Unraveling the Molecular Roles of Novel Proteins that Influence Excision of Genomic DNA in *Paramecium*

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Lilia Häußermann

aus Stuttgart

Tübingen

2023

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	31.01.2024
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Dr. Estienne Swart
2. Berichterstatter/-in:	Prof. Dr. Marja Timmermans
3. Berichterstatter/-in:	Prof. Dr. Martin Simon

Table of contents

Zusammenfassung

Viele Genome bestehen zu einem Großteil aus Sequenzen wie transponierbaren Elementen, die, wenn exprimiert, Gene unterbrechen können. Während die meisten Organismen die Genexpression in diesen Regionen inaktivieren, um die Vermehrung dieser Elemente zu verhindern, neigen Ciliaten dazu, Transposons und deren Überreste, sogenannte Internal Eliminated Sequences (IESs), aus dem somatischen Genom zu entfernen. Der drastische Schritt, große Teile des Genoms zu eliminieren, erfordert eine zuverlässige Erkennung der IESs und ist ein hochgradig orchestrierter Prozess. Den IESs in *Paramecium tetraurelia* (*Paramecium*) fehlt ein konserviertes Motiv, das ausreichende Spezifität für die sequenzbasierte Erkennung bietet. Es wurde vorgeschlagen, dass diese Herausforderung durch einen small RNA-basierten nuklearen Informationsaustausch gemeistert wird, der zu entfernende Sequenzen identifiziert, indem das vollständig reorganisierte somatische Genom mit dem nicht reorganisierten Keimbahngenom verglichen wird. Bisher ist jedoch unklar, wie alle IESs präzise eliminiert werden können, da nur einige IESs auf die small RNAs für ihre effiziente Entfernung angewiesen sind. Offensichtlich spielen zusätzliche, bisher nicht identifizierte Faktoren bei der Rekrutierung der Entfernungsmaschinerie zu den IESs eine Rolle. Daher ist das Hauptziel dieser Studie die Charakterisierung von unbekannten Proteinen, die zur Entfernung der IESs beitragen.

Zuerst wird ISWI1, ein Mitglied der hochkonservierten Imitation Switch (ISWI)-Familie ATP-abhängiger Chromatinremodeler, charakterisiert. ISWI1 ist das erste Protein, von dem berichtet wird, dass es die Präzision der IES-Eliminierung beeinflusst, vermutlich durch die Positionierung von Nukleosomen. In anderen Eukaryoten hängt ISWIs Aktivität immer von seinen Komplexpartnern ab. Nach der Identifizierung zweier Komplexpartner von ISWI1, ISWI1 Complex Protein (ICOP) 1 und ICOP2, zeigen wir, dass ISWI1 und die ICOPs in die reifenden somatischen Zellkerne, wo die Eliminierung von IESs stattfindet, lokalisieren. Die ICOPs interagieren sowohl *in vitro* als auch *in vivo* mit ISWI1. Die beobachteten phänotypischen Ähnlichkeiten in Gen-Knockdowns deuten auf eine gemeinsame Funktion der drei Protein hin: der Verbleib von IESs im reorganisierten somatischen Genom, die ungenaue Entfernung der IESs an alternativen Grenzen und die Veränderungen der Nukleosomendichte auf IESs.

Zusätzlich haben wir unter den durch Massenspektrometrie identifizierten ISWI1-assoziierten Proteinen nach neuen Kandidaten gesucht und zwei paraloge PHD finger charakterisiert: Development-specific PHD finger (DevPF) 1 und DevPF2. Trotz ihrer Ähnlichkeit auf Nukleotid- und Aminosäureebene, tragen diese Paraloge unterschiedlich zur Entfernung von IESs bei. Der früh exprimierte DevPF1 lokalisiert in einige, jedoch nicht alle, der meiotischen Zellkerne. Dieses Lokalisationsmuster wurde bisher noch nicht für Proteine in *Paramecium* berichtet. In späteren Stadien lokalisiert DevPF1 in die reifenden somatischen Zellkerne, genau wie der spät exprimierte DevPF2. Der Gen-Knockdown von DevPF1 führt zum vollständigen Verlust entwicklungsspezifischer small RNAs, während der Gen-Knockdown von DevPF2 hauptsächlich die Population spät hergestellter small RNAs betrifft. Wir zeigen auch, dass der Gen-Knockdown von DevPF1 keine Präferenz hinsichtlich der Länge der IES aufweist, während beim Gen-Knockdown von DevPF2 vorzugsweise lange IESs zurückgehalten werden.

Zusammenfassend präsentiere ich Erkenntnisse über fünf neue Schlüsselakteure, die zur Aufrechterhaltung der Genomintegrität während der Zellkernreifung beitragen, und vertiefe damit unser Verständnis der Entfernung von IESs in *Paramecium*. Diese und zeitgenössische Studien legen nahe, dass es noch weitere Proteine zu entdecken gilt, die an der DNA-Eliminierung in *Paramecium* beteiligt sind – es bleibt nur abzuwarten, wie viele. Das Aufdröseln des molekularen Netzwerks, das DNA-Eliminierung reguliert, hat auch für jene Implikationen, die an der langfristigen Einführung von unnatürlichen Genom-Editing-Komponenten (wie CRISPR) in Eukaryoten arbeiten. Weiteres Wissen über grundlegende Prozesse, wie die Genomreorganisation, ist für unser Verständnis der Zellbiologie im Allgemeinen unerlässlich und kann möglicherweise auf andere Arten übertragen werden.

Abstract

A significant portion of many genomes comprises sequences such as transposable elements, which carry the potential risk of disrupting genes when they are expressed. While most organisms silence these regions to prevent their proliferation, ciliates tend to eliminate transposable elements and their remnants, Internal Eliminated Sequences (IESs), from the somatic genome. This drastic step of removing substantial portions of the genome necessitates reliable IES targeting, making it a highly orchestrated process. In *Paramecium tetraurelia* (henceforth *Paramecium*), IESs lack a conserved motif that provides sufficient specificity for sequence-based targeting alone. To address this challenge, small RNA-guided nuclear crosstalk has been proposed to identify sequences for excision by comparing the fully reorganized somatic genome with the non-reorganized germline genome. However, it is unclear how precise elimination is achieved for all IESs, as only a subset of IESs relies on the small RNA pathway for efficient excision. It is evident that additional, as yet unidentified factors play a role in facilitating the recruitment of the excision machinery to IESs. Therefore, the primary objective of this study is to characterize novel proteins that contribute to IES excision.

First, ISWI1, a member of the highly conserved Imitation Switch (ISWI) family of ATP-dependent chromatin remodelers, is characterized. ISWI1 is the first protein to be reported that influences the precision of cutting the exact boundaries of IESs, presumably by nucleosome positioning. In other eukaryotes, ISWI always relies on complex partners for its full activity. After we identified two complex partners of ISWI1, ISWI1 Complex Protein (ICOP) 1 and ICOP2, we show that ISWI1 and the ICOPs localize to the maturing somatic nuclei, where IES excision occurs. The ICOPs interact with ISWI1 both *in vitro* and *in vivo*. In knockdown experiments, all three proteins show phenotypic similarities including IES retention in the reorganized somatic genome, imprecise excision at alternative IES boundaries and alterations in the nucleosome densities on IESs, suggesting shared functionality.

Additionally, we screened for novel candidates among ISWI1-associated proteins identified by mass spectrometry and characterize two paralogous PHD finger proteins: development-specific PHD finger (DevPF) 1 and DevPF2. Despite their high similarity

at the nucleotide and amino acid level, we show that these paralogs contribute differently to IES excision. The early-expressed DevPF1 localizes into some, though not all, of the meiotic germline nuclei, a localization pattern not yet reported for *Paramecium* proteins. In later stages, DevPF1 localizes to the maturing somatic nuclei, as does the late-expressed DevPF2. The knockdown of DevPF1 completely abolishes development-specific small RNA production, while DevPF2 knockdown mainly affects the late-produced small RNA population. We also demonstrate that DevPF1 knockdown exhibits no preference regarding IES length while in DevPF2 knockdown, preferably long IESs are retained.

Taken together, I present work that characterizes five new players essential for maintaining genome integrity during nuclear maturation, adding to our picture of IES excision in *Paramecium*. These and contemporary studies suggest that additional proteins involved in *Paramecium* DNA elimination remain to be discovered – it just remains to be seen how many. Unravelling the molecular systems regulating DNA elimination has implications for anyone considering long-term introductions of unnatural genome editing components (such as CRISPR) into eukaryotes. Further knowledge of fundamental processes like genome reorganization are vital for our understanding of cell biology in general and can potentially be transferred to other species.

List of publications

Accepted manuscript

Singh, A., Maurer-Alcalá, X. X., Solberg, T., **Häußermann, L.**, Gisler, S., Ignarski, M., Swart, E. C., & Nowacki, M. (2022). Chromatin remodeling is required for sRNA-guided DNA elimination in *Paramecium*. The EMBO Journal, 41(22), e111839. https://doi.org/10.15252/embj.2022111839

Published manuscripts

Singh, A.*, **Häußermann, L.***, Emmerich, C., Nischwitz, E., Seah, B. K., Butter, F. K. B., Nowacki, M., & Swart, E. C. (2023). ISWI1 complex proteins facilitate developmental genome editing in *Paramecium*. BioRxiv. https://doi.org/10.1101/2023.08.09.552620

* indicates equal contribution

Häußermann, L., Singh, A., & Swart, E. C. (2024). Two paralogous PHD finger proteins participate in Paramecium tetraurelia's natural genome editing. BioRxiv. https://doi.org/10.1101/2024.01.23.576875

1 Introduction

1.1 Significance and Aims

All organisms face the potential threat of mobile elements that have invaded their genome at some point in the evolutionary history. To guarantee proper gene expression, such transposons must be kept in check and prevented from jumping into regulatory or coding regions. But how is this achieved? A widespread approach across the eukaryotic tree of life is to silence transposable elements on the transcriptional or post-transcriptional level. However, ciliates have taken a different path. These unicellular organisms separate their germline and somatic genome in two distinct types of nuclei: the micronucleus (MIC) and the macronucleus (MAC), respectively. Ciliates generally remove most transposons and their remnants, Internal Eliminated Sequences (IESs), during sexual development from the MAC genome. Evidently, a tightly regulated mechanism has evolved that reliably eliminates the target sequences that account for a substantial portion of the genome.

Paramecium tetraurelia (*Paramecium*) is a particularly interesting model organism to investigate genome reorganization as its IESs are often located in coding regions. Therefore, precise recruitment of the excision machinery to these unique sequences and flawless subsequent DNA repair are essential. There is no conserved sequence motif known that specifically marks IESs for excision (just a short semi-degenerate motif of a ~ 5 bp) and the exact mechanism of precise IES targeting remains poorly understood. With *Paramecium* being a well-established model organism, many techniques are available to study genome reorganization. Several key processes leading to successful IES excision, and some of the participating proteins, have thus already been identified. A major breakthrough was the discovery of a small RNA-guided pathway (scnRNA pathway) that was proposed to identify sequences for elimination based on a comparison of the reorganized somatic genome with the germline genome. However, the picture is far from complete since the small RNAs are crucial for the excision of only a subset of IESs.

To fill in the gaps in our knowledge, this study aims to identify novel proteins assisting precise IES excision in *Paramecium* and I present the findings of five newly characterized development-specific proteins. *Paramecium's* IESs are typically very

short – in fact, shorter than the DNA wrapped around a nucleosome. Together with the exceptionally short linker DNA on *Paramecium's* somatic genome, it raises the question how the nucleosomal landscape influences the accessibility of IESs for excision. We therefore began our investigations by characterizing an ATP-dependent chromatin remodeler involved in IES excision. Next, we investigated two complex partners of this chromatin remodeler and lastly, we characterized two PHD finger proteins, one of those identified in co-immunoprecipitations (IP) of the chromatin remodeler.

1.2 Overview of the thesis structure

Here, I provide an overview of the thesis structure and how each chapter contributes to the overall aim described above.

Chapter 2, Streamlining the somatic genome: How ciliates remove interrupting sequences, reviews the literature on the progress in the field of IES excision in ciliates. First, the general contribution of ciliates in research is briefly discussed. Next, the concept of separating the germline and somatic line in two nuclei within the same cell is introduced, followed by general features of the germline and somatic genome in *Paramecium*. Subsequently, major processes occurring during genome reorganization are presented with a detailed focus on our current knowledge of IESs and their excision in *Paramecium*. Next, a brief overview of IES excision in other ciliates is given. Lastly, methods used to investigate genome reorganization in *Paramecium* are briefly described.

Chapter 3, Chromatin remodeling is required for sRNA-guided DNA elimination in *Paramecium* [published in EMBO Journal, 2022], demonstrates the importance of the chromatin remodeler ISWI1 for precise IES excision in *Paramecium*. ISWI1 localized to the new developing MACs during IES excision and interacted with PTIWI01, a core component of the scnRNA pathway. ISWI1 knockdown not only causes IES retention but also elevated excision at alternative boundaries, thereby generating excised IES fragments of the "forbidden" length. We propose that misplaced nucleosomes cause excision at the wrong IES boundaries.

Chapter 4, ISWI1 complex proteins facilitate developmental genome editing in *Paramecium* [published on BioRxiv, 2023], builds on the findings in Chapter 3, and investigates ICOP1 and ICOP2, two putative complex partners of the chromatin remodeler ISWI1. ICOP1 and ICOP2 localized to the new MACs and interacted with ISWI1 during IES excision. Their knockdown revealed high functional similarity to ISWI1: reduction of these proteins caused alternative IES excision and their effect on nucleosome density on IESs resembled ISWI1's but is distinct from proteins which are not part of the remodeling complex.

Chapter 5, Two paralogous PHD finger proteins participate in *Paramecium tetraurelia*'s natural genome editing [published on BioRxiv, 2024], characterizes two paralogous PHD finger proteins, DevPF1 and DevPF2. They were identified through RNAi screening of ISWI1-associated proteins (based on the findings in Chapter 3). Both proteins exhibited different expression profiles and showed distinct effects on IES excision. The late-expressed DevPF2 localized to the new MAC and predominantly affected the excision of long IESs. Most intriguing is the selective localization of the early-expressed DevPF1 to certain gametic and post-zygotic nuclei in early stages of sexual development, a behavior never reported before in *Paramecium*. At later stages, it was observed in the developing new MACs. Upon *DevPF1* knockdown, development-specific small RNA production was abolished, and the excision of IESs was impaired, regardless of their length. Moreover, we found evidence of DevPF1 and DevPF2 positively regulating the transcription of at least two genes exclusively expressed from the new MACs.

Chapter 6, Discussion, discusses the contribution of the findings presented in this study to our current model of IES excision in *Paramecium* in the context of findings obtained from research in other organisms.

2 Streamlining the somatic genome: How ciliates remove interrupting sequences

In this chapter, I review the literature on genome reorganization in *Paramecium*.

2.1 Why ciliates?

2.1.1 Ciliates in the world

Ciliates are unicellular eukaryotes, named after one of their external characteristics, namely the abundant cilia covering their surfaces (Figure 1). This large and highly diverse group is widespread throughout the world. The habitats range from freshwater over saltwater to sediments and soil (Wang *et al.* 2022; Pierce and Turner 1992; Foissner 1997), including extreme environments with respect to temperature, pH or salt concentrations (Hu 2014). Ciliates account for a large proportion of the world's biomass and the largest blooms of some species, like the bloom of *Mesodinium rubrum* in the Atlantic Ocean, can even be observed from space (Dierssen *et al.* 2015). Despite their abundance worldwide, ciliates are rarely pathogenic for humans, animals or plants. However, some ciliate fish parasites cause mass fatalities and have relevance for the fish industry (Treasurer 2002; Dickerson and Clark 1996; Li *et al.* 2023).

Ciliates display a broad morphological spectrum and range in their sizes from tens of micrometers to several millimeters. With their different morphologies, they have adopted a variety of life strategies (Finlay *et al.* 1996). Ciliates feed on bacteria, algae, other protists or some metazoans (Weisse 2017). In some species, like *Blepharisma*, cannibalism occurs when food resources are limited (Giese 1938). Furthermore, many species profit from various forms of symbiosis (Dziallas *et al.* 2012). Many ciliates, including *Paramecium*, are freely swimming organisms that utilize their cilia for motility and food uptake with roundish or spindle-like shapes (Van Houten 2019). Other ciliates attach to solid surfaces, including fish skin (Li *et al.* 2023), and filter the water with their cilia for food. One example for this behavior is the model organism *Stentor*, that can switch between stationary and free-swimming phases (Tartar 1961) and adapts a characteristic trumpet-like shape once attached to a surface. Ciliates also occur in complex biofilm communities (Xu *et al.* 2014).

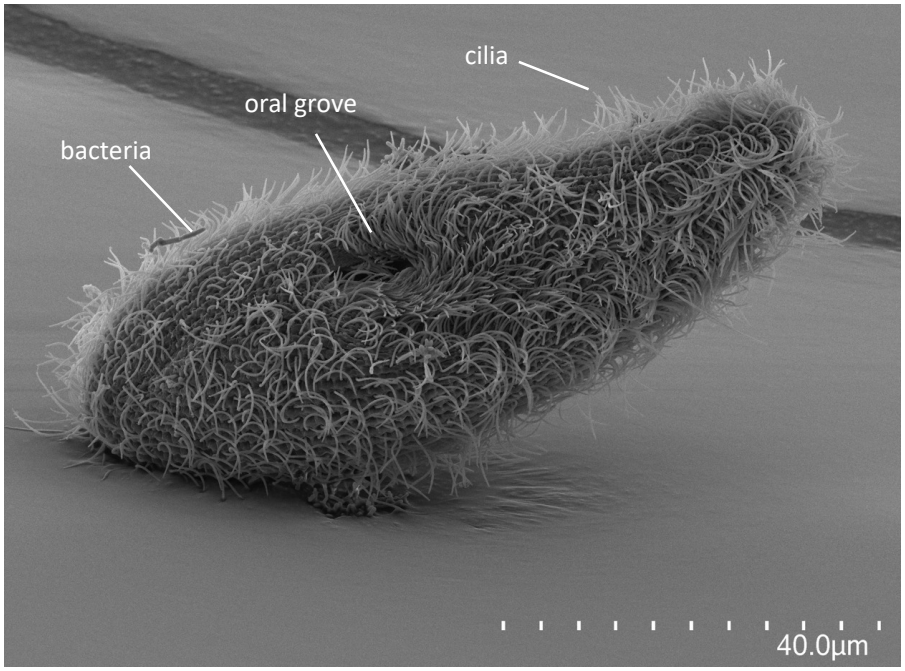


Figure 1: Scanning electron microscope (SEM) image of Paramecium tetraurelia. Kindly provided by Jürgen Berger from MPI for Biology Tübingen.

2.1.2 Ciliates in research

The high abundance of ciliates in water samples combined with their considerable size led to their early discovery in the late 17th century. Antoni van Leeuwenhoek was testing self-crafted microscope precursors and for the first time discovered microscopic life, including ciliates (Dobell 1932). Since that initial discovery, ciliates have remained a fascinating object of observation and revelation.

Ciliates harbor many features that can be exploited for science (Plattner 2022). While the list of examples is extensive, I will provide only a selection to illustrate the variety of research fields. Their most obvious trait, the many cilia, make them attractive model organisms for cilia and cell motility research (Bayless *et al.* 2019; Soh and Pearson 2022). As eukaryotes, they possess a complex chromatin structure. Ciliate research contributes to understand histone modifications and their role in gene regulation (Brownell *et al.* 1996; Strahl *et al.* 1999) as well as the mechanisms underlying complex genome reorganization for the construction of functional genomes (Drews *et al.* 2022; Rzeszutek *et al.* 2020). Due to their distant branching in the eukaryotic tree of life, ciliates are particularly valuable for evolutionary studies and the investigation of ancestral states. Since fundamental molecules are conserved, we can draw conclusions from ciliates for multicellular organisms, including insights into neuronal

systems (Plattner and Verkhratsky 2018), learning (Dussutour 2021) and regeneration (Marshall 2021).

The fundamental contributions of research conducted in ciliates to our general scientific understanding were highlighted by two Nobel prizes won for the discovery of telomerases (Greider and Blackburn 1985) and the identification of RNA's enzymatic activity in ribozymes (Cech 1990), both conducted in the ciliate *Tetrahymena*.

2.2 Ciliate genomes

Apart from their eponymous cilia, ciliates share another striking characteristic: they have two distinct types of nuclei that differ in morphology, function and genomic content (Prescott 1994). The high divergence of ciliates is reflected in the considerable variability they display in terms of nuclei number and forms. However, the functional assignment is shared by all: one type of nucleus propagates genetic material to the next progeny while the other type is optimized for gene expression. Here, I focus primarily on *Paramecium*, as the model organism most relevant for this research.

2.2.1 The separation of the germline and somatic genome

Despite being single-cell organisms, ciliates separate their germline and soma into two types of nuclei. The germline nucleus is typically substantially smaller than its somatic counterpart, resulting in the former being named the micronucleus (MIC), and the latter the macronucleus (MAC). MICs are considered the germline because they are transcriptionally silent throughout most of the life cycle and produce gametic nuclei during meiosis. In contrast, the somatic MAC is responsible for gene expression and is degraded during sexual development, not transferring any genetic material to the next generation (Prescott 1994). This functional separation between the two nuclei strikingly resembles the functional separation of germline cells and somatic cells in multicellular organisms.

Though there are a number of hypotheses about the evolution of ciliate nuclear dimorphism, it remains uncertain how this has evolved (Boscaro and Keeling 2023). As a first step, the ancestral cell acquired at least two nuclei by decoupling nuclear division from cell division. Large cells like ciliates have a high demand for protein

production. Gene expression can be boosted by an increase in the genome copy number. This genome amplification can occur in two ways in eukaryotes: either accompanied by an increase in the number of nuclei, as seen in the slime mold *Myxogastria* (Adl *et al.* 2005) and karyolict ciliates (Raikov 1985), or within the same nucleus, leading to high polyploidy and an increase in nucleus size (Lee *et al.* 2009). There are different types of polyploidy (Stebbins 1947). For the ciliate macronucleus the term "ampliploid" was proposed (Schwartz 1978), since only parts of the chromosomes are amplified.

Paramecium has two MICs and one MAC (Figure 2). The morphological changes of the nuclei throughout the life cycle of *Paramecium* are represented in Figure 3. In the asexual life cycle (also referred to as vegetative growth), all three nuclei duplicate and are distributed into the daughter cells. The MICs divide mitotically (closed orthomitosis, since the nuclear envelop remains intact), while the MAC DNA is duplicated and divided in an amitotic fission. The MAC elongates through intranuclear microtubule deployment; however, no spindle apparatus is formed and the nuclear envelop remains intact (Tucker *et al.* 1980). The daughter cells resulting from the asexual division are considered genetically identical and clones of each other. After a limited number of asexual divisions, the cells undergo senescence and eventually die (Sonneborn 1954), an event that can be prevented by entering the sexual cycle, thereby resetting the cell's age.

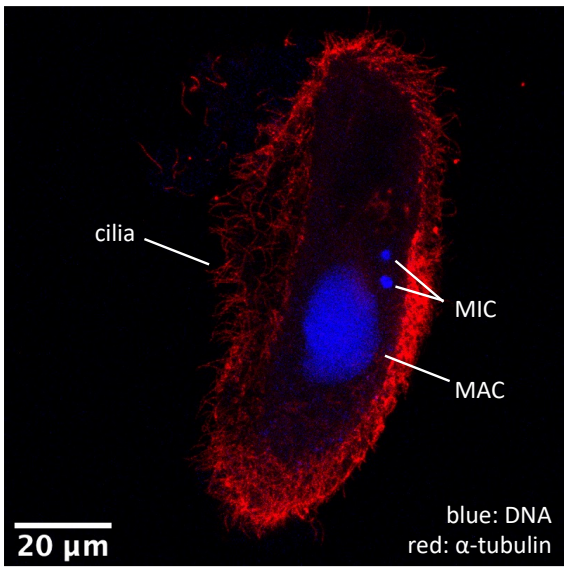


Figure 2: Nuclei of vegetative Paramecium cell. Confocal fluorescence microscopy image of an immunostained cell. Cilia were stained with a primary antibody against α-tubulin and secondary antibody conjugated to alexa488 (represented in red). DNA was stained with DAPI (represented in blue).

Paramecium is capable of both conjugation (Figure 3) and autogamy (Figure 4). Conjugation involves the pairing of two suitable mating partners (Sonneborn 1938) that form physical contacts at their oral groove. The cilia at the oral groove disintegrate

(deciliation) in order to establish cytoplasmatic bridges, enabling the exchange of gametic nuclei between the cells (Inaba *et al.* 1966; Fujishima 1988). If no mating partner is available, *Paramecium* can undergo a self-fertilizing process called autogamy. During autogamy, no deciliation is observed.

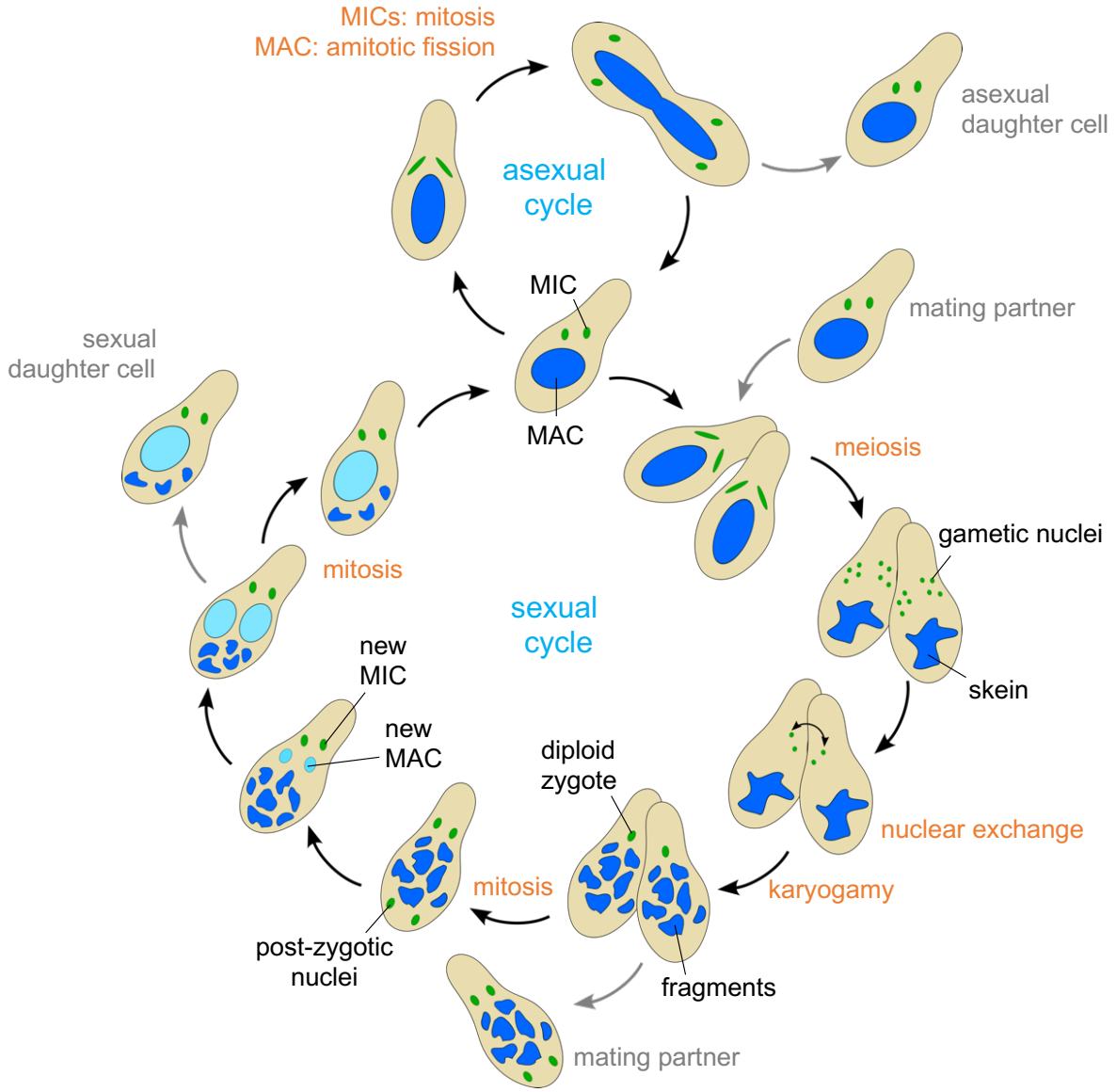


Figure 3: Life cycle of Paramecium. Asexual cycle: vegetative growth. Sexual cycle: conjugation. The life cycles are indicated in blue font, the nuclei types in black font and nuclear processes in orange font.

Both conjugation and autogamy can be induced by starvation. Sexual development starts with meiosis of the two MICs, resulting in eight haploid nuclei. Only one of the nuclei survives. It is selected by its cytoplasmic position in the paroral region, the area surrounding the oral apparatus, where it is protected from degradation that eliminates the other seven haploid nuclei (Yanagi and Hiwatashi 1985; Yanagi 1987). The selected nucleus undergoes an extra mitotic division, giving rise to the two identical

gametic nuclei which then, in the case of autogamy, fuse together to produce the entirely homozygous zygotic nucleus. In the case of conjugation, one of the identical gametic nuclei is transported through the established cytoplasmatic bridges into the cytoplasm of the mating partner, in exchange for one of the partner's gametic nuclei (Inaba *et al.* 1966). Therefore, the zygote after conjugation is a heterozygous product. The zygote divides mitotically into four identical copies, two of which remain the new MICs of the cell and the other two develop into new MACs. Again, the nuclear fate depends on their subcellular localization (Grandchamp and Beisson 1981).

During their development, the new MACs increase in size. Once the new MACs have sufficiently matured, the MICs divide mitotically to a total of four, and the cell divides. Each daughter cell will be provided with a complete set of nuclei: two MICs and one newly developed MAC. Throughout the whole process, the old MAC is lost. During meiosis, the old MAC starts to disintegrate by changing its shape into structures called skein that eventually break apart into smaller fragments. These fragments are slowly degraded. However, some fragments still remain after the final division and are further diluted and lost in subsequent vegetative divisions.

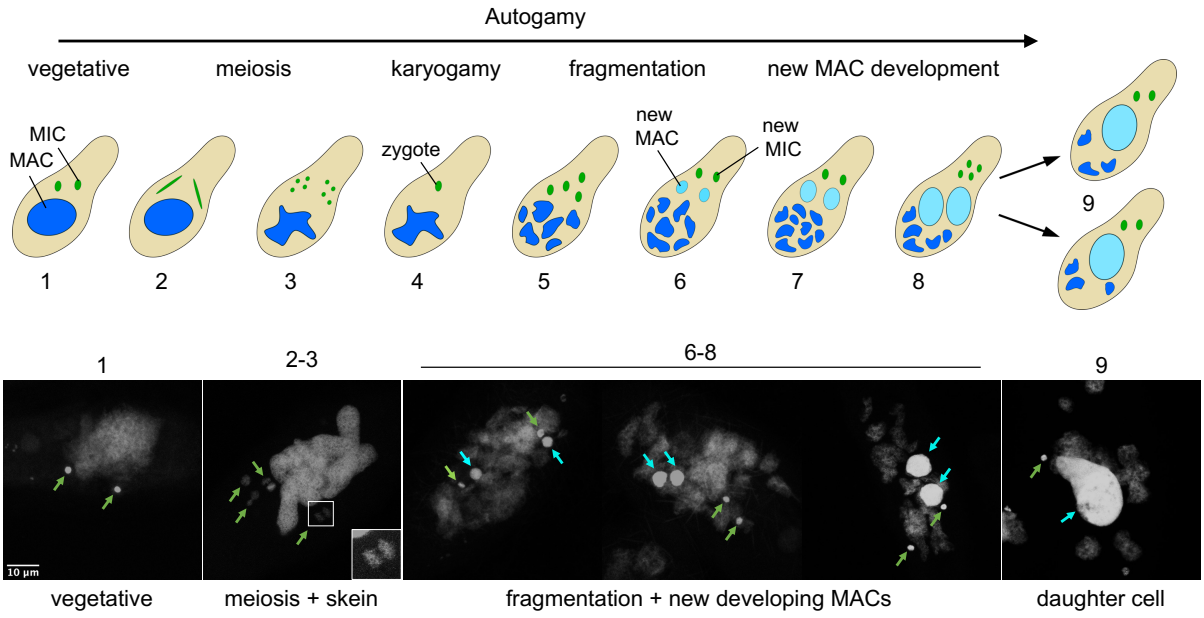


Figure 4: Autogamy in Paramecium. Green arrow: MIC. Cyan arrow: new MAC. Scale bar 10 um.

2.2.1 The genome structure in *Paramecium*

The *Paramecium* MAC genome has approximately 40,000 genes derived from multiple successive rounds of whole-genome duplications (Aury *et al.* 2006). The GC content is relatively low at 28% of the MAC genome. Due to their distinct functions, the MICs and MACs differ significantly in their genome structure. The MAC genome is amplified to a copy number of approximately 800n (Drews *et al.* 2022), while the MIC is diploid. This high number of chromosomes in the MAC results in its massive size and presumably precludes segregation of chromosomes by a spindle apparatus during nuclear division. Active centromeres are restricted to the MICs (Lhuillier-Akakpo *et al.* 2016).

MAC chromosomes are shorter than MIC chromosomes due to deletions and fragmentation. About 25% of the MIC genome (98 megabases (Mb)) (Arnaiz *et al.* 2012; Guérin *et al.* 2017) is removed to develop the MAC genome (72 Mb) (Aury *et al.* 2006). Many of these MIC-limited sequences comprise repetitive elements, satellites, and transposons. Their imprecise excision results in slight variations between multiple chromosomes. Excision sites are healed by either DNA repair or *de novo* telomere attachment, resulting in chromosome fragmentation (Baroin *et al.* 1987; Forney and Blackburn 1988; Le Mouël *et al.* 2003). On average, each MIC chromosome is fragmented into two MAC chromosomes.

Paramecium's MAC genome exhibits some unusual features. Its extremely high coding density (78%) is not only caused by the removal of MIC-limited sequences, but also by short intergenic regions and introns (average length of 352 bp and 25 bp, respectively) (Aury *et al.* 2006). Also, the linker DNA between histones was shown to be exceptionally short, only a few base pairs (Gnan *et al.* 2022), and the regulation of gene expression by the nucleosomal landscape is different from what is known in other organisms. Typically, active transcription is associated with nucleosome-free promotor regions and well-positioned nucleosomes carrying activating histone marks near the transcription start site. Silenced regions, on the other hand, are nucleosome dense and organized in condensed heterochromatin. In contrast, recent research has shown that transcribed genes in *Paramecium* show nucleosome occupancy with a broad distribution of histone marks instead of sharp peaks, while non-transcribed regions are void of nucleosomes (Drews *et al.* 2022). In accordance with this, no evidence for the

presence of the linker histone H1 was found in *Paramecium* (Drews *et al.* 2022) that assists gene silencing by heterochromatin formation in other organisms (Nalabothula *et al.* 2014). Apparently, *Paramecium* has evolved a different mechanism for gene regulation, rather based on activation than suppression.

2.3.1 IESs in *Paramecium*

IESs, a class of MIC-limited sequences, are removed precisely from a MIC genome copy that becomes the MAC genome during development. The ~45,000 IESs comprise mainly of single-copy sequences of short length (93% are shorter than 150 bp) that make up about 3.55 Mb in total (Arnaiz *et al.* 2012). In contrast to other MIC-limited sequences, IESs are found in both non-coding and coding regions with no clear preference for either. However, due to the high coding density, most IESs (77%) are located in exons (Arnaiz *et al.* 2012).

IESs are believed to be remnants of transposable elements. Several lines of evidence support this theory: Firstly, all IESs are excised by the domesticated transposase PiggyMAC (PGM), which belongs to the PiggyBac transposase family (Baudry *et al.* 2009). Secondly, all IESs are flanked by two 5'-TA-3' dinucleotides, that are part of a less conserved inverted consensus motif (Klobutcher and Herrick 1995; Mayer and Forney 1999; Arnaiz *et al.* 2012). These flanks might derive from Tc1/mariner transposons that preferably insert at and duplicate the TA dinucleotides (Dubois *et al.* 2012). Thirdly, Tc1/mariner transposons are found in *Paramecium* and imprecisely excised during new MAC development (Arnaiz *et al.* 2012; Le Mouël *et al.* 2003). Lastly, some evolutionarily young multi-copy IESs are present in multiple loci and one class showed similarity to Tc1/mariner transposons, indicating their invasion and duplication within the genome (Sellis *et al.* 2021; Arnaiz *et al.* 2012). Miniature Inverted-repeat Transposable Elements (MITEs) are non-autonomous transposons found in *Paramecium* and *Blepharisma stoltei* (*Blepharisma*) that have been proposed as transposon/IES intermediates (Sellis *et al*. 2021; Seah *et al*. 2023). However, Tc1/mariner transposons typically leave a footprint after excision, whereas after IES elimination, only one TA dinucleotide remains (Steele *et al.* 1994). Such “scarless” excision is a hallmark of PiggyBac transposase activity (Elick *et al.* 1996), which is a

useful property that has been exploited for genetic manipulations (Ding *et al.* 2005; Kim and Pyykko 2011).

It is thought that with transposons or MITEs that are excised precisely, i.e. without a footprint from the MAC genome, selective pressure was only on their efficient excision but no longer on their exact location or sequence content (Seah and Swart 2023). Therefore, IESs were tolerated in coding regions, gained extremely high sequence variability, and shortened considerably in length. They exhibit a declining periodic length distribution with a 10 to 11 bp distance between the peaks (Figure 5; (Arnaiz *et al.* 2012)), corresponding to the step width of the DNA double-helix. It is thought that the twist of the DNA restricts the sizes of excisable fragments necessary during coordinated action between PGM subunits. Additionally, the second, so-called "forbidden" peak, consists of approximately 34 – 44 bp fragment lengths and is highly underrepresented. This is thought to be caused by the necessity of PGM dimerization (Arnaiz *et al.* 2012) and the inability to alter DNA geometry to accommodate excision.

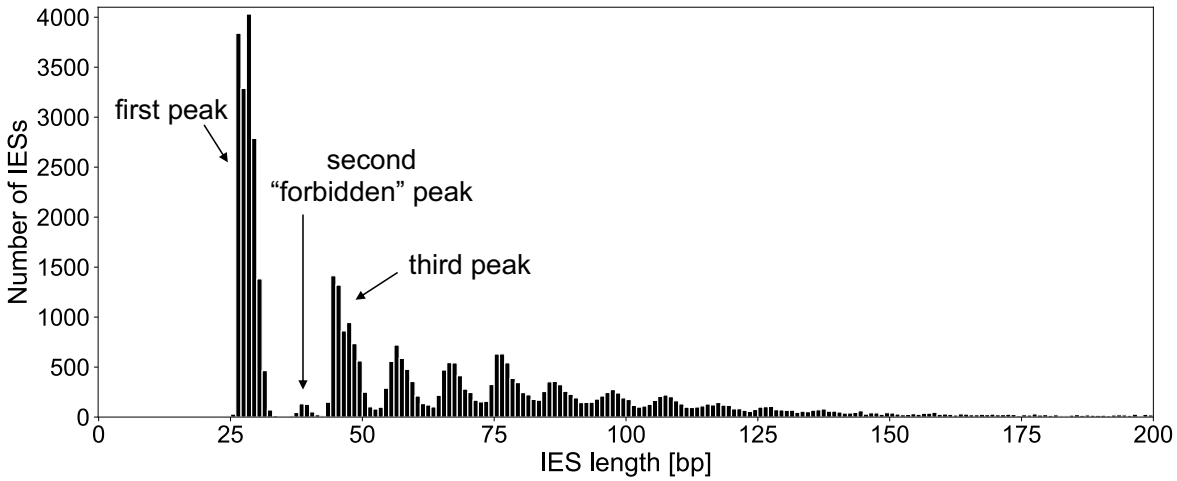


Figure 5: IES length distribution. The first three peaks are labeled. Adapted from (Arnaiz et al. 2012).

2.3 Genome reorganization

Although structurally and functionally different, the MIC and the MAC share the same origin: the zygotic nucleus, a product of MIC undergoing meiosis. The MIC genome remains unaltered, while the generation of the MAC genome requires extensive genome reorganization. In this section, I describe how the MIC genome is reorganized to develop a streamlined MAC genome. It is beyond the scope of this thesis to delve

into the high diversity of genome reorganization forms that have originated in other ciliates. The main focus lies on *Paramecium*; however, comparisons to other ciliates will be made in the last section.

2.3.1 General mechanisms of genome reorganization in *Paramecium*

There are three key events involved in genome reorganization during new MAC development: (i) endoreplication, (ii) DNA elimination (imprecise and precise excision), and (iii) chromosome fragmentation (Figure 6). Endoreplication, from 2n to about 800n, partially precedes DNA elimination. Therefore, IESs need to be removed from multiple loci. IES excision occurs between the 4th and 5th round of endoreplication (32 to 64n) (Zangarelli *et al.* 2022). As a general trend, the precise excision of first evolutionary old IESs and then evolutionary young IESs is followed by the imprecise elimination of intragenic MIC-limited sequences which starts at 64n (Bétermier *et al.* 2000; Zangarelli *et al.* 2022). Chromosome fragmentation is a consequence of DNA elimination. While IES excision is always followed by non-homologous end joining (NHEJ) (Kapusta *et al.* 2011), imprecise excision sites are either repaired or capped by *de novo* telomere attachment. Notably, at the same imprecise excision site both scenarios can occur (Forney and Blackburn 1988).

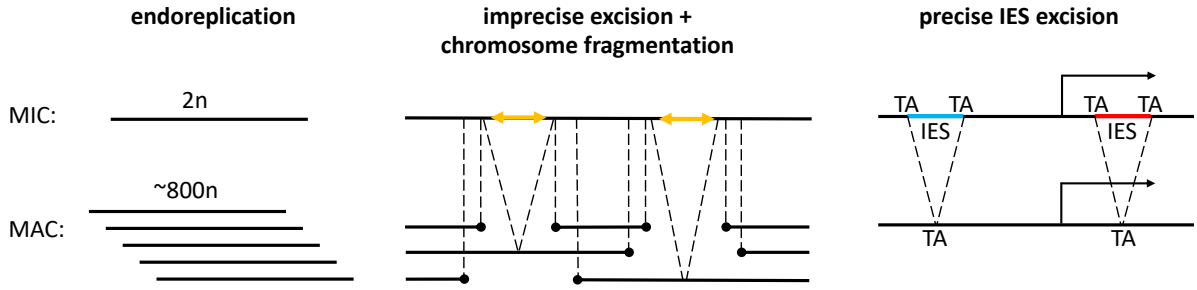


Figure 6: The three key events during genome reorganization in Paramecium. Adapted from (Coyne et al. 2012). Imprecise excision only occurs in intergenic, precise IES excision in inter- and intragenic regions.

2.3.2 IES excision in *Paramecium*

All IESs are excised at their 5'-TA-3' boundaries. Two PGM proteins work in concert and each introduces a double-strand break, centered on the TA dinucleotide, with 4-base 5′ overhangs (Gratias and Bétermier 2003; Gratias *et al.* 2008). In the "naked" DNA model, the PGM interaction was proposed to explain the size distribution of IESs

(Figure 5; (Arnaiz *et al.* 2012)). Varying between 26 and 28 bp in length, most IESs allow the interaction of two adjacent PGM subunits. For longer IES, DNA looping is required to bring the PGMs together. Therefore, IESs of ~34 – 44 bp length (second "forbidden" peak) are highly disfavored because they are too short to form DNA loops.

PGM cleavage depends on the presence of catalytically inactive PiggyMAC-like proteins (PGMLs) (Bischerour *et al.* 2018) and KU70/KU80 proteins (Marmignon *et al.* 2014; Abello *et al.* 2020). The latter are part of the NHEJ pathway and provide a direct link to DNA repair. Ligase IV and XRCC4, core components of NHEJ, fuse the double-strand breaks both in the chromosome and the excised fragments (Kapusta *et al.* 2011). PGM cleavage and the removal of the 5' terminal nucleotide precedes ligase IV recruitment since double-strand breaks accumulate in its absence. However, the nucleotide addition at the 3′ ends prior to ligation depends on ligase IV.

It still remains unclear how the PGM excision machinery can be targeted so precisely to IES boundaries. The weak consensus at the IES flanks fails to provide sufficient specificity for sequence-based recruitment. Therefore, other factors assist precise targeting. Of note, there is no general mechanism to target all IESs. Early excised, evolutionary old IESs appear to depend only on PGM for excision while late excised IESs require additional factors (Sellis *et al.* 2021).

Two classes of small non-coding RNAs are components an epigenetic mechanism targeting MIC-limited sequences: scan RNAs (scnRNAs) and IES-matching RNAs (iesRNAs) (Figure 7). scnRNAs mediate genome scanning and enable a comparison of the MIC and the MAC genome to identify MIC-limited sequences. In a first step, a set of scnRNAs is generated that comprises the sequence content of the MIC genome. During meiosis, the MIC genome is bidirectionally transcribed, presumably with the assistance of the transcription elongation factor SPT5m (Gruchota *et al.* 2017), into long non-coding double-stranded RNAs. Dicer-like proteins DCL2 and DCL3 cleave them into small RNAs duplexes with 2 nt 3′ overhangs at both ends. DCL3 produces the characteristic 5'-UNG end on one strand while DCL2 defines the length of 25 bp (Lepère *et al.* 2009; Sandoval *et al.* 2014). The 5'-U strand is preferably loaded on Piwi proteins PTIWI01/09 and thereby protected from degradation (Bouhouche *et al.* 2011; Furrer *et al.* 2017). The proteins PTIWI01/09, thought to interact with NOWA1/2

(Nowacki *et al.* 2005; Bouhouche *et al.* 2011), transport the single-stranded scnRNAs to the old MAC where scnRNA selection occurs. This process is referred to as scanning (hence scnRNAs) and is thought to be based on base pairing of the scnRNAs with transcripts from the old MAC genome. Only MAC-matching scnRNAs find complementary sequences on the transcripts. The zinc-finger protein GTSF1 was recently proposed to promote the degradation of PTIWI01/09 complexes harboring MAC-matching scnRNAs (Charmant *et al.* 2023; Wang *et al.* 2023). scnRNAs matching MIC-limited sequences (including IESs and transposons) are transported into the developing new MAC where they scan the new MAC genome, indirectly via RNA transcripts, for their complementary sequences. It is thought that TFIIS4, a RNA polymerase II elongation factor localizing to the new MACs (Maliszewska-Olejniczak *et al.* 2015), assists the transcription necessary for scnRNA pairing during new MAC genome development.

The scnRNA pathway is linked to the development-specific deposition of H3K27me3 and H3K9me3 in the new MAC (Lhuillier-Akakpo *et al.* 2014; Ignarski *et al.* 2014). Both histone marks are deposited by EZL1, the catalytically active methyltransferase of the polycomb repressive complex 2 (PRC2) (Lhuillier-Akakpo *et al.* 2014; Miró-Pina *et al.* 2022). PRC2 has been shown to directly interact with PTIWI01/09 (Miró-Pina *et al.* 2022; Wang *et al.* 2022). Knockdown of PRC2 complex components leads to IES retention. However, the co-occurrence of both marks has only been shown on transposable elements so far and represses their expression (Miró-Pina *et al.* 2022).

iesRNAs are produced from excised IESs and participate in a positive feedback loop in the new developing MAC (Sandoval *et al.* 2014). Since genome amplification partially precedes IES excision, the IESs have to be excised from multiple loci. Excised IES fragments form concatemers (Bétermier *et al.* 2000; Allen *et al.* 2017) that are transcribed into dsRNA. These serve as substrates for the Dicer-like protein DCL5 that produces small RNA duplexes. Their 2 nt 3’ overhangs have a 5′-UAG signature and a 3’-CNAU signature. The exclusively IES-matching small RNAs (hence iesRNAs) vary in length (~26 – 30 nt) (Sandoval *et al.* 2014). They are bound by PTIWI10/11 proteins that remove the passenger strand (Furrer *et al.* 2017). In contrast to many other development-specific genes, PTIWI10 is expressed from the new MAC instead

of the old MAC and its expression timing is thought to be controlled by IES excision. An IES is located in its regulatory regions. Hence, PTIWI10 expression is initiated once IES excision starts (Furrer *et al.* 2017).

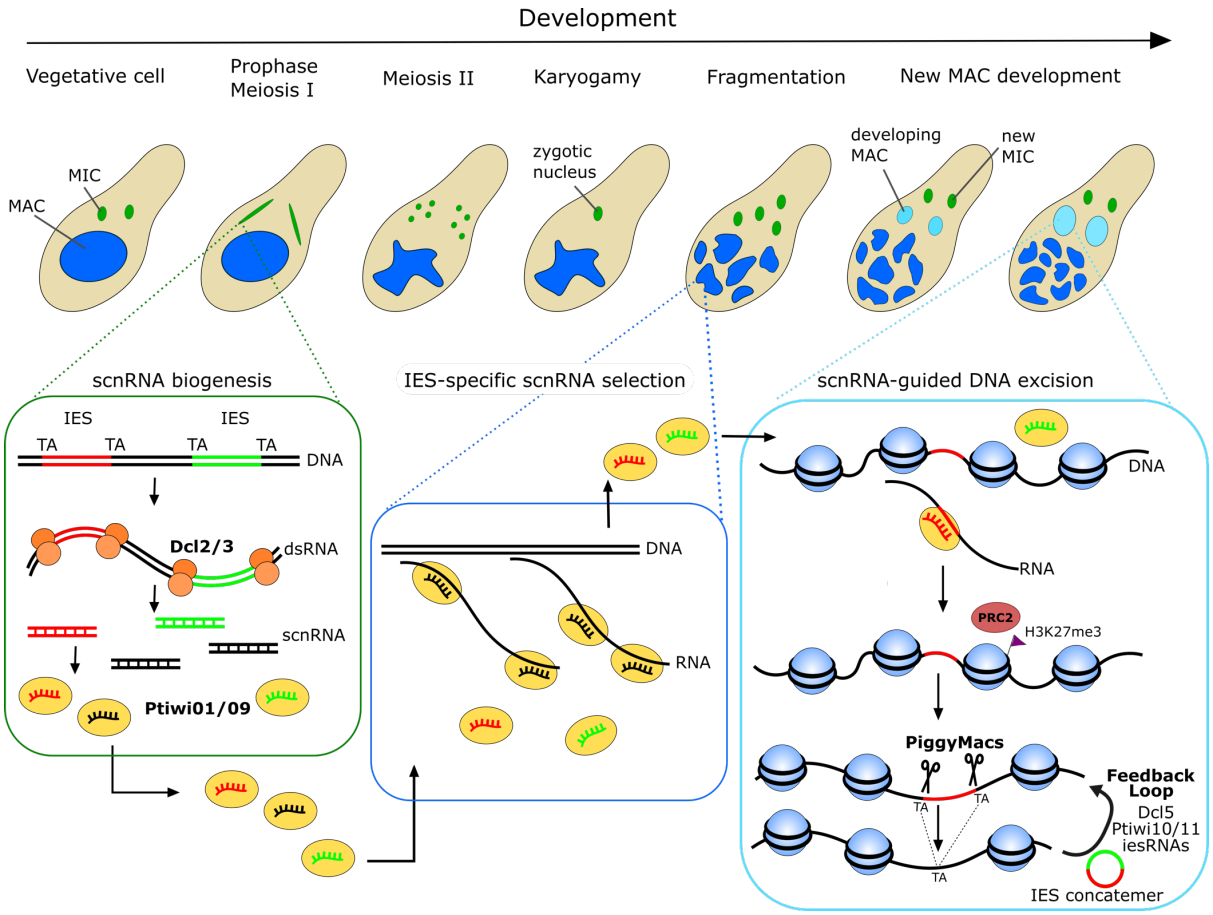


Figure 7: The scnRNA pathway in Paramecium. 25 nt scnRNAs are produced from dsRNA in the MICs, transported into the old MAC for scnRNA selection and scnRNAs matching MIC-limited sequences are transported into the new MAC, targeting their complementary sequences for excision. In the new MAC, scnRNA-dependent deposition of histone modifications occurs. iesRNAs are produced from excised IESs and participate in a positive feedback loop to efficiently eliminate all IES copies.

Despite the insights we have into the scnRNA and iesRNA pathways, only a small subset of IESs depends on these small RNAs for their excision (Sandoval *et al.* 2014; Swart *et al.* 2014). Hence, how the majority of IESs is targeted for elimination remains unclear.

2.3.3 Genome reorganization in other ciliates

The mechanisms of genome reorganization are quite diverse in ciliates. In *Tetrahymena thermophila* (*Tetrahymena*), IES removal is mostly imprecise; hence,

they are predominantly eliminated from intergenic regions (Hamilton *et al.* 2016). DNA to be eliminated is marked by H3K9 methylation (Taverna *et al.* 2002) and forms heterochromatin before excision. The scnRNA pathway targeting DNA for elimination was first reported in *Tetrahymena* (Mochizuki *et al.* 2002; Malone *et al.* 2005) which facilitated the discovery of a similar pathway in *Paramecium*. *Tetrahymena* also produces secondary small RNAs from IESs (Mutazono *et al.* 2019). In contrast to *Paramecium's* iesRNAs, their biogenesis is linked to heterochromatin formation and happens before IES excision. A domesticated PiggyBac transposase was found in *Tetrahymena* (Cheng *et al.* 2010) and in the distantly related ciliate *Blepharisma* (Singh *et al.* 2021). IESs in *Blepharisma* bear striking resemblances to those in *Paramecium*, including being abundant within genes, showing the characteristic 10-11 bp periodicity, and precise excision (Seah *et al*. 2023). Therefore, PiggyBac transposases were proposed as the ancestral excisase for ciliate IESs (Singh *et al.* 2021).

Oxytricha trifallax (*Oxytricha*) and *Stylonychia lemnae* (*Stylonychia*) are more distantly related to *Paramecium* than *Tetrahymena* and show greater differences in their programmed genome reorganization (Yerlici and Landweber 2014). Both species eliminate almost all (more than 90%) of their germline genome during MAC maturation (Prescott 1994). Their MAC genome is highly fragmented with tiny chromosomes that mostly carry only one gene (Swart *et al.* 2013; Prescott 1994). Another striking difference is their massively scrambled MIC genome (Chen *et al.* 2014). Coding regions are not only disrupted by IESs, as it is the case in *Paramecium*, but the genes are also broken into multiple, possibly inverted pieces that can be separated into multiple loci that are several kilobases apart (Chen *et al.* 2014; Ardell *et al.* 2003). The unscrambling of the genome is thought to be guided by long non-coding RNAs (Nowacki *et al.* 2008). In contrast to *Paramecium* and *Tetrahymena*, the small Piwi-associated piRNAs in *Oxytricha* select their complementary sequences for retention (Fang *et al.* 2012). Of note, recently at least three scrambled genes have been validated in *Tetrahymena* (Sheng *et al.* 2020).

2.4 Methods in *Paramecium*

Paramecium has been a model organism for several decades. Thus, many tools are available to study its genes. Here, I summarize two key methods used in this study.

2.4.1 RNAi by feeding

RNA interference (RNAi) through feeding is a well-established silencing technique in *Paramecium* (Galvani and Sperling 2002) adapted from *Caenorhabditis elegans* (Fire *et al.* 1998) that enables the functional investigation of a gene using a reverse genetics approach. Phenotypic effects resulting from reduced expression levels are analyzed. RNAi exploits the endogenous siRNA silencing pathway by introducing double-stranded (ds) RNAs that match the target gene's sequence (Carradec *et al.* 2015). Cells have developed countermeasures to degrade dsRNAs, as they are commonly of viral origin. The 23 nt long siRNAs are produced by RNA-dependent RNA polymerases and Dicer protein Drc1 before loading onto Piwi proteins PTIWI12 to PTIWI15 (Bouhouche *et al.* 2011; Lepère *et al.* 2009; Marker *et al.* 2010). These siRNAs then target complementary RNAs for degradation, thus reducing the mRNA levels of the target gene.

RNAi by feeding involves providing dsRNAs in the diet of the target organism (Figure 8). In the case of *Paramecium*, *E. coli* cells can be utilized for this purpose. To conduct an RNAi experiment, a 300 to 800 bp fragment of the target gene is cloned between two inverted inducible promoters. The expression plasmid is then introduced into *E. coli* cells, and once bidirectional expression is induced, paramecia are exposed to this silencing medium. The uptake of the *E. coli* cells introduces the dsRNAs into the paramecia, triggering siRNA-dependent silencing.

To assess a gene's involvement in IES excision, a limited number of dsRNA-producing bacteria are supplied to induce autogamy through starvation while the siRNA pathway is active. Typically, two readouts are employed: survival tests and IES retention PCRs (Polymerase Chain Reactions). Survival tests evaluate the ability of knockdown cells to produce viable progeny. After sexual development is complete, cells are isolated into fresh growth medium and their growth is monitored over several divisions. If a gene crucial to genome reorganization is downregulated during autogamy, the cells fail to generate a functional new MAC genome and subsequently die. However, survival tests alone are insufficient to confirm IES retention. A non-functional MAC genome can arise not only due to retained IESs but also from errors in meiosis, failed excision of transposable elements and repeats, defects in cell division, and numerous other factors.

One specific method for testing IES retention is through IES retention PCRs. In this process, genomic DNA from a post-autogamous cell population is extracted and subjected to PCRs using primers flanking IES loci. Properly excised IESs yield a smaller fragment (IES-), while IES-containing loci produce a larger fragment (IES+). Nonetheless, there will always be an IES- band present, even when IES retention is observed. This is due to post-autogamous cells still containing fragments of the old MAC. The old MAC is correctly reorganized with almost all IESs excised, except for the background levels. Additionally, due to the high ploidy of the new MAC, several copies of a given IES can be retained while other copies are excised properly. In knockdowns, residual levels of the target gene are still expressed and contribute to IES excision. As IES- fragments are shorter, they are more efficiently amplified in PCR reactions.

When using IES retention PCRs, only a limited number of IES loci can be assessed. To examine the approximately 45,000 IESs, the DNA from enriched new MACs is typically analyzed based on whole-genome sequencing data. The ParTIES pipeline (Denby Wilkes *et al.* 2016) was specifically developed for analyzing IES excision from short read whole-genome sequencing data. It aligns the reads to the MAC and the MAC+IES genome, calculating an IES retention score (IRS) for each IES. The IRS is determined by counting how many reads span an IES locus and contain the IES sequence, relative to the total number of reads spanning the locus.

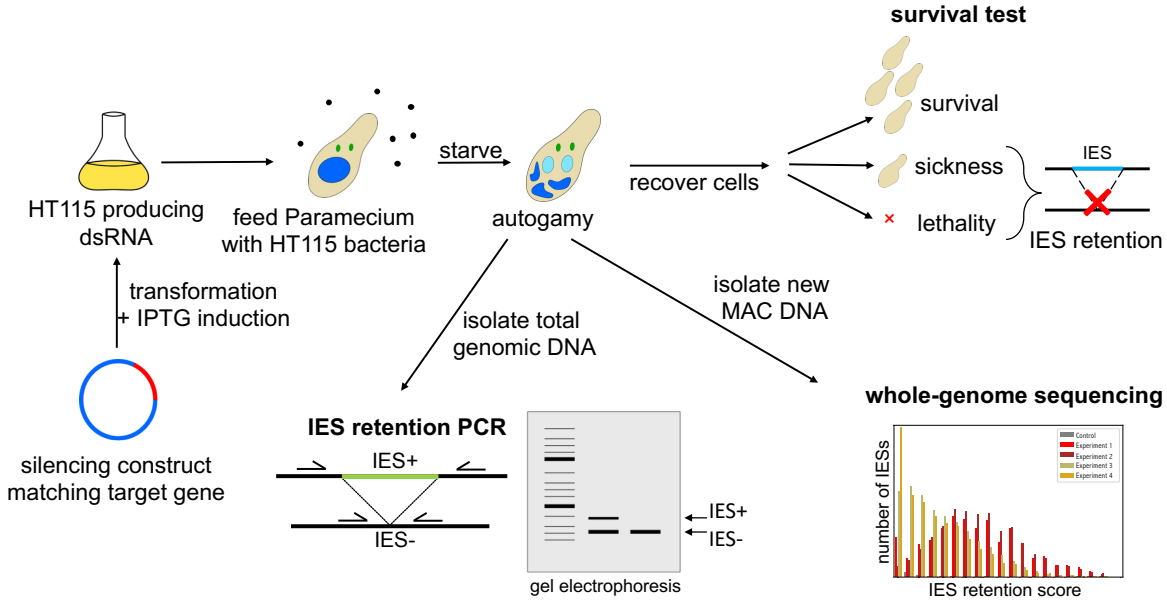


Figure 8: RNAi by feeding and subsequent analysis for IES retention.

2.4.2 Microinjection

Microinjection is utilized to transform *Paramecium* cells (Beisson *et al.* 2010). Linearized DNA is injected directly into the MAC of vegetative paramecia. The introduced DNA molecules are treated as endogenous chromosomes, as telomeres are attached. These molecules are replicated and propagated throughout vegetative growth (Gilley *et al.* 1988). Promotors and flanking regulatory regions in *Paramecium* are not known. Typically, about 200 bp upstream and downstream the gene are cloned, as the regulatory regions may extend into adjacent genes. Alternatively, regulatory regions of expressible genes showing a similar expression profile can be used. The transgene remains stable during vegetative growth but will be lost once the cell enters the sexual cycle, as only the MAC is transformed.

3 Chromatin remodeling is required for sRNA-guided DNA elimination in *Paramecium*

3.1 Citation

Singh, A., Maurer-Alcalá, X. X., Solberg, T., **Häußermann, L.**, Gisler, S., Ignarski, M., Swart, E. C., & Nowacki, M. (2022). Chromatin remodeling is required for sRNA-guided DNA elimination in *Paramecium*. The EMBO Journal, 41(22), e111839. https://doi.org/10.15252/embj.2022111839

3.2 Abstract

Small RNAs mediate the silencing of transposable elements and other genomic loci, increasing nucleosome density and preventing undesirable gene expression. The unicellular ciliate *Paramecium* is a model to study dynamic genome organization in eukaryotic cells, given its unique feature of nuclear dimorphism. Here, the formation of the somatic macronucleus during sexual reproduction requires eliminating thousands of transposon remnants (IESs) and transposable elements scattered throughout the germline micronuclear genome. The elimination process is guided by Piwi-associated small RNAs and leads to precise cleavage at IES boundaries. Here we show that IES recognition and precise excision are facilitated by recruiting ISWI1, a *Paramecium* homolog of the chromatin remodeler ISWI. ISWI1 knockdown substantially inhibits DNA elimination, quantitatively similar to development-specific sRNA gene knockdowns but with much greater aberrant IES excision at alternative boundaries. We also identify key development-specific sRNA biogenesis and transport proteins, PTIWI01 and PTIWI09, as ISWI1 cofactors in our co-immunoprecipitation studies. Nucleosome profiling indicates that increased nucleosome density correlates with the requirement for ISWI1 and other proteins necessary for IES excision. We propose that chromatin remodeling together with small RNAs is essential for efficient and precise DNA elimination in *Paramecium*.

3.3 Author contributions

Author	Author position	Scientific ideas [%]	Data generation [%]	Analysis & interpretation [%]	Paper writing [%]
Lilia Häußermann	4	0	5	5	0
Aditi Singh	1	60	40	40	50
Xyrus Maurer-Alcalá	2	0	10	8	20
Therese Solberg	3	0	7	2	0
Silvan Gisler	5	0	10	3	0
Michael Ignarski	6	0	2	2	0
Estienne Swart	7	20	26	30	30
Mariusz Nowacki	8	20	0	10	0
Title of paper:	Chromatin remodeling is required for sRNA-guided DNA elimination in *Paramecium*				
Status in publication process:	peer-reviewed and published				

4 ISWI1 complex proteins facilitate developmental genome editing in *Paramecium*

4.1 Citation

Singh, A.*, **Häußermann, L.***, Emmerich, C., Nischwitz, E., Seah, B. K., Butter, F. K. B., Nowacki, M., & Swart, E. C. (2023). ISWI1 complex proteins facilitate developmental genome editing in *Paramecium*. BioRxiv. https://doi.org/10.1101/2023.08.09.552620

* indicates equal contribution

4.2 Abstract

Chromatin remodeling is required for essential cellular processes, including DNA replication, DNA repair, and transcription regulation. The ciliate germline and soma are partitioned into two distinct nuclei within the same cell. During a massive editing process that forms a somatic genome, ciliates eliminate thousands of DNA sequences from a germline genome copy in the form of internal eliminated sequences (IESs). Recently we showed that the chromatin remodeler ISWI1 is required for somatic genome development in the ciliate *Paramecium tetraurelia*. Here we describe two paralogous proteins, ICOP1 and ICOP2, essential for DNA elimination. ICOP1 and ICOP2 are highly divergent from known proteins; the only domain detected showed distant homology to the WSD motif. We show that both ICOP1 and ICOP2 interact with the chromatin remodeler ISWI1. Upon *ICOP* knockdown, changes in alternative IES excision boundaries and nucleosome densities are similar to those observed for *ISWI1* knockdown. We thus propose that a complex comprising ISWI1 and either or both ICOP1 and ICOP2 are needed for chromatin remodeling and accurate DNA elimination in *Paramecium*.

4.3 Author contributions

* indicates equal contribution

Author	Author position	Scientific ideas [%]	Data generation [%]	Analysis & interpretation [%]	Paper writing [%]
Lilia Häußermann*	2	20	40	40	45
Aditi Singh*	1	50	40	40	35
Christiane Emmerich	3	0	3	0	0
Emily Nischwitz	4	0	6	5	5
Brandon Seah	5	0	1	3	0
Falk Butter	6	0	0	2	0
Mariusz Nowacki	7	5	0	0	0
Estienne Swart	8	25	10	10	15

Title of paper:	ISWI1 complex proteins facilitate developmental genome editing in *Paramecium*
Status in publication process:	published on bioRxiv and submitted to Genome Research; revised manuscript

5 Two paralogous PHD finger proteins participate in *Paramecium tetraurelia*'s natural genome editing

5.1 Citation

Häußermann, L., Singh, A., & Swart, E. C. (2024). Two paralogous PHD finger proteins participate in Paramecium tetraurelia's natural genome editing. BioRxiv. https://doi.org/10.1101/2024.01.23.576875

5.2 Abstract

The unicellular eukaryote *Paramecium tetraurelia* contains functionally distinct nuclei: germline micronuclei (MICs) and a somatic macronucleus (MAC). During sexual reproduction, the MIC genome is reorganized into a new MAC genome and the old MAC is lost. Almost 45,000 unique Internal Eliminated Sequences (IESs) distributed throughout the genome require precise excision to guarantee a functional new MAC genome. Here, we characterize a pair of paralogous PHD finger proteins involved in DNA elimination. DevPF1, the early-expressed paralog, is present in only some of the gametic and post-zygotic nuclei during meiosis. Both DevPF1 and DevPF2 localize in the new developing MACs, where IESs excision occurs. In DevPF2 knockdown (KD) long IESs are preferentially retained and late-expressed small RNAs decrease; no length preference for retained IESs was observed in DevPF1-KD and development-specific small RNAs were abolished. The expression of at least two genes from the new MAC seems to be influenced by DevPF1- and DevPF2-KD. Thus, both PHD fingers are crucial for new MAC genome development, with distinct functions, potentially via regulation of non-coding and coding transcription in the MICs and new MACs.

5.3 Author contributions

Author	Author position	Scientific ideas [%]	Data generation [%]	Analysis & interpretation [%]	Paper writing [%]
Lilia Häußermann	1	20	85	80	95
Aditi Singh	2	60	15	10	0
Estienne Swart	3	20	0	10	5

Title of paper:	Two paralogous PHD finger proteins participate in *Paramecium tetraurelia*'s natural genome editing
Status in publication process:	published on bioRxiv and submitted to Journal of Cell Science

6 Discussion

Ciliates have developed a sophisticated mechanism that removes thousands of sequences forming a functional somatic genome. It has long been a question as to how IESs in *Paramecium* can be precisely targeted within and outside of coding regions, but the complete answer has not yet been found. In this study, I summarize our recent findings on five newly characterized proteins contributing to different aspects of IESs excision.

We propose that the chromatin remodeler ISWI1 and its putative complex partners, ICOP1 and ICOP2, ensure the correct targeting of IES boundaries through nucleosome positioning in the new MAC. As the first proteins reported to mediate excision precision, ISWI1 and the ICOPs pave the way for deciphering the details of the underlying mechanisms.

The other two proteins, DevPF1 and DevPF2, influence development-specific small RNA populations and presumably regulate transcription during sexual development. While DevPF2 is active only during new MAC development, DevPF1 also contributes to very early processes during meiosis, such as scnRNA biogenesis. Our findings suggest that ISWI1, ICOP1/2, and DevPF2 primarily contribute to the excision of long, evolutionarily younger IESs, while DevPF1 is a more general factor essential for IES excision regardless of IES length.

The five proteins investigated in this study function at different stages during sexual development, all aiding in the excision of IESs. Four of the five proteins (ISWI1, ICOP1, ICOP2, and DevPF2) are expressed late with the characteristic localization pattern observed for this group of development-specific proteins: they localize to the new MACs (Figure 9). The localization pattern of the fifth protein, the early-expressed DevPF1, is unique and combines aspects of typical localization patterns observed for other proteins involved in IES excision (Figure 9): cytosolic localization in early stages (e.g. PTIWI01/09 (Furrer *et al.* 2017; Bouhouche *et al.* 2011)), localization to the MICs before the first meiotic division (e.g. DCL2 (Lepère *et al.* 2009)) and localization to the new developing MACs (e.g. PGM (Baudry *et al.* 2009)). What stands out is this protein's selective localization to certain gametic and post-zygotic nuclei, as all other

MIC-localizing proteins are observed in all nuclei simultaneously (e.g. SPT5m (Gruchota *et al.* 2017)).

