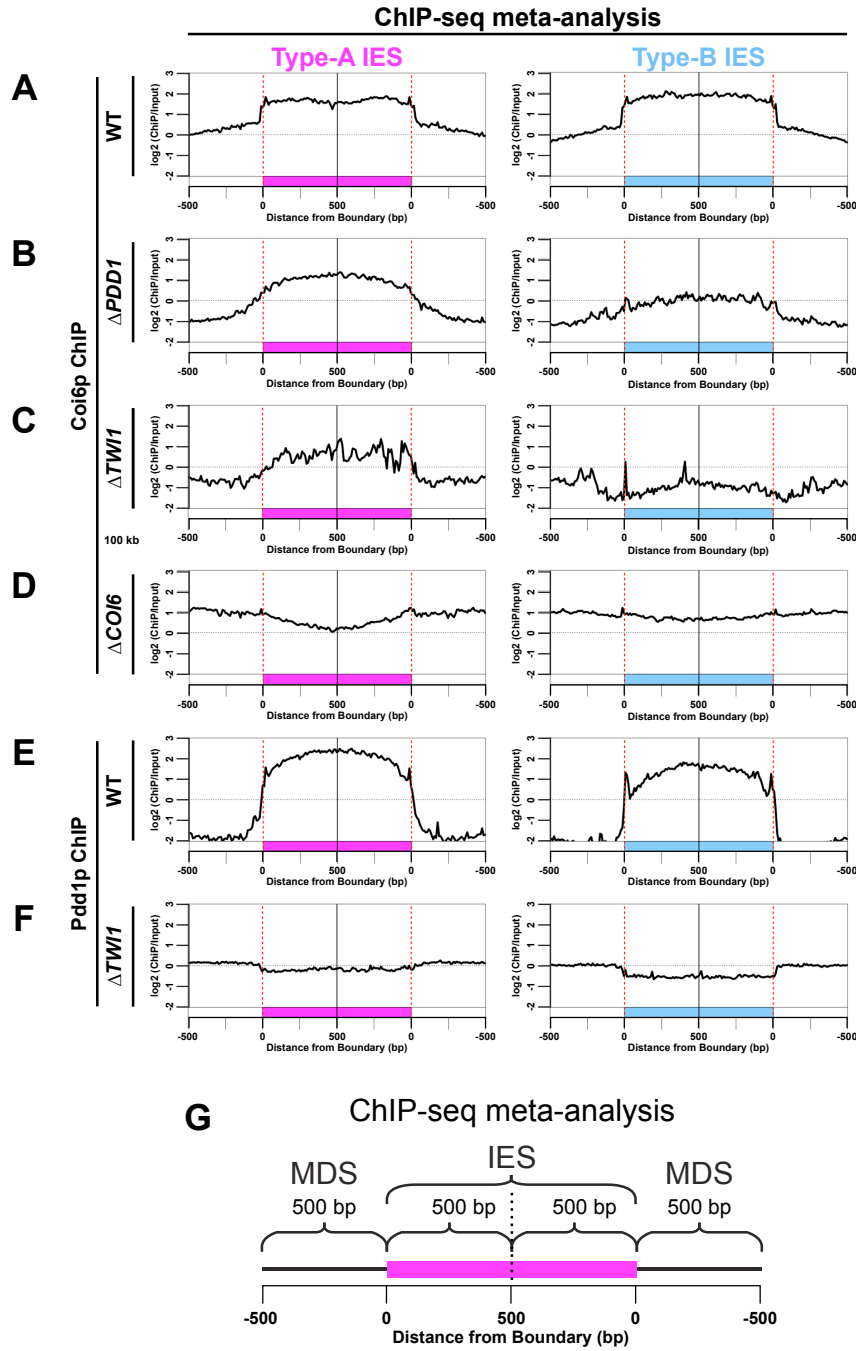


Figure 27: Meta-analysis - Coi6p IES binding depends on Twi1p and scnRNAs, but not on Pdd1p

(A, B, C and D) IES boundary-focused meta-analysis of Coi6p ChIP-seq in WT, $\Delta PDD1$, $\Delta TWI1$ and $\Delta COI6$ cells, respectively. (E and F) Pdd1p ChIP-seq in WT and $\Delta TWI1$ cells. Analysis was performed as described above and depicted in (H). Right and left panel show meta-analysis for Type-A (purple) or Type-B (light blue) IESs, respectively. All Y-axes show $\log_2(\text{ChIP}/\text{Input})$ values.

2.11 Late-scnRNA production occurs outside IES boundaries in $\Delta COI6$ cells

Because heterochromatin formation and scnRNA production are interdependent, I next asked if Late-scnRNA production is also affected in the absence of *CoI6p*. ScnRNAs derived of conjugating WT or $\Delta COI6$ cells were subjected to deep sequencing (scnRNA-seq). The time points at which scnRNAs were sampled were chosen such that both MIC expressed Early-scnRNAs and new MAC derived Late-scnRNAs could be observed (3, 6, 8, 10.5 and 12 hpm). The first 25 nt of each scnRNA sequence were used for genome mapping, while allowing only unique reads to be mapped. Then, the distribution of scnRNAs was either analyzed for individual IES loci (adding plus and minus strand reads) or using a meta-analysis similar to the one previously described for ChIP-seq (zooming in at the most proximal 500 bp left and right of each IES boundary and analysing Type-A and Type-B IESs separately) (Fig. 28 and 29, respectively). In the meta-analysis plus and minus strand mapping scnRNAs were kept separately and are shown on top or bottom, respectively. For WT cells previous observations were successfully recapitulated: At 3 hpm Early-scnRNAs, expressed from the MIC, map to Type-A IESs and to MDS regions surrounding them, while Type-B IESs are almost completely devoid of scnRNAs (Fig. 28 and 29, 3 hpm, WT and $\Delta COI6$) (Noto et al. 2015). The Early-scnRNAs mapping to MDS regions are removed from the pool of scnRNAs by the process of scanning (marked by asterisks in Fig. 28 and 29, compare signal on MDSs of 3 vs. 6 and 8 hpm in both WT and $\Delta COI6$) (Schoeberl et al. 2012; Mochizuki and Gorovsky 2004a). And with the onset of new MAC development at 8 hpm and more pronounced at 10.5 hpm Late-scnRNAs appear that specifically map to Type-B IESs (Fig. 28 and 29; 8, 10.5 and 12 hpm, in WT and $\Delta COI6$) (Noto et al. 2015).

RESULTS

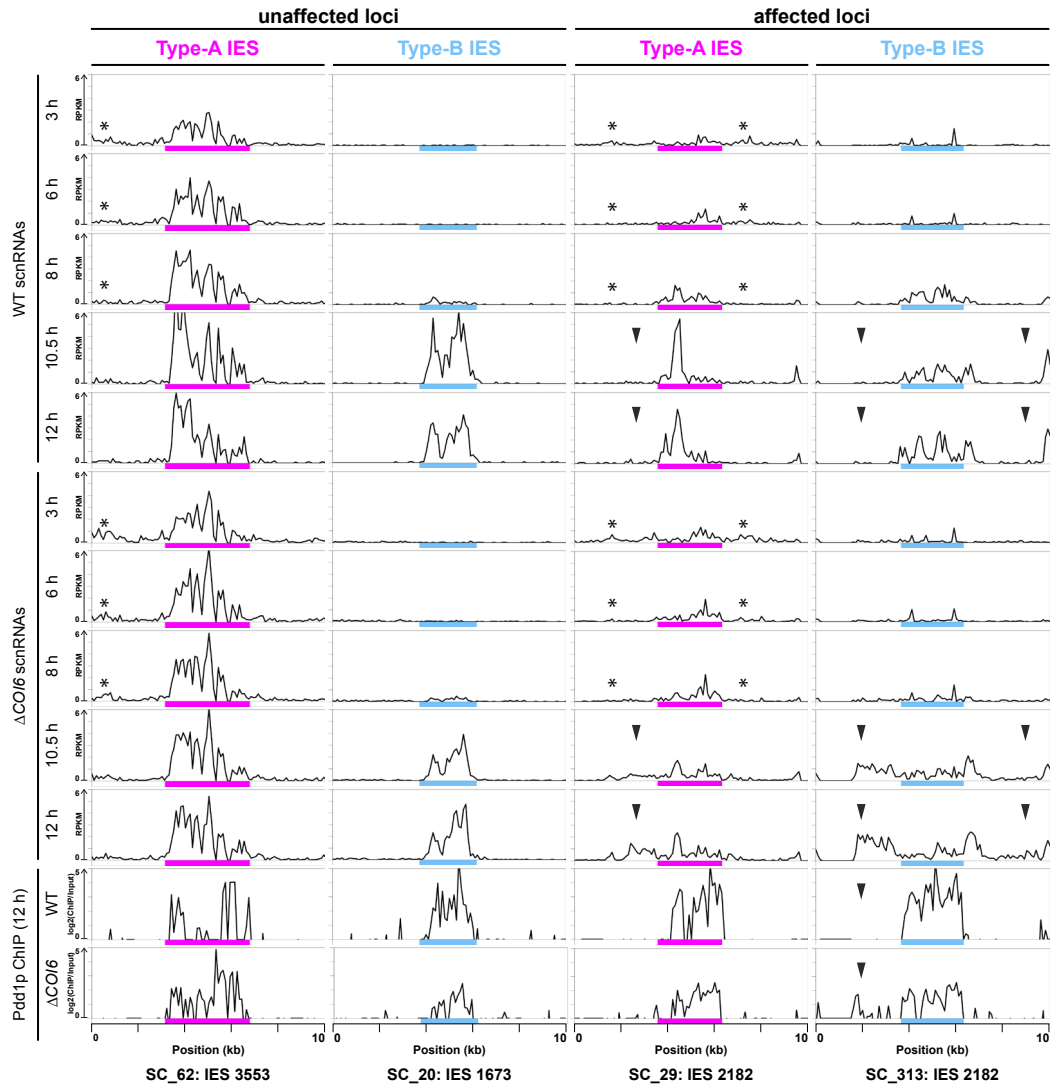


Figure 28: Loci unaffected and affected by scnRNA spreading in $\Delta COI6$ cells

For WT or $\Delta COI6$ cells scnRNAs-seq time course data (time point indicated by hpm) or Pdd1p ChIP-seq signal (lowest two rows) was analyzed for individual IES loci. In $\Delta COI6$ cells some IESs show scnRNAs levels similar to WT cells throughout conjugation (unaffected, left two panels), while others show spreading of Late-scnRNAs (affected, right two panels). One representative Type-A (purple) and Type-B (light blue) IES was chosen per category. The lowest two rows show Pdd1p ChIP seq enrichment on the respective loci in WT and $\Delta COI6$ cells. Regions where scanning of Early-scnRNAs is apparent are marked by asterisks, regions of Late-scnRNA spreading are marked by arrowheads.

Comparing the WT scnRNAs-seq profiles to those from $\Delta COI6$ cells, it was clear that both production and scanning of the Early-scnRNA pool functioned normally in the absence of *Coi6p* (Fig. 28 and 29; compare regions marked by asterisks in WT and $\Delta COI6$, pre-scanning Early-scnRNAs: 3 hpm, post-scanning Early-scnRNAs: 6 and 8 hpm). However, in meta-analysis of scnRNAs derived of $\Delta COI6$ cells at 10.5 hpm

scnRNAs were detected that map to the MDS regions surrounding both Type-A and Type B IESs (Fig. 29, $\Delta COI6$ 10.5 hpm, arrowheads). Furthermore, at 12 hpm the amount of scnRNAs detected on MDS regions increased considerably reaching levels comparable to pre-scanning 3 hpm scnRNAs for Type-A IESs. This indicates that spreading scnRNAs are products of a dynamic ongoing process (Fig. 29, compare $\Delta COI6$, 3 (asterisks) and 12 hpm (arrowheads)). The fact that these scnRNAs, which spread outside the IES boundaries, are detected exclusively at late conjugation time points (10.5 and 12 hpm) and are observed for both types of IESs, suggests that in absence of Coi6p Late-scnRNA production expands beyond IES boundaries. Moreover, analysis of scnRNAs-seq profiles of single loci revealed that this is not a ubiquitous phenomenon. While a large number of IESs show the described spreading effect, other IES loci seem to be unaffected (Fig. 28, compare affected and unaffected loci for $\Delta COI6$). Furthermore and as indicated by the meta-analysis, whether an individual IES exhibits Late-scnRNA spreading in absence of Coi6p is independent of its classification as Type-A or Type-B IES.

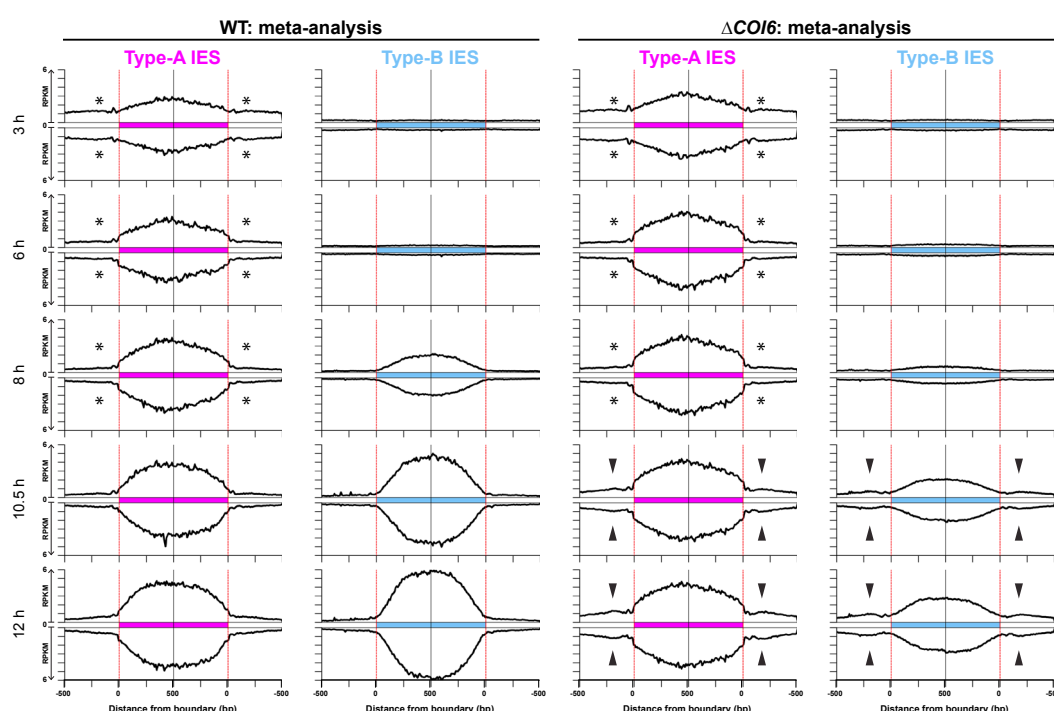


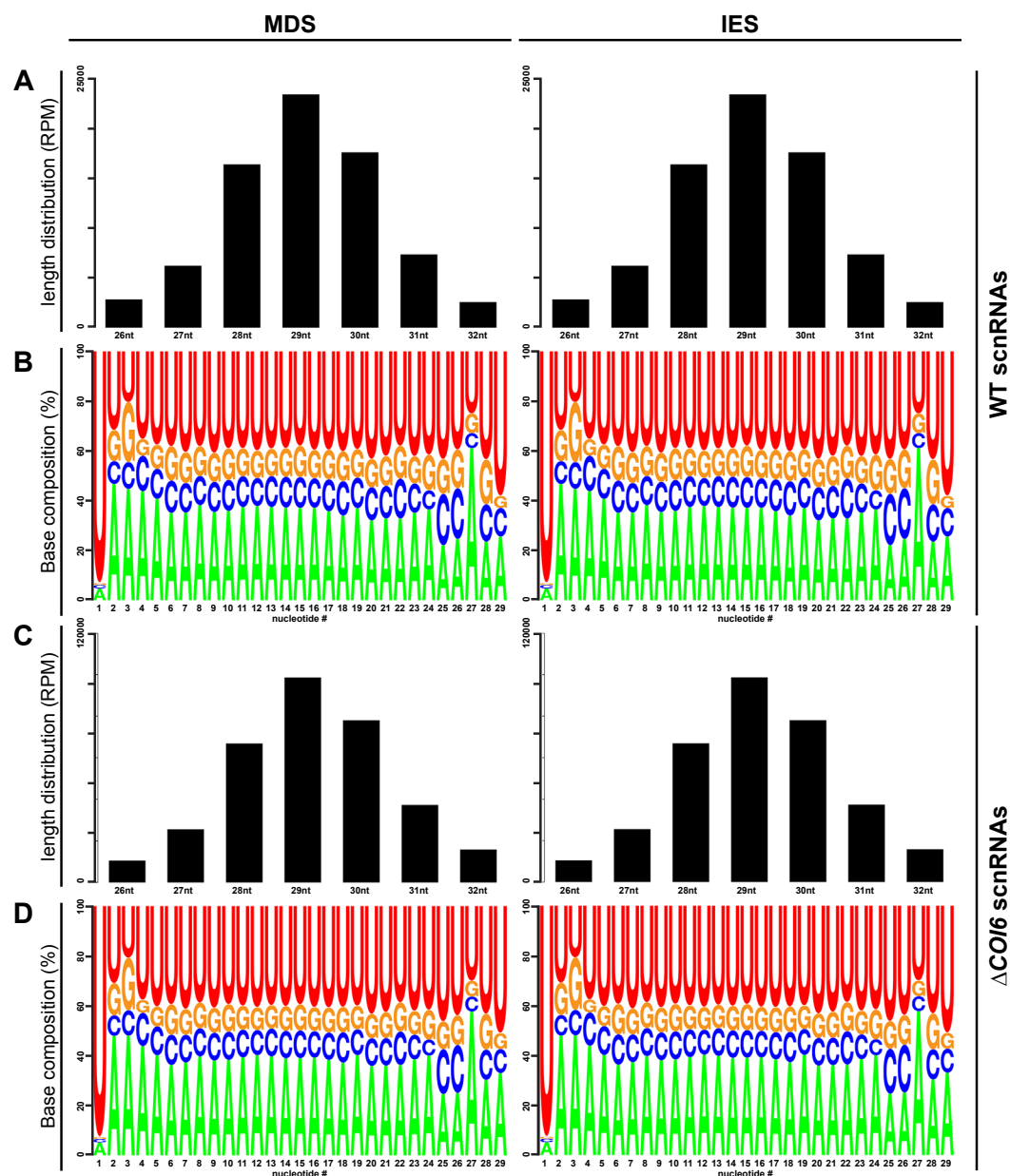
Figure 29: Meta-analysis of WT and in $\Delta COI6$ scnRNA-seq time courses

For each point in WT or in $\Delta COI6$ time courses (3, 6, 8, 10.5 and 12 hpm) scnRNAs-seq data was analyzed by boundary-focused meta-analysis. Type-A (purple) and Type-B (light blue) IESs were separated and only the regions around the two boundaries of each IES were analyzed (500 bp of MDS and IES, respectively) and compiled on one virtual Type-A or Type-B IES. (+) and (-) strand mapping scnRNAs are shown on top and bottom of each graph, respectively. Regions where scanning of Early-scnRNAs occurs are marked by asterisks, regions of Late-scnRNA spreading are marked by arrowheads.

The meta-analysis of Late-scnRNAs in $\Delta COI6$ cells might give the impression that spreading occurs on both boundaries of any given individual IES. However, analysis of single loci indicates that Late-scnRNA spreading at one IESs boundary is not predictive of spreading at the second boundary. Therefore, three kinds of IESs can be described with respect to their scnRNA spreading behavior: 1. no detectable spreading, 2. spreading at one IES boundary and 3. spreading at both IES boundaries.

Late-scnRNAs are most likely produced from double-stranded precursor RNAs produced in the new MAC and processed by Dcl1p (Noto et al. 2015). Consistent with the idea that spreading small RNAs are *bona fide* scnRNAs produced by Dcl1p from double-stranded RNAs, the scnRNAs-seq meta-analysis profiles for $\Delta COI6$ cells at 10.5 and 12 hpm revealed spreading of scnRNAs on both (+) and (-) strand in a symmetric fashion. In addition, separately analyzing small RNAs mapping IESs and MDSs revealed that the spreading small RNAs in $\Delta COI6$ cells share the same length distribution (26 – 32 nt) (Fig. 30, compare A and C) and base composition (Fig. 30, compare B and D) with WT scnRNAs. Particularly, the base distribution is mechanistically informative as the uracil bias of the first 5' base reflects both Dicer cleavage by Dcl1p and preferential incorporation of such scnRNAs by Twi1p. Additionally the predominant adenine at the third base from the 3' end is attributed to Dcl1p cleavage (Mochizuki and Kurth 2013). As these features are present in $\Delta COI6$ cells at 10.5 hpm for both IES-mapping and for MDS-mapping (spreading) scnRNAs and indistinguishable from the WT signature, it is likely that spreading Late-scnRNAs in $\Delta COI6$ cells are products of Dcl1p cleavage.