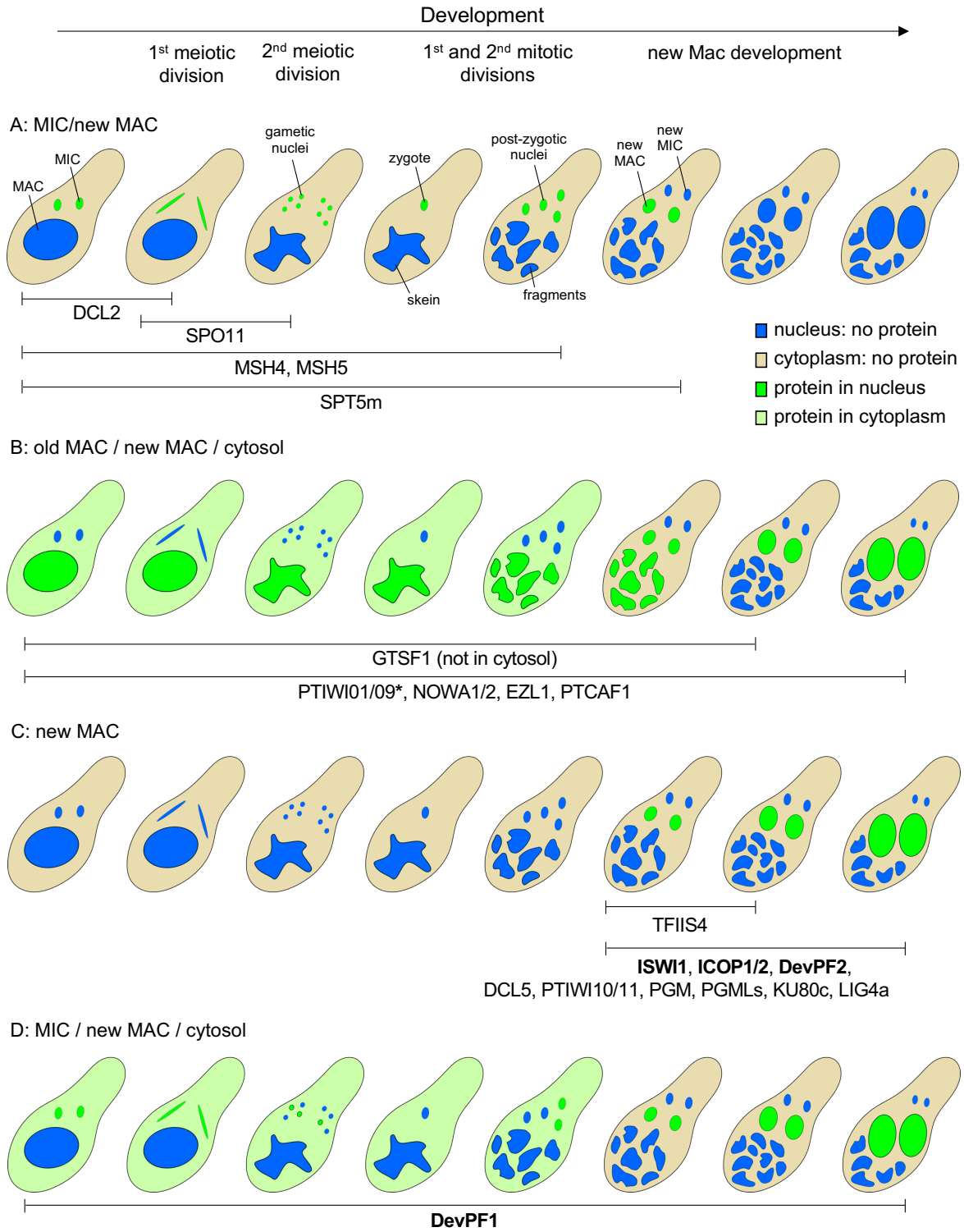


Figure 9: Schematic representation of different localization patterns (A to D) observed for development-specific GFP- or HA-tagged fusion proteins. Temporal restriction of specific proteins is indicated by capped lines. Proteins identified in this study are highlighted in bold font. Asterisk (*): PTIWI09 also localizes to MICs during S-Phase, as demonstrated in Chapter 5.

6.1 Discussion and future directions for ISWI1 and the ICOP proteins

So far, little is known about how the nucleosome landscape influences IES excision accessibility. The short length of most IESs (Arnaiz *et al.* 2012) and the extremely short linker DNA between nucleosomes in the MAC (Gnan *et al.* 2022) suggest that chromatin remodeling is needed to render IESs accessible. There are four families of ATP-dependent chromatin remodelers, differing in their complex composition and functional specialization (Clapier and Cairns 2009): switch/sucrose non-fermentable (SWI/SNF), imitation switch (ISWI), chromodomain helicase DNA-binding (CHD), and inositol 80 (INO80). ISWI1, the chromatin remodeler investigated in this study, belongs to the ISWI family.

ISWI-containing complexes were first discovered in embryos of *Drosophila melanogaster* (*Drosophila*), where three different complexes were isolated: NURF (nucleosome remodeling factor; (Tsukiyama and Wu 1995)), ACF (ATP-dependent chromatin assembly and remodelling factor; (Ito *et al.* 1997)) and CHRAC (chromatin accessibility complex; (Varga-Weisz *et al.* 1997)). All share ISWI as their ATP-dependent catalytic subunit, while the context of the chromatin remodeling activity is modulated by the other subunits. Both ACF and CHRAC are capable of generating regularly spaced nucleosome arrays and are involved in chromatin assembly during replication and transcription regulation (Yang *et al.* 2006; Sun *et al.* 2001; Fyodorov *et al.* 2004; Erdel and Rippe 2011), while NURF creates randomly spaced nucleosome arrays for transcription regulation (Varga-Weisz *et al.* 1997; Hamiche *et al.* 1999; Barak *et al.* 2003). Chromatin remodeling is a fundamental process in all eukaryotes and therefore, the role of ISWI chromatin remodelers has been studied in a wide range of species, including mammals (Barisic *et al.* 2019), plants (Li *et al.* 2014), yeast (Mellor and Morillon 2004) and ciliates (Fukuda *et al.* 2022).

We showed that ISWI1 binds to ICOP1 and ICOP2 during IES excision in *Paramecium*; however, our data are not conclusive regarding the question of whether the three proteins interact in a single complex or whether ISWI1 forms separate complexes with either ICOP1 or ICOP2. Attempts in our lab to co-inject HA-ICOP1 and ICOP2-GFP for co-immunoprecipitation repeatedly failed. Injected cells did not survive, indicating that overexpression of the ICOPs might be lethal. Diluting the injected DNA to reduce

expression levels likely negatively impacts the microinjection efficiency. Alternatively, silencing-resistant constructs of the ICOPs, with the same amino acid sequence but synomyous codons that differ from the endogenous ones, could be generated. Once injected, the endogenous ICOP expression could be silenced to prevent overexpression and lethality.

We computationally investigated the ISWI1-ICOP complex composition with AlphaFold2 predictions. Although AlphaFold2 provides computationally realistic models, the predictions need to be tested experimentally to validate their reliability. This could not be addressed in the work presented in Chapter 4 and remains to be done in future. AlphaFold2 predicted interactions of the ICOP proteins with the N-terminal part of ISWI1. As a preliminary experiment, I generated ISWI1 truncations according to the truncations given as input to AlphaFold2. In the context of the expression system used in Chapter 4, I co-expressed the His-tagged truncations with GST-tagged ICOP1 or ICOP2 constructs and performed a His-pulldown using nickel beads. Figure 10 below shows a Coomassie gel of an IP on His-ISWI1 truncations.

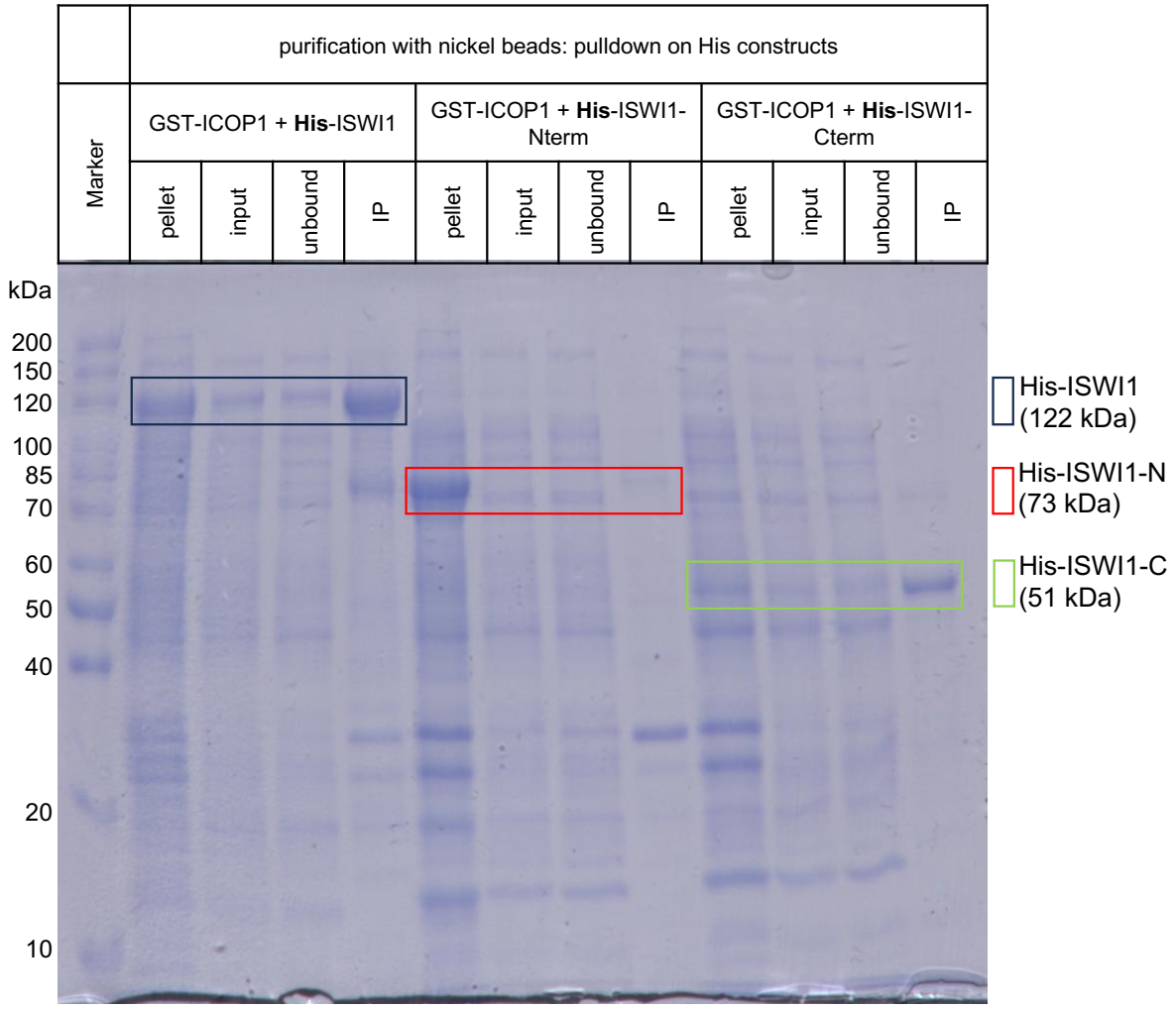


Figure 10: Coomassie gel on a pulldown on His-ISWI1 truncations (indicated on the right). His-tagged constructs were pulled with nickel beads. The constructs co-expressed in E. coli are indicated at the top. For each combination, the pellet, input unbound and IP fraction are loaded. Whole cell lysate was separated into pellet and input by centrifugation. Unbound is the supernatant of the cleared lysate after beads incubation. IP is the eluate from the washed beads. Size reference in kDa for the marker is indicated to the left. His-ISWI1-truncations and their expected localization in the gel are highlighted.

The full-length version (His-ISWI1) and the C-terminal version (His-ISWI1-C) of His-ISWI1 were successfully enriched. However, the N-terminal version (His-ISWI1-N) was predominantly in the pellet fraction and was not enriched with the beads. The same kind of experiment with ICOP2 is not shown, but the Coomassie gel resembled that of ICOP1. More optimization would be needed to solubilize the N-terminal ISWI1 to perform IP experiments. This can be achieved by changes in the buffer conditions for the purification method. Also, the choice of different split points to generate the N-terminal truncation ISWI1 might help enhance folding and solubility.

Ultimately, it would be most beneficial to purify the ISWI1-ICOP complex in high quality and quantity to solve its structure either through crystallography, nuclear magnet resonance or cryo-electron microcopy. With a successfully solved structure, we could gain valuable insights into the details of the interaction interface and the conformation of the complex. To date, ISWI protein structures have only been solved for truncations comprising either the N-terminal part containing the ATPase domain (Yan *et al.* 2016; Yan *et al.* 2019) or the C-terminal part containing the HAND-SANT-SLIDE domain (Yamada *et al.* 2011; Grüne *et al.* 2003). Additionally, crosslinking mass spectrometry (Piersimoni *et al.* 2022) can be used to map the interacting residues within a protein complex.

The function of the ISWI1 complex during IES excision is currently not fully understood. The ICOPs do resemble typical ISWI-binding partners due to their distant homology to Williams–Beuren Syndrome DDT (WSD) motif-containing proteins (Aravind and Iyer 2012). The WSD or D-TOX E motif is found in WSTF (Williams Syndrome Transcription Factor) (Lu *et al.* 1998; Sharif *et al.* 2021), a subunit in the ISWI-containing complex WICH (WSTF–ISWI chromatin remodeling complex) isolated from *Xenopus* (Bozhenok *et al.* 2002). However, other domains found in WSTF, such as the PHD and bromodomain for histone modification recognition, are lacking in the current ISWI1 complex. It is possible that the ICOPs harbor such domains, but their high divergence from other species may prevent their prediction with currently available tools. Consequently, any putative functionality of the ICOPs cannot be concluded based on their domain architecture alone.

ISWI has intrinsic binding capability for nucleosomes and DNA (Grüne *et al.* 2003), but its binding affinity is further modulated by the interaction of its binding partners with DNA and nucleosomes (Fyodorov and Kadonaga 2002). Electrophoretic mobility shift assays (EMSAs) (Hellman and Fried 2007) on purified proteins could be employed to test the DNA binding affinity of ISWI1 and the ICOPs.

Based on our findings, we proposed that the ICOP paralogs influence the directionality of the ISWI1 chromatin remodeling complex. It has already been shown that binding partners can modulate this feature. *In vitro* studies revealed that ISWI alone slides nucleosomes from the center of a DNA fragment to the end but fails to mobilize

nucleosomes positioned at the end of the DNA. However, the addition of Acf1, a subunit of ACF (Ito *et al.* 1999) and CHRAC (Eberharter *et al.* 2001), changes the sliding direction so that end-positioned nucleosomes are moved (Eberharter *et al.* 2001). To confirm a similar role for the ICOPs, these proteins need to be purified and tested in similar *in vitro* experiments. Additionally, since ISWI1's sliding activity has not been demonstrated yet, these experiments could determine whether ISWI1 is a catalytically active chromatin remodeler. Ideally, the activity of the complex should be tested on nucleosomes assembled on DNA fragments containing IES sequences to investigate the influence of the IES sequence on nucleosome positioning. Furthermore, ISWI1-sensitive IESs (i.e., IESs with strong retention upon *ISWI1* knockdown) could be compared to IESs unaffected by *ISWI1* depletion.

To understand how a protein fits into a complex process like genome reorganization, it is important to elucidate its nano-environment. Knowing which proteins it interacts with or which proteins are in close proximity to it provides valuable insights into the interplay of key factors. In the course of this study, multiple IPs have been performed on ISWI1, ICOP1, and ICOP2. Based on this, we confirmed several interactions with western blots (ISWI1 and PTIWI01 with crosslinking, as discussed in Chapter 3; ISWI1 and ICOP1 with/without crosslinking, as discussed in Chapter 4; ISWI1 and ICOP2 with/without crosslinking, as discussed in Chapter 4). However, the mass spectrometry datasets can be further exploited to investigate the interaction network.

Figure 11 displays the volcano plots visualizing the mass spectrometry datasets created in Chapter 4, with selected proteins known to be involved in IES excision highlighted. Additionally, the unpublished dataset of an HA-affinity IP in ICOP-HA + ISWI1-GFP co-injected cells is shown (Figure 11D). Table 1 summarizes the detectability of the selected proteins.

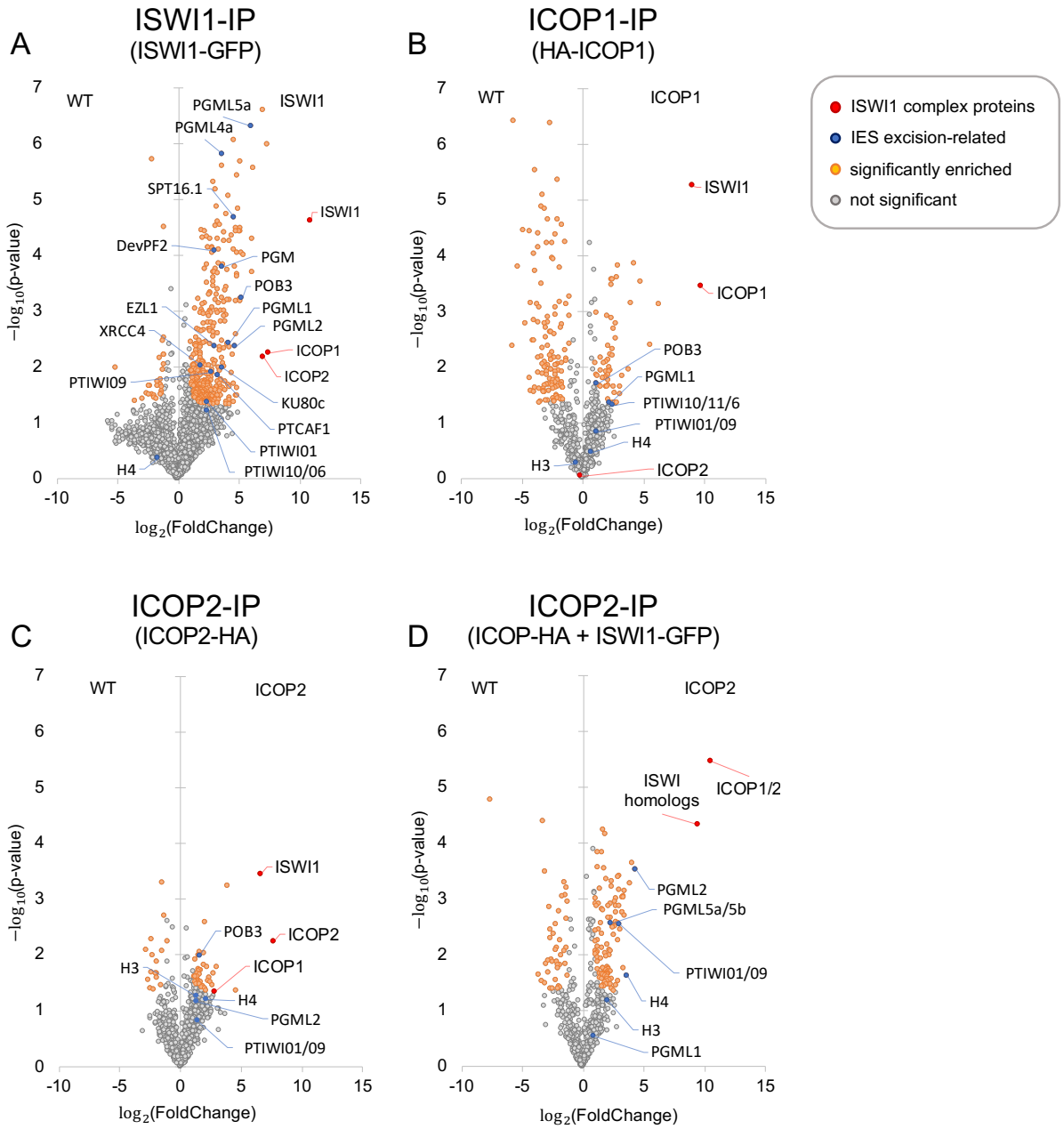


Figure 11: Proteins identified by mass spectrometry in pulldowns of tagged ISWI1, ICOP1 and ICOP2 proteins. Comparison of injected cells (the constructs are given in parentheses) and wild type (WT). A to C: published datasets generated in Chapter 4. D: unpublished dataset generated in Chapter 4. Dataset from ISWI1-HA pulldown (Chapter 3) not shown.

The ISWI1-IP showed a higher abundance of enriched proteins than the ICOP paralogs (Figure 11A). Among these, many proteins are known to be active in the new MAC, consistent with ISWI1's localization to the new MACs. For example, members of the excision complex, PGM, PGMLs, and KU80c, were among the enriched proteins. Some of the PGMLs were also detected in the ICOP-IPs. This observation is plausible, as we expect that the ISWI1 chromatin remodeling complex prepares the

chromatin around IESs for PGM complex binding. Other proteins, like the DCLs, are not detected in either of the IPs, indicating that they do not act in close proximity to the ISWI1 complex.

Table 1: Proteins detected by mass spectrometry in ISWI1-, ICOP1- and ICOP2-IP (see Figure 11). Status of detection is given for selected proteins. Enriched hits (unique protein identified): "+", enriched hits (peptides assigned to multiple proteins): "(+)", no hit: "n") and not detected "-".

	ISWI1-IP	ICOP1-IP	ICOP2-IP	ICOP2-IP (+ISWI1)
ISWI1	+	+	+	(+)
ICOP1	+	+	+	(+)
ICOP2	+	n	+	(+)
POB3	+	+	+	-
SPT16.1	+	-	-	-
PTIWI01	+	n	n	(+)
PTIWI09	+	n	n	(+)
DCL2	-	-	-	-
DevPF1	-	-	-	-
DevPF2	+	-	-	-
EZL1	+	-	-	-
PTCAF1	+	-	-	-
PGM	+	-	-	-
PGML1	+	+	n	n
PGML2	+	n	n	+
PGML4a	+	-	-	-
PGML5a	+	-	-	(+)
KU80c	+	-	-	-
XRCC4	+	-	-	-
DCL5	-	-	-	-
PTIWI10/11	n	n	n	-
H4	n	n	n	(+)
H3	n	n	n	n

One protein enriched in all three published IPs is POB3 ([DNA] Polymerase One Binding protein 3). POB3 and SPT16 form the histone chaperone complex FACT (FAcilitates Chromatin Transcription) (Gurova *et al.* 2018; Orphanides *et al.* 1998). Histone chaperones assist in the formation and dismantling of nucleosomes for replication and transcription (Hammond *et al.* 2017). In contrast to chromatin remodelers, the activity of histone chaperones is independent of ATP. Upon binding, the yeast FACT complex reversibly uncoils the DNA from the nucleosome without removing or shifting the histones (Valieva *et al.* 2016). Interfaces of FACT compete with the DNA for the histone binding sites and thereby stabilize the histones in an otherwise thermodynamically disfavored state (Hondele and Ladurner 2013). Histone

chaperones work together with chromatin remodelers to modulate chromatin accessibility; for example, they remove nucleosomes from promoter regions for transcription initiation (Erkina and Erkine 2015; Gkikopoulos *et al.* 2009; Philpott *et al.* 2000).

In *Drosophila*, the ISWI-containing remodeling complex ACF works closely with the histone chaperones NAP-1 or CAF-1 for chromatin assembly and transcriptional activation (Ito *et al.* 1997). NAP-1 and CAF-1 can integrate histones into DNA in an ATP-independent manner to form randomly distributed nucleosomes and the ATP-dependent activity of ACF subsequently generates regularly spaced nucleosome arrays. Hence, POB3 is a likely candidate for an additional complex member or direct interaction partner of the ISWI1 chromatin remodeling complex in *Paramecium*.

The role of FACT during *Paramecium* IES excision has already been investigated (de Vanssay *et al.* 2020). Although *POB3* is upregulated during sexual development, its knockdown does not affect the survival of the new progeny (de Vanssay *et al.* 2020). This observation does not exclude POB3 functioning during genome reorganization. DCL2 and 3 single knockdowns do not affect cell survival and show lethality only in the double-knockdown (Lepère *et al.* 2009; Sandoval *et al.* 2014). Since only two IESs were tested for IES retention in a POB3 knockdown (de Vanssay *et al.* 2020), more subtle effects on genome reorganization might have been missed. A homolog of the second subunit of the FACT complex, SPT16.1, has been studied in more detail and showed strong effects on genome reorganization, including cell death in the new progeny, strong retention of IESs, and reduced production of iesRNAs (de Vanssay *et al.* 2020). The interaction between POB3 and SPT16 has not been verified in *Paramecium*, but in *Tetrahymena* both proteins appeared in IPs of tagged histone H2A and H2B proteins (Ashraf *et al.* 2019). SPT16.1 was enriched in the ISWI1-IP, too, but in neither of the ICOP-IPs. Clearly, more IP experiments are needed to test for direct interactions between FACT and the ISWI1 complex.

Since ISWI complexes typically slide nucleosomes, I checked for histones in the mass spectrometry datasets. The only enriched hit was histone 4 (H4) in the unpublished ICOP2-IP (Figure 11D; Table 1). Histones H4 and H3 were also detected in the published ICOP2-IP, although not classified as enriched. H4 showed more of a

depletion in the ISWI1-IP. This is interesting since *in vitro* studies using reconstituted *Xenopus laevis* histone octamers revealed that the tail of H4, but none of the other histones, directly regulates the catalytic activity of recombinant ISWI complexes (Clapier *et al.* 2001).

ISWI has an autoregulatory region, AutoN, that is conserved across multiple species and binds to the ATPase domain in the absence of a nucleosome, thereby inhibiting the ATPase activity (Clapier and Cairns 2012; Yan *et al.* 2016). It was proposed that AutoN and its adjacent regions compete with the H4 tail for an overlapping binding interface, leading to the replacement of AutoN upon nucleosome binding, allowing for remodeling activity (Clapier and Cairns 2012; Hwang *et al.* 2014; Ludwigsen *et al.* 2017). Arginine 17, which is part of the basic patch within the H4 tail, was identified as the most crucial residue for the binding of H4 to ISWI isolated from the yeast species *Myceliophthora thermophila* and *Saccharomyces cerevisiae* (Yan *et al.* 2019; Yan *et al.* 2016). Furthermore, the replacement of AutoN by H4 was proposed to regulate the linker DNA length. When nucleosomes are tightly packed, the H4 tail interacts with neighboring nucleosomes (Luger *et al.* 1997), allowing AutoN to inhibit the ATPase activity when short linker DNA is encountered (Hwang *et al.* 2014).

In the published structure of the yeast ISWI ATPase domain with a nucleosome (Yan *et al.* 2019), the H4 tail extends towards ISWI precisely at the interface AlphaFold2 predicted the interaction between ISWI1 and the ICOP paralogs (Figure 12). More strikingly, arginine residues from the ICOPs extend towards the same pocket as arginine 17 from the H4 tail. Taking this prediction into account, along with the mass spectrometry data, it is plausible that ICOP1 (or 2) competes with H4 for binding. In this scenario H4 would not show an interaction but a depletion in ISWI1-IP and instead would be enriched in the ICOP-IPs, as observed. Further experiments would be needed to investigate the binding capability of the ICOPs to the H4 tail and how this modulates ISWI1 activity.

The basic patch in the H4 tail (RHRK) is extremely conserved across eukaryotes, ranging from mammals to plants (Kayne *et al.* 1988). In the alignment produced by Kayne et al, the ciliate *Tetrahymena* was the only species showing an insertion in the basic patch, that is also present in *Paramecium* (RHARK). This raises the question

how easily the findings acquired in other organisms regarding the basic patch can be transferred to *Paramecium*. Considering *Paramecium*'s unusual genome structure with nucleosome-free intragenic regions (Drews *et al.* 2022) and extremely short linker DNA in the MAC (Gnan *et al.* 2022), it also would need to be determined whether *Paramecium*'s ISWI complexes are regulated in the same way as in other organisms and how interaction partners modulate the remodeling activity to accommodate the nucleosomal differences associated with the tremendous evolutionary spans involved.

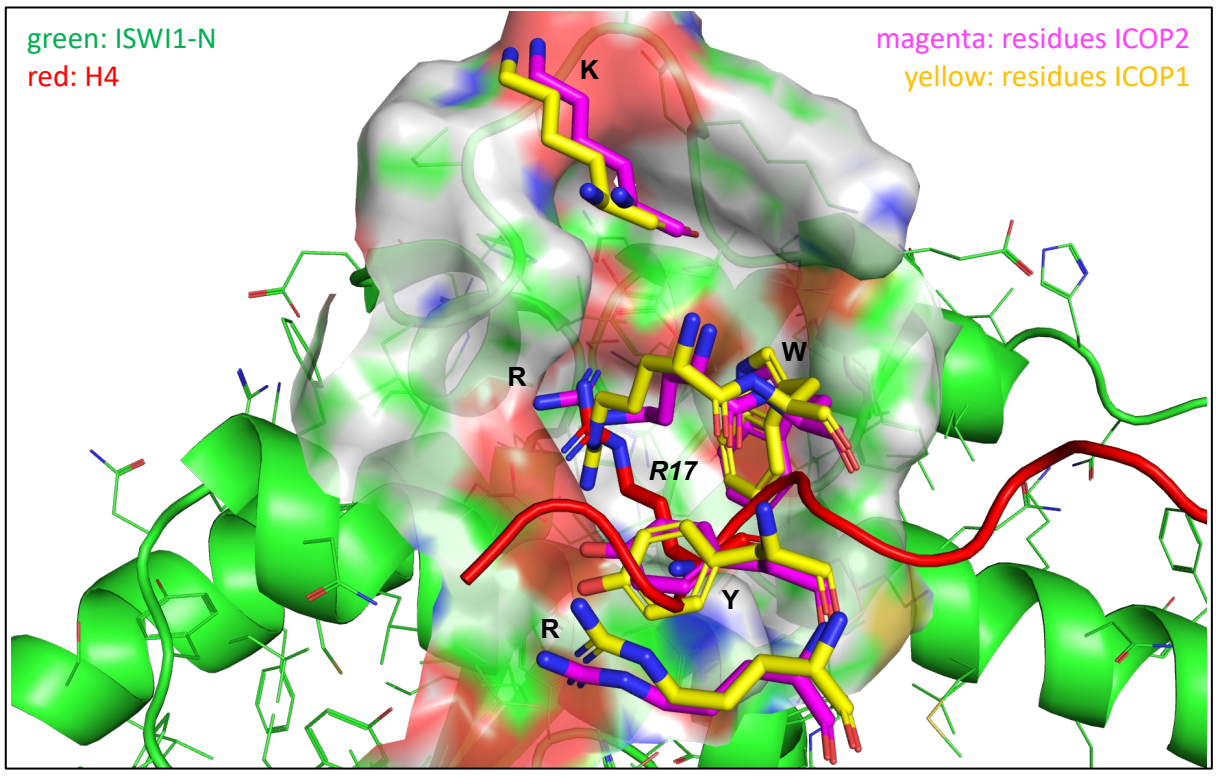


Figure 12: The predicted interaction interface of ISWI1 and ICOPs in relation to H4. The heterodimer prediction (AlphaFold2) of N-terminal ISWI1 (ISWI1-N; green) and ICOP1 (yellow) or ICOP2 (magenta) are superimposed with the published structure from yeast ISW1 in the context of a reconstituted nucleosome (PDB accession number 6JYL). The binding pocket in the ISWI1-N is displayed as a surface. For the ICOPs, only the interacting residues are displayed as sticks and the amino acids are specified in one letter code. From the published structure, only the histone 4 (red) is shown. Arginine 17 is displayed as sticks and labeled in italic one letter code. Structural alignment was performed with the ISWI N-terminal domain.

6.2 Discussion and future directions for the DevPF proteins

Many ISWI-containing chromatin remodelers harbor PHD finger subunits (Wysocka *et al.* 2006; Tan *et al.* 2020; Bozhenok *et al.* 2002). Acf1, present in both ACF and CHRAC

complexes, harbors a bromodomain, two PHD fingers, and WAC/WAKZ motifs (Ito *et al.* 1999) and shares a closely related domain architecture with WSTF (Bozhenok *et al.* 2002). In our ISWI1-IPs, DevPF2 was repeatedly detected (mass spectrometry datasets from Chapter 3 and 4; Figure 11A; Table 1). However, based on the findings in Chapter 5, neither DevPF1 nor DevPF2 seem to be functionally related to the ISWI1 complex: ISWI1 was not detected in the DevPF-IPs and the DevPFs do not cause enhanced alternative excision. In fact, very few of the know proteins involved in IES excision were identified in the in the DevPF-IPs, indicating they might be involved in less well characterized processes contributing to genome reorganization.

Most reported PHD finger proteins bind to histone modifications (Sanchez and Zhou 2011) and thereby recruit chromatin-regulating enzymes, such as chromatin remodelers or histone-modifying enzymes (Taverna *et al.* 2006), context-specific to their site of action. Our data suggest that DevPF1 and DevPF2 regulate non-coding and coding transcription in the MICs (only DevPF1) and in the new MAC (DevPF1 and DevPF2). Histone-matching peptides were identified in the DevPF-IPs, indicating a potential interaction with histone modifications. Since histones are highly abundant nuclear proteins, the specificity of these interactions needs to be investigated in more detail. To identify the histone modifications to which DevPF1 and DevPF2 might bind, the proteins could be purified and subjected to peptide microarrays.

Chromatin immunoprecipitation (ChIP) followed by high-throughput sequencing provides information about DNA sequences associated with a protein and has, therefore, been a standard method to investigate binding sites for transcription factors (Weinmann and Farnham 2002; Gade and Kalvakolanu 2012). ChIP has been established in *Paramecium* and was used to investigate the chromatin landscape of different histone modifications in the MAC genome (Cheaib and Simon 2013; Drews *et al.* 2022).

ChIP works well on highly expressed proteins that bind strongly to DNA, like histones. Transcription factors, however, bind indirectly and transiently to DNA and are challenging targets. Advanced methods like Cut-and-Run (Skene *et al.* 2018; Kong *et al.* 2021) may overcome some of these issues since they only analyze genomic sequences proximate to the protein of interest. This should drastically reduce the

signal-to-noise ratio, the required sequencing depth, and the input material, making this method more suitable for low-expressed proteins and transcription factors. The DevPF constructs produced in this study can be used as the basis to establish one of the above-mentioned methods to identify DNA sequences associated with DevPF1 or DevPF2. DevPF binding sites on chromatin can be associated with the differential gene expression data acquired in Chapter 5 to see how the genes located at the DevPF binding sites respond to *DevPF1* or *DevPF2* depletion.

The selective localization of the DevPF1-GFP fusion protein to certain gametic and post-zygotic nuclei (Figure 9D) is not immediately linked to nuclear division or nuclear fate decisions. This raises the question of how the nuclei individually recruit DevPF1-GFP and how DevPF1's presence influences the nuclei. There are other examples of multinucleated cells where the nuclei exhibit individual behavior. In many plant species, female gametophyte development includes multinuclear cell stages, where the nuclei give rise to functionally different cell types (Yadegari and Drews 2004). In the multinuclear cells of fungi (Kokkoris *et al.* 2020), the nuclei can divide or migrate asynchronously (Gladfelter 2006; Evangelisti *et al.* 2019; Stein *et al.* 2020) and show nucleus-specific gene expression (Gehrmann et al. 2018). The *Drosophila* embryo is one of the most extensively studied models for gene expression patterns across multinucleated cells. After fertilization, the nuclei divide without cytokinesis, resulting in the syncytial blastoderm that encloses ~6000 nuclei in a shared cytoplasm (Foe and Alberts 1983). The shared cytoplasm allows the establishment of gradients that regulate gene expression patterns, depending on the nuclei's localization. In this way, the position for tissue development is determined. Due to the non-uniform distribution exhibited by the participating regulatory proteins, they are referred to as morphogens (Rogers and Schier 2011).

The standard example of morphogens is Bicoid, a homeodomain-containing transcription factor that establishes the anterior-posterior axis in the embryo (Frohnhöfer and Nüsslein-Volhard 1986; Struhl *et al.* 1989; Frigerio *et al.* 1986). The protein localizes in a gradient, with high concentrations at the anterior of the embryo and no presence at the posterior (Driever and Nüsslein-Volhard 1988). Bicoid regulates gene expression in a concentration-dependent manner: high concentrations are required to develop the anteriormost region, intermediate concentrations promote

the development of the head, and low concentrations generate the thorax of the fruit fly (Driever and Nüsslein-Volhard 1988). The protein gradient is proposed to be established by mRNA transcription at the anterior end and the diffusion of both mRNA and protein towards the posterior end, combined with degradation of both mRNA and protein towards the posterior end (Spirov *et al.* 2009).

Another well-studied morphogen in *Drosophila* is the transcription factor Dorsal (Roth *et al.* 1989; Jiang *et al.* 1992), responsible for the dorsal-ventral axis. In contrast to Bicoid, the gradient regulating Dorsal activity is not a gradient of Dorsal localization but rather of its ability to translocate into nuclei. In the early stages of embryo development, Dorsal is evenly distributed across the cytoplasm, but it is selectively imported into nuclei on the dorsal side at later stages (Roth *et al.* 1989). In the cytoplasm, Dorsal forms a complex with Cactus, which prevents its import into nuclei (Morisato and Anderson 1995). The Spätzle-Toll signaling pathway transmits an extracellular signal into specific cytoplasmic regions on the dorsal side, where Dorsal is released from Cactus and enters the nuclei (Morisato and Anderson 1995).

These studies demonstrate the importance of nuclei-specific localization of nuclear proteins in differentiation and development. Of the two described regulatory mechanisms, the DevPF1-GFP localization pattern exhibits more similarities with Dorsal. It is uniformly distributed in the cytoplasm, and nuclei-specific localization may be controlled by selective import into the nuclei. However, in contrast to Dorsal, we did not observe a general trend indicating that the subcellular localization of the nuclei correlates with DevPF1-GFP translocation. A more extensive localization study of cells at the relevant developmental stages is needed to complete the picture. While translocation is likely not controlled by extracellular signals, the general mechanism of Dorsal regulation might still apply: a signaling pathway that locally enables the import of DevPF1 into the nucleus.

Nuclear import is typically mediated by importins that guide the protein through the nuclear pore complex (Christophe *et al.* 2000). In *Tetrahymena*, it was shown that the MICs and MACs have both shared and nucleus-specific nuclear pore complex subunits, with the latter generating distinct permeability for nuclear proteins (Malone *et al.* 2008; Iwamoto *et al.* 2009; Iwamoto *et al.* 2015). Also, importin α-like proteins

localize in a MIC-specific manner (Malone *et al.* 2008). Clearly, the nuclear import machinery regulates the nuclei-specific proteome and influences nuclear differentiation and regulation (Yasuhara *et al.* 2007; Yasuhara *et al.* 2013). There are different possibilities for how the interaction of a nuclear protein with importins might be regulated: by the degradation of a cytoplasmic anchor (as is the case for Dorsal and Cactus), by the interaction with an adapter protein, or by post-translational modifications, such as phosphorylation.

In future, separating the cytosolic and nuclear fractions of DevPF1-GFP-injected cells during development will allow for the investigation of differences in binding partners and/or post-translational modifications between cytoplasmic and nuclear DevPF1. However, since DevPF1-GFP localizes to only some of the MICs, it might be challenging to produce enough material for subsequent analysis. As a first step, it might be useful to compare the cytosolic fraction of cells at early developmental stages (where most of the DevPF1-GFP is in the cytoplasm) with the new MAC fraction of cells at late developmental stages (where most of the DevPF1-GFP is in the new MACs). To investigate the nuclear proteome specific to DevPF1-GFP positive nuclei, gametic and post-zygotic nuclei could be separated based on the presence or absence of DevPF1-GFP applying fluorescence-activated nuclear sorting (FANS), a technique that has already been established in *Paramecium* (Guérin *et al.* 2017; Zangarelli *et al.* 2022).

A comparative analysis between DevPF1-GFP positive or negative nuclei based on mass spectrometry could provide insights into their specific nuclear proteomes. The MIC and MAC proteomes of *Oxytricha* were analyzed for the first time in a ciliate (Lu *et al.* 2023), but these nuclei were separated by a traditional discontinuous sucrose gradient method rather than flow sorting. As expected, due to their distinct functions, the two types of nuclei differ in their chromatin structure, with specific H3 variants associated with either MIC or MAC. In the MAC, transcription-related proteins are enriched. Identifying MIC-specific nuclear proteins is more challenging. This is partly due to the smaller MIC size (hence, less material) and partly because the authors mentioned impurities in the MIC enrichment, particularly from mitochondria. The use of flow sorting to purify DevPF1-GFP positive or negative nuclei should overcome this issue.

Both Bicoid and Dorsal regulate gene expression, and we also found indications that DevPF1 and DevPF2 might function as transcription regulators. First, DevPF1 likely regulates non-coding transcription for scnRNA production which occurs during the S-phase before the first meiotic division. Later, DevPF1 and DevPF2 seem to be involved in gene expression from the new MAC. However, through changes in the chromatin structure they might also influence chromatin-related processes other than transcription. DNA repair, essential for chromatin reorganization, is tightly controlled by histone modifications and their variants (Ferrand *et al.* 2021). Also, chromatin changes are expected to occur in gametic cells before degradation, as it was shown for apoptotic cells (Koukalová *et al.* 1997; Füllgrabe *et al.* 2010). More insights into DevPFs binding partners and their associated DNA sequences might provide answers to their exact function in genome reorganization.

6.3 Conclusion

Genome reorganization is a fundamental process underlying functional gene expression from protists to mammals and plants. In humans, programmed DNA reorganization is highly relevant. V(D)J recombination (Bassing *et al.* 2002; Schatz and Ji 2011; Rooney *et al.* 2004) and class-switch recombination (Chaudhuri and Alt 2004) are essential for the diverse substrate recognition of the immune system. Aberrant genome reorganization is also associated with some cancers (Mani and Chinnaiyan 2010; Jones and Jallepalli 2012; Forment *et al.* 2012; Zhang *et al.* 2013).

Ciliates are unique in several ways and their genome reorganization likely evolved independently well before that in multicellular organisms, given that their proposed IES excisases exist in distantly related lineages (Singh *et al.* 2021). Nevertheless, the conservation and use of some of the same fundamental molecules, such as those involved in NHEJ repair of DNA and chromatin remodeling, allow comparison and investigation of potential mechanistic generality and evolvability. The ciliate *Paramecium* continues to be an excellent model organism to study genome reorganization as it undergoes a massive amount of genome reorganization each sexual cycle to generate a streamlined somatic genome. This study has provided novel insights into multiple aspects of the precise excision of IESs in *Paramecium*, like

small RNA biogenesis, chromatin remodeling, and gene expression, and paves the way for a better understanding of development-specific DNA elimination.

References

Abello, A., Régnier, V., Arnaiz, O., Le Bars, R., Bétermier, M., & Bischerour, J. (2020). Functional diversification of Paramecium Ku80 paralogs safeguards genome integrity during precise programmed DNA elimination. *PLoS Genetics*, *16*(4), e1008723. https://doi.org/10.1371/journal.pgen.1008723

Adl, S. M., Simpson, A. G. B., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., Bowser, S. S., Brugerolle, G., Fensome, R. A., Fredericq, S., James, T. Y., Karpov, S., Kugrens, P., Krug, J., Lane, C. E., Lewis, L. A., Lodge, J., Lynn, D. H., Mann, D. G., … Taylor, M. F. J. R. (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *The Journal of Eukaryotic Microbiology*, *52*(5), 399–451. https://doi.org/10.1111/j.1550-7408.2005.00053.x

Allen, S. E., Hug, I., Pabian, S., Rzeszutek, I., Hoehener, C., & Nowacki, M. (2017). Circular Concatemers of Ultra-Short DNA Segments Produce Regulatory RNAs. *Cell*, *168*(6), 990-999.e7. https://doi.org/10.1016/j.cell.2017.02.020

Aravind, L., & Iyer, L. M. (2012). The HARE-HTH and associated domains: novel modules in the coordination of epigenetic DNA and protein modifications. *Cell Cycle*, *11*(1), 119–131. https://doi.org/10.4161/cc.11.1.18475

Ardell, D. H., Lozupone, C. A., & Landweber, L. F. (2003). Polymorphism, recombination and alternative unscrambling in the DNA polymerase alpha gene of the ciliate Stylonychia lemnae (Alveolata; class Spirotrichea). *Genetics*, *165*(4), 1761–1777. https://doi.org/10.1093/genetics/165.4.1761

Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.-M., Denby Wilkes, C., Garnier, O., Labadie, K., Lauderdale, B. E., Le Mouël, A., Marmignon, A., Nowacki, M., Poulain, J., Prajer, M., Wincker, P., Meyer, E., Duharcourt, S., Duret, L., Bétermier, M., & Sperling, L. (2012). The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal

eliminated sequences. *PLoS Genetics*, *8*(10), e1002984. https://doi.org/10.1371/journal.pgen.1002984

Ashraf, K., Nabeel-Shah, S., Garg, J., Saettone, A., Derynck, J., Gingras, A.-C., Lambert, J.-P., Pearlman, R. E., & Fillingham, J. (2019). Proteomic Analysis of Histones H2A/H2B and Variant Hv1 in Tetrahymena thermophila Reveals an Ancient Network of Chaperones. *Molecular Biology and Evolution*, *36*(5), 1037–1055. https://doi.org/10.1093/molbev/msz039

Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B. M., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Câmara, F., Duharcourt, S., Guigo, R., Gogendeau, D., … Wincker, P. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, *444*(7116), 171–178. https://doi.org/10.1038/nature05230

Barak, O., Lazzaro, M. A., Lane, W. S., Speicher, D. W., Picketts, D. J., & Shiekhattar, R. (2003). Isolation of human NURF: a regulator of Engrailed gene expression. *The EMBO Journal*, *22*(22), 6089–6100. https://doi.org/10.1093/emboj/cdg582

Barisic, D., Stadler, M. B., Iurlaro, M., & Schübeler, D. (2019). Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature*, *569*(7754), 136–140. https://doi.org/10.1038/s41586-019-1115-5

Baroin, A., Prat, A., & Caron, F. (1987). Telomeric site position heterogeneity in macronuclear DNA of Paramecium primaurelia. *Nucleic Acids Research*, *15*(4), 1717–1728. https://doi.org/10.1093/nar/15.4.1717

Bassing, C. H., Swat, W., & Alt, F. W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. *Cell*, *109 Suppl*, S45-55. https://doi.org/10.1016/s0092-8674(02)00675-x

Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E., & Bétermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in

programmed genome rearrangements in the ciliate Paramecium tetraurelia. *Genes & Development*, *23*(21), 2478–2483. https://doi.org/10.1101/gad.547309

Bayless, B. A., Navarro, F. M., & Winey, M. (2019). Motile cilia: innovation and insight from ciliate model organisms. *Frontiers in Cell and Developmental Biology*, *7*, 265. https://doi.org/10.3389/fcell.2019.00265

Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J. R., & Sperling, L. (2010). DNA microinjection into the macronucleus of paramecium. *Cold Spring Harbor Protocols*, *2010*(1), pdb.prot5364. https://doi.org/10.1101/pdb.prot5364

Bétermier, M., Duharcourt, S., Seitz, H., & Meyer, E. (2000). Timing of developmentally programmed excision and circularization of Paramecium internal eliminated sequences. *Molecular and Cellular Biology*, *20*(5), 1553–1561. https://doi.org/10.1128/MCB.20.5.1553-1561.2000

Bischerour, J., Bhullar, S., Denby Wilkes, C., Régnier, V., Mathy, N., Dubois, E., Singh, A., Swart, E., Arnaiz, O., Sperling, L., Nowacki, M., & Bétermier, M. (2018). Six domesticated PiggyBac transposases together carry out programmed DNA elimination in Paramecium. *ELife*, *7*. https://doi.org/10.7554/eLife.37927

Boscaro, V., & Keeling, P. J. (2023). How ciliates got their nuclei. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(7), e2221818120. https://doi.org/10.1073/pnas.2221818120

Bouhouche, K., Gout, J.-F., Kapusta, A., Bétermier, M., & Meyer, E. (2011). Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Research*, *39*(10), 4249–4264. https://doi.org/10.1093/nar/gkq1283

Bozhenok, L., Wade, P. A., & Varga-Weisz, P. (2002). WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. *The EMBO Journal*, *21*(9), 2231–2241. https://doi.org/10.1093/emboj/21.9.2231

Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., & Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*, *84*(6), 843–851. https://doi.org/10.1016/s0092-8674(00)81063-6

Carradec, Q., Götz, U., Arnaiz, O., Pouch, J., Simon, M., Meyer, E., & Marker, S. (2015). Primary and secondary siRNA synthesis triggered by RNAs from food bacteria in the ciliate Paramecium tetraurelia. *Nucleic Acids Research*, *43*(3), 1818–1833. https://doi.org/10.1093/nar/gku1331

Cech, T. R. (1990). Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena. *Bioscience Reports*, *10*(3), 239–261. https://doi.org/10.1007/BF01117241

Charmant, O., Gruchota, J., Arnaiz, O., Zangarelli, C., Bétermier, M., Nowak, K., Legros, V., Chevreux, G., Nowak, J., & Duharcourt, S. (2023). The nuclear PIWI-interacting protein Gtsf1 controls the selective degradation of small RNAs in *Paramecium*. *BioRxiv*. https://doi.org/10.1101/2023.09.19.558372

Chaudhuri, J., & Alt, F. W. (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nature Reviews. Immunology*, *4*(7), 541–552. https://doi.org/10.1038/nri1395

Cheaib, M., & Simon, M. (2013). Dynamic chromatin remodelling of ciliate macronuclear DNA as determined by an optimized chromatin immunoprecipitation (ChIP) method for Paramecium tetraurelia. *Applied Microbiology and Biotechnology*, *97*(6), 2661–2670. https://doi.org/10.1007/s00253-013-4708-1

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. *Molecular Biology of the Cell*, *21*(10), 1753–1762. https://doi.org/10.1091/mbc.e09-12-1079

Chen, X., Bracht, J. R., Goldman, A. D., Dolzhenko, E., Clay, D. M., Swart, E. C., Perlman, D. H., Doak, T. G., Stuart, A., Amemiya, C. T., Sebra, R. P., &

Landweber, L. F. (2014). The architecture of a scrambled genome reveals massive levels of genomic rearrangement during development. *Cell*, *158*(5), 1187–1198. https://doi.org/10.1016/j.cell.2014.07.034

Christophe, D., Christophe-Hobertus, C., & Pichon, B. (2000). Nuclear targeting of proteins. *Cellular Signalling*, *12*(5), 337–341. https://doi.org/10.1016/S0898-6568(00)00077-2

Clapier, C R, Längst, G., Corona, D. F., Becker, P. B., & Nightingale, K. P. (2001). Critical role for the histone H4 N terminus in nucleosome remodeling by ISWI. *Molecular and Cellular Biology*, *21*(3), 875–883. https://doi.org/10.1128/MCB.21.3.875-883.2001

Clapier, Cedric R, & Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annual Review of Biochemistry*, *78*, 273–304. https://doi.org/10.1146/annurev.biochem.77.062706.153223

Clapier, Cedric R, & Cairns, B. R. (2012). Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature*, *492*(7428), 280–284. https://doi.org/10.1038/nature11625

Coyne, R. S., Lhuillier-Akakpo, M., & Duharcourt, S. (2012). RNA-guided DNA rearrangements in ciliates: is the best genome defence a good offence? *Biology of the Cell*, *104*(6), 309–325. https://doi.org/10.1111/boc.201100057

Denby Wilkes, C., Arnaiz, O., & Sperling, L. (2016). ParTIES: a toolbox for *Paramecium* interspersed DNA elimination studies. *Bioinformatics*, *32*(4), 599–601. https://doi.org/10.1093/bioinformatics/btv691

de Vanssay, A., Touzeau, A., Arnaiz, O., Frapporti, A., Phipps, J., & Duharcourt, S. (2020). The Paramecium histone chaperone Spt16-1 is required for Pgm endonuclease function in programmed genome rearrangements. *PLoS Genetics*, *16*(7), e1008949. https://doi.org/10.1371/journal.pgen.1008949

Dickerson, H. W., & Clark, T. G. (1996). Immune response of fishes to ciliates. *Annual Review of Fish Diseases*, *6*, 107–120. https://doi.org/10.1016/S0959-8030(96)90008-3

Dierssen, H., McManus, G. B., Chlus, A., Qiu, D., Gao, B.-C., & Lin, S. (2015). Space station image captures a red tide ciliate bloom at high spectral and spatial resolution. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(48), 14783–14787. https://doi.org/10.1073/pnas.1512538112

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., & Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell*, *122*(3), 473–483. https://doi.org/10.1016/j.cell.2005.07.013

Dobell, C. (1932). *Antony van Leeuwenhoek and his "Little Animals".* John Bale, Sons & Danielsson, Ltd.

Drews, F., Boenigk, J., & Simon, M. (2022). Paramecium epigenetics in development and proliferation. *The Journal of Eukaryotic Microbiology*, *69*(5), e12914. https://doi.org/10.1111/jeu.12914

Drews, F., Salhab, A., Karunanithi, S., Cheaib, M., Jung, M., Schulz, M. H., & Simon, M. (2022). Broad domains of histone marks in the highly compact Paramecium macronuclear genome. *Genome Research*, *32*(4), 710–725. https://doi.org/10.1101/gr.276126.121

Driever, W., & Nüsslein-Volhard, C. (1988a). A gradient of bicoid protein in Drosophila embryos. *Cell*, *54*(1), 83–93. https://doi.org/10.1016/0092-8674(88)90182-1

Driever, W., & Nüsslein-Volhard, C. (1988b). The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell*, *54*(1), 95–104. https://doi.org/10.1016/0092-8674(88)90183-3

Dubois, E., Bischerour, J., Marmignon, A., Mathy, N., Régnier, V., & Bétermier, M. (2012). Transposon invasion of the paramecium germline genome countered by a

domesticated piggybac transposase and the NHEJ pathway. *International Journal of Evolutionary Biology*, *2012*, 436196. https://doi.org/10.1155/2012/436196

Dussutour, A. (2021). Learning in single cell organisms. *Biochemical and Biophysical Research Communications*, *564*, 92–102. https://doi.org/10.1016/j.bbrc.2021.02.018

Dziallas, C., Allgaier, M., Monaghan, M. T., & Grossart, H.-P. (2012). Act together-implications of symbioses in aquatic ciliates. *Frontiers in Microbiology*, *3*, 288. https://doi.org/10.3389/fmicb.2012.00288

Eberharter, A., Ferrari, S., Längst, G., Straub, T., Imhof, A., Varga-Weisz, P., Wilm, M., & Becker, P. B. (2001). Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. *The EMBO Journal*, *20*(14), 3781–3788. https://doi.org/10.1093/emboj/20.14.3781

Elick, T. A., Bauser, C. A., & Fraser, M. J. (1996). Excision of the piggyBac transposable element in vitro is a precise event that is enhanced by the expression of its encoded transposase. *Genetica*, *98*(1), 33–41. https://doi.org/10.1007/BF00120216

Erdel, F., & Rippe, K. (2011). Chromatin remodelling in mammalian cells by ISWI-type complexes--where, when and why? *The FEBS Journal*, *278*(19), 3608–3618. https://doi.org/10.1111/j.1742-4658.2011.08282.x

Erkina, T. Y., & Erkine, A. (2015). ASF1 and the SWI/SNF complex interact functionally during nucleosome displacement, while FACT is required for nucleosome reassembly at yeast heat shock gene promoters during sustained stress. *Cell Stress & Chaperones*, *20*(2), 355–369. https://doi.org/10.1007/s12192-014-0556-x

Evangelisti, E., Shenhav, L., Yunusov, T., Le Naour-Vernet, M., Rink, P., & Schornack, S. (2019). Hydrodynamic shape changes underpin nuclear rerouting in branched hyphae of an oomycete pathogen. *MBio*, *10*(5). https://doi.org/10.1128/mBio.01516-19

Fang, W., Wang, X., Bracht, J. R., Nowacki, M., & Landweber, L. F. (2012). Piwi-interacting RNAs protect DNA against loss during *Oxytricha* genome rearrangement. *Cell*, *151*(6), 1243–1255. https://doi.org/10.1016/j.cell.2012.10.045

Ferrand, J., Plessier, A., & Polo, S. E. (2021). Control of the chromatin response to DNA damage: Histone proteins pull the strings. *Seminars in Cell & Developmental Biology*, *113*, 75–87. https://doi.org/10.1016/j.semcdb.2020.07.002

Finlay, B. J., Corliss, J. O., Esteban, G., & Fenchel, T. (1996). Biodiversity at the Microbial Level: The Number of Free-Living Ciliates in the Biosphere. *The Quarterly Review of Biology*, *71*(2), 221–237. https://doi.org/10.1086/419370

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, *391*(6669), 806–811. https://doi.org/10.1038/35888

Foe, V. E., & Alberts, B. M. (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. *Journal of Cell Science*, *61*, 31–70. https://doi.org/10.1242/jcs.61.1.31

Foissner, W. (1997). Soil ciliates (Protozoa: Ciliophora) from evergreen rain forests of Australia, South America and Costa Rica: diversity and description of new species. *Biology and Fertility of Soils*, *25*(4), 317–339. https://doi.org/10.1007/s003740050322

Forment, J. V., Kaidi, A., & Jackson, S. P. (2012). Chromothripsis and cancer: causes and consequences of chromosome shattering. *Nature Reviews. Cancer*, *12*(10), 663–670. https://doi.org/10.1038/nrc3352

Forney, J. D., & Blackburn, E. H. (1988). Developmentally controlled telomere addition in wild-type and mutant paramecia. *Molecular and Cellular Biology*, *8*(1), 251–258. https://doi.org/10.1128/mcb.8.1.251-258.1988

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., & Noll, M. (1986). Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. *Cell*, *47*(5), 735–746. https://doi.org/10.1016/0092-8674(86)90516-7

Frohnhöfer, H. G., & Nüsslein-Volhard, C. (1986). Organization of anterior pattern in the Drosophila embryo by the maternal gene bicoid. *Nature*, *324*(6093), 120–125. https://doi.org/10.1038/324120a0

Fujishima, M. (1988). Conjugation. In H.-D. Görtz (Ed.), *Paramecium* (pp. 70–84). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-73086-3_5

Fukuda, Y., Akematsu, T., Bando, H., & Kato, K. (2022). Snf2 Proteins Are Required to Generate Gamete Pronuclei in Tetrahymena thermophila. *Microorganisms*, *10*(12). https://doi.org/10.3390/microorganisms10122426

Füllgrabe, J., Hajji, N., & Joseph, B. (2010). Cracking the death code: apoptosis-related histone modifications. *Cell Death and Differentiation*, *17*(8), 1238–1243. https://doi.org/10.1038/cdd.2010.58

Furrer, D. I., Swart, E. C., Kraft, M. F., Sandoval, P. Y., & Nowacki, M. (2017). Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA. *Cell Reports*, *20*(2), 505–520. https://doi.org/10.1016/j.celrep.2017.06.050

Fyodorov, D. V., Blower, M. D., Karpen, G. H., & Kadonaga, J. T. (2004). Acf1 confers unique activities to ACF/CHRAC and promotes the formation rather than disruption of chromatin in vivo. *Genes & Development*, *18*(2), 170–183. https://doi.org/10.1101/gad.1139604

Fyodorov, D. V., & Kadonaga, J. T. (2002). Binding of Acf1 to DNA involves a WAC motif and is important for ACF-mediated chromatin assembly. *Molecular and Cellular Biology*, *22*(18), 6344–6353. https://doi.org/10.1128/MCB.22.18.6344-6353.2002

Gade, P., & Kalvakolanu, D. V. (2012). Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity. *Methods in Molecular Biology*, *809*, 85–104. https://doi.org/10.1007/978-1-61779-376-9_6

Galvani, A., & Sperling, L. (2002). RNA interference by feeding in Paramecium. *Trends in Genetics*, *18*(1), 11–12. https://doi.org/10.1016/s0168-9525(01)02548-3

Gehrmann, T., Pelkmans, J. F., Ohm, R. A., Vos, A. M., Sonnenberg, A. S. M., Baars, J. J. P., Wösten, H. A. B., Reinders, M. J. T., & Abeel, T. (2018). Nucleus-specific expression in the multinuclear mushroom-forming fungus Agaricus bisporus reveals different nuclear regulatory programs. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(17), 4429–4434. https://doi.org/10.1073/pnas.1721381115

Giese, A. C. (1938). Cannibalism and gigantism in blepharisma. *Transactions of the American Microscopical Society*, *57*(3), 245. https://doi.org/10.2307/3222693

Gilley, D., Preer, J. R., Aufderheide, K. J., & Polisky, B. (1988). Autonomous replication and addition of telomerelike sequences to DNA microinjected into Paramecium tetraurelia macronuclei. *Molecular and Cellular Biology*, *8*(11), 4765–4772. https://doi.org/10.1128/mcb.8.11.4765-4772.1988

Gkikopoulos, T., Havas, K. M., Dewar, H., & Owen-Hughes, T. (2009). SWI/SNF and Asf1p cooperate to displace histones during induction of the saccharomyces cerevisiae HO promoter. *Molecular and Cellular Biology*, *29*(15), 4057–4066. https://doi.org/10.1128/MCB.00400-09

Gladfelter, A. S. (2006). Nuclear anarchy: asynchronous mitosis in multinucleated fungal hyphae. *Current Opinion in Microbiology*, *9*(6), 547–552. https://doi.org/10.1016/j.mib.2006.09.002

Gnan, S., Matelot, M., Weiman, M., Arnaiz, O., Guérin, F., Sperling, L., Bétermier, M., Thermes, C., Chen, C.-L., & Duharcourt, S. (2022). GC content, but not nucleosome positioning, directly contributes to intron splicing efficiency in Paramecium. *Genome Research*, *32*(4), 699–709. https://doi.org/10.1101/gr.276125.121

Grandchamp, S., & Beisson, J. (1981). Positional control of nuclear differentiation in paramecium. *Developmental Biology*, *81*(2), 336–341. https://doi.org/10.1016/0012-1606(81)90297-9

Gratias, A., & Bétermier, M. (2003). Processing of double-strand breaks is involved in the precise excision of paramecium internal eliminated sequences. *Molecular and Cellular Biology*, *23*(20), 7152–7162. https://doi.org/10.1128/MCB.23.20.7152-7162.2003

Gratias, A., Lepère, G., Garnier, O., Rosa, S., Duharcourt, S., Malinsky, S., Meyer, E., & Bétermier, M. (2008). Developmentally programmed DNA splicing in Paramecium reveals short-distance crosstalk between DNA cleavage sites. *Nucleic Acids Research*, *36*(10), 3244–3251. https://doi.org/10.1093/nar/gkn154

Greider, C. W., & Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell*, *43*(2 Pt 1), 405–413. https://doi.org/10.1016/0092-8674(85)90170-9

Gruchota, J., Denby Wilkes, C., Arnaiz, O., Sperling, L., & Nowak, J. K. (2017). A meiosis-specific Spt5 homolog involved in non-coding transcription. *Nucleic Acids Research*, *45*(8), 4722–4732. https://doi.org/10.1093/nar/gkw1318

Grüne, T., Brzeski, J., Eberharter, A., Clapier, C. R., Corona, D. F. V., Becker, P. B., & Müller, C. W. (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Molecular Cell*, *12*(2), 449–460. https://doi.org/10.1016/s1097-2765(03)00273-9

Guérin, F., Arnaiz, O., Boggetto, N., Denby Wilkes, C., Meyer, E., Sperling, L., & Duharcourt, S. (2017). Flow cytometry sorting of nuclei enables the first global characterization of *Paramecium* germline DNA and transposable elements. *BMC Genomics*, *18*(1), 327. https://doi.org/10.1186/s12864-017-3713-7

Gurova, K., Chang, H.-W., Valieva, M. E., Sandlesh, P., & Studitsky, V. M. (2018). Structure and function of the histone chaperone FACT - Resolving FACTual issues. *Biochimica et Biophysica Acta. Gene Regulatory Mechanisms*. https://doi.org/10.1016/j.bbagrm.2018.07.008

Hamiche, A., Sandaltzopoulos, R., Gdula, D. A., & Wu, C. (1999). ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. *Cell*, *97*(7), 833–842. https://doi.org/10.1016/s0092-8674(00)80796-5

Hamilton, E. P., Kapusta, A., Huvos, P. E., Bidwell, S. L., Zafar, N., Tang, H., Hadjithomas, M., Krishnakumar, V., Badger, J. H., Caler, E. V., Russ, C., Zeng, Q., Fan, L., Levin, J. Z., Shea, T., Young, S. K., Hegarty, R., Daza, R., Gujja, S., … Coyne, R. S. (2016). Structure of the germline genome of *Tetrahymena thermophila* and relationship to the massively rearranged somatic genome. *ELife*, *5*. https://doi.org/10.7554/eLife.19090

Hammond, C. M., Strømme, C. B., Huang, H., Patel, D. J., & Groth, A. (2017). Histone chaperone networks shaping chromatin function. *Nature Reviews. Molecular Cell Biology*, *18*(3), 141–158. https://doi.org/10.1038/nrm.2016.159

Hellman, L. M., & Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols*, *2*(8), 1849–1861. https://doi.org/10.1038/nprot.2007.249

Hondele, M., & Ladurner, A. G. (2013). Catch me if you can: how the histone chaperone FACT capitalizes on nucleosome breathing. *Nucleus (Austin, Tex.)*, *4*(6), 443–449. https://doi.org/10.4161/nucl.27235

Hu, X. (2014). Ciliates in extreme environments. *The Journal of Eukaryotic Microbiology*, *61*(4), 410–418. https://doi.org/10.1111/jeu.12120

Hwang, W. L., Deindl, S., Harada, B. T., & Zhuang, X. (2014). Histone H4 tail mediates allosteric regulation of nucleosome remodelling by linker DNA. *Nature*, *512*(7513), 213–217. https://doi.org/10.1038/nature13380

Ignarski, M., Singh, A., Swart, E. C., Arambasic, M., Sandoval, P. Y., & Nowacki, M. (2014). Paramecium tetraurelia chromatin assembly factor-1-like protein PtCAF-1 is involved in RNA-mediated control of DNA elimination. *Nucleic Acids Research*, *42*(19), 11952–11964. https://doi.org/10.1093/nar/gku874

Inaba, F., Imamoto, K., & Suganuma, Y. (1966). Electron-microscopic Observations on Nuclear Exchange during Conjugation in Paramecium multimicronucleatum. *Proceedings of the Japan Academy*, *42*(4), 394–398. https://doi.org/10.2183/pjab1945.42.394

Ito, T., Bulger, M., Pazin, M. J., Kobayashi, R., & Kadonaga, J. T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell*, *90*(1), 145–155. https://doi.org/10.1016/s0092-8674(00)80321-9

Ito, T., Levenstein, M. E., Fyodorov, D. V., Kutach, A. K., Kobayashi, R., & Kadonaga, J. T. (1999). ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes & Development*, *13*(12), 1529–1539. https://doi.org/10.1101/gad.13.12.1529

Iwamoto, M., Hiraoka, Y., & Haraguchi, T. (2015). The nuclear pore complex acts as a master switch for nuclear and cell differentiation. *Communicative & Integrative Biology*, *8*(4), e1056950. https://doi.org/10.1080/19420889.2015.1056950

Iwamoto, M., Mori, C., Kojidani, T., Bunai, F., Hori, T., Fukagawa, T., Hiraoka, Y., & Haraguchi, T. (2009). Two distinct repeat sequences of Nup98 nucleoporins characterize dual nuclei in the binucleated ciliate tetrahymena. *Current Biology*, *19*(10), 843–847. https://doi.org/10.1016/j.cub.2009.03.055

Jiang, J., Rushlow, C. A., Zhou, Q., Small, S., & Levine, M. (1992). Individual dorsal morphogen binding sites mediate activation and repression in the Drosophila embryo. *The EMBO Journal*, *11*(8), 3147–3154. https://doi.org/10.1002/j.1460-2075.1992.tb05387.x

Jones, M. J. K., & Jallepalli, P. V. (2012). Chromothripsis: chromosomes in crisis. *Developmental Cell*, *23*(5), 908–917. https://doi.org/10.1016/j.devcel.2012.10.010

Kapusta, A., Matsuda, A., Marmignon, A., Ku, M., Silve, A., Meyer, E., Forney, J. D., Malinsky, S., & Bétermier, M. (2011). Highly precise and developmentally programmed genome assembly in Paramecium requires ligase IV-dependent end joining. *PLoS Genetics*, *7*(4), e1002049. https://doi.org/10.1371/journal.pgen.1002049

Kayne, P. S., Kim, U. J., Han, M., Mullen, J. R., Yoshizaki, F., & Grunstein, M. (1988). Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. *Cell*, *55*(1), 27–39. https://doi.org/10.1016/0092-8674(88)90006-2

Kim, A., & Pyykko, I. (2011). Size matters: versatile use of PiggyBac transposons as a genetic manipulation tool. *Molecular and Cellular Biochemistry*, *354*(1–2), 301–309. https://doi.org/10.1007/s11010-011-0832-3

Klobutcher, L. A., & Herrick, G. (1995). Consensus inverted terminal repeat sequence of *Paramecium* IESs: resemblance to termini of Tc1-related and *Euplotes* Tec transposons. *Nucleic Acids Research*, *23*(11), 2006–2013. https://doi.org/10.1093/nar/23.11.2006

Kokkoris, V., Stefani, F., Dalpé, Y., Dettman, J., & Corradi, N. (2020). Nuclear dynamics in the arbuscular mycorrhizal fungi. *Trends in Plant Science*, *25*(8), 765–778. https://doi.org/10.1016/j.tplants.2020.05.002

Kong, N. R., Chai, L., Tenen, D. G., & Bassal, M. A. (2021). A modified CUT&RUN protocol and analysis pipeline to identify transcription factor binding sites in human cell lines. *STAR Protocols*, *2*(3), 100750. https://doi.org/10.1016/j.xpro.2021.100750

Koukalová, B., Kovarík, A., Fajkus, J., & Siroký, J. (1997). Chromatin fragmentation associated with apoptotic changes in tobacco cells exposed to cold stress. *FEBS Letters*, *414*(2), 289–292. https://doi.org/10.1016/s0014-5793(97)01008-9

Le Mouël, A., Butler, A., Caron, F., & Meyer, E. (2003). Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia. *Eukaryotic Cell*, *2*(5), 1076–1090. https://doi.org/10.1128/EC.2.5.1076-1090.2003

Lee, H. O., Davidson, J. M., & Duronio, R. J. (2009). Endoreplication: polyploidy with purpose. *Genes & Development*, *23*(21), 2461–2477. https://doi.org/10.1101/gad.1829209

Lepère, G., Nowacki, M., Serrano, V., Gout, J.-F., Guglielmi, G., Duharcourt, S., & Meyer, E. (2009). Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Research*, *37*(3), 903–915. https://doi.org/10.1093/nar/gkn1018

Lhuillier-Akakpo, M., Frapporti, A., Denby Wilkes, C., Matelot, M., Vervoort, M., Sperling, L., & Duharcourt, S. (2014). Local effect of enhancer of zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements. *PLoS Genetics*, *10*(9), e1004665. https://doi.org/10.1371/journal.pgen.1004665

Lhuillier-Akakpo, M., Guérin, F., Frapporti, A., & Duharcourt, S. (2016). DNA deletion as a mechanism for developmentally programmed centromere loss. *Nucleic Acids Research*, *44*(4), 1553–1565. https://doi.org/10.1093/nar/gkv1110

Li, G., Liu, S., Wang, J., He, J., Huang, H., Zhang, Y., & Xu, L. (2014). ISWI proteins participate in the genome-wide nucleosome distribution in Arabidopsis. *The Plant Journal*, *78*(4), 706–714. https://doi.org/10.1111/tpj.12499

Li, M., Bastos Gomes, G., Zhao, W., Hu, G., Huang, K., Yoshinaga, T., Clark, T. G., Li, W., Zou, H., Wu, S., & Wang, G. (2023). Cultivation of fish ciliate parasites: Progress and prospects. *Reviews in Aquaculture*, *15*(1), 142–162. https://doi.org/10.1111/raq.12708

Lu, M. W., Beh, L. Y., Yerlici, V. T., Fang, W., Kulej, K., Garcia, B. A., & Landweber, L. F. (2023). Exploration of the Nuclear Proteomes in the Ciliate Oxytricha trifallax. *Microorganisms*, *11*(2). https://doi.org/10.3390/microorganisms11020343

Lu, X., Meng, X., Morris, C. A., & Keating, M. T. (1998). A novel human gene, WSTF, is deleted in Williams syndrome. *Genomics*, *54*(2), 241–249. https://doi.org/10.1006/geno.1998.5578

Ludwigsen, J., Pfennig, S., Singh, A. K., Schindler, C., Harrer, N., Forné, I., Zacharias, M., & Mueller-Planitz, F. (2017). Concerted regulation of ISWI by an autoinhibitory domain and the H4 N-terminal tail. *ELife*, *6*. https://doi.org/10.7554/eLife.21477

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Maliszewska-Olejniczak, K., Gruchota, J., Gromadka, R., Denby Wilkes, C., Arnaiz, O., Mathy, N., Duharcourt, S., Bétermier, M., & Nowak, J. K. (2015). TFIIS-Dependent Non-coding Transcription Regulates Developmental Genome Rearrangements. *PLoS Genetics*, *11*(7), e1005383. https://doi.org/10.1371/journal.pgen.1005383

Malone, C. D., Anderson, A. M., Motl, J. A., Rexer, C. H., & Chalker, D. L. (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*(20), 9151–9164. https://doi.org/10.1128/MCB.25.20.9151-9164.2005

Malone, C. D., Falkowska, K. A., Li, A. Y., Galanti, S. E., Kanuru, R. C., LaMont, E. G., Mazzarella, K. C., Micev, A. J., Osman, M. M., Piotrowski, N. K., Suszko, J. W., Timm, A. C., Xu, M.-M., Liu, L., & Chalker, D. L. (2008). Nucleus-specific importin alpha proteins and nucleoporins regulate protein import and nuclear division in the binucleate Tetrahymena thermophila. *Eukaryotic Cell*, *7*(9), 1487–1499. https://doi.org/10.1128/EC.00193-08

Mani, R.-S., & Chinnaiyan, A. M. (2010). Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. *Nature Reviews. Genetics*, *11*(12), 819–829. https://doi.org/10.1038/nrg2883

Marker, S., Le Mouël, A., Meyer, E., & Simon, M. (2010). Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in Paramecium tetraurelia. *Nucleic Acids Research*, *38*(12), 4092–4107. https://doi.org/10.1093/nar/gkq131

Marmignon, A., Bischerour, J., Silve, A., Fojcik, C., Dubois, E., Arnaiz, O., Kapusta, A., Malinsky, S., & Bétermier, M. (2014). Ku-mediated coupling of DNA cleavage and repair during programmed genome rearrangements in the ciliate Paramecium

tetraurelia. *PLoS Genetics*, *10*(8), e1004552. https://doi.org/10.1371/journal.pgen.1004552

Marshall, W. F. (2021). Regeneration in Stentor coeruleus. *Frontiers in Cell and Developmental Biology*, *9*, 753625. https://doi.org/10.3389/fcell.2021.753625

Mayer, K. M., & Forney, J. D. (1999). A mutation in the flanking 5'-TA-3' dinucleotide prevents excision of an internal eliminated sequence from the Paramecium tetraurelia genome. *Genetics*, *151*(2), 597–604. https://doi.org/10.1093/genetics/151.2.597

Mellor, J., & Morillon, A. (2004). ISWI complexes in Saccharomyces cerevisiae. *Biochimica et Biophysica Acta*, *1677*(1–3), 100–112. https://doi.org/10.1016/j.bbaexp.2003.10.014

Miró-Pina, C., Charmant, O., Kawaguchi, T., Holoch, D., Michaud, A., Cohen, I., Humbert, A., Jaszczyszyn, Y., Chevreux, G., Del Maestro, L., Ait-Si-Ali, S., Arnaiz, O., Margueron, R., & Duharcourt, S. (2022). Paramecium Polycomb repressive complex 2 physically interacts with the small RNA-binding PIWI protein to repress transposable elements. *Developmental Cell*, *57*(8), 1037-1052.e8. https://doi.org/10.1016/j.devcel.2022.03.014

Mochizuki, K., Fine, N. A., Fujisawa, T., & Gorovsky, M. A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell*, *110*(6), 689–699. https://doi.org/10.1016/s0092-8674(02)00909-1

Morisato, D., & Anderson, K. V. (1995). Signaling pathways that establish the dorsal-ventral pattern of the Drosophila embryo. *Annual Review of Genetics*, *29*, 371–399. https://doi.org/10.1146/annurev.ge.29.120195.002103

Mutazono, M., Noto, T., & Mochizuki, K. (2019). Diversification of small RNA amplification mechanisms for targeting transposon-related sequences in ciliates. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(29), 14639–14644. https://doi.org/10.1073/pnas.1903491116

Nalabothula, N., McVicker, G., Maiorano, J., Martin, R., Pritchard, J. K., & Fondufe-Mittendorf, Y. N. (2014). The chromatin architectural proteins HMGD1 and H1 bind reciprocally and have opposite effects on chromatin structure and gene regulation. *BMC Genomics*, *15*, 92. https://doi.org/10.1186/1471-2164-15-92

Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T. G., & Landweber, L. F. (2008). RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature*, *451*(7175), 153–158. https://doi.org/10.1038/nature06452

Nowacki, M., Zagorski-Ostoja, W., & Meyer, E. (2005). Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in Paramecium tetraurelia. *Current Biology*, *15*(18), 1616–1628. https://doi.org/10.1016/j.cub.2005.07.033

Orphanides, G., LeRoy, G., Chang, C. H., Luse, D. S., & Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell*, *92*(1), 105–116. https://doi.org/10.1016/s0092-8674(00)80903-4

Philpott, A., Krude, T., & Laskey, R. A. (2000). Nuclear chaperones. *Seminars in Cell & Developmental Biology*, *11*(1), 7–14. https://doi.org/10.1006/scdb.1999.0346

Pierce, R. W., & Turner, J. T. (1992). Ecology of planktonic ciliates in marine food webs. *Rev. Aquat. Sci*, *6*(2), 139–181.

Piersimoni, L., Kastritis, P. L., Arlt, C., & Sinz, A. (2022). Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein-Protein Interactions─A Method for All Seasons. *Chemical Reviews*, *122*(8), 7500–7531. https://doi.org/10.1021/acs.chemrev.1c00786

Plattner, H., & Verkhratsky, A. (2018). The remembrance of the things past: Conserved signalling pathways link protozoa to mammalian nervous system. *Cell Calcium*, *73*, 25–39. https://doi.org/10.1016/j.ceca.2018.04.001

Plattner, H. (2022). Ciliate research: From myth to trendsetting science. *The Journal of Eukaryotic Microbiology*, *69*(5), e12926. https://doi.org/10.1111/jeu.12926

Prescott, D. M. (1994). The DNA of ciliated protozoa. *Microbiological Reviews*, *58*(2), 233–267. https://doi.org/10.1128/mr.58.2.233-267.1994

Raikov, I. B. (1985). Primitive never-dividing macronuclei of some lower ciliates. *International Review of Cytology*, *95*, 267–325. https://doi.org/10.1016/s0074-7696(08)60584-7

Rogers, K. W., & Schier, A. F. (2011). Morphogen gradients: from generation to interpretation. *Annual Review of Cell and Developmental Biology*, *27*, 377–407. https://doi.org/10.1146/annurev-cellbio-092910-154148

Rooney, S., Chaudhuri, J., & Alt, F. W. (2004). The role of the non-homologous end-joining pathway in lymphocyte development. *Immunological Reviews*, *200*, 115–131. https://doi.org/10.1111/j.0105-2896.2004.00165.x

Roth, S., Stein, D., & Nüsslein-Volhard, C. (1989). A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the Drosophila embryo. *Cell*, *59*(6), 1189–1202. https://doi.org/10.1016/0092-8674(89)90774-5

Rzeszutek, I., Maurer-Alcalá, X. X., & Nowacki, M. (2020). Programmed genome rearrangements in ciliates. *Cellular and Molecular Life Sciences*, *77*(22), 4615–4629. https://doi.org/10.1007/s00018-020-03555-2

Sanchez, R., & Zhou, M.-M. (2011). The PHD finger: a versatile epigenome reader. *Trends in Biochemical Sciences*, *36*(7), 364–372. https://doi.org/10.1016/j.tibs.2011.03.005

Sandoval, P. Y., Swart, E. C., Arambasic, M., & Nowacki, M. (2014). Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting. *Developmental Cell*, *28*(2), 174–188. https://doi.org/10.1016/j.devcel.2013.12.010

Schatz, D. G., & Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. *Nature Reviews. Immunology*, *11*(4), 251–263. https://doi.org/10.1038/nri2941

Schwartz, V. V. (1978). Struktur und Entwicklung des Makronucleus von *Paramecium bursaria*. *Archiv Für Protistenkunde*, *120*(3), 255–277. https://doi.org/10.1016/S0003-9365(78)80002-5

Seah, Brandon Kwee Boon, & Swart, E. C. (2023). When cleaning facilitates cluttering - genome editing in ciliates. *Trends in Genetics*, *39*(5), P344-346. https://doi.org/10.1016/j.tig.2023.02.016

Seah, Brandon K B, Singh, M., Emmerich, C., Singh, A., Woehle, C., Huettel, B., Byerly, A., Stover, N. A., Sugiura, M., Harumoto, T., & Swart, E. C. (2023). MITE infestation accommodated by genome editing in the germline genome of the ciliate Blepharisma. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(4), e2213985120. https://doi.org/10.1073/pnas.2213985120

Sellis, D., Guérin, F., Arnaiz, O., Pett, W., Lerat, E., Boggetto, N., Krenek, S., Berendonk, T., Couloux, A., Aury, J.-M., Labadie, K., Malinsky, S., Bhullar, S., Meyer, E., Sperling, L., Duret, L., & Duharcourt, S. (2021). Massive colonization of protein-coding exons by selfish genetic elements in *Paramecium* germline genomes. *PLoS Biology*, *19*(7), e3001309. https://doi.org/10.1371/journal.pbio.3001309

Sharif, S. B., Zamani, N., & Chadwick, B. P. (2021). BAZ1B the protean protein. *Genes*, *12*(10). https://doi.org/10.3390/genes12101541

Sheng, Y., Duan, L., Cheng, T., Qiao, Y., Stover, N. A., & Gao, S. (2020). The completed macronuclear genome of a model ciliate *Tetrahymena thermophila* and its application in genome scrambling and copy number analyses. *Science China. Life Sciences*, *63*(10), 1534–1542. https://doi.org/10.1007/s11427-020-1689-4

Singh, M., Seah, B. K. B., Emmerich, C., Singh, A., Woehle, C., Huettel, B., Byerly, A., Stover, N. A., Sugiura, M., Harumoto, T., & Swart, E. C. (2021). Genome editing excisase origins illuminated by somatic genome of *Blepharisma*. *BioRxiv*. https://doi.org/10.1101/2021.12.14.471607

Skene, P. J., Henikoff, J. G., & Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature Protocols*, *13*(5), 1006–1019. https://doi.org/10.1038/nprot.2018.015

Soh, A. W. J., & Pearson, C. G. (2022). Ciliate cortical organization and dynamics for cell motility: Comparing ciliates and vertebrates. *The Journal of Eukaryotic Microbiology*, *69*(5), e12880. https://doi.org/10.1111/jeu.12880

Sonneborn, T. M. (1938). Mating types in Paramecium aurelia: diverse conditions for mating in different stocks; occurrence, number and interrelations of the types. *Proceedings of the American Philosophical Society*, 411–434.

Sonneborn, T. M. (1954). The Relation of Autoǵamy to Senescence and Rejuvenescence in Paramecium aurelia*. *The Journal of Protozoology*, *1*(1), 38–53. https://doi.org/10.1111/j.1550-7408.1954.tb00792.x

Spirov, A., Fahmy, K., Schneider, M., Frei, E., Noll, M., & Baumgartner, S. (2009). Formation of the bicoid morphogen gradient: an mRNA gradient dictates the protein gradient. *Development*, *136*(4), 605–614. https://doi.org/10.1242/dev.031195

Stebbins, G. L. (1947). Types of polyploids: their classification and significance. In *Advances in Genetics* (Vol. 1, pp. 403–429). Elsevier. https://doi.org/10.1016/S0065-2660(08)60490-3

Steele, C. J., Barkocy-Gallagher, G. A., Preer, L. B., & Preer, J. R. (1994). Developmentally excised sequences in micronuclear DNA of Paramecium. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(6), 2255–2259. https://doi.org/10.1073/pnas.91.6.2255

Stein, V., Blank-Landeshammer, B., Müntjes, K., Märker, R., Teichert, I., Feldbrügge, M., Sickmann, A., & Kück, U. (2020). The STRIPAK signaling complex regulates dephosphorylation of GUL1, an RNA-binding protein that shuttles on endosomes. *PLoS Genetics*, *16*(9), e1008819. https://doi.org/10.1371/journal.pgen.1008819

Strahl, B. D., Ohba, R., Cook, R. G., & Allis, C. D. (1999). Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(26), 14967–14972. https://doi.org/10.1073/pnas.96.26.14967

Struhl, G., Struhl, K., & Macdonald, P. M. (1989). The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell*, *57*(7), 1259–1273. https://doi.org/10.1016/0092-8674(89)90062-7

Sun, F. L., Cuaycong, M. H., & Elgin, S. C. (2001). Long-range nucleosome ordering is associated with gene silencing in Drosophila melanogaster pericentric heterochromatin. *Molecular and Cellular Biology*, *21*(8), 2867–2879. https://doi.org/10.1128/MCB.21.8.2867-2879.2001

Swart, E. C., Bracht, J. R., Magrini, V., Minx, P., Chen, X., Zhou, Y., Khurana, J. S., Goldman, A. D., Nowacki, M., Schotanus, K., Jung, S., Fulton, R. S., Ly, A., McGrath, S., Haub, K., Wiggins, J. L., Storton, D., Matese, J. C., Parsons, L., … Landweber, L. F. (2013). The *Oxytricha trifallax* macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. *PLoS Biology*, *11*(1), e1001473. https://doi.org/10.1371/journal.pbio.1001473

Swart, E. C., Wilkes, C. D., Sandoval, P. Y., Arambasic, M., Sperling, L., & Nowacki, M. (2014). Genome-wide analysis of genetic and epigenetic control of programmed DNA deletion. *Nucleic Acids Research*, *42*(14), 8970–8983. https://doi.org/10.1093/nar/gku619

Tan, L.-M., Liu, R., Gu, B.-W., Zhang, C.-J., Luo, J., Guo, J., Wang, Y., Chen, L., Du, X., Li, S., Shao, C.-R., Su, Y.-N., Cai, X.-W., Lin, R.-N., Li, L., Chen, S., Du, J., & He, X.-J. (2020). Dual recognition of h3k4me3 and DNA by the ISWI component ARID5 regulates the floral transition in arabidopsis. *The Plant Cell*, *32*(7), 2178–2195. https://doi.org/10.1105/tpc.19.00944

Tartar, V. (1961). *The Biology of Stentor* (Vol. 5). Pergammon Press.

Taverna, S. D., Coyne, R. S., & Allis, C. D. (2002). Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. *Cell*, *110*(6), 701–711. https://doi.org/10.1016/s0092-8674(02)00941-8

Taverna, S. D., Ilin, S., Rogers, R. S., Tanny, J. C., Lavender, H., Li, H., Baker, L., Boyle, J., Blair, L. P., Chait, B. T., Patel, D. J., Aitchison, J. D., Tackett, A. J., & Allis, C. D. (2006). Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Molecular Cell*, *24*(5), 785–796. https://doi.org/10.1016/j.molcel.2006.10.026

Treasurer, J. W. (2002). A review of potential pathogens of sea lice and the application of cleaner fish in biological control. *Pest Management Science*, *58*(6), 546–558. https://doi.org/10.1002/ps.509

Tsukiyama, T., & Wu, C. (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell*, *83*(6), 1011–1020. https://doi.org/10.1016/0092-8674(95)90216-3

Tucker, J. B., Beisson, J., Roche, D. L., & Cohen, J. (1980). Microtubules and control of macronuclear “amitosis” in *Paramecium*. *Journal of Cell Science*, *44*, 135–151. https://doi.org/10.1242/jcs.44.1.135

Valieva, M. E., Armeev, G. A., Kudryashova, K. S., Gerasimova, N. S., Shaytan, A. K., Kulaeva, O. I., McCullough, L. L., Formosa, T., Georgiev, P. G., Kirpichnikov, M. P., Studitsky, V. M., & Feofanov, A. V. (2016). Large-scale ATP-independent nucleosome unfolding by a histone chaperone. *Nature Structural & Molecular Biology*, *23*(12), 1111–1116. https://doi.org/10.1038/nsmb.3321

Van Houten, J. (2019). Paramecium Biology. *Results and Problems in Cell Differentiation*, *68*, 291–318. https://doi.org/10.1007/978-3-030-23459-1_13

Varga-Weisz, P. D., Wilm, M., Bonte, E., Dumas, K., Mann, M., & Becker, P. B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature*, *388*(6642), 598–602. https://doi.org/10.1038/41587

Wang, C., Lv, L., Solberg, T., Wen, Z., Zhang, H., & Gao, F. (2023). Conservation of the ancestral function of GTSF1 in transposon silencing in the unicellular eukaryote *Paramecium tetraurelia*. *BioRxiv*. https://doi.org/10.1101/2023.10.06.561219

Wang, C., Solberg, T., Maurer-Alcalá, X. X., Swart, E. C., Gao, F., & Nowacki, M. (2022). A small RNA-guided PRC2 complex eliminates DNA as an extreme form of transposon silencing. *Cell Reports*, *40*(8), 111263. https://doi.org/10.1016/j.celrep.2022.111263

Wang, Z., Chi, Y., Li, T., Song, W., Wang, Y., Wu, T., Zhang, G., Liu, Y., Ma, H., Song, W., Al-Rasheid, K. A. S., Warren, A., & Lu, B. (2022). Biodiversity of freshwater ciliates (Protista, Ciliophora) in the Lake Weishan Wetland, China: the state of the art. *Marine Life Science & Technology*, *4*(4), 429–451. https://doi.org/10.1007/s42995-022-00154-x

Weinmann, A. S., & Farnham, P. J. (2002). Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods*, *26*(1), 37–47. https://doi.org/10.1016/S1046-2023(02)00006-3

Weisse, T. (2017). Functional diversity of aquatic ciliates. *European Journal of Protistology*, *61*(Pt B), 331–358. https://doi.org/10.1016/j.ejop.2017.04.001

Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M., Tackett, A. J., Chait, B. T., Badenhorst, P., Wu, C., & Allis, C. D. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, *442*(7098), 86–90. https://doi.org/10.1038/nature04815

Xu, H., Zhang, W., Jiang, Y., & Yang, E. J. (2014). Use of biofilm-dwelling ciliate communities to determine environmental quality status of coastal waters. *The Science of the Total Environment*, *470–471*, 511–518. https://doi.org/10.1016/j.scitotenv.2013.10.025

Yadegari, R., & Drews, G. N. (2004). Female gametophyte development. *The Plant Cell*, *16 Suppl*(Suppl), S133-41. https://doi.org/10.1105/tpc.018192

Yamada, K., Frouws, T. D., Angst, B., Fitzgerald, D. J., DeLuca, C., Schimmele, K., Sargent, D. F., & Richmond, T. J. (2011). Structure and mechanism of the chromatin remodelling factor ISW1a. *Nature*, *472*(7344), 448–453. https://doi.org/10.1038/nature09947

Yanagi, A, & Hiwatashi, K. (1985). Intracellular positional control of survival or degeneration of nuclei during conjugation in Paramecium caudatum. *Journal of Cell Science*, *79*, 237–246. https://doi.org/10.1242/jcs.79.1.237

Yanagi, Akira. (1987). Positional control of the fates of nuclei produced after meiosis in Paramecium caudatum: Analysis by nuclear transplantation. *Developmental Biology*, *122*(2), 535–539. https://doi.org/10.1016/0012-1606(87)90317-4

Yang, J. G., Madrid, T. S., Sevastopoulos, E., & Narlikar, G. J. (2006). The chromatin-remodeling enzyme ACF is an ATP-dependent DNA length sensor that regulates nucleosome spacing. *Nature Structural & Molecular Biology*, *13*(12), 1078–1083. https://doi.org/10.1038/nsmb1170

Yan, L., Wang, L., Tian, Y., Xia, X., & Chen, Z. (2016). Structure and regulation of the chromatin remodeller ISWI. *Nature*, *540*(7633), 466–469. https://doi.org/10.1038/nature20590

Yan, L., Wu, H., Li, X., Gao, N., & Chen, Z. (2019). Structures of the ISWI-nucleosome complex reveal a conserved mechanism of chromatin remodeling. *Nature Structural & Molecular Biology*, *26*(4), 258–266. https://doi.org/10.1038/s41594-019-0199-9

Yasuhara, N., Shibazaki, N., Tanaka, S., Nagai, M., Kamikawa, Y., Oe, S., Asally, M., Kamachi, Y., Kondoh, H., & Yoneda, Y. (2007). Triggering neural differentiation of ES cells by subtype switching of importin-alpha. *Nature Cell Biology*, *9*(1), 72–79. https://doi.org/10.1038/ncb1521

Yasuhara, N., Yamagishi, R., Arai, Y., Mehmood, R., Kimoto, C., Fujita, T., Touma, K., Kaneko, A., Kamikawa, Y., Moriyama, T., Yanagida, T., Kaneko, H., & Yoneda, Y. (2013). Importin alpha subtypes determine differential transcription factor

localization in embryonic stem cells maintenance. *Developmental Cell*, *26*(2), 123–135. https://doi.org/10.1016/j.devcel.2013.06.022

Yerlici, V. T., & Landweber, L. F. (2014). Programmed genome rearrangements in the ciliate oxytricha. *Microbiology Spectrum*, *2*(6). https://doi.org/10.1128/microbiolspec.MDNA3-0025-2014

Zangarelli, C., Arnaiz, O., Bourge, M., Gorrichon, K., Jaszczyszyn, Y., Mathy, N., Escoriza, L., Bétermier, M., & Régnier, V. (2022). Developmental timing of programmed DNA elimination in Paramecium tetraurelia recapitulates germline transposon evolutionary dynamics. *Genome Research*, *32*(11–12), 2028–2042. https://doi.org/10.1101/gr.277027.122

Zhang, C.-Z., Leibowitz, M. L., & Pellman, D. (2013). Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes & Development*, *27*(23), 2513–2530. https://doi.org/10.1101/gad.229559.113

Appendix

I. Chromatin remodeling is required for sRNA-guided DNA elimination in *Paramecium*

II. ISWI1 complex proteins facilitate developmental genome editing in *Paramecium*

III. Two paralogous PHD finger proteins participate in *Paramecium tetraurelia*'s natural genome editing

Article

SOURCE DATA | TRANSPARENT PROCESS | OPEN ACCESS

THE EMBO JOURNAL

Chromatin remodeling is required for sRNA-guided DNA elimination in *Paramecium*

Aditi Singh[1,2,3], Xyrus X Maurer-Alcalá[1], Therese Solberg[1,2], Lilia Häußermann[3], Silvan Gisler[1], Michael Ignarski[1], Estienne C Swart[1,3,†] & Mariusz Nowacki[1,*]

Abstract

Small RNAs mediate the silencing of transposable elements and other genomic loci, increasing nucleosome density and preventing undesirable gene expression. The unicellular ciliate *Paramecium* is a model to study dynamic genome organization in eukaryotic cells, given its unique feature of nuclear dimorphism. Here, the formation of the somatic macronucleus during sexual reproduction requires eliminating thousands of transposon remnants (IESs) and transposable elements scattered throughout the germline micronuclear genome. The elimination process is guided by Piwi-associated small RNAs and leads to precise cleavage at IES boundaries. Here we show that IES recognition and precise excision are facilitated by recruiting ISWI1, a *Paramecium* homolog of the chromatin remodeler ISWI. ISWI1 knockdown substantially inhibits DNA elimination, quantitatively similar to development-specific sRNA gene knockdowns but with much greater aberrant IES excision at alternative boundaries. We also identify key development-specific sRNA biogenesis and transport proteins, Ptiwi01 and Ptiwi09, as ISWI1 cofactors in our co-immunoprecipitation studies. Nucleosome profiling indicates that increased nucleosome density correlates with the requirement for ISWI1 and other proteins necessary for IES excision. We propose that chromatin remodeling together with small RNAs is essential for efficient and precise DNA elimination in *Paramecium*.

Keywords chromatin remodeler; genome editing; nucleosomes; small RNAs; transposable elements

Subject Categories Chromatin, Transcription & Genomics; Microbiology, Virology & Host Pathogen Interaction

DOI 10.15252/embj.2022111839 | Received 8 June 2022 | Revised 14 September 2022 | Accepted 19 September 2022 | Published online 11 October 2022

The EMBO Journal (2022) 41: e111839

Introduction

Ciliates, such as *Paramecium tetraurelia* (class Oligohymenophora), provide excellent model systems to understand the dynamic genome organization in eukaryotic cells due to their unique feature of nuclear dimorphism. The formation of *Paramecium*'s somatic nucleus during sexual reproduction involves DNA endoreplication, DNA elimination, DNA repair, and transcription of genes that are specifically expressed when these processes occur (Chalker & Yao, 2011). Hence, the chromatin needs to be in a tightly controlled dynamic state. The germline micronuclear (MIC) genome contains regions that are removed during the development of the somatic macronuclear (MAC) genome (Beisson *et al*, 2010a) in a sophisticated process of genome reorganization, a natural form of genome editing. During this event, about 45,000 unique, noncoding Internal Eliminated Sequences (IES) are typically precisely excised (Arnaiz *et al*, 2012).

IES elimination is carried out by a catalytically active domesticated transposase PiggyMac (*PGM*; Baudry *et al*, 2009) in concert with catalytically inactive PGM homologs (Bischerour *et al*, 2018). Precise elimination of IESs is crucial for forming a functional somatic genome since these sequences would otherwise frequently interrupt exonic coding sequences. IESs have a distinctive, periodic size distribution and a weak end consensus sequence that probably reflects the preferences of the excision machinery (Baudry *et al*, 2009; Swart *et al*, 2014). However, the presence of consensus sequences is not enough for precise IES excision (Duret *et al*, 2008).

Currently, the proposed model for *Paramecium*'s IES excision involves two classes of small RNAs (scnRNAs and iesRNAs; Lepère *et al*, 2009; Sandoval *et al*, 2014b) that guide the process via indirectly comparing the maternal genome to the developing genome. These sRNAs are produced by Dicer-like proteins (Dcl2/3 and Dcl5, respectively; Lepère *et al*, 2009; Sandoval *et al*, 2014b) and Piwi proteins (Ptiwi01/09 and Ptiwi10/11, respectively; Bouhouche *et al*, 2011; Furrer *et al*, 2017b). However, as judged by the effects of gene knockdowns, most IESs in *P. tetraurelia* are efficiently excised independently of scnRNAs and iesRNAs (Sandoval *et al*, 2014b; Swart *et al*, 2017a). Other proteins also cooperate in IES excision, with substantial differences in the effects of knockdowns of their genes, suggesting it is far more complicated than can be explained by a single linear pathway (Nowacki *et al*, 2005; Kapusta *et al*, 2011; Dubois *et al*, 2012; Sandoval *et al*, 2014b; Maliszewska-Olejniczak *et al*, 2015; Swart *et al*, 2017a; Abello *et al*, 2020).

1 Institute of Cell Biology, University of Bern, Bern, Switzerland
2 Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland
3 Max Planck Institute for Biology, Tubingen, Germany
*Corresponding author. Tel: +41 31 684 46 54; Fax: +41 31 684 44 31; E-mail: mariusz.nowacki@unibe.ch
†This author contributed equally to this work as senior author

Based on research in the ciliate *Tetrahymena* (class Oligohymenophora), one proposal for *PGM* recruitment and IES elimination suggests histone modifications mark IES boundaries, recruiting PGM for IES excision (Liu *et al*, 2007). Indeed, alteration of histone modifications, specifically H3K27me3 and H3K9me3, is associated with knockdowns of EZL1 and PTCAF1, affecting the excision of most IESs in *Paramecium* (Ignarski *et al*, 2014; Lhuillier-Akakpo *et al*, 2014b). In addition, both these modifications are scnRNA-dependent in *Paramecium* (Ignarski *et al*, 2014).

Nevertheless, fundamental differences exist between *Tetrahymena* and *Paramecium* IESs. Firstly, IES excision is predominantly precise in *Paramecium* and imprecise in *Tetrahymena* (Arnaiz *et al*, 2012; Coyne *et al*, 2012; Hamilton *et al*, 2016). Secondly, in contrast to *Tetrahymena*, the majority of IESs in *Paramecium* are scattered throughout the coding regions. Thirdly, the majority of *Paramecium* IESs are much shorter (median ~ 50 bp) than the size of a nucleosome (~146 bp; Arnaiz *et al*, 2012) or linker regions between MAC nucleosomes (several base pairs; Gnan *et al*, 2022). *Tetrahymena* IESs are much longer (hundreds of bp to kbp; Hamilton *et al*, 2016). Thus *Paramecium* DNA would often be expected to be wrapped around nucleosomes, making it difficult to access IESs for excision. Therefore, there is no particular expectation that the model for *Tetrahymena*, proposing the formation of heterochromatic DNA is necessary for IES excision, is applicable to the excision of most *Paramecium* IESs.

DNA elimination, carried out by *Paramecium*'s PGM, requires IES boundary accessibility. One way to do so would be through the action of ATP-dependent remodelers, such as SNF2-related proteins, that can restructure the chromatin providing access to DNA (Sadeh & Allis, 2011; Rando & Winston, 2012). Suggesting such activity, in *Tetrahymena*, an SNF2/brahma-related gene, TtBRG1 is known to be essential for nuclear development during conjugation (Fillingham *et al*, 2006). Numerous homologs of SNF2-related genes are conserved in *Paramecium tetraurelia* as well. Among these are homologs of ISWI, an SNF2-related, ATP-dependent chromatin remodeler (Pazin & Kadonaga, 1997). ISWI proteins form different complexes interacting with several conserved domains, with each complex modulating a discrete function (Dirscherl & Krebs, 2004). Although ISWI complexes have distinct functions, the general mechanism underlying their various roles is based on altering nucleosome spacing. By moving around nucleosomes, ISWI proteins help DNA-binding proteins access previously unavailable sites (Clapier & Cairns, 2009). To the best of our knowledge, there is currently no information regarding how nucleosomal positioning influences ciliate DNA excision. To this end, we studied the putative role of *Paramecium* ISWI, an SNF2-related protein (Pazin & Kadonaga, 1997), and its influence on both nucleosomes and DNA excision.

Results

We identified ISWI1 in a preliminary RNAi screening of genes with differential upregulation of expression during an autogamy (self-fertilization) time course (Arnaiz *et al*, 2010; Arnaiz & Sperling, 2011). *Paramecium tetraurelia* has five putative ISWI homologs with the characteristic SWI/SNF family ATPase core domain as well as SANT and SLIDE domains towards their C-termini (Fig 1A). Out of these, two pairs of paralogs arose from the well-characterized whole genome duplication (WGD, Fig 1B) events in *Paramecium* (Aury *et al*, 2006). Among these, the gene of the homolog characterized here, *ISWI1*, shows substantial differential upregulation during the macronuclear development, peaking during fragmentation and Dev1 stages of the autogamy time course (Fig 1C). In contrast, the other three paralogs; *ISWI2*, *ISWI3*, and *ISWI4* tend to be repressed during autogamy. The remaining *ISWI* homolog, *ISWI5*, also shows substantial differential expression, peaking during meiosis and fragmentation of the parental MAC before decreasing in abundance for the remainder of development (Fig 1C).

Knockdown of *ISWI1* affects cell survival and DNA elimination

We induced knockdown (KD) of *ISWI1* by feeding *Paramecium* with an *ISWI1*-specific sequence, triggering the cell's internal RNAi machinery (Fig EV1A). In a survival test of the post-autogamous progeny after *ISWI1*-KD over 3 days, 86% of the cells did not survive beyond the first day after cells were re-fed and allowed to resume vegetative division (Fig 1D). The remaining 14% of cells did not go through the usual rate of four vegetative divisions per day. In the control culture of *ND7*–KD (a gene required for exocytotic membrane fusion trichocyst discharge; Skouri & Cohen, 1997), the division rate of all the progeny remained unchanged. In the positive control of *PGM–KD*, 90% of the cells did not survive as expected. In contrast to *ISWI1–KD*, for *ISWI5*–KD, 90% of the cells showed no substantial difference in division rate or mortality compared to the control cells (Fig EV1B and C).

To test if the knockdown of *ISWI1* and *ISWI5* affect DNA elimination, we determined the retention status of germline-specific DNA elements in the newly developed MAC genome. We tested for IES retention from a well-characterized locus using PCR with IES-flanking primers (Appendix Table S1). For *ISWI1*–KD, most of the IESs we analyzed were retained (Fig 1F and Appendix Table S3). For *ISWI5*–KD, no retention of any of the IESs was observed (Fig EV1D). In *ISWI1–KD*, there was greater Sardine and Thon transposons retention, respectively, compared to the control *ND7*–KD (Fig 1E).

We also investigated the knockdown of other ISWI paralogs (*ISWI2*, *ISWI3*, and *ISWI4*; not upregulated during autogamy). In knockdown experiments for each of these paralogs, we did not observe growth defects or IES retention (Fig EV1E and F). To focus our investigations on genome reorganization, all further experiments were, therefore, carried out for *ISWI1* only.

ISWI1 is required for the complete excision of most IESs

To gain a genome-wide perspective on IES retention, we analyzed high-throughput sequencing of genomic DNA isolated from the developing macronucleus (anlagen) from *ISWI1*–KD cell cultures (two biological replicates). As a control, we used genomic DNA from the developing macronucleus after *ND7*–KD (also a pair of biological replicates). IES retention scores (IRSs) vary from 0.0 (complete IES excision) to 1.0 (complete failure of IES excision) upon knockdown. Approximately 35,000 (78%) IESs are sensitive to *ISWI1*–KD with a right-skewed retention score distribution (Fig 2A). IES retention scores of the biological replicates correlated well with each other (Pearson correlation coefficient: $r = 0.91$). Generally,

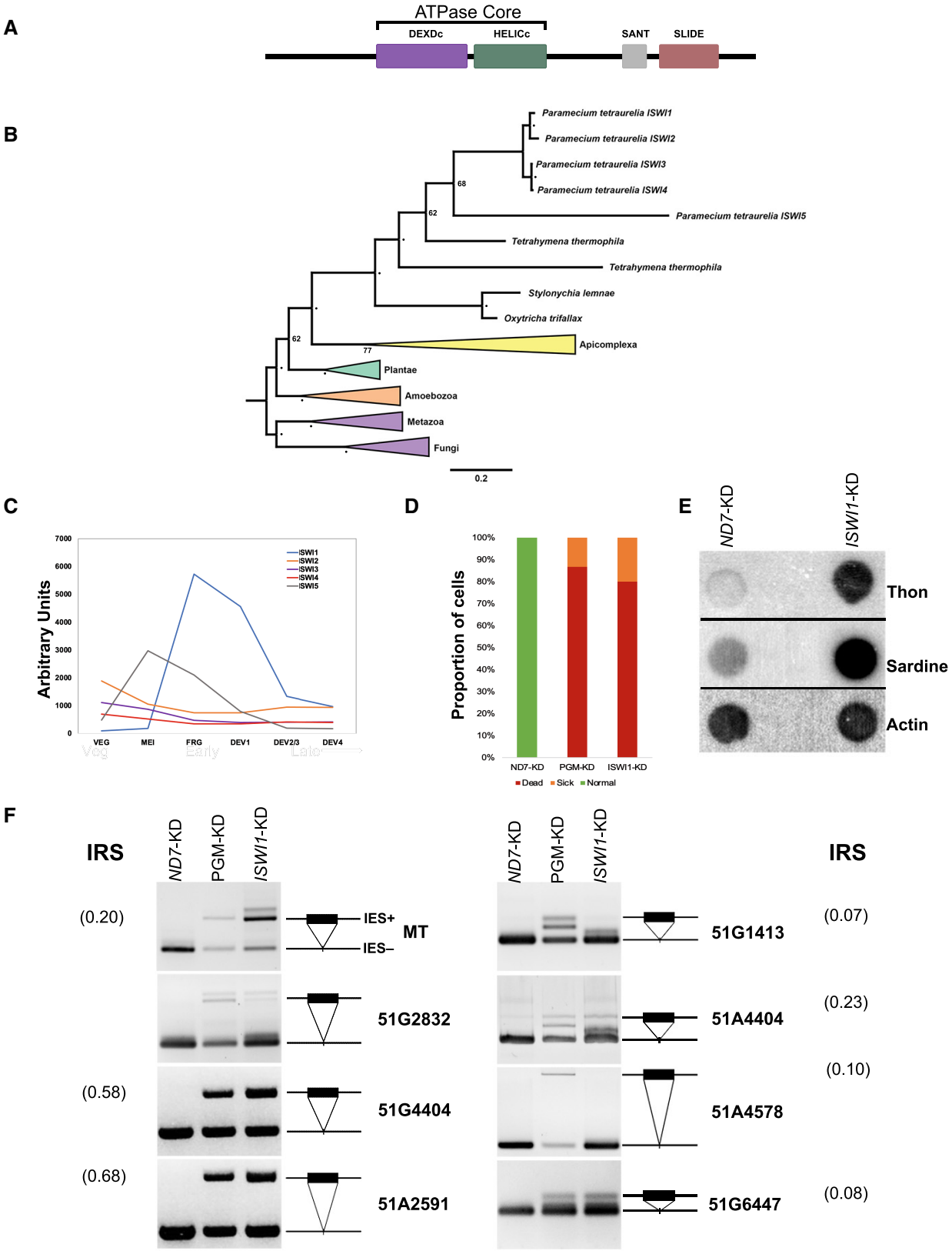


Figure 1.

Figure 1. Properties of ISWI1 and ISWI1-KD effects on DNA elimination.

A Predicted protein domains in ISWI1.
B Phylogenetic analysis of ISWI proteins in selected organisms. Node bootstrap values below ≥ 80 are indicated by '•' or are otherwise labeled.
C Gene expression profile (in arbitrary units) of ISWI genes based on published RNA-seq data (Arnaiz *et al*, 2010). Veg: cells undergoing vegetative division; Early: ~50% of cells with fragmented parental macronucleus (our early time point); Late: the majority of cells with a visible anlagen (our late timepoint).
D Survival test graph. Dead cells are represented in red, sick in orange, and normally dividing cells in green. *PGM*–KD is a positive control, and *ND7*–KD is a negative control.
E Dot blot analysis to check the effect of *ISWI1*-KD on transposon elimination. Probes against transposons Sardine and Thon were used, while a probe against Actin was used as a loading control.
F IES retention PCR (cropped inverted images). Four maternally-controlled IES and four non-maternally controlled IESs are shown. The IES+ band represents retained IES; the IES− band represents an excised IES; additional bands are likely PCR artifacts. IRS is IES retention Score for the IESs calculated after whole genome sequencing.

Source data are available online for this figure.

ISWI1–KD IES retention scores are modestly correlated with other known factors of excision machinery, correlating best with *DCL2/3/5*–KD ($r = 0.74$) and *NOWA1/2*–KD ($r = 0.72$; Fig EV2A). *ISWI1*–KD retention scores do not correlate as well with chromatin-related factors, *PTCAF1* ($r = 0.59$) and *EZL1* ($r = 0.52$).

As for most genes that influence IES excision, *ISWI1*–KD IES retention is length dependent (Fig EV2B). No periodicity of IES retention scores with respect to IES length is present. Similar to other gene knockdowns, IES sub-terminal base frequency changes relative to IES retention scores for *ISWI1*–KD, i.e., base frequencies are relatively constant for the shortest and most common IESs but differ considerably in relation to IES retention scores for longer IESs (Fig EV2C; Swart *et al*, 2014).

ISWI1-KD enhances excision of IESs at alternative boundaries

Excised IESs in *Paramecium* have a highly distinctive periodic length distribution (Fig 2D), proposed to reflect the periodicity of DNA and cooperation of transposase subunits during excision (Arnaiz *et al*, 2012). As can be seen in Fig 2D, the so-called "forbidden" second IES length peak (at ~40 bp; Arnaiz *et al*, 2012) is barely noticeable compared to the flanking IES length peaks. This was hypothesized as not being permitted by the biophysical constraints of DNA of this length, which prevents the two components of a conventional, domesticated PiggyBac transposase dimer from coming into the correct orientation needed for coordinated cleavage at both boundaries (Arnaiz *et al*, 2012). Since ISWI homologs are involved in nucleosome positioning in other organisms, we sought to determine if and how *ISWI1*–KD might impact IES excision precision.

First, we examined cryptic IESs, i.e., off-target IES-like sequences that are randomly excised at low levels throughout DNA, typically destined to become macronuclear during development (Duret *et al*, 2008; Swart *et al*, 2014). Such erroneous excision in *ISWI1*–KD was comparable to other knockdowns (Fig EV3C and D).

Next, we examined the excision of IESs at alternative boundaries. Natural excision of IESs using alternative boundaries occurs at low frequency, impacting ~16% of IESs in our negative control, *ND7*–KD (Fig 2B). In contrast, in *ISWI1*–KD, alternative boundary excision occurs at ~65% of IESs (supported by one or more mapped reads; Fig 2C). This is also substantially greater than for knockdowns of other genes necessary for IES excision, where the use of alternative IES boundaries is essentially the same as the control (Fig 2B). In general, though the amount of alternative IES excision for any given IES in *ISWI1*–KD is low (median 4.6%, mean 9.2%), it is substantially higher than that of other knockdowns (median 0%; mean 1.5–2.4%; Fig 2B).

The length distribution of alternatively excised IESs, irrespective of the knockdown, follows a similar periodic pattern to normal IESs, with smaller IESs more likely to result than larger ones (Fig 2D). Compared to normal IES excision, there is not as strong a preference for excision of the shortest IESs in alternative excision after *ISWI1*–KD.

Interestingly, there are substantially more alternatively excised IESs in *ISWI1*–KD in the second, "forbidden" length peak around 35 bp than conventional IESs (Fig 2D). We see a peak at this length of alternative excision events, regardless of whether they occurred internally versus externally (Fig EV3A and B). As for conventional IES excision, in other knockdowns, alternative IES excision in the forbidden length range was low (Fig 2D). Thus, enhanced alternative IES excision is a distinctive feature of *ISWI1*–KD. We also observe that most alternative excision events are close to the canonical IES boundaries, i.e., within 20 bp, or one or two turns of dsDNA (Fig 2E). In other words, *ISWI1*–KD leads to erroneous DNA excision at the next closest available sites.

ISWI1 protein localizes exclusively to the developing MAC

A C-terminal GFP fusion construct was made with *ISWI1* under the control of the putative *ISWI1* regulatory region and injected into *Paramecium* vegetative macronucleus. The transformed cell line was then cultured, and cells at different developmental stages (Fig 3A) were collected for confocal microscopy. When the early developing MACs (anlagen) were seen using DAPI staining, the GFP signal of the fusion protein also accumulated in the developing MAC and remained there throughout the late developmental stages (Fig 3B). The GFP signal was lost from the developing MAC after the developmental stages before karyogamy. Our observations suggest that the ISWI1 is expressed exclusively in the developing MAC at the time when genome reorganization takes place in *Paramecium*.

PTIWI01 and ISWI1 proteins interact *in vivo*

We sought to determine interacting partners of *Paramecium* ISWI1. First, we transformed *P. tetraurelia* cells with ISWI1 under its endogenous promoter and tagged it with a 3XFlagHA at its C-terminal. We then co-immunoprecipitated (IP) ISWI1 to analyze the associated proteins by label-free mass spectrometry. As a control, we performed the same experiment on wild-type cells where we did not expect to see any pulldown of proteins with the HA affinity

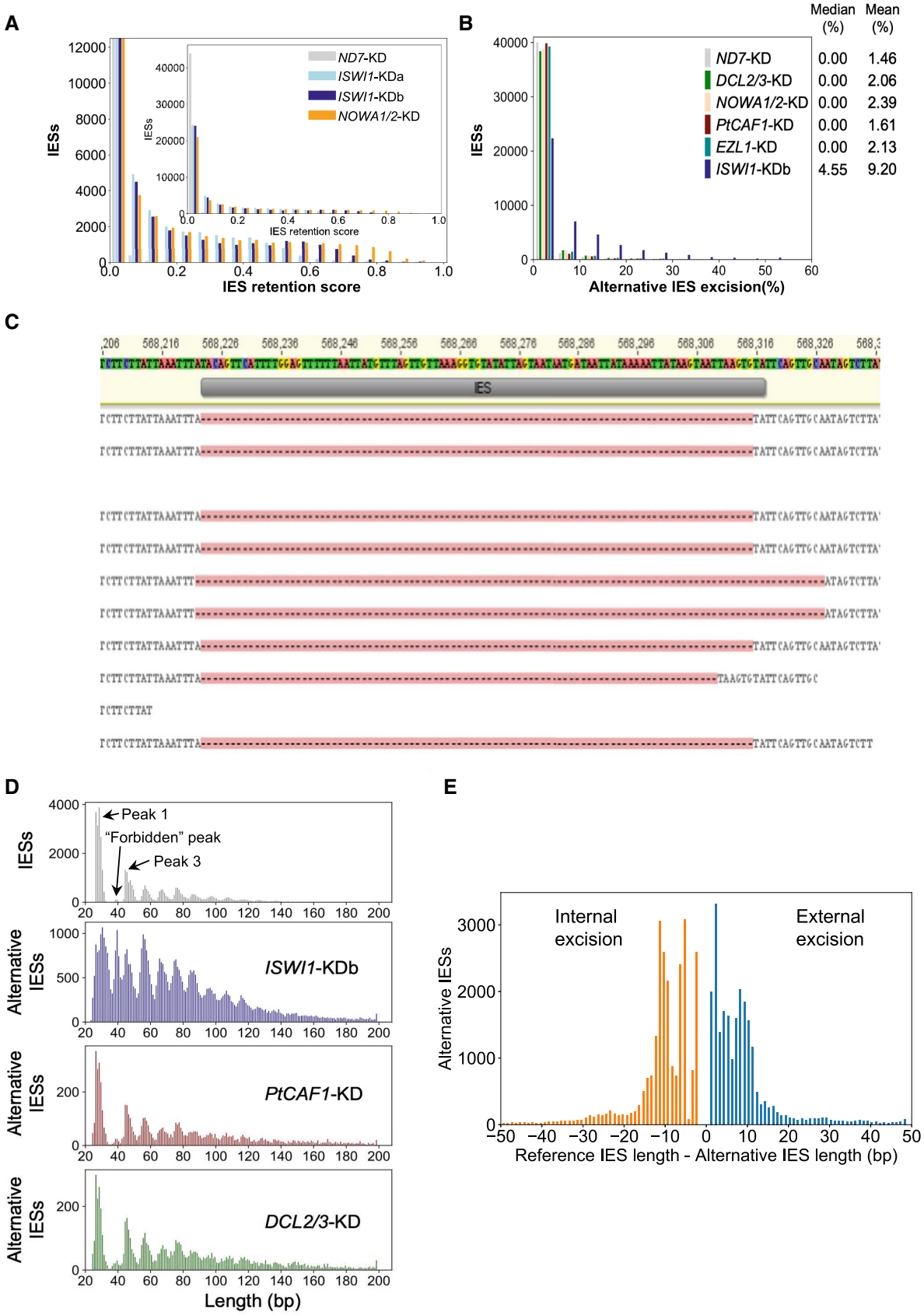


Figure 2.

Figure 2. Genome-wide analysis of IES excision upon ISWI1-KD.

A IES Retention Score (IRS) distributions for *ISWI1*–KD replicates and NOWA1/2-KD. ND7-KD was used as a negative control.

B Genome-wide analysis of alternative boundary excision in *ND7*-KD, *DCL2/3*-KD, *NOWA1/2*-KD, *EZL1*-KD, *PTCAF1*-KD, and *ISWI1*-KDb. Alternative excision (%) = 100*(alternative excised reads)/(alternatively + correctly excised reads).

C Reads mapped to an IES (IESPGM.PTET51.1.7.550914) showing both external (2 reads) and internal (1 read) alternatively excision; gaps opened in reads with excised IESs are indicated by dashes on a pink background.

D Length distribution of conventional IESs compared to alternatively excised IESs in knockdowns of ISWI1, PtCAF1, and DCL2/3.

E Difference in alternative IES lengths from the reference IES length.

Source data are available online for this figure.

matrix. The control cell samples and the cell samples transformed with ISWI1–3XFLAGHA were collected in two biological replicates during the developmental stage when ISWI1 localizes in the developing new MAC, as observed in Fig 3B. Before IP experiments, the samples were crosslinked with 1% PFA (see Materials and Methods).

We analyzed our IP samples by loading 1% of the total input and 20% of the IP fraction on an SDS gel. We detected a signal on a Western blot using an anti-HA antibody at the expected size of ~124 KDa (Fig 3C). The total IP samples were further analyzed using mass spectrometry (MS), where about 1,500 proteins were detected (Dataset EV1). Aside from proteins with peptides exclusively identified from cells expressing tagged ISWI1 protein, our analysis identified Ptiwi01 (or Ptiwi09, since most peptides from the mass-spectrometry analysis are shared between these almost identical proteins) as one of the proteins with the greatest fold enrichment in the ISWI1 IP (*P*-value: 0.0049; Figs 3D and EV4E, and Appendix Table S2).

We transformed *Paramecium* cells with 3XFLAGHA-tagged Ptiwi01 and GFP-tagged ISWI1 to test whether these proteins interact *in vivo*. IP samples were collected when ISWI1 localizes in the developing MAC. Cells were transformed with ISWI1–GFP as a negative control to check whether the GFP tag and HA affinity matrix could interact non-specifically. We performed the IP of the GFP-fused protein using the HA-affinity beads. GFP-fused ISWI1 (~150 KDa) was observed in the input but not in the IP (Fig 3E, lower panel). In addition, immunostaining was used to confirm the absence of HA signal in cells transformed with only ISWI1-GFP (Fig EV4F). Therefore, no cross-reactivity between the GFP and HA tags on their own was expected.

We observed no growth defects or IES retention in the transformed cells, either in single or in co-transformed cells (Fig EV4A–D). We succeeded in co-immunoprecipitating Ptiwi01 fused with 3XFLAGHA (expected size ~90 KDa) at the developmental stage when ISWI1 is expressed (Fig 3E, upper panel). IP samples were probed with an antibody against GFP, and a signal for ISWI1–GFP was detected at the expected size (~150 KDa, Fig 3E, lower panel). Our data suggest an interaction between ISWI1 and Ptiwi01, and most likely with Ptiwi09 (since they are 99% identical), in *Paramecium*. Since all our samples were crosslinked before the IP assays, we cannot exclude the possibility that this interaction might have been indirect via chromatin.

Nucleosomal densities increase with IES dependence on ISWI1 and other genes involved in *Paramecium* IES excision

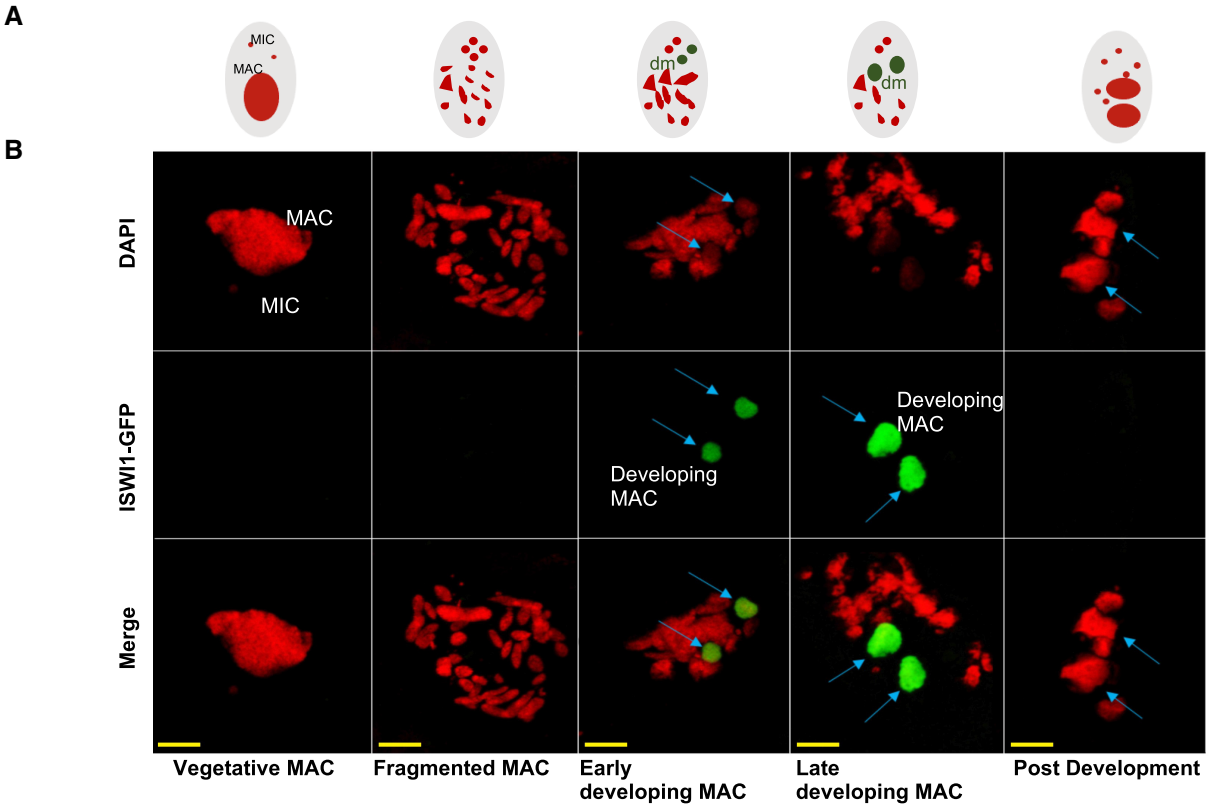
We sought to determine whether nucleosome density changes occur around an IES during DNA elimination and whether this is ISWI1 dependent. For this, we isolated developing macronuclear DNA from *ND7/PGM*–KD and *ISWI1/PGM*-KD cultures either with or without Atlantis dsDNase treatment. Atlantis dsDNase cleaves phosphodiester bonds in double-stranded DNA and yields homogeneous populations of core nucleosomes. As *PGM* is a key component of the core endonuclease that cleaves IESs (Baudry *et al*, 2009; Arnaiz *et al*, 2012; Bischerour *et al*, 2018), we used *ND7/PGM*-KD as the control for our experiment, mapping the nucleosome density around IESs. A double knockdown of *ISWI1* with *PGM* is necessary to retain the majority of IESs to map the nucleosome density across them.

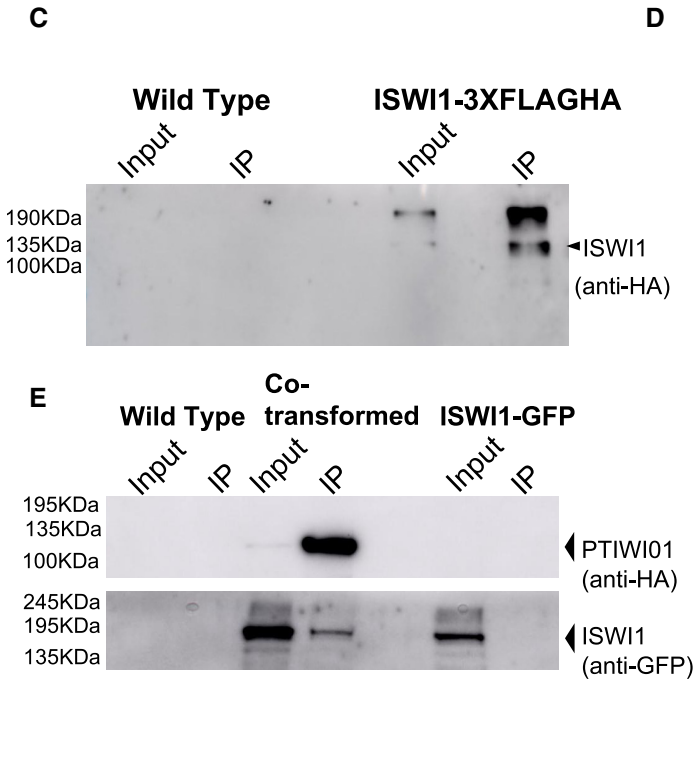
Given the constraint that a minimum of 9 bp of a read needs to match to an IES, and that some reads mapping to the flanking MDS regions may be derived from the parental MAC, it does not currently seem prudent to obtain accurate nucleosomal positioning for short IESs. We, therefore, examined a simpler measure of nucleosome densities for IESs: mapped nucleosome profiling (DNase-seq) reads, normalized by DNA-seq isolated from new MACs (Fig 4A–D).

In general, we observe that IESs, which are more strongly retained in any knockdown (e.g., *ISWI1*-KD IRS > 0.2), have higher nucleosome densities (Fig 4A and B). To rule out that this effect was not merely a consequence of more strongly retained IESs tending to be longer (e.g., Fig EV2B; Swart *et al*, 2014), we examined nucleosome density distributions of IESs of the same length, corresponding to the first IES length peak (26–31 bp). For these IESs, too, nucleosome densities are substantially higher for more strongly retained IESs (Fig 4C and D). Longer IESs (150–200 bp) show similar trends (Fig 4E and F), with higher nucleosome densities for more strongly retained IESs. Kolmogorov–Smirnov (KS) statistics between the distributions of IESs with IRS < 0.2 and IRS ≥ 0.2 vary between 0.33 and 0.38, with *P*-values < 1e-30 (Fig 4A–F).

ISWI1/PGM–KD alters the distribution of nucleosome densities compared to *ND7/PGM*–KD, for IESs in general and 26–31 bp IESs (Fig 4A–D; KS statistics between 0.048 and 0.089, with *P*-values < 1e-9). We also examined similar distributions for *NOWA1/2/PGM*–KD vs. *ND7/PGM*–KD (Fig EV5D–I), as both genes are required for the sRNA-mediated genome scanning (Nowacki *et al*, 2005), and their IES retention scores correlate more strongly with *ISWI1*–KD and *DCL2/3/5*–KD than *PTWI01/09*–KD. Though the coverage of DNase-seq was lower, as the number of nucleosomal reads mapping from the libraries was smaller, for a separate experiment with *ND7/PGM*–KD and *NOWA1/2/PGM*–KD, we also observed differences in distributions of nucleosomes for IESs, particularly those with *ISWI1*–KDb IRS ≥ 0.2 (KS statistics 0.17–0.19; *P*-values < 1e-7; Fig EV5D–I).

In summary, there appear to be differences in nucleosome density distributions between both *ISWI1/PGM*–KD and *ND7/PGM*–KD, and *NOWA1/2/PGM*–KD and *EV/PGM*–KD. However, these are much less pronounced than the difference in nucleosome density





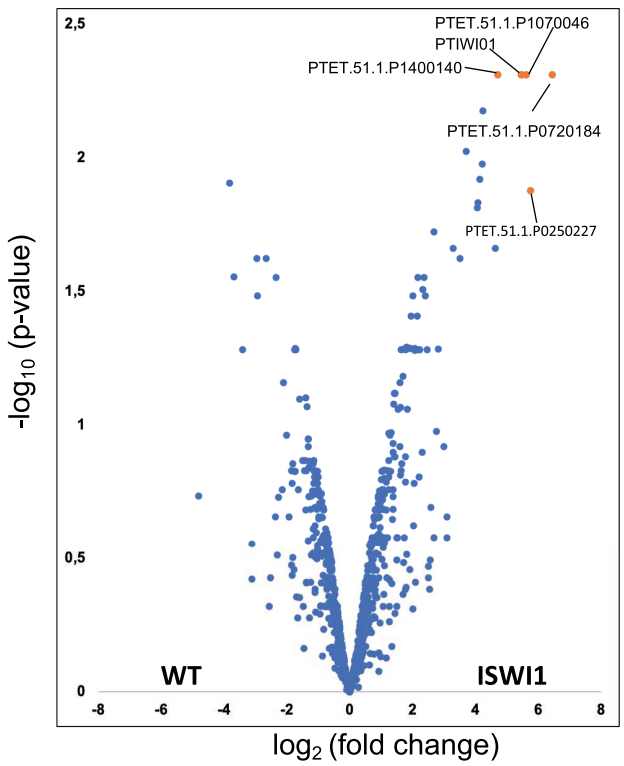


Figure 3.

Figure 3. Localization, Co-immunoprecipitation, and mass spectrometry analysis.

A Schematic drawing of the life cycle stages of *Paramecium tetraurelia*. MIC and parental MAC are represented in red, representing the DAPI signal, and developing MAC (dm) is represented in green until fully developed, representing the GFP signal.

B ISWI1-tagged C-terminally with GFP localizes in the developing MAC as soon as developing new MACs (panel Early Development) become visible and remain there throughout late MAC development (panel Late Development). Red: DAPI, Green: ISWI1–GFP. Blue arrows identify developing MAC; scale bar 10 μm.

C Western blot analysis using anti-HA antibody after coimmunoprecipitation of ISWI1-3XFlagHA fusion protein. Non-transformed cells (WT) of the same strain were used as the negative control. 1% of the total lysate was loaded as Input, and 20% of co-immunoprecipitated samples were loaded on 12% SDS gel.

D Volcano plot illustrating the distribution of proteins identified in label-free MS in WT Vs. ISWI1-3XFlagHA. Significantly abundant proteins (fold change ≥ 4) are highlighted in orange.

E Western blot analysis using anti-HA and anti-GFP antibodies after coimmunoprecipitation of Ptiwi01-3XFlagHA fusion protein co-transformed with ISWI1-GFP. Non-transformed cells (WT) of the same strain and ISWI1-GFP fusion protein transformation were used as negative controls. 1% of the total lysate was loaded as Input, and 20% of co-immunoprecipitated samples were loaded on 10% SDS gel.

Source data are available online for this figure.

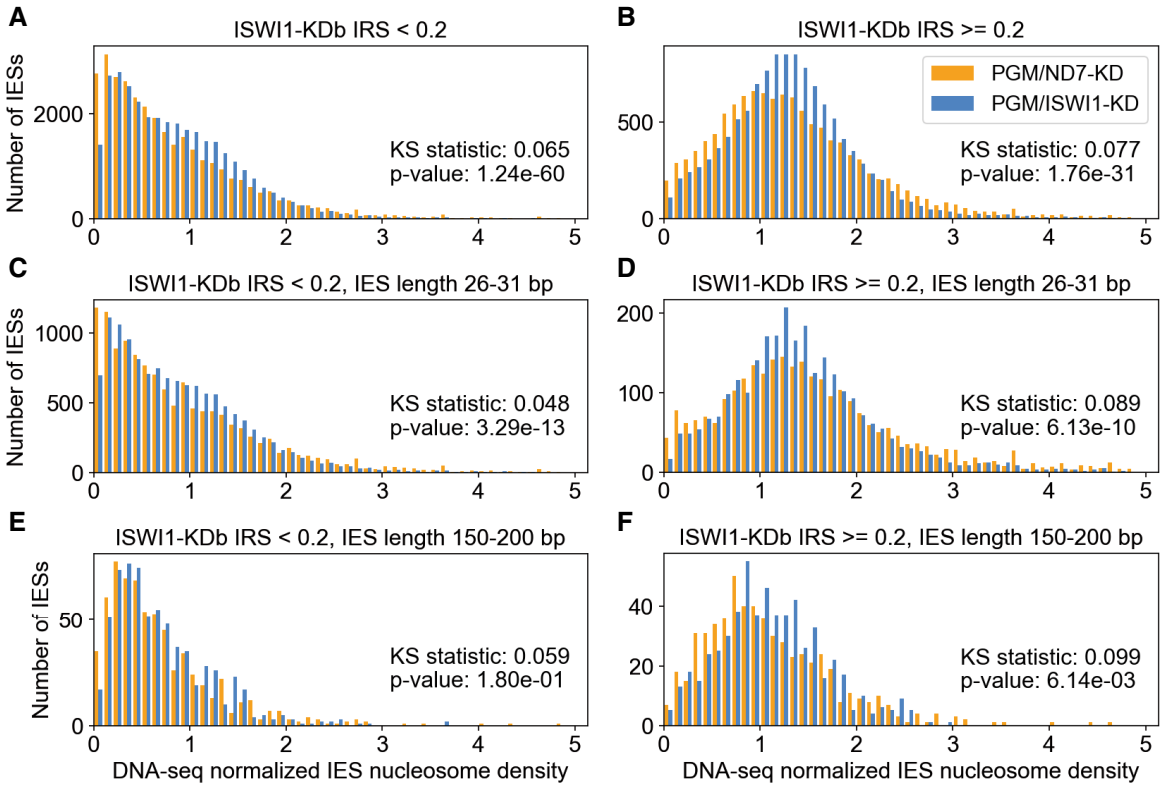


Figure 4. Nucleosome density increases with IES retention in ISWI1-KD.

A, B Nucleosome density histograms for IESs weakly (IRS < 0.2) or more strongly retained in *ISWI1*-Kdb (IRS ≥ 0.2). Kolmogorov–Smirnov statistics and their *P*-values are provided.

C–F Histograms as in (A&B), including additional length constraints, corresponding to the first IES length peak (26–31 bp; C and D) or the first non-periodic length IESs (150–200 bp; E and F).

distributions between IESs that are more weakly and more strongly retained in knockdowns like *ISWI1*–KD.

Discussion

Paramecium depends on efficient and accurate whole genome reorganization to produce a functional somatic nucleus during sexual reproduction. The excision of numerous IESs requires scnRNAs for their excision. Identification of additional proteins required for the excision of IESs (Arambasic *et al*, 2014; Data ref: Lhuillier-Akakpo *et al*, 2014a; Wasmuth & Lima, 2017) suggests additional or alternative mechanisms beyond those envisaged in earlier models of RNA scanning and heterochromatin formation contributing to IES targeting and excision.

In this study, we have identified a homolog of ISWI, an ATP-dependent chromatin remodeler, that is required for the precise elimination of IESs. ISWI proteins are highly conserved

ATP-dependent chromatin remodelers (Corona *et al*, 1999) that regulate several biological processes (Yadon & Tsukiyama, 2011), and now, as we have shown, also in genome editing in *Paramecium*. *Paramecium*'s ISWI1 is exclusively present in the developing macronucleus (Fig 3A) when the molecules responsible for genome reorganization cooperate to eliminate DNA. We also show that, in *Paramecium*, ISWI1 can interact with PIWI proteins (Fig 3E) that are known to guide genome reorganization in ciliates in an sRNA-dependent manner (Bouhouche *et al*, 2011; Furrer *et al*, 2017b). Our data, therefore, suggest that the shifting action of ISWI1 occurs in conjunction with an sRNA–Piwicomplex that guides subsequent precise excision.

Histone modification and heterochromatin formation are proposed to be a prerequisite for programmed DNA elimination in ciliates. The most evidence in support of this has been obtained for *Tetrahymena thermophila* (Liu *et al*, 2007; Xu *et al*, 2021). A similar model was proposed for IES excision in *Paramecium* (Coyne *et al*, 2012). It has been shown that histone modifications, in particular H3K27me3 and H3K9me3, are required for targeting the excision of at least a subset of IESs (Ignarski *et al*, 2014; Data ref: Lhuillier-Akakpo *et al*, 2014a). Indeed, the knockdown of EZL1, a histone methyltransferase (Frapporti *et al*, 2019), affects the excision of the majority of IESs. Since heterochromatin regions generally spread across several kilobases in the genomes of other organisms (Margueron & Reinberg, 2011; Huang *et al*, 2012), it was suggested that in *Paramecium*, H3K27me3 and H3K9me3 marks are placed locally (Lhuillier-Akakpo *et al*, 2014b). Although it was recently shown that the transposable elements are enriched with nucleosomes bearing these modifications (Frapporti *et al*, 2019), currently, there is no published information on H3K27me3 or H3K9me3 nucleosome association with IESs. Moreover, these modifications are not limited to the developing macronucleus and are also present in the fragments of the parental macronucleus (Ignarski *et al*, 2014; Lhuillier-Akakpo *et al*, 2014b; Frapporti *et al*, 2019). The inhibition of IES excision and the resultant cell lethality due to *EZL1*–KD and/or *PTCAF1*–KD may arise due to alteration in gene expression and failure to repress transposable elements by the PRC complex that also interacts with Ptiwi01/09 proteins (Miró-Pina *et al*, 2022). Thus, further experiments will be necessary to disentangle possible indirect effects of these knockdowns from direct ones.

ISWI1–KD IES retention correlates better with *DCL2/3/5*–KD than with *DCL2/3*–KD or *DLC5*–KD (Fig EV2A), suggesting ISWI1 is necessary for excision of IESs requiring either scnRNAs or iesRNAs. In addition, we also observed an interaction between Ptiwi01 (or Ptiwi09) and ISWI1 *in vivo* in our co-immunoprecipitation assay, though this may be an indirect action with chromatin intervening (Fig 3E). We also observed Ptiwi11 in our mass spectrometry analysis (Appendix Table S2). Taken together with a stronger correlation between *DCL2/3/5*-KD and *ISWI1*–KD, we suggest ISWI1 also cooperates with iesRNAs in targeting IESs.

IESs most sensitive to *ISWI1*–KD and other knockdowns, like *DCL2/3/5*–KD, are substantially more nucleosome rich (Fig 4C–H). Like *ISWI1/PGM*–KD, *NOWA1/2/PGM*–KD alters the distribution of nucleosome densities across IESs. However, this is certainly more subtle than the large differences in these densities observed between weaker and more strongly retained IESs upon *ISWI1*–KD (compare Figs 4C–H and EV5D–I) and other genes we examined that are involved in genome reorganization. A plausible explanation could be that local nucleosome density changes are required to govern accessibility and possibly activating the endonuclease for DNA elimination. A similar explanation has been proposed for V(D)J recombination, where nucleosome location and occupancy changes were observed to regulate DNA recombination (Pulivarthy *et al*, 2016).