The observation of Late-scnRNAs spreading outside IESs boundaries in absence of Coi6p could explain the spreading of H3K9me3, H3K27me3 and Pdd1p detected in $\Delta COI6$ cells, as such scnRNAs could use the normal heterochromatin targeting machinery to guide these heterochromatin marks to different genomic regions.



2.12 Coi6p interacts with Coi7p and Lia5p

In order to gain a better understanding of the function of Coi6p I aimed to identify interaction partners of Coi6p. Coi6p was immunoprecipitated (IPed) from cell lysate derived of WT conjugating cells at 8 hpm using the α -Coi6p(r) antibody. As a negative control, an IP with pre-immune serum was performed. Both samples were then subjected to high resolution MS/MS analysis. The data retrieved indicated that the Coi6p IP worked efficiently and specifically, as high amounts of Coi6p peptides were detected in the IP sample, compared to no peptides in the pre-immune control IP (Peptide spectrum matches (PSM): control IP: 0; Coi6p IP: 458) (Fig. 31A). Interestingly, two proteins were specifically co-precipitated with Coi6p and yielded a high number of peptides in MS/MS analysis: Coi7p (PSM: control IP: 0; Coi6p IP: 146) and Lia5p (PSM: control IP: 4; Coi6p IP: 436) (Fig. 31A).

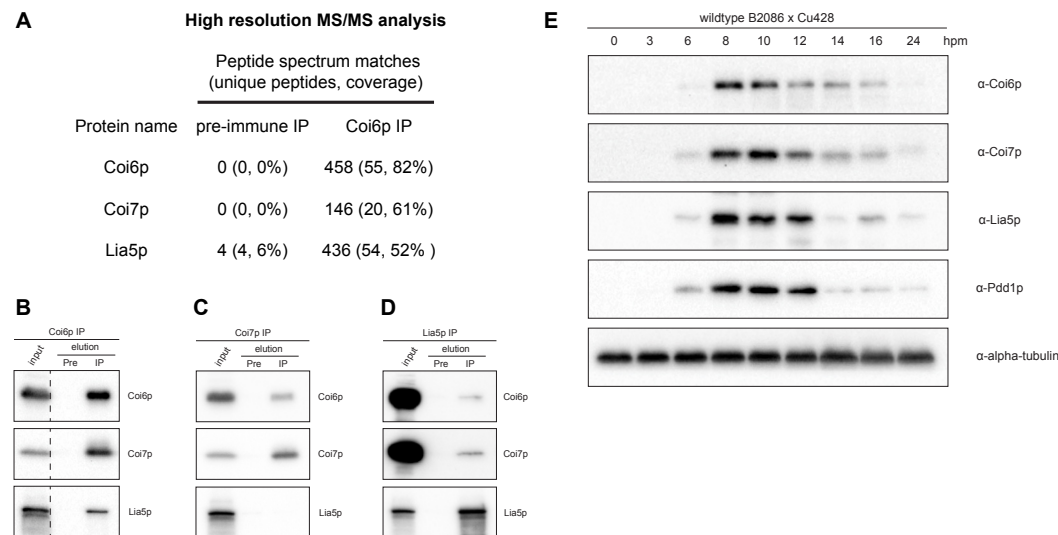


Figure 31: Interaction and co-expression of Coi6p, Coi7p and Lia5p
(A) Results of high resolution MS/MS analysis of Coi6p IPs from WT cells at 8 hpm are displayed, showing peptide spectrum matches, unique peptides and protein coverage for each protein. Pre-immune serum was used as a negative control (pre-immune IP). (B) Repeat of (A) using western blot analysis with α -Coi6p(r), α -Coi7p and α -Lia5p as a readout. (C and D) Reverse IPs of Coi7p and Lia5p, respectively, using α -Coi7p and α -Lia5p with WT cell lysate at 8 hpm followed by western blot detection. (E) Time course western blot analysis of protein expression of Coi6p, Coi7p, Lia5p, Pdd1p and alpha-tubulin (as loading control) in WT cells during conjugation. Time points are indicated in hours post mixing (hpm).

In order to verify these results, first, rabbit antibodies were raised against recombinant MBP-tagged Coi7p and recombinant MBP-tagged Lia5p. These antibodies (designated α -Coi7p and α -Lia5p, respectively) detected their epitopes Coi7p or Lia5p specifically (Fig.

34A and C). Then, the Coi6p IP was repeated and analyzed by western blotting. Consistent with mass spectrometry results, Coi6p was co-precipitated with both Coi7p and Lia5p (Fig. 31B). Subsequently, Coi7p or Lia5p IPs were performed using the α -Coi7p or α -Lia5p antibodies and pre-immune serum as control. In the Coi7p IPs Coi7p was clearly detected together with co-precipitating Coi6p, while Lia5p could not be detected (Fig. 31C). Furthermore, Coi6p and Coi7p were co-precipitated with Lia5p, albeit in lesser amounts compared to the precipitated Lia5p (Fig. 31D). Therefore, the Coi6p-Coi7p interaction seems to be the most robust, while a strong interaction with Lia5p was only detected in Coi6p IPs. Thus, it is likely that Coi6p, Coi7p and Lia5p interact with each other in vivo, however the molecular mode of these interactions may not be stoichiometric. Nevertheless, these results point to a potential concerted role of Coi6p, Coi7p and Lia5p in controlling precision of the Late-scnRNA production and heterochromatin formation.

2.13 Conservation of Coi7p and Lia5p and co-expression with Coi6p

Interestingly, both interaction partners of Coi6p, Coi7p and Lia5p, were previously reported to be expressed conjugation-specifically.

For Coi7p (conjugation-induced protein 7) its conjugation-specific expression had previously been described (Woehrer et al. 2015), yet no analysis of *COI7* was undertaken, as conventional methods of generating *COI7* KO strains by homologous recombination failed (unpublished data). *COI7* encodes a 264 amino acid protein (~ 31.1 kDa) that belongs to the acidic (leucine-rich) nuclear phosphoprotein 32 kDa (ANP32) protein family. This protein family is conserved in most organisms, including humans, and ANP32 proteins typically consist of an N-terminal Leucine-rich repeat domain (LRR) followed by an acidic low-complexity region of variable length. Diverse functions have been attributed to ANP32 proteins (reviewed in Reilly et al. 2014).

Lia5p has been previously described as conjugation-specific transposon-related protein that shares similarity with the IS4 transposase family and thus likely arose from a domestication event (Yao et al. 2007; Shieh and Chalker 2013). As the catalytic DDD residues found in active transposases could not be identified in Lia5p, this protein is presumed to be catalytically inactive (Shieh and Chalker 2013). The detailed analysis of Lia5p published by Shieh et al. also described its localization in new MAC heterochromatic foci much like Coi6p. Intriguingly, this study described a phenotype for Δ LIA5 cells that parallels the phenotype of Δ COI6 cells in several aspects: 1. A severe

block of programmed DNA elimination, 2. Judged by IF stainings histone H3K9/K27 methylations are established in the new Mac and 3. failure to form heterochromatin foci (Shieh and Chalker 2013).

In addition to the already-published evidence, conjugation-specific expression of Coi7p and Lia5p was verified using α -Coi7p and α -Lia5p antibodies (Fig. 31E). It is noteworthy that all three proteins Coi6p, Coi7p and Lia5p are co-expressed with the peak of expression around 8 - 12 hpm, thus at the time of new MAC development, Late-scnRNA production and DNA elimination. Co-expression is further supported by publicly available microarray based mRNA expression analysis (Fig. 14A).

2.14 Coi7p and Lia5p are enriched on IESs

As Coi7p and Lia5p interact with Coi6p, which was shown before to enrich on IESs in ChIP-seq experiments, I hypothesized that also Coi7p and Lia5 might bind to IESs.

To test this ChIP-seq experiments were performed with WT cells at 12 hpm and α -Coi7p or α -Lia5p antibodies. Analysis of ChIP-seq data on individual genomic loci as well as by meta-analysis revealed enrichment of Coi7p and Lia5p on both Type-A and Type-B IESs (Fig. 32C, D, G and H). However, more ChIP-seq reads were detected from MDS region in Coi7p and Lia5p ChIP than observed in Coi6p or Pdd1p ChIP (compare MDS regions in Fig. 32A and B vs C and D). Furthermore, meta-analysis showed a lower level of IES enrichment for Coi7p and particularly for Lia5p (Fig. 31G and H). This lower enrichment of both proteins might either reflect technical shortcomings of the ChIP experiments, such as lower antibody quality, or it might indeed reflect less abundant binding of, especially, Lia5p to IESs. Despite these concerns, it is clear that both Coi7p and Lia5p have the capacity to interact with IES DNA much like Coi6p, which indicates that these three proteins not only interact in cell lysate, but also on chromatin.

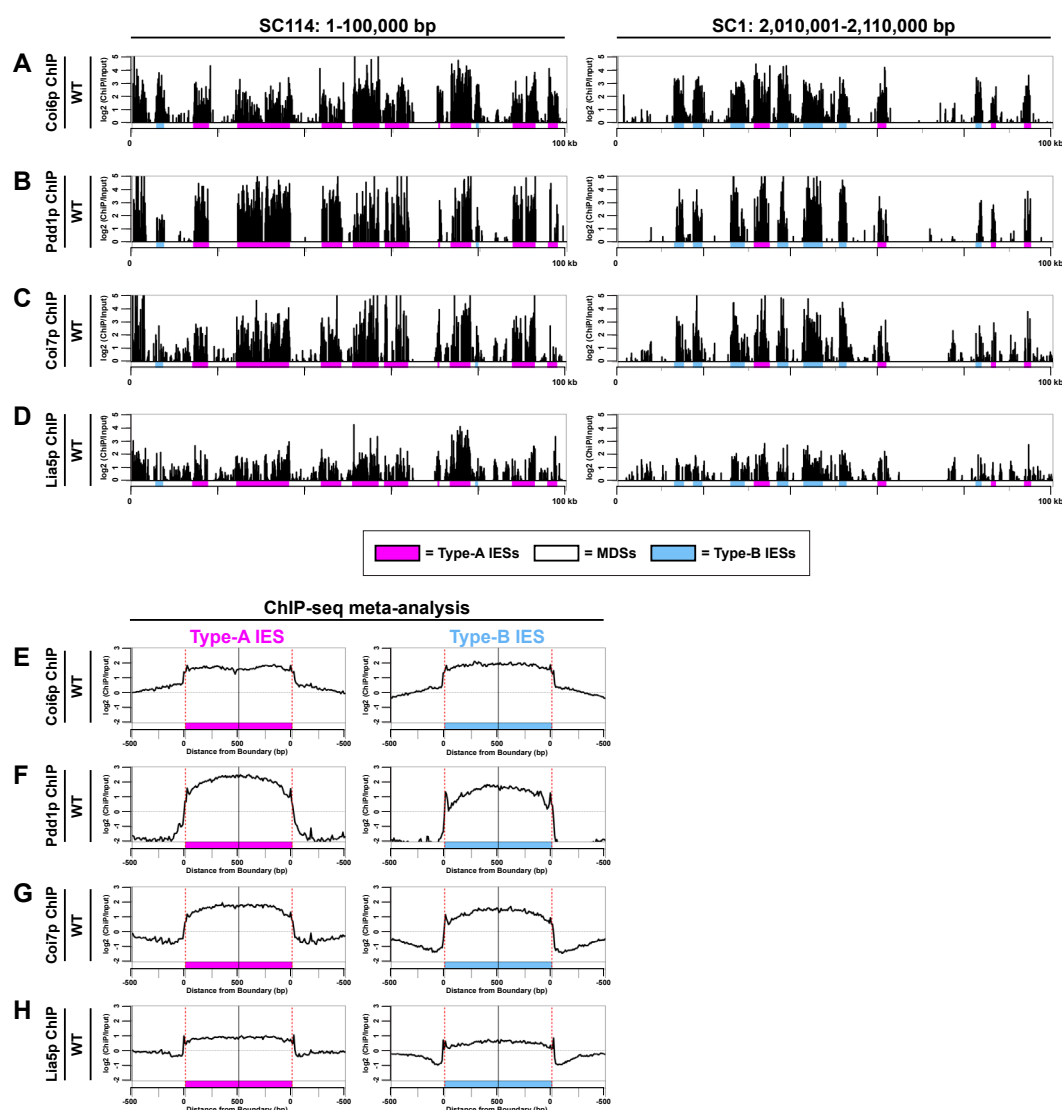


Figure 32: Coi7p and Lia5p are enriched on IESs
 (A-D) Analysis of ChIP-seq data in two 100 kb genomic regions for Coi6p, Pdd1p, Coi7p and Lia5p ChIP in WT, respectively. The left panel shows a region mostly containing Type-A IESs, while the right panel shows a region containing mostly Type-B IESs. (E-H) IES boundary-focused meta-analysis (as described before) of the same ChIP experiments as in (A-D). Type-A (purple) and Type-B (light blue) IESs were analyzed separately.

2.15 Generation of $\Delta COI7$ strains using CRISPR/Cas9

As highlighted above Coi7p is a conjugation-specific protein that was identified as a Coi6p interaction partner and binds to IESs. This made Coi7p a great candidate for playing an important role during DNA elimination, yet no $\Delta COI7$ strains were available

wild-type cells were mutagenized. Mutagenized cells were immediately outcrossed to non-mutagenized wild-type cells to ask if the mutation can be transmitted to the next sexual generation and thus was successfully introduced into the MIC. Heterozygote mutants carrying mutant alleles in both MIC and MAC were recovered and crossed in order to generate homozygous *COI7* mutant strains (henceforth simply termed $\Delta COI7$) lacking all WT copies of *COI7* in MIC and MAC (Fig. 33A). Two almost identical heterozygote mutants were used to generate $\Delta COI7$ cells: *COI7FS1* and *COI7FS2*. Both lead to nonsense mutation of *COI7* by a 1 bp deletion in codon 14 of the *COI7* ORF (*COI7FS1*: T40 deletion, *COI7FS2*: A41 deletion), which facilitates a downstream frameshift and a premature stop-codon in codon position 59 (Fig. 33B and C). The resulting 58 amino acid long mutated proteins *Coi7p_fs1* and *Coi7p_fs2* are identical except for amino acid 14 and retain only the first 13 amino acids of WT *Coi7p* (Fig. 33C and D). Thus they are most likely non-functional even if they are stably expressed. Sanger sequencing of genomic DNA derived of $\Delta COI7$ cells showed a characteristic double A-T peak at *COI7* nucleotide position 40, indicating the presence of both *COI7FS1* and *COI7FS2* alleles (Fig. 33B). By α -*Coi7p* antibody western blot analysis of 8 hpm cell lysate derived of WT or $\Delta COI7$ cells the complete absence of *Coi7p* from $\Delta COI7$ cells was furthermore verified on protein level (Fig. 33E). Thus, using the CRISPR/Cas9 system it was possible to establish $\Delta COI7$ strains, although other techniques had previously failed for this purpose. Importantly, this is to my knowledge the first successful demonstration of CRISPR/Cas9 mediated gene knock-out in *Tetrahymena*.

2.16 *Coi7p* and *Lia5p* are necessary for *Tlr1* elimination

Since *Coi7p* and *Lia5p* were identified as interaction partners of *Coi6p*, a comparison of the severity of DNA elimination defects in KOs of these proteins was desirable. To this end $\Delta COI7$ cells and $\Delta LIA5$ cells (kindly provided by Douglas Chalker and described in (Shieh and Chalker 2013)) were analyzed using DNA FISH for the moderately repeated *Tlr1* IES at 34 hpm. $\Delta COI7$ cells and $\Delta LIA5$ cells show a severe block of *Tlr1* elimination. In the absence of *Coi7p* approximately 84% of cells showed homogeneous *Tlr1* signal in the new MACs, 15% partial *Tlr1* signal and 1% no signal, therefore recapitulating the phenotype observed in $\Delta COI6$ cells (Fig. 18, $\Delta COI6$). For $\Delta LIA5$ cells a phenotype more severe than that of $\Delta COI6$ and $\Delta COI7$ cells was detected with 95% of cells showing homogeneous *Tlr1* signal in new MACs and 5% partial signal. Cells eliminating *Tlr1* completely could not be detected in $\Delta LIA5$ cells (Fig. 17C, $\Delta LIA5$).

Although differences in terms of Tlr1 FISH signal were observed between $\Delta COI6$ $\Delta COI7$ and $\Delta LIA5$ cells, these results are in agreement with these proteins fulfilling an important concerted role during DNA elimination in *Tetrahymena*.

2.17 Coi6p is unstable in absence of Coi7p

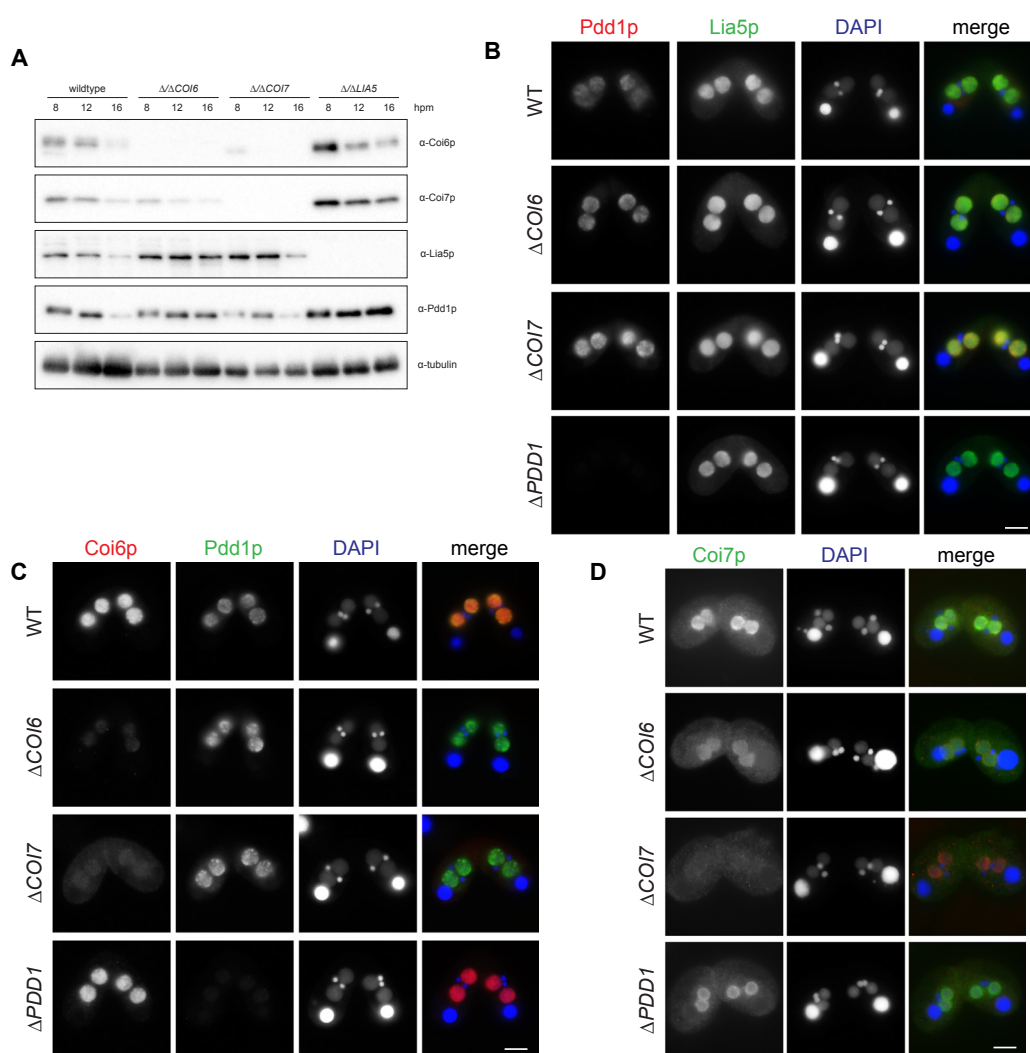


Figure 34: Stable accumulation of Coi6p depends on Coi7p

(A) Western blot analysis of Coi6p, Coi7p, Lia5p and Pdd1p expression in WT, $\Delta COI6$, $\Delta COI7$ and $\Delta LIA5$ cells between onset and end of heterochromatin formation (8, 12, 16 hpm). Self raised α -Coi6p(r), α -Coi7p and α -Lia5p antibodies as well as commercial α -Pdd1p and α -alpha-tubulin antibodies (loading control) were used. (B, C and D) IF stainings of formaldehyde fixed (B and C) or *Schaudin* fixed (D) WT, $\Delta COI6$, $\Delta COI7$ and $\Delta PDD1$ cells at 8hpm, stained with α -Lia5p, α -Coi6p(r) and α -Coi7p antibodies respectively. In (B and C) α -Pdd1p (gp) was used as a control. In all IF stainings DNA was visualized using DAPI.