In the future, detailed DNase-seq experiments with variable nuclease digestion conditions and deeper sequencing may be able to obtain greater resolution and examination of dynamics. Furthermore, it will be necessary to conduct additional experiments to resolve the possible contributions of non-nucleosomal proteins to protecting DNA from DNase digestion. Nevertheless, as nucleosomal proteins are the most abundant nuclear DNA-binding proteins, we believe they are the largest contributors to the differences in DNase-seq read distributions we observed, hence why we refer to them as nucleosome density distributions.

Recent research into *Paramecium* MAC chromatin has revealed notable differences from other eukaryotes, including the ciliate *Tetrahymena*, including the absence of linker histones in *Paramecium* (Drews *et al*, 2022; Gnan *et al*, 2022). In particular, *Paramecium* has extremely average short internucleosomal distances (~151 bp). This would correspond to tiny linker sequences of several bases, rather than tens of bases in other eukaryotes, including *Tetrahymena* (Drews *et al*, 2022; Gnan *et al*, 2022). Thus, we expect *Tetrahymena* IES excision constraints to differ fundamentally from *Paramecium's*.

Uniquely among *Paramecium* proteins involved in IES excision investigated thus far, *ISWI1* gene silencing leads to elevated alternative IES excision (Fig 2B), suggesting that the endonuclease complex is not always able to correctly target the boundaries of an IES in the absence of ISWI1. The commonly accepted mechanism underlying ISWI function is that it controls the length of linker DNA and the chromatin architecture by altering nucleosome spacing (Xiao *et al*, 2001; Corona *et al*, 2007; Bartholomew, 2014). Global nucleosome density changes are known to occur across genomes during cell lineage commitment as an additional regulatory mechanism (Erdel *et al*, 2011; Li *et al*, 2012).

We propose that the presence of nucleosomes on, or partially overlapping, an IES may be crucial for its targeting and accessibility to the excision machinery (Fig 5). In contrast to the current "naked" DNA model for IES excision (Fig 5A), we propose a "clothed" DNA model with nucleosomes present. Crucially, in our model, IES boundaries need to be accessible to their excesses. We propose that "forbidden" length DNA is cut when nucleosomes have not been displaced from IES boundaries by ISWI1, as happens with *ISWI1*–KD (Fig 2D). In the absence of the usual required nucleosomal shift, IESs can be excised at alternative TA boundaries, though they are still most frequently cut at the conventional boundaries (Fig 5B and C). In other words, *ISWI1*–KD assists in properly positioning nucleosomes around an IES, preventing alternative excision errors.

In Fig 5, we do not indicate the involvement of any histone modifications in *Paramecium* IES excision. Until more detailed analyses can be performed, showing the exact positioning of specific histone modifications in relation to IESs, we would prefer to avoid speculating about their role. On the other hand, it may also be possible for an alternative mechanism for IES targeting that does not invoke such modifications. Instead, it might also be possible that longer RNA transcripts across IESs promote binding of scnRNAs/iesRNAs,

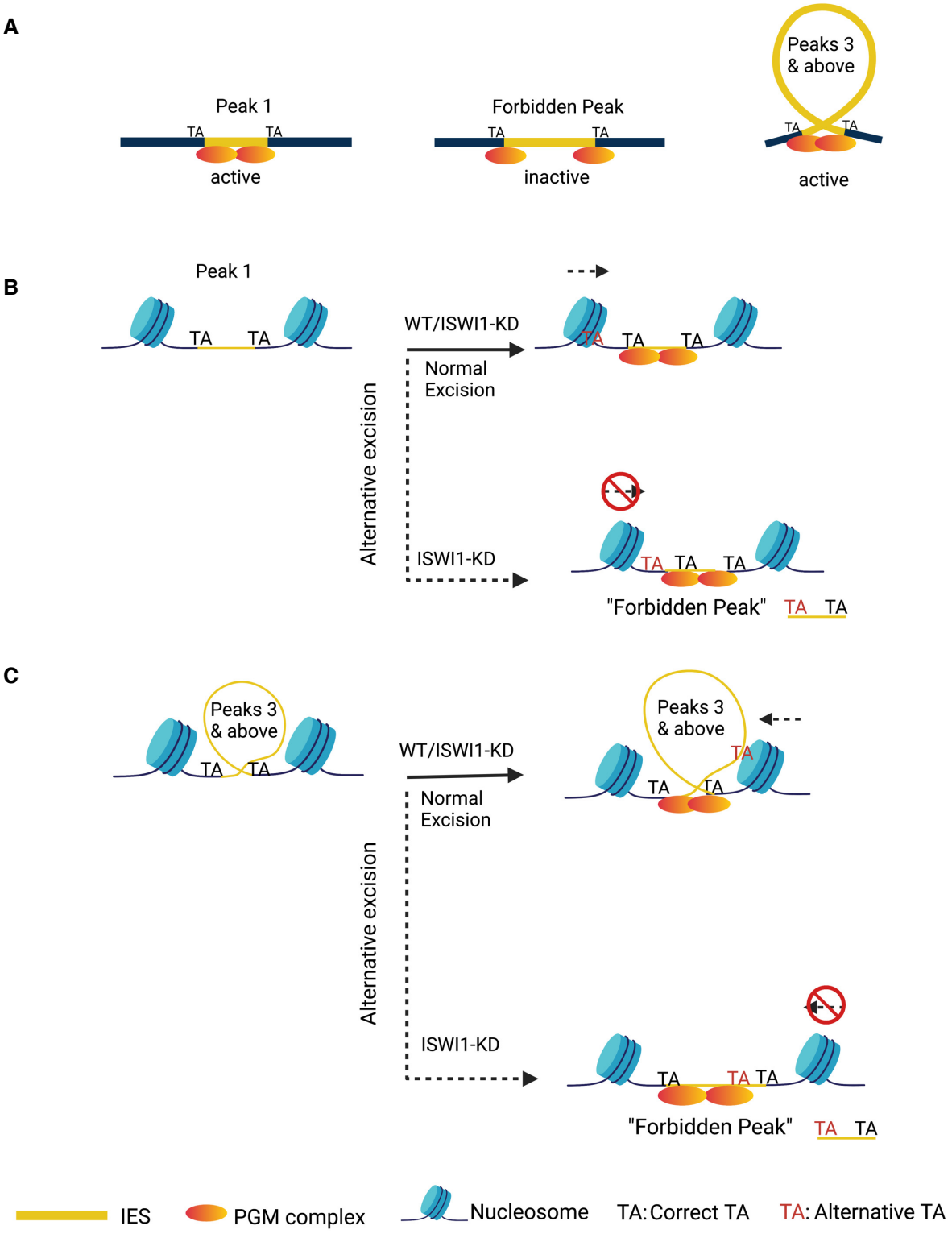


Figure 5. Assembly of active PiggyMac (PGM) excision complex on IESs.

A–C (A) "Naked" model proposed by Arnaiz *et al*, 2012; (B and C) Revised "clothed" model, which accounts for accessibility of IES boundaries in the presence of nucleosomes. If nucleosomes are not properly positioned, IESs can be cut at alternative boundaries, leading to IES accumulation of the "forbidden peak" length. Image created with BioRender.com.

and more direct recruitment of the IES excisases. In any event, more detailed experiments will be necessary to examine nucleosomal properties, including positioning and modifications, and how they might influence IES excision.

Taken together, our investigations provide evidence of an interplay between chromatin remodeling and sRNA-complexes during *Paramecium* genome development. Typically chromatin remodelers do not operate in isolation in other organisms but as multicomponent complexes, performing a range of sophisticated functions. In the future, it would be necessary to closely examine the mechanistic details of the interplay of ISWI1 with sRNAs and other proteins in *Paramecium* and how they are involved in massive, accurate genome editing.

Materials and Methods

Paramecium cultivation

Mating type 7 of *Paramecium* strain 51 was used in different experiments. Cells were cultured in Wheat Grass Powder (WGP; Pines International, Lawrence, KS) medium bacterized either with nonvirulent *Klebsiella pneumoniae* or with *Escherichia coli*, strain HT115 and supplemented with 0.8 mg/l of β-sitosterol (567152, Calbiochem). Cells were either cultured at 27°C or at 18°C as per requirement. Clonal cell lines of *Paramecium* transformed with recombinant genes were maintained at 18°C as previously described (Beisson *et al*, 2010b).

Knockdown experiments, survival test, and IES retention PCR

The silencing (gene knockdown) construct of *ISWI1* (Genbank accession: XM_001431568, XM_001431569) was made by cloning a 704 bp construct from its C-terminal and cloned into an L4440 plasmid (using GGGTCTCACCTAAGATGAACG and TCACTTTCTTAACAGACTCAGATCC). *ISWI2* (Genbank accession: XM_001447087.1) a 584 bp long region (using GGAGGAGCGTTAAGAACAA, CACAAGAGATCTTCCCATAG) was used for generating the silencing construct. For *ISWI3* (Genbank accession: XM_001442140), using CTT AGCTAGTCATCTCTTT and CTTTTCATAAGCATCCTTG oligonucleotides, a 500 bp long region was cloned, and for *ISWI4* (Genbank accession: XM_001446844.1) a 394 bp long region was cloned (using CAATTGCTAATCATCATTTC, GAGAGTTTTGGATTTAACG) for the knockdown experiments. For *ISWI5* (Genbank accession: XM_001432642), the silencing construct was made by cloning an 1,106 bp long fragment into an L4440 plasmid (using ATGAGTGAAAGTGAAGATGAG and AGATTTCGTCCTTCTTAACAT). The plasmids were then transformed into HT1115 (DE3) *E. coli* strain.

Cells were seeded into the silencing medium at a density of 100 cells/ml, and silencing was carried out according to a previously described protocol (Beisson *et al*, 2010c). After the cells finished autogamy, 30 post-autogamous cells were transferred individually to threewell glass slides containing the medium bacterized with avirulent *K. pneumoniae* for the survival test. Cells were monitored for 3 days (approximately 12 divisions) and categorized into three groups according to their observed phenotype. In parallel, a 100 ml culture was harvested for DNA extraction using GeneElute–Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich). PCRs were done on different genomic regions flanking an IES (Appendix Table S1).

In regards to co-silencing performed to analyze nucleosomal densities, the distribution of retention scores in *PGM/ND7*–KD is shifted and skewed to the left (lower IES retention) compared to the reference *PGM*–KD data sets (Arnaiz *et al*, 2012; Swart *et al*, 2014), whereas, the IRS of *ISWI1/PGM*-KD is more similar to the knockdown expected for *PGM*–KD (Fig EV5A). Previous experiments have shown that weakened IES retention due to dilution of gene knockdown can occur in *Paramecium* due to gene co-silencing (Bischerour *et al*, 2018). The weaker silencing effect can be explained by the dilution of the *PGM* silencing medium with the *ND7* silencing medium. This was done to ensure that the RNAi effects from the *PGM/ND7* and the *ISWI1/PGM* knockdowns would be directly comparable.

Related to this, for *NOWA1/2/PGM* silencing, *NOWA1/2*-KD also minimizes potential dilution effects since one silencing construct silences both genes (Nowacki *et al*, 2005), whereas *PTIWI01/09*–KD requires two silencing constructs in addition to the necessary *PGM* silencing construct.

Dot blot

Dot blot assays were conducted following standard protocols (Brown, 2001). Briefly, 3 μg of DNA from post-autogamous cultures were blotted onto a nylon membrane (Hybond N+XL membrane, Amersham). Probes specific to Sardine and Thon transposons and actin (first 240 bp of the gene) labeled with α-32P dATP (3,000 Ci/mmol) using RadPrime DNA Labeling System (Invitrogen) were used. The signal was quantified with ImageJ 1.48e.

Northern blot

Ten microgram of RNA were run in a 1.2% agarose denaturing gel and transferred to a nylon membrane (Hybond N+XL membrane, Amersham) by capillary blotting. After transfer, the membrane was crosslinked twice with UV (120,000 μJ/cm^2). Specific probes labeled with α-32P dATP (3,000 Ci/mmol) using RadPrime DNA Labeling System (Invitrogen) for ISWI1, ISWI5, and rRNA were used for hybridization. Membranes were screened using the Typhoon Imaging system (GE Healthcare).

GFP tagging, microinjection, and GFP localization experiment

A set of specific *ISWI1* specific primers (5′-GTA GAA TCC TAT TGA TAG GAG GAG-3′ and 5′-TGG CTC TAA GAA ATT CAT TTA T-3′) were used for the amplification of full gene including 227 bp upstream and 62 bp downstream of the coding region. *ISWI1* was tagged with GFP on its C-terminus. The construct was linearized using the NaeI restriction enzyme (R0190S, New England Biolabs) and injected into the macronucleus of the vegetative cells as previously described (Beisson *et al*, 2010d). Cells positive for GFP expression were collected during different stages of autogamy and either stored with 70% ethanol at −20°C or directly fixed with 2% PFA in PHEM and then washed in 5% BSA with 0.1% Triton X-100. Cells were then counterstained with DAPI (4,6-diamidino-2-2phenylindole) in 5% BSA with 0.1% Triton X-100 and mounted with Prolong Gold Antifade mounting medium (Invitrogen). Images

were then acquired with Olympus Fluoview FV1000 confocal microscope system with PLAPON 60× O SC NA 1.40. Images were analyzed and given pseudo-color on Imaris software.

Immunofluorescence analysis

Immunostaining of cells transformed with ISWI1-GFP was done after the cells were fixed in 70% ethanol at −20°C. Cells were first washed in 1× PBS pH7.4 twice for 5 min to remove any traces of ethanol. Cells were then permeabilized with 1% Triton X-100 in Pipes–Hepes–EGTA–MgCl2 (PHEM) buffer for 20 min at room temperature. Afterward, cells were fixed in 2% paraformaldehyde for 10 min and washed with 1× PBS for 5 min. Cells were then blocked in 3% BSA in TBSTEM buffer for 1 h at room temperature. Primary antibody incubation was done at 4°C overnight using mouse anti-HA (sc-7,392, Santa Cruz) with a 1:50 dilution factor. After washing the cells three times for 10 min in 3% BSA in TBSTEM, cells were incubated in goat anti-mouse Alexa-594 (dilution 1:200, BLD–405326, Biozol) for 1 h in dark conditions. The cells were further washed three times for 10 min in 3% BSA in TBSTEM. In the last wash, DAPI was added to the BSA, and cells were incubated for 5 min. The cells were then mounted with Prolong® Gold Antifade mounting medium (Life Technologies), and sealed with a coverslip. Images were acquired on TCS SP8 with a 63×/1.40 oil objective, zoom factor 3 and step size of 1.0. Images were analyzed using Fiji with maximum intensity projection.

Co-immunoprecipitation assay

ISWI1 specific primers (5′-GTA GAA TCC TAT TGA TAG GAG GAG-3′ and 5′-TGG CTC TAA GAA ATT CAT TTA T-3′) were used for the amplification of the full gene with regulatory regions. The gene was tagged with 3XFLAGHA at its C-terminal. 4.5×10^5 cells were harvested and crosslinked with 1% Paraformaldehyde by incubating for 10 min (min) at room temperature. Cells were then quenched using 100 µl of 1.25 M Glycine and incubated at room temperature for 5 min. Cells were washed once with PBS for 2 min at 500 *g*. Further steps were carried out on ice or at 4°C. Two milliliter of lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 1%Triton X100, 1× Protease inhibitor (Roche,11836170001), 10% glycerol) was added and the cells were sonicated (Branson Digital Sonifier) with 55% amplitude for 15 s. The lysate was then centrifuged for 30 min at 13,000 *g* or until the lysate was clear. Fifty microliter of bead slurry (HA High-Affinity Matrix,11815016001, clone 3F10, Roche) was used per IP sample and was washed thrice by centrifuging for 2 min at 500 *g*. After washing the beads, 1 ml of the lysate was mixed with the beads and incubated overnight with agitation at 4°C. After the incubation, the beads were washed five times with the IP buffer (10 mM Tris pH8.0, 150 mM NaCl, 0.01% NP-40, 1 mM $MgCl_2$, 1× Protease inhibitor (Roche,11836170001), 5% Glycerol) for 2 min at 500 *g*. NP-40 was added freshly to the buffer. Proteins were then eluted by adding 50 µl of the 2× loading buffer (10% SDS, 0.25 M Tris pH 6.8, 50% Glycerol, 0.2 M DTT, 0.25% Bromophenol blue).

For co-transformation with ISWI1-GFP, PTIWI01 with primers in its regulatory regions was (CATTTTTAAGAGATTTCAATAAAA-CAATTATCC and GTGCTTTGAAAATCAATGAAAATCA) amplified, and 3XFLAGHA was fused at its N-terminal. After linearisation with NaeI, both constructs were mixed in equal proportions for microinjection. Co-immunoprecipitation assay was performed as explained above with a slight modification. Sonication was done with 52% amplitude for 20 s using MS72 tip on Bandelin Sonopulse.

Mass spectrometry analyses

Mass spectrometry data processing and statistics were provided by the Proteomics & Mass Spectrometry Core Facility (PMSCF), University of Bern. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Deutsch *et al*, 2020) via the PRIDE (Perez-Riverol *et al*, 2019) partner repository with the dataset identifier PXD027206.

Differential expression tests were performed for proteins detected in the control and ISWI1 pulldown groups by applying the empirical Bayes moderated *t*-test (Kammers *et al*, 2015) as implemented in the R limma package. Bayes statistics were only applied where there were two valid LFQ (label-free quantitative intensity) values. The adjusted *P*-values were calculated following the Benjamini & Hochberg (1995) method to correct for multiple testing.

Western blot

Western blotting on IP samples was done by running a 10% SDS–PAGE gel, and the proteins were transferred on a 0.45 µm nitrocellulose membrane (10600002 Amersham, GE Healthcare). 1% of Input and 20% of IP fraction were used for the samples to be run on the gel. The membrane was blocked with 5% BSA in PBS for 1 h at room temperature. The membrane was then incubated overnight at 4°C with anti-HA (sc805, Santa Cruz, RRID: AB_631618) at a dilution of 1:500. A goat anti-rabbit HRP conjugate (sc2004, Santa Cruz, RRID: AB_631746) in a dilution of 1:5,000 was used after washing the membrane with PBS/0.1% Tween-20 for 10 min (three times). For PTIWI01-3XFLAGHA IP, the membrane was incubated with either anti-HA (sc-7,392 HRP, Santa Cruz, RRID: AB_627809) in a dilution of 1:500 or with anti-GFP (ab290, Abcam, RRID: AB_303395) in a dilution of 1:1,000. The secondary antibody incubation was done for 1 h at room temperature, and the membrane was washed thrice with PBS/0.1% Tween-20 for 10 min. The membrane was then washed once for 5 min with 1× PBS before imaging. The membrane was scanned using chemiluminescence settings on an Amersham Imager 600 (GE Healthcare).

Phylogenetic analyses

ISWI proteins were identified (OG5_127117) and retrieved using PhyloToL (Cerón-Romero *et al*, 2019). Briefly, multi-sequence alignments were constructed using MAFFT (Katoh & Standley, 2013) and then iteratively refined with GUIDANCE2 (Sela *et al*, 2015), which identifies and removes spurious sequences and columns, preserving phylogenetically informative regions in the alignment. This refined alignment was then passed to RAxML (Stamatakis, 2014) and used to generate 200 bootstrap replicates.

Macronuclear isolation and Illumina DNA-sequencing

The samples for MAC isolation were collected from *ND7–KD*, *ISWI1–KD*, and *PTCAF1–KD* cultures from the cultures 3 days post

autogamy, as described previously (Arnaiz *et al*, 2012). Paired-end libraries (Illumina TruSeq DNA, PCR-free) were made according to the standard Illumina protocol. Library preparation and sequencing were done at the NGS platform, University of Bern.

Reference genomes

The following reference genomes were used for analyzing DNA-seq data.

MAC: https://paramecium.i2bc.paris-saclay.fr/files/Paramecium/tetraurelia/51/sequences/ptetraurelia_mac_51.fa

MAC + IES: https://paramecium.i2bc.paris-saclay.fr/files/Paramecium/tetraurelia/51/sequences/ptetraurelia_mac_51_with_ies.fa

IES retention and alternative boundary analysis

IES retention scores were calculated with the MIRET component of ParTIES (Denby Wilkes *et al*, 2016). IES retention scores are provided as Source Data for Figure 2 (ISWI1_MIRET.tab).

The MILORD component (default parameters) of a pre-release version (13 August 2015) of ParTIES was used to annotate alternative and cryptic IES excision. For each IES with alternative or cryptic excision, the identifiers for the supporting reads are recorded. Output for this is provided as Dataset EV2 (CAF1_MILORD.gff3.gz, DCL23_MILORD.gff3.gz, ISWI1-b_MILORD.gff3.gz, ND7-b_MILORD.gff3.gz, and NOWA1_MILORD.gff3.gz). IRS correlations, the relationship of IRS with length, and sub-terminal frequencies were calculated as described previously (Swart *et al*, 2014). IES retention scores for *PGM/ISWI1*-KD and *PGM/ND7*–KD are provided in Source Data for Expanded View (PGM_ND7_ISWI_MIRET.tsv). The DNA-seq data for IRS correlations and alternative excision analysis apart from *ISWI1*-KDs and their corresponding controls were obtained from previous studies (ENA PRJEB12406 (Data ref: Swart *et al*, 2017b); ENA ERA309409 (Data ref: Lhuillier-Akakpo *et al*, 2014a); ENA ERS1656548 (Data ref: Furrer *et al*, 2017a) SRA SRX215498 (Data ref: Sandoval *et al*, 2014a)).

Nucleosomal DNA isolation and Illumina DNA-sequencing

Cultures for nucleosomal DNA isolation were harvested approximately 16 h after the developing macronucleus were seen. Macronuclear DNA isolation protocol was followed up to the stage of ultracentrifugation. After ultracentrifugation, the pellet containing macronucleus was washed twice with chilled 1× PBS pH 7.4, and the excess PBS was removed by centrifuging at 200 *g* for 2 min at 4°C. Half of the nuclear pellet was then recovered in 100 µl of resuspension buffer (10 mM Tris–HCl, pH 7.4, 10 mM $MgCl_2$) for MAC DNA isolation and sequencing (DNA-seq). The other half of the pellet was used for nucleosomal DNA isolation (DNase-seq). All the steps from here were optimized from the standard protocol provided with the EZ Nucleosomal DNA Prep Kit (D5220, Zymo Research). Briefly, 1 ml of chilled Nuclei Prep Buffer was used to resuspend the cell pellet before incubating on ice for 5 min. The nuclear pellet was then centrifuged at 200 *g* for 2 min at 4°C. After washing twice with Atlantis Digestion buffer, the pellet was resuspended in 1 ml of Atlantis Digestion Buffer. Five hundred microliter of the reaction was then used for DNA isolation without digestion as a control. The remaining 500 µl of the reaction was used for nucleosomal DNA isolation and 35 µl of the Atlantis dsDNAse. The reaction was incubated at 42°C for 20 min. After 20 min, the reaction was stopped by adding MN Stop Buffer, and the nucleosomal DNA isolation was carried out according to the kit protocol (D5220, Zymo Research). Note that we used Atlantis DNase for nucleosomal DNA isolation (provided in the kit D5220, Zymo Research). Atlantis dsDNase is an endonuclease specific to double-stranded DNA that cleaves phosphodiester bonds yielding oligonucleotides with 5′-phosphate and 3′-hydroxyl termini.

For *ISWI1/PGM*–KD and its control *ND7&/PGM*–KD, Illumina TruSeq PCR-free DNA library was prepared without bead-based size selection, followed by a preparative size selection on the PippinHT to remove non-ligated adaptors and library molecules with inserts > 500 bp (refer to Fig EV5B). The samples were sequenced at the NGS platform, University of Bern. For *NOWA1/2/PGM*–KD and its control *EV/PGM*–KD, the Illumina DNA Nano library preparation protocol without size selection was used. Adapter ligation was followed by bead purification to remove the non-ligated adapters. The libraries were then amplified with a library size of > 200 bp (insert + adapters; refer to Fig EV5C). The sequencing was done at Fasteris, Genesupport SA (Switzerland).

Histograms of outer distances of PE reads (Fig EV5B and C) were generated for a single representative scaffold ("scaffold51_9_with_IES") from the reference *P. tetraurelia* strain 51 MAC + IES genome, bamPEFragmentSize (with switches: "--maxFragmentLength 500 -n 1000") from the deepTools2 (Ramírez *et al*, 2016) software package was used. To obtain bins of 1 bp, in deepTools2 bamPEFragmentSize.py, for the function "getDensity", the line to generate the histogram was changed from "n, bins, patches = plt.hist(lengths, bins=100, range=(minVal, maxVal), density=True)" to "n, bins, patches = plt.hist(lengths, bins=range(-maxVal), range=(minVal, maxVal), density=True)".

DNase-seq analyses

For general nucleosome density distribution analyses, we use HISAT2 (Kim *et al*, 2019) for read mapping of nucleosomal and new MAC DNA preparations with parameters "--min-intronlen 24" and "--max-intronlen 20,000" to the reference *P. tetraurelia* strain 51 "MAC + IES" genome (Arnaiz *et al*, 2012). For nucleosome profiling, "properly paired" (as defined by the samtools (Li *et al*, 2009) flag "2") paired-end reads with an outer distance between 100 and 175 bp, in the range expected for mononucleosomes were selected for further analysis. Bedtools (Quinlan & Hall, 2010) was used to extract reads with at least 9 bp of sequences matching IESs with the parameters "-f 0.06 -split". htseq-count from the HTSeq package (Anders *et al*, 2015) was used to count IES-matching reads.

DNA-seq normalized IES nucleosome densities (dimensionless quantities, since IES length normalizations for DNA-seq and nucleosome profiling, cancel each other out), r_c and r_e (subscript c = control; subscript e = experiment), for each IES, IES_i (i = 1 to 44,925), were calculated according to the following:

$$r_c = (n_c/N_c) \div (d_c/D_c).$$

$$r_e = (n_e/N_e) \div (d_e/D_e).$$

Control = *ND7/PGM*-KD (for *ISWI1*) or *ND7/PGM*-KD (for *NOWA1/2*).

Experiment = *ISWI1/PGM*-KD or *NOWA1/2/PGM*-KD.

d_c = number of DNA reads mapped to a particular IES from control.

n_c = number of nucleosomal reads mapped to a particular IES from control.

d_e = number of DNA reads mapped to a particular IES from experiment.

n_e = number of nucleosomal reads mapped to a particular IES from experiment.

D_c = number of mapped DNA reads from control.

N_c = number of mapped nucleosomal reads from control.

D_e = downsampled number of mapped DNA reads from the experiment.

N_e = downsampled number of mapped nucleosomal reads from the experiment.

Since the number of reads between the libraries differed, we downsampled the larger ones to equivalent total numbers to the smaller ones, using the samtools (Li *et al*, 2009) version 1.7 command "samtools view -s" with the suitable fraction for the "-s" switch. The *ND7/PGM*–KD MAC DNA library (for ISWI1) was 0.7777 times the size of *ISWI1/PGM*–KD, and the *ISWI1/PGM*–KD nucleosomal library was 0.5875 times the size of *ND7/PGM*–KD. The total mapped IES read counts after downsampling were D_c = 939,549, D_e = 1,522,345; N_c = 1,017,091, N_e = 2,484,586. For *ND7/PGM*-KD (for NOWA1/2) and *NOWA1/2/PGM*-KD: D_c = 594,577, D_e = 860,348; N_c = 203,593, N_e = 231,555.

Note that the amount of IES DNA from the parental MAC is negligible compared to that from the knockdowns (compare Figs 2A vs. EV5A), both of which use the same nuclear isolation procedure. Thus, no explicit normalizations were applied to account for parental MAC DNA.

Calculations and the graphs generated are available as a Jupyter notebook (Dataset EV3; "DNase-seq_analysis.ipynb"), together with the necessary read count data (Dataset EV4, "ISWI1.IES_read_counts.txt" and, Dataset EV5, "NOWA1.IES_read_counts.txt") and IES retention score table (Dataset EV6, "ies_retention_plus_ISWI1.tab").

Data availability

The genomic datasets are available in the following databases:

- DNA-seq data: All raw sequencing data are available at the European Nucleotide Archive under the accession number PRJEB21344 (http://www.ebi.ac.uk/ena/data/view/PRJEB21344). Accession numbers for individual experiments are as follows:
  - DNA-seq for *ISWI1*-KD(a): ERR2010817
  - DNA-seq for *ISWI1*-KD(b): ERR2010816
  - DNA-seq for *PTCAF1*-KD: ERR2010818
  - DNA-seq for *ND7*-KD: ERR2010819
  - DNA-seq for *PGM/ND7*-KD: ERR2798685 DNA
  - DNA-seq for for *PGM/ISWI1*-KD: ERR2798686
  - DNase-seq for *PGM/ISWI1*-KD: ERR2798687
  - DNase-seq for *PGM/ND7*-KD: ERR2798688
  - DNA-seq for for *ND7/PGM*-KD (control for *NOWA1/2/PGM*-KD): ERS12021512
  - DNA-seq for *NOWA1/2/PGM*-KD: ERS12021513
  - DNase-seq for *NOWA1/2/PGM*-KD: ERS12021514
  - DNase-seq for *ND7/PGM*-KD (control for *NOWA1/2/PGM*-KD): ERS12021515

Expanded View for this article is available online.

Acknowledgements

We are thankful to Dr. Nasikhat Stahlberger for her technical support and to all the members of the Nowacki lab for their suggestions and fruitful discussions. We thank Christiane Emmerich for her technical support. We are also very thankful to Prof. Dr. Manfred Heller and Dr. Anne-Christine Uldry for their help with mass spectrometry and data analysis. This research was supported by grants from the European Research Council (ERC): 260358 "EPIGENOME" and 681178 "G-EDIT," Swiss National Science Foundation: 31003A_146257 and 31003A_166407 and from the National Center of Competence in Research (NCCR) RNA and Disease to MN, and the Max Planck Society (ECS).

Author contributions

Aditi Singh: Conceptualization; data curation; formal analysis; validation; investigation; visualization; methodology; writing – original draft; writing – review and editing. **Xyrus X Maurer-Alcalá:** Formal analysis; validation; investigation; writing – original draft. **Therese Solberg:** Validation; investigation; writing – review and editing. **Lilia Häußermann:** Validation; investigation; writing – review and editing. **Silvan Gisler:** Validation; investigation; writing – review and editing. **Michael Ignarski:** Validation; investigation; writing – review and editing. **Estienne C Swart:** Resources; data curation; formal analysis; supervision; funding acquisition; validation; investigation; visualization; methodology; writing – review and editing. **Mariusz Nowacki:** Conceptualization; resources; supervision; funding acquisition; methodology; project administration; writing – review and editing.

Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

References

Abello A, Régnier V, Arnaiz O, Le Bars R, Bétermier M, Bischerour J (2020) Functional diversification of paramecium Ku80 paralogs safeguards genome integrity during precise programmed DNA elimination. *PLoS Genet* 16: e1008723

Anders S, Pyl PT, Huber W (2015) HTSeq — a python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166–169

Arambasic M, Sandoval PY, Hoehener C, Singh A, Swart EC, Nowacki M (2014) Pdsg1 and Pdsg2, novel proteins involved in developmental genome remodelling in paramecium. *PLoS ONE* 9: e112899

Arnaiz O, Goût J-F, Bétermier M, Bouhouche K, Cohen J, Duret L, Kapusta A, Meyer E, Sperling L (2010) Gene expression in a paleopolyploid: a transcriptome resource for the ciliate *Paramecium tetraurelia*. *BMC Genomics* 11: 547

Arnaiz O, Mathy N, Baudry C, Malinsky S, Aury J-M, Denby Wilkes C, Garnier O, Labadie K, Lauderdale BE, Le Mouël A *et al* (2012) The *Paramecium* germline genome provides a niche for intragenic parasitic DNA:

evolutionary dynamics of internal eliminated sequences. *PLoS Genet* 8: e1002984

Arnaiz O, Sperling L (2011) *Paramecium*DB in 2011: new tools and new data for functional and comparative genomics of the model ciliate *Paramecium tetraurelia*. *Nucleic Acids Res* 39: D632–D636

Aury J-M, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, Anthouard V, Aiach N *et al* (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444: 171–178

Bartholomew B (2014) Regulating the chromatin landscape: structural and mechanistic perspectives. *Annu Rev Biochem* 83: 671–696

Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, Bétermier M (2009) PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev* 23: 2478–2483

Beisson J, Bétermier M, Bré M-H, Cohen J, Duharcourt S, Duret L, Kung C, Malinsky S, Meyer E, Preer JR *et al* (2010a) *Paramecium tetraurelia*: the renaissance of an early unicellular model. *Cold Spring Harb Protoc* 2010: pdb.emo140

Beisson J, Bétermier M, Bré M-H, Cohen J, Duharcourt S, Duret L, Kung C, Malinsky S, Meyer E, Preer JR *et al* (2010b) Maintaining clonal *Paramecium tetraurelia* cell lines of controlled age through daily reisolation. *Cold Spring Harb Protoc* 2010: pdb.prot5361

Beisson J, Bétermier M, Bré M-H, Cohen J, Duharcourt S, Duret L, Kung C, Malinsky S, Meyer E, Preer JR *et al* (2010c) Silencing specific *Paramecium tetraurelia* genes by feeding double-stranded RNA. *Cold Spring Harb Protoc* 2010: pdb.prot5363

Beisson J, Bétermier M, Bré M-H, Cohen J, Duharcourt S, Duret L, Kung C, Malinsky S, Meyer E, Preer JR *et al* (2010d) DNA microinjection into the macronucleus of *Paramecium*. *Cold Spring Harb Protoc* 2010: pdb.prot5364

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol* 57: 289–300

Bischerour J, Bhullar S, Denby Wilkes C, Régnier V, Mathy N, Dubois E, Singh A, Swart E, Arnaiz O, Sperling L *et al* (2018) Six domesticated PiggyBac transposases together carry out programmed DNA elimination in *Paramecium*. *eLife* 7: e37927

Bouhouche K, Gout J-F, Kapusta A, Bétermier M, Meyer E (2011) Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res* 39: 4249–4264

Brown T (2001) Dot and slot blotting of DNA. *Curr Protoc Mol Biol* Chapter 2: Unit 2.9B

Cerón-Romero MA, Maurer-Alcalá XX, Grattepanche J-D, Yan Y, Fonseca MM, Katz LA (2019) PhyloToL: A taxon/gene-rich phylogenomic pipeline to explore genome evolution of diverse eukaryotes. *Mol Biol Evol* 36: 1831–1842

Chalker DL, Yao M-C (2011) DNA elimination in ciliates: transposon domestication and genome surveillance. *Annu Rev Genet* 45: 227–246

Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem* 78: 273–304

Corona DF, Längst G, Clapier CR, Bonte EJ, Ferrari S, Tamkun JW, Becker PB (1999) ISWI is an ATP-dependent nucleosome remodeling factor. *Mol Cell* 3: 239–245

Corona DFV, Siriaco G, Armstrong JA, Snarskaya N, McClymont SA, Scott MP, Tamkun JW (2007) ISWI regulates higher-order chromatin structure and histone H1 assembly in vivo. *PLoS Biol* 5: e232

Coyne RS, Lhuillier-Akakpo M, Duharcourt S (2012) RNA-guided DNA rearrangements in ciliates: Is the best genome defence a good offence? *Biol Cell* 104: 309–325

Denby Wilkes C, Arnaiz O, Sperling L (2016) ParTIES: A toolbox for *Paramecium* interspersed DNA elimination studies. *Bioinformatics* 32: 599–601

Deutsch EW, Bandeira N, Sharma V, Perez-Riverol Y, Carver JJ, Kundu DJ, García-Seisdedos D, Jarnuczak AF, Hewapathirana S, Pullman BS *et al* (2020) The ProteomeXchange consortium in 2020: enabling "big data" approaches in proteomics. *Nucleic Acids Res* 48: D1145–D1152

Dirscherl SS, Krebs JE (2004) Functional diversity of ISWI complexes. *Biochem Cell Biol* 82: 482–489

Drews F, Salhab A, Karunanithi S, Cheaib M, Jung M, Schulz MH, Simon M (2022) Broad domains of histone marks in the highly compact *Paramecium* macronuclear genome. *Genome Res* 32: 710–725

Dubois E, Bischerour J, Marmignon A, Mathy N, Régnier V, Bétermier M (2012) Transposon invasion of the *Paramecium* germline genome countered by a domesticated piggybac transposase and the NHEJ pathway. *Int J Evol Biol* 2012: 436196

Duret L, Cohen J, Jubin C, Dessen P, Gout J-F, Mousset S, Aury J-M, Jaillon O, Noël B, Arnaiz O *et al* (2008) Analysis of sequence variability in the macronuclear DNA of *Paramecium tetraurelia*: a somatic view of the germline. *Genome Res* 18: 585–596

Erdel F, Krug J, Längst G, Rippe K (2011) Targeting chromatin remodelers: signals and search mechanisms. *Biochim Biophys Acta* 1809: 497–508

Fillingham JS, Garg J, Tsao N, Vythilingum N, Nishikawa T, Pearlman RE (2006) Molecular genetic analysis of an SNF2/brahma-related gene in *Tetrahymena thermophila* suggests roles in growth and nuclear development. *Eukaryot Cell* 5: 1347–1359

Frapporti A, Miró Pina C, Arnaiz O, Holoch D, Kawaguchi T, Humbert A, Eleftheriou E, Lombard B, Loew D, Sperling L *et al* (2019) The Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in *Paramecium*. *Nat Commun* 10: 2710

Furrer DI, Swart EC, Kraft MF, Sandoval PY, Nowacki M (2017a) European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/text-search?query=ERS1656548). [DATASET]

Furrer DI, Swart EC, Kraft MF, Sandoval PY, Nowacki M (2017b) Two sets of Piwi proteins are involved in distinct sRNA pathways leading to elimination of germline-specific DNA. *Cell Rep* 20: 505–520

Gnan S, Matelot M, Weiman M, Arnaiz O, Guérin F, Sperling L, Bétermier M, Thermes C, Chen C-L, Duharcourt S (2022) GC content, but not nucleosome positioning, directly contributes to intron splicing efficiency in paramecium. *Genome Res* 32: 699–709

Hamilton EP, Kapusta A, Huvos PE, Bidwell SL, Zafar N, Tang H, Hadjithomas M, Krishnakumar V, Badger JH, Caler EV *et al* (2016) Structure of the germline genome of *Tetrahymena thermophila* and relationship to the massively rearranged somatic genome. *eLife* 5: e19090

Huang B, Pan X, Li Z, Wang Z, Yu Y, Dou Z (2012) Polycomb group proteins and their roles in regulating stem cell development. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao* 34: 281–285

Ignarski M, Singh A, Swart EC, Arambasic M, Sandoval PY, Nowacki M (2014) *Paramecium tetraurelia* chromatin assembly factor-1-like protein PtCAF-1 is involved in RNA-mediated control of DNA elimination. *Nucleic Acids Res* 42: 11952–11964

Kammers K, Cole RN, Tiengwe C, Ruczinski I (2015) Detecting significant changes in protein abundance. *EuPA Open Proteom* 7: 11–19

Kapusta A, Matsuda A, Marmignon A, Ku M, Silve A, Meyer E, Forney JD, Malinsky S, Bétermier M (2011) Highly precise and developmentally programmed genome assembly in *Paramecium* requires ligase IV-dependent end joining. *PLoS Genet* 7: e1002049

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30: 772–780

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37: 907–915

Lepère G, Nowacki M, Serrano V, Gout J-F, Guglielmi G, Duharcourt S, Meyer E (2009) Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res* 37: 903–915

Lhuillier-Akakpo M, Frapporti A, Denby Wilkes C, Matelot M, Vervoort M, Sperling L, Duharcourt S (2014a) European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/view/ERA309409). [DATASET]

Lhuillier-Akakpo M, Frapporti A, Denby Wilkes C, Matelot M, Vervoort M, Sperling L, Duharcourt S (2014b) Local effect of enhancer of zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements. *PLoS Genet* 10: e1004665

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079

Li Z, Gadue P, Chen K, Jiao Y, Tuteja G, Schug J, Li W, Kaestner KH (2012) Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. *Cell* 151: 1608–1616

Maliszewska-Olejniczak K, Gruchota J, Gromadka R, Denby Wilkes C, Arnaiz O, Mathy N, Duharcourt S, Bétermier M, Nowak JK (2015) TFIIS-dependent non-coding transcription regulates developmental genome rearrangements. *PLoS Genet* 11: e1005383

Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. *Nature* 469: 343–349

Miró-Pina C, Charmant O, Kawaguchi T, Holoch D, Michaud A, Cohen I, Humbert A, Jaszczyszyn Y, Chevreux G, Del Maestro L *et al* (2022) Paramecium Polycomb repressive complex 2 physically interacts with the small RNA-binding PIWI protein to repress transposable elements. *Dev Cell* 57: 1037–1052

Nowacki M, Zagorski-Ostoja W, Meyer E (2005) Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in paramecium tetraurelia. *Curr Biol* 15: 1616–1628

Pazin MJ, Kadonaga JT (1997) SWI2/SNF2 and related proteins: ATP-driven motors that disrupt protein-DNA interactions? *Cell* 88: 737–740

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M *et al* (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47: D442–D450

Pulivarthy SR, Lion M, Kuzu G, Matthews AGW, Borowsky ML, Morris J, Kingston RE, Dennis JH, Tolstorukov MY, Oettinger MA (2016) Regulated large-scale nucleosome density patterns, and precise nucleosome positioning correlate with V(D)J recombination. *Proc Natl Acad Sci USA* 113: E6427–E6436

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T (2016) deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44: W160–W165

Rando OJ, Winston F (2012) Chromatin and transcription in yeast. *Genetics* 190: 351–387

Sadeh R, Allis CD (2011) Genome-wide "re"-modeling of nucleosome positions. *Cell* 147: 263–266

Sandoval PY, Swart EC, Arambasic M, Nowacki M (2014a) European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/view/SRX309864?show=reads). [DATASET]

Sandoval PY, Swart EC, Arambasic M, Nowacki M (2014b) Functional diversification of dicer-like proteins and small RNAs required for genome sculpting. *Dev Cell* 28: 174–188

Sela I, Ashkenazy H, Katoh K, Pupko T (2015) GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res* 43: W7–W14

Skouri F, Cohen J (1997) Genetic approach to regulated exocytosis using functional complementation in paramecium: Identification of the ND7 gene required for membrane fusion. *Mol Biol Cell* 8: 1063–1071

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313

Swart EC, Denby Wilkes C, Sandoval PY, Hoehener C, Singh A, Furrer DI, Arambasic M, Ignarski M, Nowacki M (2017a) Identification and analysis of functional associations among natural eukaryotic genome editing components [version 1; peer review: 1 approved, 1 approved with reservations]. *F1000Res* 6: 1374

Swart EC, Denby Wilkes C, Sandoval PY, Hoehener C, Singh A, Furrer DI, Arambasic M, Ignarski M, Nowacki M (2017b) European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/view/PRJEB12406?show=reads). [DATASET]

Swart EC, Wilkes CD, Sandoval PY, Arambasic M, Sperling L, Nowacki M (2014) Genome-wide analysis of genetic and epigenetic control of programmed DNA deletion. *Nucleic Acids Res* 42: 8970–8983

Wasmuth EV, Lima CD (2017) The Rrp6 C-terminal domain binds RNA and activates the nuclear RNA exosome. *Nucleic Acids Res* 45: 846–860

Xiao H, Sandaltzopoulos R, Wang HM, Hamiche A, Ranallo R, Lee KM, Fu D, Wu C (2001) Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol Cell* 8: 531–543

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

Yadon AN, Tsukiyama T (2011) SnapShot: chromatin remodeling: ISWI. *Cell* 144: 453

Expanded View Figures

Figure EV1. Knockdown effects of ISWI paralogs in *Paramecium tetraurelia*.

A, B Northern blot analysis using ISWI1-specific and ISWI5-specific probes, respectively. rRNA probe was used as a loading control against ribosomal RNA. Early: ~50% of cells with fragmented parental macronucleus; Late: the majority of cells with a visible anlagen. *ND7*-KD is used as a control to confirm mRNA expression.

C–F (C and E) Survival test graph. Dead cells are represented in red, sick in orange and cells diving at a normal rate in green. *PGM*–KD is used as a positive control, and *ND7*–KD as a negative control. (D and F) IES retention PCR (cropped inverted images). Five maternally controlled IES and five non-maternally controlled IESs are shown. The IES+ band represents retained IES; the IES-band represents excised IES; additional bands are likely PCR artifacts or primer dimers.

Source data are available online for this figure.

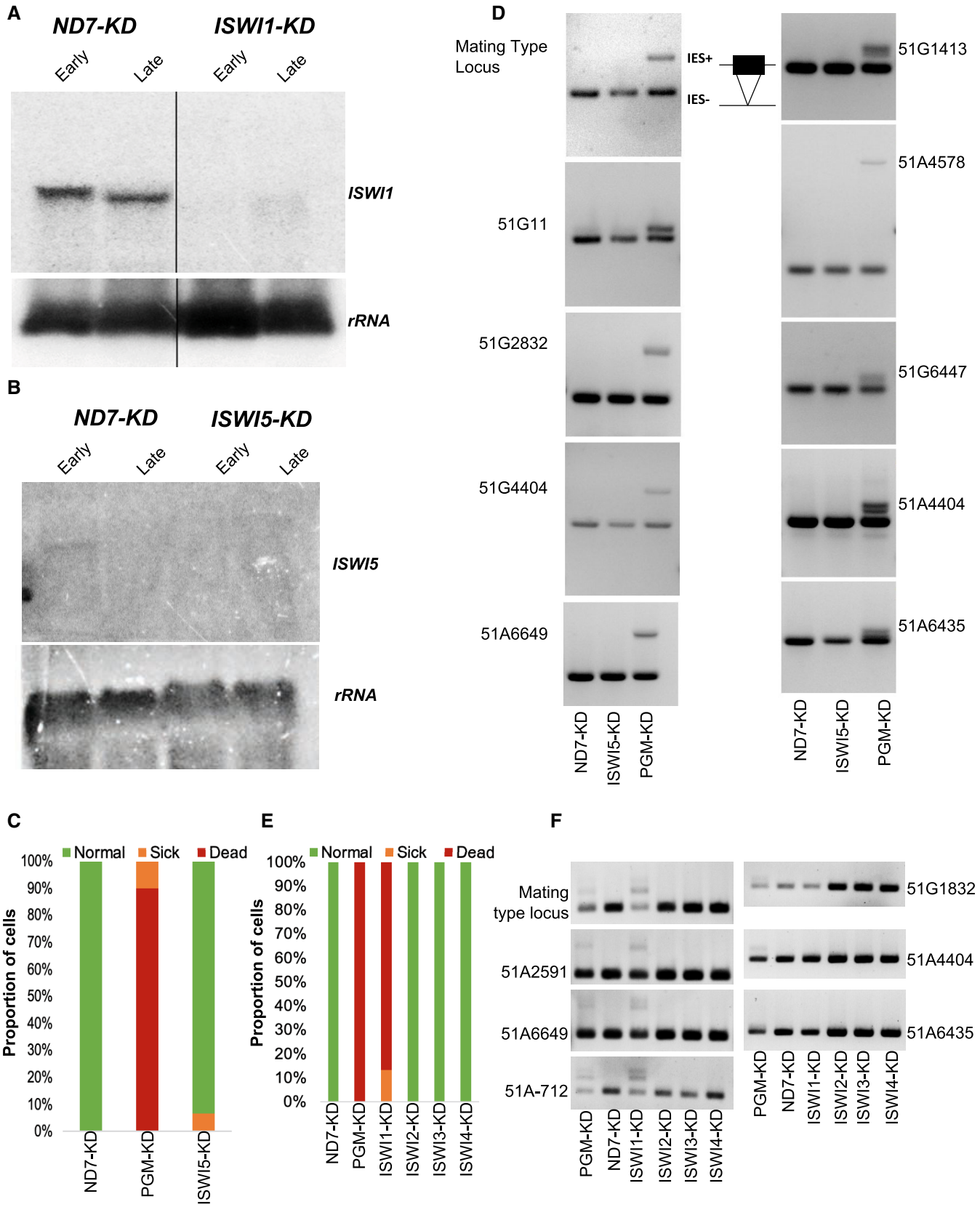


Figure EV1.

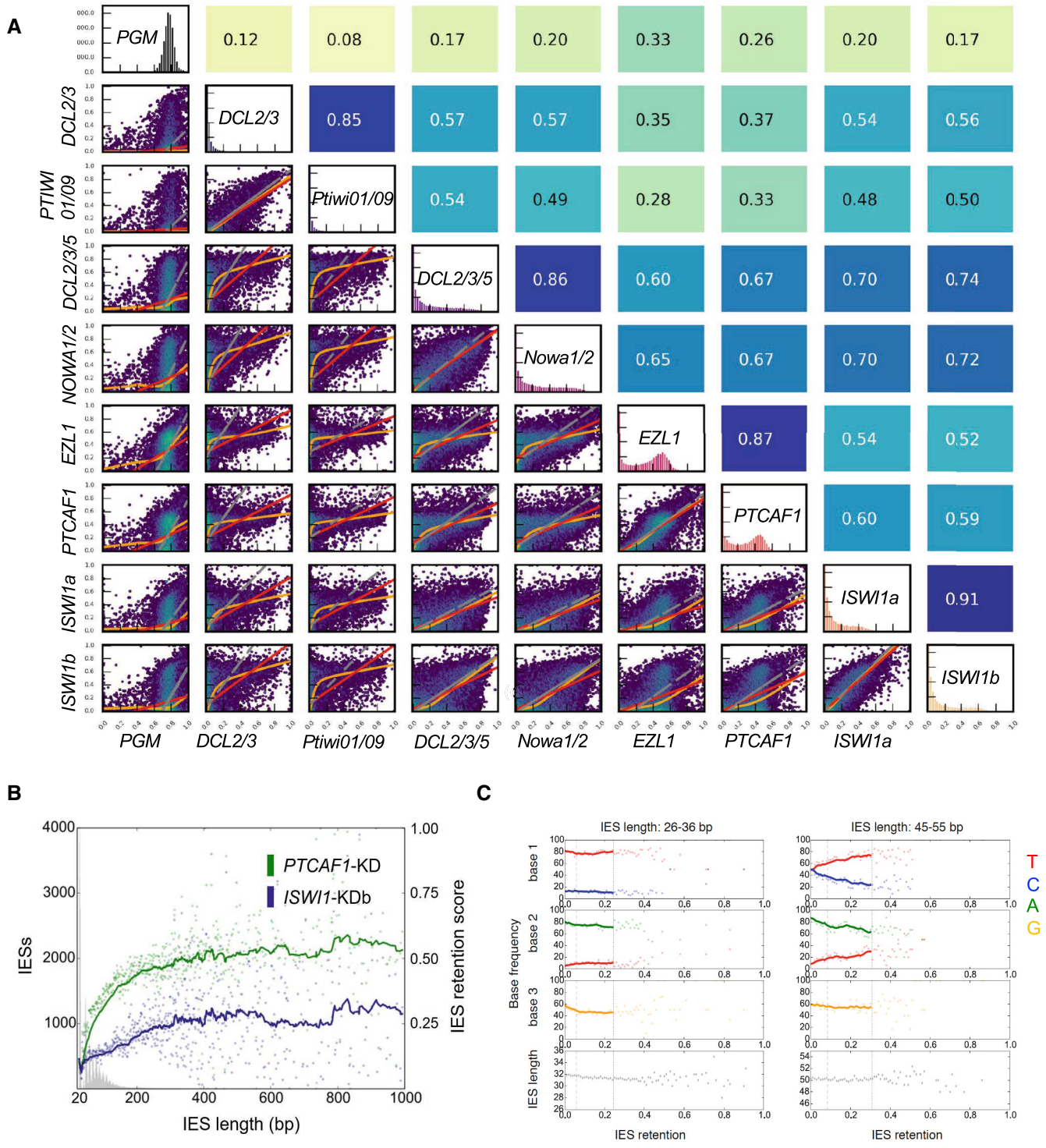


Figure EV2. Relationship between IES retention scores, IES length, and base sequences.

A Relationships in IES retention among knockdown pairs. Hexagonal binning of IES retention scores was used to generate the plots. Pearson's correlation coefficients are given above each subgraph. Red lines are for ordinary least-squares (OLS) regression, orange lines are for LOWESS, and gray lines are for orthogonal distance regression (ODR).

B IRSs versus IES length as described previously.

C Base frequencies of the first three bases after the TA repeat relative to the IRS of *ISWI1*-KDb from the first and third *Paramecium tetraurelia* IES length peak.

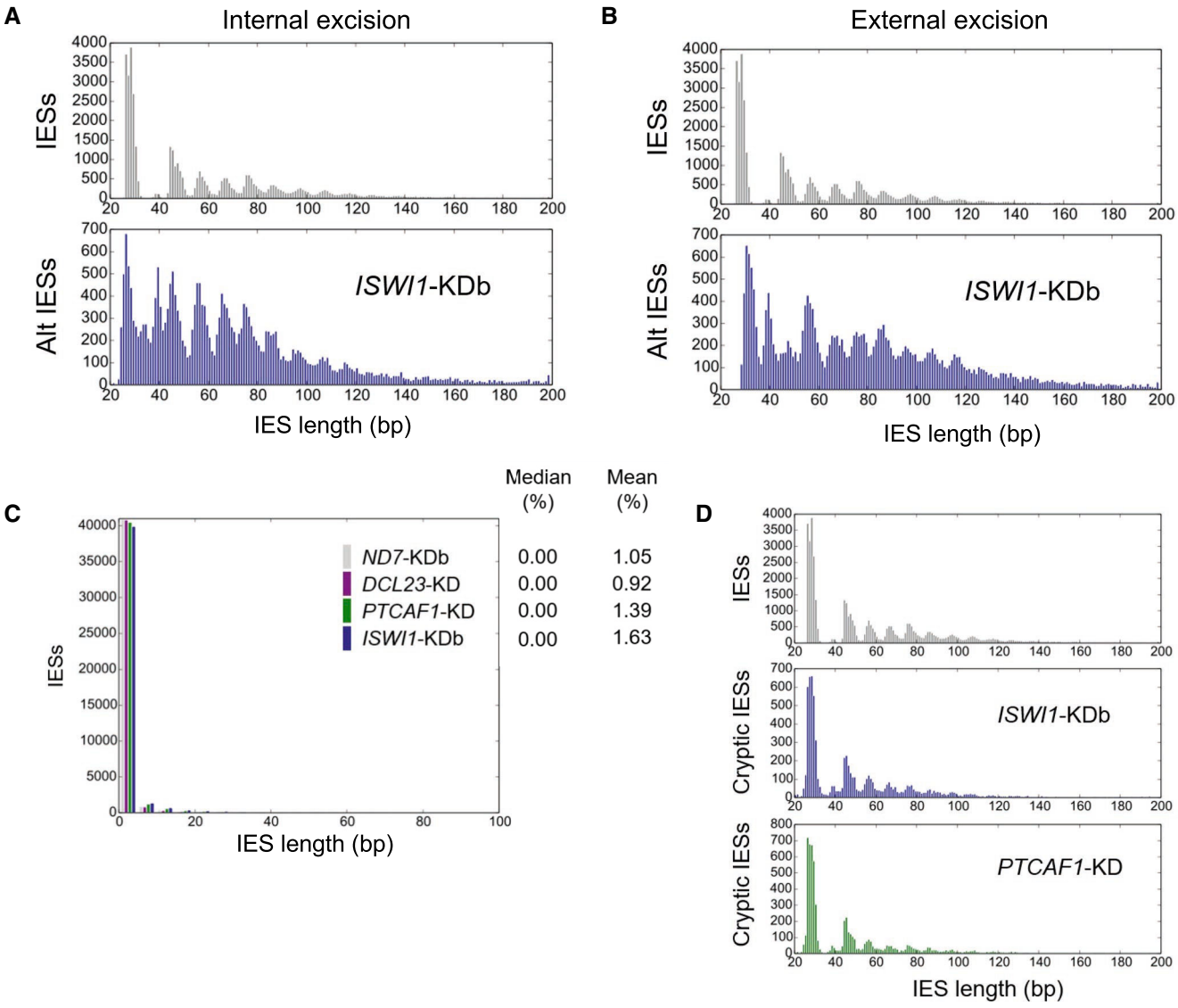
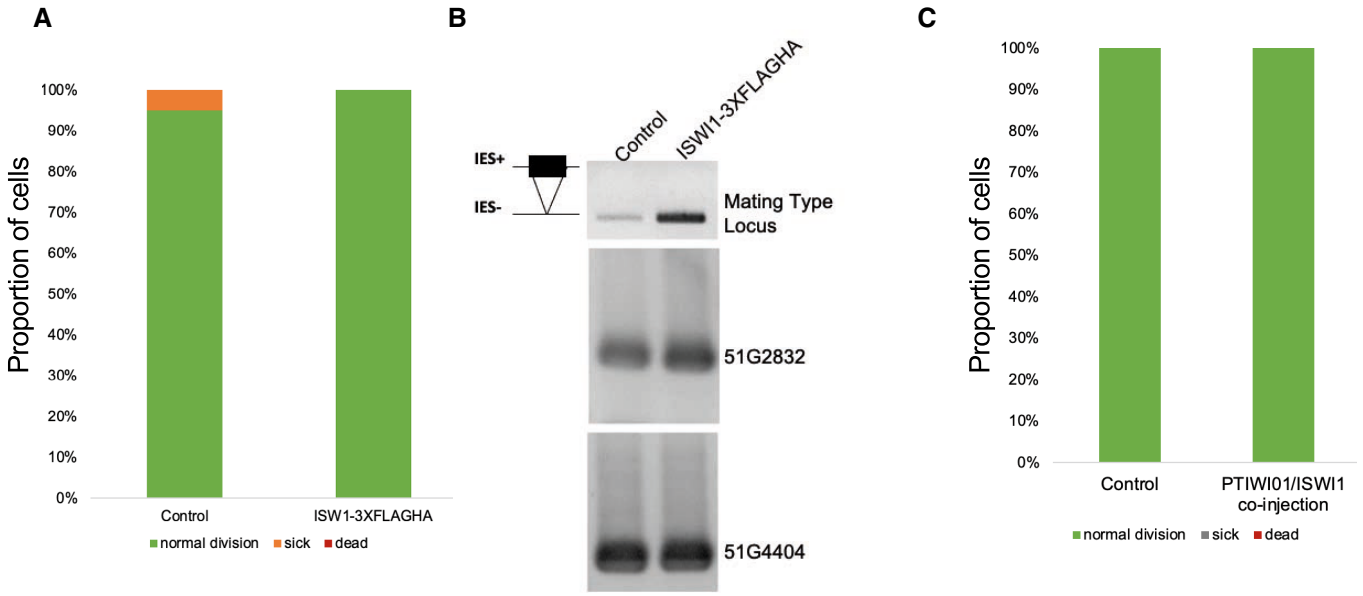
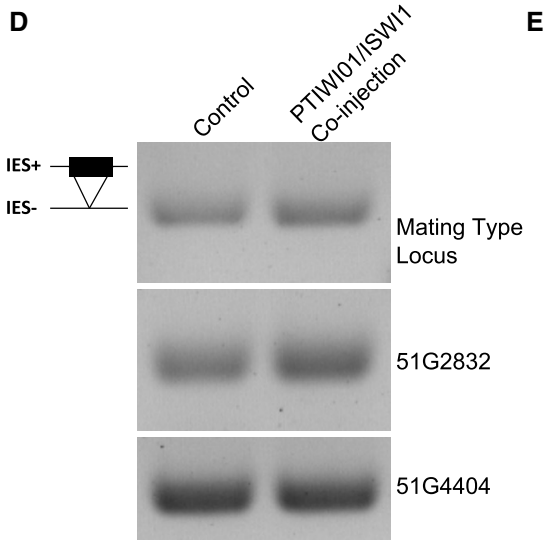


Figure EV3. Size distribution of Alternatively excised IESs and Cryptic IESs.

A Length distribution of internally excised alternative (Alt) IES boundaries.
B Length distribution of externally alternative (Alt) excised IES boundaries, respectively.
C Genome-wide analysis of cryptic IES excision. Cryptic excision (%) = 100*(cryptically excised reads)/(all reads).
D Length distribution of cryptically excised IES.





E

Majority protein IDs	Peptide Count ISWI1	Peptide count ISWI2	Peptide count WT1	Peptide count WT2	*p* BH adjusted
ISWI1	301	329	0	0	n.d.
Uncharacterized protein	90	104	0	0	n.d.
DDRP	62	61	0	0	n.d.
SPT6	49	45	0	0	n.d.
PTIWI01	26	43	2	3	0,004
NOWA2	11	9	0	0	n.d.

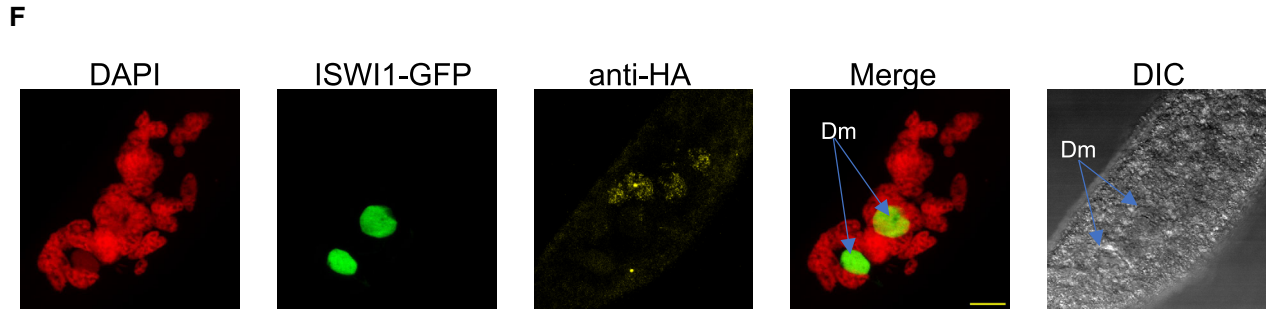


Figure EV4.

Figure EV4. Overexpression of fusion proteins do not show any adverse effects.

A–D (A and C) Survival test graph. Dead cells are represented in red, sick in orange and cells diving at a normal rate in green. (B and D) IES retention PCR (cropped inverted images). Mating type, 51G2832, and 51G4404 IESs are shown. The IES+ band represents retained IES; the IES- band represents excised IES; additional bands are likely PCR artifacts or primer dimers.

E Most abundant proteins in the ISWI1-3XFLAGHA MS analysis ISWI1 & ISWI2 are two biological replicates for ISWI1-3XFLAGHA, while WT1 and WT2 are biological replicates for control in MS analysis. Peptide count refers to the number of peptides detected in MS. Adjusted *P*-values were calculated following the Benjamini and Hochberg correction for multiple testing.

F ISWI1-GFP localization to developing macronucleus seen in green during development; merge is an overlay of DAPI (red) staining parental and developing macronucleus, ISWI1-GFP (green), and anti-HA (yellow); scale bar = 10 μm.

Source data are available online for this figure.

Figure EV5. Nucleosome density measurements after DNase-seq.

A IES Retention Score (IRS) distributions for *PGM/ND7*-KD and *PGM/ISWI1*-KD. (B, C) Histograms of outer paired-end distances of mapped DNase-seq reads.

D–I Normalized nucleosome density histograms for IESs weakly (IRS < 0.2) or more strongly retained in *ISWI1*-Kdb (IRS ≥ 0.2), either for *ND7/PGM*-KD or *NOWA1/PGM*-KD. Kolmogorov–Smirnov statistics and their *P*-values are provided. Titles for graphs give criteria for IES selection.

Source data are available online for this figure.

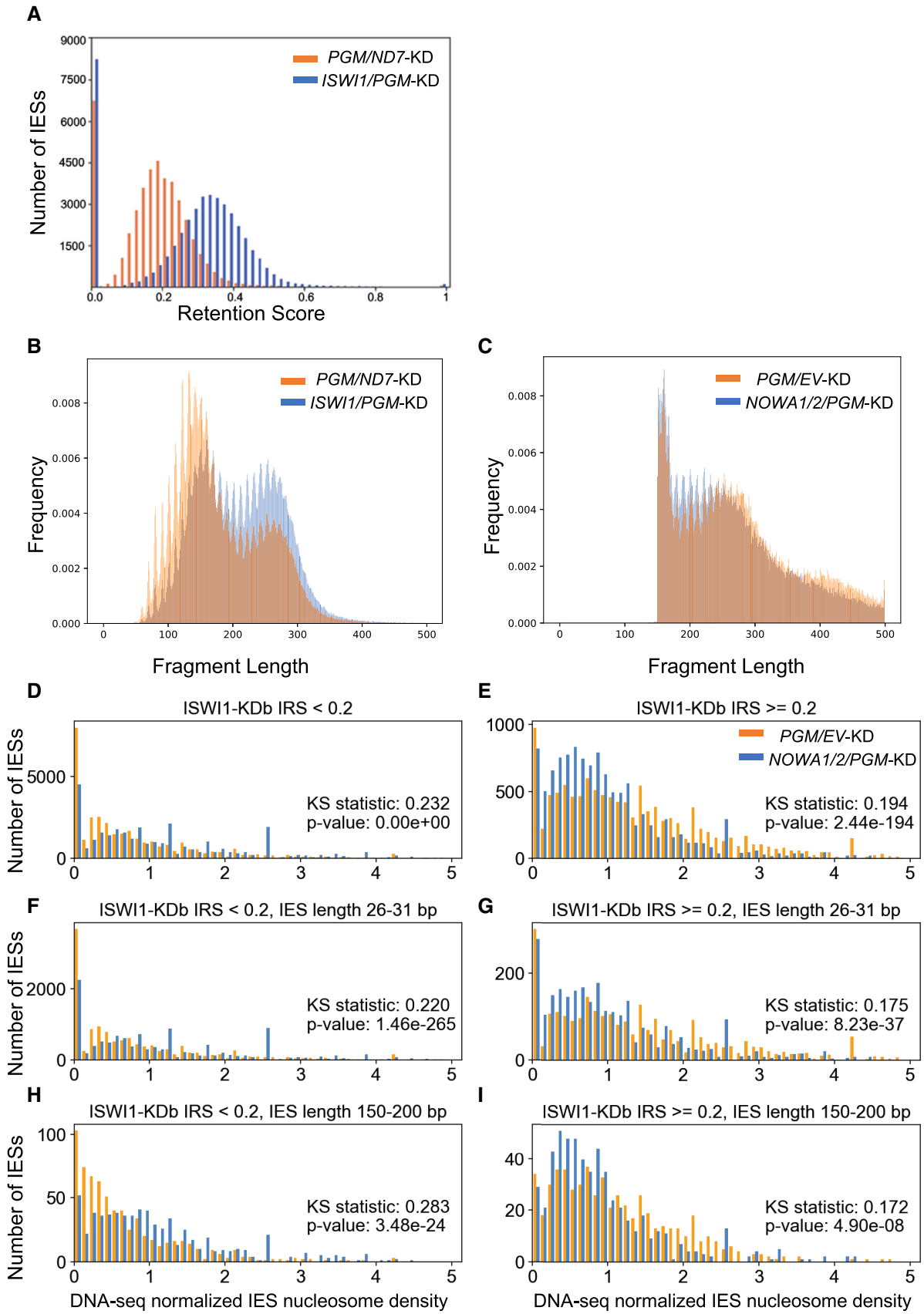


Figure EV5.

Appendix

Table of Contents **Page number**

Appendix Table S1

IES	Primer sequence (5' to 3' orientation)
51G--11F	ATCATAAGATTGATATCTTCTCCCTTCTCC
51G--11R	ACTTGCTACTAAAGCAAGAAACATTGAGAG
51G1413F	GAAGCTGCTTGTGTTAAGAATTCTACTGG
51G1413R	GCATCCAGCACTAGTTGAATTTACTGTAC
51G1832F	CTATAACTCTTGAAGCTGCTTGTAATATG
51G1832R	TTGTCAATGAGCCATTAACAGTTGCTGGAT
51G2832F	GAGCAGGATGTACAAATACTGGTGG
51G2832R	AGCTGATTAGATAACAATACAACCAGTACC
51G4404F	CTGTTGCTACACATTGTGCATATGTTACT
51G4404R	GCTGTAAGATTAACATTGAGCATGATCAAG
51G6447F	AATGCATCAAATGTAGTAACTACTCCTGCT
51G6447R	AATTTGTAAAGTATCCAGCGCAGGCAG
MT Locus F	GGTGTTTATATCTTAATTGTTGACCCTCAC
MT Locus R	CCATCTATACTCCATTCTTTATCTTAATTCAT
51A--712F	TTTGTCAAAAAGACATGTATCAAAATGCAG
51A--712R	TAGAATACTAAGAGATTCAATACAACAAAC
51A1835F	TAATGTATTGATAAGGCTTGCTCTACAGCC
51A1835R	ATCCTAACATCCTTGAATAGTTACTGATCC
51A2591F	ATGTGTTTGGACTGGATTGGCATGTAGAAG
51A2591R	GATGTAGCATAACATTTATCAACAATCCAT
51A4404F	TGGAATAGTGCTGCATCACCAGCTGCTTGC
51A4404R	CCAGTTATTGAACTGCAACTTACTGCAGTG
51A4578F	CACTGCAGTAAGTTGCAGTTCAATAACTGG
51A4578R	TGTAGTCTTAAAATCTTAGCATGTTGTACC
51A6435F	CAAATTGTGTCACTAGAGGTACATGTTTCC
51A6435R	GCGACATCAATAGTAACAGCTGAGCATGAG
51A6649F	ACTGCACCTCTAACTTTAACAAGCGAAGCA
51A6649R	CAGCAGTACATCCAGCTCTCTAAGTTTAGC
51-429F	GTT GGA TAT GCA TCC ACA TC
51-429R	CTG CTT CGA TAT GCA TAA GAA AG

Appendix Table S1: Primers used in IES retention PCRs. F: forward primer; R: reverse primer.