The observation of the three interacting proteins Coi6p, Coi7p and Lia5p lead me to investigate whether these proteins might influence each other's expression, stability or nuclear import, as is often observed for protein complexes.

First, we tested expression of all these proteins in KO of *COI6*, *COI7* and *LIA5* during the stage of heterochromatin formation (8, 12 16 hpm) by western blotting. Interestingly we observed that although Lia5p accumulated to normal levels in $\Delta COI7$ cells, Coi6p expression was barely detectable (Fig. 33A). This suggests that Coi7p is necessary for the stable accumulation of Coi6p. Consistent with this observation, IF stainings performed at 8 hpm using the a-Coi6(r) antibody detected only minute amounts of Coi6p in the new MAC of $\Delta COI7$ cells (Fig. 33C). This stability effect furthermore did not seem to be applicable in an inverse manner, as Coi7p levels were at most mildly reduced in absence of Coi6p (Fig. 33A and D). Moreover, Lia5p levels were not affected by absence of either Coi6p or Coi7p (Fig. 34A and B), indicating that the stabilizing effect that Coi7p exerts on Coi6p is specific and is not extended to Lia5p.

The observation of Coi6p stabilization by Coi7p is a strong argument for a close interaction of the two proteins, yet precludes a detailed analysis of Coi7p function beyond this stabilization effect as *COI7* KO cells *de facto* closely resemble hypothetical *COI7 COI6* double-KO cells and the contribution of absence of Coi6p and Coi7p to the observed phenotype cannot be delineated. The strong Tlr1 elimination defect observed in $\Delta COI7$ cells can be explained solely by greatly reduced Coi6p levels.

2.18 Coi7p and Lia5p are important for precise Late-scnRNA production

Coi6p is important for precise Late-scnRNA production from IESs and interacts with Coi7p and Lia5p. This suggests that Coi7p and Lia5p might also be important to prevent spreading of Late-scnRNA production outside IESs boundaries.

To test this possibility scnRNAs-seq was performed for both $\Delta COI7$ and $\Delta LIA5$ cells at 10.5 hpm. Meta-analysis of scnRNA production from compiled IESs for $\Delta COI6$ scnRNA-seq was carried out as described above (Fig. 35). Furthermore, to gain a comprehensive overview of effects on single IES level all IESs were analyzed separately and then compiled into a heat map where each line represents one IES locus. As for the meta-analysis the 500 bp inside and outside of each IES boundary were used for the heat map representation. The ranking of IESs loci was determined by scnRNAs signal intensity on IESs in WT cells at 10.5 hpm (Fig. 36).

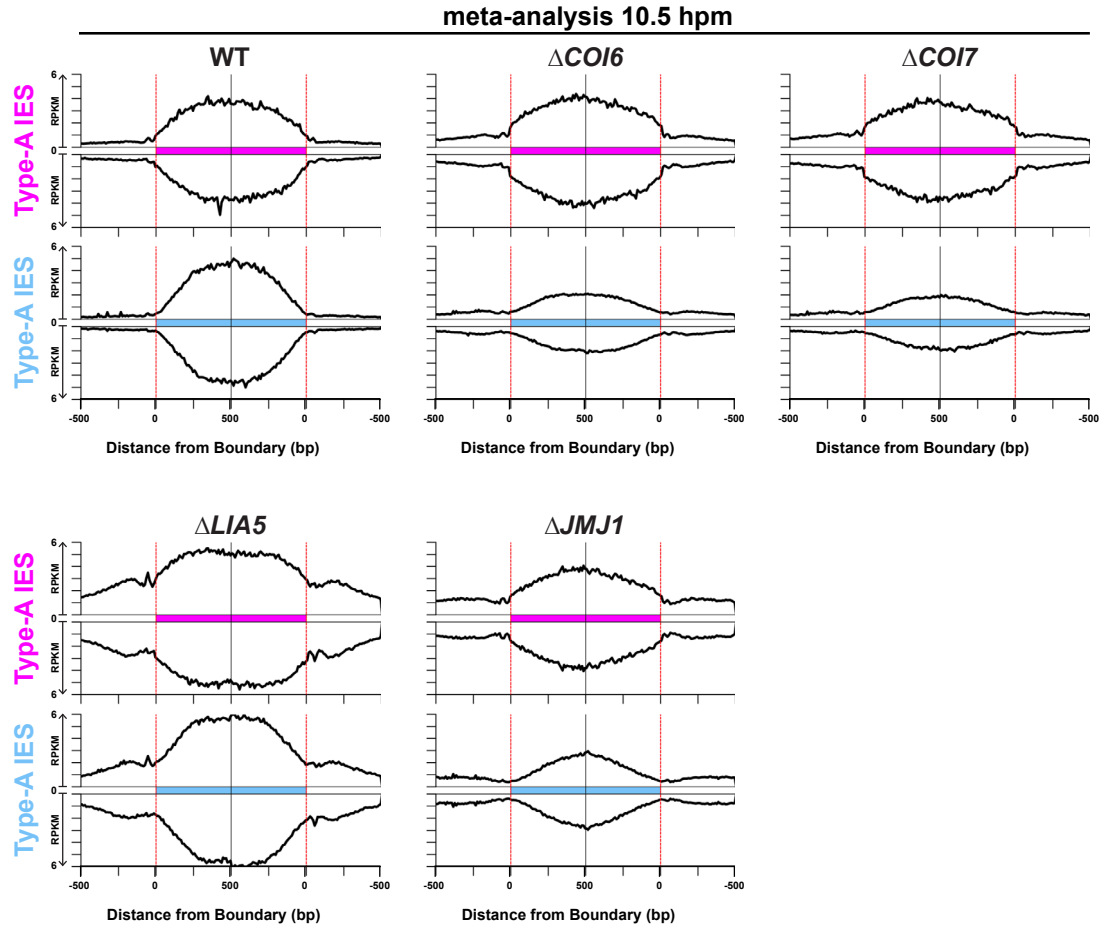


Figure 35: Meta-analysis of 10.5 hpm scnRNAs in WT, $\Delta COI6$, $\Delta COI7$ and $\Delta LIA5$ and $\Delta MJM1$ cells

IES boundary-focused meta-analysis of WT, $\Delta COI6$, $\Delta COI7$ and $\Delta LIA5$ and $\Delta MJM1$ cells, performed as described before. Type-A (purple) and Type-B (light blue) IESs were analyzed separately for each strain and (+) and (-) strand mapping scnRNAs are shown on top and bottom of each graph, respectively.

For $\Delta COI7$ cells a phenotype similar to that of $\Delta COI6$ cells was observed in meta-analysis, signified by spreading of Late-scnRNAs from IESs to MDS regions for both Type-A and Type-B IES, while the overall production of Late-scnRNAs for Type-B IESs was reduced (Fig. 35, compare WT, $\Delta COI6$ and $\Delta COI7$). This was also confirmed by the analysis of individual IES loci in heat map format, which reveals abundant, yet locus dependent, loss of IES boundary maintenance for $\Delta COI7$ cells and highlights once again a similar phenotype as for $\Delta COI6$ cells (Fig. 36, compare WT, $\Delta COI6$ and $\Delta COI7$). This is in stark contrast to WT cells where IES signal is remarkably clearly focused on IESs genome-wide (Fig. 36, WT).

Analysis of scnRNAs-seq data derived of $\Delta LIA5$ cells also reveals spreading of Late-scnRNAs outside IES boundaries. Once again, this can be observed indiscriminately for Type-A and Type-B IESs and in both meta-analysis and on single loci level (Fig. 35 and 36, $\Delta LIA5$). In comparison to the phenotype of $\Delta COI6$ and $\Delta COI7$ cells, the defect of $\Delta LIA5$ cells seems to diverge at least in a quantitative manner as scnRNAs signal detected outside the IES boundaries is much more strongly elevated than that observed in $\Delta COI6$ and $\Delta COI7$ cells, especially in the immediate vicinity of the an IES (Fig. 35, $\Delta LIA5$). Furthermore, and in contrast to the $\Delta COI6$ and $\Delta COI7$ phenotype scnRNAs signal on Type-B IESs did not decrease in $\Delta LIA5$ cells, but might even be elevated (Fig. 35, $\Delta LIA5$).

Collectively, this data shows that both *Coi7p* and *Lia5p* are of crucial importance to ensure precision of Late-scnRNA production. The observed similarities between defects in $\Delta COI6$, $\Delta COI7$ and $\Delta LIA5$ cells suggest that these three proteins at least partially act together to confine scnRNA production to IESs. The similarity of the $\Delta COI6$ and $\Delta COI7$ phenotypes can be explained by the destabilization of *Coi6p* in the absence of *Coi7p*, while differences between $\Delta COI6$, $\Delta COI7$ cells on one hand and $\Delta LIA5$ on the other hand might hint at additional or differential functions between *Coi6p/Coi7p* and *Lia5p*.

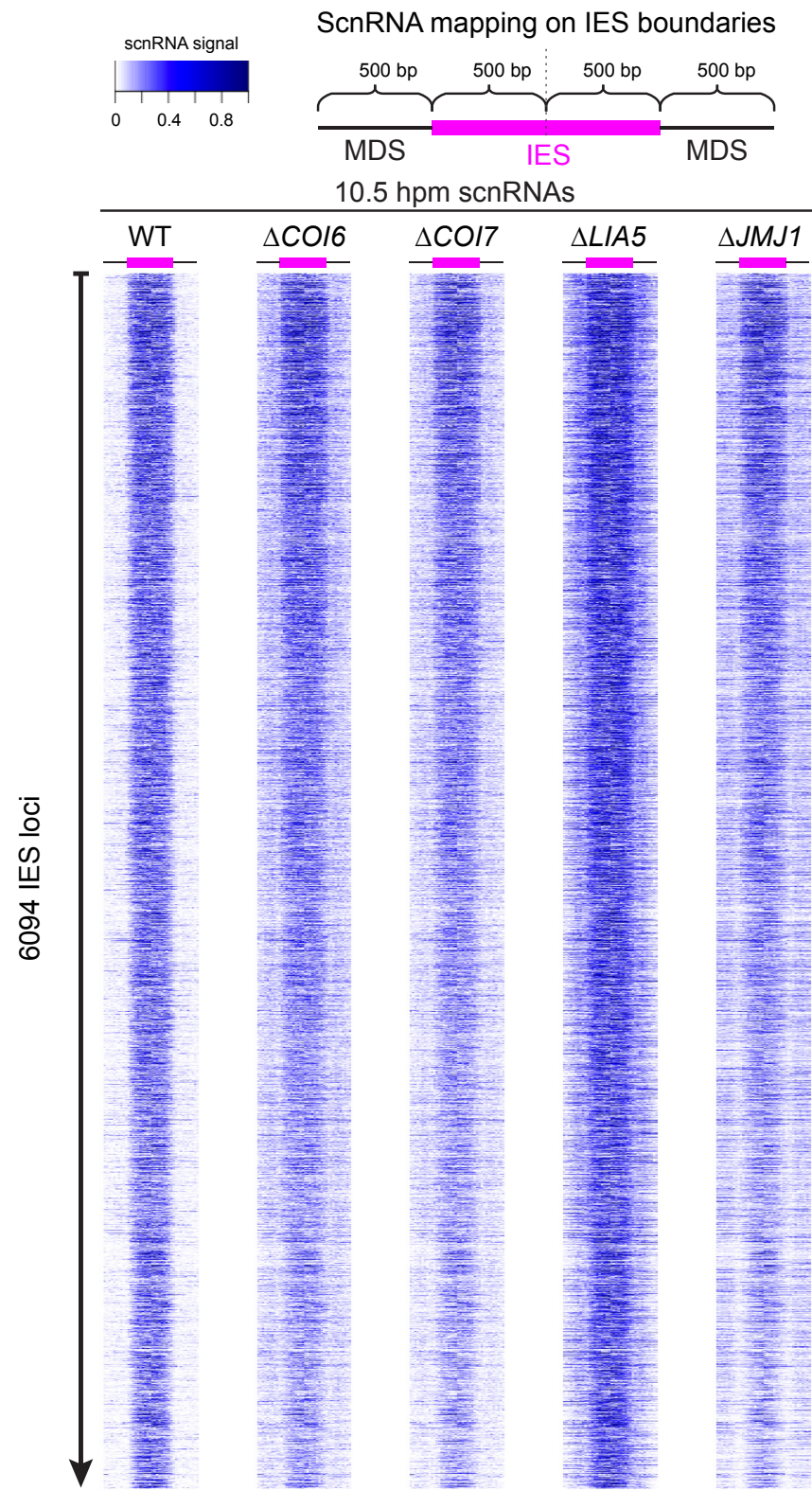


Figure 36: Heat map analysis of 10.5 hpm scnRNAs in WT, $\Delta COI6$, $\Delta COI7$ and $\Delta LIA5$ and $\Delta JMJ1$ cells revealing genome-wide scnRNAs spreading in all KOs

The IES boundary regions (500 bp up- and downstream of the two IES boundaries) of 6094 reliably mappable (> 200 bp mappable in each 500 bp window) IESs loci were plotted as a heat map for WT, $\Delta COI6$, $\Delta COI7$ and $\Delta LIA5$ and $\Delta JMJ1$ cells. ScnRNA signal intensity was calculated for each bin, assuming the value of the bin with the most reads as 1 (intense blue) and the one with the lowest reads as 0 (white) for each strain independently. IES loci were ranked by scnRNAs signal intensity on the IESs body in WT cells at 10.5 hpm (highest intensity on top).

2.19 The histone H3K27 demethylase Jmj1p is involved in maintaining precision of the Late-scnRNA production

The results presented so far indicate that heterochromatin is not only required for the production of Late-scnRNAs, but also plays a crucial role in facilitating precision of this process, particularly via its components Coi6p, Coi7p and Lia5p. In $\Delta COI6$ cells spreading of scnRNAs outside IES boundaries furthermore coincided at least partially with expansion of histone H3K9/K27me3 and Pdd1p to MDS regions, consistent with the primary requirement of Pdd1p for Late-scnRNA production (Fig. 22, 23).

This raised the question whether aberrant targeting of heterochromatin during new MAC development would generally impact precision of Late-scnRNA production. A previous report described *JMJ1* as a conjugation-specific histone demethylase, which likely turns over Ezi2p catalyzed H3K27me3, presumably on genic loci, during new MAC development and in parallel to DNA elimination (Chung and Yao 2012). The study by Chung et al. found that knock-down of *JMJ1* resulted only in mild defects of DNA elimination, although progeny viability was heavily impaired. However, they also reported a decrease of ChIP H3K27me3 and Pdd1p signal on IESs in *JMJ1* knock-downs, accompanied by an increase of H3K27me3 on several genes. Given the close spacing of genes and IESs in the *Tetrahymena* genome, I speculated that this genomic redistribution of H3K27me3 to regions outside IESs could impact Late-scnRNA production.

To approach this issue knock-out strains of the histone H3K27me3 demethylase *JMJ1* ($\Delta JMJ1$) were obtained (kindly provided by Dr. Yifan Liu, University of Michigan, Ann Arbor).

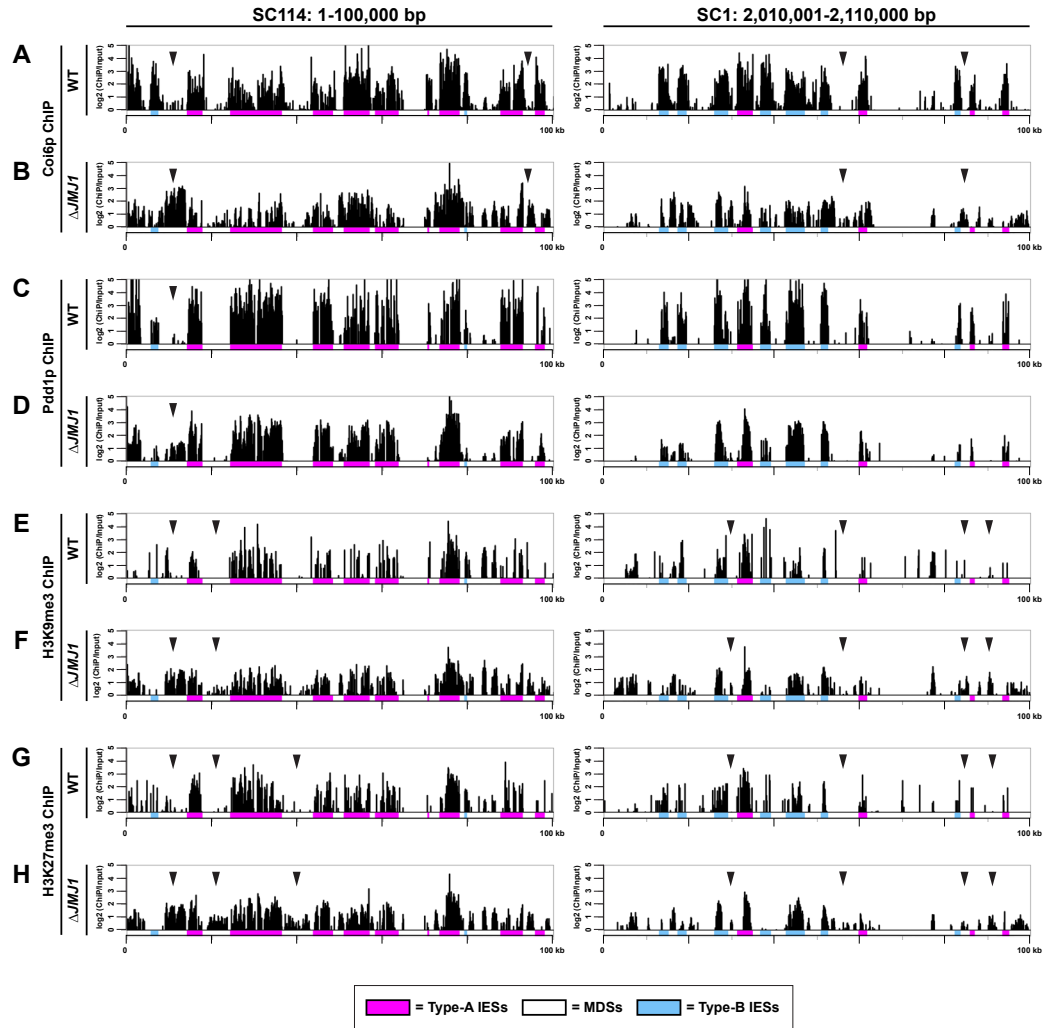


Figure 37: ChIP-seq analysis of heterochromatin spreading in $\Delta JMJ1$ cells

ChIP-seq analysis of Coi6p, Pdd1p, H3K9me3 and H3K27me3 in WT and $\Delta JMJ1$ cells. (A and B) ChIP-seq analysis of Coi6p in WT and $\Delta JMJ1$ cells, respectively. (C and D) ChIP-seq analysis of Pdd1p in WT and $\Delta JMJ1$ cells, respectively. (E and F) ChIP-seq analysis of H3K9me3 in WT and $\Delta JMJ1$ cells, respectively. (G and H) ChIP-seq analysis of H3K27me3 in WT and $\Delta JMJ1$ cells, respectively. For each experiment two genomic 100 kb loci are shown containing either predominantly Type-A IESs (left) or Type-B IESs (right). The Y-axis depicts ChIP results as $\log_2(\text{ChIP}/\text{Input})$. Arrowheads indicate positions of heterochromatin spreading.

Using the Tlr1 DNA FISH assay it could be confirmed that $\Delta JMJ1$ cells showed a similarly mild block of Tlr1 elimination (18% of cells contain few foci of Tlr1 in their new MACS, whereas 82% eliminate Tlr1 completely), as reported previously (Fig. 18B, $\Delta JMJ1$). Seeking to confirm the results of Chung et al. using $\Delta JMJ1$ cells, ChIP-seq experiments were performed, as described above, precipitating Pdd1p, H3K27me3 and additionally Coi6p and H3K9me3 using specific antibodies. As above, ChIP-seq data was mapped on two 100 kb genomic regions and analyzed by IES boundary-focused meta-analysis (Fig.

37 and 38, respectively). As predicted, the H3K27me3 ChIP showed a redistribution of H3K27me3 from IESs to MDS regions in $\Delta JMJ1$ cells when looking at meta-analysis with clearly decreased signal on IESs and elevated levels on MDSs (compare Fig. 38G and H). This effect was observed regardless of Type-A and Type-B IES classifications and new H3K27me3 accumulations in MDS regions were easily detectable in the 100 kb genomic regions (Fig. 37G and H, arrowheads). As Jmj1p had been described as H3K27me3 demethylase, I did not expect any effects on H3K9me3 in $\Delta JMJ1$ cells. However, ChIP-seq analysis indicates that H3K9me3 mirrors the behavior of H3K27me3, with decreased IES signal and increased MDS signal in $\Delta JMJ1$ cells (Fig. 37E and F, Fig. 38E and F). This contradicts a simple model, in which Jmj1p simply counteracts Ezh2p, which has been characterized as H3K27 di- and trimethylase, yet not as H3K9 methylase (Zhang et al. 2012; Chung and Yao 2012). Moreover, Pdd1p, which is known to bind to both H3K9me and H3K27me, did not follow exactly the pattern of these histone marks, but localized remarkably precisely to both Type-A and Type-B IESs even in the absence of Jmj1p (compare signal on IESs in Fig. 37C and D, Fig. 38C and D). However, at some loci clear redistribution of Pdd1p to MDS regions occurred (Fig. 37D, arrowhead). Interestingly, this parallels the behavior of Pdd1p in $\Delta COI6$ cells, where histone modifications spread outside IESs, yet Pdd1p follows them only partially (Fig. 23B and C and 24B and C). In contrast to Pdd1p, Coi6p seemed to track more closely with H3K9/K27me3 in $\Delta JMJ1$ cells, as it clearly redistributed more strongly from IESs to MDSs in meta-analysis, with almost equal levels between IESs and MDS regions (compare Fig. 38A and B). Also mapping Coi6p ChIP-seq on 100 kb genomic regions revealed more clear peaks on MDSs than for Pdd1p (compare Figure 37B and D, arrowheads). The above results clearly show that heterochromatin is aberrantly targeted in the absence of Jmj1p. Therefore it was interesting to assess whether Late-scnRNA production is affected in such a situation. To this end scnRNAs-seq experiments were carried out for $\Delta JMJ1$ cells as previously for other KOs. Meta-analysis of these data indicated that Late-scnRNAs did spread outside IES boundaries and heat map analysis showed that this spreading could be observed for thousands of IESs, yet not for all (Fig. 34 and 35, $\Delta JMJ1$). The effects observed, particularly in meta-analysis, were similar to the phenotype of $\Delta COI6$, $\Delta COI7$, $\Delta LIA5$, yet seemed less centered on IESs, consistent with the idea that the primary target of Jmj1p for histone demethylase activity might not be IESs, but genes or MDS regions in a more general way. All together, this indicates that misguided heterochromatin indeed impacts accuracy of the Late-scnRNA production. Given the close spacing of IESs and genes in the *Tetrahymena* genome, this might be a general phenomenon.

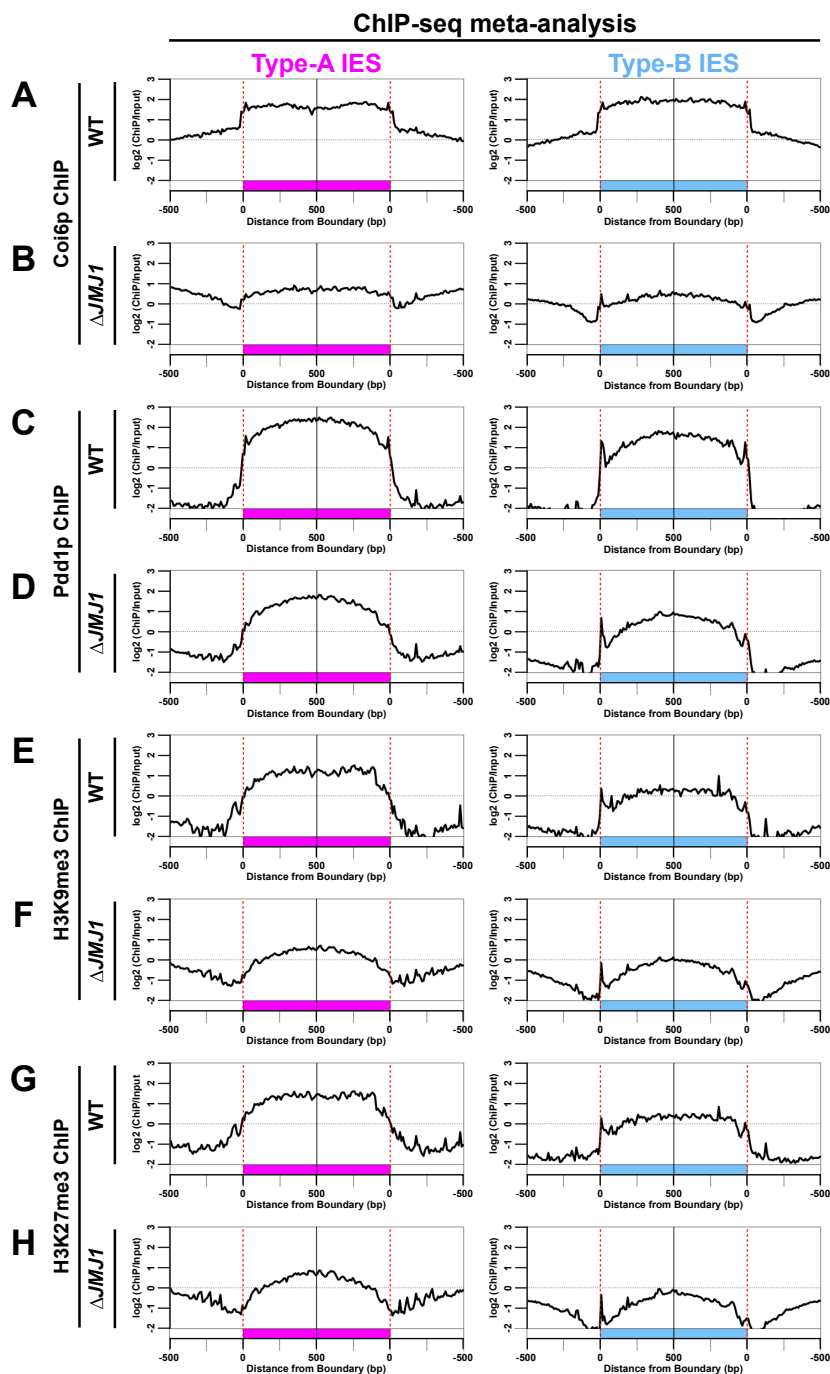


Figure 38: ChIP-seq meta-analysis of WT and $\Delta JMJ1$ cells

IES boundary-focused meta-analysis of Coi6p, Pdd1p, H3K9me3 and H3K27me3 ChIP-seq in WT and $\Delta JMJ1$ cells. (A and B) ChIP-seq analysis of Coi6p in WT and $\Delta JMJ1$ cells, respectively. (C and D) ChIP-seq analysis of Pdd1p in WT and $\Delta JMJ1$ cells, respectively. (E and F) ChIP-seq analysis of H3K9me3 in WT and $\Delta JMJ1$ cells, respectively. (G and H) ChIP-seq analysis of H3K27me3 in WT and $\Delta JMJ1$ cells, respectively. The Y-axis depicts ChIP results as $\log_2(\text{ChIP}/\text{Input})$. Type-A (purple) and Type-B (light blue) IESs were analyzed separately.

2.20 Imprecise production of Late-scnRNAs causes abnormal elimination of IESs along with MDS regions

Visual inspection of scnRNAs-seq tracts mapped on individual IES loci in $\Delta COI6$ cells, led me to realize that not only did the boundaries of many IESs become blurry on the level of scnRNAs, but many neighboring IESs were virtually joined based on scnRNAs mapping (Fig. 39A). Also in $\Delta COI7$, $\Delta LIA5$ and $\Delta JMJ1$ cells, similar loci were observed and while scnRNAs profiles between $\Delta COI6$, $\Delta COI7$ and $\Delta LIA5$ were similar in most loci, some IESs were joint on scnRNAs level in $\Delta JMJ1$ cells, but not $\Delta COI6$, $\Delta COI7$ and $\Delta LIA5$ cells or vice versa (Fig. 38A). Furthermore, in some of these loci, we detected spreading of Pdd1p into MDS region between affected IESs in $\Delta COI6$ cells (Fig. 39A, lowest two rows). Together with the fact that partial Tlr1 elimination had been observed, particularly in $\Delta COI6$, $\Delta COI7$ and $\Delta JMJ1$ cells, this prompted the idea that the downstream consequence at such IES loci, might be the elimination of the MDS region between the two IESs. I coined the term overelimination for such events of extended DNA elimination. To test the prevalence of Late-scnRNA spreading induced overelimination I sought to develop a PCR-based assay, which would detect possible products of overelimination. Several loci were chosen, which comprise two IESs that were joined on scnRNAs level either in $\Delta COI6$ or in $\Delta JMJ1$ cells (Fig. 39A and B). Primers were designed in MDS regions close the “left” boundary of the first IES and close to the “right” boundary of the second IES, such that overelimination would yield small, efficiently amplified PCR products. This PCR-based overelimination assay was performed for five loci with $\Delta COI6$, $\Delta COI7$, $\Delta LIA5$ and $\Delta JMJ1$ cells using total genomic DNA derived of exconjugants at 24 hpm or vegetative background strains to control for naturally occurring polymorphisms. WT and $PDD1$ cells were furthermore used as controls, since in WT all elimination should occur precisely and in $\Delta PDD1$ cells all DNA elimination, including overelimination, should be blocked. In all of the loci tested we detected PCR products corresponding to DNA shorter than the normally-rearranged MAC genome (marked with stars) in $\Delta COI6$ and $\Delta COI7$ cells at 24 hpm (Fig. 39B). To verify the identity of overelimination PCR products, $\Delta COI6$ cells derived products of two of those loci were cloned and subjected to Sanger-sequencing (Fig. 40). Importantly, all of the sequenced clones were fusions of MDSs sequences that correlated to the expected overelimination events. Microheterogeneity at the fusion point (also found in WT cells at normal IES boundaries) furthermore suggested that many different in vivo overelimination events were detected (Fig. 40, alignments).

RESULTS

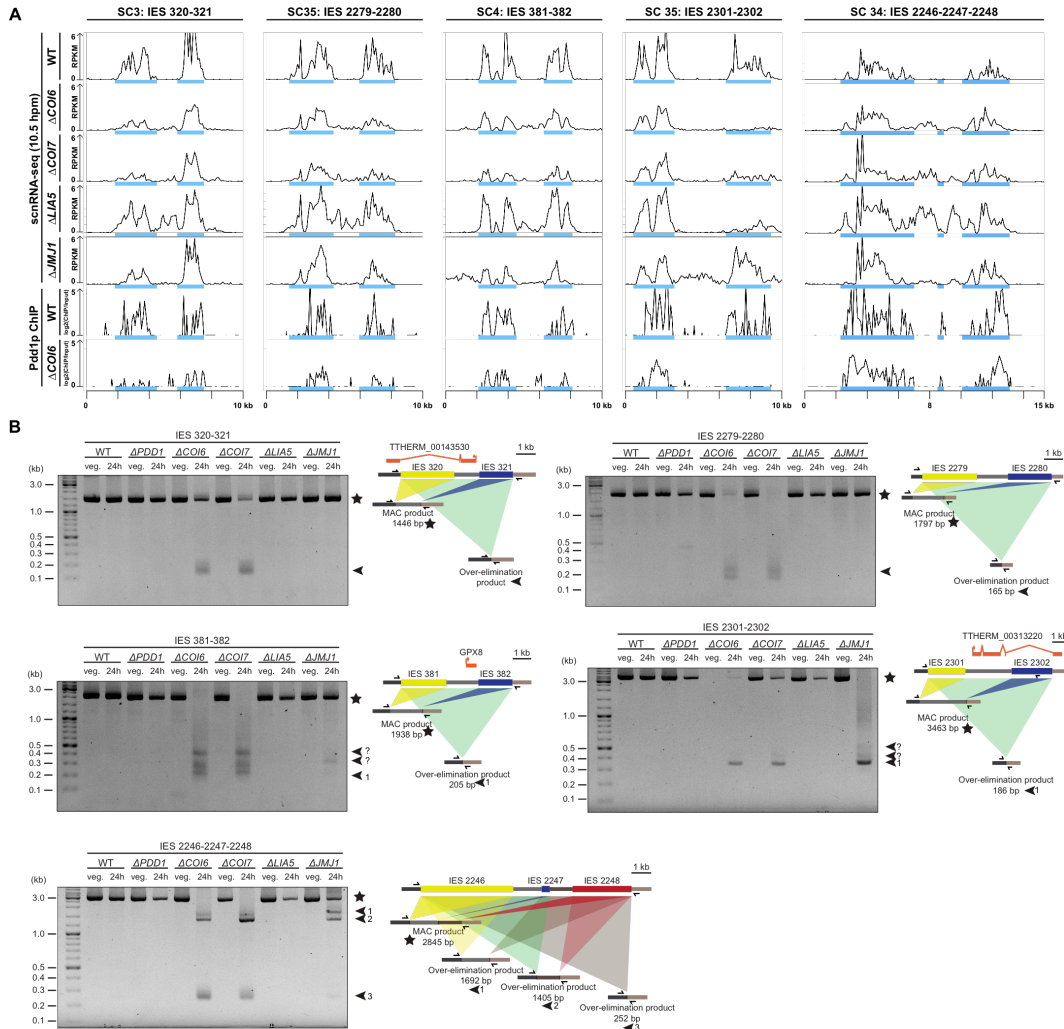


Figure 39: Spreading of Late-scnRNAs leads to elimination of MDS regions and genes

(A) ScnRNA-seq profiles at 10.5 hpm are shown for overelimination IES loci in WT, $\Delta COI6$, $\Delta COI7$, $\Delta LIA5$ and $\Delta MJM1$ cells (upper five rows). Moreover, Pdd1p ChIP-seq profiles in WT, $\Delta COI6$ at 12 hpm are shown (lower two rows). IESs are indicated by light blue bars beneath the x-axis. (B) PCR based overelimination analysis of the same five selected loci shown in (A) for WT, $\Delta PDD1$, $\Delta COI6$, $\Delta COI7$, $\Delta LIA5$ and $\Delta MJM1$ cells. PCR products were run on 2% agarose gels and detected by Ethidiumbromid staining. For each strain PCR was performed on a mix of genomic DNA of vegetatively growing background strains (veg.) or on genomic DNA derived from a mix of the same strains during conjugation at 24 hpm (24h). On the right hand side of each agarose gel a scheme depicts the MIC genome locus, and arrangement after WT DNA elimination (MAC product) and overelimination events (overelimination product). Genes present are indicated in orange above the scheme.

In contrast to $\Delta COI6$ and $\Delta COI7$, DNA elimination in $\Delta MJM1$ was only affected at 3 of the 5 loci tested (IES 381-382, IES 2301-2302 and IES 2246-2247-2248) (Fig. 39B, middle and lower row). The size of overelimination PCR products mostly fitted the predicted sizes, however additional bands were detected in some instances. As these were usually

RESULTS

locus, spreading Late-scnRNAs seems to induce alternative overelimination of two MDS regions, yielding three different overelimination PCR products (Fig. 39B, lowest row). This shows that overelimination can afflict larger regions extending further than two IESs and suggests that the new MAC genome might be altered severely by overelimination in $\Delta COI6$ and $\Delta COI7$ and $\Delta JMJ1$ cells.

3. Discussion

It has long been believed that during *Tetrahymena* DNA elimination a linear pathway of RNAi-induced heterochromatin formation is active, i.e., RNAi induces heterochromatin formation, whereas there is no feedback from heterochromatin to the RNAi machinery. However, the recent discovery that Late-scnRNAs are produced in parallel to heterochromatin formation, and that their accumulation crucially depends on the heterochromatin components Ezl1p and Pdd1p, pointed clearly to a self-reinforcing feedback loop between RNAi and heterochromatin as it is found in many other organisms (Noto et al. 2015; Holoch and Moazed 2015). The presence of such a feedback system predicts the necessity of mechanisms that prevent the occurrence of RNAi-heterochromatin feedback loops outside of IES borders. This study for the first time presents direct evidence for the presence of such a border-protection mechanism. Analysis of the heterochromatin component Coi6p, an HP1-like protein, and its interaction partners Coi7p and Lia5p uncovered that for a subset of IESs these components are crucial to confine the production of Late-scnRNAs and the formation of heterochromatin to IESs. Similarly, the histone H3K27me3 demethylase Jmj1p safeguards the boundaries of an overlapping subset of IESs. These activities are critical for genome integrity as inaccurate targeting of heterochromatin leads to aberrant DNA elimination, a phenomenon for which I coined the term “overelimination”.

3.1 The HP1-like protein Coi6p is a new heterochromatin component in *Tetrahymena*

The conjugation-specific HP1-like protein Coi6p is a new component of RNAi-induced heterochromatin on IESs. This is reflected by its co-localization with Pdd1p in heterochromatin foci and its enrichment on IESs in ChIP-seq experiments (Fig. 15B, 16A). In cells lacking Coi6p DNA elimination is severely, but not completely, inhibited, whereas $\Delta PDD1$ cells show a complete block of elimination, indicating that these two HP1-like proteins fulfill non-redundant functions (Fig. 18B). The Coi6p amino acid sequence indicates that it likely binds to H3K9me2/3 only, in contrast to Pdd1p, which has dual H3K9/K27me3 specificity (Fig. 12B) (Liu et al. 2007). This is consistent with the requirement of Early-scnRNAs and Twi1p for Coi6p IESs binding, as accumulation of

H3K9/K27me in developing new MACs is also Twi1p-dependent (Liu et al. 2004; Liu et al. 2007). The severe DNA elimination defect in chromodomain aromatic-cage mutants of Coi6p further supports such a model (Fig. 21B, *COI6_CDmut*). Although a direct test for this prediction is necessary, I failed to detect interaction between recombinantly expressed Coi6p and any modified histone tail peptide. Together with Pdd1p, Pdd3p and Tcd1, Coi6p is the fourth HP1-like protein involved in DNA elimination (Madireddi et al. 1996; Nikiforov et al. 2000; Xu et al. 2015). Thus it will be interesting to explore whether these proteins' binding affinities for H3K9me2/3 and H3K27me2/3 are coordinated in a way that fosters their cooperation as observed for Swi6 and Clr4 in *S. pombe* or if they are in competition with each other (Al-Sady et al. 2013).

3.2 Coi7p and Lia5p are Coi6p interaction partners

Coi7p and Lia5p were identified as interaction partners of Coi6p (Fig. 31A and B). However, IP of Coi7p yielded only Coi6p and IPs of Lia5p gave only small amounts of Coi6p and Coi7p (Fig 31C and D). These results may indicate that there might be distinct pools of these proteins with differential interactions and potentially additional binding partners.

Coi7p and Lia5p are also heterochromatin components and are accumulated on IESs, indicating that Coi6p-Coi7p-Lia5p interaction might take place on chromatin (Fig. 32) (Shieh and Chalker 2013). As HP1 proteins are considered to be adapter proteins that recruit additional activities to chromatin, an attractive model is that Coi6p binds to RNAi-induced H3K9me2/3 and then recruits Coi7p and Lia5p to IESs (Fig. 42i). ChIP-seq of Coi7p and Lia5p in $\Delta COI6$ cells will shed more light on this in future.

3.3 Potential functions of Coi7p and Lia5p

The interaction between Coi6p and Coi7p was the most robust observed (Fig. 31A-C). Close interaction between Coi6p and Coi7p is further supported by the dramatic destabilization of Coi6p in absence of Coi7p, whereas Lia5p remained stable in this condition (Fig. 34A-C). Thus one role of Coi7p is to stabilize Coi6p. Consistent with the above observations, highly similar phenotypes were observed for $\Delta COI6$ and $\Delta COI7$ cells in terms of Tlr1 DNA elimination and Late-scnRNA spreading, whereas $\Delta LIA5$ cells have in both cases more severe phenotypes (Fig. 18B, 35, 36, compare $\Delta COI6$, $\Delta COI7$,

$\Delta LIA5$). This argues that Lia5p has functions exceeding the ones it has in conjunction with Coi6p-Coi7p. Since the defect observed in $\Delta COI7$ cells phenocopies that observed in $\Delta COI6$ cells and Coi6p is greatly reduced in $\Delta COI7$ cells, it cannot be deduced whether Coi7p has a function beyond stabilizing Coi6p. However, if such a function exists, it must act in the same pathway as Coi6p, because the lack of Coi6p in $\Delta COI7$ cells caused no apparent additive phenotype.