Appendix Table S2

Majority protein IDs	MS/MS count ISWI1	MS/MS count ISWI2	MS/MS count WT1	MS/MS count WT2	Entrez_Protein	Putative Description
PTET.51.1.P0140243	243	261	0	0	XP_001431605	ISWI1 (this manuscript)
PTET.51.1.P0440186	90	104	0	0	XP_001447805.1	Uncharacterized
PTET.51.1.P0140243	58	68	0	0	XP_001431606	ISWI1 (C-terminal)
PTET.51.1.P1370127	62	61	0	0	XP_001431060	DDRP
PTETP2700007001	61	50	0	0	XP_001441411	NAD () ADP-ribosyltransferase-3
PTETP300037001	49	45	0	0	XP_001442677	SPT6
PTET.51.1.P0420126	42	49	0	0	XP_001447124	ISWI2 (this manuscript
PTETP7100004001	26	43	2	3	XP_001456124	PTIWI01
PTET.51.1.P0180124	33	36	0	0	XP_001437349	Uncharacterized
PTET.51.1.P0070284	39	28	0	0	XP_001455634	Poly (ADP-ribose) Polymerase
PTET.51.1.P0480005	34	29	2	0	XP_001448777	DDRP2
PTETP10500011001	31	33	0	0	XP_001424861	SPT16
PTET.51.1.P0110326	19	26	3	8	XP_001425954	S-adenosyl methionine synthetase
PTET.51.1.P0330075	26	20	0	0	XP_001443869	SMC2
PTET.51.1.P0560063	17	22	2	2	XP_001451492	RuvB-like 2
PTET.51.1.P0720184	19	20	0	1	XP_001456414	Exportin
PTET.51.1.P0990120	17	22	0	0	XP_001462544	MCM6
PTETP13700005001	19	17	0	0	XP_001431029	PARP
PTET.51.1.P0040036	18	17	0	0	XP_001446004	sequence specific DNA binidng
PTET.51.1.P1720081	19	15	0	0	XP_001436697	Myb-like protein
PTET.51.1.P0610231	17	16	0	0	XP_001453395	structural constituent of ribososme
PTET.51.1.P0380048	17	15	0	0	XP_001445418	SPT5v
PTET.51.1.P0200322	14	16	0	0	XP_001439097	PHD-type
PTET.51.1.P1060046	11	16	0	0	XP_001425020	WD40-repeat like
PTET.51.1.P0030037	9	15	2	1	XP_001442334	RuvB-like 1
PTET.51.1.P1600114	9	16	0	0	XP_001435111	Erythrocyte membrane protein 3
PTET.51.1.P0470202	10	15	0	0	XP_001448725	Helicase W08D27
PTET.51.1.P0410183	12	13	0	0	XP_001446881	Uncharacterized protein
PTET.51.1.P1180147	7	16	0	0	XP_001427491	MCM2
PTET.51.1.P006001	9	10	3	0	XP_001452484	Uncharacterized protein
PTET.51.1.P0620215	10	11	0	0	XP_001453643	PTIWI11
PTET.51.1.P0170077	11	9	0	0	XP_001449252	NOWA2
PTET.51.1.P0400053	12	8	0	0	XP_001436180	DNA directed DNA polymerase
PTET.51.1.P0450314	6	12	0	0	XP_001448195	Uncharacterized protein
PTET.51.1.P0080258	8	10	0	0	XP_001458158	Uncharacterized protein

Appendix Table S2: Mass spectrometry analysis of ISWI1-3XFLAGHA co-immunoprecipitation (IP). Majority protein IDs correspond to the *Paramecium* Database (Arnaiz and Sperling, 2011) accession numbers of the proteins identified by MS. MS/MS count ISWI1 & MS/MS count ISWI1 represents total peptide count in ISWI1-3XFLAGHA IP replicates. MS/MS count WT1 & MS/MS WT2 represents total peptide count in negative control to IP. MS/MS count represents combined peptide counts in the replicates. Putative description is retrieved from Paramecium Database (https://paramecium.i2bc.paris-saclay.fr/)

Appendix Table S3

IES ID	IES Name	*PGM* -KD	*DCL2/3* -KD	*DCL5*KD	*DCL2/3/5* -KD	*NOWA1/2* -KD	*ISWI1* -KD
IESPGM.PTET51.1.51.451201	51G-11	0.79	0.05	0.19	0.56	0.53	0.48
IESPGM.PTET51.1.51.452624	51G1413	0.78	0.01	0.01	0.04	0.15	0.07
IESPGM.PTET51.1.51.453043	51G1832	0.73	0.00	0.00	0.02	0.02	0.04
IESPGM.PTET51.1.51.454043	51G2832	0.77	0.22	0.00	0.71	0.59	0.20
IESPGM.PTET51.1.51.455615	51G4404	0.84	0.60	0.00	0.75	0.80	0.58
IESPGM.PTET51.1.51.457658	51G6447	0.77	0.04	0.00	0.00	0.05	0.08
IESPGM.PTET51.1.106.281631	51A-712	0.78	0.07	0.05	0.71	0.75	0.62
IESPGM.PTET51.1.106.284157	51A1835	0.80	0.00	0.00	0.04	0.03	0.03
IESPGM.PTET51.1.106.284913	51A2591	0.91	0.54	0.00	0.80	0.81	0.68
IESPGM.PTET51.1.106.286750	51A4404	0.82	0.00	0.00	0.00	0.00	0.23
IESPGM.PTET51.1.106.286924	51A4578	0.77	0.04	0.00	0.07	0.20	0.10
IESPGM.PTET51.1.106.288781	51A6435	0.77	0.00	0.00	0.00	0.01	0.04
IESPGM.PTET51.1.106.288995	51A6649	0.81	0.55	0.01	0.73	0.73	0.61

Appendix Table S3: Comparison of IES retention scores. IESs are those that were used to test retention using standard primers (Table T1) against IESs described in (Duharcourt et al. 1998).

ISWI1 complex proteins facilitate developmental genome editing in *Paramecium*

Aditi Singh[1,*,#], Lilia Häußermann[1,*], Christiane Emmerich[1], Emily Nischwitz[2], Brandon KB Seah[1], Falk Butter[2,3], Mariusz Nowacki[4], Estienne C. Swart[1,#]

[1]Max Planck Institute for Biology, Max-Planck-Ring 5, 72076, Tübingen, Germany

[2]Institute of Molecular Biology, Ackermannweg 4, 55128 Mainz, Germany

[3]Institute of Molecular Virology and Cell Biology (IMVZ), Friedrich Loeffler Institut, Greifswald, Germany

[4]Institute of Cell Biology, University of Bern, Baltzerstrasse 4, 3012, Bern, Switzerland

[*]equal contribution

[#]Corresponding author: estienne.swart@tuebingen.mpg.de, aditi.singh@tuebingen.mpg.de

Abstract

One of the most extensive forms of natural genome editing occurs in ciliates, a group of microbial eukaryotes. Ciliate germline and somatic genomes are contained in distinct nuclei within the same cell. During the massive reorganization process of somatic genome development, ciliates eliminate tens of thousands of DNA sequences from a germline genome copy. Recently, we showed that the chromatin remodeler ISWI1 is required for somatic genome development in the ciliate *Paramecium tetraurelia*. Here, we describe two high similarity paralogous proteins, ICOP1 and ICOP2, essential for their genome editing. ICOP1 and ICOP2 are highly divergent from known proteins; the only domain detected showed distant homology to the WSD (WHIM2+WHIM3) motif. We show that both ICOP1 and ICOP2 interact with the chromatin remodeler ISWI1. Upon *ICOP* knockdown, changes in alternative DNA excision boundaries and nucleosome densities are similar to those observed for *ISWI1* knockdown. We thus propose that a complex comprising ISWI1 and either or both ICOP1 and ICOP2 are needed for *Paramecium*'s precise genome editing.

Keywords

ISWI, genome editing, chromatin remodelling, nucleosome, DNA

Introduction

Chromatin's underlying subunit, the nucleosome, comprises ~146 base pairs of DNA wrapped around an octamer of histone proteins (Kornberg 1977). The presence of a nucleosome alters DNA's geometry and physically shields it, affecting interactions with other DNA-binding proteins (Piña et al. 1990; Pryciak and Varmus 1992; Morgunova and Taipale 2021). The nucleosome thus participates in and regulates numerous molecular processes (Campos and Reinberg 2009; Bai and Morozov 2010; Alabert and Groth 2012; Price and D'Andrea 2013).

Nucleosomes can be moved, ejected, or reconstructed with alternative histone variants by four families of ATP-dependent chromatin remodelers (Clapier and Cairns 2009). The imitation switch (ISWI) family of chromatin remodelers forms several complexes capable of nucleosome sliding (Längst et al. 1999) in different organisms, each serving a distinct role. ISWI contains an N-terminal SNF2 ATPase domain that provides energy to move the nucleosome (Li et al. 2019). The C-terminal HAND-SANT-SLIDE domain (HSS) is essential for substrate recognition (Grüne et al. 2003). ISWI complex partners determine the context of the complex activity and alter its remodeling efficiency (Längst et al. 1999; Toto et al. 2014). ISWI complexes have been shown to regulate DNA replication, transcription, DNA repair, and V(D)J cleavage of polynucleosomal DNA (Patenge et al. 2004; Clapier and Cairns 2009; Aydin et al. 2014).

Like other ciliates, *Paramecium* has distinct nuclei: germline micronuclei (MICs) and the somatic macronucleus (MAC). MICs produce gametic nuclei that form a diploid zygotic nucleus, which generates new MICs and MACs. The zygotic genome developing into a new MAC genome undergoes massive editing, excising thousands of germline-limited sequences and also amplification to a high polyploidy (~800n) (Drews et al. 2022a; Zangarelli et al. 2022). *Paramecium*'s internal eliminated sequences (IESs) are distributed throughout the germline genome, including coding sequences (Arnaiz et al. 2012). IES removal thus requires precise excision and DNA repair (Kapusta et al. 2011; Dubois et al. 2012).

Each of *Paramecium*'s 45,000 unique IESs is flanked by conserved 5'-TA-3' dinucleotides, which are part of a less well-conserved ~5 bp terminal inverted repeat (Klobutcher and Herrick 1995; Arnaiz et al. 2012; Bischerour et al. 2018). PiggyMAC (PGM), a domesticated PiggyBac transposase (Baudry et al. 2009; Bischerour et al. 2018), is responsible for the excision of IESs and other germline-specific DNA. The IES length distribution monotonically declines with a characteristic 10/11 bp periodicity, except for a ~34-44 bp "forbidden" peak, where IESs appear largely absent, prevented by DNA's topological constraints and necessity of proper PGM subunit orientation for interaction (Arnaiz et al. 2012). IESs lack motifs for precise recognition and excision necessitating additional molecules for their removal.

Paramecium germline-limited sequences are thought to be targeted by two small non-coding RNA classes: scnRNAs and iesRNAs. scnRNAs are produced by Dicer-

like proteins Dcl2 and Dcl3 in the MICs and loaded on Piwi proteins Ptiwi01/09 and transported to the old MAC where “scanning” is thought to subtract non-IES matching molecules (Bouhouche et al. 2011); (Lepère et al. 2009; Sandoval et al. 2014). In the new MAC non-coding transcripts produced by RNA polymerase II with distinct, nucleus-specific TFIIS4 subunits, are proposed to bind to scnRNAs and facilitate IES targeting (Maliszewska-Olejniczak et al. 2015). iesRNAs, produced by Dcl5 and Ptiwi10/11 proteins, are proposed to form a positive feedback loop that efficiently excises most IES copies after IES excision onset (Sandoval et al. 2014; Furrer et al. 2017). As a general trend, shorter IESs tend to be older and primarily iesRNA- and scnRNA-independent, whereas younger, longer IESs require these and other molecules for excision (Sellis et al. 2021). Ptiwi01/09 was recently also suggested to interact with Polycomb repressive complex 2 (PRC2) that represses transposable elements (Miró-Pina et al. 2022; Wang et al. 2022), and also with ISWI1, assisting precise IES excision (Singh et al. 2022).

We recently showed that an ISWI homolog, ISWI1, is required for precise genome editing in *Paramecium tetraurelia* (henceforth, *Paramecium*) (Singh et al. 2022). ISWI1 depletion is lethal, leading to two distinct errors: failure of excision of numerous IESs, and excision of IESs at alternative TA boundaries (Singh et al. 2022). In the latter case, excision precision was proposed to be compromised by inappropriate nucleosome positioning. A distinctive characteristic of ISWI1 depletion is substantial alternative “forbidden” length IES excision. Here, we identified and investigated the contribution to IES excision of ISWI1 complex subunits.

Results

Identifying putative components of the ISWI1 complex

Previously, we performed co-immunoprecipitation (co-IP) of 3XFLAG-HA-tagged ISWI1 (Singh et al. 2022). After ISWI1, the most abundant protein candidate detected by protein mass spectrometry (MS), with more than five-fold enrichment in peptides identified relative to the input, is a 779 amino acid-long uncharacterised protein (ParameciumDB identifier: PTET.51.1.P0440186). The ohnolog of this protein (PTET.51.1.P0180124, 783 amino-acid long; 92.04% amino acid sequence identity) is also present in the subset of peptides identified as unique to ISWI1-IP replicates in the same MS dataset (Singh et al. 2022).

We next checked if the candidate proteins have homologs that form ISWI complexes in other organisms (Dirscherl and Krebs 2004). Since HMMER3 Pfam database searches failed to identify any domain (Finn et al. 2003), we searched for more distantly associated domains using HHpred (Zimmermann et al. 2018). HHpred generates a Hidden Markov Model (HMM) for the query using the iterative search and alignment functionality provided by HHblits (Remmert et al. 2011). The HHpred results indicated a probability of 91.68% for the “D-TOX E motif, Williams-Beuren syndrome DDT (WSD) motif” (Pfam model PF15613; 65 aa long, spanning almost the complete model length) (Fig. 1A,B). This domain corresponds to the entire

126 WHIM2 and half of the succeeding WHIM3, i.e. two of three “motifs” in a series of so-
127 called WHIM motif proteins (Aravind and Iyer 2012).

128

129 ICOP1 and ICOP2 proteins had no other detectable domains but contained three
130 amino acid residues, called the GxD signature (Aravind and Iyer 2012) (Fig. 1A-C),
131 within the WSD motif. The WSD motif (InterPro ID: IPR028941) is known to interact
132 with linker DNA and the SLIDE domain in ISWI proteins (Mukherjee et al. 2009;
133 Yamada et al. 2011; Aravind and Iyer 2012). Proteins with WHIM motifs often have
134 multiple domain architectures (Aravind and Iyer 2012). Public databases like PFAM
135 may annotate proteins as single-domain despite having other domains, due to limited
136 detection sensitivity. For example, IOC3 in yeast (Uniprot identifier: P43596) is
137 annotated with WHIM1 alone (Fig. 1D) but also has WHIM2 and WHIM3 (Aravind
138 and Iyer 2012). Based on this and subsequent experiments, we named our putative
139 interacting candidates ISWI1 Complex Protein 1 (ICOP1) and its closely related
140 ohnolog ISWI1 Complex Protein 2 (ICOP2).

141

142 *ICOP1* and *ICOP2* gene expression is upregulated during autogamy with a profile
143 similar to *ISWI1*'s (Fig. 1E). Furthermore, phylogenetic analysis of proteins with the
144 WSD motif suggests that ICOP1 and ICOP2 are highly divergent relative to other
145 WSD motif-containing proteins (Supplemental Fig. 1). As shown subsequently, the
146 ICOP ohnologs appeared functionally equivalent.

147

ICOP proteins localize to the developing MACs during autogamy

We co-transformed paramecia with either N-terminally tagged HA-ICOP1 or C-terminally tagged ICOP2-HA with ISWI1-GFP to check ICOP localization. Similar to ISWI1 (Singh et al. 2022), ICOPs localized exclusively to the developing MACs during autogamy (Fig. 2A; Supplemental Fig. 2A). We observed no growth defects in the co-transformed cells during vegetative growth or in the F1 progeny (Supplemental Fig. 3A). The ICOP paralog localization thus suggests they function at the same stage as ISWI1 during new MAC development.

ISWI1 and ICOP paralogs form a complex *in vivo* during autogamy

Using the co-transformed HA-ICOP1/ISWI1-GFP or ICOP2-HA/ISWI1-GFP lysates, we performed reciprocal co-IPs to assess ICOP1 and ICOP2 interactions with ISWI1. As controls, lysates of non-transformed (wild type, WT) and ISWI1-GFP transformed cells were used. WT cells showed no protein pulldown signal with either HA- or GFP-conjugated beads (Fig. 2B,C; Supplemental Fig. 3B). ISWI1-GFP signal was detected only in the "input" fraction when using the HA-conjugated beads (Fig 2B, lower panel) in ISWI1-GFP transformants. ISWI1-GFP was successfully co-purified with HA-ICOP1 or ICOP2-HA from the co-transformed cell lysates (Fig. 2B,C; Supplemental Fig. 3B). Co-IPs with ISWI1-GFP, HA-ICOP1, and ICOP2-HA single

transformants were further analyzed using protein MS (Supplemental Fig. 3C). ISWI1 was among the most highly enriched proteins, along with either or both of the ICOPs (Supplemental Fig. 3D). Therefore, we conclude that both ICOP paralogs can interact with ISWI1 in *Paramecium*.

Assessment of GxD signature requirement for interaction ISWI1-ICOP interaction

We tested whether ICOP1 and ICOP2 could bind directly to ISWI1 by co-expressing them in *E. coli*. GST or His N-terminal fusion proteins or untagged proteins were used for the pulldown. Pulldown specificity was validated using glutathione agarose (GST) or nickel-IMAC agarose (Ni_2+NTA) beads. Unspecific binding or cross-reactivity of tagged proteins in the IP fraction of the pulldowns was not observed (Supplemental Fig. 3E-G). ISWI1, ICOP1, and ICOP2 were co-expressed in different combinations to perform pulldowns using GST beads. The expected protein interactions were observed in all the pulldown combinations tested (Fig. 2D-F).

Since the GxD signature in WSD motif-containing proteins was proposed to mediate interactions with ISWI1 in diverse eukaryotes (Aravind and Iyer 2012), we assessed whether this signature is needed to form the ISWI1-ICOP complex. ICOP1 and ICOP2 have two GxDs (Fig. 3A). As aspartate was proposed to drive the interaction in the GxD signature (Aravind and Iyer 2012), ICOP mutants with a D to A substitution (GxA mutants) were generated. In addition, mutants with the complete

GxD deletion (delGxD mutants) were also generated (Fig. 3B). Both mutants co-immunoprecipitated with His-ISWI1 (Fig. 3C,D). A 2x del ICOP1 mutant (GSD and GFD removed) pulldown was inefficient (Fig. 3D), barring which His-ISWI1 co-purified with all the other ICOP mutants. Therefore, we found no evidence that ISWI1-ICOP interaction requires a GxD signature.

We then explored ISWI1-ICOP interaction using AlphaFold2. ISWI1's predicted structure was of high confidence, and its domains were similar to published structures of yeast ISWI (Yamada et al. 2011; Yan et al. 2019) (Supplemental Fig. 4A,B). However, ICOP structure predictions were low confidence (Supplemental Fig. 4B), likely due to their high divergence compared to other WSD proteins that generated a less informative multiple sequence alignment for structure prediction. We detected large interaction interfaces between ISWI1 and the ICOPs using AlphaFold version 2.3.0 in all the tested combinations. In contrast, AlphaFold2 version 2.2.0 predicted an interaction of ICOPs only with the ISWI1 N-terminus (residues 1-603, including the ATPase domain but not the HSS domain) (Fig. 3E-H, Supplemental Table S1). In these models, the ICOPs bound with a defined helix-loop-helix motif (ICOP1: residues 556-597; ICOP2: residues 560-603) (Fig. 3H). Irrespective of the AlphaFold2 version, neither of the GxD signatures were predicted to participate in the interaction (Fig. 3F,G; Supplemental Table S1). Ioc3, a WSD motif-containing ISWI complex protein in yeast, binds to ISW1a C-terminus (Yamada et al. 2011; Aravind and Iyer 2012) without any polar interactions between the GxD

signature of its WSD and ISW1a (Fig. 3I). Hence, the GxD signature does not appear to be necessary for ISWI1-ICOP interaction.

ICOP1/2 knockdown affects cell survival and genome editing

ICOP1 and *ICOP2* knockdown using RNAi by feeding, individually or combined, was performed to assess the ICOP roles. Knockdown (KD) of *ND7*, a gene involved in trichocyst discharge (exocytosis) (Skouri and Cohen 1997) was used as negative control (CTRL). Previously published *ISWI1*-KD data (Singh et al. 2022) was used as positive control and for comparison. Knockdown efficiency was confirmed using RNA-seq: the target gene expression was substantially reduced compared to the controls in all KDs (Fig. 4A). As the ICOPs are 92% identical at the nucleotide level, we checked for paralog co-silencing. Allowing no mismatches, ParameciumDB's off-target tool predicted a 24 bp window in *ICOP2* that could lead to co-silencing with the *ICOP1* RNAi construct (*Paramecium* siRNAs are typically 23 nt). We observed co-silencing of the opposing paralogs in the single KDs but to a lesser extent than the target gene (Fig. 4A). *ICOP1*-KD led to 30% lethality and *ICOP2*-KD led to ~20% lethality; a double KD of *ICOP1* and *ICOP2* led to ~65% lethality in the F1 generation (Fig. 4B). Additionally, most single knockdown cells failed to grow at a standard division rate ("sick" cells; Fig. 4B).

PCRs on post-autogamous cell genomic DNA were used to check whether the *ICOP* KDs affect IES excision (Fig. 4C). Longer fragments containing IESs (IES^+) were

amplified in all KD permutations, suggesting ICOPs are essential during genome editing.

Next, we investigated the genome-wide effect of *ICOP* KDs on IES retention using whole-genome sequencing of new MAC-enriched DNA. For each IES a retention score (IRS) was calculated as $IES^+/(IES^+)+(IES^-)$ (IES^+=reads with IES; IES^-=reads without IES). Both single and double KDs caused substantially more IES retention than *ND7*-KD. IRS score distributions of *ICOP* KDs were similar to those of *ISWI1*-KD (Fig. 4D) and right-shifted towards higher IRS compared to knockdowns of *PTIWI01/09*-KD (Fig. 4E, diagonal histograms). In addition, transposon retention was also observed when sequencing reads were mapped against Sardine and Thon transposons (ENA identifier: HE774469) (Supplemental Fig. 5A).

Strong IRS correlation suggests close cooperation between different genome editing molecules. For example, *EZL1* and *PTCAF1*, genes of the *Paramecium* PRC2 complex (Miró-Pina et al. 2022; Wang et al. 2022) have a strong IRS correlation (Swart et al. 2017) when knocked down. Like *ISWI1*-KD, *ICOP1/2*-KD IRSs correlated modestly with other gene knockdown IRSs (e.g., Fig. 4E). The correlation of *ICOP1/2*-KD was strongest with *ISWI1*-KD (Pearson correlation=0.75) (Fig. 4E).

ICOP1/2-KD affects IES excision precision

IES excision errors can naturally manifest as alternative excision, occurring at *Paramecium* TA dinucleotides that are not the predominant boundaries (Duret et al. 2008) (Fig. 5A). Generally, natural alternative excision levels are low (CTRL-KD, Fig. 5B,C). *ISWI1*-KD substantially enhances alternative excision versus KDs of other genome-editing genes (Singh et al. 2022). Similar to but less than *ISWI1*-KD, *ICOP1*-KD and *ICOP2*-KD increased imprecise excision (Fig. 5; Supplemental Table S2). Previously (Singh et al. 2022), we did not measure alternative excision of IESs where 100% of the mapped reads were alternatively excised (Supplemental Table S2), thus underestimating alternative excision. Nevertheless, by the old estimation method, the percentage of alternative excision events per IES was highest in *ICOP1*-KD (mean 7%) and similar between *ICOP2*-KD (mean 4.2%) and *ICOP1/2*-KD (mean 4.7%). With the exception of *ISWI1*-KD (mean 9.2% (Singh et al. 2022); Supplemental Table S2), this is higher than other KDs (mean 1.5-2.4% (Singh et al. 2022)).

The use of alternative TA boundaries changes the excised fragment lengths. The maximum and minimum excised IES lengths were shifted towards the extremes, and alternatively excised IESs were generally longer than the reference length. The alternatively excised IES length distribution resembled the ~10 bp periodicity characteristic of *Paramecium* IESs, with the striking exception that the “forbidden” peak (Arnaiz et al. 2012) was present in all three *ICOP* KDs, as in *ISWI1*-KD (Fig. 5C). In *ISWI1*-KD, alternative IESs in the “forbidden” peak mainly originated from the

first and third peaks, while they primarily originated from the third peak in *ICOP* KDs (Fig. 5D). The similarity in alternative excision effects of *ISWI1* and *ICOP* KDs suggests that ISWI1 and ICOP proteins cooperate in precise IES excision.

Furthermore, we examined five possible alternative IES excision events: “partial internal”, “partial external”, “overlap”, “internal,” and “external” (Fig. 5A). Generally, “internal” and “external” are low-frequency events in all KDs (Supplemental Fig. 5B). In the negative control KD, “overlap”, “partial external” and “partial internal” events were approximately equal at around 30% each (Supplemental Fig. 5C). This contrasts with *ICOP*s and *ISWI1* KDs, where “overlap” was infrequent, while “partial internal” and “partial external” comprised the largest share of erroneous excision events (Fig. 5E, Supplemental Fig. 5C; Supplemental Table S3). In *ISWI1*-KD, "partial internal" (- 43%) and "partial external" (42%) events contributed equally, while “partial internal” dominated the *ICOP* KDs. The preference was more pronounced in the single KDs (“partial internal” - 57%; “partial external” - 28% for *ICOP1*- and *ICOP2*-KD) than in *ICOP1/2*-KD (“partial internal” - 47%; “partial external” - 34%) (Supplemental Fig. 5C).

ICOP1/2-KD does not alter ISWI1-GFP or GFP-Ptiwi09 localization but affects scnRNAs and iesRNAs

We knocked down *ICOP1* and/or *ICOP2* to check whether their expression is required for ISWI1-GFP localization. As in control cells with no RNAi (Fig. 6A),

ISWI1-GFP localization was not impaired in *ICOP* KDs (Fig. 6C-E). Only in *ISWI1*-KD was the GFP signal entirely lost from the new MAC (Fig. 6B). Conversely, HA-ICOP1 and ICOP2–HA localized to the new MACs upon *ISWI1*-KD (Supplemental Fig. 2B-D) as in non-knockdown cells. In *Paramecium*, the excision of a subset of IESs is suggested to depend on scnRNAs (Garnier et al. 2004). We tested the dependence of ISWI1-GFP and ICOP-HA localization on genome scanning by knocking down *PTIWI01/09* (scnRNA Piwis). ISWI1-GFP, HA-ICOP1 and ICOP2-HA localized to the new MAC upon *PTIWI01/09*-KD (Fig. 6F, Supplemental Fig. 2B-D). This suggests ISWI1 localization is independent of the ICOP proteins and genome scanning.

Conversely, we checked whether *ICOP1/2*-KD influences the small RNA population and consequently genome scanning. scnRNAs are generated in MICs well before new MAC development (Lepère et al. 2009). Consequently, scnRNA production should only be affected by the silencing of genes involved in their biogenesis. As expected, in early development (~40% of cells with fragmented parental MACs), we did not observe a pronounced effect on scnRNA production in *ICOP1/2*-KD compared to the negative control *ND7*-KD (*CTRL*-KD) (Fig. 6G).

Knockdowns of genes like *PTIWI10/11* and *DCL5* directly involved in iesRNA biogenesis inhibit iesRNA production (Sandoval et al. 2014; Furrer et al. 2017). Knockdowns of other genes that inhibit IES excision, like *NOWA1/2*, *PDSG1/2*, *EZL1*-KD, and *PTCAF1* (Arambasic et al. 2014; Ignarski et al. 2014; Lhuillier-Akakpo

et al. 2014; Swart et al. 2017) also inhibit iesRNA production, presumably as iesRNAs require excised IESs as substrates for the transcription of their dsRNA precursors (Allen et al. 2017). Here, we also observed inhibition of iesRNA production for *ICOP1/2*-KD (Fig. 6H) in late development.

Comparing the MAC-matching scnRNAs normalized to MAC-matching siRNAs, there was a greater quantity of MAC-matching scnRNAs in the late time point (~90% of cells with visible new MACs) for *ICOP1/2*-KD than for *CTRL*-KD (Fig. 6I,J). This suggests that MAC-matching scnRNA subtraction, as proposed in the RNA scanning model, was impaired by *ICOP1/2*-KD (Fig. 6J). We also examined sRNA biogenesis-related gene transcription (i.e. *PTIWI*, *DCL* and *NOWA*) in *ICOP1/2*-KD vs *CTRL*-KD. In late development, *PTIWI10* and *PTIWI11* expression was almost completely lost upon *ICOP1/2*-KD (Fig. 6K), whereas *PTIWI01, PTIWI09*, *DCL2*, *DCL3* and *NOWA1/2* were upregulated (Fig. 6K,L). Hence, MAC-matching scnRNA enrichment might be caused by scnRNA-associated gene dysregulation.

We also investigated Ptiwi09-GFP localization upon *ISWI1*-KD and *ICOP1/2*-KD (Supplemental Fig. 6). Without knockdown, Ptiwi09-GFP localizes first to the parental MAC and cytosol, and later to the new MACs during development. Additionally, we observed Ptiwi09-GFP transiently in the swelling MICs before the first meiotic division (Supplemental Fig. 6A). Upon *DCL2/3*-KD, Ptiwi09-GFP failed to enter the parental MAC, remaining in the cytosol, whereas its localization to the new MACs was unimpaired (Supplemental Fig. 6B). In contrast to the obvious changes in

localization due to *DCL2/3*-KD, in the *ISWI1*-KD and *ICOP1/2*-KD, we observed a tendency for Ptiwi09-GFP to remain longer in parental MAC fragments and an enhanced localization around the MICs during new MAC development (Supplemental Fig. 6C,D).

ICOP1/2-KD IES nucleosome density changes are similar to those of *ISWI1*-KD

To further investigate the functional contribution of the *ICOP* paralogs to the ISWI1 complex, we analyzed the *ICOP* KD effects on IES nucleosome densities. IESs with high retention in *ICOP1/2*-KD (IRS ≥ 0.2) tended to have higher nucleosome densities (Fig. 7A; Supplemental Fig. 7C,D) in both *ICOP1/2/PGM*-KD and *CTRL/PGM*-KD, similar to our previous observations with other knockdowns including *ISWI1/PGM*-KD (Singh et al. 2022). The nucleosome density differences (experiment minus control) for *ICOP1/2/PGM*-KD and *ISWI1/PGM*-KD had similar distributions with a narrow peak centered around 0 (Fig. 7B; Supplemental Table S4). The comparable distributions for *NOWA1/2/PGM*-KD and *PTCAF1/PGM*-KD were similar to one another but broader and flatter than *ICOP1/2/PGM*-KD (Fig. 7B). This suggests distinct effects of the ISWI1 complex on nucleosome densities and would accord with ICOP1/2 and ISWI1 being present in a distinct complex from PTCAF1 and NOWA1/2.

To check effects of IES length and *ICOP1/2*-KD IRS on nucleosome density, IESs were grouped according to these properties. In *ICOP1/2/PGM*-KD and *ISWI1/PGM*-KD, nucleosome density differences were most prominent for long and/or ICOP1/2-dependent IESs (Fig. 7C). In the *ISWI1/PGM*-KD, there was no clear trend towards higher or lower nucleosome densities, whereas, in *ICOP1/2/PGM*-KD, there tended to be higher nucleosome densities in the experimental sample (Fig. 7C; Supplemental Table S4). This shift towards higher nucleosome densities was also observed for *PTCAF1/PGM*-KD (Supplemental Fig. 8; Supplemental Table S4), indicating this effect is not specific to components of the ISWI1 complex.

Discussion

In this study, we identified and analyzed the role of the ICOP1 and ICOP2 subunits, that, together with ISWI1 protein subunit, appear to form a developmental genome editing complex in *Paramecium*. ICOP1 and ICOP2 are highly divergent from other proteins, lacking homology or additional domains detectable by routine search methods. In such cases, it may be helpful to use software like HHpred, using pairwise HMM comparisons for distant homology searches (Zimmermann et al. 2018). Thus, we identified a highly divergent WSD motif in ICOPs (Fig. 1). WSD motif is found in proteins that are subunits of the ISWI complex in several organisms (Toto et al. 2014) (Fig. 1D).

We provided evidence using *Paramecium* and *E. coli* that ISWI1 forms a complex with the ICOP proteins (Fig. 2). The observations of proteins overexpressed in *E.*

389 *coli*, lacking *Paramecium* proteins, support direct ISWI1-ICOP binding. ISWI1 co-IPs
390 with both paralogs. ICOP2 was not enriched in HA-ICOP1 co-IP, while ICOP2-HA
391 co-IP has low ICOP1 enrichment (Supplemental Fig. 3D). Thus, despite their ability
392 to interact directly (Fig. 2D), it is likely that ISWI1 might typically form complexes with
393 either ICOP subunit.

394

395 Though WSD's GxD aspartate is proposed to determine ISWI-WSD motif-containing
396 protein interaction (Aravind and Iyer 2012), to our knowledge, no supporting
397 experimental evidence exists for this. In an Ioc3 crystal structure, a WSD motif-
398 containing protein of yeast, the GIQ lacks the third acidic residue and forms no polar
399 interactions with ISW1a (Fig. 3I). Our heterologous expression studies show that
400 GxD signature mutation or deletion does not completely abolish ICOP-ISWI
401 interaction (Fig. 3C,D). Furthermore, AlphaFold2 modeling predicted the interaction
402 of the ICOPs at the ISWI1 N-terminus, mediated by a helix-turn-helix motif rather
403 than the GxD (Fig. 3F,G). Better structural prediction software and experimental
404 approaches will be needed to determine precisely how the proteins interact in this
405 complex.

406

407 Along with strong iesRNA production inhibition, *PTIWI10*/*11* expression was
408 abolished by the *ICOP* KDs. As these genes are transcribed in the developing MAC
409 (Furrer et al. 2017), the loss of *PTIWI10/11* expression could either be due to IES
410 retention in their promoters or to nonsense-mediated decay (NMD) of mRNA
411 triggered by IES retention in the CDS (Sandoval et al. 2014; Furrer et al. 2017;

412 Bazin-Gélis et al. 2023). sRNA sequencing also revealed that the MAC-specific
413 scnRNAs are elevated in *ICOP1/2*-KD compared to the control (Fig. 6H,J). The same
414 phenomenon has been observed in *NOWA1/2*-KD (Swart et al. 2017) and *PTCAF1*-
415 KD (Ignarski et al. 2014). NOWA1/2 is involved in genome scanning (Nowacki et al.
416 2005), whereas PTCAF1 is a part of the PRC2 complex employed in H3K27me3
417 deposition during IES excision (Ignarski et al. 2014; Miró-Pina et al. 2022; Wang et
418 al. 2022). Previous studies suggest that elevated MAC-specific scnRNA levels are
419 due to PTCAF1 inhibition in the old MAC during scanning (Ignarski et al. 2014).

420

421 With the caveat of the lack of replicates, unlike *PTIWI10/11* (iesRNA Piwis), genes
422 associated with scnRNAs, notably *PTIWI01*/*09*, were modestly upregulated in the
423 late developmental stage upon *ICOP1/2*-KD, potentially by inhibiting MAC-genome-
424 matching scnRNAs from degradation. Though we observed subtle differences in
425 PTIWI09-GFP localization after *ICOP1/2*-KD, this would have to be replicated and
426 the possible mechanism investigated in the future. Furthermore, it would also be
427 worth revisiting the RNA scanning model in *Paramecium*, rigorously examining key
428 details not yet directly established, e.g. what substrates the scnRNAs pair with.

429

430 It would also be worth investigating the expression of *PTIWI01*/*09* and related
431 genome editing genes (e.g., *NOWA1/2* and *PTCAF1*) for other knockdowns to
432 compare to those in *ICOP1/2*-KD. However, it is clear that the IES retention in
433 *ICOP1/2*-KD is substantially stronger than in *PTIWI* knockdowns (Fig. 4E) and also
434 exhibits enhanced alternative excision (Fig. 5). Thus, altered expression of the

435 *PTIWIs* and other genome editing genes cannot account for most of the observed
436 *ICOP1/2*-KD effects, irrespective of whether the development-specific sRNA levels
437 or their MAC:IES ratios are altered.

438

439 Most IESs are likely transposon remnants (Sellis et al. 2021; Seah et al. 2023) that
440 decayed beyond recognition due to their efficient developmental excision (Sellis et
441 al. 2021). A third of all IESs are 26 to 28 bp in length and are proposed to be short
442 enough to allow direct interaction of two PGMs without DNA looping (Arnaiz et al.
443 2012). Longer IESs require DNA looping, causing 34 to 44 bp IESs in the “forbidden"
444 peak to be highly underrepresented, either too long for direct PGM subunits’
445 interaction without looping or too short for DNA looping to permit this interaction.
446 Similar to ISWI1, *ICOP*-KDs caused both IES retention and elevated alternative IES
447 excision (Fig. 4, 5).

448

449 Generally in genome editing gene knockdowns, alternative excision levels do not
450 exceed the background (Singh et al. 2022), but are enhanced by ISWI1 knockdown.
451 This led to the emergence of “forbidden” peak length IESs. In the *ICOP* KDs, the
452 alternatively excised IESs in the “forbidden” peak mainly originated from the
453 subsequent peak. This aligns with the observation that partial internal excision,
454 leading to shorter lengths, dominated alternative excision events in *ICOP* KDs
455 (mainly in the single KDs). In *ISWI1*-KD, partial internal and external excision
456 contributed equally to the alternatively excised IESs and the “forbidden” peak. The
457 difference in excision preference might be caused by ISWI’s ability to move

458 nucleosomes on its own (Havas et al. 2000; Längst and Becker 2001). Some
459 nucleosome repositioning may still happen via ISWI1 in the *ICOP* KDs, although not
460 as effectively as with the ICOPs. However, in *ISWI1*-KD, where nucleosome
461 repositioning fails, IES removal occurs at the next available TA, whether internal or
462 external.

463

464 We observed nucleosome density difference distributions for *ICOP1/2/PGM*-KD and
465 *ISWI1/PGM*-KD were sharply peaked, indicating generally little difference in
466 nucleosome density on IESs irrespective of the ISWI1 complex presence (Fig. 7B).
467 However, *NOWA1/2/PGM*-KD and *PTCAF1/PGM*-KDs showed broader distributions,
468 implying that IES nucleosome densities are less influenced by ISWI1 complex
469 components' downregulation than by the downregulation of other genes. Since
470 nucleosome densities do not capture exact nucleosomal positions, nucleosome
471 positions rather than the number of nucleosomes may change in *ICOP1/2/PGM*-KD
472 and *ISWI1/PGM*-KD. However, this cannot be properly investigated by current
473 computational methods due to the inability to distinguish between most old and new
474 MAC sequences, and read alignment accuracy limitations at IES boundaries.

475

476 *NOWA1/2/PGM*-KD and *PTCAF1/PGM*-KDs might have stronger effects on
477 nucleosome density differences because *NOWA1* and *PTCAF1* are expressed
478 earlier than the ISWI1 complex and localize to both maternal and developing MACs
479 (Nowacki et al. 2005; Ignarski et al. 2014). Therefore, observed nucleosome density
480 differences could either be due to disruption of events downstream of NOWA1 and

PTCAF1 functions or due to inter-generational nuclear crosstalk effects on gene regulation as proposed recently (Bazin-Gélis et al. 2023). Irrespective, a clear difference on both chromatin and IES excision can be observed between the ISWI1 complex and other genome editing components, indicating a distinct role for ICOPs and ISWI1 on nucleosomes.

ICOP paralogs might contribute to the directionality of the remodeling complex, as shown for *Drosophila* Acf1, a protein that regulates ISWI-containing complex CHRAC directionality (Eberharter et al. 2001). In contrast to ISWI1, ICOP KDs caused a preference both for partial internal excision (Supplemental Fig. 5C) and for higher nucleosome densities on long/highly retained IESs (Fig. 7C). Higher nucleosome densities might be a direct cause for preferred partial internal excision.

We previously proposed a “clothed” model for IES excision, where mispositioned nucleosomes change the accessibility of the IES boundaries to the PGM excision complex (Singh et al. 2022). Assuming that the cooperating PGMs cannot interact across a nucleosome without a sufficiently long DNA loop, partial internal excision might be preferred if a nucleosome is located on a TA boundary since an alternative TA lying within the IES might be more easily accessible than one outside.

Besides nucleosome positioning, precise IES boundary targeting might also depend on the DNA topology, which influences protein binding and can be exploited in

503 regulation (Baranello et al. 2012). Some ISWI family chromatin remodelers can
504 change the DNA topology (Havas et al. 2000), which might cause the PGM complex
505 to recognize the wrong TA dinucleotides if alterations in chromatin remodeling occur.
506 This would also explain how the “forbidden” peak can emerge. According to the
507 original “naked” DNA model, the PGM excision machinery struggles to excise 34-44
508 bp fragments (Arnaiz et al. 2012). However, if the DNA helix conformation changes,
509 the PGM complex working distance might correspond to the forbidden length. It
510 seems that the ICOPs can partially compensate for each other since their double KD
511 resembled the *ISWI1*-KD more than their single KDs in terms of cell survival
512 (Supplemental Fig. 3A) and IES retention and alternative excision effects (Fig. 4D;
513 Fig. 5B; Supplemental Fig. 5C). We thus propose that the ICOP proteins assist
514 ISWI1’s function in precise genome editing, either by nucleosome sliding or DNA
515 topology changes.

516

517 In *Paramecium*, linker DNA between somatic nucleosomes was shown to be
518 extremely short at just a few bp (Gnan et al. 2022), and no linker histone H1 was
519 detected (Drews et al. 2022b). Furthermore, histone modifications characteristic of
520 eu- and heterochromatin in other eukaryotes did not show the expected relations
521 with active and repressive gene expression (Drews et al. 2022b). *Paramecium* MIC
522 and MAC nucleosome properties, like their distribution and dynamics, still need more
523 thorough investigation. Future studies enabling more precise nucleosome
524 positioning, potentially via isolation from flow-sorted MACs, will be essential to

525 determine how nucleosome occupancy and movements by complexes like ISWI1-
526 ICOP affect the targeting of natural genome editing.

527

528 Materials and methods

529 Culture cultivation and RNAi assays

530 Culture cultivation and RNAi assays are described in Supplemental Methods.

531

532 DNA microinjection and localization

533 The standard DNA microinjection protocol was followed (Beisson et al. 2010). Since
534 ICOP1 and ICOP2 fusion gene expression with endogenous flanking regulatory
535 regions failed, those of ISWI1 (Singh et al. 2022) were used instead. Human
536 influenza hemagglutinin (HA) was fused N-terminally to ICOP1 and C-terminally to
537 ICOP2. ISWI1-GFP plasmid is described in (Singh et al. 2022). GFP-PTIWI09
538 plasmid was a gift from the Nowacki lab (Bern, Switzerland). Cells were collected
539 during different stages of autogamy and either stored in 70% ethanol at -20 °C or
540 directly fixed with 2% formaldehyde (PFA) in PHEM (PIPES, HEPES, EGTA,
541 Magnesium Sulphate), washed (2 × 5 min at room temperature (RT)). 5% BSA with
542 0.1% Triton X-100 in Tris-buffered saline with 10 mM EGTA and 2 mM MgCl2
543 (TBSTEM) was used for blocking (1 h, RT). Cells were stained overnight at 4 °C with
544 a primary anti-HA antibody (sc-7392, Santa Cruz) followed by washing and

secondary anti-mouse Alexa-594 (BLD-405326, Biozol) or anti-mouse Alexa-568 (A11004, Thermo Fisher Scientific) incubation (1h, RT). After washing, cells were counterstained with DAPI (4,6-diamidino-2-2-phenylindole) in 5% BSA with 0.1% Triton X-100-TBSTEM. Cells were mounted with 40 µl of Prolong Gold Antifade mounting medium (Invitrogen). Images were acquired with a Leica SP8 confocal microscope system with a 60× oil objective (NA 1.4). Images were analyzed using Fiji (version 2.9.0/1.53t) (Schindelin et al. 2012). Macros used for image analysis are available from https://github.com/Swart-lab/ICOP_code/tree/main/Postprocessing_IF.

Co-immunoprecipitation and western blotting

Co-immunoprecipitations and western blots were done as previously described (Singh et al. 2022) using late-stage lysates. Sonication used an MS72 tip on a Bandelin Sonopulse device with 52% amplitude for 15 s. For non-crosslinked samples, cells were lysed using sonication on ice after washing with 10 mM Tris pH 7.4 in a resuspension of 2 ml lysis buffer. Pulldown fractions were resolved on 12% SDS-PAGE gels. 1% of total lysates were loaded as input, optionally 1% of supernatant after beads incubation as unbound, and 30% (Fig. 2) or 20% (Supplemental Fig. 3) of the total IP samples were loaded.

An anti-HA antibody (1:500, sc-7392 HRP, Santa Cruz) and anti-GFP antibody (1:2000, ab290, Abcam) incubation was done overnight at 4 °C. The secondary antibody, goat-anti-Rabbit HRP conjugated (12-348, Merck Millipore), was incubated

for 1 h at room temperature. Membranes were screened using AI600 (GE Healthcare) after incubation with an HRP substrate (42029053, Millipore) for 2-5 mins.

Protein expression in *E. coli*

Plasmids used for *E. coli* expression are detailed in Supplemental Methods. 50 ml of ZY medium (Studier 2014) containing appropriate antibiotics was inoculated with 100 µl of transformed *E. coli* culture. Cultures were grown at 37 °C at 180 rpm until an OD600 of 2 was reached. Afterwards, the temperature was decreased to 20 °C for overnight protein expression. 2 ml of culture was centrifuged at 4000 g at 4 °C, and pellets were frozen at -80 °C.

Recombinant protein co-precipitation

Cell pellets were resuspended in 1 ml of lysis buffer: 20 mM Tris pH 7.5, 100 mM NaCl for GST pulldown or 20 mM Tris pH 7.5, 100 mM NaCl, 20 mM Imidazole, 1 mM DTT for His pulldown. 20% amplitude (0.5 s on, 0.5 s off) with an MS72 tip (Bandelin Sonopulse) was used for sonication, followed by centrifugation (21130 g, 15 min, 4 °C) to recover the supernatant for pulldown. 30 µl of beads (42172.01/ 42318.01, Serva) were washed, equilibrated with lysis buffer, loaded with protein supernatant and incubated for 1 h or overnight at 4 °C using gentle shaking. After three washes in lysis buffer, the enriched protein was eluted from beads by adding

30 µl of 2× protein loading Buffer (100 mM Tris-HCl pH 6.8, 4% (w/v) SDS, 20% Glycerol, 0.2 M DTT) and boiling for 10 min. The supernatant was loaded on a 10-12% SDS-PAGE gel. 1% of the total lysate was loaded as input, and 20% of the total pulldown was loaded in the IP fraction. 1:4000 rabbit anti-GST antibody (G7781, Sigma) and mouse anti-His (1:2500, 362601, BioLegend) were diluted in 5% BSA in 1× PBS + 0.2% Tween20 for blotting. 1:5000 reciprocal secondary antibody incubation was done for 1 h at room temperature.

DNA and total RNA extraction and sequencing

Standard methods were used to isolate macronuclear DNA and total RNA for sequencing (details in Supplemental Methods).

IES retention and alternative boundary analysis

IES retention scores and alternative excision were calculated as previously (Singh et al. 2022) (see Supplemental Methods for details).

Nucleosome Density Analysis

See Supplemental Methods for nucleosomal DNA isolation and sequencing procedures. Nucleosome densities were calculated as previously described (Singh et al. 2022). As previously, we focused on IES-mapping reads since the old and new

MAC sources of MAC-mapping reads are indistinguishable. Due to the experimental nuclear/nucleosome isolation procedure, most IES-mapping reads should be from the developing new MAC. For further details see Supplemental Methods.

sRNA analysis

sRNA-seq processing and analysis are described in Supplemental methods.

Structure prediction with AlphaFold

Protein structures were predicted with AlphaFold multimer version 2.2.0 and 2.3.0 (Evans et al. 2021; Jumper et al. 2021). Protein sequences provided as input are listed in Supplemental Table S8. All predictions were computed on the high-performance computer “Raven”, operated by the Max-Planck Computing and Data Facility in Garching, Munich, Germany. PDB files are available as SourceData_Fig3 (Singh 2023).

Data access

All original images corresponding to gels and microscopy can be accessed from EDMOND (https://doi.org/10.17617/3.ZBOLU8). Whole-genome-sequencing data, small RNA sequencing data, and mRNA sequencing data can be obtained from European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home): BioProject

PRJEB64685. Raw protein mass spectrometry data is available from ProteomExchange (https://www.proteomexchange.org/): PXD044340.

Competing Interest Statement

The authors declare no competing interests.

Acknowledgments

We thank: the MPI for Biology (Tübingen, Germany) core facilities for microscopy and sequencing assistance; Vikram Alva for helpful discussions on using HHpred; Andre Noll for computer system administration; Therese Solberg (Nowacki lab, Bern, Switzerland) for GFP-PTIWI09 plasmid.

Author contributions: A.S., L.H., E.C.S. designed research; A.S., L.H., C.E., E.N., B.K.B.S., performed research; A.S., L.H., E.C.S. analyzed data; M.N., E.C.S. contributed reagents/analytical tools; L.H., A.S., E.C.S., wrote the paper; A.S., F.B., E.C.S supervision.

References

Alabert C, Groth A. 2012. Chromatin replication and epigenome maintenance. *Nat Rev Mol Cell Biol* **13**: 153–167.

Allen SE, Hug I, Pabian S, Rzeszutek I, Hoehener C, Nowacki M. 2017. Circular Concatemers of Ultra-Short DNA Segments Produce Regulatory RNAs. *Cell* **168**: 990-999.e7.

Arambasic M, Sandoval PY, Hoehener C, Singh A, Swart EC, Nowacki M. 2014. Pdsg1 and Pdsg2, novel proteins involved in developmental genome remodelling in Paramecium. *PLoS ONE* **9**: e112899.

Aravind L, Iyer LM. 2012. The HARE-HTH and associated domains: novel modules in the coordination of epigenetic DNA and protein modifications. *Cell Cycle* **11**: 119–131.

Arnaiz O, Mathy N, Baudry C, Malinsky S, Aury J-M, Denby Wilkes C, Garnier O, Labadie K, Lauderdale BE, Le Mouël A, et al. 2012. The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet* **8**: e1002984.

Aydin ÖZ, Vermeulen W, Lans H. 2014. ISWI chromatin remodeling complexes in the DNA damage response. *Cell Cycle* **13**: 3016–3025.

Bai L, Morozov AV. 2010. Gene regulation by nucleosome positioning. *Trends Genet* **26**: 476–483.

Baranello L, Levens D, Gupta A, Kouzine F. 2012. The importance of being supercoiled: how DNA mechanics regulate dynamic processes. *Biochim*

Biophys Acta **1819**: 632–638.

Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, Bétermier M. 2009. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. *Genes Dev* **23**: 2478–2483.

Bazin-Gélis M, Eleftheriou E, Zangarelli C, Lelandais G, Sperling L, Arnaiz O, Bétermier M. 2023. Inter-generational nuclear crosstalk links the control of gene expression to programmed genome rearrangements during the *Paramecium* sexual cycle. *BioRxiv*.

Beisson J, Bétermier M, Bré M-H, Cohen J, Duharcourt S, Duret L, Kung C, Malinsky S, Meyer E, Preer JR, et al. 2010. DNA microinjection into the macronucleus of paramecium. *Cold Spring Harb Protoc* **2010**: pdb.prot5364.

Bischerour J, Bhullar S, Denby Wilkes C, Régnier V, Mathy N, Dubois E, Singh A, Swart E, Arnaiz O, Sperling L, et al. 2018. Six domesticated PiggyBac transposases together carry out programmed DNA elimination in Paramecium. *eLife* **7**.

Bouhouche K, Gout J-F, Kapusta A, Bétermier M, Meyer E. 2011. Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res* **39**: 4249–4264.

Campos EI, Reinberg D. 2009. Histones: annotating chromatin. *Annu Rev Genet* **43**: 559–599.

688 Clapier CR, Cairns BR. 2009. The biology of chromatin remodeling complexes. *Annu*
689 *Rev Biochem* **78**: 273–304.

690 Dirscherl SS, Krebs JE. 2004. Functional diversity of ISWI complexes. *Biochem Cell*
691 *Biol* **82**: 482–489.

692 Drews F, Boenigk J, Simon M. 2022a. Paramecium epigenetics in development and
693 proliferation. *J Eukaryot Microbiol* **69**: e12914.

694 Drews F, Salhab A, Karunanithi S, Cheaib M, Jung M, Schulz MH, Simon M. 2022b.
695 Broad domains of histone marks in the highly compact Paramecium
696 macronuclear genome. *Genome Res* **32**: 710–725.

697 Dubois E, Bischerour J, Marmignon A, Mathy N, Régnier V, Bétermier M. 2012.
698 Transposon invasion of the paramecium germline genome countered by a
699 domesticated piggybac transposase and the NHEJ pathway. *Int J Evol Biol*
700 **2012**: 436196.

701 Duret L, Cohen J, Jubin C, Dessen P, Goût J-F, Mousset S, Aury J-M, Jaillon O,
702 Noël B, Arnaiz O, et al. 2008. Analysis of sequence variability in the
703 macronuclear DNA of *Paramecium tetraurelia*: a somatic view of the germline.
704 *Genome Res* **18**: 585–596.

705 Eberharter A, Ferrari S, Längst G, Straub T, Imhof A, Varga-Weisz P, Wilm M,
706 Becker PB. 2001. Acf1, the largest subunit of CHRAC, regulates ISWI-induced
707 nucleosome remodelling. *EMBO J* **20**: 3781–3788.

708 Evans R, O'Neill M, Pritzel A, Antropova N, Senior AW, Green T, Žídek A, Bates R,
709 Blackwell S, Yim J, et al. 2021. Protein complex prediction with AlphaFold-

Multimer. *BioRxiv*.

Finn R, Griffiths-Jones S, Bateman A. 2003. Identifying protein domains with the Pfam database. *Curr Protoc Bioinformatics* **Chapter 2**: Unit 2.5.

Furrer DI, Swart EC, Kraft MF, Sandoval PY, Nowacki M. 2017. Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA. *Cell Rep* **20**: 505–520.

Garnier O, Serrano V, Duharcourt S, Meyer E. 2004. RNA-mediated programming of developmental genome rearrangements in Paramecium tetraurelia. *Mol Cell Biol* **24**: 7370–7379.

Gnan S, Matelot M, Weiman M, Arnaiz O, Guérin F, Sperling L, Bétermier M, Thermes C, Chen C-L, Duharcourt S. 2022. GC content, but not nucleosome positioning, directly contributes to intron splicing efficiency in Paramecium. *Genome Res* **32**: 699–709.

Grüne T, Brzeski J, Eberharter A, Clapier CR, Corona DFV, Becker PB, Müller CW. 2003. Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol Cell* **12**: 449–460.

Havas K, Flaus A, Phelan M, Kingston R, Wade PA, Lilley DM, Owen-Hughes T. 2000. Generation of superhelical torsion by ATP-dependent chromatin remodeling activities. *Cell* **103**: 1133–1142.

Ignarski M, Singh A, Swart EC, Arambasic M, Sandoval PY, Nowacki M. 2014. Paramecium tetraurelia chromatin assembly factor-1-like protein PtCAF-1 is involved in RNA-mediated control of DNA elimination. *Nucleic Acids Res* **42**:

11952–11964.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583–589.

Kapusta A, Matsuda A, Marmignon A, Ku M, Silve A, Meyer E, Forney JD, Malinsky S, Bétermier M. 2011. Highly precise and developmentally programmed genome assembly in Paramecium requires ligase IV-dependent end joining. *PLoS Genet* **7**: e1002049.

Klobutcher LA, Herrick G. 1995. Consensus inverted terminal repeat sequence of *Paramecium* IESs: resemblance to termini of Tc1-related and *Euplotes* Tec transposons. *Nucleic Acids Res* **23**: 2006–2013.

Kornberg RD. 1977. Structure of chromatin. *Annu Rev Biochem* **46**: 931–954.

Längst G, Becker PB. 2001. ISWI induces nucleosome sliding on nicked DNA. *Mol Cell* **8**: 1085–1092.

Längst G, Bonte EJ, Corona DF, Becker PB. 1999. Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. *Cell* **97**: 843–852.

Lepère G, Nowacki M, Serrano V, Gout J-F, Guglielmi G, Duharcourt S, Meyer E. 2009. Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res* **37**: 903–915.

Lhuillier-Akakpo M, Frapporti A, Denby Wilkes C, Matelot M, Vervoort M, Sperling L, Duharcourt S. 2014. Local effect of enhancer of zeste-like reveals cooperation

754 of epigenetic and cis-acting determinants for zygotic genome rearrangements.
755 *PLoS Genet* **10**: e1004665.

756 Li M, Xia X, Tian Y, Jia Q, Liu X, Lu Y, Li M, Li X, Chen Z. 2019. Mechanism of DNA
757 translocation underlying chromatin remodelling by Snf2. *Nature* **567**: 409–413.

758 Maliszewska-Olejniczak K, Gruchota J, Gromadka R, Denby Wilkes C, Arnaiz O,
759 Mathy N, Duharcourt S, Bétermier M, Nowak JK. 2015. TFIIS-Dependent Non-
760 coding Transcription Regulates Developmental Genome Rearrangements.
761 *PLoS Genet* **11**: e1005383.

762 Miró-Pina C, Charmant O, Kawaguchi T, Holoch D, Michaud A, Cohen I, Humbert A,
763 Jaszczyszyn Y, Chevreux G, Del Maestro L, et al. 2022. Paramecium
764 Polycomb repressive complex 2 physically interacts with the small RNA-binding
765 PIWI protein to repress transposable elements. *Dev Cell* **57**: 1037-1052.e8.

766 Morgunova E, Taipale J. 2021. Structural insights into the interaction between
767 transcription factors and the nucleosome. *Curr Opin Struct Biol* **71**: 171–179.

768 Mukherjee K, Brocchieri L, Bürglin TR. 2009. A comprehensive classification and
769 evolutionary analysis of plant homeobox genes. *Mol Biol Evol* **26**: 2775–2794.

770 Nowacki M, Zagorski-Ostoja W, Meyer E. 2005. Nowa1p and Nowa2p: novel
771 putative RNA binding proteins involved in trans-nuclear crosstalk in
772 Paramecium tetraurelia. *Curr Biol* **15**: 1616–1628.

773 Patenge N, Elkin SK, Oettinger MA. 2004. ATP-dependent remodeling by SWI/SNF
774 and ISWI proteins stimulates V(D)J cleavage of 5 S arrays. *J Biol Chem* **279**:
775 35360–35367.

776 Piña B, Brüggemeier U, Beato M. 1990. Nucleosome positioning modulates
777 accessibility of regulatory proteins to the mouse mammary tumor virus
778 promoter. *Cell* **60**: 719–731.

779 Price BD, D'Andrea AD. 2013. Chromatin remodeling at DNA double-strand breaks.
780 *Cell* **152**: 1344–1354.

781 Pryciak PM, Varmus HE. 1992. Nucleosomes, DNA-binding proteins, and DNA
782 sequence modulate retroviral integration target site selection. *Cell* **69**: 769–780.

783 Remmert M, Biegert A, Hauser A, Söding J. 2011. HHblits: lightning-fast iterative
784 protein sequence searching by HMM-HMM alignment. *Nat Methods* **9**: 173–
785 175.

786 Sandoval PY, Swart EC, Arambasic M, Nowacki M. 2014. Functional diversification
787 of Dicer-like proteins and small RNAs required for genome sculpting. *Dev Cell*
788 **28**: 174–188.

789 Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T,
790 Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: an open-source
791 platform for biological-image analysis. *Nat Methods* **9**: 676–682.

792 Seah BKB, Singh M, Emmerich C, Singh A, Woehle C, Huettel B, Byerly A, Stover
793 NA, Sugiura M, Harumoto T, et al. 2023. MITE infestation accommodated by
794 genome editing in the germline genome of the ciliate Blepharisma. *Proc Natl*
795 *Acad Sci USA* **120**: e2213985120.

796 Sellis D, Guérin F, Arnaiz O, Pett W, Lerat E, Boggetto N, Krenek S, Berendonk T,
797 Couloux A, Aury J-M, et al. 2021. Massive colonization of protein-coding exons

798 by selfish genetic elements in *Paramecium* germline genomes. *PLoS Biol* **19**:
799 e3001309.

800 Singh A, Maurer-Alcalá XX, Solberg T, Häußermann L, Gisler S, Ignarski M, Swart
801 EC, Nowacki M. 2022. Chromatin remodeling is required for sRNA-guided DNA
802 elimination in Paramecium. *EMBO J* **41**: e111839.

803 Singh A. 2023. ISWI complex proteins facilitate genome editing and development .
804 *Edmond*. https://doi.org/10.17617/3.ZBOLU8 (Accessed July 28, 2023).

805 Skouri F, Cohen J. 1997. Genetic approach to regulated exocytosis using functional
806 complementation in Paramecium: identification of the ND7 gene required for
807 membrane fusion. *Mol Biol Cell* **8**: 1063–1071.

808 Studier FW. 2014. Stable expression clones and auto-induction for protein
809 production in E. coli. *Methods Mol Biol* **1091**: 17–32.

810 Swart EC, Denby Wilkes C, Sandoval PY, Hoehener C, Singh A, Furrer DI,
811 Arambasic M, Ignarski M, Nowacki M. 2017. Identification and analysis of
812 functional associations among natural eukaryotic genome editing components
813 [version 1; peer review: 1 approved, 1 approved with reservations]. *F1000Res*
814 **6**: 1374.

815 Toto M, D'Angelo G, Corona DFV. 2014. Regulation of ISWI chromatin remodelling
816 activity. *Chromosoma* **123**: 91–102.

817 Wang C, Solberg T, Maurer-Alcalá XX, Swart EC, Gao F, Nowacki M. 2022. A small
818 RNA-guided PRC2 complex eliminates DNA as an extreme form of transposon
819 silencing. *Cell Rep* **40**: 111263.

820 Yamada K, Frouws TD, Angst B, Fitzgerald DJ, DeLuca C, Schimmele K, Sargent
821 DF, Richmond TJ. 2011. Structure and mechanism of the chromatin
822 remodelling factor ISW1a. *Nature* **472**: 448–453.

823 Yan L, Wu H, Li X, Gao N, Chen Z. 2019. Structures of the ISWI-nucleosome
824 complex reveal a conserved mechanism of chromatin remodeling. *Nat Struct*
825 *Mol Biol* **26**: 258–266.

826 Zangarelli C, Arnaiz O, Bourge M, Gorrichon K, Jaszczyszyn Y, Mathy N, Escoriza L,
827 Bétermier M, Régnier V. 2022. Developmental timing of programmed DNA
828 elimination in Paramecium tetraurelia recapitulates germline transposon
829 evolutionary dynamics. *Genome Res* **32**: 2028–2042.

830 Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding
831 J, Lupas AN, Alva V. 2018. A Completely Reimplemented MPI Bioinformatics
832 Toolkit with a New HHpred Server at its Core. *J Mol Biol* **430**: 2237–2243.

Figure 1

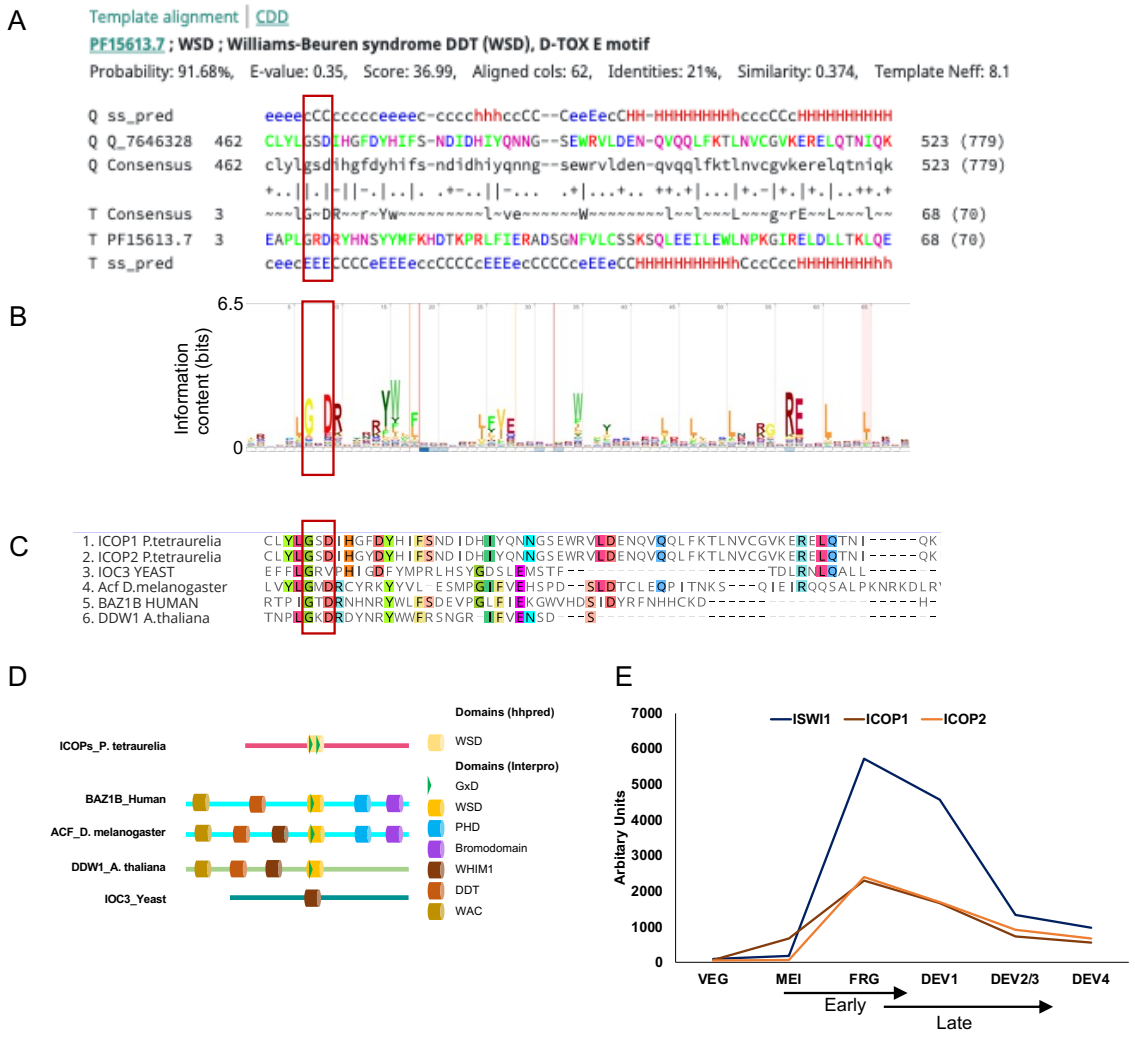


Figure 1: Identification of ISWI Complex Proteins (ICOP).

(A) Template alignment generated by HHpred analysis of ICOP1 showing 91.68% probability match (E-value 0.35) with Williams-Beuren syndrome DDT(WSD) or D-TOX E motif. The conserved GxD signature is highlighted with a red bar.; Q= Query (ICOP1); ss_pred: secondary structure prediction; T= template. For more detailed output description please consult https://toolkit.tuebingen.mpg.de/tools/hhpred. (B) Signature of PFAM model PF15613 from InterPro. (C) Multiple sequence alignment of ICOP1 and ICOP2 WSD/WHIM2+3 domain with ISWI complex/WHIM domain proteins from other organisms. Ioc3: ISWI one complex protein 3 in yeast; ACF: ATP-

dependent chromatin assembly factor large subunit (Acf) from *D. melanogaster*; BAZ1B: human Tyrosine-protein kinase BAZ1B, DDW1: DDT domain-containing protein in *A. thaliana;* red box: GxD signature; highlighted amino acids with >= 50% residues identical to the consensus residue. (D) Domain architecture comparison of ICOPs with ISWI1 complex proteins with WHIM domains. (E) mRNA expression profile (arbitrary units) of ICOP1 and ICOP2 in comparison to ISWI1 during autogamy. VEG: vegetative, MEI: the stage where MICs undergo meiosis and maternal MAC begins to fragment, FRG: about 50% of cells with fragmented maternal MAC, Dev1: the earliest stage with visible developing macronuclei (anlage), Dev2/3: most cells with macronuclear anlage, Dev4: most cells with distinct anlage. MEI and FRG constitute the "Early" time point, and the "Late" time point consists of Dev1 and Dev2/3 stages.

Figure 2

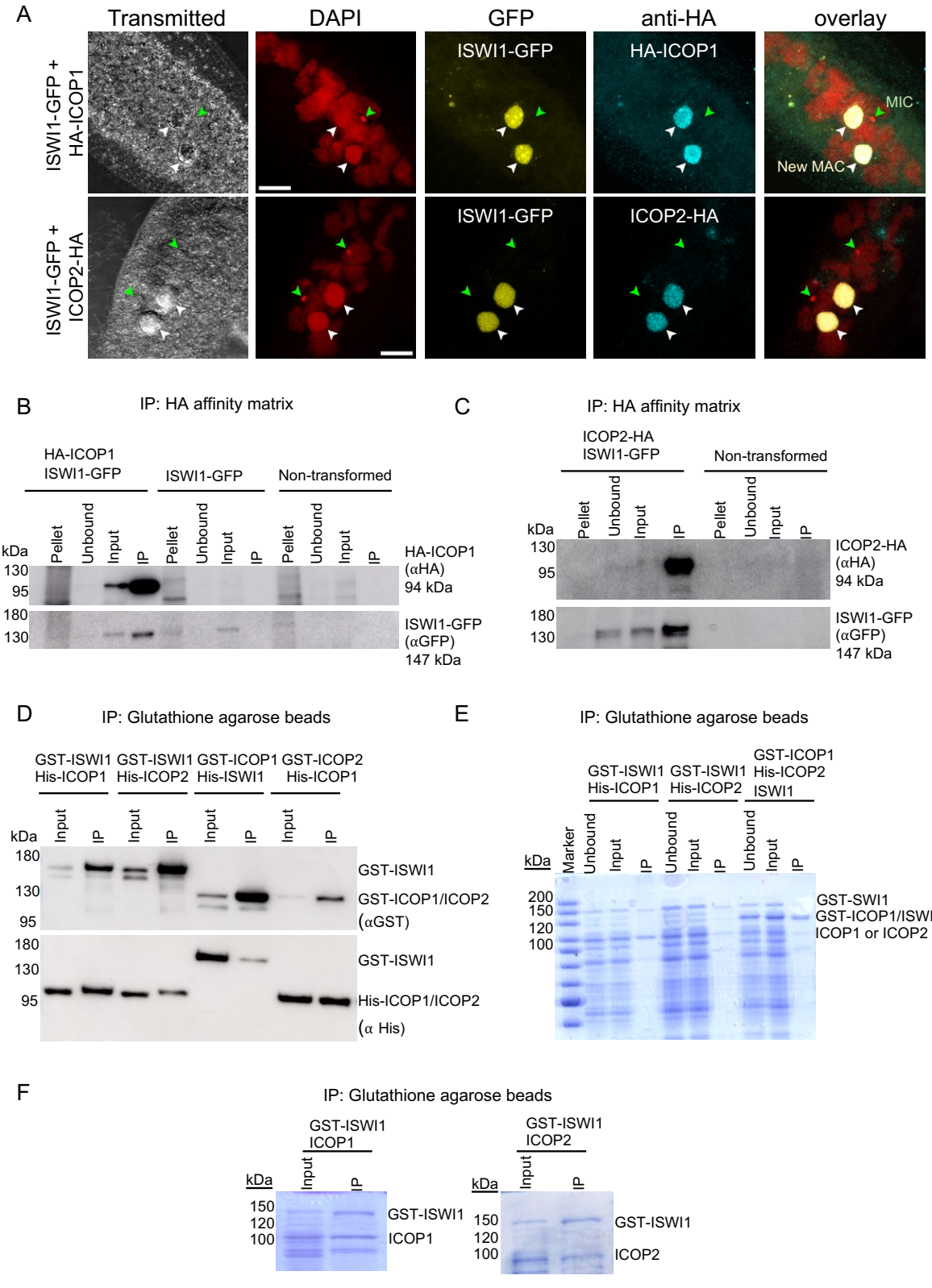

A
Transmitted
DAPI
GFP
anti-HA
overlay
ISWI1-GFP + HA-ICOP1
ISWI1-GFP + ICOP2-HA
ISWI1-GFP
HA-ICOP1
ICOP2-HA
MIC
New MAC
B
IP: HA affinity matrix
HA-ICOP1 ISWI1-GFP
ISWI1-GFP
Non-transformed
Pellet
Unbound
Input
IP
kDa
130
95
180
130
HA-ICOP1 (αHA) 94 kDa
ISWI1-GFP (αGFP) 147 kDa
C
IP: HA affinity matrix
ICOP2-HA ISWI1-GFP
Non-transformed
ICOP2-HA (αHA) 94 kDa
ISWI1-GFP (αGFP) 147 kDa
D
IP: Glutathione agarose beads
GST-ISWI1 His-ICOP1
GST-ISWI1 His-ICOP2
GST-ICOP1 His-ISWI1
GST-ICOP2 His-ICOP1
Input
IP
kDa
180
130
95
GST-ISWI1
GST-ICOP1/ICOP2 (αGST)
GST-ISWI1
His-ICOP1/ICOP2 (α His)
E
IP: Glutathione agarose beads
GST-ISWI1 His-ICOP1
GST-ISWI1 His-ICOP2
GST-ICOP1 His-ICOP2 ISWI1
Marker
Unbound
Input
IP
kDa
200
150
120
100
GST-SWI1
GST-ICOP1/ISWI1
ICOP1 or ICOP2
F
IP: Glutathione agarose beads
GST-ISWI1 ICOP1
GST-ISWI1 ICOP2
Input
IP
kDa
150
120
100
GST-ISWI1
ICOP1
ICOP2

Figure 2: Interaction of ICOP1 and ICOP2 with ISWI1 in new MACs.