Coi7p belongs to the conserved ANP32 family of proteins. In mammals the conserved paralogous proteins ANP32A, ANP32B and ANP32E have been attributed with chromatin modulating or histone chaperone activities, although knowledge remains scarce (Reilly et al. 2014). It is tempting to speculate that Coi7p might contribute to IESs boundary formation by similar activities, for instance by increasing histone turnover at IES boundaries, which could counteract continuous methylation of histone H3K9/K27 by Ezi1p (Wang et al. 2014).

Lia5p is likely a domesticated transposase originating from the IS4 transposase family (Shieh and Chalker 2013). While the protein-fold of the Lia5p catalytic domain is conserved, Lia5p lacks the critical catalytic DDD residues, likely rendering Lia5p catalytically inactive. Nevertheless, the mutated catalytic domain might retain some affinity for DNA or might even have evolved to recognize features of IES boundaries. Such properties should be addressed in future experiments.

3.4 A crucial role for Coi6p and its interaction partners in maintaining IES boundaries

In agreement with the idea of Coi6p-Coi7p-Lia5p fulfilling a concerted role on chromatin, absence of either of these proteins compromises the integrity of the IES-MDS boundary, as demonstrated by spreading of Late-scnRNAs (Fig. 28, 29, 35, 36), histone modifications H3K9/K27me3 and to lesser degree Pdd1p outside IESs (Fig. 23, 24). Although this clearly affects only a subset of IESs, the defect is frequent and can be considered genome-wide. Furthermore, this defect does not correlate with Type-A or Type-B IES classifications, arguing that not only on Type-B IESs, but also on Type-A IESs substantial Late-scnRNA production takes place that needs to be restricted. Interestingly, the two boundaries of each IES act like independent entities, as spreading at one boundary is not predictive of spreading at the other. Given the relatively small average size and the absence of any obvious feature that conveys directionality to IESs, this result seems surprising. Small-scale differences of molecular environment or DNA

sequence within an IES or its flanking regions must determine the vulnerability of each individual boundary to spreading events.

Currently the exact biochemical role(s) of the Coi6p-Coi7p-Lia5p complex in maintaining IES boundaries are unknown. Also given the self-reinforcing nature of RNAi and heterochromatin, it is unclear at which step these proteins are primarily involved: spreading of Late-scnRNAs or spreading of heterochromatin.

3.5 Jmj1p restricts heterochromatin and Late-scnRNA production to IESs

Jmj1p was initially described as conjugation specific H3K27me3 demethylase counteracting presumed genic (MDS) disposition of H3K27me3 by the E(Z)-related methyltransferase Ezi2p during new MAC development (Chung and Yao 2012). Knock-down of *JMJ1* inhibited DNA elimination to a small degree, a result which was reproduced using $\Delta JMJ1$ cells (Fig. 18B) (Chung and Yao 2012). Using ChIP-seq analysis I showed that H3K27me3 indeed accumulates on MDS regions in absence of Jmj1p (Fig. 37G and H, 38G and H). Moreover, accumulations of H3K9me3, Coi6p and, to a lesser extent, Pdd1p on MDSs were also detected in $\Delta JMJ1$ cells (Fig. 37, 38, each A-F). This aberrant targeting of heterochromatin causes Late-scnRNA biogenesis outside of the IES boundaries for a large subset of IESs, which is only partially overlapping with the $\Delta COI6$ -subset (Fig. 35, 36, $\Delta JMJ1$). As Jmj1p is attributed with an obvious molecular function, that of a histone demethylase, it is clear that spreading of heterochromatin must be the primary defect, which triggers subsequent spreading of Late-scnRNA production. General features of Late-scnRNA spreading in absence of Jmj1p mirror effects observed in $\Delta COI6$ cells, e.g. the negligible relevance of Type-A and Type-B IESs classifications and the independence of both IES boundaries (Fig. 39A).

Previous studies indicated that Ezi2p is an H3K27-specific methyltransferase and is not required for the DNA elimination process (Chung and Yao 2012; Zhang et al. 2012; Gao et al. 2013; Papazyán et al. 2014). Therefore, mis-accumulation of H3K9me3 on MDSs in $\Delta JMJ1$ cells is most probably catalyzed by Ezi1p, the only known H3K9 methyltransferase in *Tetrahymena*, which normally acts exclusively during conjugation on IESs (Liu et al. 2007). Because it has been suggested that H3K27me3 is the prerequisite of H3K9me3 in Ezi1p-mediated catalysis (Liu et al. 2007), aberrant accumulation of Ezi2p-deposited H3K27me3 in $\Delta JMJ1$ cells may trigger subsequent Ezi1p-mediated H3K9me3 and redirection of IES-specific heterochromatin factors like Coi6p and Pdd1p

to MDS regions (Fig. 41, Model 1). An alternative possibility is that Jmj1p may remove H3K27me3 deposited by both Ezi2p and Ezi1p (Fig. 41, Model 2). In this model, Jmj1p may confine heterochromatin formation on IESs by directly counteracting Ezi1p activity at the border of IESs.

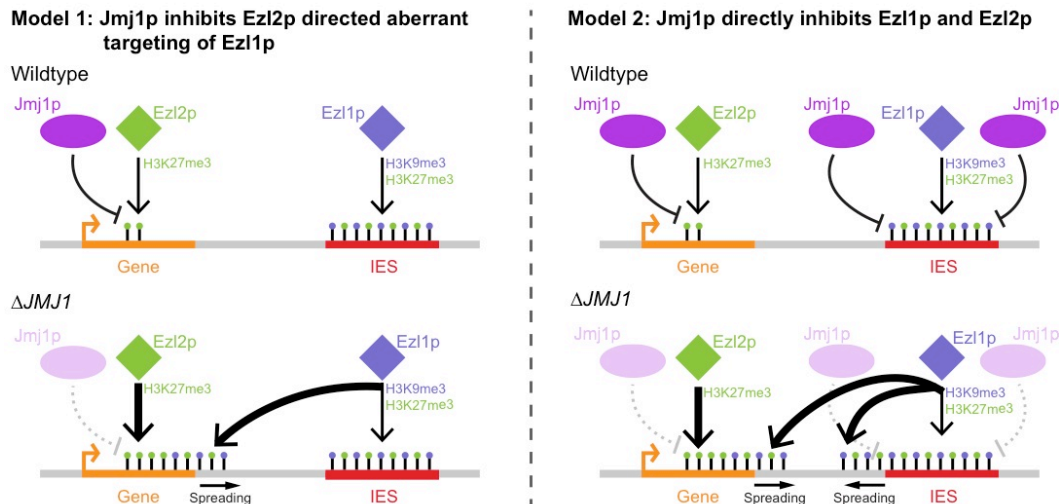


Figure 41: Two models for Jmj1p activity during DNA elimination

Model 1 (left panel): The role of Jmj1p is restricted to gene regulation during *Tetrahymena* new MAC development at genic loci. At these loci it counteracts Ezi2p-mediated H3K27me3. In absence of Jmj1p (lower row) an increase in Ezi2p-mediated H3K27me3 redirects Ezi1p H3K9/K27 methylation activity to MDS regions (including genic regions) outside of IESs and subsequently leads to aberrant targeting of heterochromatin and Late-scnRNA production. Model 2 (right panel): Jmj1p has two roles during new MAC development, counteracting both Ezi2p-mediated genic H3K27me3 and Ezi1p-mediated H3K9/K27me3 on IESs or IES boundaries. Absence of Jmj1p (lower row) leads to spreading of Ezi1p directed H3K9/K27me3 outside IESs. Additionally an increase in Ezi2p-mediated H3K27me3 might redirect Ezi1p to genic regions. In consequence heterochromatin and Late-scnRNA production spreads outside IESs.

3.6 RNAi-heterochromatin spreading-induced overelimination impacts genome integrity

In $\Delta COI6$, $\Delta COI7$ and $\Delta JMJ1$ cells I detected overelimination phenotypes for a number of loci concomitant with abnormal scnRNA spreading (Fig. 39B). Notably, $\Delta LIA5$ cells constituted an exception to these observations. Although Late-scnRNA profiles generally followed the pattern of $\Delta COI6$ and $\Delta COI7$, I did not detect any overelimination in $\Delta LIA5$ cells (Fig. 39A and B). This is in agreement with the more severe DNA elimination block of $\Delta LIA5$ cells (Fig. 18B). The more severe phenotype in terms of scnRNA spreading in $\Delta LIA5$ cells might block DNA elimination almost completely, which could prevent overelimination despite the spreading of scnRNAs and presumably heterochromatin. In

$\Delta COI6$, $\Delta COI7$ and $\Delta JMJ1$ cells, the milder scnRNA and heterochromatin spreading might allow for some DNA elimination to occur, resulting in overelimination at some loci. Occurrence of overelimination was only tested at five genomic loci, however it likely occurs on a genome-wide scale, as a large proportion of IESs are affected by Late-scnRNA spreading. In the *Tetrahymena* genome, IESs and genes are closely spaced and although IESs are rarely located in exons, they are frequently present in introns and promoter regions (Fass et al. 2011). Therefore, overelimination should have deleterious consequences for genome integrity and controlling RNAi-heterochromatin feedback loops must have essential importance for *Tetrahymena*. This view is supported by the severely reduced progeny viability caused by down-regulation of Jmj1p (Chung and Yao 2012). On the other hand, in an evolutionary context, rare overelimination events might also be an additional source of genomic variation. There may be a mechanism that down-regulates Jmj1p and/or the Coi6p-complex in some environmental conditions to actively increase the occurrence of overelimination and thereby boost the phenotypic variations of progeny.

3.7 A model for IES boundary maintenance

Although detailed knowledge about the relationship between heterochromatin formation and Late-scnRNA production in the new MAC remains scarce, a preliminary model can be delineated (Fig. 42, compare to Fig. 10 and 11 for a complete overview of the pathway). In this model initiation of both heterochromatin formation on IESs and Late-scnRNA production crucially depends on maternal (old MAC expressed) Twi1p loaded with Early-scnRNAs, which have been preselected to exclusively match IESs by scanning (Fig. 42a) (Liu et al. 2004; Aronica et al. 2008; Noto et al. 2015). Base-pairing of Early-scnRNAs with non-coding IES-transcripts triggers recruitment of histone methyltransferase Ezi1p, which trimethylates H3K9/K27 (Fig. 42b). Pdd1p is recruited to IESs through its binding to H3K9/K27me3 (Fig. 42c). Then, although the mechanism is unclear, IESs are bidirectionally transcribed (Fig. 42d) and transcripts form double stranded RNAs (Fig. 42e). These are likely the substrate to Dcl1p-mediated cleavage, which leads to the formation of double stranded Late-scnRNA duplexes (Fig. 42f) (Noto et al. 2015). These are loaded into zygotic (new MAC expressed) Twi1p and Twi11p (Fig 42g). Subsequent removal of the passenger strand generates active Twi1/11p-RITS, which initiates spreading of heterochromatin and Late-scnRNA production by starting another cycle of Ezi1p recruitment (Fig. 42h).

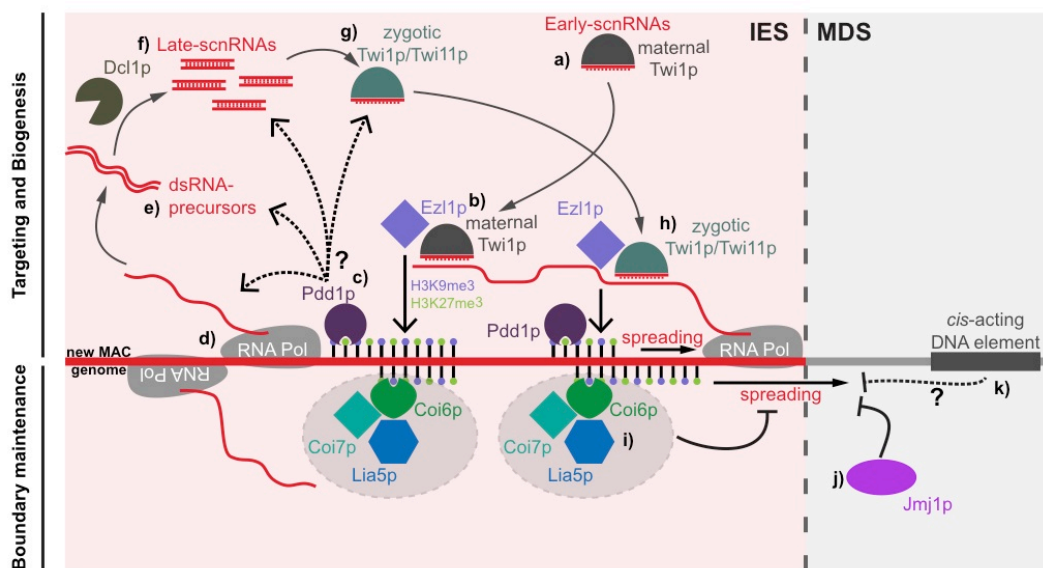


Figure 42: Model for initiation and IES-confinement of Late-scnRNAs and heterochromatin

Schematic depiction of an IES-MDS boundary in the new MAC. IES genomic sequence is indicated in red (with red background) and MDS genomic sequence in grey (with grey background). As IES boundary maintenance is required for both Type-A and Type-B IESs, only one generic IES is depicted. (a) Early-scnRNAs loaded into maternal (old MAC) expressed Twi1p (dark grey) are imported into the new MAC and (b) recognize ncRNA transcribed by an RNA polymerase, which triggers Ezl1p-mediated (blue diamond) H3K9me3 (blue small sphere) and H3K27me3 (green small sphere). (c) Pdd1p (dark purple) binds H3K9/K27me3 and has an essential, but elusive, function for subsequent Late-scnRNA production (dashed arrows with question mark). (d) Bidirectional transcription of IESs leads to (e) dsRNA-precursor formation. (f) The precursors are likely cleaved by Dcl1p to form Late-scnRNA duplexes. (g) By unknown means Late-scnRNAs are loaded into zygotically (new MAC) expressed Twi1p and Twi11p (olive-green) and the passenger strand is removed. (h) Late-scnRNA-loaded Twi1p and Twi11p can then further recognize complementary IESs and potentially MDS-regions and foster the spreading of heterochromatin by further recruiting Ezl1p and downstream Pdd1p. (i) Heterochromatin and Late-scnRNA production spreading is counteracted by Coi6p (dark green) with its interaction partners Coi7p (cyan) and Lia5p (blue) via an unknown mechanism. Coi6p likely recognizes H3K9me3. (j) Jmj1p (purple) mediated H3K27me3 demethylation also counteracts heterochromatin spreading by stopping expansion of heterochromatic histone modifications. (h) Moreover, it is possible that genomic *cis*-acting DNA elements (dark grey box) contribute to IES-MDS boundary formation by unknown mechanisms.

Although Ezl1p and Pdd1p were shown to be essential for the accumulation of Late-scnRNAs, it remains enigmatic which of the Late-scnRNA production steps is impacted by Ezl1p and Pdd1p (Fig. 42c) (Noto et al. 2015). The similarity in phenotype of $\Delta EZL1$ and $\Delta PDD1$ indicates that Ezl1p is likely required in its role as obligatory recruitment factor for Pdd1p (Liu et al. 2007). Pdd1p might then impact either bidirectional IES transcription or downstream processing/loading of Late-scnRNAs (Fig. 42c-g). This could involve the Pdd1p interaction partner Lia1p, which lacks detectable homology to any other protein, but is required for DNA elimination (Rexer and Chalker 2007). The spreading of Late-scnRNA production and heterochromatin formation is counteracted by

Coi6p, which is likely recruited to IESs via binding to H3K9me3 (Fig. 42i). Coi6p attracts its interaction partners Coi7p and Lia5p, which are also necessary to stop spreading (Fig. 42i). Coi6p might contribute to IES boundary formation by recruiting the yet-undefined biochemical activities of Coi7p and Lia5p. Alternatively, Coi6p may have boundary-enforcing activity itself similar to its *S. pombe* homologue Swi6. Swi6 is required for spreading and maintenance of RNAi and heterochromatin, however its RNA binding activity is also crucial for restricting heterochromatin (Hall et al. 2002; Hayashi et al. 2012; Keller et al. 2013). In addition to the Coi6p-Coi7p-Lia5p complex, Jmj1p also counteracts spreading by demethylation of H3K27me3 (Fig. 42j), a role similar to that of H3K9me demethylase *LSD1* in *S. pombe* (Lan et al. 2007). Whether Jmj1p removes H3K27me3 specifically at IES-MDS boundaries or enhances its turnover in general remains to be tested. However, in support of the former scenario histone demethylases such as *S. pombe* LSD1 and *N. crassa* DMM-1, have been detected specifically at the border of heterochromatin (Lan et al. 2007; Honda et al. 2010). The presence of IES boundaries that are neither affected by absence of Coi6p or of Jmj1p may suggest that either these two proteins act redundantly at these boundaries or that another boundary-forming mechanism independent of these two proteins is in place. The *cis*-acting DNA elements, which have been described for few IESs, might act as a boundary-forming mechanism in the latter scenario (Fig. 42k) (Godiska et al. 1993; Chalker et al. 1999). Also, regulation of histone acetylation might be a promising mechanism to prevent spreading, as indeed inhibition of HDACs seemed to partially inhibit DNA elimination (Duharcourt and Yao 2002).

3.8 Potential primary borders for Late-scnRNA production?

Building on the model outlined above, a speculative primary layer of IES boundary regulation can be proposed. This is based in two related observations and based on previously published data (Noto et al. 2015, Fig. 6A and B): Firstly, Late-scnRNAs can be detected on Type-B IESs as early as 7.5 hpm and appear to be expressed simultaneously on the whole IES just shortly after new MAC formation. This seems to contradict heterochromatin-based *cis*-spreading of Late-scnRNA production, which would predict an initial peak of Late-scnRNA production around the trigger site (the A-repeat) with subsequent spreading across the IES. Secondly, even at 7.5 hpm the boundaries of Late-scnRNA production are remarkably sharp, which argues against a solely nucleosome-based chromatin process.

I therefore speculate that there might be two functionally distinct boundary-forming mechanisms. The initial, primary boundary may act at ~7.5 hpm before the powerful RNAi-heterochromatin feedback loop (described above) is fully active and confines initial Late-scnRNA production to IESs. The secondary boundary may function later to restrict the RNAi-heterochromatin feedback loop to IESs and may be based on chromatin, including the Coi6p-complex and Jmj1p. This “dual boundary model” makes two testable predictions: Firstly, the initial wave of Late-scnRNA production might only depend to the Early-scnRNA trigger, not on Ezl1p and Pdd1p. Secondly, the “secondary boundary proteins”, such as Coi6p and Jmj1p, are not required for the precise production of Late-scnRNA from IESs in the initial phase (~7.5 hpm), but in absence of these proteins the primary boundary is later (~ 10.5 hpm) “overwritten” and Late-scnRNAs start to be expressed from outside IESs. Testing these predictions and understanding the biology of IES boundaries in *Tetrahymena* will be an exciting challenge for the future.

4. Materials and Methods

4.1 Growth and mating conditions for *Tetrahymena*

Wild-type strains used were B2086 and Cu428, available from the *Tetrahymena* Stock Center at Cornell University (<http://tetrahymena.vet.cornell.edu>). All genetically modified strains generated in this study are derived of these strains. Cells were stored unfrozen in soy bean stocks at room temperature (RT) or frozen in liquid nitrogen for long term storage (Sweet and Allis 2006; Cassidy-Hanley et al. 1995). In preparation of matings vegetative cells were grown with gentle agitation (90 rpm) at 30 °C in SPP medium to a concentration of $3.5 - 5 \times 10^5$ cells/ml, washed with 10 mM Tris pH 7.5 and starved for 8 – 24 h at 30°C with a density of $\leq 8.0 \times 10^5$ cells/ml (Gorovsky et al. 1975). Mating was induced by mixing equal numbers of cells derived of two strains with different mating types and incubation at 30°C without agitation. Mating efficiency was assessed at 8 – 10 hpm by fixing cells and staining with DAPI. A cell counter was used to determine cell concentrations (Beckman Coulter).