(A) Confocal fluorescence microscopy images of HA-ICOP1, ICOP2-HA, and ISWI1-GFP localization: maximum intensity projections of z-planes. Red = DAPI. Yellow = GFP. Cyan = HA. Green arrow = MIC. White arrow = new MAC. All channels were optimized individually for the best visual representation. DAPI channel of ICOP2-HA: Gamma factor = 0.8. Scale bar = 10 µm. (B) & (C) Western blot, co-immunoprecipitation (co-IP) of HA-ICOP1/ISWI1-GFP and ICOP2-HA/ISWI1-GFP in *Paramecium*. Controls: non-transformed and ISWI1-GFP transformed. (D-F) co-IP after *E. coli* expression and pulldown; (D) Western blot, (E&F) Coomassie staining. GST-ISWI: 147 kDa, His-ISWI1: 122 kDa, His-ICOP1 & His-ICOP2: 95 kDa, GST-ICOP1/ICOP2: 119 kDa, untagged ISWI1: 120 kDa, untagged ICOP1 & ICOP2: 93 kDa.

Figure 3

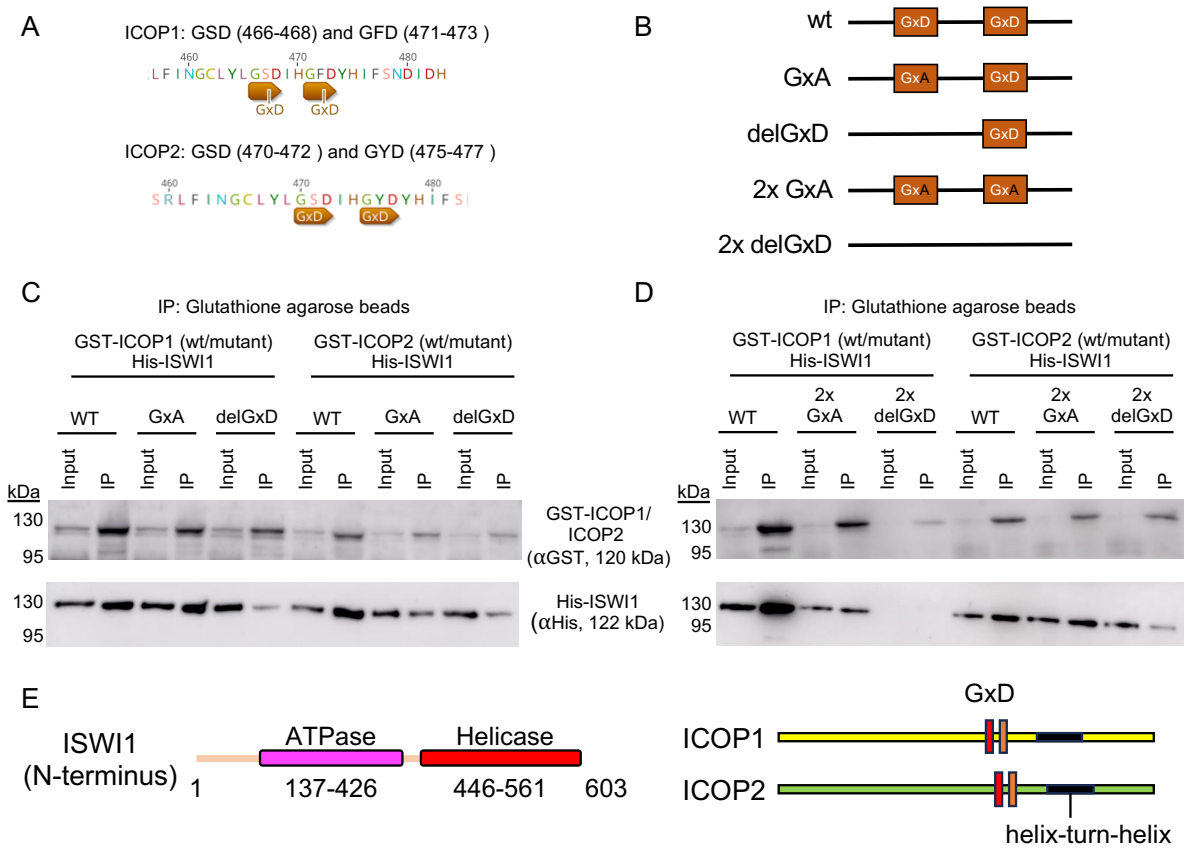

A
ICOP1: GSD (466-468) and GFD (471-473)
460 470 480
LF INGCLYLGSDIHGFDYHIFSNDIDH
GxD GxD
ICOP2: GSD (470-472) and GYD (475-477)
460 470 480
SRLFINGCLYLGSDIHGYDYHIFSI
GxD GxD
B
wt GxD GxD
GxA GxA GxD
delGxD GxD
2x GxA GxA GxA
2x delGxD
C
IP: Glutathione agarose beads
GST-ICOP1 (wt/mutant) His-ISWI1
GST-ICOP2 (wt/mutant) His-ISWI1
WT GxA delGxD WT GxA delGxD
Input IP Input IP Input IP Input IP Input IP Input IP
kDa
130
95
GST-ICOP1/ ICOP2 (αGST, 120 kDa)
His-ISWI1 (αHis, 122 kDa)
D
IP: Glutathione agarose beads
GST-ICOP1 (wt/mutant) His-ISWI1
GST-ICOP2 (wt/mutant) His-ISWI1
WT 2x GxA 2x delGxD WT 2x GxA 2x delGxD
Input IP Input IP Input IP Input IP Input IP Input IP
kDa
130
95
E
ISWI1 (N-terminus)
ATPase
Helicase
1
137-426
446-561
603
GxD
ICOP1
ICOP2
helix-turn-helix
F
ISWI1 (N-terminus) – ICOP1
predicted interaction
GSD
GFD
G
ISWI1 (N-terminus) – ICOP2
predicted interaction
GSD
GYD
H
ISWI1 (N-terminus) – ICOP1/ICOP2
I
ISW1a (N-terminus) – Ioc3
Q
G
I

Figure 3: Investigation of the GxD signature in ICOP/ISWI1 interaction.

(A) Screenshots from Geneious Prime (version 2023.1.1) showing GxD signature in ICOP1 and ICOP2. (B) Schematic representation of GxD mutants generated. (C) & (D) Western blot on co-IP of GST-ICOP GxD mutants and His-ISWI1 overexpressed in *E. coli* probed with anti-GST and anti-His antibodies; GST-ICOP wild-type is a control. (E) Schematic representation of the sequences used for complex predictions in (F) & (G). (F-H) Structure prediction of multimers (ISWI1 N-terminus (residues 1-603) with ICOP1 or ICOP2) with AlphaFold (version 2.2.0). ICOP1: yellow, ICOP2: green, GSD signature: red, GFD/GYD signature: orange, ISWI1: wheat, ISWI1 ATPase domain: magenta, ISWI1 helicase domain: red. (F) & (G) ISWI1-ICOP1 and ISWI1-ICOP2 interaction, respectively. Predicted interaction interface and GxD signatures are circled. (H) ISWI1 N-terminus with interacting helices of ICOP paralogs (ICOP1: residues 556-597; ICOP2: residues 560-603). Proximate residues on ISWI1 are shown in blue. Proximate residues of ICOPs are shown as sticks. (I) GxD signature in the published crystal structure (PDB accession number 2Y9Y): ISW1a (del_ATPase; cyan) and Ioc3 (WHIM containing protein; dark salmon) from yeast. GxD signature (GIQ in Ioc3) and spatially close residues in ISW1a are shown as sticks, polar contacts between the proteins are yellow.

Figure 4

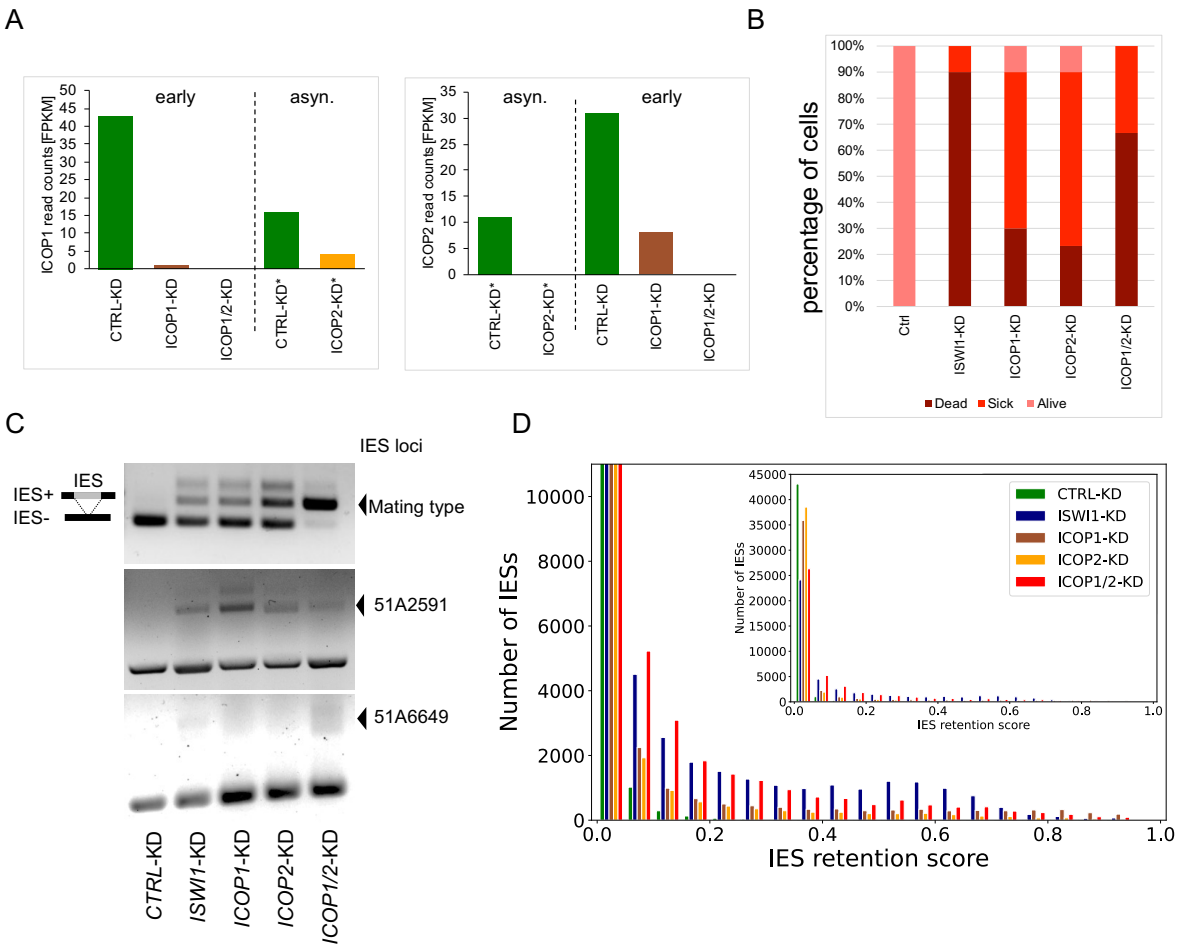

A
ICOP1 read counts [FPKM]
early
asyn.
CTRL-KD
ICOP1-KD
ICOP1/2-KD
CTRL-KD*
ICOP2-KD*
ICOP2 read counts [FPKM]
asyn.
early
CTRL-KD*
ICOP2-KD*
CTRL-KD
ICOP1-KD
ICOP1/2-KD
B
percentage of cells
Ctrl
ISWI1-KD
ICOP1-KD
ICOP2-KD
ICOP1/2-KD
Dead
Sick
Alive
C
IES loci
IES
IES+
IES-
Mating type
51A2591
51A6649
CTRL-KD
ISWI1-KD
ICOP1-KD
ICOP2-KD
ICOP1/2-KD
D
Number of IESs
IES retention score
CTRL-KD
ISWI1-KD
ICOP1-KD
ICOP2-KD
ICOP1/2-KD
E
PGM
TFIIS4
PTIWI01/09
ISWI1
ICOP1/2
0.21
0.07
0.16
0.13
0.43
0.67
0.59
0.50
0.50
0.75

Figure 4: Effects of *ICOP* knockdowns on DNA excision.

(A) mRNA expression levels in FPKM (Fragments per kilobase per million mapped reads) compared between ICOP1 and ICOP2 knockdowns for transcripts early in development (40% old MAC fragmentation) or asynchronous (asyn.) culture (*). (B) Survival of recovered post-autogamous knockdown cells followed for several vegetative divisions. Alive (pink): normal division. Sick (red): slower division rate. Dead (cayenne): no cells. (C) Retention of individual IESs, *ISWI1*-KD = positive control. Retained IESs (IES+) result in a larger amplicon (D) Genome-wide IES retention across ˜45,000 IESs in different KDs. For each IES, IES retention score, IRS = IES+ reads/(IES+ reads + IES- reads). (E) Correlation of IRSs among KDs. Diagonal: IRS distributions of individual KDs. Below diagonal: correlation graphs of pairwise comparisons. Above diagonal: corresponding Pearson correlation coefficients. Red lines: ordinary least-squares (OLS) regression, orange lines: LOWESS, and grey lines: orthogonal distance regression (ODR).

Figure 5

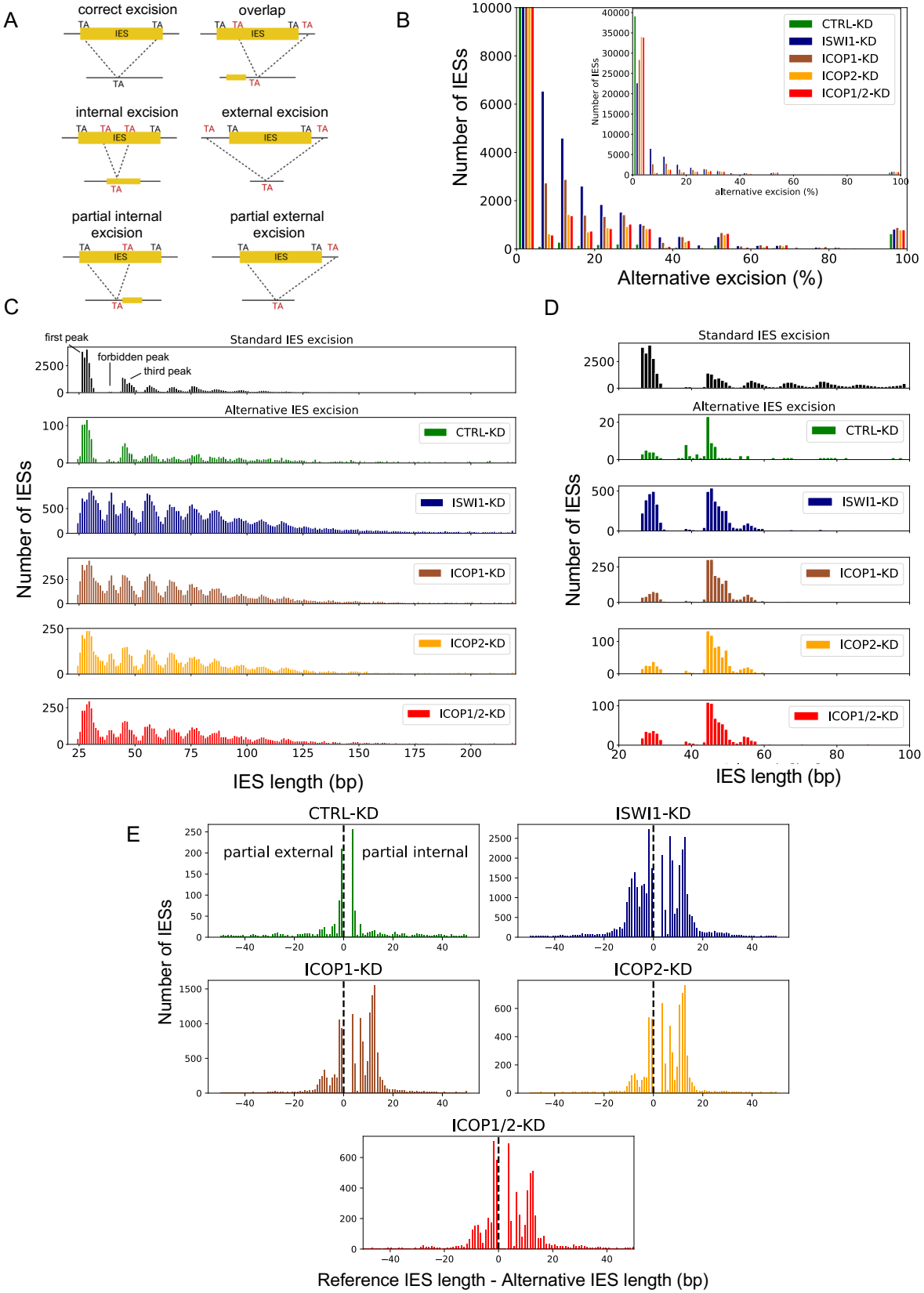

A
correct excision
overlap
internal excision
external excision
partial internal excision
partial external excision
IES
TA
B
Number of IESs
Alternative excision (%)
CTRL-KD
ISWI1-KD
ICOP1-KD
ICOP2-KD
ICOP1/2-KD
C
D
first peak
forbidden peak
third peak
Standard IES excision
Alternative IES excision
IES length (bp)
E
CTRL-KD
ISWI1-KD
ICOP1-KD
ICOP2-KD
ICOP1/2-KD
partial external
partial internal
Reference IES length - Alternative IES length (bp)

Figure 5: Alternative IES excision in *ICOP* and other relevant knockdowns.

(A) Schematic representation of analyzed IES excision events. (B) Distribution of genome-wide alternative IES excision (percent per IES) for different KDs. (C) Length distribution of alternatively excised IESs for each KD. The reference length distribution for all IESs is given above ("Standard IES excision"). (D) Origin of alternatively excised IESs in the "forbidden" peak. The reference length is plotted for all alternatively excised 34 – 44 bp IESs. (E) Length distribution of partial external and partial internal alternative excision events for the KDs.

Figure 6

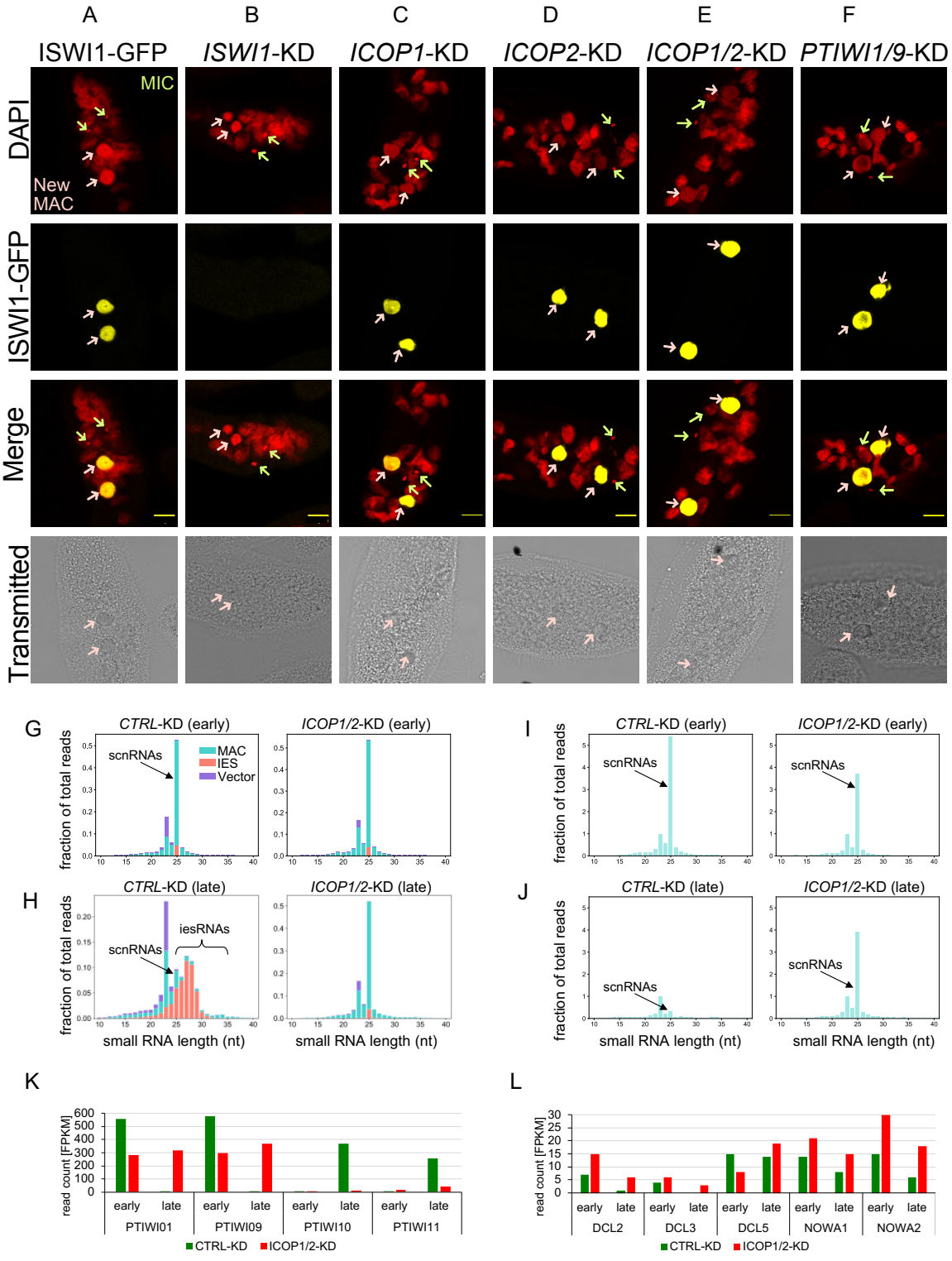

A
B
C
D
E
F
ISWI1-GFP
ISWI1-KD
ICOP1-KD
ICOP2-KD
ICOP1/2-KD
PTIWI1/9-KD
DAPI
ISWI1-GFP
Merge
Transmitted
MIC
New MAC
G
CTRL-KD (early)
ICOP1/2-KD (early)
fraction of total reads
scnRNAs
MAC
IES
Vector
H
CTRL-KD (late)
ICOP1/2-KD (late)
iesRNAs
small RNA length (nt)
I
J
K
L
read count [FPKM]
early
late
PTIWI01
PTIWI09
PTIWI10
PTIWI11
CTRL-KD
ICOP1/2-KD
DCL2
DCL3
DCL5
NOWA1
NOWA2

Figure 6: Effects of ICOP1 and ICOP2 knockdowns on ISWI-GFP localization, sRNAs and gene expression.

(A-F) Confocal fluorescence microscopy of ISWI1-GFP localization under gene knockdowns. (A) Positive control: ISWI1-GFP transformed cells without RNAi; Red = DAPI, Yellow = GFP. Green arrow = MIC; pink arrow = new MAC, scale bar = 10 μm. (G-J) sRNA histograms. (G & H) sRNA reads mapped to the L4440 plasmid sequence (Vector, purple), macronuclear genome (MAC, cyan), and IESs (IES, pink). (I & J) Histogram of MAC genome-matching sRNAs normalized against MAC genome-matching siRNAs. Early = 40% of cells have fragmented MAC, Late = most cells with visible new MAC. (K & L) Histogram of mRNA expression levels in FPKM (Fragments per kilobase per million mapped reads) for different development-specific genes.

Figure 7

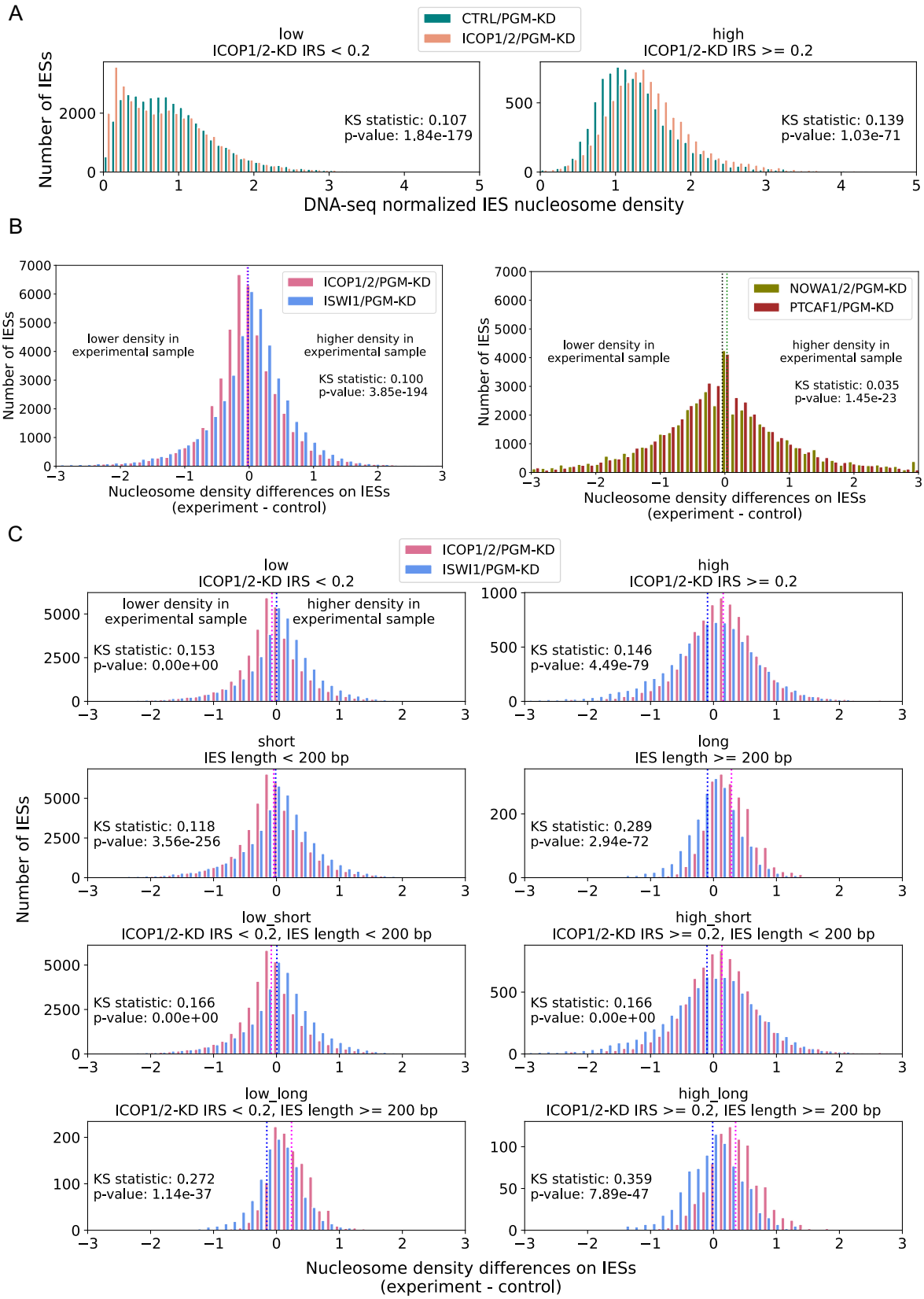
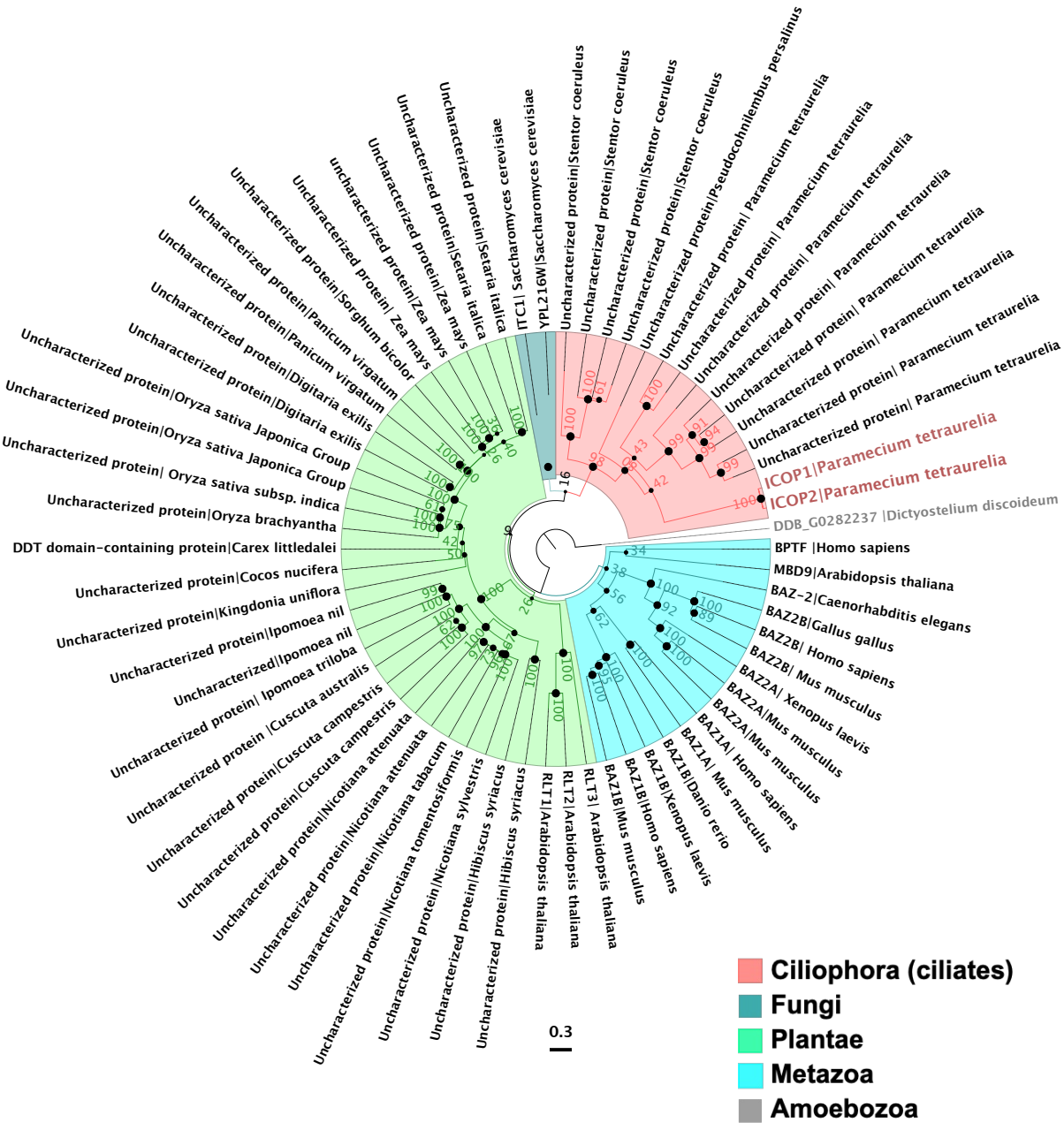

A
low
ICOP1/2-KD IRS < 0.2
high
ICOP1/2-KD IRS >= 0.2
CTRL/PGM-KD
ICOP1/2/PGM-KD
Number of IESs
KS statistic: 0.107
p-value: 1.84e-179
KS statistic: 0.139
p-value: 1.03e-71
DNA-seq normalized IES nucleosome density
B
ICOP1/2/PGM-KD
ISWI1/PGM-KD
lower density in experimental sample
higher density in experimental sample
KS statistic: 0.100
p-value: 3.85e-194
Nucleosome density differences on IESs (experiment - control)
NOWA1/2/PGM-KD
PTCAF1/PGM-KD
KS statistic: 0.035
p-value: 1.45e-23
C
low
ICOP1/2-KD IRS < 0.2
high
ICOP1/2-KD IRS >= 0.2
KS statistic: 0.153
p-value: 0.00e+00
KS statistic: 0.146
p-value: 4.49e-79
short
IES length < 200 bp
long
IES length >= 200 bp
KS statistic: 0.118
p-value: 3.56e-256
KS statistic: 0.289
p-value: 2.94e-72
low_short
ICOP1/2-KD IRS < 0.2, IES length < 200 bp
high_short
ICOP1/2-KD IRS >= 0.2, IES length < 200 bp
KS statistic: 0.166
p-value: 0.00e+00
KS statistic: 0.166
p-value: 0.00e+00
low_long
ICOP1/2-KD IRS < 0.2, IES length >= 200 bp
high_long
ICOP1/2-KD IRS >= 0.2, IES length >= 200 bp
KS statistic: 0.272
p-value: 1.14e-37
KS statistic: 0.359
p-value: 7.89e-47
Nucleosome density differences on IESs (experiment - control)

Figure 7: Nucleosome density changes associate with *ICOP* knockdowns.

(A) Normalized nucleosome densities across IESs for *ICOP1/2/PGM*-KD and *CTRL/PGM*-KD. IESs are grouped as low (IRS < 0.2) or high (IRS ≥ 0.2) IRS in *ICOP1/2*-KD. (B) Nucleosome density differences for all IESs. Means are dashed lines (*ICOP1/2/PGM*-KD: magenta; *ISWI1/PGM*-KD: blue; *NOWA1/2/PGM*-KD: green; *PTCAF1/PGM*-KD: black). (C) Comparison of *ICOP1/2/PGM*-KD and *ISWI1/PGM*-KD in selected IES groups: IESs were grouped by IRS in *ICOP1/2*-KD (low: IRS < 0.2; high: IRS ≥ 0.2) and IES length (short: IES length < 200 bp; long: IES length ≥ 200 bp). IES group is given above the diagrams. Means are dashed lines (*ICOP1/2/PGM*-KD: magenta; *ISWI1/PGM*-KD: blue).

Supplemental Figure S1

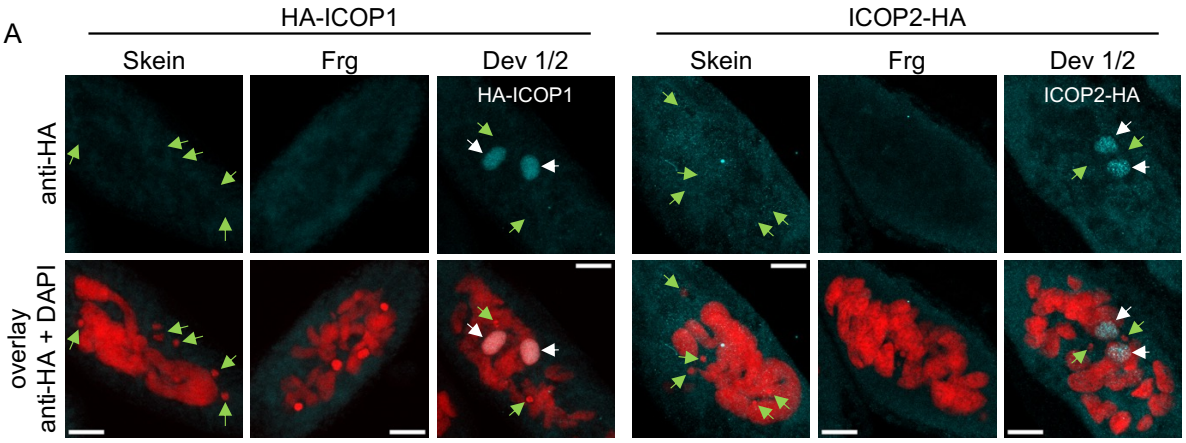


Supplemental Figure S1: WSD-containing proteins are highly diverse.

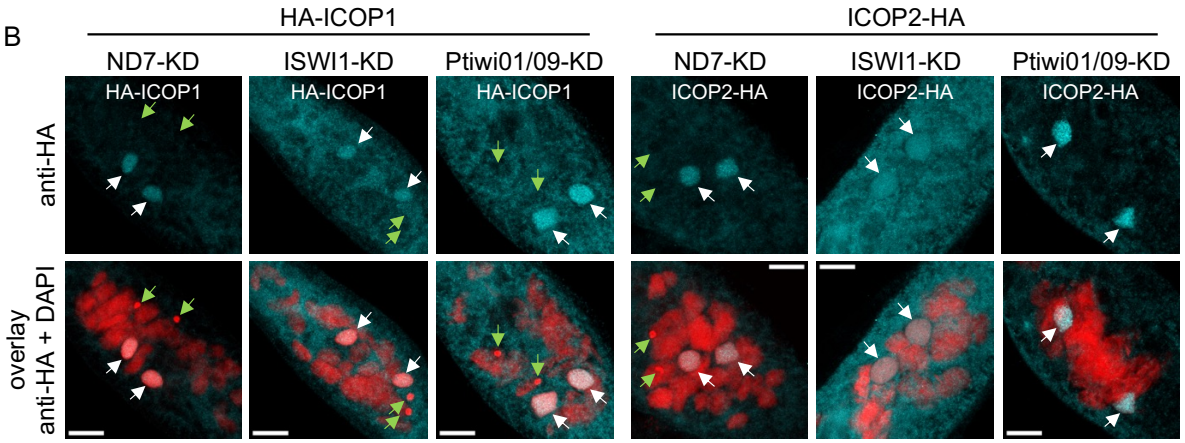
Phylogenetic analysis of proteins with WHIM2 domain in selected organisms. Node bootstrap values are labeled, and the '•' size corresponds to node values. The tree is rooted at the *Dictyostelium discoideum* sequence, labeled in gray. Scale bar is 0.3 amino acid substitutions per site. ICOP1 and ICOP2 are labeled in salmon.

Supplemental Figure S2

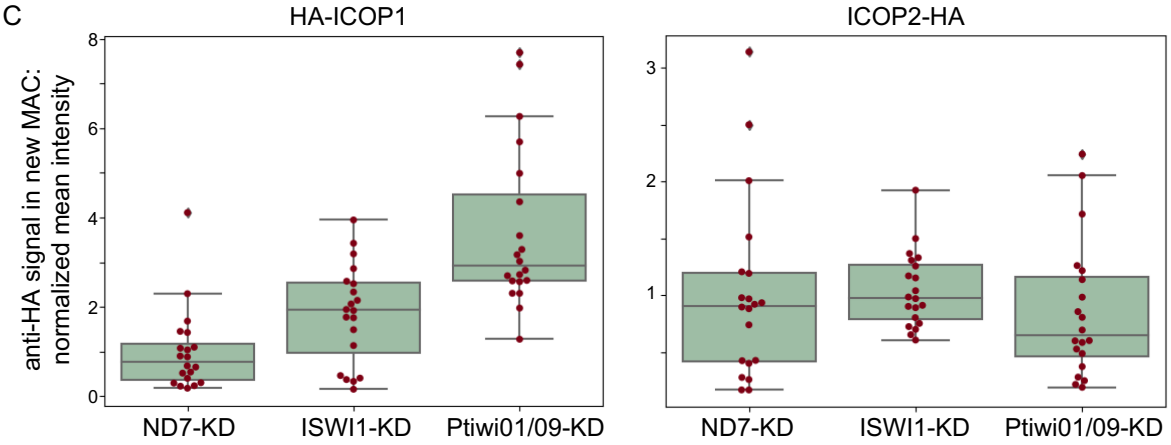
A



B



C



D

HA-ICOP1

group 1	group 2	diff	p-value
ND7-KD	ISWI1-KD	0.840	0.125
ND7-KD	Ptiwi1/9-KD	2.666	≤ 0.001
ISWI1-KD	Ptiwi1/9-KD	1.826	≤ 0.001

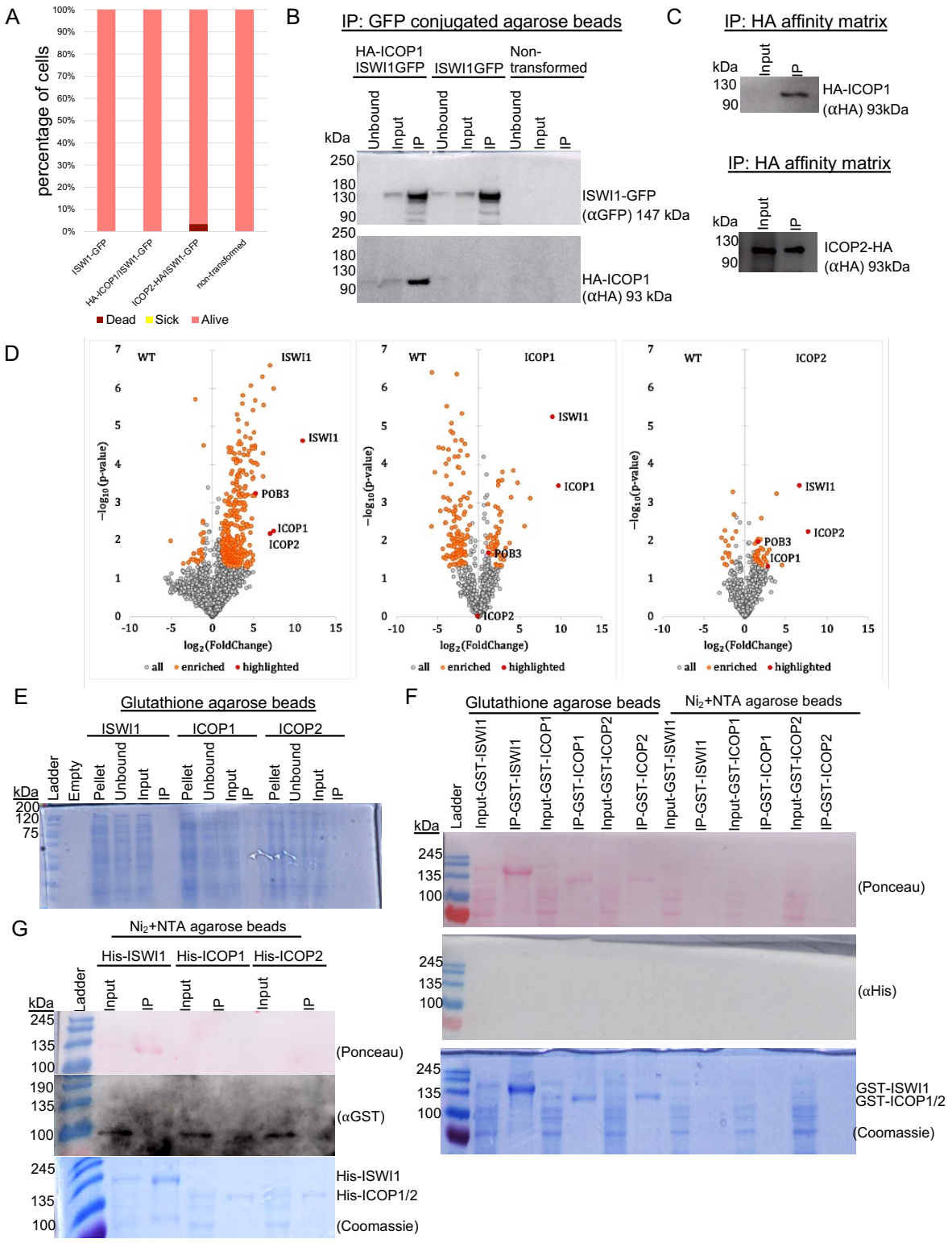
ICOP2-HA

group 1	group 2	diff	p-value
ND7-KD	ISWI1-KD	0.047	≥ 0.9
ND7-KD	Ptiwi1/9-KD	0.147	0.705
ISWI1-KD	Ptiwi1/9-KD	0.194	0.564

Supplemental Figure S2: ICOP localization.

(A) & (B) Confocal fluorescence microscopy images of HA-ICOP1 and ICOP2-HA localization at different developmental stages (A) or upon KDs (B). Maximum intensity projections of z-planes. Red = DAPI. Cyan = HA. Green arrow = MIC. White arrow = new MAC. Brightness and contrast in HA-channel are constant across all HA-ICOP1 and across all ICOP2-HA images. Scale bar = 10 µm. (C) anti-HA signal quantification in new MACs: Mean fluorescence intensity for different KDs normalized to ND7-KD. Black line: median. Whiskers: 1.5 times the interquartile range of the lower or upper quartile. Brown dots: data points (sample size = 20). (D) Tukey post-hoc test statistics: comparisons (group 1 and group 2), difference between means (diff), and p-value.

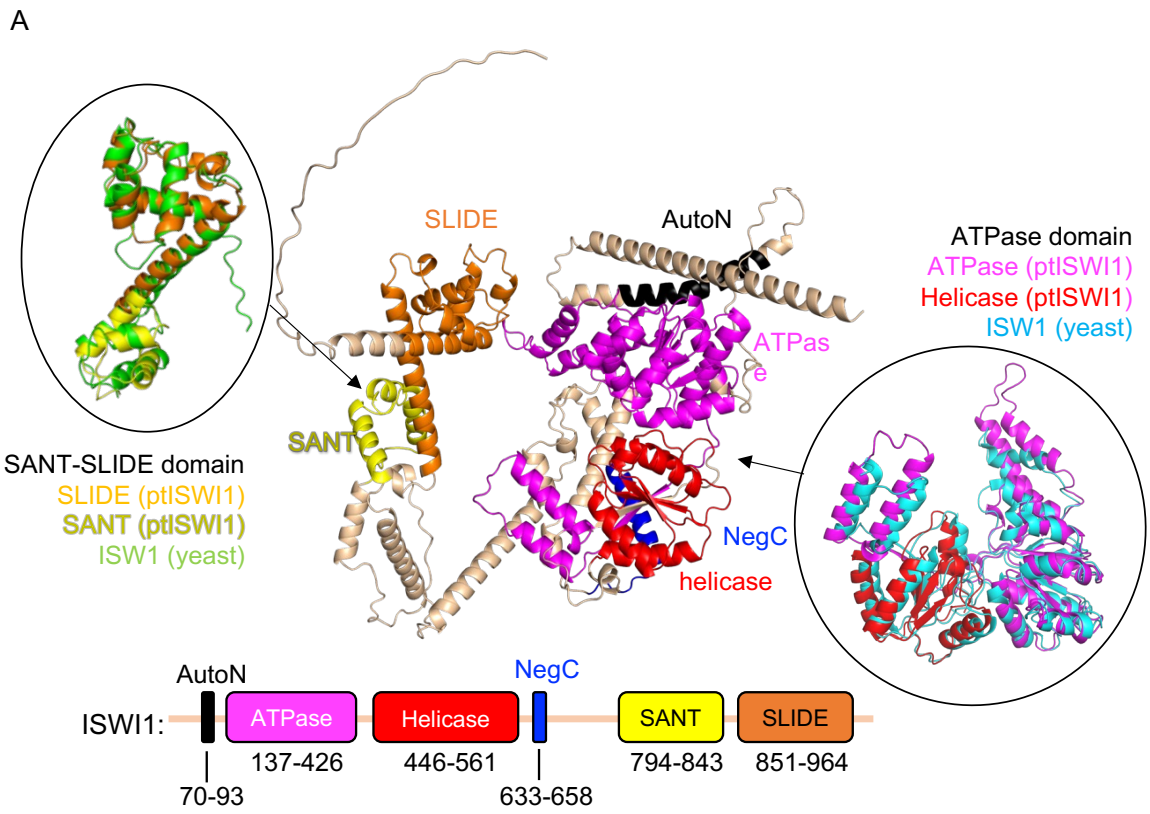
Supplemental Figure S3

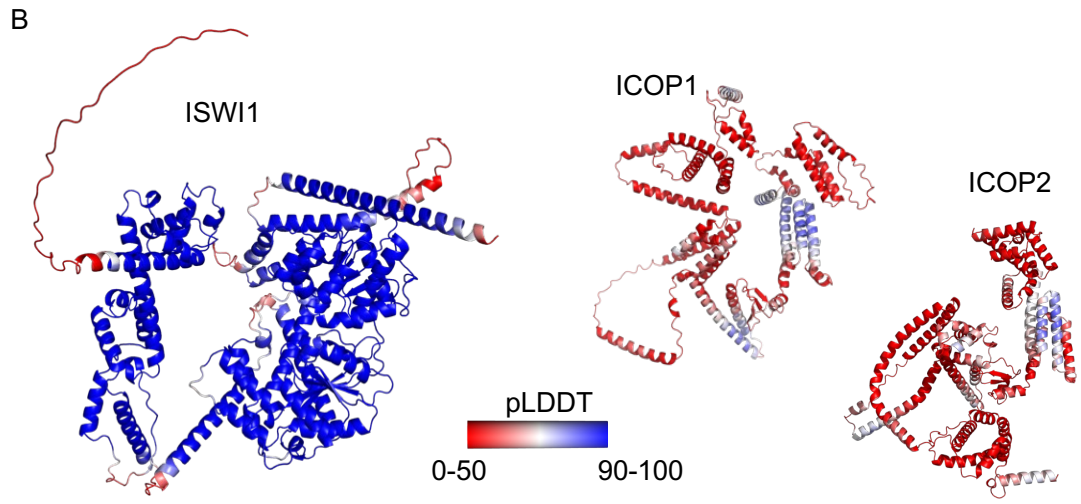


Supplemental Figure S3: ICOP paralogs interact with ISWI1.

(A) Survival assay of F1 generation after knockdown. Alive (pink): normal division. Sick (yellow): slower division rate. Dead (Cayenne): no cells. (B) Western blot on co-IP of HA-ICOP1/ISWI1-GFP co-transformed, ISWI1-GFP transformed and non-transformed, wild-type *Paramecium*. (C) Western blot on co-IP of HA-ICOP1 and ICOP2-HA overexpressed in paramecia. (D) Volcano plots showing protein enrichment of mass spectrometry (MS) analysis for ISWI1-GFP (left), HA-ICOP1 (middle), and ICOP2-HA (right) co-IP. (E) to (F): Pulldowns on overexpressed recombinant proteins in *E. coli*. (E) Coomassie staining of untagged ISWI1, ICOP1 and ICOP2. (F) Western blot and Coomassie staining of GST-tagged recombinant protein pulldowns; Ponceau-stained membranes probed with anti-His antibody. (G) Western blot and Coomassie staining of His-tagged recombinant proteins; Ponceau-stained membranes probed with anti-GST antibody.

Supplemental Figure S4

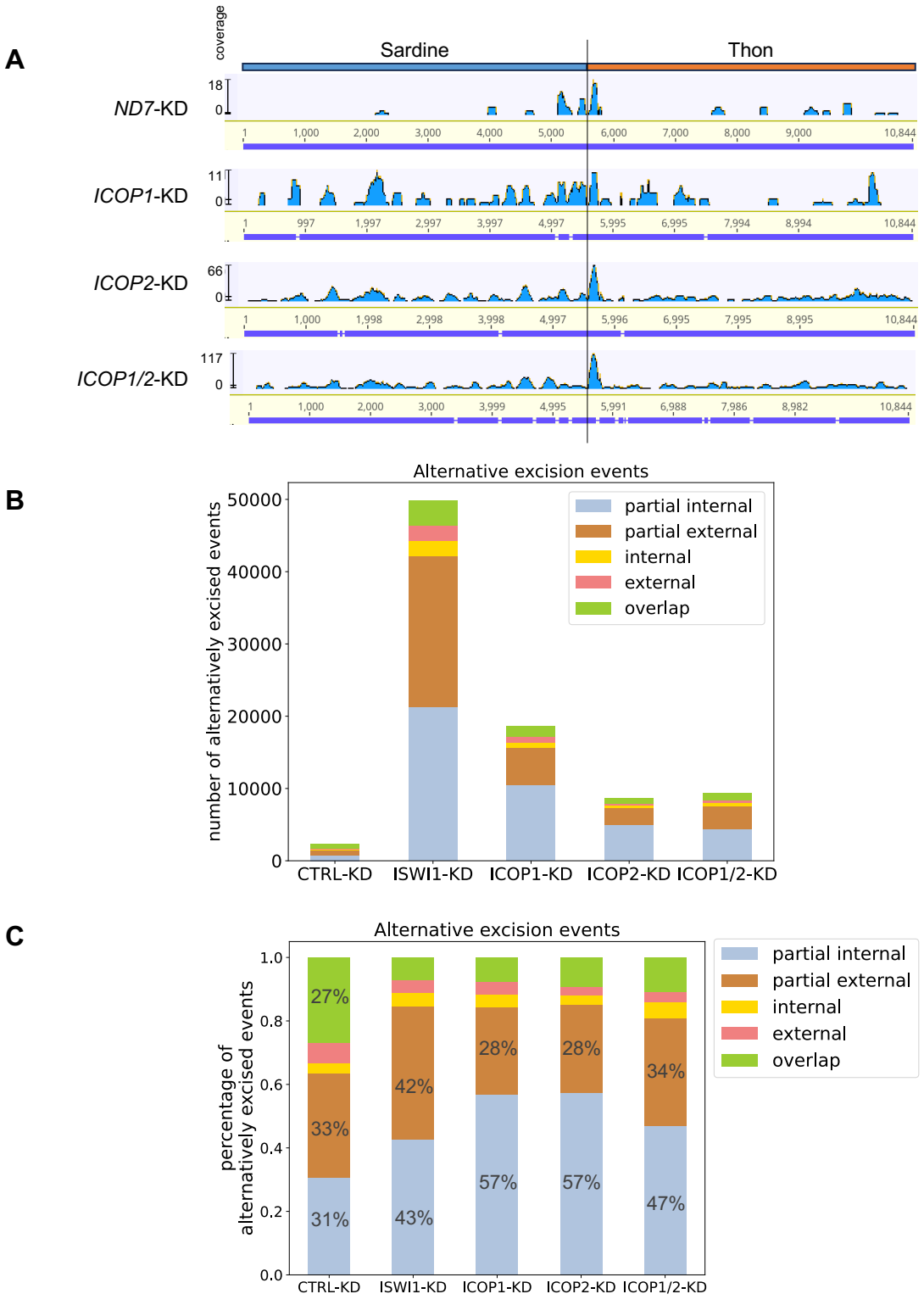




Supplemental Figure S4: ISWI1 and ICOP structure predictions.

(A) and (B) AlphaFold (version 2.2.0) structure predictions. (A) Domains in *Paramecium* ISWI1. ATPase and Helicase are superimposed with a published structure of N-terminal ISWI from yeast (PDB accession number 6JYL) (color: cyan) and SANT-SLIDE domains are superimposed with ISW1a (del_ATPase) from yeast (PDB accession number 2Y9Y) (color: green). (B) Structure prediction confidence for ISWI1, ICOP1, and ICOP2. Models are colored by predicted local distance difference test (pLDDT). pLDDT ≤ 50 are represented in red. pLDDT ≥ 90 are represented in blue.

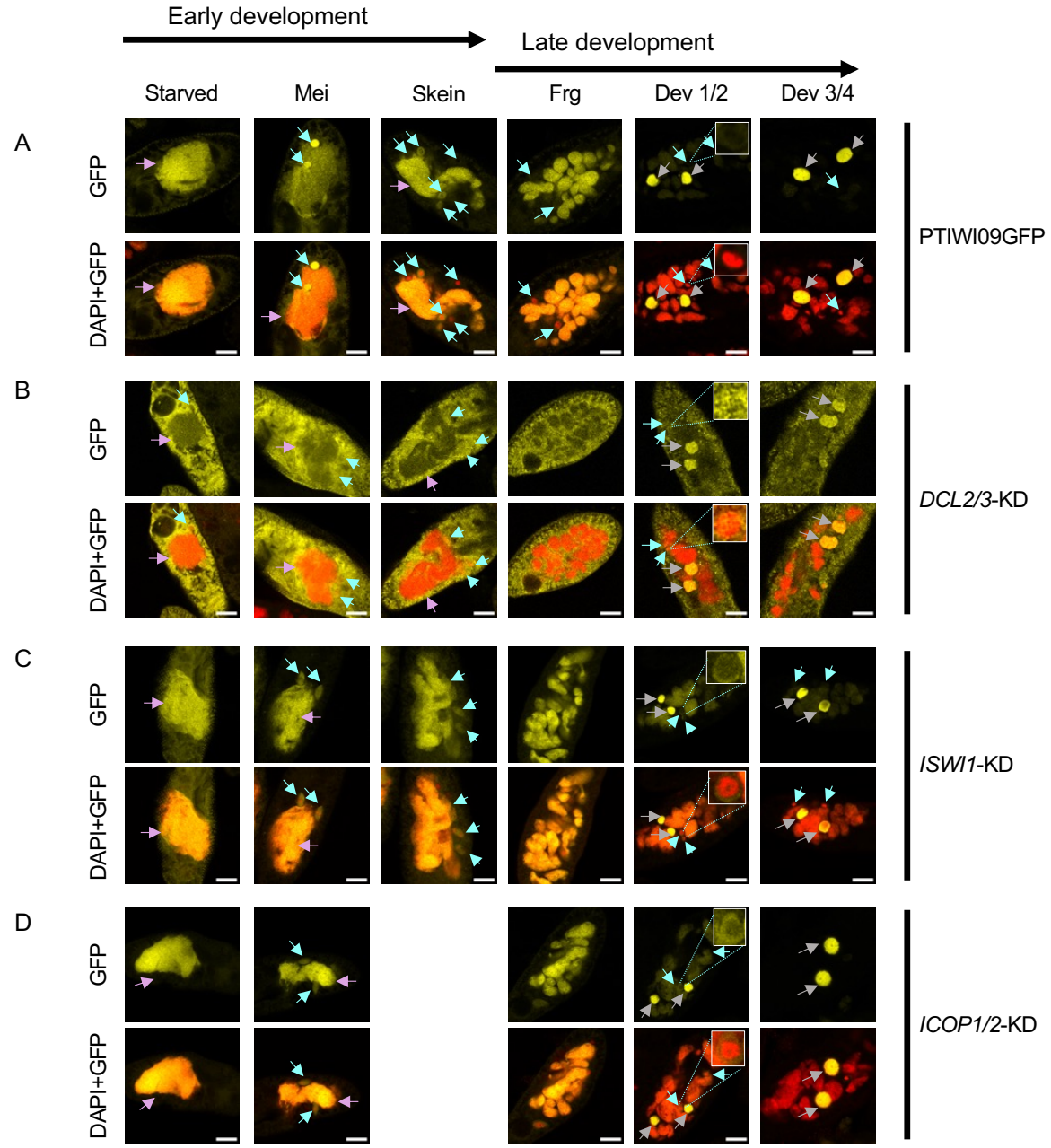
Supplemental Figure S5



Supplemental Figure S5: DNA elimination events.

(A) Geneious screenshot of reads mapped to Sardine and Thon transposons (ENA HE774469) upon different knockdowns. (B & C) Stacked bar graphs of alternative excision events detected in *ISWI1*-KD, *ICOP1*-KD, *ICOP2*-KD and *ICOP1/2*-KD. *ND7*-KD was used as a control. (B) Absolute and (C) relative abundance of alternative excision events occurring upon KDs.

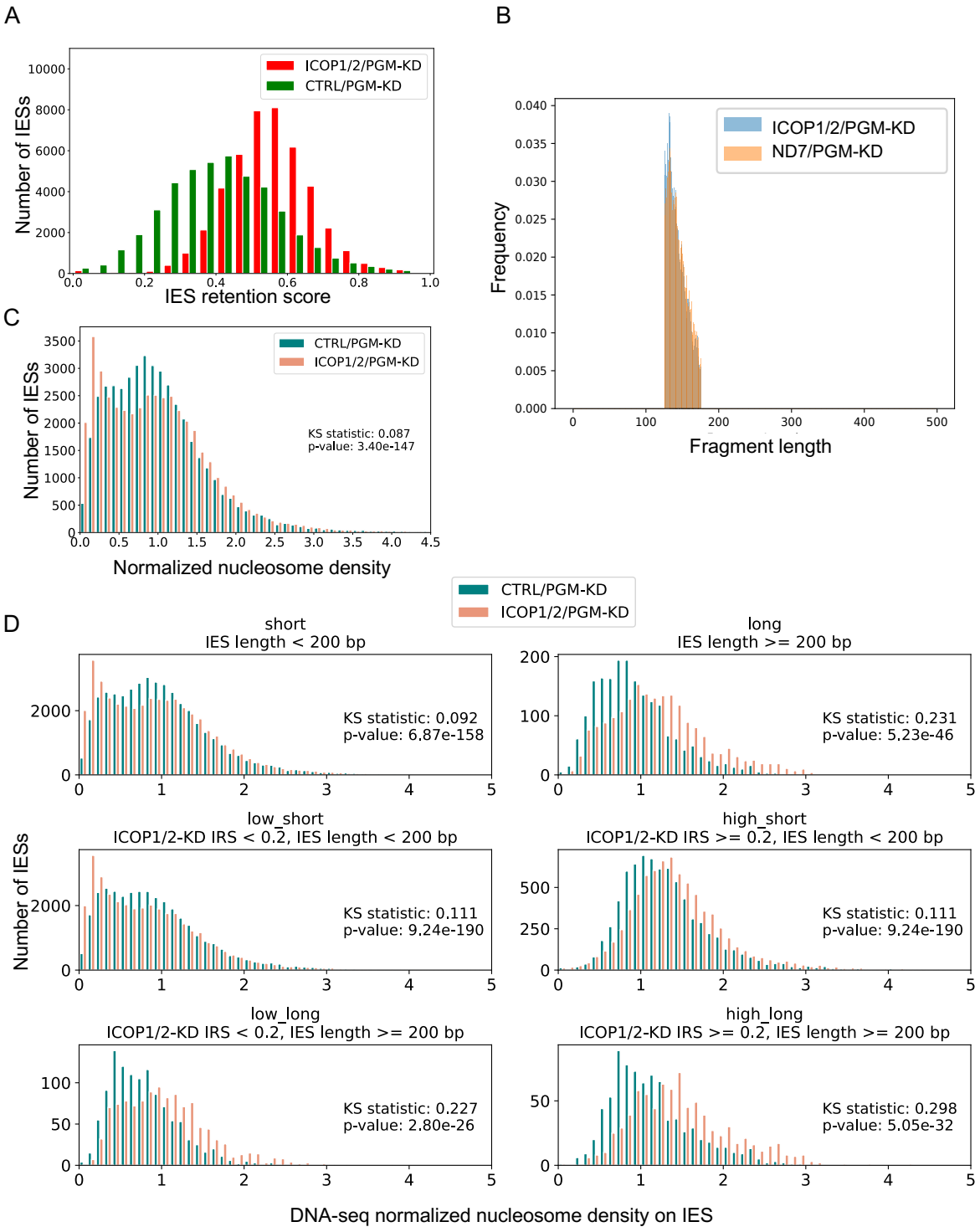
Supplemental Figure S6



Supplemental Figure S6: Confocal microscopy images of PTIWI09GFP localization upon different knockdowns.

PTIWI09GFP localization without any knockdown (A), upon *DCL2/3*-KD (B), upon *ISWI1*-KD (C), upon *ICOP1/2*-KD (D). MIC is cropped, magnified, and contrast adjusted in Dev1/2 panel for better visualization. Red = DAPI. Yellow = GFP. Purple arrow = Maternal MAC. Cyan arrow = MICs. Grey Arrow = Developing MACs. All channels were optimized for the best visual representation. Scale bar = 10μm. Starved = induction of autogamy, Mei = MIC meiosis, Skein = beginning of macronuclear fragmentation, Frg = fragmentation of maternal MAC, Dev1/2 = visible new MAC, Dev3/4 = larger new MAC

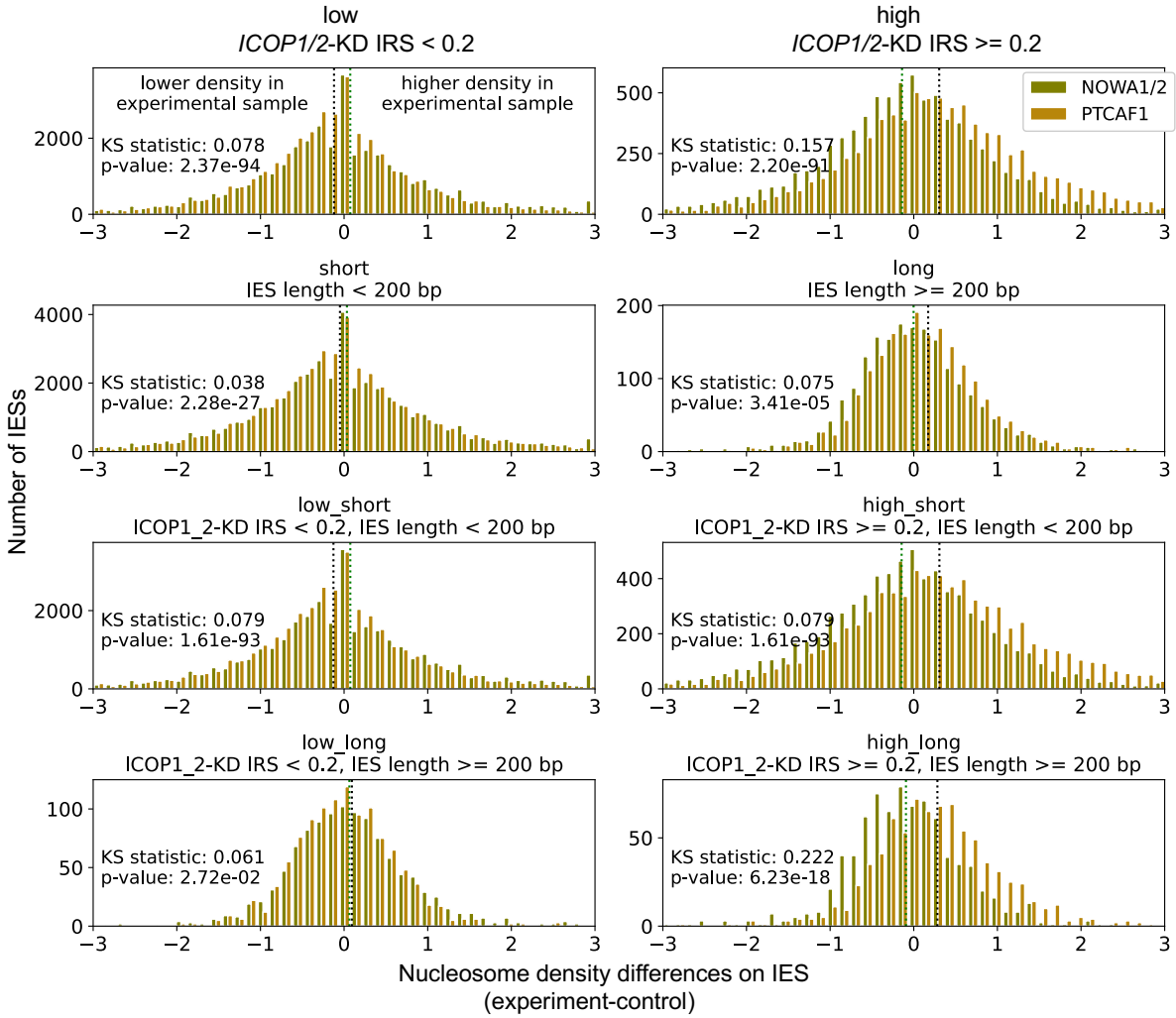
Supplemental Figure S7



Supplemental Figure S7: Nucleosome densities for *ICOP1/2/PGM*-KD and *CTRL/PGM*-KD.

(A) IRS histogram for *ICOP1/2/PGM*-KD and *CTRL/PGM*-KD. (B) Size distribution of reads mapped to scaffold51_9 for *ICOP1/2/PGM*-KD and *CTRL/PGM*-KD. (C) Nucleosome densities of all IESs in *ICOP1/2/PGM*-KD and *CTRL/PGM*-KD. (D) Nucleosome densities of selected IES groups in *ICOP1/2/PGM*-KD and *CTRL/PGM*-KD. IESs were grouped by IES retention score (IRS) in *ICOP1/2*-KD (low: IRS < 0.2; high: IRS ≥ 0.2) and IES length (short: IES length < 200 bp; long: IES length ≥ 200 bp). IES group is given above the diagrams.

Supplemental Figure S8



Supplemental Figure S8: Nucleosome density differences for *NOWA1/2/PGM*-KD and *PTCAF1/PGM*-KD.

Comparison of *NOWA1/2/PGM*-KD and *PTCAF1/PGM*-KD nucleosome density differences in selected IES groups: IESs were grouped by IES retention score (IRS) in *ICOP1/2*-KD (low: IRS < 0.2; high: IRS ≥ 0.2) and IES length (short: IES length < 200 bp; long: IES length ≥ 200 bp). The specification for each IES group is given above the individual diagrams. Means are indicated as dashed lines (*NOWA1/2/PGM*-KD: green; *PTCAF1/PGM*-KD: black).