4.2 Generation of transgenic strains

Somatic (MAC) and germline (MIC) transformations were performed using homologous recombination by creating constructs that contain 5'- and 3'-flanking regions of the target locus and a selection cassette conferring neomycin resistance. Linearized constructs were introduced into *Tetrahymena* cells using biolistic gene transformation with a Biolistic Particle Delivery System (Bio-Rad) (Bruns and Cassidy-Hanley 2000). Transformants were selected with paromomycin (Sigma-Aldrich) and verified by PCR.

4.3 Genomic DNA extraction

Genomic DNA was extracted from vegetative or mating cells using CHAOS cell lysis buffer (25 mM Tris pH 8.0, 4 M guanidine thiocyanate, 0.5% SDS, 0.1 M β -mercaptoethanol) and an equal volume PEB buffer (100 mM Tris pH 8.0, 10 mM EDTA, 1% SDS). DNA was extracted from the resulting mixture by Phenol-chloroform extraction, precipitated using 2-Propanol and once washed with 70 % Ethanol. The DNA pellet was

resuspended in TE buffer (10 mM Tris pH 7.5, 1 mM EDTA) and subsequently treated with RNase A (50 µg/ml) and Proteinase K (50 µg/ml). DNA extraction by Phenol-chloroform and 2-Propanol precipitation were repeated once more and the DNA pellet was resuspended in H₂O.

4.4 Polymerase Chain Reaction (PCR)

Genotyping, transformation verification and overelimination PCR assays were performed using Taq polymerase (5 PRIME) according to company recommendations. Genotyping and transformation PCR verifications requiring amplification of large products were carried out with Extender Polymerase (5 Prime). For large-scale genotyping 20µl Taq polymerase PCRs were performed, replacing genomic DNA for 1 µl densely grown cell culture (referred to as direct PCR). In such cases the initial denaturing PCR step was extended to 10 min. For molecular cloning Primestart polymerase (Takara) was used.

4.5 Overelimination PCR assay

20 µl PCR was performed with Taq polymerase (5 Prime) for 35 cycles and with annealing temperature 50°C. PCR elongation time was chosen such that the MAC band was efficiently amplified. PCR reactions were mixed with 6x Purple Gel Loading Dye (NEB) and 5 µl were loaded on a 2% LE Agarose (Biozyme) gel. Gels were run for 25 min at 135 V in 0.5x TAE and then stained for 40 min at RT with Ethidiumbromide (0.5 µg/ml) dissolved in 0.5x TAE. Gels were imaged on a UV Gel-Imager (INTAS).

4.6 TOPO TA cloning of overelimination PCR products

100 µl overelimination PCR (with 30 min final elongation) were run on a 2% Agarose gel as described above. Overelimination PCR products were purified from the gel using the Nucleospin Gel and PCR clean-up Kit (Macherey-Nagel) and eluted in 35 µl H₂O. PCR products were cloned using the TOPO TA Cloning Kit (Invitrogen) according to manufacturer's recommendations, but replacing One Shot E.coli cells for competent DH5 α E. coli. Selection of clones was done with Ampicillin and blue/white selection using 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal). Plasmid DNA of selected clones

was purified using the Quiagen Spin Miniprep Kit and subjected to Sanger-sequencing (using UNI and REV primers).

4.7 Site-directed mutagenesis

Point mutations were introduced by two-step PCR-directed mutagenesis. Primestart Polymerase (Takara) was used for this purpose. In the first step PCR primers were used such that two nested PCR products were amplified. Mutations were introduced with the “inner” primers of each primer pair. In the second PCR nested PCR products were joined by using flanking primers. PCR products were digested by restriction enzymes and cloned into the respective vector. Plasmid sequences were verified by Sanger-sequencing.

4.8 *COI6* knock-out strains

Complete (MIC+MAC) Δ *COI6* strains were published by Woehrer et al. and used in this study (Fig. 17 and Woehrer et al. 2015, Suppl. Fig. 2). The germline *COI6* locus was targeted for homologous recombination using biolistic gene transformation. The *COI6* locus was replaced with a *NEO4* cassette (*MTT1* promoter driven neomycin resistance cassette). Full replacement of the *COI6* locus was tested by PCR (Fig. 17B).

4.9 CRISPR/Cas9-mediated knock-out of *COI7*

Standard molecular cloning techniques were used to introduce DNA oligonucleotides containing the guide RNA sequence 5'-AGGCTCTCAAATAAATTAGT-3' into the BbsI digested pBNMB1-HA-Cas9Tti-U6gRNA vector, creating pBNMB1-HA-Cas9Tti-U6gRNA-COI7T1 (containing *BTU1* homology flanks, H4 promoter driven neomycin resistance, *MTT1* promoter driven codon optimized Cas9, U6 promoter and terminator driven COI7-guideRNA-scaffold fusion). The vector was linearized and introduced to the MAC *BTU1* locus of Cu428 WT cells by biolistic transformation (Bruns and Cassidy-Hanley 2000). Transformants were assorted until growing well in 10 mg/ml paromomycin. Cas9 expression was induced by addition of 1 μ g/ml CdCl₂ for 5 h during vegetative growth. CdCl₂ was washed out using 10 mM Tris pH 7.5 and cells were starved over night (o/n).

To ensure MIC transmission of mutations cells were mated with the B2086 WT strain and single mating pairs were isolated at 6 – 9 hpm. Progeny were selected using 15µg/ml 6-Methylpurine (Byrne et al. 1978). To identify mutations a ~400 bp fragment of the *COI7* locus was amplified using direct PCR (Primers: Seq59, *COI7_CasMut_cFW1*), purified using the Illustra ExoProStar 1-Step Kit (GE Healthcare) and subjected to Sanger-Sequencing (with Seq59 as sequencing primer). Mutant heterozygote strains were crossed to generate homozygote $\Delta COI7$ strains. These were genotyped in two direct-PCR based rounds: The first round was performed as described above, the second round made use of a nearby IES to confirm MIC *COI7* locus mutations (Extender PCR Polymerase (5 Prime), Primers: *COI7_CasMut_cFW1* and *COI7_Mic_CasMut_cRV3*). PCR products were purified using the Illustra ExoProStar 1-Step Kit as above and subjected to Sanger sequencing using Seq59 as sequencing primer.

4.10 *COI6* rescue strain generation

$\Delta COI6$ strains A2 and B6 were used as background strains for transformation of different rescue constructs. Constructs were created using the pBNMB1-HA plasmid (Shang et al. 2002). cDNA of WT *COI6* was cloned into this vector generating pBNMB1-HA-*COI6*. Alternatively, site-directed mutated *COI6_CDmut*, *COI6_CSDmut1* and *COI6_CSDmut2* were introduced to the same plasmid (Fig. 19). This puts all *COI6* expression cassettes under control of the Cadmium-inducible *MTT1* promoter. Furthermore, the cassette contains the *NEO5* resistance marker, which expresses the neomycin resistance gene under control of the constitutively expressed H4 promoter. This cassette was introduced into the *BTU1* locus of the parental MAC of $\Delta COI6$ cells by using homologous flanking regions and biolistic gene transformation. Transformants were selected using a previously tested concentrations of paramomycin, that are sufficient to kill $\Delta COI6$ background strains A2 and B6 in absence of Cadmium. Highly assorted cells were used for subsequent assays as assessed by PCR.

4.11 SDS-PAGE and Western blotting analysis

For preparation of whole cell extract (WCE) from starved or conjugating cells trichloroacetic acid (TCA) was added directly to the cell suspension at 10 % (e.g. 900 µl of 5×10^5 cell/ml + 100 µl TCA). Vegetative cells were first pre-washed once with 10 mM

Tris pH 7.5. Cells were put on ice for 5 min and then centrifuges at 4°C at 10,000 x g for 5 min. Precipitate was resuspended in 75 µl Mono Q H₂O and then supplemented with 25 µl 4x SDS loading buffer (125 mM Tris-HCl pH 6.8, 2.3% SDS, 10% Glycerol, 4% beta-mercaptoethanol, 10 µg/ml BPB). To adjust the pH to neutral ~3 µl 2 M Tris-Base was added. Protein samples were denatured for 10 min at 95°C and frozen at – 20°C.

For western blot analysis proteins were separated in self-made SDS-polyacrylamide gels according to standard procedures or in precast 8-16% gels (Amersham). Proteins were separated at 80-140 V in SDS-running buffer (25 mM Tris-Base, 192 mM glycine, 0.1% SDS). PageRuler Prestained Protein Ladder (Fermentas) was used as size marker. Using the semi-dry Trans Blot device (Bio-Rad) and blotting buffer (50 mM Tris-Base, 192 mM glycine, 20% methanol, 0.1% SDS) proteins were transferred from gels to PVDF membranes, that were pre-rinsed in 100% Methanol (≥ 3 min) and in blotting buffer (≥ 30 min). After transfer, membranes were rinsed in PBS for 10 min and then blocked for 40 – 120 min with blocking solution (1% BSA, 1% milkpowder PBST (PBS + 0.1% Tween-20), all w/v). Blocked membranes were incubated with primary antibody o/n at 4°C. Primary antibodies were diluted in blocking solution as follows: α - alpha-tubulin (12G10, Developmental Studies Hybridoma Bank) 1:20000; α -Coi6p(r) IgG purified/ HBS (rabbit #477, Eurogentec) 1:2000; α -Coi7p IgG purified/HBS (rabbit #487, Eurogentec) 1:2000; α -Lia5p IgG purified/PBS 1:5000 (rabbit #4813, Gramsch); α -Pdd1p (ab5338, abcam) 1:2000. Membranes were washed 3x 10 min with PBST and then incubated with respective secondary HRP-coupled antibody (Jackson ImmunoResearch Laboratories Inc.) 1:10000 diluted for 1 h at RT. Membranes were washed again 3x 10 min in PBST and then developed using ECL Western Blotting Detection Reagent (Amersham) using either exposure to film (Hyperfilm ECL, GE Healthcare) or the ChemiDoc MP Imaging System (Bio-rad).

4.12 DMP cross-linking of antibodies to protein A dynabeads

All steps were carried out at RT. Per immunoprecipitation (IP) 50 µl Dynabeads protein A (Invitrogen) were cross-linked to 30 µl antibody serum. Dynabead slurry was washed twice briefly with IP buffer (20 mM Tris pH 7.5, 100 mM NaCl, 2 mM MgCl₂, 2 mM CaCl₂, 0.1% Tween-20) and then incubated for 30 min with 200 µl IP buffer and 30 µl antibody serum. Unbound antibody was washed out by three 3 min washes with 20 bead volumes (bv) IP buffer. Dynabeads were equilibrated with three 1 min 10 bv washes of 0.2 M Na-

borate, pH 9.2 and then incubated with 20 bv of freshly prepared 20 mM DMP (dimethyl pimelimidate dehydrochloride, Sigma-Aldrich) in 0.2 M Na-borate pH 9.2 for exactly 30 min. The cross-linking reaction was stopped by two 10 min washes with 10 bv of 0.2 M Tris-HCl, pH 8.0, followed by two 3 min washes with IP buffer. Non-cross-linked antibody was washed out by two 1 min washes with 10 bv 0.1 M glycine, pH 2.0. Dynabeads were re-equilibrated by three 5 min washes with 10 -20 bv of IP buffer and stored at 4 °C until use (≥ 3 days).

4.13 Cell lysate preparation for immunoprecipitation

Per IP $\sim 1 \times 10^7$ mating cells (15 ml of 7×10^5 cells/ml) were harvested at 8 hpm and immediately put on ice. All subsequent procedures were carried out at 4°C. Cell volume was estimated and 1 volume (vol.) of 2x IP buffer was added. The cell suspension was filled up to 1 ml with 1x IP buffer and supplemented with Protease Inhibit Cocktail Tablets (without EDTA) (Roche) dissolved in H₂O ad 2x concentration and 0.5 mM PMSF (Phenylmethylsulfonyl Fluoride) dissolved in 2-Propanol. Cells were lysed by sonication and afterwards again 0.5 mM PMSF was added. Insoluble material was pelleted by 20 min centrifugation at 17000 x g. The supernatant was kept as cell lysate, frozen in liquid nitrogen and stored at -80°C.

4.14 Coi6p immunoprecipitation for mass spectrometry

IPs were carried out at 4°C. For mass spectrometry analysis of Coi6p IPs, 5x amounts of cell lysate (5 ml of 1×10^7 cells/ml in IP buffer) and dynabeads (250 μ l dynabeads + 150 μ l pre-immune or antibody serum (α -Coi6p(r), rabbit #477, eurogentec) were used. Antibody was cross-linked to beads using DMP as described above. IP buffer was: 20 mM Tris pH 7.5, 100 mM NaCl, 2 mM MgCl₂, 2 mM CaCl₂, 0.1% Tween-20. Cell lysate and dynabeads were incubated for 2 h at 4°C and then washed at 4 °C, 3x 5 min with 40 bv and 6x ~ 1 min with 4 bv of IP buffer without detergent. Proteins were eluted using 100 μ l of 0.1 M glycine HCl pH 2.0 for 10 min at 25°C, shaking at 1200 rpm. 10 μ l 1 M Tris pH 9.0 was added to the elution to neutralize the pH. Samples were then digested with Trypsin and analyzed by mass spectrometry.

4.15 NanoLC-MS Analysis.

The UltiMate 3000 HPLC RSLC nano system (Thermo Scientific) was utilized coupled to a Q Exactive Plus mass spectrometer (Thermo Scientific) and equipped with a Proxeon nanospray source (Thermo Scientific). Peptides were applied onto a trap column (Thermo Fisher Scientific, Amsterdam, Netherlands, PepMap C18, 5 mm × 300 µm ID, 5 µm particles, 100 Å pore size) at a flow rate of 25 µL min⁻¹ using 0.1% TFA as mobile phase. After 10 min, the trap column was switched in line with the analytical column (Thermo Fisher Scientific, Amsterdam, Netherlands, PepMap C18, 500 mm × 75 µm ID, 2 µm, 100 Å). Peptide elution was performed using a flow rate of 230 nL min⁻¹, and a binary 2h gradient, respectively 165 min. The gradient started with the mobile phases: 98% A (water/formic acid, 99.9/0.1, v/v) and 2%B (water/acetonitrile/formic acid, 19.92/80/0.08, v/v/v) and increases to 35%B over the next 120min, followed by a gradient to 90%B in 5 min. 90%B was applied for five min and decreases in 2min to the starting concentration of 98%A and 2%B for equilibration at 30°C.

The Q Exactive mass spectrometer was operated in data-dependent mode, using a full scan (m/z range 380-1650, nominal resolution of 70 000, target value 3E6) with subsequent MS/MS scans of the twelve most abundant ions. MS/MS spectra were acquired using normalized collision energy of 27%, isolation width of 2 and the target value was set to 1E5. Precursor ions selected for fragmentation (charge state 2 and higher) were put on a dynamic exclusion list for 30 s. In addition the underfill ratio was set to 20% resulting in an intensity threshold of 4E4. The peptide match feature and the exclude isotopes feature were enabled.

4.16 MS/MS Data Analysis

Peptide identification was initiated by loading the RAW-files into Proteome Discoverer (version 1.4.0.288, Thermo Scientific). Thereby created MS/MS spectra were searched using Mascot 2.2.07 (Matrix Science, London, UK) against the NCBI non-redundant protein sequence database, using the taxonomy *Tetrahymena*. Search parameters used were: Beta-methylthiolation on cysteine was set as a fixed modification, oxidation on methionine, acetylation on lysine, phosphorylation on serine, threonine and tyrosine, mono- and dimethylation on lysine and arginine, trimethylation on lysine and ubiquitinylation on lysine were set as variable modifications. Monoisotopic masses were

searched within unrestricted protein masses for tryptic peptides. The peptide mass tolerance was set to ± 5 ppm and the fragment mass tolerance to ± 0.03 Da. The maximal number of missed trypsin cleavages was set to 2. The result was filtered to 1% FDR using Percolator algorithm integrated in Proteome Discoverer. The localization of sites of variable modifications within a given peptide was performed with the tool ptmRS, integrated in Proteome Discoverer and based on phosphoRS (Taus et al. 2011).

4.17 Small scale Coi6p, Coi7 and Lia5p immunoprecipitation

Protease sensitivity of Lia5p led to adaptations of the above IP protocol to enhance Lia5p stability. IP buffer used was 50 mM Tris pH 7.5, 100 mM NaCl, 20 mM EDTA, 0.1% Tween-20. During cell lysate preparation (as above) 1 mM PMSF was added before and after sonication. Furthermore 3x Ultra Protease Inhibitor Cocktail (Roche) was added. After thawing the cell lysate for IP 1 mM fresh PMSF was supplemented. For Coi7p IPs 150 μ l dynabeads were cross-linked to 90 μ l serum (α -Coi7p serum, rabbit #487, eurogentec) as described above, for Coi6p IPs (α -Coi6p(r) serum, rabbit #477, eurogentec) and Lia5p IPs (α -Lia5p serum, rabbit #4813) 50 μ l dynabeads were cross-linked to 30 μ l serum. Pre-immune serum of the respective rabbit was used as negative control in all cases. Proteins were IPed from 1 ml of cell lysate for 2 h at 4 °C and washed 4x 5 min with 1 ml IP buffer, proteins were eluted in 35 μ l 0.1 M glycine pH 2.0 for 10 min at 25°C, shaking 1200 rpm. To adjust pH and salt concentration for SDS-PAGE 1/10th vol. of 0.75 M Tris-HCl pH 9.0, 1.25 M NaCl was added. Then 4x SDS-loading buffer was added (1/3rd vol.), samples were denatured for 10 min at 95°C and stored at -20°C.

4.18 scnRNA analysis

ScnRNA analysis by high throughput sequencing was performed as described by Schoeberl et al. (Schoeberl et al. 2012). Minor alterations were the reduction of input material to 1 x 10⁶ cells and analysis using multiplexed sequencing (TruSeq index sequence in reverse primer of reverse transcription step) by a HiSeq2000 instrument (Illumina). Mating pairs used for scnRNA-sequencing were: 1. B2086 x Cu428 (WT), 2. A2 x B15 (Δ COI6), 3. F20-1 x F35-4 (Δ COI6), 4. M1-42-1 x M4-64-2 (Δ COI7), 5. M2-30-2 x M3-124-1 (Δ COI7), 6. Ms x 4-2 (Δ LIA5) (courtesy of Douglas Chalker), 7. 1.2 x 2.2 (Δ JMJ1) (courtesy of Yifan Liu). Utilizing the Illumina HiSeq2000 50 nt single reads were

sequenced and 25 nt of each scnRNA-read were used for mapping to the MIC genome. scnRNA-seq reads were normalized as reads per kilobase of genome per million mapped reads (RPKM). The positions of the 5' ends of each scnRNA-read were mapped on 10 kb genomic loci (bin size 100bp) or analyzed using boundary-focused meta-analysis, which compiles results for 500 bp inside and outside each boundaries of 3,722 Type-A and 2,863 Type-B IESs. The position of the 13th nucleotide of each scnRNA was mapped in 10 bp bins for this analysis.

4.19 Immunofluorescence

General fixation for immunofluorescence (IF) was performed directly in starvation medium (10 mM Tris pH 7.5) for conjugating cells. 1/10 vol. of 37% formaldehyde was added and incubated for 5 min at RT. Then Triton-X was added to 0.25% (w/v) and incubation was continued for 25 min at RT. Cells were pelleted, resuspended in 3.7% formaldehyde, 3.4% sucrose solution, distributed on Superfrost Ultra-Plus glass slides (Thermo Scientific) and dried at RT for ~ 30 min. Slides were used immediately for IF stainings or stored at -20°C.

α -Coi7p antibody (IgG purified/HBS, rabbit #487, eurogentec) showed only specific staining in an adaptation of Schaudin's fixation method (Fig. 34D). 6 ml of conjugating cells were mixed with 20 μ l of *Partial Schaudin's* fixative (2 vol. of Saturated HgCl₂ (0.074g/mL) and 1 vol. of 100% EtOH). and incubated 5 min at RT, cells were washed twice with 6 ml 100% methanol and then resuspended in 1.5 ml 100% methanol. Cells were spread on Superfrost Ultra-Plus glass slides (Thermo Scientific), or were stored in suspension at -20°C.