Supplemental Table S1: Predicted interaction in AlphaFold2 models

AF2 version	model	interaction predicted	GxD involved	interaction interface
v2.2.0	ISWIN + ICOP1	yes	no	ICOP1 resi 556-597; ISWI1 resi 425-426+431+434-437+474+477-478+481
	ISWIN + ICOP2	yes	no	ICOP2 resi 560-603; ISWI1 resi 425-426+431+434-437+474+477-478+481
	ISWIC + ICOP1	no	-	-
	ISWIC + ICOP2	no	-	-
	ISWI1 + ICOP1	no	-	-
	ISWI1 + ICOP2	no	-	-
	ICOP1 + ICOP2	no	-	-
	ISWIN + ICOP1/2	no	-	-
	ISWIN + Ptiwi01	no	-	-
	ISWIC + Ptiwi01	yes	-	-
v2.3.0	ISWIN + ICOP1	yes	no	large interaction interface
	ISWIN + ICOP2	yes	no	large interaction interface
	ISWIC + ICOP1	yes	no	large interaction interface
	ISWIC + ICOP2	yes	no	large interaction interface
	ISWI1 + ICOP1	yes	no	mostly ISWI1 N-terminus
	ISWI1 + ICOP2	yes	no	mostly ISWI1 N-terminus
	ICOP1 + ICOP2	yes	-	-

Predicted interactions between multimers with AlphaFold2. ISWI1 was either predicted as full length (ISWI1) or split version (ISWIN or ISWIC, referring to N-terminus and C-terminus, respectively). All other proteins were provided as full length. For detailed input sequences refer to Supplemental Table 4.

Supplemental Table S2: Percentage of alternatively excised IESs

	including IES with 100% alternative excision			excluding IES with 100% alternative excision		
	median	mean	IESs	median	mean	IESs
CTRL-KD	0.00 %	2.59 %	41311	0.00 %	1.09 %	40680
ISWI1-KD	4.55 %	10.86 %	43983	4.35 %	9.18 %	43171
ICOP1-KD	0.00 %	8.97 %	42237	0.00 %	7.01 %	41349
ICOP2-KD	0.00 %	6.00 %	41573	0.00 %	4.18 %	40785
ICOP1/2-KD	0.00 %	6.53 %	41767	0.00 %	4.71 %	40972

Median and mean percentage (in %) of alternative excision for all IESs in the KDs. The number of IESs included in the analysis is given ("IESs"). IESs with 100% alternative excision are either included (left) or excluded (right).

Supplemental Table S3: Length differences in partial external and partial internal alternative excision

	partial external -(reference length – alternative length)				partial internal (reference length – alternative length)			
	min	max	mean	median	min	max	mean	median
CTRL-KD	1	4933	88,29	8	3	3272	62,15	7
ISWI1-KD	1	9538	78,82	8	3	4357	21,80	11
ICOP1-KD	1	8148	110,81	5	3	2394	15,62	10
ICOP2-KD	1	7701	52,06	4	3	2049	15,66	10
ICOP1/2-KD	1	9586	125,17	4	3	3056	26,57	11

Length differences of alternatively excised fragment to the IES reference length in partial external (left) and partial internal (right) alternative excision events in the KDs.

Supplemental Table S4: Means of nucleosome density differences

IES group	IRS in ICOP1/2-KD	IES length	ICOP1_2	ISWI1	NOWA1_2	PtCAF1
total	all	all	-0.03 ± 0.78	-0.02 ± 1.17	0.03 ± 1.69	-0.04 ± 1.45
low	< 0.2	all	-0.07 ± 0.77	0.00 ± 1.07	0.07 ± 1.75	-0.12 ± 1.43
high	>= 0.2	all	0.15 ± 0.81	-0.09 ± 1.50	-0.14 ± 1.40	0.30 ± 1.48
short	all	< 200 bp	-0.04 ± 0.79	-0.01 ± 1.02	0.03 ± 1.72	-0.05 ± 1.44
long	all	>= 200 bp	0.29 ± 0.39	-0.09 ± 2.88	0.00 ± 0.92	0.17 ± 1.75
low_short	< 0.2	< 200 bp	-0.08 ± 0.78	0.01 ± 0.86	0.07 ± 1.77	-0.13 ± 1.40
low_long	< 0.2	>= 200 bp	0.24 ± 0.38	-0.15 ± 3.74	0.06 ± 1.04	0.09 ± 2.19
high_short	>= 0.2	< 200 bp	0.13 ± 0.84	-0.10 ± 1.57	-0.15 ± 1.46	0.31 ± 1.54
high_long	>= 0.2	>= 200 bp	0.35 ± 0.41	-0.01 ± 0.49	-0.09 ± 0.71	0.28 ± 0.78

Supplemental Table S5: Primers used for IES Retention PCR analysis

IES	Primer sequence (5' to 3' orientation)
MT Locus F	GGTGTTTATATCTTAATTGTTGACCCTCAC
MT Locus R	CCATCTATACTCCATTCTTTATCTTAATTCAT
51A2591F	ATGTGTTTGGACTGGATTGGCATGTAGAAG
51A2591R	GATGTAGCATAACATTTATCAACAATCCAT
51A6649F	ACTGCACCTCTAACTTTAACAAGCGAAGCA
51A6649R	CAGCAGTACATCCAGCTCTCTAAGTTTAGC

Supplemental Table S6: Reads used for adapter trimming

Read 1	AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
Read 2	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Supplemental Table S7: Changes to the lib/PARTIES/Map.pm file

old line	system("$bowtie2 --threads $self->{THREADS} --quiet --local -x $self->{BT2_INDEX} -1 $self->{FASTQ1} -2 $self->{FASTQ2} -X $self->{MAX_INSERT_SIZE} | $samtools view -uS - | $samtools sort - -o $out_bam > /dev/null 2>&1");
new line	system("$bowtie2 --threads $self->{THREADS} --quiet --local -x $self->{BT2_INDEX} -1 $self->{FASTQ1} -2 $self->{FASTQ2} -X $self->{MAX_INSERT_SIZE} | $samtools view - -uS | $samtools sort - -o notneeded.bam > $out_bam ");

Supplemental Table S8: Input Sequences for AlphaFold2 modeling

ISWI1 full length	MSNQSDDENEVLQVELASDEEQRAEEEDERIKKLEQDKKSFMSQIKSTGRMNTNIKFDNIESKINTLLENAEKYAMFLLH RHKRTQESKQKVQGQQRGKHRQIVEDGSEEEDFDDTPTVLEKQPTILKGGQLKSYQLTGLNWMISLFEEGINGILADEM GLGKTIQTIGFLAFLKEYKKISGPYLIVAPKSTLGNWMREFKIWMPCMRVVKLIAIKEERDEILNRYFQPGKFDVCLTSYEG VNICLKHIRRFQYKYIIIDEAHKIKNEDAIISQNLRKIRTNYKLLLTGTPLQNTPHELWSLLNYLLPDLFDSSEVFDKWFEVNT EAKLKEGNETIHQDELEQRNLEMVQKFQKILRPFMLRRTKAEVERMLPPKQEIHLFIKMSNLQKSMYQNILIHNNPHEGD DKGFYMNKLMQLRKICLHPYLFPEVEDKSLPALGEHLVDVSGKMRVLDKFLQKLSEGQHQILIFSQFTMMLNILEDYCNF RGYEYCRIDGETEIQSRDDQIAEFTAPDSKKFIFLLSTRAGGLGINLATADTVIIYDSDFNPQMDMQAMDRAHRIGQKSRV MVYRMACEHTVEEKIIERQQIKLRWDSLMVQQGRLQQKQNGKLLSKEDLKELTTYGASQIFKLDGDDIKDEDIDILLKRG EQLTKEMNERIEKKFENFKDKVQSLDLGLGQINIFDYFDEAKRNKEDEDALEDALVNHLMQDNKTRNRDKRAMMIGTNS KKIQGKQIKLSEHHLYENKDRLQYLLQKEEDFLAQQKTQKKANENDENVDFGGLTQDERQEQKRLLETGFKNWNKQEF QDFITANEKYGKDAYEKIQEVIKTKSQDEVKAYAQAFWERIDGLSEKDKIVKQIERGQKLIEQKTNGQKLIEEKCKHFHQP KYELVFTPQLYNKFKSKYFSLENDKFLIYMTNEVGYGNWAQLKQSIRKEPMFRFDHAFKCKSENELKNRVISLVKVLDKE KENNSMGRSLVKNTYIEKPKVLQESQKKKAKNDEEDVQDGSESVKKVKV
ISWI1 N-terminus	MSNQSDDENEVLQVELASDEEQRAEEEDERIKKLEQDKKSFMSQIKSTGRMNTNIKFDNIESKINTLLENAEKYAMFLLH RHKRTQESKQKVQGQQRGKHRQIVEDGSEEEDFDDTPTVLEKQPTILKGGQLKSYQLTGLNWMISLFEEGINGILADEM GLGKTIQTIGFLAFLKEYKKISGPYLIVAPKSTLGNWMREFKIWMPCMRVVKLIAIKEERDEILNRYFQPGKFDVCLTSYEG VNICLKHIRRFQYKYIIIDEAHKIKNEDAIISQNLRKIRTNYKLLLTGTPLQNTPHELWSLLNYLLPDLFDSSEVFDKWFEVNT EAKLKEGNETIHQDELEQRNLEMVQKFQKILRPFMLRRTKAEVERMLPPKQEIHLFIKMSNLQKSMYQNILIHNNPHEGD DKGFYMNKLMQLRKICLHPYLFPEVEDKSLPALGEHLVDVSGKMRVLDKFLQKLSEGQHQILIFSQFTMMLNILEDYCNF RGYEYCRIDGETEIQSRDDQIAEFTAPDSKKFIFLLSTRAGGLGINLATADTVIIYDSDFNPQMDMQAMDRAHRIGQKSRV MVYRMACEHTVEEKIIERQQIKLRWDSLMVQQGRLQ
ISWI1 C-terminus	QKQNGKLLSKEDLKELTTYGASQIFKLDGDDIKDEDIDILLKRGEQLTKEMNERIEKKFENFKDKVQSLDLGLGQINIFDYF DEAKRNKEDEDALEDALVNHLMQDNKTRNRDKRAMMIGTNSKKIQGKQIKLSEHHLYENKDRLQYLLQKEEDFLAQQK TQKKANENDENVDFGGLTQDERQEQKRLLETGFKNWNKQEFQDFITANEKYGKDAYEKIQEVIKTKSQDEVKAYAQAF WERIDGLSEKDKIVKQIERGQKLIEQKTNGQKLIEEKCKHFHQPKYELVFTPQLYNKFKSKYFSLENDKFLIYMTNEVGYG NWAQLKQSIRKEPMFRFDHAFKCKSENELKNRVISLVKVLDKEKENNSMGRSLVKNTYIEKPKVLQESQKKKAKNDEED VQDGSESVKKVKV

ICOP1	MDNKENEKQAKLKEFQRRFPNYMNGKKVIFPIMDEIIVQFSQIFPQSNQFGKMREGNVIKTFSPIPLNQQIEITTFINAFPY NQQFAAMSDSQNQFQLLLSPLLKLTGHLIKDYHTNATFSNSYQYSTFDTQEQSDRINFKLAAFIVEDIFKYKMGLLKPIDIA KQRQVKDEIKKKGNKPQKMSLLSIVEKQQEEEKDLNQDLTIQSQKYEFENYLDPTCQFFWKIAYESFNKIIKKSQRTKLQ DMDIEQDSDEQPETIQNVNKETLSQDNIEMRRQQIISKYQQLCKLKDQSKKKSKQTKQYSQIKSFKIKDRYTDLEMLRFN NLFKFLVQNWPSFLLQSIKLPYVQSLFTDQELRNIKSIGQNELGYFGLKAKQRADITSNLIEGVRETDSFKVIIEYRQEVTE AVGLVLENVQDELNSINQSLQKKDSQYTQQQQQQYRKVYKQYLQLVEFNRLFINGCLYLGSDIHGFDYHIFSNDIDHIYQ NNGSEWRVLDENQVQQLFKTLNVCGVKERELQTNIQKLMACELFNDQETKELITIKNVEQSQVQAGNRSPKQLIVKILLE VVQKYTDILMVRKLRWESYKIREKFQNTIKTLENPLDMVDFMKILIEQFETAQVLIIDQQKMQNGSQYDQRDLKEFQQRL RIYENKLKGLKEPEKILFYDTQLFQMMESREHIKPNGVKCNTKFWQQSLGLEVKEALMNFANKVDKEHQQYDVVFMAS TLLLAVQEYELSSSAQDNEDDELLRQIVKDVKVDNSNFPSKIDNHQKNNQIIELD
ICOP2	MDNKENEKQAKLKEFQRRFPNYMNGKKVIFPIMDEIIVQFSQIFPQSNQFGKMREGNVIKTFSPIPLNQQIEITTFINAFPY NQQFAAMSDSQNQFQLLLSPLLKLTGHLIKDYHSNATFSNSYQYSTFDTQEQSDRINFKLTAFIIEDIFKYKMGLLKPIDIA KQRQVKDDIKKKANKAQKISLLTMVEKQQQQEEEKDVNQDLTLQSQKYEFENYLDPTCQSFWKIAYEAFNKIIKKSQRTK LQDMDIEQESDEQPDTIQNVNNQKGNLNQESIELRRQQIISKYQQLGKLKDQNKKKSKSTKQYTQIKSFKIKDKYTDLEM LRFNNLFKFLIQNWPAMLLQSIKLPYVQSLFNDQELRNIKSIGQNDLGYFGLKAKQRADITSNLIEGVRETDSFKIIIEYRQE VTEAVGLVLENVQDELNSINQALQKKDSQYTLQQQQQYRKIYKQYLQLVEFSRLFINGCLYLGSDIHGYDYHIFSNDIDHI YQNNGSEWRVLDENQVQQLFKTLNVCGVKERELQTNIQKLMACELFNDQDTKELITIKNVEQSQVQAGNRSPKQLIVKI LLEVVQKYTDTLMIRKLRWESYKIREKFQNTIKTLENPLDMVDFMKILIEQFETAQVLVIDQQKMQNGNQYDQKDLKEFQ QRIRIYENKLQGLKEPEKILFYDSQLFQLMESKEYIKPNGIKSNTKFWQQSLGVEVKEALMNFANKVDKENQQYDVVFMA STLLLAVQEYELSSSSQDNEDDELLRQIVKEVKFDNQNLLLKTDNHQVNNQIIELD
Ptiwi01	MFQNIQLKANFHEMRLNPSRPVYQYKLEITDSSPEKVSEALKKFRPQLQTQLILFMSLNQNIYSPKLIQEADNGLVLGSLS GNETNQDTATLKLVGKIENKADLNIIISRLFKQVIRSQMQMVSVGNKGQKLFWSSRAQQFKDQNLEIWPGVECIFRPGEG GAQNPTLVIDCAFKMLRYRSALEELNQTRNPACIQDQIVMTTYNKKFYKVEAVDVNLKPASTFTNEKGETISFAQYYEQR YKVKVDGNQPLIRATVRSKQDKTEKTIHLIPQLCQLTGLTDAIRNDFNAMKNLAVVTKPGADQRMKMAQEFANQLASTEI VNKKLGTKRQIFKEWGVEINPGSMDVPARRIHPGNMLMGNGLKLDLSSPQTNLDRQTQTQMFSTPPQQLILGIIYNKKT GQQTMDSLMQNFQAACNDFKFQAFMAPKVFPIEQDRDEDLERVLDGFQKQAEANKAKVGFLLFLLPGQKKKARLYKTA KKISMQKFGCASQVVVEKTLAKNTRSIVNKILIQLNAKVGGTPWAIDSLPTTFQNQPTMICGTDCFVKSGRKNQLAFCST VDRNLSRYYSQVVTSGEFSQHLQQVFKASLLAFKEQNGIFPKLIIIYRDGVGDGQQAVVLANELPQYKQALEELQITDTKI SLVVCNKRVSAKFYTGGNARPDNPQPGTCVDNPKVVEQSNPNFYLISQVTRQGTVTPSLYKIIHSDQAGLDDDIKVLTFK LCWLFYNFTGSIKIPAPVRYAHCLCNFIGDNYDDRDQVKFLPLPDLVKQKVLFYI

Supplemental Table S9: Codon-optimized sequences for recombinant protein expression

ISWI1	MSNQSDDENEVLQVELASDEEQRAEEEDERIKKLEQDKKSFMSQIKSTGRMNTNIKFDNIESKINTLLENAEKYAMFLLH RHKRTQESKQKVQGQQRGKHRQIVEDGSEEEDFDDTPTVLEKQPTILKGGQLKSYQLTGLNWMISLFEEGINGILADEM GLGKTIQTIGFLAFLKEYKKISGPYLIVAPKSTLGNWMREFKIWMPCMRVVKLIAIKEERDEILNRYFQPGKFDVCLTSYEG VNICLKHIRRFQYKYIIIDEAHKIKNEDAIISQNLRKIRTNYKLLLTGTPLQNTPHELWSLLNYLLPDLFDSSEVFDKWFEVNT EAKLKEGNETIHQDELEQRNLEMVQKFQKILRPFMLRRTKAEVERMLPPKQEIHLFIKMSNLQKSMYQNILIHNNPHEGD DKGFYMNKLMQLRKICLHPYLFPEVEDKSLPALGEHLVDVSGKMRVLDKFLQKLSEGQHQILIFSQFTMMLNILEDYCNF RGYEYCRIDGETEIQSRDDQIAEFTAPDSKKFIFLLSTRAGGLGINLATADTVIIYDSDFNPQMDMQAMDRAHRIGQKSRV MVYRMACEHTVEEKIIERQQIKLRWDSLMVQQGRLQQKQNGKLLSKEDLKELTTYGASQIFKLDGDDIKDEDIDILLKRGE QLTKEMNERIEKKFENFKDKVQSLDLGLGQINIFDYFDEAKRNKEDEDALEDALVNHLMQDNKTRNRDKRAMMIGTNSK KIQGKQIKLSEHHLYENKDRLQYLLQKEEDFLAQQKTQKKANENDENVDFGGLTQDERQEQKRLLETGFKNWNKQEFQ DFITANEKYGKDAYEKIQEVIKTKSQDEVKAYAQAFWERIDGLSEKDKIVKQIERGQKLIEQKTNGQKLIEEKCKHFHQPK YELVFTPQLYNKFKSKYFSLENDKFLIYMTNEVGYGNWAQLKQSIRKEPMFRFDHAFKCKSENELKNRVISLVKVLDKEK ENNSMGRSLVKNTYIEKPKVLQESQKKKAKNDEEDVQDGSESVKKVKV
ICOP1	MDNKENEKQAKLKEFQRRFPNYMNGKKVIFPIMDEIIVQFSQIFPQSNQFGKMREGNVIKTFSPIPLNQQIEITTFINAFPY NQQFAAMSDSQNQFQLLLSPLLKLTGHLIKDYHTNATFSNSYQYSTFDTQEQSDRINFKLAAFIVEDIFKYKMGLLKPIDIA KQRQVKDEIKKKGNKPQKMSLLSIVEKQQEEEKDLNQDLTIQSQKYEFENYLDPTCQFFWKIAYESFNKIIKKSQRTKLQD MDIEQDSDEQPETIQNVNKETLSQDNIEMRRQQIISKYQQLCKLKDQSKKKSKQTKQYSQIKSFKIKDRYTDLEMLRFNNL FKFLVQNWPSFLLQSIKLPYVQSLFTDQELRNIKSIGQNELGYFGLKAKQRADITSNLIEGVRETDSFKVIIEYRQEVTEAV GLVLENVQDELNSINQSLQKKDSQYTQQQQQQYRKVYKQYLQLVEFNRLFINGCLYLGSDIHGFDYHIFSNDIDHIYQNN GSEWRVLDENQVQQLFKTLNVCGVKERELQTNIQKLMACELFNDQETKELITIKNVEQSQVQAGNRSPKQLIVKILLEVV QKYTDILMVRKLRWESYKIREKFQNTIKTLENPLDMVDFMKILIEQFETAQVLIIDQQKMQNGSQYDQRDLKEFQQRLRIY ENKLKGLKEPEKILFYDTQLFQMMESREHIKPNGVKCNTKFWQQSLGLEVKEALMNFANKVDKEHQQYDVVFMASTLLL AVQEYELSSSAQDNEDDELLRQIVKDVKVDNSNFPSKIDNHQKNNQIIELD

ICOP2	MDNKENEKQAKLKEFQRRFPNYMNGKKVIFPIMDEIIVQFSQIFPQSNQFGKMREGNVIKTFSPIPLNQQIEITTFINAFPY NQQFAAMSDSQNQFQLLLSPLLKLTGHLIKDYHSNATFSNSYQYSTFDTQEQSDRINFKLTAFIIEDIFKYKMGLLKPIDIAK QRQVKDDIKKKANKAQKISLLTMVEKQQQQEEEKDVNQDLTLQSQKYEFENYLDPTCQSFWKIAYEAFNKIIKKSQRTKL QDMDIEQESDEQPDTIQNVNNQKGNLNQESIELRRQQIISKYQQLGKLKDQNKKKSKSTKQYTQIKSFKIKDKYTDLEML RFNNLFKFLIQNWPAMLLQSIKLPYVQSLFNDQELRNIKSIGQNDLGYFGLKAKQRADITSNLIEGVRETDSFKIIIEYRQEV TEAVGLVLENVQDELNSINQALQKKDSQYTLQQQQQYRKIYKQYLQLVEFSRLFINGCLYLGSDIHGYDYHIFSNDIDHIY QNNGSEWRVLDENQVQQLFKTLNVCGVKERELQTNIQKLMACELFNDQDTKELITIKNVEQSQVQAGNRSPKQLIVKILL EVVQKYTDTLMIRKLRWESYKIREKFQNTIKTLENPLDMVDFMKILIEQFETAQVLVIDQQKMQNGNQYDQKDLKEFQQRI RIYENKLQGLKEPEKILFYDSQLFQLMESKEYIKPNGIKSNTKFWQQSLGVEVKEALMNFANKVDKENQQYDVVFMASTL LLAVQEYELSSSSQDNEDDELLRQIVKEVKFDNQNLLLKTDNHQVNNQIIELD

Supplemental Table S10: Samples used for nucleosome density analyses

Sample	Reference
ICOP1/2/PGM-KD	This study
ND7/PGM-KD (control for this study)	This study
ISWI1/PGM-KD	(Singh et al. 2022)
NOWA1/2/PGM-KD	(Singh et al. 2022)
PTCAF1/PGM-KD	(Wang et al. 2022)
empty vector/PGM-KD (control for *PTCAF1/PGM*-KD & *NOWA1/2/PGM*-KD)	(Singh et al. 2022; Wang et al. 2022)

ISWI1 complex proteins facilitate developmental genome editing in *Paramecium*

Aditi Singh[1,*,#], Lilia Häußermann[1,*], Christiane Emmerich[1], Emily Nischwitz[2], Brandon KB Seah[1], Falk Butter[2,3], Mariusz Nowacki[4], Estienne C. Swart[1,#]

Supplemental Methods

Culture Cultivation

Mating type 7 cells (strain 51) of *Paramecium tetraurelia* were grown according to the standard protocol (Beisson et al. 2010b, 2010a). *E. coli* strain HT115 was used for feeding, and the cultures were maintained either at 27 °C or at 18 °C.

RNAi assays

ICOP1 and ICOP2 RNAi constructs were made by cloning a 538 bp (2708-3246) and a 1089 bp gene fragment (3349-4527), respectively, into the L4440 plasmid. The plasmids were transformed into *E. coli* HT1115 (DE3). Knockdown experiments were performed as previously (Beisson et al. 2010c). Isopropyl ß-D-1-thiogalactopyranoside (IPTG) induction was done at 30 °C. Cells were collected at early (~40% of cells with fragmented parental MAC) and late (~90% of cells with visible new MAC) developmental time points. After three days, 30 post-autogamous cells were fed with a non-induced feeding medium to assay survival. Genomic DNA was extracted from post-autogamous cultures using the standard kit protocol (G1N350, Sigma-Aldrich). PCRs were done on different genomic regions flanking an IES (Supplemental Table S5) to test IES retention.

HHpred identification of protein domains

ICOP1 (XM_001447768.1, PTET.51.1.P0440186) and ICOP2 (XM_001437312.1, XM_001437313.1, PTET.51.1.P0180124) protein sequences were analyzed using HHpred from the MPI bioinformatics toolkit (https://toolkit.tuebingen.mpg.de/#/) across the COG (Tatusov et al. 2000), Pfam (Finn et al. 2003, 2016), NCBI Conserved Domain Database (CDD) (Marchler-Bauer et al. 2017), and ECOD (Cheng et al. 2014, 2015) databases. PSI-BLAST (Bhagwat and Aravind 2007) with 4 iterations was used to identify further proteins with WSD domains, and multiple alignments were done using MAFFT (Katoh and Standley 2013; Kuraku et al. 2013) provided as a plugin within Geneious (version 2022.1.1, http://www.geneious.com)

(Kearse et al. 2012). InterProScan (Paysan-Lafosse et al. 2023) within Geneious was used to identify domains within MAFFT-aligned protein sequences. For Fig. 1C Clustal Omega (1.2.3) from Geneious was used (default parameters). Proteins in the alignment are ICOP1 (this study), ICOP2 (this study), IOC3 (Yamada et al. 2011), ACF (Ito et al. 1997), BAZ1B (Bozhenok et al. 2002), DDW1 (Yamada et al. 2011; Tan et al. 2020).

Phylogenetic analysis of ICOP1 and ICOP2

Trimal-auto (Capella-Gutiérrez et al. 2009) was used to select well-aligned columns from the MAFFT-aligned protein sequences of PSI-BLAST identified (see HHpred of protein domains) proteins. PHYML version 2.2.4 (Guindon and Gascuel 2003) provided as a plugin in Geneious (version 2022.1.1, http://www.geneious.com) was used to generate a maximum likelihood phylogeny with 100 bootstrap replicates. FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) was used to inspect, manipulate and generate the graphical tree.

Fluorescence intensity quantification

The signal of HA-ICOP1 or ICOP2-HA in the new MACs was quantified in *ND7*-KD, *ISWI1*-KD and *PTIWI01/09*-KD cells. Background levels were acquired in wild-type cells. For 20 cells per sample, nicely visible new MACs were chosen in the DAPI channel, and DAPI and anti-HA signal was captured in one z-plane through at least one of the new MACs. Constant laser settings were used across all samples. In the subsequent analysis, HA-ICOP1 and ICOP2-HA injected cells were treated

separately. Brightness and contrast settings were kept constant for anti-HA signals. A constant area was positioned in the DAPI channel in the new MAC and the mean fluorescence intensity was measured in the anti-HA channel. After subtracting the mean of all wild-type cells from all measurements, the values were normalized by the mean of all *ND7*-KD cells. This quantification does not consider the high variability of background levels within cells. Stronger staining might lead to stronger signals in the new MACs; however, background levels within cells were not acquired due to the non-homogeneous distribution and danger of non-representative results. Boxplots were generated with matplotlib and seaborn packages. Data was analyzed with ANOVA and a Tukey post-hoc test.

Mass spectrometry analysis

Samples were separated on a 4%–12% NOVEX NuPage gradient gel (Thermo) for 10 minutes at 180 V in 1 X MES buffer (Thermo). Proteins were fixed and stained with Coomassie G250 brilliant blue (Carl Roth). The gel lanes were cut, and each lane was minced into approximately 1x1 mm pieces. Gel pieces were destained with a 50% ethanol/50 mM ammonium bicarbonate (ABC) solution. Proteins were reduced in 10 mM DTT (Sigma-Aldrich) for 1 hour at 56 °C and then alkylated with 50 mM iodoacetamide (Sigma-Aldrich) for 45 min at room temperature. Proteins were digested with 1 µg mass spectrometry grade trypsin (Sigma) overnight at 37 °C. Peptides were extracted from the gel by two incubations with 30% ABC/acetonitrile and three subsequent incubations with pure acetonitrile. The acetonitrile was subsequently evaporated in a concentrator (Eppendorf) and loaded on StageTips (Rappsilber et al. 2007) for desalting and storage.

Peptides were eluted from the StageTips using 80% acetonitrile / 0.1% formic acid and concentrated before loading on an uHPLC nLC-1200 system coupled to an Exploris 480 mass spectrometer (Thermo). The peptides were loaded on a 50 cm (Exploris 480) column (75 µm inner diameter) in-house packed with Reprosil C18 (Dr. Maisch GmbH) and eluted with a 73 or 88 min optimized gradient increasing from 3% to 40% mixture of 80% acetonitrile/0.1% formic acid at a flow rate of 225 nl/min or 250 nl/min. The Exploris 480 was operated in positive ion mode with a data-dependent acquisition strategy of one MS full scan (scan range 300 - 1,650 m/z; 60,000 resolution; normalized AGC target 300%; max IT 28 ms) and up to twenty MS/MS scans (15,000 resolution; AGC target 100%, max IT 40 ms; isolation window 1.4 m/z) with peptide match preferred using HCD fragmentation.

MS raw data were searched using the Andromeda search engine (Cox et al. 2011) integrated into MaxQuant suite 1.6.5.0 (Cox and Mann 2008) using the *Paramecium* predicted proteins as the database (ptetraurelia_mac_51_annotation_v2.0). In all analyses, carbamidomethylation at cysteine was set as a fixed modification, while methionine oxidation and protein N-acetylation were considered as variable modifications. The match between run option was activated. Prior to bioinformatics analysis, reverse hits, proteins only identified by site, protein groups based on one unique peptide, and known contaminants were removed.

For further bioinformatics analysis, the label-free quantitation (LFQ) values were $\log_2$ transformed, and the median across the replicates was calculated. This enrichment was plotted against the $-\log_{10}$ transformed p-value (Welch t-test) using the ggplot2 package in the R environment.

Plasmids and vectors for recombinant protein expression assay

DNA sequences coding for *Paramecium* proteins ISWI1, ICOP1, and ICOP2 were codon-optimized (Supplemental Table S9) for expression in *E. coli* using the GENEius tool of Eurofins (Luxembourg). Gene synthesis was performed at Eurofins Genomics Germany GmbH (Ebersberg, Germany). The synthetic constructs were cloned into pET-MCN vectors (Romier et al. 2006), expressing proteins with either no tag, a hexahistidine (His), or a GST tag. Plasmids were co-transformed in different combinations into *E. coli* strain Gold pLysS.

Total RNA extraction and mRNA sequencing.

Approximately 1.2×10^6 cells were collected from the early (approx. 40% of cells with a fragmented MAC) and the late (majority of cells with visible anlagen) developmental stages using an oil centrifuge at 280 g for 2 minutes. The cells were washed twice in 10 mM Tris-HCl (pH 7.4), flash frozen using liquid nitrogen dropped gently over the pellets, and stored at -80°C. Total RNA extraction was performed by adding 6 ml of Tri reagent (Sigma-Aldrich, T9424) per sample and following the standard protocol provided with the reagent. mRNA and sRNA libraries were prepared and sequenced at Genewiz International (Leipzig, Germany).

Knockdown efficiency validation using RNA-seq

Total RNA was sequenced by Genewiz (Germany, GmbH) using poly-A enrichment with NovaSeq 2×150 bp reads. Illumina adapter sequences were trimmed from reads with TrimGalore (Krueger 2019) (Supplemental Table S6). Reads were mapped to the *Paramecium tetraurelia* strain 51 transcriptome. Mapping showed high coverage on the silencing regions, most likely caused by RNAs of the siRNA silencing pathway. For each knockdown, the target gene was replaced by three split transcripts (the silencing region, the 5' upstream non-silencing region and the 3' downstream non-silencing region), and only the 5' upstream region was considered for analysis. FPKM (fragments per kilobase transcript per million mapped reads) values were calculated using eXpress (Roberts and Pachter 2013) (SourceData_Fig4; (Singh 2023)) and rounded by the standard Python method to integers. Scripts are available from https://github.com/Swart-lab/ICOP_code/tree/main/KD-efficiency.

Macronuclear isolation and Illumina DNA-sequencing

Samples for MAC isolation were collected from *ICOP1*-K*D, ICOP2*-KD*,* and *ICOP1/2*-KD cultures three days post autogamy as described previously (Arnaiz et al. 2012). *ICOP1/2/PGM*-KD and *ND7/PGM*-KD cultures were collected ca. 12 h after developing MACs were visible in most of the cells. DNA libraries were prepared using the FS DNA Library Prep kit (E7805, NEB). Paired-end (2×150 bp) sequencing was done on NextSeq 2000 at MPI for Biology, Tübingen.

Reference genomes and predicted genes

Reference genome to analyze DNA-seq data:

MAC: https://paramecium.i2bc.paris-saclay.fr/files/Paramecium/tetraurelia/51/sequences/ptetraurelia_mac_51.fa

MAC+IES: https://paramecium.i2bc.paris-saclay.fr/files/Paramecium/tetraurelia/51/sequences/ptetraurelia_mac_51_with_ies.fa

Reference CDS + UTR sequences used to analyze mRNA-seq data:

https://paramecium.i2bc.paris-saclay.fr/files/Paramecium/tetraurelia/51/annotations/ptetraurelia_mac_51/ptetraurelia_mac_51_annotation_v2.0.transcript.fa

IES retention and alternative boundary analysis

Illumina adapter sequences were trimmed from whole genome sequencing (WGS) reads of enriched new MAC DNA after knockdown using TrimGalore (Krueger 2019) (Supplemental Table S6). ParTIES (Denby Wilkes et al. 2016) v1.05 was used to map reads to MAC and MAC+IES genomes and calculate IRSs.

To accommodate changes in a newer version of samtools (Li et al. 2009), the /lib/PARTIES/Map.pm file was changed (Supplemental Table S7). IRSs are provided in SourceData_Fig4 (Singh 2023) as ICOP_IRS.tab.gz. IRS correlations using IRSs from published knockdown data ((*ISWI1*-KD (Singh et al. 2022), *PGM*-KD (Arnaiz et al. 2012), *TFIIS4*-KD (Maliszewska-Olejniczak et al. 2015) and *PTIWI01/09*-KD

(Furrer et al. 2017)) were calculated with After_ParTIES (option --use_pearson (https://github.com/gh-ecs/After_ParTIES)).

Since alternative excision analysis depends on IES coverage, to ensure a fair comparison, libraries were adjusted to similar sizes by downsampling. Downsampling factors relative to the smallest library used were calculated according to the number of properly paired and mapped reads to the MAC+IES reference genome (ND7 = 0.686; ICOP1 = 0.512; ICOP2 = 0.453; ISWI1 = 0.698; ICOP1_2 = 1.0). The “MILORD” module of a ParTIES pre-release version (13 August 2015) was used to annotate alternative and cryptic IES excision (SourceData_Fig5; (Singh 2023)).

All scripts are available from https://github.com/Swart-lab/ICOP_code/tree/main/Alternative_excision.

Nucleosomal DNA Isolation and Illumina DNA-sequencing

Nucleosomal DNA was isolated with the EZ Nucleosomal DNA Prep Kit (D5220, Zymo Research) as previously, utilizing a sucrose cushion to isolate nuclei once new MACs are visible (Singh et al. 2022), except that digested DNA was size-selected with SPRIselect magnetic beads (Beckman Coulter) to enrich for mono- and di-nucleosomal fragments (0.7× volume right-side size selection). Libraries were prepared with NEBNext Ultra II DNA library prep kit (E7645S, NEB), size-selected for 150 bp insert. 2×100 bp paired-end sequencing was performed on an Illumina NextSeq 2000 instrument with P3 chemistry at MPI for Biology, Tübingen.

Nucleosome Density Analysis

Illumina adapter sequences were trimmed from reads with TrimGalore (Krueger 2019) (Supplemental Table S6).

Reads were mapped to the MAC+IES genome, then properly paired and mapped reads overlapping IESs were extracted and counted. IRS distributions are provided in Supplemental Fig. 7A. DNase reads were size selected (100 - 175 bp outer distance). Library sizes to calculate downsampling factors were obtained by the "samtools stats" command on the .sorted.bam files. The length distribution of outer distances of PE reads mapping to scaffold51_9 was plotted (Supplemental Fig. 7B).

Samples used for nucleosome density analysis are in Supplemental Table S10. Nucleosome density differences (re_rc) were calculated for each IES by subtracting the nucleosome density of the control (r_c) from the experimental sample (r_e).

re_rc = r_e - r_c

IES with infinite ("inf") or not available "nan" values were excluded, resulting in 43,409 (in *NOWA1/2/PGM*-KD) and 44,448 (in *ICOP1/2/PGM*-KD) IESs used for analysis. Kolmogorov-Smirnov (KS) statistics and associated p-values for two sample tests were calculated to assess distribution differences.

All scripts are available from https://github.com/Swart-lab/ICOP_code/tree/main/Nucleosome_density.

Read counts for IESs are available in SourceData_Fig7 (Singh 2023).

sRNA analysis

sRNA-seq reads were mapped to the *Paramecium tetraurelia* strain 51 MAC + IES genome and L4440 silencing vector with bwa version 0.7.17-r1188 (Li and Durbin 2009). 10-49 bp long, uniquely mapped reads (possessing the flags “XT:A:U”) were selected by grep in a shell script. sRNA length histograms were generated by a Python script. Shell scripts for the RNA mapping, post-processing, and histogram are available from https://github.com/Swart-lab/ICOP_code/tree/main/sRNA_analysis.

Supplemental References

Arnaiz O, Mathy N, Baudry C, Malinsky S, Aury J-M, Denby Wilkes C, Garnier O, Labadie K, Lauderdale BE, Le Mouël A, et al. 2012. The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet* **8**: e1002984.

Beisson J, Bétermier M, Bré M-H, Cohen J, Duharcourt S, Duret L, Kung C, Malinsky S, Meyer E, Preer JR, et al. 2010a. Maintaining clonal *Paramecium tetraurelia* cell lines of controlled age through daily reisolation. *Cold Spring Harb Protoc* **2010**: pdb.prot5361.

Beisson J, Bétermier M, Bré M-H, Cohen J, Duharcourt S, Duret L, Kung C, Malinsky S, Meyer E, Preer JR, et al. 2010b. Mass culture of *Paramecium tetraurelia*. *Cold Spring Harb Protoc* **2010**: pdb.prot5362.

Beisson J, Bétermier M, Bré M-H, Cohen J, Duharcourt S, Duret L, Kung C, Malinsky S, Meyer E, Preer JR, et al. 2010c. Silencing specific Paramecium tetraurelia genes by feeding double-stranded RNA. *Cold Spring Harb Protoc* **2010**: pdb.prot5363.

Bhagwat M, Aravind L. 2007. PSI-BLAST tutorial. *Methods Mol Biol* **395**: 177–186.

Bozhenok L, Wade PA, Varga-Weisz P. 2002. WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. *EMBO J* **21**: 2231–2241.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Cheng H, Liao Y, Schaeffer RD, Grishin NV. 2015. Manual classification strategies in the ECOD database. *Proteins* **83**: 1238–1251.

Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, Shi S, Kim B-H, Grishin NV. 2014. ECOD: an evolutionary classification of protein domains. *PLoS Comput Biol* **10**: e1003926.

Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*

26: 1367–1372.

Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**: 1794–1805.

Denby Wilkes C, Arnaiz O, Sperling L. 2016. ParTIES: a toolbox for *Paramecium* interspersed DNA elimination studies. *Bioinformatics* **32**: 599–601.

Finn R, Griffiths-Jones S, Bateman A. 2003. Identifying protein domains with the Pfam database. *Curr Protoc Bioinformatics* **Chapter 2**: Unit 2.5.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**: D279-85.

Furrer DI, Swart EC, Kraft MF, Sandoval PY, Nowacki M. 2017. Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA. *Cell Rep* **20**: 505–520.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.

Ito T, Bulger M, Pazin MJ, Kobayashi R, Kadonaga JT. 1997. ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* **90**: 145–155.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Krueger F. 2019. *TrimGalore: A wrapper around Cutadapt and FastQC to consistently apply*

adapter and quality trimming to FastQ files, with extra functionality for RRBS data. GitHub.

Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res* **41**: W22-8.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Maliszewska-Olejniczak K, Gruchota J, Gromadka R, Denby Wilkes C, Arnaiz O, Mathy N, Duharcourt S, Bétermier M, Nowak JK. 2015. TFIIS-Dependent Non-coding Transcription Regulates Developmental Genome Rearrangements. *PLoS Genet* **11**: e1005383.

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* **45**: D200–D203.

Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, et al. 2023. InterPro in 2022. *Nucleic Acids Res* **51**: D418–D427.

Rappsilber J, Mann M, Ishihama Y. 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**: 1896–1906.

Roberts A, Pachter L. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods* **10**: 71–73.

Romier C, Ben Jelloul M, Albeck S, Buchwald G, Busso D, Celie PHN, Christodoulou E, De

Marco V, van Gerwen S, Knipscheer P, et al. 2006. Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. *Acta Crystallogr D Biol Crystallogr* **62**: 1232–1242.

Singh A, Maurer-Alcalá XX, Solberg T, Häußermann L, Gisler S, Ignarski M, Swart EC, Nowacki M. 2022. Chromatin remodeling is required for sRNA-guided DNA elimination in Paramecium. *EMBO J* **41**: e111839.

Singh A. 2023. ISWI complex proteins facilitate genome editing and development . *Edmond*. https://doi.org/10.17617/3.ZBOLU8 (Accessed July 28, 2023).

Tan L-M, Liu R, Gu B-W, Zhang C-J, Luo J, Guo J, Wang Y, Chen L, Du X, Li S, et al. 2020. Dual recognition of h3k4me3 and DNA by the ISWI component ARID5 regulates the floral transition in arabidopsis. *Plant Cell* **32**: 2178–2195.

Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**: 33–36.

Wang C, Solberg T, Maurer-Alcalá XX, Swart EC, Gao F, Nowacki M. 2022. A small RNA-guided PRC2 complex eliminates DNA as an extreme form of transposon silencing. *Cell Rep* **40**: 111263.

Yamada K, Frouws TD, Angst B, Fitzgerald DJ, DeLuca C, Schimmele K, Sargent DF, Richmond TJ. 2011. Structure and mechanism of the chromatin remodelling factor ISW1a. *Nature* **472**: 448–453.

Two paralogous PHD finger proteins participate in *Paramecium tetraurelia*'s natural genome editing

Lilia Häußermann[1], Aditi Singh[1,#], Estienne C. Swart[1,#]

[1]Max Planck Institute for Biology, Max-Planck-Ring 5, 72076, Tübingen, Germany

[#]Corresponding author: estienne.swart@tuebingen.mpg.de,

aditi.singh@tuebingen.mpg.de

Keywords: genome reorganization; PHD finger proteins; small RNAs

Abstract

The unicellular eukaryote *Paramecium tetraurelia* contains functionally distinct nuclei: germline micronuclei (MICs) and a somatic macronucleus (MAC). During sexual reproduction, the MIC genome is reorganized into a new MAC genome and the old MAC is lost. Almost 45,000 unique Internal Eliminated Sequences (IESs) distributed throughout the genome require precise excision to guarantee a functional new MAC genome. Here, we characterize a pair of paralogous PHD finger proteins involved in DNA elimination. DevPF1, the early-expressed paralog, is present in only some of the gametic and post-zygotic nuclei during meiosis. Both DevPF1 and DevPF2 localize in the new developing MACs, where IESs excision occurs. In *DevPF2* knockdown (KD) long IESs are preferentially retained and late-expressed small RNAs decrease; no length preference for retained IESs was observed in *DevPF1*-KD and development-specific small RNAs were abolished. The expression of at least two genes from the new MAC with roles in genome reorganization seems to be influenced by *DevPF1*- and *DevPF2*-KD. Thus, both PHD fingers are crucial for new MAC genome development, with distinct functions, potentially via regulation of non-coding and coding transcription in the MICs and new MACs.

Introduction

A unique feature shared by all ciliates is the presence of nuclear dimorphism. In *Paramecium tetraurelia* (henceforth *Paramecium*) the two micronuclei (MICs) resemble the germline of multicellular organisms, being transcriptionally silent throughout most of the life cycle and generating haploid nuclei during meiosis that develop and give rise to all nuclei in the subsequent generation. Also similar to the multicellular soma, the macronucleus (MAC) is optimized for most gene expression, and originates from a MIC copy. The old MAC is fragmented during sexual division and subsequently diluted across cell divisions, with the new MAC completely taking over somatic expression. The development from the MIC genome to the MAC genome in *Paramecium* is a natural

form of genome editing that requires extensive reorganization, including genome amplification (~800n), chromosome fragmentation and the elimination of about 25% of the sequence content (Arnaiz et al., 2012; Guérin et al., 2017). These MIC genome-specific sequences comprise repeats, transposable elements and Internal Eliminated Sequences (IESs).

In contrast to other elimination events, IES elimination requires precise excision in *Paramecium*. Precise IES excision is not characteristic of all ciliates. Notably, in *Paramecium*'s oligohymenophorean relative *Tetrahymena*, IESs are predominantly imprecisely excised and only tolerated in intergenic regions (Hamilton et al., 2016). The roughly 45,000 IESs in *Paramecium* are scattered throughout the genome in both non-coding and coding regions and vary from tens to thousands of base pairs in length (Arnaiz et al., 2012). Since the coding density of the *Paramecium* MAC genome is high, most IESs are intragenic (Arnaiz et al., 2012). *Paramecium* IESs are flanked by conserved 5'-TA-3' dinucleotides (Klobutcher & Herrick, 1995) and excised by PiggyMAC (Pgm). Pgm is a domesticated transposase derived from PiggyBac transposases (Baudry et al., 2009) like the excisase responsible for IES excision in *Tetrahymena* (Cheng et al., 2010). The weakly conserved ~5 bp long inverted repeats at *Paramecium* IES ends (Klobutcher & Herrick, 1995) fail to provide enough specificity for reliable Pgm recruitment (Arnaiz et al., 2012). This suggests that other factors are needed for precise IES targeting.

The targeting of MIC-specific sequences for elimination is thought to be assisted by small non-coding RNAs, first characterized in *Tetrahymena* (Chalker & Yao, 2001; Mochizuki et al., 2002). Like *Tetrahymena*, the biogenesis of the 25 nucleotide (nt) scan RNAs (scnRNAs) occurs during meiosis in the *Paramecium* MICs. Bidirectional non-coding transcription of the MIC genome is thought to be initiated by the putative transcription elongation factor Spt5m (Gruchota et al., 2017) and followed by the cleavage of long double-stranded RNA (dsRNA) by the closely related Dicer-like protein paralogs Dcl2 and Dcl3 (Hoehener et al., 2018; Lepère et al., 2009; Sandoval et al., 2014). Argonaute/Piwi proteins Ptiwi01/09 (also close paralogs) process the resulting

short dsRNAs, removing one of the two strands, and stabilize single-stranded scnRNAs throughout the selection process in the parental MAC and targeting of MIC-specific sequences in the new MACs (Bouhouche et al., 2011; Furrer et al., 2017). In the parental MAC, Gtsf1 was recently proposed to promote ubiquitination and subsequent degradation of the Ptiwi01/09 complexes harboring MAC-matching scnRNAs (Charmant et al., 2023; Wang et al., 2023). In the new MACs, the putative transcription elongation factor TFIIS4 was proposed to promote non-coding transcription required for scanning the developing genome (Maliszewska-Olejniczak et al., 2015).

In *Tetrahymena*, H3K9 and H3K27 methylation precede IES excision (Y. Liu et al., 2007; Taverna et al., 2002) and it was shown in *Paramecium* that development-specific H3K9me3 and H3K27me3 histone mark deposition by the PRC2 complex depends on scnRNAs and is essential for the elimination of transposons and IESs (Frapporti et al., 2019; Ignarski et al., 2014; Lhuillier-Akakpo et al., 2014; Miró-Pina et al., 2022; Wang et al., 2022). We recently showed that the ISWI1 chromatin remodeling complex is necessary for IES excision precision and Ptiwi01/09 co-immunoprecipitated with ISWI1 in a crosslinked treatment (Singh et al., 2022, 2023). After the initial onset of IES excision, additional single-stranded sRNAs, iesRNAs, ranging in size from ~26 to 30 bp, are produced by Dcl5 from excised IES fragments and stabilized on Ptiwi10/11 (Furrer et al., 2017; Sandoval et al., 2014). iesRNAs were proposed to participate in a positive feedback loop for the efficient removal of all IES copies (Sandoval et al., 2014). Nevertheless, only a fraction of IES excision appears to depend on scnRNAs or iesRNAs (Nowacki et al., 2005; Sandoval et al., 2014).

Despite the knowledge gained in the past decades, the picture of IES excision is far from complete. To identify novel genes involved in IES excision, we examined proteins potentially associated with ISWI1, a chromatin remodeler we recently showed to facilitate precise IES excision (Singh et al., 2022).

Results

Identification of a novel protein involved in IES excision

Recently, we reported evidence supporting the formation of a protein complex involving ISWI1 and the ICOP proteins (Singh et al., 2023). We conducted an RNAi screen of additional genes that were unique in the ISWI1 co-immunoprecipitation (IP)-mass spectrometry (MS) and exhibited upregulation in a developmental gene expression time course from ParameciumDB (Arnaiz et al., 2017) (Fig. S1A).

In the screening, we sought phenotypic evidence for failed genome reorganization in the form of growth defects (assessed by survival tests), and substantial IES retention (assessed by IES retention PCRs). *ND7*, a gene involved in trichocyst discharge (Lefort-Tran et al., 1981), was used as a negative control as its silencing does not impair genome reorganization (Nowacki et al., 2005). *Nowa1*-KD, which affects the excision of scnRNA-dependent IESs (Nowacki et al., 2005), was used as a positive control. Candidate 2 (PTET.51.1.G0620188) displayed both IES retention and lethality in the new progeny, whereas candidate 1 (PTET.51.1.G0990120) showed high lethality without IES retention (Fig. S1B,C). Therefore, candidate 2 was selected for further investigations.

DevPF2 and DevPF1 are paralogous PHD finger proteins

The *Paramecium aurelia* species complex, to which *P. tetraurelia* belongs, underwent multiple whole-genome duplications, with many closely related paralogs generated from the most recent of these (Sellis et al., 2021). The chosen candidate has a closely related paralog (PTET.51.1.G0240213) with which it shares 86.6% identity at both the nucleotide and amino acid levels. The paralog is upregulated during sexual development as well, although earlier (Fig. 1A). HMMER3 searches of the Pfam database (Finn et al., 2003) predicted two domains in both proteins: a PHD and a PHD zinc-finger-like domain (Fig. 1B,D,E).

The highly conserved PHD domain has often been reported to mediate the interaction of nuclear proteins with histone modifications (Sanchez & Zhou, 2011), but other binding affinities have also been described (see Discussion). PHD domains possess a well-conserved motif consisting of eight cysteine and histidine residues (C4HC3) that coordinate two zinc ions, thereby providing it with structural stability. The presence of the C4HC3 motif in both paralogs was confirmed using a multiple sequence alignment with PHD domains from well-established PHD finger proteins from *Homo sapiens* and *Drosophila melanogaster* (Fig. 1C).

AlphaFold2 predicted the structures of both paralogs with high confidence for the domains (Fig. 1F,G). We compared the PHD predictions with the published structure of the WSTF (Williams Syndrome Transcription Factor) PHD finger (Pascual et al., 2000). WSTF, associated with the Williams Syndrome (Lu et al., 1998), is a subunit of the ISWI-containing chromatin remodeling complex WICH (Bozhenok et al., 2002). The superimposition confirmed the orientation of the eight C4HC3 residues in the DevPFs towards the two zinc ions (Fig. 1H), supporting the idea that both paralogs function as PHD finger proteins. Since they show development-specific upregulation (Fig. 1A), we named the paralogs development-specific PHD finger 1 (DevPF1; early-expressed paralog) and 2 (DevPF2; late-expressed paralog).

DevPF1 and DevPF2 show distinct nuclear localization

To determine the localization of both paralogs, we injected DNA constructs encoding DevPF1 and 2 C-terminally tagged with green fluorescent protein (GFP) into MACs of vegetative paramecia. The cells were collected during *Paramecium* sexual development for confocal microscopy. The injected cultures displayed no growth defects compared to non-transformed cells (Fig. S2A). However, we observed variable numbers for gametic MICs (Figs 2, 3) and new MACs (Fig. S2C) in some cells, which has been observed frequently for transgenes (e.g, Nowa1-GFP fusion; (Nowacki et al., 2005)).

Consistent with DevPF1's early peak in mRNA expression from the developmental time course in ParameciumDB, DevPF1-GFP was expressed during the onset of sexual development, but not in vegetative cells with food vacuoles containing bacteria (Figs 2A, S2B). DevPF1-GFP was distributed throughout the cytoplasm and localized in both MICs before and during the S-phase of meiosis, when these nuclei swell (Fig. 2A). Throughout the subsequent meiotic divisions, DevPF1-GFP localized to only a few of the gametic MICs (Fig. 3A). Its micronuclear localization appeared independent of nuclear division as detected by the presence of the spindle apparatus (Fig. 3A,B). During post-zygotic mitotic divisions, DevPF1-GFP was observable in certain post-zygotic nuclei, but not in all of them (Fig. 3B). Later during development, DevPF1-GFP was present in the early new MACs and remained in the new MACs throughout development up to very late stages (Fig. 2A) despite the drop in its mRNA levels (Fig. 1B). During new MAC development there was also comparatively little cytoplasmic DevPF1-GFP compared to that during meiosis.

Consistent with its mRNA expression profile, DevPF2-GFP emerged after the onset of new MAC development and localized within the new MACs, where it remained up to the late stages (Fig. 2B).

Silencing constructs partially co-silence both paralogs

We utilized RNAi by feeding to investigate the influence of the DevPFs on IES excision. Two silencing regions were selected (a and b) on each *DevPF1* and *DevPF2* (Fig. 4A). Due to the lack of regions with sufficient specificity for either of the paralogs, co-silencing was predicted (see Methods). Hence, we first experimentally verified the possibility of co-silencing with mRNA and protein levels using silencing region a, since it exhibited less off-target hits.

The mRNA levels of *DevPF1* and *DevPF2* were examined during a time course experiment (more details and further analysis follow in subsequent sections) (Fig. 4B). Consistent with the published expression profiles (Arnaiz et al., 2017), *DevPF1*

179 expression in the *ND7* control knockdown (KD) was highest during onset of
180 development and gradually declined to almost no expression at the “very late” time
181 point. The late-expressed *DevPF2* peaked at the “late” time point in the control KD. The
182 expression of both genes was strongly reduced upon their respective KDs (*DevPF1*
183 mRNA levels were reduced upon *DevPF1*-KD; *DevPF2* mRNA levels were reduced
184 upon *DevPF2*-KD). A lesser reduction was also observed upon silencing of the
185 respective paralog (*DevPF1* levels were reduced in *DevPF2*-KD and vice versa). Thus,
186 the *DevPF1* and *DevPF2* silencing constructs lead to co-silencing which is less efficient
187 than the target gene silencing.
188
189 To investigate how the changes in mRNA levels affect protein levels, we checked the
190 localization of the GFP-tagged DevPFs upon KDs. Since *DevPF1* is expressed
191 throughout the whole development, multiple developmental time points were collected
192 (Fig. S3A). For the late-expressed *DevPF2*, only cell stages with clearly visible new
193 MACs were considered (Fig. S3B). In addition to *ND7*-KD, the knockdown of *PGM*, the
194 gene encoding the PiggyMac IES excisase (Baudry et al., 2009), was performed to test
195 whether the disturbance of IES excision alters DevPF localization. Neither the
196 localization of DevPF1-GFP nor of DevPF2-GFP was impaired by either of the control
197 KDs. In contrast, the GFP signals were almost completely lost upon *DevPF1*- or
198 *DevPF2*-KD. To quantify this observation, GFP fluorescence signals were measured in
199 new MACs (Fig. 4C) as both paralogs exclusively localize to the new MACs during late
200 stages. In line with the observed reduction in mRNA levels, DevPF1-GFP expression
201 was efficiently reduced upon *DevPF1*-KD, whereas *DevPF2*-KD led to a weaker
202 reduction. For DevPF2-GFP, the levels were almost equally reduced in *DevPF1*- and
203 *DevPF2*-KD. Thus, we confirmed co-silencing on both mRNA and protein levels with
204 reduced silencing efficiency compared to the targeted KD. Therefore, all results
205 obtained in KD experiments must be considered, at least in part, as a combined effect
206 of silencing both DevPF1 and DevPF2, albeit with only a partial contribution from the
207 non-targeted gene silencing.
208

To further investigate the impact of co-silencing on the KD analysis we examined IES retention score (IRS) correlations of multiple KD replicates (more details and further analysis in subsequent sections). The *DevPF2*-KD replicates showed high to moderate correlation among each other while they correlated less well with two out of four *DevPF1*-KD replicates (Fig. 4D). This suggests that despite the partial co-silencing, individual KD effects might be observed.

***DevPF1* and *DevPF2* affect IES excision genome-wide**

The influence of the DevPFs on genome reorganization was initially investigated with survival tests and IES retention PCRs upon KDs. Reduced protein levels during sexual development can induce errors including IES retention, impacting the survival of the subsequent generation. For survival tests, the growth of the cells that completed their sexual development was followed for several divisions. IES retention PCRs test for the presence (failed excision) of specific IESs in the new MAC genome. *ND7*-KD and *PGM*-KD were used as negative and positive control, respectively. To investigate the possibility that the observed effects result from off-target silencing of an unrelated gene, two silencing probes (a and b) were tested for each paralog (Fig. 4A). *DevPF1* and *DevPF2* KDs with either of the silencing probes resembled *PGM*-KD, with high lethality in the new progeny (Fig. 5A) and retention of selected IESs (Fig. 5B). This indicates that both *DevPF1* and *DevPF2* contribute to IES excision.

Next, we tested genome-wide IES retention in enriched new MAC DNA samples. We observed considerably elevated levels of retained IESs in both *DevPF1*- and *DevPF2*-KD (Fig. 5C,D). Notably, differences between replicates of the same KDs were observed, whereas replicate pairs processed in parallel (see Methods) exhibited similar profiles. Correlations among the paralog replicates indicated that despite varying IES retention distributions, *DevPF2*-KD replicates demonstrated high correlations among themselves (Fig. 4D). *DevPF1*-KD replicates correlated less well with each other, and *DevPF1*-KD replicate 3 (3) showed a high correlation with the *DevPF2*-KDs. This

238 indicates that *DevPF2*-KD replicates were more consistent than the *DevPF1*-KD
239 replicates.
240
241 Genes that work closely together are expected to show similar KD effects on IES
242 retention. To identify functionally related genes, *DevPF1*-KD and *DevPF2*-KD IRS data
243 was correlated with published data from other gene KDs (Fig. 5E). *DevPF2*-KD (4) was
244 selected from the *DevPF2* replicates. *DevPF1*-KD (2) and *DevPF1*-KD (4) were
245 selected as representative of the variability observed in the *DevPF1*-KDs. *DevPF2*-KD
246 (4) displayed high correlation with other KDs, such as *TFIIS4* and *DCL2/3/5* (Fig. 5E).
247 Moderate correlation was observed for *DevPF1*-KD (4) with *SPT5m*, whereas *DevPF1*-
248 KD (2) did not correlate well with any of the tested KDs.
249
250 Short IESs are proposed to predominantly rely on the excision complex (specifically
251 Pgm (Baudry et al., 2009) and Ku80c (Marmignon et al., 2014)) for removal, while long
252 IESs tend to require additional molecules for excision (Sellis et al., 2021). To determine
253 whether *DevPF1*- and *DevPF2*-KD preferentially affect long IESs, the length distribution
254 of the top 10% of highly retained IESs in each KD was plotted (Fig. S4A,B, Table S1). In
255 comparison to the length distribution of all IESs, *DevPF2*-KD (4) showed an
256 overrepresentation of long IESs, similar to observations in *EZL1*-KD, silencing of the
257 catalytic subunit of the PCR2 complex (Frapporti et al., 2019; Lhuillier-Akakpo et al.,
258 2014), or *DCL2/3/5*-KD, silencing of the scnRNA and iesRNA biogenesis proteins
259 (Lepère et al., 2009; Sandoval et al., 2014), (Fig. S4A). Conversely, the highly retained
260 IESs in *DevPF1*-KD (2) did not show the same overrepresentation and resembled the
261 profile in *PGM*- and *KU80c*-KD, silencing of two members of the excision complex.
262 Again, the replicates of the *DevPF* KDs exhibited variation in the extent of the observed
263 effect (Fig. S4B).
264
265 Defects in IES excision not only result in the retention of IESs but can also lead to
266 excision at alternative TA boundaries. So far, alternative excision above background
267 levels has only been reported for silencing of ISWI1 and its complex partners (Singh et

al., 2022, 2023). Neither *DevPF1*-KD nor *DevPF2*-KD resulted in elevated levels of alternative excision (Fig. S4; Table S2).

***DevPF1*- and *DevPF2*-KD alter the small RNA population**

The early-produced 25 nt scnRNAs and the late-produced 26-30 nt iesRNAs have been proposed to assist MIC-specific sequence targeting in the new MACs (Sandoval et al., 2014). Therefore, the small RNA populations across developmental time points upon DevPF KDs were analyzed (Figs 6A, S6A). In *DevPF1*-KD (2), scnRNA production was completely abolished, an effect also observed in the KD of genes proposed to be involved in scnRNA production: the two scnRNA-processing genes *DCL2* and *DCL3* (Sandoval et al., 2014), and *STP5m*, involved in the generation of the transcripts serving as substrates for Dcl2/3 cleavage (Gruchota et al., 2017). The KD of the late-expressed *DevPF2* showed a much weaker reduction of scnRNA production, which might be caused by co-silencing of *DevPF1*.

To further investigate DevPF1's effect on the scnRNA pathway, we observed Ptiwi09-GFP localization upon *DevPF1*-KD. Ptiwi09, together with Ptiwi01, stabilizes the scnRNAs throughout scnRNA selection in the parental MAC and targeting of MIC-specific sequences in the new MACs (Bouhouche et al., 2011; Furrer et al., 2017). As previously described (Bouhouche et al., 2011; Singh et al., 2023), Ptiwi09-GFP localizes first to the cytoplasm and parental MAC with a transient localization in the swelling MICs before shifting to the new MAC (Fig. 6B). Upon *DevPF1*-KD (Fig. 6B), the localization to the MICs before meiosis I is not impaired; however, the translocation into the parental MAC is strongly reduced and Ptiwi09-GFP predominantly remains in the cytoplasm throughout meiosis II and MAC fragmentation. We have reported a similar change in Ptiwi09-GFP localization upon *DCL2/3*-KD (Singh et al., 2023), suggesting that the loss of scnRNAs is responsible for the failed protein transfer into the parental MAC. Similar to *DCL2/3*-KD, *DevPF1* depletion does not affect Ptiwi09-GFP's localization to the new MACs (Fig. 6B).

Interestingly, DevPF1-HA IP at two developmental time points (early: about 30% fragmentation; late: visible new MACs in fragmented cells) identified Ptiwi01/09 as potential interaction partners of DevPF1 with a higher enrichment in the early than the late time point (Fig. S6B, Table S3). None of the other small RNA-related proteins were detected (Dcls, Spt5m, TFIIS4 or Ptiwi10/11).

For both *DevPF1*- and *DevPF2*-KD, iesRNA production was impaired. iesRNAs are proposed to derive from dsRNAs transcribed from excised IESs (Allen et al., 2017; Sandoval et al., 2014). Hence, failed excision of IESs in *DevPF1*- or *DevPF2*-KD contributes to reduced iesRNA levels, as has consistently been observed for many other KDs of genes involved in *Paramecium* genome editing (Charmant et al., 2023; de Vanssay et al., 2020; Ignarski et al., 2014; Maliszewska-Olejniczak et al., 2015; Singh et al., 2022; Wang et al., 2023). The lack of scnRNAs in the *DevPF1*-KD cannot explain the absence of iesRNAs, as these accumulate even if the preceding scnRNA production is blocked (Sandoval et al., 2014). In the late time point analyzed for DevPF IPs, peptides mapping to Ptiwi10/11/06 were detected in DevPF2-IP (Fig. S6C, Table S3), but not DevPF1-IP (Fig. S6B). Therefore, DevPF2 might contribute to iesRNA biogenesis by an interaction with Ptiwi proteins.

***DevPF1*- and *DevPF2*-KD affect mRNA expression**

Since PHD fingers have often been reported to be involved in gene expression regulation (Aasland et al., 1995; Sanchez & Zhou, 2011) we sought to investigate whether the *DevPF* KDs alter mRNA expression levels during development. Batch effects had a major influence on the variance within the replicates (Fig. S7A), as observed for IES retention (Fig. 5C,D).

DevPF1-KD showed almost no differentially expressed genes compared to *ND7*-KD during onset of development (Fig. 7A). During this early stage, genes are transcribed solely from the parental MAC, where DevPF1-GFP does not localize (Figs 2A, 3). Surprisingly, in *DevPF2*-KD, a high number of genes were differentially expressed

during the onset of development (Fig. 7A). Since *DevPF2* is late-expressed and *DevPF1*-KD showed no effect, the observed difference might be caused by differing cell stages within the collected populations of *DevPF2*-KD and *ND7*-KD. During the “early”, “late” and “very late” time points, *DevPF1*- and *DevPF2*-KD showed similar changes in mRNA expression.

The abolishment of development-specific small RNAs in the *DevPF*-KDs might result from downregulation of genes involved in scnRNA or iesRNA production. We observed no general trend indicating a drastic reduction of expression of scnRNA-related genes, like *DCL2*, *PTIWI01* or *SPT5m* (Figs 7B, S7B, Table S4, S5). However, these trends in expression should be considered with the caveat of considerable expression variability and limitation of the number of replicates that could practically be obtained. At least for Ptiwi09, the localization experiments upon *DevPF1*-KD confirmed no loss in protein levels (Fig. 6B).

The expression of iesRNA-related genes was altered in both *DevPF1*- and *DevPF2*-KD compared to *ND7*-KD (Figs 7D, S7B, Table S4, S5). *DCL5*, the Dicer-like protein responsible for the initial cleavage of IES derived dsRNAs into small iesRNAs (Sandoval et al., 2014), was downregulated (Table S4, S5) in early stages, but tended to be upregulated in the very late stage (Table S4, S5). *PTIWI10* and *PTIWI11*, the Piwi proteins responsible for further processing and stabilization of iesRNAs during the positive feedback loop (Furrer et al., 2017), were downregulated in both *DevPF1*- and *DevPF2*-KD (Table S4, S5). Successful expression of *PTIWI10/11* has been proposed to depend on IES excision since both genes are expressed from the new MAC and harbor IESs in their flanking/coding regions (Furrer et al., 2017) (Fig. S7C). If IES retention was the only cause for downregulation, one would expect higher IRSs for these IESs in KDs with lower mRNA levels. While the mRNA reduction is stronger in *DevPF1*-KD than in *DevPF2*-KD (Fig. 7D, Table S4, S5), this trend is not reflected in the IRSs of the IESs whose retention is proposed to interfere with *PTIWI10/11* expression (Table 1). In most of the KD replicates, there is no or low retention (IRS < 0.1) and the replicates showing moderate to high retention (0.1 < IRS < 0.3) belong to both *DevPF1*-

and *DevPF2*-KD. Hence, the reduced mRNA levels of *PTIWI10/11* cannot only be explained by IES retention.

Discussion

Implications of the PHD domain for DevPF1 and DevPF2 functions

Genome reorganization is a fundamental process underlying cell and immune system development and some diseases (Bassing et al., 2002; Forment et al., 2012; Mani & Chinnaiyan, 2010; Rooney et al., 2004). Ciliates undergo massive genome reorganization during the maturation of their somatic genome. This makes them excellent models for studying the complex mechanisms involved in the targeted elimination of genomic sequences (Beisson et al., 2010d). In the present study, we described two paralogous PHD finger proteins, DevPF1 and DevPF2, involved in IES excision in *Paramecium*. Both paralogs harbor a PHD and a PHD zinc finger-like domain (Fig. 1). These domains belong to the zinc-finger family and the PHD domain is characterized by a well-conserved C4HC3 motif (Aasland et al., 1995; Schindler et al., 1993). The eight core amino acids of this motif coordinate two zinc ions and thereby provide structural stability to the domain (Pascual et al., 2000). Among other histone-binding domains, such as bromodomains or PWWP domains, PHD fingers are the smallest (Fleck et al., 2021; Miller et al., 1985). Multiple sequence alignment and structure predictions confirmed the presence of the characteristic C4HC3 motif in both DevPF1 and DevPF2 (Fig. 1), suggesting that both PHDs might be functional.

PHD fingers, mainly nuclear proteins, are often considered epigenetic readers, recognizing histone modifications, primarily on the histone 3 (H3) N-terminal tail (Sanchez & Zhou, 2011). Peptides matching to histones were enriched in the DevPF-IPs of late developmental time points (Fig. S6B,D, Table S3), however none of them was specific to H3. PHD fingers have been reported to bind non-H3 partners, like DNA, histone 4, or other proteins (Bienz, 2006; Black & Kutateladze, 2023; Gaurav & Kutateladze, 2023; L. Liu et al., 2012; Oppikofer et al., 2017). The combination of the

PHD and PHD-zinc-finger-like domain in the DevPFs may enable the paralogs to simultaneously recognize two adjacent histone modifications, as demonstrated for tandem PHD domains (Zeng et al., 2010). PHD domains are also found in various chromatin associated proteins involved in gene regulation. Notably, ISWI-containing chromatin remodeling complexes often include a subunit with a PHD domain, such as the ACF (Eberharter et al., 2004), NURF (Haitao Li et al., 2006; Wysocka et al., 2006) or WICH (Bozhenok et al., 2002) complexes.

DevPF2 was initially identified in pulldowns of the ISWI1 protein, and, thus far, no PHD-containing protein has been shown to be a part of this remodeling complex (Singh et al., 2022, 2023). It is intriguing to consider that DevPF2 might contribute PHD functionality to the ISWI1 chromatin remodeling complex. However, *DevPF2*-KD does not show elevated levels of alternative excision (Fig. S4C-E) that are characteristic of other members of the complex so far (Singh et al., 2022, 2023) and ISWI1 was not identified as a potential interaction partner in the DevPF2-IP (Fig. S6C). If DevPF2 interacts with the ISWI1 complex, we infer that it may not be a core complex component, particularly as it does not contribute to excision precision.

A potential role for DevPF1 and DevPF2 as transcription factors?

A potential role in non-coding transcription in the MICs (for scnRNA production)

DevPF1's localization in the MICs (Figs 2A, 3) and its importance for scnRNA production (Fig. 6A) could point towards its involvement in the bidirectional transcription of the MIC genome for scnRNA production. Spt5m (Gruchota et al., 2017) and TFIIS2/3 (Maliszewska-Olejniczak et al., 2015) are proposed to be involved in this micronuclear transcription. One of the *DevPF1*-KD replicates showed moderate IRS correlation with *SPT5m* (Fig. 5E) (to our knowledge, no IRS data exists for TFIIS2 or TFIIS3) and *SPT5m*-KD also reduces scnRNA production. The localization of Dcl2-GFP (Lepère et al., 2009), Ptiwi09-GFP (Fig. 6B) and DevPF1-GFP (Fig. 2A) in the swelling MICs suggests that scnRNA biogenesis occurs during the S-phase of meiosis. Ptiwi09 and

DevPF1 may interact in the MICs or the cytoplasm. Non-crosslinked IP's would be needed to further verify this interaction. However, *PTIWI01/09*-KD does not completely abolish scnRNAs (Furrer et al., 2017), indicating that DevPF1 acts upstream of scnRNA loading and guide strand removal. Future investigations of bi-directional transcription and scnRNA biogenesis will allow to identify how all these molecules cooperate.