All washing steps during stainings were performed at RT. Slides were washed in PBST (PBS + 0.1% Tween-20) for 10 min and incubated with blocking solution (3% BSA, 10% Normal Goat Serum, 0.05% Sodium Azide in PBST) for 2 h at RT. Primary antibodies (α -Coi6p(r) IgG purified/ HBS (rabbit #477, Eurogentec) 1:2000; α -Coi6p(gp) IgG purified/ HBS (guinea pig SATC54, Eurogentec) 1:2000; α -Coi7p IgG purified/HBS (rabbit #487, Eurogentec) 1:2000; α -Lia5p IgG purified/PBS 1:5000 (rabbit #4813, Gramsch); α -Pdd1p (ab5338, abcam) 1:2000; α -Pdd1p(gp) (guinea pig, kindly provided by Kensuke Kataoka) 1:2000) were diluted in blocking solution and incubated with slides o/n at 4°C. Slides were washed 3x 10 min with PBST and then incubated for 1 h at RT with the respective secondary antibody, which was coupled to Alexa-488, -568 or -647

(Invitrogen). Slides were washed 3x 5 min with PBST and then incubated with PBST + 20 ng/ml DAPI for 30 min and destained for 3 min in PBST. Stained cells were covered with a coverslip in ProLong Gold solution (Invitrogen). Pictures were acquired using a wide-field microscope (Axioplan2, Zeiss).

4.20 Chromatin Spreading

Conjugating WT cells at 14 hpm were used for chromatin spreading. 5 ml of cells in starvation medium were cooled on ice for 30 sec. Then 500 μ l of ice-cold formaldehyde/Triton-X mix were added (450 μ l 10% Triton-X + 50 μ l 37% formaldehyde). And incubated on ice for 25 min with occasional gentle agitation by inverting. Then 450 μ l 37% formaldehyde were added to the solution and incubated for 5 min on ice. Cells were pelleted and resuspended in 3.7% Formaldehyde, 3.4% Sucrose and distributed on Superfrost Ultra-Plus glass slides (Thermo Scientific). IF stainings were performed as described above.

4.21 DNA fluorescence in situ hybridization

Production of Tlr1 FISH probe: 2 μ g plasmid mix (equal amounts of pMBR-Tlr1, pMBR-IntBTlr, pMBR-Tlr1-4C1) were simultaneously digested and labeled with Cy3 in the following reaction (vol. = 200 μ l): 2 μ g plasmid was incubated at 15°C for 2 h in a solution containing 50 μ M dNTP mixture (without dTTP), 0.1 nM Cy3-dUTP (GE healthcare), 0.2 U DNaseI, 50 U DNA polymerase I (NEB) in nick translation buffer (50 mM Tris pH 7.5, 10 mM MgSO₄, 0.1 mM DTT, 0.05% BSA). Labeled DNA was purified from the reaction by Ethanol precipitation and resuspended in 400 μ l hybridization buffer (50% formamide, 10 % dextran sulfate in 2x SSC). The Tlr1 FISH probe was denatured for 3 min at 95°C and then stored at -20°C.

Tlr1 FISH stainings: Cells were fixed using formaldehyde (as described for normal IF) at 34 hpm. Cells on glass slides were hydrated in Mono Q H₂O for $\geq$ 3 min. 1 M Na-thiocyanate was applied to the glass slide with a HybriWell (Invitrogen) for 15 min at 90°C, followed by a quick wash in Mono Q H₂O and incubation with 70% Formamide, 2x SCC at 70°C for 2 min. After another quick wash with Mono Q H₂O, water was removed and 7 μ l Tlr1 FISH probe were applied to the cells, covered with a cover slip and sealed

by rubber cement (Marabu). Slides were incubated at 80°C for 10 min, and then for ≥ 16 h at 37°C in a dark, moist chamber.

4.22 ChIP-sequencing

All steps were performed at 4°C unless otherwise noted.

Preperation of input material by nuclear isolation: Conjugating cells at 12 hpm were resuspended in TMSN buffer (10 mM Tris-HCl pH 7.5, 250 mM Sucrose, 10 mM MgCl₂, 3 mM CaCl₂, 0.016% NP-40, 1 mM PMSF, 1x Complete proteinase inhibitor cocktail (PI) (Roche) and 0.1 µg/ml DAPI) and disrupted by 15 strokes with a dounce homogenizer. Nuclei were pelleted at 4500 x g for 5 min. The upper white part of the nuclei pellet was resuspended in TMSN w/o PI and centrifuged at 4500 x g for 3 min. This step was repeated once more. Nuclei were fixed at RT in PBSN (PBS + 0.016% NP-40) supplemented with 1.25 mg/ml Di(N-succinimidyl) glutarate (DSG) (Sigma) for 30 min. Nuclei were pelleted at 9000 x g for 5 min and washed twice with PBSN. A second step of fixation was then conducted at RT with 1% paraformaldehyde (PFA) (Sigma) in PBSN for 10 min. The fixation reaction was stopped by addition of 0.25 M glycine at RT for 5 min. Nuclei were pelleted at 9000 x g for 5 min and washed twice with PBSN containing 0.1 µg/ml DAPI. Nuclei were aliquoted in PBSN to 1.5 ml tubes, centrifuged at 4500 x g for 5 min and nuclei pellets were frozen in liquid nitrogen and stored at -80°C. Enrichment of new MACs was achieved by FACS sorting nuclei prepared as above in PBSN with 0.1 µg/ml DAPI with a FACSAria III device (BD biosciences) according to forward scatter (size) and DAPI intensity. Purity of new MAC fractions retrieved was tested using Tlr1 DNA FISH in combination with α-H3K4me3 antibody (Abcam) (Tlr1- nuclei = parental MAC; Tlr1+ and H3K4me3+ nuclei = new MAC; Tlr1+ and H3K4me3- nuclei = MIC). Samples of high purity (≥ 85% new MAC with few ~ 1 % parental MAC) were used for subsequent chromatin preparation.

Preperation of chromatin: High purity new MAC samples were sheered by sonication to 150-500 bp fragments by 5x 30 pulses (power 20%, pulses 50%) of a Omni Ruptor 250 probe sonicator (OMNI International Inc.) in sonication buffer (50 mM Tris, 10 mM EDTA, 0.1 % SDS, 1x Complete protease inhibitor, 1 mM PMSF). Chromatin was aliquoted, frozen in liquid nitrogen and stored at -80°C until use.

ChIP: Per sample 20 mg of Protein A Sepharose Cl-4B (GE healthcare) were resuspendend in TE buffer and washed twice, followed by three washes with ChIP buffer (25 mM Tris-HCl pH 8.0, 167 mM NaCl, 5 mM EDTA, 0.05% SDS, 1% Triton X-100, 1x

Complete protease inhibitor, 1 mM PMSF). Beads were blocked for 1 h in 90 μ l ChIP buffer supplemented with 10 μ l of 10 mg/ml BSA (NEB). Then 10 μ l of beads suspension were used for coupling ChIP-antibodies to it and 25 μ l were used to pre-clear the respective ChIP-input material. For Pre-clearing, the vol. of chromatin derived of 2×10^6 new MACs was adjusted to 500 μ l with sonication buffer, 2x ChIP buffer (334 mM NaCl, 2% Triton X-100, 1x Complete protease inhibitor) was added to 1 ml total vol., and sample was incubated for 2 h. For binding antibodies to 10 μ l blocked sepharose beads, these were resuspended in 500 μ l ChIP buffer supplemented with the respective antibody (50 μ l (α -Coi6p(r) IgG purified/ HBS (rabbit #477, Eurogentec); 50 μ l α -Coi7p IgG purified/HBS (rabbit #487, Eurogentec); 50 μ l α -Lia5p IgG purified/PBS (rabbit #4813, Gramsch), 10 μ l α -Pdd1p (ab5338, abcam); 20 μ l α -H3K9me3 (millipore); 10 μ l α -H3K27me3 (gift of Thomas Jenuwein)) and incubated for 2 h. Beads were washed with ChIP buffer three times. Then the pre-cleared chromatin solution was applied to the antibody-bound sepharose beads and incubated o/n. This was followed by extensive wash steps, each 10 min: 2x with low-salt wash buffer (20mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), 2x High-salt wash buffer (20mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), 2x LiCl wash buffer (10mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 1% Na-deoxycholate, 1% NP-40), and 2 x with TE. Precipitated chromatin was eluted twice with 250 μ l of 100 mM NaHCO₃, 1% SDS for 15 min. For decross-linking NaCl was added to the sample ad 200 mM and then incubated o/n at 65°C. DNA was extracted by 2-Propanal precipitation after RNaseA and ProteinaseK treatment and resuspended in 30 μ l EB buffer (Quiagen). ChIP-sequencing libraries were prepared from immunoprecipitated DNA and from 20 ng of input DNA by using the NEBNext DNA Library Prep Reagent Set for Illumina (NEB) and KAPA real-Time Library Amplification Kits (KAPA Biosystems). 50 nt single reads were sequenced using the Illumina HiSeq2000 device. The first 36 nt of each read were used for mapping to the MIC genome. ChIP reads were divided by input reads. And the resulting quotient was transformed using log₂. Reads were mapped on 100 kb genomic loci or 10 kb genomic loci (bin size 100 bp) or analyzed using boundary focused meta-analysis which compiles results for 500 bp inside and outside each boundaries of 3,715 Type-A and 2,863 Type-B IESs (bin size 10 bp).

4.23 List of antibodies

Name of antibody	Immunogen	Company (ID#)
α -Coi6p(r)	recombinant full-length MBP-Coi6p	Eurogentec
α -Coi6p(gp)	recombinant full-length MBP-Coi6p	Eurogentec
α -Coi7p	recombinant full-length MBP-Coi7p	Eurogentec
α -Lia5p	recombinant full-length MBP-Lia5p	Gramsch
α -Pdd1p	Synthetic peptide derived from residues 300 - 400 of Tetrahymena sp. Pdd1.	Abcam (ab5338)
α -Pdd1p(gp)	Synthetic peptide CLGSKGQSQQVEKEQATNS, corresponding to aa 315-332 of Pdd1p	
α -H3K9me3	KLH-conjugated, 2X-branched synthetic peptide containing the sequence ...AR[me3K]ST	Millipore(#07-442)
α -H3K27me3	Synthetic methylated peptide corresponding to H3K27me3 and surrounding amino acids	Gift of Thomas Jenuwein
α - alpha-tubulin	alpha tubulin, Tetrahymena thermophila and Tetrahymena pyriformis (mixture)	Developmental Studies Hybridoma Bank (12G10)

4.24 List of oligonucleotides

Personal Index	Name	5'-3' sequence
Sanger sequencing and genotyping		
Seq010	BTU1_3UTR_Rv	TCTCTACTATTACAACAACAGTAAGATGC
Seq011	MTT1seqFor	GCTACGTGATTCACGATTTATGC
Seq014	Coi6_tetORF_RV1	TAATTTGGTTATCTGATATTCCACC
Seq015	Coi6_tetORF_FW1	GAGCCTTCTTAGGCATAAAACCACC
Seq016	Coi6_tetORF_RV2	CATTGCTGTTTAATTCGTTTTCCAG
Seq017	Coi6_tetORF_FW2	TATCTTAAAAGCAATTACCAAAGCC
Seq020	MBP_3'_FW	AGACAAACCGCTGGGTGCCG
Seq021	MBP_5'_RV	TAGCCACCAAAGCGGTCGTG
Seq022	ecCoi6_FW1	AAATACGAAAGAACAGATCGAACCG
Seq023	ecCoi6_RV1	GTTCTGACTGTTTTGTTGTTACCG
Seq024	ecCoi6_FW2	ACAATAACCAGCAACAGAATAACGG
Seq025	ecCoi6_FW3	TCCGCTGAAGATCCTGAAACATGCG
Seq026	ecCoi6_RV2	ACGATCTTTCAGTTCGTTGAAAGCG
Seq027	ecCoi6_RV3	AAACGGCGGCAGTTTCTTTTTGGC
Seq040	BTU1_3'seqRV	AAAAGTTATTTGTTGATTTAGAGAG
Clon301	COI6-KO-3RV2-SacII	CCGCCGCGGTTTTTAATCAGCTTAATTTAGGTC
Clon302	COI6-KO-5FW-XhoI	GCGCTCGAGACTGTAGCTATTAATAACTATTAG
Seq055	ecCoi7_FW1	GGTTGCAATATCTCAAGCTATAGCC
Seq056	ecCoi7_RV1	TCTAATTGAGCTAAGTCCGGAAGCG
Seq059	ttCoi7_Cas9gt_RV2	ACATATCTCTAATCTTAGCAGTCTC
-	COI7_CasMut_cFW1	AGTACCTTAATAATTGGATCTATCTC
-	COI7_Mic_CasMut_cRV3	AAAGTTAGTAATATAGCTGTTCGTGC
	UNI	CCCAGTCACGACGTTG
	REV	CACAGGAAACAGCTATG
	malE	GGTCGTCAGACTGTCGATGAAGCC
Cloning recombinant proteins		
Clon212	TEV-ecCoi7_BamHI_FW	CGCGGATCCGAAAATCTTTATTTTCAAGGCATGGAAC
		TCCTTGACTCCGACATCGG
Clon213	ecCoi7_2xSTOP_HindIII	CCCAAGCTTTTATTATTTATAGCGGCTAATTTTGTCT
		GC
COI6 Rescue constructs		

MATERIALS AND METHODS

Clon053	tetCoi6_BlnI_RV	GCCGCCCTAGGGCCACGCGTGCCATCGATG
Clon054	tetCoi6Y39A_FW	CCGGTTTATCCCATAGAAGAAGAAGCTAATGTTGAAT ATTTATATGG
Clon055	tetCoi6Y39A_RV	CCATATAAATATTCAACATTAGCTTCTTCTCTATGGG ATAAACCGG
Clon056	tetCoi6_W60AY63A_FW	GGTTAAATTAAGTATTGTGTAAAAGCTGAAAATGCTA CTTTTGAAGAAAGTTCATTGAGC
Clon057	tetCoi6_W60AY63A_RV	GCTCGAATGAACCTTTCTTCAAAAGTAGCATTTTCAGC TTTTACACAATACTTAATTTAACC
Clon058	tetCoi6_F438AK439A_FW	GCAGTCTTTACTTTAAAGTTATAGCTGCTCCTAGACC TTCTGGTACAGTTCC
Clon059	tetCoi6_F438AK439A_RV	GGAAGTGTACCAGAAGGTCTAGGAGCAGCTATAACT TTAAAGTAAAGACTGC
Clon060	tetCoi6_YYE468-470A_FW	CCCAAGAGTTTTAATGGAAGCTGCTGCTTAACATGCA ATTTTGTTAGATCC
Clon061	tetCoi6_YYE468-470A_RV	GGATCTAACAAAATTGCATGTTAAGCAGCAGCTTCCA TTAAACTCTTGGG
Clon066	tetCoi6_BamH1_noNheI_FW	GCCGGATCCATGGCTAAGATAAAATACGAAGGAAG

Overelimination assay

Clon258	bIES320-321_FW2	GAAAAAGTATAATCTTATATATGCAAG
Clon259	bIES320-321_RV2	TAAAAACAATTAAAGATATATACCTG
Clon260	bIES2246-2248_FW2	ATATTTTTTAAAGTTTTAGATTGGAG
Clon261	bIES2246-2248_RV2	ATGCTTATTTAATTTACAAAAGGTG
Clon262	bIES2279-2280_FW2	TAATCATAATATTATAAATTGTGGG
Clon263	bIES2279-2280_RV2	GAATGTTTTTTTTTAATGAATAGAG
Clon268	bIES381-382_FW3	GTAAAAATAAGCAATTTAATAATTCCTAC
Clon269	bIES381-382_RV3	CATAAAAAAGTGAAAGGTCTATACCAG
Clon295	bIES2301-2302_FW1	TAAATTTAGTAACCAAAGCATTATGC
Clon296	bIES2301-2302_RV2	ATTTAGTTTTGGTAACTGTAATTTGGG

Abbreviations

A. thaliana	-	Arabidopsis thaliana
aa	-	Amino acids
Ago	-	Argonaute
ANP32	-	Acidic (leucine-rich) nuclear phosphoprotein 32
	kDa	
bp	-	Base pairs
BTU1	-	Beta-tubulin 1
C. elegans	-	Caenorhabditis elegans
CD	-	Chromodomain
CDS	-	Chromoshadow domain
ChIP-seq	-	Chromatin immunoprecipitation sequencing
COI6	-	Conjugation-induced gene 6
D. melanogaster	-	Drosophila melanogaster
DMP	-	Dimethyl pimelimidate
ds	-	Double stranded
DSG	-	Disuccinimidyl glutarate
FACS	-	Fluorescence-activated cell sorting
FISH	-	Fluorescence in situ hybridization
gRNA	-	Guide RNA
H. sapiens	-	Homo sapiens
HDAC	-	Histone deacetylase
HKMT	-	Histone lysine methyltransferase
HP1	-	Heterochromatin protein 1
IES	-	Internal eliminated sequence
IP	-	Immunoprecipitation
IP	-	Immunoprecipitation
KO	-	Gene knock-out
KO	-	Knock-out
LRR	-	Leucin-rich repeat
MBP	-	Maltose binding protein
MDS	-	Macronuclear destined sequence
miRNA	-	Micro RNA

ABBREVIATIONS

mRNA	-	Messenger RNA
MTT1	-	Metallothionein
N. crassa	-	Neurospora crassa
ncRNA	-	None-coding RNA
nt	-	Nucleotides
o/n	-	Over night
ORF	-	Open reading frame
PBS	-	Phosphate buffered saline
PCR	-	Polymerase chain reaction
PFA	-	Paraformaldehyde
piRNA	-	Piwi interacting RNA
Piwi	-	P-element induced wimpy testis
RISC	-	RNA-induced silencing complex
RITS	-	RNA-induced transcriptional silencing
RNAi	-	RNA interference
RT	-	Room temperature
S. pombe	-	Schizosaccharomyces pombe
scnRNA	-	scanRNA
scnRNA-seq	-	scnRNA-sequencing
siRNA	-	small interfering RNA
T. thermophila	-	Tetrahymena thermophila
Twi	-	Tetrahymena Piwi
WT	-	Wild-type

References

- Al-Sady B, Madhani HD, Narlikar GJ. 2013. Division of Labor between the Chromodomains of HP1 and Suv39 Methylase Enables Coordination of Heterochromatin Spread. *Molecular Cell* **51**: 80–91.
- 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 & Development* **22**: 2228–2241.
- Baldauf SL. 2003. The deep roots of eukaryotes. *Science* **300**: 1703–1706.
- Bi X, Yu Q, Sandmeier JJ, Zou Y. 2004. Formation of boundaries of transcriptionally silent chromatin by nucleosome-excluding structures. *Molecular and Cellular Biology* **24**: 2118–2131.
- 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, Cassidy-Hanley D. 2000. Biolistic transformation of macro- and micronuclei. *Methods Cell Biol* **62**: 501–512.
- Buhler M, Verdel A, Moazed D. 2006. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**: 873–886.
- Byrne BC, Brussard TB, Bruns PJ. 1978. Induced resistance to 6-methylpurine and cycloheximide in *tetrahymena*. I. Germ line mutants of *T. thermophila*. *Genetics* **89**: 695–702.
- Canzio D, Larson A, Narlikar GJ. 2014. Mechanisms of functional promiscuity by HP1 proteins. *Trends in Cell Biology* **24**: 377–386.
- Cassidy-Hanley D, Smith HR, Bruns PJ. 1995. A simple, efficient technique for freezing *Tetrahymena thermophila*. *J Eukaryot Microbiol* **42**: 510–515.
- 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.
- Chalker DL, La Terza A, Wilson A, Kroenke CD, Yao MC. 1999. Flanking regulatory sequences of the *Tetrahymena* R deletion element determine the boundaries of DNA rearrangement. *Molecular and Cellular Biology* **19**: 5631–5641.
- Chalker DL, Yao MC. 2001. Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in *Tetrahymena thermophila*. *Genes & Development* **15**: 1287–1298.

REFERENCES

- 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.
- Chung P-H, Yao M-C. 2012. *Tetrahymena thermophila* JMJD3 Homolog Regulates H3K27 Methylation and Nuclear Differentiation. *Eukaryot Cell* **11**: 601–614.
- Couvillion MT, Lee SR, Hogstad B, Malone CD, Tonkin LA, Sachidanandam R, Hannon GJ, Collins K. 2009. Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of *Tetrahymena thermophila*. *Genes & Development* **23**: 2016–2032.
- Coyne RS, Nikiforov MA, Smothers JF, Allis CD, Yao MC. 1999. Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation. *Molecular Cell* **4**: 865–872.
- Doudna JA, Charpentier E. 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**: 1258096–1258096.
- Duharcourt S, Yao M-C. 2002. Role of histone deacetylation in developmentally programmed DNA rearrangements in *Tetrahymena thermophila*. *Eukaryot Cell* **1**: 293–303.
- 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.
- Eissenberg JC, Elgin SCR. 2014. HP1a: a structural chromosomal protein regulating transcription. *Trends Genet* **1**–8.
- Elgin SCR, Reuter G. 2013. Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harbor Perspectives in Biology* **5**: a017780.
- Fabian MR, Sonenberg N, Filipowicz W. 2010. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* **79**: 351–379.
- Fass JN, Joshi NA, Couvillion MT, Bowen J, Gorovsky MA, Hamilton EP, Orias E, Hong K, Coyne RS, Eisen JA, et al. 2011. Genome-Scale Analysis of Programmed DNA Elimination Sites in *Tetrahymena thermophila*. *G3 (Bethesda)* **1**: 515–522.
- 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.
- Frankel J. 2000. Cell biology of *Tetrahymena thermophila*. *Methods Cell Biol* **62**: 27–125.
- Gao S, Xiong J, Zhang C, Berquist BR, Yang R, Zhao M, Molascon AJ, Kwiatkowski SY, Yuan D, Qin Z, et al. 2013. Impaired replication elongation in *Tetrahymena* mutants deficient in histone H3 Lys 27 monomethylation. *Genes & Development* **27**: 1662–1679.

- Gao Z, Liu H-L, Daxinger L, Pontes O, He X, Qian W, Lin H, Xie M, Lorkovic ZJ, Zhang S, et al. 2010. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* **465**: 106–109.
- Gibbons IR, Rowe AJ. 1965. Dynein: A Protein with Adenosine Triphosphatase Activity from Cilia. *Science* **149**: 424–426.
- Godiska R, James C, Yao MC. 1993. A distant 10-bp sequence specifies the boundaries of a programmed DNA deletion in *Tetrahymena*. *Genes & Development* **7**: 2357–2365.
- Gorovsky MA, Yao MC, Keevert JB, Pleger GL. 1975. Isolation of micro- and macronuclei of *Tetrahymena pyriformis*. *Methods Cell Biol* **9**: 311–327.
- Greider CW, Blackburn EH. 1985. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**: 405–413.
- Grewal SI, Jia S. 2007. Heterochromatin revisited. *Nat Rev Genet* **8**: 35–46.
- Grossniklaus U, Paro R. 2014. Transcriptional Silencing by Polycomb-Group Proteins. *Cold Spring Harbor Perspectives in Biology* **6**: a019331–a019331.
- Ha M, Kim VN. 2014. Regulation of microRNA biogenesis. *Nature Publishing Group* **15**: 509–524.
- Hall IM, Shankaranarayana GD, Noma K-I, Ayoub N, Cohen A, Grewal SIS. 2002. Establishment and maintenance of a heterochromatin domain. *Science* **297**: 2232–2237.
- Hamilton AJ, Baulcombe DC. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**: 950–952.
- Hayashi A, Ishida M, Kawaguchi R, Urano T, Murakami Y, Nakayama J-I. 2012. Heterochromatin protein 1 homologue Swi6 acts in concert with Ers1 to regulate RNAi-directed heterochromatin assembly. *Proc Natl Acad Sci U S A* **109**: 6159–6164.
- Heard E, Martienssen RA. 2014. Transgenerational Epigenetic Inheritance: Myths and Mechanisms. *Cell* **157**: 95–109.
- Heitz E. 1928. Das Heterochromatin der Moose. *Jahrbuch der Wissenschaften Botanik* **69**: 762–818.
- Holoch D, Moazed D. 2015. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet* **16**: 71–84.
- Honda S, Lewis ZA, Huarte M, Cho LY, David LL, Shi Y, Selker EU. 2010. The DMM complex prevents spreading of DNA methylation from transposons to nearby genes in *Neurospora crassa*. *Genes & Development* **24**: 443–454.

REFERENCES

- Horrell SA, Chalker DL. 2014. LIA4 encodes a chromoshadow domain protein required for genomewide DNA rearrangements in *Tetrahymena thermophila*. *Eukaryot Cell* **13**: 1300–1311.
- Howard-Till RA, Yao M-C. 2006. Induction of gene silencing by hairpin RNA expression in *Tetrahymena thermophila* reveals a second small RNA pathway. *Molecular and Cellular Biology* **26**: 8731–8742.
- Huang XA, Yin H, Sweeney S, Raha D, Snyder M, Lin H. 2013. A major epigenetic programming mechanism guided by piRNAs. *Developmental Cell* **24**: 502–516.
- Iwasaki YW, Siomi MC, Siomi H. 2015. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem* **84**: 405–433.
- Jacobs SA, Khorasanizadeh S. 2002. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**: 2080–2083.
- Jia S, Noma K-I, Grewal SIS. 2004. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. *Science* **304**: 1971–1976.
- Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE. 2007. The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr Biol* **17**: 379–384.
- Jones-Rhoades MW, Bartel DP, Bartel B. 2006. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol* **57**: 19–53.
- Karrer KM. 2012. *Nuclear Dualism*. Elsevier Inc.
- Kataoka K, Mochizuki K. 2011. Programmed DNA elimination in *Tetrahymena*: a small RNA-mediated genome surveillance mechanism. *Adv Exp Med Biol* **722**: 156–173.
- Kaustov L, Ouyang H, Amaya M, Lemak A, Nady N, Duan S, Wasney GA, Li Z, Vedadi M, Schapira M, et al. 2011. Recognition and specificity determinants of the human cbx chromodomains. *Journal of Biological Chemistry* **286**: 521–529.
- Keller C, Adaixo R, Stunnenberg R, Woolcock KJ, Hiller S, Bühler M. 2012. HP1(Swi6) Mediates the Recognition and Destruction of Heterochromatic RNA Transcripts. *Molecular Cell*.
- Keller C, Kulasegaran-Shylini R, Shimada Y, Hotz H-R, hler MBU. 2013. Noncoding RNAs prevent spreading of a repressive histone mark. *Nat Struct Mol Biol* 1–8.
- Kimura A, Horikoshi M. 2004. Partition of distinct chromosomal regions: negotiable border and fixed border. *Genes Cells* **9**: 499–508.
- Kurth HM, Mochizuki K. 2009. 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in *Tetrahymena*. *RNA* **15**: 675–685.
- Kwon SH, Workman JL. 2011. The changing faces of HP1: From heterochromatin formation and gene silencing to euchromatic gene expression: HP1 acts as a positive regulator of transcription. *Bioessays* **33**: 280–289.

- Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**: 116–120.
- Lan F, Zaratiegui M, Villén J, Vaughn MW, Verdel A, Huarte M, Shi Y, Gygi SP, Moazed D, Martienssen RA, et al. 2007. *S. pombe* LSD1 homologs regulate heterochromatin propagation and euchromatic gene transcription. *Molecular Cell* **26**: 89–101.
- Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, Strahl BD, Patel DJ, Jacobsen SE. 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**: 385–389.
- 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 & Development* **27**: 390–399.
- Lee SR, Collins K. 2007. Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. *Nat Struct Mol Biol* **14**: 604–610.
- Lee SR, Collins K. 2006. Two classes of endogenous small RNAs in *Tetrahymena thermophila*. *Genes & Development* **20**: 28–33.
- Lee SR, Talsky KB, Collins K. 2009. A single RNA-dependent RNA polymerase assembles with mutually exclusive nucleotidyl transferase subunits to direct different pathways of small RNA biogenesis. *RNA* **15**: 1363–1374.
- Li G, Zhou L. 2013. Genome-wide identification of chromatin transitional regions reveals diverse mechanisms defining the boundary of facultative heterochromatin. *PLoS ONE* **8**: e67156.
- 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 U S A* **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 & Development* **21**: 1530–1545.
- Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**: 251–260.
- 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.
- Madireddi MT, Davis MC, Allis CD. 1994. Identification of a novel polypeptide involved in the formation of DNA-containing vesicles during macronuclear development in *Tetrahymena*. *Dev Biol* **165**: 418–431.

REFERENCES

- 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*. *Molecular and Cellular Biology* **25**: 9151–9164.
- Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. *Cell* **136**: 656–668.
- Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ, Voigt P, Martin SR, Taylor WR, De Marco V, et al. 2009. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**: 762–767.
- Martienssen R, Moazed D. 2015. RNAi and Heterochromatin Assembly. *Cold Spring Harbor Perspectives in Biology* **7**: a019323.
- Meneghini MD, Wu M, Madhani HD. 2003. Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* **112**: 725–736.
- Miao W, Xiong J, Bowen J, Wang W, Liu Y, Braguinets O, Grigull J, Pearlman RE, Orias E, Gorovsky MA. 2009. Microarray analyses of gene expression during the *Tetrahymena thermophila* life cycle. *PLoS ONE* **4**: e4429.
- 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 & Development* **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 & Development* **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. *Eukaryot Cell* **3**: 1233–1240.
- Mochizuki K, Gorovsky MA. 2004c. Small RNAs in genome rearrangement in *Tetrahymena*. *Current Opinion in Genetics & Development* **14**: 181–187.
- Mochizuki K, Kurth HM. 2013. Loading and pre-loading processes generate a distinct siRNA population in *Tetrahymena*. *Biochemical and Biophysical Research Communications* **436**: 497–502.
- 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.
- Motl JA, Chalker DL. 2011. Zygotic Expression of the Double-Stranded RNA Binding Motif Protein, Drb2p, is required for DNA Elimination in the Ciliate, *Tetrahymena thermophila*. *Eukaryot Cell*.

- Muller HJ. 1930. Types of visible variations induced by X-rays in *Drosophila*. *Journ of Gen* **22**: 299–334.
- Nakayama T, Nishioka K, Dong Y-X, Shimojima T, Hirose S. 2007. *Drosophila* GAGA factor directs histone H3.3 replacement that prevents the heterochromatin spreading. *Genes & Development* **21**: 552–561.
- Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, Reinberg D. 2015. Transcription. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**: 1017–1021.
- Nielsen PR, Nietlispach D, Mott HR, Callaghan J, Bannister A, Kouzarides T, Murzin AG, Murzina NV, Laue ED. 2002. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**: 103–107.
- Nikiforov MA, Gorovsky MA, Allis CD. 2000. A novel chromodomain protein, pdd3p, associates with internal eliminated sequences during macronuclear development in *Tetrahymena thermophila*. *Molecular and Cellular Biology* **20**: 4128–4134.
- Noma K-I, Cam HP, Maraia RJ, Grewal SIS. 2006. A role for TFIIIC transcription factor complex in genome organization. *Cell* **125**: 859–872.
- 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. *Molecular Cell* 1–30.
- 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.
- Okamura K, Lai EC. 2008. Endogenous small interfering RNAs in animals. *Nature Publishing Group* **9**: 673–678.
- Papazyan R, Voronina E, Chapman JR, Luperchio TR, Gilbert TM, Meier E, Mackintosh SG, Shabanowitz J, Tackett AJ, Reddy KL, et al. 2014. Methylation of histone H3K23 blocks DNA damage in pericentric heterochromatin during meiosis. *Elife* **3**: e02996.
- Pikaard CS, Mittelsten Scheid O. 2014. Epigenetic regulation in plants. *Cold Spring Harbor Perspectives in Biology* **6**: a019315.
- Reilly PT, Yu Y, Hamiche A, Wang L. 2014. Cracking the ANP32 whips: Important functions, unequal requirement, and hints at disease implications. *Bioessays* **36**: 1062–1071.
- Rexer CH, Chalker DL. 2007. Lia1p, a novel protein required during nuclear differentiation for genome-wide DNA rearrangements in *Tetrahymena thermophila*. *Eukaryot Cell* **6**: 1320–1329.
- Sabin LR, Delás MJ, Hannon GJ. 2013. Dogma Derailed: The Many Influences of RNA on the Genome. *Molecular Cell* **49**: 783–794.

REFERENCES

- Saksouk N, Simboeck E, Déjardin J. 2015. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & Chromatin* **8**: 3.
- 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 & Development* **26**: 1729–1742.
- Schotta G, Ebert A, Krauss V, Fischer A, Hoffmann J, Rea S, Jenuwein T, Dorn R, Reuter G. 2002. Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J* **21**: 1121–1131.
- Shang Y, Song X, Bowen J, Corstanje R, Gao Y, Gaertig J, Gorovsky MA. 2002. A robust inducible-repressible promoter greatly facilitates gene knockouts, conditional expression, and overexpression of homologous and heterologous genes in Tetrahymena thermophila. *Proc Natl Acad Sci U S A* **99**: 3734–3739.
- Shieh AWY, Chalker DL. 2013. LIA5 is required for nuclear reorganization and programmed DNA rearrangements occurring during tetrahymena macronuclear differentiation. *PLoS ONE* **8**: e75337.
- Sienski G, Dönertas D, Brennecke J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**: 964–980.
- Smothers JF, Mizzen CA, Tubbert MM, Cook RG, Allis CD. 1997. Pdd1p associates with germline-restricted chromatin and a second novel anlagen-enriched protein in developmentally programmed DNA elimination structures. *Development* **124**: 4537–4545.
- Stargell LA, Bowen J, Dadd CA, Dedon PC, Davis M, Cook RG, Allis CD, Gorovsky MA. 1993. Temporal and spatial association of histone H2A variant hv1 with transcriptionally competent chromatin during nuclear development in Tetrahymena thermophila. *Genes & Development* **7**: 2641–2651.
- Stargell LA, Gorovsky MA. 1994. TATA-binding protein and nuclear differentiation in Tetrahymena thermophila. *Molecular and Cellular Biology* **14**: 723–734.
- Sugai T, Hiwatashi K. 1974. Cytologic and autoradiographic studies of the micronucleus at meiotic prophase in Tetrahymena pyriformis. *J Protozool* **21**: 542–548.
- Suka N, Luo K, Grunstein M. 2002. Sir2p and Sas2p opposingly regulate acetylation of yeast histone H4 lysine16 and spreading of heterochromatin. *Nature Genetics* **32**: 378–383.
- Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, Patel DJ, van der Oost J. 2014. The evolutionary journey of Argonaute proteins. *Nat Struct Mol Biol* **21**: 743–753.
- Sweet MT, Allis CD. 2006. Long-term storage of unfrozen tetrahymena cultures in soybean medium. *CSH Protoc* **2006**.

- Talsky KB, Collins K. 2012. Strand-asymmetric endogenous Tetrahymena small RNA production requires a previously uncharacterized uridylyltransferase protein partner. *RNA* **18**: 1553–1562.
- Taus T, Köcher T, Pichler P, Paschke C, Schmidt A, Henrich C, Mechtler K. 2011. Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res* **10**: 5354–5362.
- Taverna SD, Coyne RS, Allis CD. 2002. Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. *Cell* **110**: 701–711.
- Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SIS, Moazed D. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**: 672–676.
- Vogt A, Mochizuki K. 2013. A domesticated PiggyBac transposase interacts with heterochromatin and catalyzes reproducible DNA elimination in Tetrahymena. *PLoS Genet* **9**: e1004032.
- Volpe T, Schramke V, Hamilton GL, White SA, Teng G, Martienssen RA, Allshire RC. 2003. RNA interference is required for normal centromere function in fission yeast. *Chromosome Res* **11**: 137–146.
- 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 J, Lawry ST, Cohen AL, Jia S. 2014. Chromosome boundary elements and regulation of heterochromatin spreading. *Cell Mol Life Sci* **71**: 4841–4852.
- Wang J, Tadeo X, Hou H, Tu PG, Thompson J, Yates JR, Jia S. 2013. Epe1 recruits BET family bromodomain protein Bdf2 to establish heterochromatin boundaries. *Genes & Development* **27**: 1886–1902.
- Weick EM, Miska EA. 2014. piRNAs: from biogenesis to function. *Development* **141**: 3458–3471.
- 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.
- Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nature Genetics* **41**: 630–634.
- Wilson RC, Doudna JA. 2013. Molecular Mechanisms of RNA Interference. *Annu Rev Biophys* **42**: 217–239.
- Woehrler 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.

REFERENCES

- 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, Yuan Y, Liang A, Wang W. 2015. Chromodomain protein Tcd1 is required for macronuclear genome rearrangement and repair in *Tetrahymena*. *Sci Rep* **5**: 10243.
- Yao M-C, Yao C-H, Halasz LM, Fuller P, Rexer CH, Wang SH, Jain R, Coyne RS, Chalker DL. 2007. Identification of novel chromatin-associated proteins involved in programmed genome rearrangements in *Tetrahymena*. *J Cell Sci* **120**: 1978–1989.
- Yap KL, Zhou M-M. 2011. Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. *Biochemistry* **50**: 1966–1980.
- Zhang C, Molascon AJ, Gao S, Liu Y, Andrews PC. 2012. Quantitative proteomics reveals that the specific methyltransferases Txr1p and Ezl2p differentially affect the mono-, di- and tri-methylation states of histone H3 lysine 27. *Mol Cell Proteomics*.
- 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.
- Tetrahymena Comparative Sequencing Project, Broad Institute of Harvard and MIT (<http://www.broadinstitute.org/>). *broadinstitute.org*.
<http://www.broadinstitute.org/annotation/genome/Tetrahymena/> (Accessed August 4, 2013).

Curriculum vitae

Jan Henrik Suhren

M.Sc. Molecular Medicine

Personal Details

Date of birth May 3rd, 1984
Place of birth Hamburg, Germany
Nationality German
Address Neustiftgasse 16/5/36
 1070 Vienna
 Austria
Telephone +43 6769 588369
E-mail jh.suhren@gmx.de

Education

- 07/2011 - Present* **PhD Thesis** in the laboratory of Dr. Kazufumi Mochizuki
Institute for Molekular Biotechnology (IMBA) and University of Vienna, Austria
- Title: "Heterochromatin components are required for limiting biogenesis of siRNAs to sequences targeted for DNA elimination in *Tetrahymena*"
- 04/2007 - 05/2011* **Studies of Molecular Medicine**
University of Freiburg, Germany
- Master grade: 1.3
 - Master thesis in the laboratory of Dr. Robert Schneider
Max Planck Institute of Immunobiology and Epigenetics, Freiburg
Title: "Characterization of the novel Histone Modifications Lysine 115 and 122 Acetylation of Histone H3"
- 10/2004 - 04/2007* **Medical School**
University of Freiburg, Germany
- Preclinical medical licensing examination; grade: 2.0
- 08/1995 - 07/2004* **German higher education entrance qualification**
Primary and Secondary school, Hamburg, Germany
- Grade: 1.2

-

Lebenslauf

Jan Henrik Suhren

Diplom Molekularmediziner

Persönliche Daten

Geburtsdatum 3. Mai 1984
Geburtsort Hamburg
Nationalität Deutsch
Adresse Neustiftgasse 16/5/36
1070 Wien
Österreich
Telefon +43 6769 588369
E-Mail jh.suhren@gmx.de

Ausbildung

- Seit 07/2011* **Promotion** im Labor von Dr. Kazufumi Mochizuki
Institut für Molekulare Biotechnologie (IMBA), Wien
– Titel: „Heterochromatinkomponenten sind dafür notwendig die siRNA Biogenese auf Sequenzen zu begrenzen, die Ziel von DNA Elimination sind“
- 04/2007 - 05/2011* **Studium der Molekularen Medizin**
Albert-Ludwigs-Universität Freiburg
– Diplomnote: 1,3 (Diplomarbeitenote: 1,3; Vordiplomnote: 1,5)
– Diplomarbeit im Labor von Dr. Robert Schneider
Max Planck Institut für Immunobiologie und Epigenetik,
Freiburg
– Titel: “Charakterisierung der neuen Histonmodifizierungen
Lysin 115 und 122 Acetylierung von Histon H3”
- 10/2004 - 04/2007* **Studium der Humanmedizin**
Albert-Ludwigs-Universität Freiburg
– Physikumsnote: 2,0
- 08/1995 - 07/2004* **Abitur**
Gymnasium Volksdorf, Hamburg
– Abiturnote: 1,2

-

Acknowledgments

The completion of this PhD thesis would not have been possible without the many great people, who supported me.

Firstly, I would like to thank Kaz for giving me this interesting and challenging project at the beginning of my PhD studies and for subsequent, constant support throughout the following year. In particular, I appreciate very much that Kaz always took time for discussions and advice and that he gave me enough space to develop my own ideas.

Secondly, I would like to thank all past and present members of the Mochizuki laboratory, whom I had the pleasure to interact with: Alex, Agnieszka, Henri, Jana, Katy, Ken, Quentin, Tomoko, Sophie and Ursi. Their enthusiastic contributions in discussions and their teaching of new techniques to me were instrumental for the success of this study. In particular, I would like to thank Ken for great support with ChIP-sequencing and Tomoko for introducing me to the technique of scRNA-sequencing.

Thirdly, I would like to acknowledge the members of my PhD committee, Julius Brennecke, Meinrad Busslinger and Alwin Köhler, for supporting me with directions and advice throughout this project.

At last, I would like to thank my friends and family for great moral support during the, sometimes trying, times of this thesis. A special thanks in this regard goes to Fangfang Xu, who always encouraged me to continue even in the most stressful situations.