Spt5m-GFP, TFIIS2/3-GFP and DevPF1-GFP are present in the MICs beyond S-phase and localize to the new MACs at later stages (Gruchota et al., 2017; Maliszewska-Olejniczak et al., 2015). Their role in the MIC during meiotic divisions remains unknown. It was speculated that Spt5m might be involved in co-transcriptional deposition of epigenetic marks that sustain meiotic processes, ultimately aiding in IES targeting. The potential of PHD domains to bind histone modifications raises a similar possibility for DevPF1. However, its role appears to be more specific, as DevPF1 is not present in all gametic and zygotic nuclei simultaneously (Fig. 2&3).

Msh4/5, homologs of proteins essential for crossover, are also present in all gametic nuclei during the first and second meiotic division, and their silencing leads to substantial IES retention (Rzeszutek et al., 2022). However, their non-canonical functions that lead to IES retention are not yet fully understood (Rzeszutek et al., 2022). Since new MACs develop in *DevPF1*-KD (Figs 6B,S3) and *MSH5*-KD cells, neither of the genes are essential for crossover or karyogamy. More research will be needed in future to decipher the functions of the DevPF proteins in the gametic nuclei.

A potential role in non-coding transcription in the new MAC (for scnRNA-based targeting and iesRNA production)

Non-coding transcription in the new MAC, which is hypothized to generate substrates for scnRNA pairing, was proposed to be regulated by the putative transcription elongation factor TFIIS4 that specifically localizes to the early new MACs (Maliszewska-Olejniczak et al., 2015). *DevPF2*-KD IRSs of some replicates correlated most strongly with *TFIIS4*-KD (Fig. 5E), pointing towards a shared functionality. Both

DevPF1 and DevPF2 have the potential to act in the same regulatory process as TFIIS4 because both their GFP-fusions localize to the new MACs. In fact, there are reports of transcription factors that combine the TFIIS and PHD domains: Bypass of Ess1 (Bye1) protein in S*accharomyces cerevisiae* harbors a PHD and a TFIIS-like domain, with the former recognizing histone 3 lysine 4 trimethylation and the latter establishing contact with polymerase II for transcriptional regulation (Kinkelin et al., 2013; Pinskaya et al., 2014). It is possible that similar functionality is separated on two individual proteins in *Paramecium*. However, TFIIS4 was not detected in either of the DevPF-IPs in the late developmental stage.

The production of iesRNAs was also proposed to depend on the non-coding transcription of concatenated excised IES fragments (Allen et al., 2017; Sandoval et al., 2014). Although it was established that IES concatemers are likely formed by DNA ligase 4 (Lig4) (Allen et al., 2017), little is known about the proposed bidirectional transcription to produce substrates for Dcl5 cleavage. Allen *et al.* speculated on the involvement of TFIIS4. Since iesRNA production is almost completely abolished in *DevPF1*- and *DevPF2*-KD, a contribution to this transcription is plausible.

The potential function of the DevPFs may extend far beyond TFIIS4-dependent transcription: whereas TFIIS4-GFP localizes transiently to early new MACs (Maliszewska-Olejniczak et al., 2015), DevPF2-GFP and DevPF1-GFP remain in the new MACs for much longer (Fig. 2).

A potential role in gene transcription in the parental and the new MAC

Early in development, the parental MAC is solely responsible for gene expression and, after genome reorganization progresses, the new MAC contributes at later stages (Berger, 1973). In *Tetrahymena*, E2F family transcription factors were shown to control the cell cycle through gene expression during meiosis (Zhang et al., 2018). DevPF1 and DevPF2 are unlikely to be active in the parental MAC since none of the GFP-fusion proteins localized there (Fig. 2). Consistently, *DevPF1*-KD showed no differential gene expression compared to *ND7*-KD during the onset of development (Fig. 7C) and

Ptiwi09-GFP expression was not impaired upon *DevPF1*-KD (Fig. 6B). However, it is difficult to reach a definite conclusion for other genes due to the high variability in expression between the replicates (Figs 7D, S7B) and the high number of differentially expressed genes in *DevPF2*-KD (Fig. 7C) observed during the onset of development. Cells in the "onset" time point are challenging to collect because cell staging relies on MAC morphology changes visualized by DAPI staining. Truly vegetative cells cannot be distinguished from cells initiating meiosis since their MACs look the same; however, the gene expression profiles are expected to differ substantially (Figs 2A, S2B). The collection of subsequent time points is more reliable because the alteration of old MAC shape as development progresses is pronounced.

At the subsequent stages, *DevPF1*- and *DevPF2*-KD affected similar genes. Either, the changes are nonspecific to the *DevPF*-KDs and result from the proposed nuclear crosstalk to adjust transcription levels to accommodate for failed IES excision (Bazin-Gélis et al., 2023) or they are specific to the *DevPF*-KDs and both paralogs exhibit similar functions in the regulation of gene expression. Interestingly, differential expression was observed at the "early" time point (Fig. 7C). *GTSF1*-KD, also causing substantial IES retention, hardly shows any differentially expressed genes at a comparable stage (*DevPF1/2*-KD: 282/231 differentially expressed genes, respectively, at about 30% fragmentation (Fig. 7C); *GTSF1*-KD: 10 differentially expressed genes at about 30-50% fragmentation; (Wang et al., 2023)). This indicates that the early change in gene expression might be specific to *DevPF*-KDs, potentially mediated by other proteins shuttling into the parental MAC. Since Ptiwi09-GFP translocates efficiently to the parental MAC upon *GTSF1*-KD (Wang et al., 2023) but not upon *DevPF1*-KD (Fig. 6B), it might be worth investigating differential expression upon *PTIWI01/09*-KD.

Late in development, gene expression starts from the new MACs (Berger, 1973), where both DevPF paralogs localized (Fig. 2). Some late-expressed genes, like *PTIWI10*, are expressed only from the new MAC after the initial onset of IES excision (Furrer et al., 2017). Indeed, *PTIWI10/11* mRNA levels are downregulated in *DevPF1*-KD or *DevPF2*-KD (Fig. 7D, S7B, Table S4, S5). This trend cannot be explained solely by the strength

of retention observed for the IESs interfering with *PTIWI10/11* expression (Table 1). It suggests that DevPF1 and DevPF2 may regulate gene expression in the new MAC, albeit specifically for some genes like *PTIWI10* and *PTIWI11.* The extent of gene expression regulation by the DevPFs beyond these genes remains uncertain. To further investigate if the DevPFs serve as transcription factors, and if so, which genes they regulate, genes associated with DevPF binding could be identified by techniques like Cut-and-Run (Skene et al., 2018) and compared to mRNA expression changes upon *DevPF*-KDs.

Potential cytoplasmic functions

In contrast to the other putative transcription factors discussed so far (Spt5m, TFIIS2/3/4, DevPF2), DevPF1-GFP exhibits a pronounced cytoplasmic distribution in the early stages of development (Fig. 2A). While most described PHD fingers are nuclear proteins, some can be recruited to the cytoplasm or plasma membrane by binding partners (Betz et al., 2004; Gozani et al., 2003). DevPF1 may play a role in transmitting signals of sensed starvation to the MICs, initiating sexual development. As *DevPF1* is not constitutively expressed during vegetative growth (Figs 1B, S2B), another factor is needed to first initiate *DevPF1*'s gene expression in the parental MAC. However, DevPF1 might interact with specific markers of starvation in the cytosol, promoting early sexual processes. If that is the case, DevPF1 is not essential for general meiotic processes, as meiosis and new MACs development show no defects in *DevPF1*-depleted cells (Figs 6B,S3). Since peptides matching Ptiwi01/09 were identified in the DevPF1-IP, the Ptiwi01/09 complex is a potential binding partner of DevPF1 in the cytoplasm. However, since Ptiwi01/09 are highly expressed proteins (Bouhouche et al., 2011), further IP experiments would be needed to verify this interaction.

DevPF1's selective localization to gametic and post-zygotic nuclei

The selective localization of DevPF1 to certain gametic and post-zygotic nuclei (Fig. 3) raises intriguing questions about its potential role in nuclear fate decisions. The survival

and destruction of the gametic nuclei depends on their subcellular positioning (Grandchamp & Beisson, 1981). DevPF1 may play a role in either promoting their movement or preparing for their degradation. However, the observed number of nuclei simultaneously containing DevPF1-GFP (zero to four) neither fits the number of nuclei selected for survival (one) nor for degradation (seven). DevPF1 may either contribute to this process successively or may not be directly related to the nuclear fate itself. The fate of the post-zygotic nuclei is decided during the second mitotic division by the subcellular localization of the division products (Grandchamp & Beisson, 1981). This means, from each post-zygotic nucleus, one of the division products will remain as MIC and one develops into a new MAC. During the second mitotic division, DevPF1-GFP was observed in one of the two dividing nuclei. Its localization in the precursor of one MIC and one MAC without being present in the precursor of the other MIC and MAC, does not imply its involvement in the nuclear fate decision. Furthermore, *DevPF1*-KD neither impaired the selection of gametic nuclei nor the differentiation of the new MACs.

The specific localization of nuclear proteins to certain nuclei in multinuclear cells has been studied extensively in insect embryos. In *Drosophila*, the transcription factors Bicoid (Driever & Nüsslein-Volhard, 1988) and Dorsal (Roth et al., 1989) establish the anterior-posterior, and dorsal-ventral axis, respectively, by initiating gene expression depending on the cytoplasmic localization of the nuclei. The activity of the transcription factors is restricted by gradients to a certain cytoplasmic region (Morisato & Anderson, 1995; Spirov et al., 2009). However, DevPF1-GFP's nuclear localization does not appear associated with subcellular localization of the nuclei and it remains unclear how DevPF1-GFP is specifically recruited.

As only fixed cells were examined, the dynamics of DevPF1-GFP localization were not captured. The fact that DevPF1-GFP localization is independent of nuclear divisions (Fig. 3B), combined with observations of cells at the meiotic or mitotic division stage with an absence of DevPF1-GFP in all nuclei (Fig. S2C), suggests that DevPF1 localization might be asynchronous and transient. Possibly it is recruited to each of the gametic nuclei at some point before the completion of the second meiotic division and to

each of the post-zygotic nuclei before completion of the second mitotic divisions. Live cell imaging could illuminate the dynamics of DevPF1 localization and its correlation with nuclear fate. However, this approach presents challenges, as it requires confocal imaging to capture the DevPF1-GFP signal in the MICs, and the observation time scale would need to span across multiple hours of *Paramecium* development.

DevPF1: a general factor for IES excision

DevPF1 plays a role throughout sexual development: from the early stages before meiosis to the very late stages (Fig. 2A). It appears to influence various aspects of genome reorganization in the MICs and the new MACs, including scnRNA production and potentially expression of certain genes. Consequently, the depletion of *DevPF1* affects the excision of a wide range of IESs (Fig. 5C). However, it is important to reiterate that we observed high batch-to-batch variability in the *DevPF* replicates in both IES retention (Fig. 5C, D) and mRNA expression (Figs 7D, S7B). The time point collection had a major influence on mRNA levels (Fig. S7A). Variable new MAC enrichment by a sucrose gradient might introduce variation into the IRS analysis, as fragments of the parental MAC add unexcised IES sequences, diluting the effect of IES retention (Charmant et al., 2023). Fluorescence-activated nuclear sorting (FANS) enables better nuclear separation in *Paramecium* (Charmant et al., 2023; Guérin et al., 2017) and should be able to eliminate most of such variation. Additionally, microinjection of DNA into macronuclei before RNAi experiments can be used to control for contaminating DNA from old MAC fragments.

Revisiting previous KD experiments with additional replicates would be worthwhile to explore the extent of batch-to-batch IRS and expression variance for other KDs. Noteworthy, variability in IES retention across replicates has recently been shown for *GTSF1* (Charmant et al., 2023; Wang et al., 2023), suggesting this phenomenon is not restricted to *DevPF1* and *DevPF2*. In general, KD experiments are challenging to tightly control for reproducibility, and more effort should be invested in generating knockouts in *Paramecium*, as established in *Tetrahymena* (Chalker, 2012).

It has been shown that evolutionarily old IESs tend to be short and are excised early in development, independent of additional factors apart from the excision machinery (Sellis et al., 2021). On the other hand, evolutionarily young IESs tend to be long, later excised and dependent on the scnRNA pathway and the deposition of histone modifications in the new MAC for their excision (Sellis et al., 2021; Swart et al., 2014; Zangarelli et al., 2022). In line with this, most gene KDs tested in this study exhibited an overrepresentation of long IESs among their most highly retained IESs, including *DevPF2* (Fig. S4A,B). Only *PGM*-KD, *KU80c*-KD and two of the *DevPF1*-KD replicates showed no preference for long IESs. Pgm and Ku80c are components of the excision machinery and are therefore expected to affect all IESs. While DevPF1 may not be a direct part of the excision machinery, it appears to have a general contribution to IES excision, regardless of the length of the IES. Consequently, we propose that DevPF2 contributes to the excision of long IESs, while DevPF1 may serve as a more general factor.

Methods

Paramecium tetraurelia cultivation

Mating type 7 (MT7) cells from strain 51 of *Paramecium tetraurelia* were grown in Wheat Grass Powder (WGP, Pines International) medium supplemented with 10 mM sodium phosphate buffer (pH 7.3). WGP medium was bacterized with *E.coli* strain HT115 to feed paramecia, and the cultures were maintained either at 27°C or at 18°C according to the standard protocol (Beisson et al., 2010b, 2010c).

Protein localization imaging by fluorescence microscopy

Plasmids for microinjection were generated by amplifying the coding and flanking sequences from MT7 genomic DNA and introducing them with the PCR-based method CPEC (Quan & Tian, 2011) into the L4440 plasmid (Addgene, USA). DevPF1 was

626 expressed with its endogenous flanking regions (304 bp upstream of the DevPF1 start
627 codon and 272 bp downstream of the DevPF1 stop codon). DevPF2 endogenous
628 flanking regions (455 bp upstream the DevPF2 start codon and 273 bp downstream of
629 the DevPF2 stop codon) yielded no expression. Therefore, as *PGM* exhibits a similar
630 expression profile to *DevPF2* (Fig. 1B), DevPF2 genomic coding sequence was inserted
631 between the *PGM* flanking regions (96 bp upstream of the *PGM* start codon and 54 bp
632 downstream of the *PGM* stop codon). Before the stop codon, the *GFP* coding sequence
633 was connected to the protein coding sequences via a glycine-serine-linker
634 (SSGGGSGGSGGGS). 60 µg of plasmid DNA was linearized with AhdI (New England
635 Biolabs, UK) and extracted with phenol-chloroform for injection.
636
637 *Paramecia* were microinjected with either C-terminally GFP-tagged DevPF1
638 (endogenous regulatory regions) or C-terminally GFP-tagged DevPF2 (*PGM* regulatory
639 regions) following the standard protocol (Beisson et al., 2010a). Sexual development
640 was induced by starvation and cells of different developmental stages were collected
641 and stored in 70% ethanol at -20°C. To stain cells with DAPI (4,6-diamidino-2-2-
642 phenylindole), cells were dried on a microscopy slide, washed twice with phosphate-
643 buffered saline (PBS) and permeabilized for 10 min at RT (room temperature) with 1%
644 Triton X-100 in PHEM (PIPES, HEPES, EGTA, magnesium sulfate), fixed with 2%
645 paraformaldehyde (PFA) in PHEM and washed once for 5 min at (RT) with 3% BSA
646 (bovine serum albumin, Merck-Sigma, Germany) in Tris-buffered saline with 10 mM
647 EGTA and 2 mM $MgCl_2$ (TBSTEM). After DAPI (2 µg/ml in 3% BSA) incubation for 7-10
648 min at RT, the cells were mounted 40 µl of ProLong Gold Antifade mounting medium
649 (Invitrogen, USA) or ProLong Glass Antifade mounting medium (Invitrogen, USA). For
650 α-tubulin staining, after permeabilization and fixation, cells were blocked for 1 h at RT
651 with 3% BSA and 0.1% Triton X-100. Primary rat anti-α-tubulin antibody (Abcam, UK)
652 was diluted 1:200 in 3% BSA and 0.1% Triton X-100 in TBSTEM and incubated
653 overnight at 4°C. After 3 washes with 3% BSA, the goat anti-rat secondary antibody
654 conjugated to Alexa fluorophore 568 (Abcam, UK) was diluted 1:500 in 3% BSA and
655 0.1% Triton X-100 in TBSTEM and incubated for 1 h at RT. After two washes, cells
656 were stained with DAPI and mounted with Prolong Glass Antifade mounting medium.

Images were acquired on a confocal SP8 Leica fluorescence microscope (60x/1.4 oil objective) with constant laser settings. The detector (photon multiplier) gain for the DAPI signal (430-470 nm) varied to accommodate differences in signal strength (500-550 V). Postprocessing was done in Fiji (version 2.14.0/1.54f) (Schindelin et al., 2012). Brightness and contrast in the GFP channel was set the same in all the images to be compared (Figs 2, S2B: DevPF1-GFP: Min 0, Max 681 and DevPF2-GFP: Min 0, Max 170; Figs 3, S2C: DevPF1-GFP: Min 0, Max 703; Fig. S3: constant settings for each cell stage).

Knockdown efficiency validation using fluorescence intensity

Cells injected with either DevPF2-GFP or DevPF1-GFP were subjected to KDs of *ND7*, *PGM*, *DevPF2* and *DevPF1* genes. Cells during new MAC development were collected (for details see methods on silencing experiments), then stained with DAPI and mounted on ProLong Glass Antifade as described above. Images of a single z-plane through the new MAC were acquired on a SP8 Leica Confocal microscope with 60x/1.4 oil objective using the same laser settings for all images. For each KD, 10 cells were imaged. In Fiji software (version 2.14.0/1.54f), the brightness and contrast in the GFP channel was set the same values for all images compared in the same analysis (DevPF1-GFP injected cells: Min 0, Max 1078; DevPF2-GFP injected cells: Min 0, Max 298). Fluorescent signal was measured in a constant area in 1 MAC of each cell and the area mean was used as intensity for this nucleus. The area was set in the DAPI channel and the fluorescence was measured in the GFP channel. Since the same area was measured for each nucleus, no normalization was used to account for nuclear size variation. To account for background fluorescence, GFP fluorescence in non-transformed wild type cells was measured and the mean of all wild type cells was subtracted from all measured intensities. All intensities were normalized to the mean of all *ND7*-KD cells in the corresponding injection. All scripts are available from https://github.com/Swart-lab/DevPF_code.

Co-immunoprecipitation

Paramecia were injected with either Human influenza hemagglutinin (HA)-tagged DevPF1 (same cloning strategy as described before) or GFP-tagged DevPF2. For DevPF1-HA, an early time point (about 30% fragmentation) and late time point (new MACs clearly visible in fragmented cells) was collected, while for DevPF2-GFP, only the late time point was collected. Non-transformed wild type cells were collected as controls. Cells were washed twice with 10 mM Tris and as much liquid was removed as possible. For 300 ml initial culture volume, cells were fixed with 1 ml 1% PFA for 10 min at RT and quenched with 100 µl of 1.25 M glycine for 5 min at RT. After one wash with PBS (centrifugation for 1 min at 4°C and 1000 g), 2 ml lysis buffer (50 mM Tris, 150 mM NaCl, 5 mM $MgCl_2$, 1% Triton X-100, 10% Glycerol and cOmplete protease inhibitor EDTA-free (Roche, Germany)) were added and cells were sonicated using an MS72 tip on a Bandelin Sonopulse device with 52% amplitude for 15 s on ice. The pellet and input fraction were separated by centrifugation (13,000 g, 4°C, 30 min).

To enrich HA-tagged proteins, 50 µl beads (Anti-HA-affinity matrix, Merck-Sigma, Germany) were washed thrice (500 g, 4°C, 2 min) in ice-cold IP buffer (10 mM Tris pH 8, 150 mM NaCl, 1 mM $MgCl_2$, 0.01% NP-40, 5% Glycerol, cOmplete protease inhibitor EDTA-free (Roche, Germany) and incubated with 1 ml of cleared input lysate overnight at 4°C. After four washes with ice-cold IP buffer, the bound proteins were eluted from the beads in 50 µl 2× PLB (10% SDS, 0.25 M Tris pH 6.8, 50% Glycerol, 0.2 M DTT, 0.25% Bromophenol blue) at 98°C for 20 min (IP fraction).

To enrich GFP-tagged proteins, 25 µl beads (GFP-Trap Agarose beads, Chromotek, Germany) were washed once with ice-cold 20 mM Tris pH 7.5 with 100mM NaCl (2,500 g, 4°C, 5 min) and thrice in ice-cold IP buffer. Beads were incubated with 1 ml cleared input lysate for 1 to 2 h at 4°C and washed four times with ice-cold IP buffer. Bound proteins were eluted in 30 µl 2× PLB at 98°C for 20 min (IP fraction).

For western blots, 0.5% of total input and 15% of total IP fraction were resolved on 10% SDS-PAGE gels and wet transferred onto a 0.45 µm nitrocellulose membrane for 2 h at 80 V and 4°C (Bio-Rad, Germany). The membrane was blocked for 1 h in 5% BSA in PBST (PBS + 0.2% Tween20). HA-tagged proteins were detected with an HRP-conjugated anti-HA antibody (sc-7392 HRP, Santa Cruz, USA) diluted 1:500 in PBST and incubated overnight at 4°C. GFP-tagged proteins were detected with an primary anti-GFP antibody (ab290, Abcam, UK) diluted 1:2000 and incubated overnight at 4°C followed by an secondary anti-rabbit HRP conjugated antibody (12-348, Merck Millipore, Germany) diluted 1:5000 in PBST and incubated for 1 h at RT. Membranes were screened using AI600 (GE Healthcare, Germany).

Samples were sent to EMBL's Proteomics Core Facility in Germany for mass spectrometry experiments and analysis. Using R, contaminants were removed from the FragPipe output files (protein.tsv, (Kong et al., 2017)), and only proteins quantified with a minimum of two razor peptides were included for subsequent analysis. After log2 transformation of raw TMT reporter ion intensities, batch effect correction (limma package's (Ritchie et al., 2015) 'removeBatchEffects' function), and variance stabilization normalization (vsn) with vsn package (Huber et al., 2002), the abundance difference in WT and DevPF samples was maintained by determining different normalization coefficients. To investigate differential protein expression (limma package), replicate information was incorporated in the design matrix with the 'lmFit' limma function. "hit" annotation: false discovery rate (FDR) smaller 5% and a fold change of at least 100%. "candidate" annotation: FDR smaller 20% and a fold change of at least 50%. Scripts to generate volcano plots are available from https://github.com/Swart-lab/DevPF_code.

Silencing experiments, survival test and IES retention PCR

Silencing constructs for *DevPF2* and *DevPF1* were generated by cloning genomic gene fragments into a T444T plasmid (Sturm et al., 2018) (Addgene, USA) using CPEC (Quan & Tian, 2011). For both *DevPF1* and *DevPF2*, two silencing regions were

selected: DevPF1 silencing region a (525 bp fragment from 3-527; position 1 is the first nucleotide of the start codon); DevPF1 silencing region b (733 bp fragment from 532-1264); DevPF2 silencing region a (525 bp fragment from 3-527); DevPF2 silencing region b (731 bp fragment from 532-1262). Co-silencing was predicted with the RNAi off-target tool from ParameciumDB (Heng Li & Durbin, 2009) for both silencing regions (*DevPF1* silencing region a and b: 19 and 30 hits, respectively, in *DevPF2* gene; *DevPF2* silencing region a and b: 19 and 30 hits, respectively, in *DevPF1* gene). The plasmids were transformed into HT1115 (DE3) *E. coli* strain and expression was induced overnight at 30°C with Isopropyl ß-D-1-thiogalactopyranoside (IPTG; Carl Roth, Germany). Paramecia were seeded into the silencing medium at a density of 100 cells/ml to induce sexual development by starvation after 4 to 6 divisions. KD experiments were performed as previously described (Beisson et al., 2010e).

After the paramecia finished sexual development, 15 cells were transferred into a regular, non-induced, feeding medium for the survival test. Paramecia were monitored for three days to observe growth effects. For IES retention PCRs, genomic DNA was extracted from cultures that finished sexual development using GeneElute – Mammalian Genomic DNA Miniprep Kit (Merck-Sigma, Germany). PCRs were done on specific genomic regions flanking an IES (Table S6) to check for the retention of IESs. 1-12.5 ng DNA was used as input andPCR products were resolved on 1-2% agarose gels.

Time course silencing experiments

The time course experiments were conducted in three batches, each processing two KD replicates in parallel (batch A: replicates 1 and 2 of *ND7*-, *DevPF1*- and *DevPF2*-KD; batch B: replicates 3 and 4 of *ND7*-, *DevPF1*- and *DevPF2*-KD; batch C: replicates 5 and 6 of *ND7*- and *DevPF2*-KD). In batch A and B, cells were collected as soon as the first meiotic cells were observed in the population (onset), between 20 to 40% fragmentation (early), at 80-90% fragmentation (late) and 6 h after the late time point (very late). In batch C, cells were collected before the onset of autogamy (vegetative), at 50% fragmentation (early), at 100% fragmentation + visible anlagen (very late) and 6 h

later (very late + 6h). Since batch C was collected at different stages, only the “very late” time point of batch C was considered for differential expression analysis. For all time course replicates, enriched new MAC DNA was analyzed for IES retention and total RNA was collected from the collected time points for sRNA and/or mRNA analysis.

Macronuclear isolation and Illumina DNA-sequencing

Samples for new MAC isolation were collected from the KD cultures of all time course experiments three days after completion of sexual development as described previously (Arnaiz et al., 2012). DNA library preparation (350 bp fragment sizes) and Illumina sequencing (paired-end, 150 bp reads) were done at Novogene (UK) Company Limited, Cambridge according to their standard protocols.

IES retention and alternative boundary analysis

For IES retention score analysis, whole genome sequencing reads of enriched new MAC DNA after KD were adaptor trimmed using TrimGalore (Krueger, 2019) if significant Illumina adapter content was observed using FastQC v0.11.9 (Andrews, 2010) (see Table S7 for adapter sequences). The “Map” module of ParTIES v1.05 pipeline was used to map the reads on MAC and MAC+IES reference genomes with changes in the /lib/PARTIES/Map.pm file as described in (Singh et al., 2023). The IES retention scores (IRS) were calculated by the “MIRET” module (provided as DevPF_IRS.tab.gz). All scripts are available from https://github.com/Swart-lab/DevPF_code. IRS correlations were calculated as described previously (Swart et al., 2014).

Alternative excision was analyzed as described previously (Singh et al., 2023). In brief, properly paired and mapped reads were selected from the output from the ParTIES "Map" module for the MAC+IES reference genome and downsampled to the same library size (DevPF1-KD (1) and DevPF2-KD (2) were excluded due to small library

size). We then employed the "MILORD" module of a pre-release version of ParTIES (13 August 2015) with default parameters to annotate alternative and cryptic IES excision. All scripts are available from https://github.com/Swart-lab/DevPF_code.

The data generate in this study was compared with data of previously published KDs: *PGM*-KD (Arnaiz et al., 2012), *TFIIS4*-KD (Maliszewska-Olejniczak et al., 2015), *SPT5m*-KD (Gruchota et al., 2017), *PTCAF1*-KD (Ignarski et al., 2014), *DCL2/3/5*-KD (Sandoval et al., 2014), *KU80c*-KD (Abello et al., 2020), *EZL1*-KD (Lhuillier-Akakpo et al., 2014) and *ISWI1*-KD (Singh et al., 2022).

RNA extraction and sequencing

Total RNA was either extracted with phenol-chloroform followed by Monarch Total RNA Miniprep kit (New England Biolabs) or with the Quick-RNA Miniprep kit (Zymo). For phenol-chloroform extraction (batch C), 300 ml cells subjected to RNAi were washed twice with 10 mM Tris pH 7.5 (RT, 280 g, 2 min) and shock frozen by dropping them directly into liquid nitrogen. 500 µl of 2× DNA/RNA protection reagent from the Monarch kit were added to the frozen pellet and the cells thawed by vortexing. After adding 10 µl proteinase K and 1 ml RNA lysis buffer, the manufacturer's instructions (RNA Binding and Elution (Cultured Mammalian Cells)) were followed. On-column DNase I treatment was included.

For RNA extraction with Quick-RNA Miniprep kit (batch A and B), 100 ml of *Paramecium* cultures subjected to RNAi by feeding were washed twice in 10 mM Tris pH 7.5 in pear-shaped oil flasks by centrifugation (RT, 280 g, 2 min). After the final wash, cells were collected on ice and spun at 2,000 g for 2 min and 4°C and as much liquid as possible was removed. 3× volume of 1× DNA/RNA Shield (Biozym) was mixed with the cells and the samples were stored at -70°C until further processing. For RNA extraction, samples were thawed at RT and mixed with 1× volume of RNA lysis buffer. The manufacturer’s instructions were followed (section: (III) Total RNA Purification).

Extracted total RNA was send to Azenta Life Sciences for library preparation (sRNA: NEBNext Small RNA Library Prep Set for Illumina; mRNA: NEBNext Ultra II RNA Library Prep Kit for Illumina) and paired-end Illumina sequencing (NovaSeq 2×150bp).

Small RNA analysis

Small RNA sequencing reads were trimmed using cutadapt (Martin, 2011) version 3.2 with the parameter -a "AGATCGGAAGAGCACACGTCTGAACTCCAGTCA" to remove the relevant Illumina adaptor sequence. Trimmed reads were mapped to the *Paramecium tetraurelia* strain 51 MAC + IES genome and L4440 (*ND7*-KD) or T444T (*DevPF1*/*DevPF2*-KD) silencing vector with bwa version 0.7.17-r1188 (Heng Li & Durbin, 2009). GNU grep (version 2.14) was used to select 10-49 bp long, uniquely mapped reads (possessing the SAM file format flags "XT:A:U") and sRNA length histograms were generated by a Python script. All scripts are available from https://github.com/Swart-lab/DevPF_code.

mRNA analysis

Illumina adapter sequences (Table S7) were trimmed from reads with TrimGalore (Krueger, 2019). Reads were mapped to the *Paramecium tetraurelia* strain 51 transcriptome with hisat2 (Kim et al., 2019) allowing 20 multimappings (-k 20). Using samtools (Heng Li et al., 2009), the properly paired and mapped reads were filtered (-f2 flag) and sorted by the read name (-n flag). Unique mapping reads were acquired with eXpress (Roberts & Pachter, 2013) with 5 additional online expectation-maximization rounds to perform on the data after the initial online round (-O 5 flag) to improve accuracy. Scripts are available from https://github.com/Swart-lab/DevPF_code.

Read counts were normalized with DEseq2 (Love et al., 2014) package in R (version 3.6.3). For plotting, DEseq2 in-build functions plotPCA, plotMA and plotCounts were combined with ggplot2 (Villanueva & Chen, 2019) package (version 3.4.3). Differentially

expressed genes were identified for each time point with a Wald test (false discovery rate (alpha) = 0.1). Differentially expressed genes were filtered with an absolute log2(Fold Change) > 2 (corresponding to a 4-fold change) and an adjusted p-value < 0.01. The time point, KD and batch were known sources of variation in the dataset (design = ~ batch + timepoint + KD+ timepoint:KD). All scripts are available from https://github.com/Swart-lab/DevPF_code.

Structure prediction with AlphaFold

Protein structures were predicted with AlphaFold2 multimer (Evans et al., 2021; Jumper et al., 2021) using the ColabFold v1.5.2-patch (Mirdita et al., 2022) in Google Colab with default parameters.

Sequence alignment

Domains were predicted using InterProScan (Paysan-Lafosse et al., 2023). The nucleotide sequence of DevPF2 and DevPF1 (including introns) were aligned with clustalOmega (Sievers et al., 2011) (version 1.2.3) pairwise sequence alignment tool in Geneious prime (version 2023.2.1) with default parameters (Fig 4A).

Multiple sequence alignment of PHD domains was done with clustalOmega (version 1.2.1) using the MPI bioinformatics toolkit's web interface (Zimmermann et al., 2018) with default parameters.

Manuscript writing

Grammar and language refinement were assisted by an AI language model developed by OpenAI (GPT-3.5 architecture) (OpenAI, 2023).

Acknowledgements

We thank the BioOptics core facility and Genome center of MPI for Biology (Tübingen, Germany) for their assistance and Andre Noll for computer system administration.

Competing interests

The authors declare no competing interests.

Funding

This work was funded by the Max Planck Society.

Data availability

Supplementary files, including uncropped blot images, microcopy raw files and IES retention scores have been deposited to the open research data repository of the Max Planck Society EDMOND (https://doi.org/10.17617/3.VKJBJ0). Sequencing raw files have been deposited to the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena/browser/home) (Leinonen et al., 2011) (accession number: PRJEB67678). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Deutsch et al., 2023) via the PRIDE (Perez-Riverol et al., 2022) partner repository (accession number: PXD046704).

References

Aasland, R., Gibson, T. J., & Stewart, A. F. (1995). The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends in Biochemical Sciences*, *20*(2), 56–59. https://doi.org/10.1016/s0968-0004(00)88957-4

Abello, A., Régnier, V., Arnaiz, O., Le Bars, R., Bétermier, M., & Bischerour, J. (2020). Functional diversification of Paramecium Ku80 paralogs safeguards genome integrity during precise programmed DNA elimination. *PLoS Genetics*, *16*(4), e1008723. https://doi.org/10.1371/journal.pgen.1008723

Allen, S. E., Hug, I., Pabian, S., Rzeszutek, I., Hoehener, C., & Nowacki, M. (2017). Circular Concatemers of Ultra-Short DNA Segments Produce Regulatory RNAs. *Cell*, *168*(6), 990-999.e7. https://doi.org/10.1016/j.cell.2017.02.020

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data.* (0.11.9) [Computer software]. Babraham Bioinformatics.

Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.-M., Denby Wilkes, C., Garnier, O., Labadie, K., Lauderdale, B. E., Le Mouël, A., Marmignon, A., Nowacki, M., Poulain, J., Prajer, M., Wincker, P., Meyer, E., Duharcourt, S., Duret, L., Bétermier, M., & Sperling, L. (2012). The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genetics*, *8*(10), e1002984. https://doi.org/10.1371/journal.pgen.1002984

Arnaiz, O., Van Dijk, E., Bétermier, M., Lhuillier-Akakpo, M., de Vanssay, A., Duharcourt, S., Sallet, E., Gouzy, J., & Sperling, L. (2017). Improved methods and resources for *Paramecium* genomics: transcription units, gene annotation and gene expression. *BMC Genomics*, *18*(1), 483. https://doi.org/10.1186/s12864-017-3887-z

Bassing, C. H., Swat, W., & Alt, F. W. (2002). The mechanism and regulation of chromosomal V(D)J recombination. *Cell*, *109 Suppl*, S45-55. https://doi.org/10.1016/s0092-8674(02)00675-x

Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E., & Bétermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. *Genes & Development*, *23*(21), 2478–2483. https://doi.org/10.1101/gad.547309

Bazin-Gélis, M., Eleftheriou, E., Zangarelli, C., Lelandais, G., Sperling, L., Arnaiz, O., & Bétermier, M. (2023). Inter-generational nuclear crosstalk links the control of gene expression to programmed genome rearrangements during the *Paramecium* sexual cycle. *BioRxiv*. https://doi.org/10.1101/2023.04.16.537068

Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J. R., & Sperling, L. (2010a). DNA microinjection into the macronucleus of paramecium. *Cold Spring Harbor Protocols*, *2010*(1), pdb.prot5364. https://doi.org/10.1101/pdb.prot5364

Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J. R., & Sperling, L. (2010b). Maintaining clonal *Paramecium tetraurelia* cell lines of controlled age through daily reisolation. *Cold Spring Harbor Protocols*, *2010*(1), pdb.prot5361. https://doi.org/10.1101/pdb.prot5361

Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J. R., & Sperling, L. (2010c). Mass culture of *Paramecium tetraurelia*. *Cold Spring Harbor Protocols*, *2010*(1), pdb.prot5362. https://doi.org/10.1101/pdb.prot5362

Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J. R., & Sperling, L. (2010d). Paramecium tetraurelia: the renaissance of an early unicellular model. *Cold Spring Harbor Protocols*, *2010*(1), pdb.emo140. https://doi.org/10.1101/pdb.emo140

Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J. R., & Sperling, L. (2010e). Silencing specific

Paramecium tetraurelia genes by feeding double-stranded RNA. *Cold Spring Harbor Protocols*, *2010*(1), pdb.prot5363. https://doi.org/10.1101/pdb.prot5363

Berger, J. D. (1973). Nuclear differentiation and nucleic acid synthesis in well-fed exconjugants of Paramecium aurelia. *Chromosoma*, *42*(3), 247–268. https://doi.org/10.1007/BF00284774

Betz, C., Schlenstedt, G., & Bailer, S. M. (2004). Asr1p, a novel yeast ring/PHD finger protein, signals alcohol stress to the nucleus. *The Journal of Biological Chemistry*, *279*(27), 28174–28181. https://doi.org/10.1074/jbc.M401595200

Bienz, M. (2006). The PHD finger, a nuclear protein-interaction domain. *Trends in Biochemical Sciences*, *31*(1), 35–40. https://doi.org/10.1016/j.tibs.2005.11.001

Black, J. C., & Kutateladze, T. G. (2023). Atypical histone targets of PHD fingers. *The Journal of Biological Chemistry*, *299*(4), 104601. https://doi.org/10.1016/j.jbc.2023.104601

Bouhouche, K., Gout, J.-F., Kapusta, A., Bétermier, M., & Meyer, E. (2011). Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Research*, *39*(10), 4249–4264. https://doi.org/10.1093/nar/gkq1283

Bozhenok, L., Wade, P. A., & Varga-Weisz, P. (2002). WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. *The EMBO Journal*, *21*(9), 2231–2241. https://doi.org/10.1093/emboj/21.9.2231

Chalker, D. L., & Yao, M. C. (2001). Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila. *Genes & Development*, *15*(10), 1287–1298. https://doi.org/10.1101/gad.884601

Chalker, D. L. (2012). Transformation and strain engineering of Tetrahymena. *Methods in Cell Biology*, *109*, 327–345. https://doi.org/10.1016/B978-0-12-385967-9.00011-6

983 Charmant, O., Gruchota, J., Arnaiz, O., Zangarelli, C., Bétermier, M., Nowak, K.,
984 Legros, V., Chevreux, G., Nowak, J., & Duharcourt, S. (2023). The nuclear PIWI-
985 interacting protein Gtsf1 controls the selective degradation of small RNAs in
986 *Paramecium*. *BioRxiv*. https://doi.org/10.1101/2023.09.19.558372

987 Cheng, C.-Y., Vogt, A., Mochizuki, K., & Yao, M.-C. (2010). A domesticated piggyBac
988 transposase plays key roles in heterochromatin dynamics and DNA cleavage during
989 programmed DNA deletion in Tetrahymena thermophila. *Molecular Biology of the*
990 *Cell*, *21*(10), 1753–1762. https://doi.org/10.1091/mbc.e09-12-1079

991 de Vanssay, A., Touzeau, A., Arnaiz, O., Frapporti, A., Phipps, J., & Duharcourt, S.
992 (2020). The Paramecium histone chaperone Spt16-1 is required for Pgm
993 endonuclease function in programmed genome rearrangements. *PLoS Genetics*,
994 *16*(7), e1008949. https://doi.org/10.1371/journal.pgen.1008949

995 Deutsch, E. W., Bandeira, N., Perez-Riverol, Y., Sharma, V., Carver, J. J., Mendoza, L.,
996 Kundu, D. J., Wang, S., Bandla, C., Kamatchinathan, S., Hewapathirana, S., Pullman,
997 B. S., Wertz, J., Sun, Z., Kawano, S., Okuda, S., Watanabe, Y., MacLean, B.,
998 MacCoss, M. J., … Vizcaíno, J. A. (2023). The ProteomeXchange consortium at 10
999 years: 2023 update. *Nucleic Acids Research*, *51*(D1), D1539–D1548.
1000 https://doi.org/10.1093/nar/gkac1040

1001 Driever, W., & Nüsslein-Volhard, C. (1988). A gradient of bicoid protein in Drosophila
1002 embryos. *Cell*, *54*(1), 83–93. https://doi.org/10.1016/0092-8674(88)90182-1

1003 Eberharter, A., Vetter, I., Ferreira, R., & Becker, P. B. (2004). ACF1 improves the
1004 effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts. *The*
1005 *EMBO Journal*, *23*(20), 4029–4039. https://doi.org/10.1038/sj.emboj.7600382

1006 Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A. W., Green, T., Žídek, A.,
1007 Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M.,
1008 Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., …

1009 Hassabis, D. (2021). Protein complex prediction with AlphaFold-Multimer. *BioRxiv*.
1010 https://doi.org/10.1101/2021.10.04.463034

1011 Finn, R., Griffiths-Jones, S., & Bateman, A. (2003). Identifying protein domains with the
1012 Pfam database. *Current Protocols in Bioinformatics*, *Chapter 2*, Unit 2.5.
1013 https://doi.org/10.1002/0471250953.bi0205s01

1014 Fleck, K., Nitz, M., & Jeffers, V. (2021). “Reading” a new chapter in protozoan parasite
1015 transcriptional regulation. *PLoS Pathogens*, *17*(12), e1010056.
1016 https://doi.org/10.1371/journal.ppat.1010056

1017 Forment, J. V., Kaidi, A., & Jackson, S. P. (2012). Chromothripsis and cancer: causes
1018 and consequences of chromosome shattering. *Nature Reviews. Cancer*, *12*(10), 663–
1019 670. https://doi.org/10.1038/nrc3352

1020 Frapporti, A., Miró Pina, C., Arnaiz, O., Holoch, D., Kawaguchi, T., Humbert, A.,
1021 Eleftheriou, E., Lombard, B., Loew, D., Sperling, L., Guitot, K., Margueron, R., &
1022 Duharcourt, S. (2019). The Polycomb protein Ezl1 mediates H3K9 and H3K27
1023 methylation to repress transposable elements in Paramecium. *Nature*
1024 *Communications*, *10*(1), 2710. https://doi.org/10.1038/s41467-019-10648-5

1025 Furrer, D. I., Swart, E. C., Kraft, M. F., Sandoval, P. Y., & Nowacki, M. (2017). Two Sets
1026 of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of
1027 Germline-Specific DNA. *Cell Reports*, *20*(2), 505–520.
1028 https://doi.org/10.1016/j.celrep.2017.06.050

1029 Gaurav, N., & Kutateladze, T. G. (2023). Non-histone binding functions of PHD fingers.
1030 *Trends in Biochemical Sciences*, *48*(7), 610–617.
1031 https://doi.org/10.1016/j.tibs.2023.03.005

1032 Gozani, O., Karuman, P., Jones, D. R., Ivanov, D., Cha, J., Lugovskoy, A. A., Baird, C.
1033 L., Zhu, H., Field, S. J., Lessnick, S. L., Villasenor, J., Mehrotra, B., Chen, J., Rao, V.
1034 R., Brugge, J. S., Ferguson, C. G., Payrastre, B., Myszka, D. G., Cantley, L. C., …

1035 Yuan, J. (2003). The PHD finger of the chromatin-associated protein ING2 functions
1036 as a nuclear phosphoinositide receptor. *Cell*, *114*(1), 99–111.
1037 https://doi.org/10.1016/s0092-8674(03)00480-x

1038 Grandchamp, S., & Beisson, J. (1981). Positional control of nuclear differentiation in
1039 paramecium. *Developmental Biology*, *81*(2), 336–341. https://doi.org/10.1016/0012-
1040 1606(81)90297-9

1041 Gruchota, J., Denby Wilkes, C., Arnaiz, O., Sperling, L., & Nowak, J. K. (2017). A
1042 meiosis-specific Spt5 homolog involved in non-coding transcription. *Nucleic Acids*
1043 *Research*, *45*(8), 4722–4732. https://doi.org/10.1093/nar/gkw1318

1044 Guérin, F., Arnaiz, O., Boggetto, N., Denby Wilkes, C., Meyer, E., Sperling, L., &
1045 Duharcourt, S. (2017). Flow cytometry sorting of nuclei enables the first global
1046 characterization of *Paramecium* germline DNA and transposable elements. *BMC*
1047 *Genomics*, *18*(1), 327. https://doi.org/10.1186/s12864-017-3713-7

1048 Hamilton, E. P., Kapusta, A., Huvos, P. E., Bidwell, S. L., Zafar, N., Tang, H.,
1049 Hadjithomas, M., Krishnakumar, V., Badger, J. H., Caler, E. V., Russ, C., Zeng, Q.,
1050 Fan, L., Levin, J. Z., Shea, T., Young, S. K., Hegarty, R., Daza, R., Gujja, S., …
1051 Coyne, R. S. (2016). Structure of the germline genome of *Tetrahymena thermophila*
1052 and relationship to the massively rearranged somatic genome. *ELife*, *5*.
1053 https://doi.org/10.7554/eLife.19090

1054 Hoehener, C., Hug, I., & Nowacki, M. (2018). Dicer-like enzymes with sequence
1055 cleavage preferences. *Cell*, *173*(1), 234-247.e7.
1056 https://doi.org/10.1016/j.cell.2018.02.029

1057 Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., & Vingron, M. (2002).
1058 Variance stabilization applied to microarray data calibration and to the quantification
1059 of differential expression. *Bioinformatics*, *18 Suppl 1*, S96-104.
1060 https://doi.org/10.1093/bioinformatics/18.suppl_1.s96

Ignarski, M., Singh, A., Swart, E. C., Arambasic, M., Sandoval, P. Y., & Nowacki, M. (2014). Paramecium tetraurelia chromatin assembly factor-1-like protein PtCAF-1 is involved in RNA-mediated control of DNA elimination. *Nucleic Acids Research*, *42*(19), 11952–11964. https://doi.org/10.1093/nar/gku874

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kinkelin, K., Wozniak, G. G., Rothbart, S. B., Lidschreiber, M., Strahl, B. D., & Cramer, P. (2013). Structures of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15277–15282. https://doi.org/10.1073/pnas.1311010110

Klobutcher, L. A., & Herrick, G. (1995). Consensus inverted terminal repeat sequence of *Paramecium* IESs: resemblance to termini of Tc1-related and *Euplotes* Tec transposons. *Nucleic Acids Research*, *23*(11), 2006–2013. https://doi.org/10.1093/nar/23.11.2006

Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., & Nesvizhskii, A. I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature Methods*, *14*(5), 513–520. https://doi.org/10.1038/nmeth.4256

Krueger, F. (2019). *TrimGalore: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data* (0.6.10) [Computer software]. GitHub.

Lefort-Tran, M., Aufderheide, K., Pouphile, M., Rossignol, M., & Beisson, J. (1981). Control of exocytotic processes: cytological and physiological studies of trichocyst mutants in Paramecium tetraurelia. *The Journal of Cell Biology*, *88*(2), 301–311. https://doi.org/10.1083/jcb.88.2.301

Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdeno-Tárraga, A., Cheng, Y., Cleland, I., Faruque, N., Goodgame, N., Gibson, R., Hoad, G., Jang, M., Pakseresht, N., Plaister, S., Radhakrishnan, R., Reddy, K., Sobhany, S., Ten Hoopen, P., Vaughan, R., … Cochrane, G. (2011). The european nucleotide archive. *Nucleic Acids Research*, *39*(Database issue), D28-31. https://doi.org/10.1093/nar/gkq967

Lepère, G., Nowacki, M., Serrano, V., Gout, J.-F., Guglielmi, G., Duharcourt, S., & Meyer, E. (2009). Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Research*, *37*(3), 903–915. https://doi.org/10.1093/nar/gkn1018

Lhuillier-Akakpo, M., Frapporti, A., Denby Wilkes, C., Matelot, M., Vervoort, M., Sperling, L., & Duharcourt, S. (2014). Local effect of enhancer of zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements. *PLoS Genetics*, *10*(9), e1004665. https://doi.org/10.1371/journal.pgen.1004665

Liu, L., Qin, S., Zhang, J., Ji, P., Shi, Y., & Wu, J. (2012). Solution structure of an atypical PHD finger in BRPF2 and its interaction with DNA. *Journal of Structural Biology*, *180*(1), 165–173. https://doi.org/10.1016/j.jsb.2012.06.014

Liu, Y., Taverna, S. D., Muratore, T. L., Shabanowitz, J., Hunt, D. F., & Allis, C. D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation

and DNA elimination in Tetrahymena. *Genes & Development*, *21*(12), 1530–1545. https://doi.org/10.1101/gad.1544207

Li, Haitao, Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D., & Patel, D. J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature*, *442*(7098), 91–95. https://doi.org/10.1038/nature04802

Li, Heng, & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, Heng, Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, X., Meng, X., Morris, C. A., & Keating, M. T. (1998). A novel human gene, WSTF, is deleted in Williams syndrome. *Genomics*, *54*(2), 241–249. https://doi.org/10.1006/geno.1998.5578

Maliszewska-Olejniczak, K., Gruchota, J., Gromadka, R., Denby Wilkes, C., Arnaiz, O., Mathy, N., Duharcourt, S., Bétermier, M., & Nowak, J. K. (2015). TFIIS-Dependent Non-coding Transcription Regulates Developmental Genome Rearrangements. *PLoS Genetics*, *11*(7), e1005383. https://doi.org/10.1371/journal.pgen.1005383

Mani, R.-S., & Chinnaiyan, A. M. (2010). Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. *Nature Reviews. Genetics*, *11*(12), 819–829. https://doi.org/10.1038/nrg2883

1137 Marmignon, A., Bischerour, J., Silve, A., Fojcik, C., Dubois, E., Arnaiz, O., Kapusta, A.,
1138 Malinsky, S., & Bétermier, M. (2014). Ku-mediated coupling of DNA cleavage and
1139 repair during programmed genome rearrangements in the ciliate Paramecium
1140 tetraurelia. *PLoS Genetics*, *10*(8), e1004552.
1141 https://doi.org/10.1371/journal.pgen.1004552

1142 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput
1143 sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

1144 Miller, J., McLachlan, A. D., & Klug, A. (1985). Repetitive zinc-binding domains in the
1145 protein transcription factor IIIA from Xenopus oocytes. *The EMBO Journal*, *4*(6),
1146 1609–1614. https://doi.org/10.1002/j.1460-2075.1985.tb03825.x

1147 Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M.
1148 (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, *19*(6),
1149 679–682. https://doi.org/10.1038/s41592-022-01488-1

1150 Miró-Pina, C., Charmant, O., Kawaguchi, T., Holoch, D., Michaud, A., Cohen, I.,
1151 Humbert, A., Jaszczyszyn, Y., Chevreux, G., Del Maestro, L., Ait-Si-Ali, S., Arnaiz, O.,
1152 Margueron, R., & Duharcourt, S. (2022). Paramecium Polycomb repressive complex
1153 2 physically interacts with the small RNA-binding PIWI protein to repress
1154 transposable elements. *Developmental Cell*, *57*(8), 1037-1052.e8.
1155 https://doi.org/10.1016/j.devcel.2022.03.014

1156 Mochizuki, K., Fine, N. A., Fujisawa, T., & Gorovsky, M. A. (2002). Analysis of a piwi-
1157 related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell*,
1158 *110*(6), 689–699. https://doi.org/10.1016/s0092-8674(02)00909-1

1159 Morisato, D., & Anderson, K. V. (1995). Signaling pathways that establish the dorsal-
1160 ventral pattern of the Drosophila embryo. *Annual Review of Genetics*, *29*, 371–399.
1161 https://doi.org/10.1146/annurev.ge.29.120195.002103

Nowacki, M., Zagorski-Ostoja, W., & Meyer, E. (2005). Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in Paramecium tetraurelia. *Current Biology*, *15*(18), 1616–1628. https://doi.org/10.1016/j.cub.2005.07.033

OpenAI. (2023). ChatGPT *[Large language model]* (GPT-3.5) [Computer software]. https://chat.openai.com/chat.

Oppikofer, M., Sagolla, M., Haley, B., Zhang, H.-M., Kummerfeld, S. K., Sudhamsu, J., Flynn, E. M., Bai, T., Zhang, J., Ciferri, C., & Cochran, A. G. (2017). Non-canonical reader modules of BAZ1A promote recovery from DNA damage. *Nature Communications*, *8*(1), 862. https://doi.org/10.1038/s41467-017-00866-0

Pascual, J., Martinez-Yamout, M., Dyson, H. J., & Wright, P. E. (2000). Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *Journal of Molecular Biology*, *304*(5), 723–729. https://doi.org/10.1006/jmbi.2000.4308

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., … Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, *51*(D1), D418–D427. https://doi.org/10.1093/nar/gkac993

Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D. J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., Walzer, M., Wang, S., Brazma, A., & Vizcaíno, J. A. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research*, *50*(D1), D543–D552. https://doi.org/10.1093/nar/gkab1038

Pinskaya, M., Ghavi-Helm, Y., Mariotte-Labarre, S., Morillon, A., Soutourina, J., & Werner, M. (2014). PHD and TFIIS-Like domains of the Bye1 transcription factor determine its multivalent genomic distribution. *Plos One*, *9*(7), e102464. https://doi.org/10.1371/journal.pone.0102464

1189 Quan, J., & Tian, J. (2011). Circular polymerase extension cloning for high-throughput
1190 cloning of complex and combinatorial DNA libraries. *Nature Protocols*, *6*(2), 242–251.
1191 https://doi.org/10.1038/nprot.2010.181

1192 Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015).
1193 limma powers differential expression analyses for RNA-sequencing and microarray
1194 studies. *Nucleic Acids Research*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007

1195 Roberts, A., & Pachter, L. (2013). Streaming fragment assignment for real-time analysis
1196 of sequencing experiments. *Nature Methods*, *10*(1), 71–73.
1197 https://doi.org/10.1038/nmeth.2251

1198 Rooney, S., Chaudhuri, J., & Alt, F. W. (2004). The role of the non-homologous end-
1199 joining pathway in lymphocyte development. *Immunological Reviews*, *200*, 115–131.
1200 https://doi.org/10.1111/j.0105-2896.2004.00165.x

1201 Roth, S., Stein, D., & Nüsslein-Volhard, C. (1989). A gradient of nuclear localization of
1202 the dorsal protein determines dorsoventral pattern in the Drosophila embryo. *Cell*,
1203 *59*(6), 1189–1202. https://doi.org/10.1016/0092-8674(89)90774-5

1204 Rzeszutek, I., Swart, E. C., Pabian-Jewuła, S., Russo, A., & Nowacki, M. (2022). Early
1205 developmental, meiosis-specific proteins - Spo11, Msh4-1, and Msh5 - Affect
1206 subsequent genome reorganization in Paramecium tetraurelia. *Biochimica et*
1207 *Biophysica Acta. Molecular Cell Research*, *1869*(6), 119239.
1208 https://doi.org/10.1016/j.bbamcr.2022.119239

1209 Sanchez, R., & Zhou, M.-M. (2011). The PHD finger: a versatile epigenome reader.
1210 *Trends in Biochemical Sciences*, *36*(7), 364–372.
1211 https://doi.org/10.1016/j.tibs.2011.03.005

1212 Sandoval, P. Y., Swart, E. C., Arambasic, M., & Nowacki, M. (2014). Functional
1213 diversification of Dicer-like proteins and small RNAs required for genome sculpting.
1214 *Developmental Cell*, *28*(2), 174–188. https://doi.org/10.1016/j.devcel.2013.12.010

1215 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T.,
1216 Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J.,
1217 Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source
1218 platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682.
1219 https://doi.org/10.1038/nmeth.2019

1220 Schindler, U., Beckmann, H., & Cashmore, A. R. (1993). HAT3.1, a novel Arabidopsis
1221 homeodomain protein containing a conserved cysteine-rich region. *The Plant Journal*,
1222 *4*(1), 137–150. https://doi.org/10.1046/j.1365-313x.1993.04010137.x

1223 Sellis, D., Guérin, F., Arnaiz, O., Pett, W., Lerat, E., Boggetto, N., Krenek, S.,
1224 Berendonk, T., Couloux, A., Aury, J.-M., Labadie, K., Malinsky, S., Bhullar, S., Meyer,
1225 E., Sperling, L., Duret, L., & Duharcourt, S. (2021). Massive colonization of protein-
1226 coding exons by selfish genetic elements in *Paramecium* germline genomes. *PLoS*
1227 *Biology*, *19*(7), e3001309. https://doi.org/10.1371/journal.pbio.3001309

1228 Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R.,
1229 McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011).
1230 Fast, scalable generation of high-quality protein multiple sequence alignments using
1231 Clustal Omega. *Molecular Systems Biology*, *7*, 539.
1232 https://doi.org/10.1038/msb.2011.75

1233 Singh, A., Häußermann, L., Emmerich, C., Nischwitz, E., Seah, B. K., Butter, F. K. B.,
1234 Nowacki, M., & Swart, E. C. (2023). ISWI1 complex proteins facilitate developmental
1235 genome editing in Paramecium. *BioRxiv*. https://doi.org/10.1101/2023.08.09.552620

1236 Singh, A., Maurer-Alcalá, X. X., Solberg, T., Häußermann, L., Gisler, S., Ignarski, M.,
1237 Swart, E. C., & Nowacki, M. (2022). Chromatin remodeling is required for sRNA-
1238 guided DNA elimination in Paramecium. *The EMBO Journal*, *41*(22), e111839.
1239 https://doi.org/10.15252/embj.2022111839

1240 Skene, P. J., Henikoff, J. G., & Henikoff, S. (2018). Targeted in situ genome-wide
1241 profiling with high efficiency for low cell numbers. *Nature Protocols*, *13*(5), 1006–
1242 1019. https://doi.org/10.1038/nprot.2018.015

1243 Spirov, A., Fahmy, K., Schneider, M., Frei, E., Noll, M., & Baumgartner, S. (2009).
1244 Formation of the bicoid morphogen gradient: an mRNA gradient dictates the protein
1245 gradient. *Development*, *136*(4), 605–614. https://doi.org/10.1242/dev.031195

1246 Sturm, Á., Saskoi, É., Tibor, K., Weinhardt, N., & Vellai, T. (2018). Highly efficient RNAi
1247 and Cas9-based auto-cloning systems for C. elegans research. *Nucleic Acids*
1248 *Research*, *46*(17), e105. https://doi.org/10.1093/nar/gky516

1249 Swart, E. C., Wilkes, C. D., Sandoval, P. Y., Arambasic, M., Sperling, L., & Nowacki, M.
1250 (2014). Genome-wide analysis of genetic and epigenetic control of programmed DNA
1251 deletion. *Nucleic Acids Research*, *42*(14), 8970–8983.
1252 https://doi.org/10.1093/nar/gku619

1253 Taverna, S. D., Coyne, R. S., & Allis, C. D. (2002). Methylation of histone h3 at lysine 9
1254 targets programmed DNA elimination in tetrahymena. *Cell*, *110*(6), 701–711.
1255 https://doi.org/10.1016/s0092-8674(02)00941-8

1256 Villanueva, R. A. M., & Chen, Z. J. (2019). ggplot2: Elegant Graphics for Data Analysis
1257 (2nd ed.). *Measurement: Interdisciplinary Research and Perspectives*, *17*(3), 160–
1258 167. https://doi.org/10.1080/15366367.2019.1565254

1259 Wang, C., Lv, L., Solberg, T., Wen, Z., Zhang, H., & Gao, F. (2023). Conservation of the
1260 ancestral function of GTSF1 in transposon silencing in the unicellular eukaryote
1261 *Paramecium tetraurelia*. *BioRxiv*. https://doi.org/10.1101/2023.10.06.561219

1262 Wang, C., Solberg, T., Maurer-Alcalá, X. X., Swart, E. C., Gao, F., & Nowacki, M.
1263 (2022). A small RNA-guided PRC2 complex eliminates DNA as an extreme form of
1264 transposon silencing. *Cell Reports*, *40*(8), 111263.
1265 https://doi.org/10.1016/j.celrep.2022.111263

1266 Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J., Kauer, M.,
1267 Tackett, A. J., Chait, B. T., Badenhorst, P., Wu, C., & Allis, C. D. (2006). A PHD finger
1268 of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling.
1269 *Nature*, *442*(7098), 86–90. https://doi.org/10.1038/nature04815

1270 Zangarelli, C., Arnaiz, O., Bourge, M., Gorrichon, K., Jaszczyszyn, Y., Mathy, N.,
1271 Escoriza, L., Bétermier, M., & Régnier, V. (2022). Developmental timing of
1272 programmed DNA elimination in Paramecium tetraurelia recapitulates germline
1273 transposon evolutionary dynamics. *Genome Research*, *32*(11–12), 2028–2042.
1274 https://doi.org/10.1101/gr.277027.122

1275 Zeng, L., Zhang, Q., Li, S., Plotnikov, A. N., Walsh, M. J., & Zhou, M.-M. (2010).
1276 Mechanism and regulation of acetylated histone binding by the tandem PHD finger of
1277 DPF3b. *Nature*, *466*(7303), 258–262. https://doi.org/10.1038/nature09139

1278 Zhang, J., Yan, G., Tian, M., Ma, Y., Xiong, J., & Miao, W. (2018). A DP-like
1279 transcription factor protein interacts with E2fl1 to regulate meiosis in Tetrahymena
1280 thermophila. *Cell Cycle*, *17*(5), 634–642.
1281 https://doi.org/10.1080/15384101.2018.1431595

1282 Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F.,
1283 Söding, J., Lupas, A. N., & Alva, V. (2018). A Completely Reimplemented MPI
1284 Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular*
1285 *Biology*, *430*(15), 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007

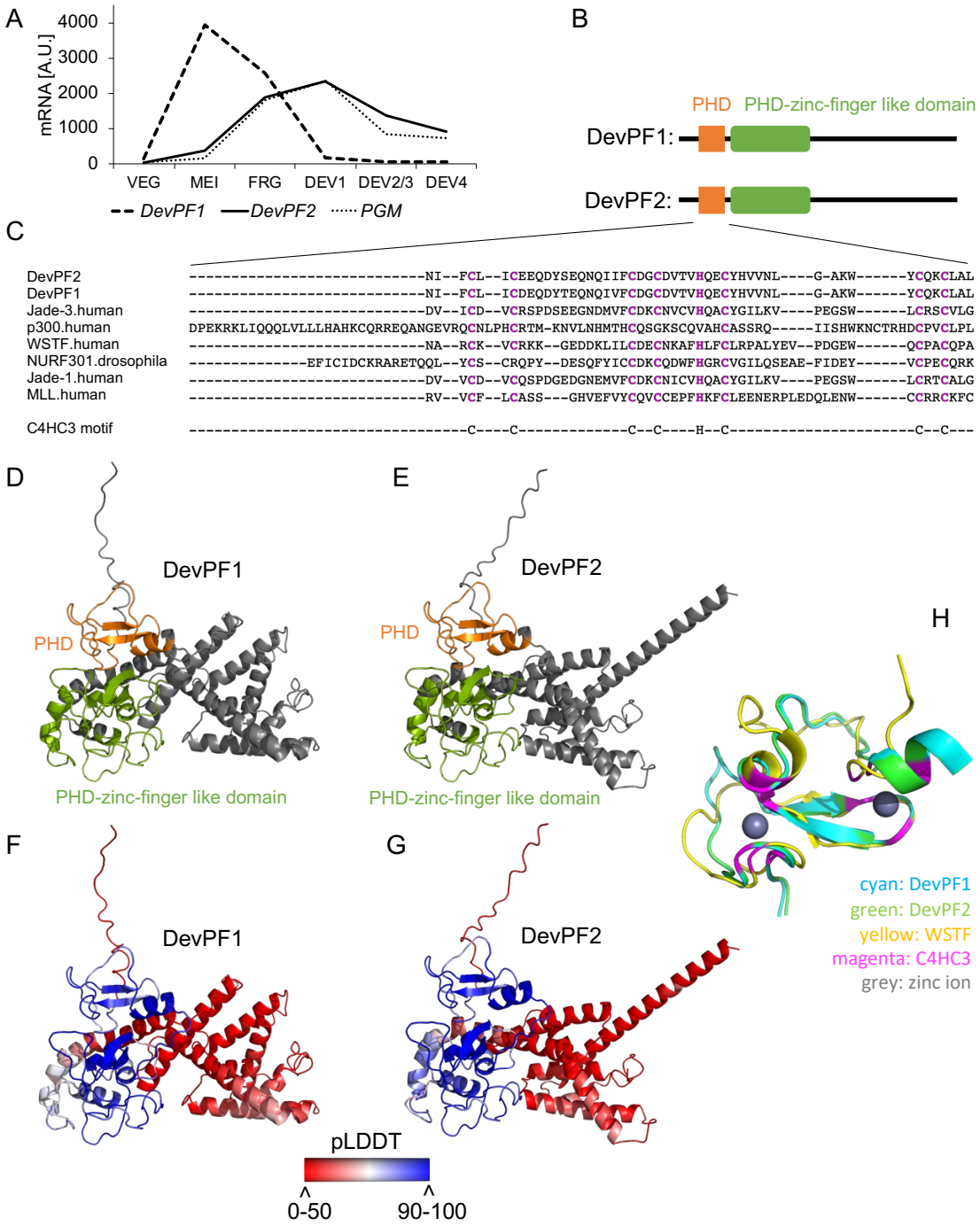


Figure 1: Features of the PHD finger proteins DevPF1 and DevPF2

(A) mRNA expression profiles for *DevPF1*, *DevPF2* and *PGM* during various developmental stages: VEG (vegetative growth), MEI (micronuclear meiosis and macronuclear fragmentation), FRG (~50% of the population with fragmented maternal MACs), DEV1 (significant proportion with visible anlagen), DEV2/3 (majority with visible

anlagen), DEV4 (majority with visible anlagen). Expression data retrieved from ParameciumDB (Arnaiz et al., 2017). (B) Schematic representation of predicted domain architecture for DevPF1 and DevPF2. (C) Multiple sequence alignment (Clustal Omega) of DevPF1 and DevPF2 amino acid sequence with PHD domains of published human and *Drosophila* PHD finger proteins. (D) to (F): Predicted protein structure (AlphaFold2) for DevPF1 and DevPF2, colored by domain (PHD: orange; PHD-zinc-finger-like domain: green) in (D) and (E), and by prediction confidence (pLDDT: predicted local distance difference test) in (F) and (G). (H) Structure predictions of DevPF1 and DevPF2 PHD domain superimposed with NMR structure of WSTF PHD domain (PDB accession number 1F62).

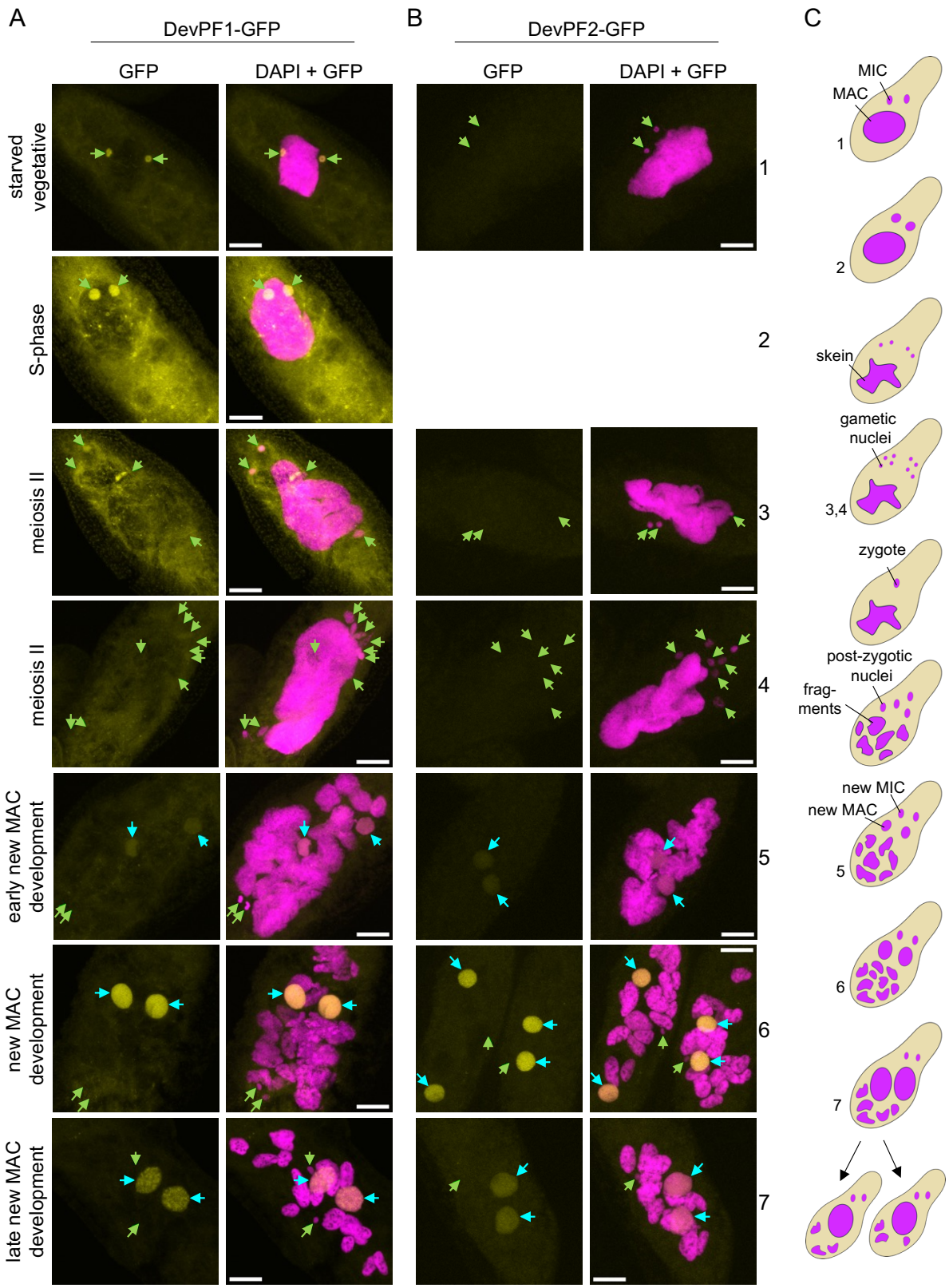


Figure 2: Subcellular localization of DevPF-GFP proteins

DevPF1-GFP (A) and DevPF2-GFP (B) localization at various developmental stages. DNA (stained with DAPI) in magenta. GFP signal in yellow. No image of DevPF2-GFP during S-phase was acquired. Green arrow: MIC. Cyan arrow: new MAC. Maximum intensity projections of multiple z-planes. Scale bar = 10 µm. (C) Schematic overview of nuclear morphology during sexual development, with corresponding cell stages in the images indicated by numbers.

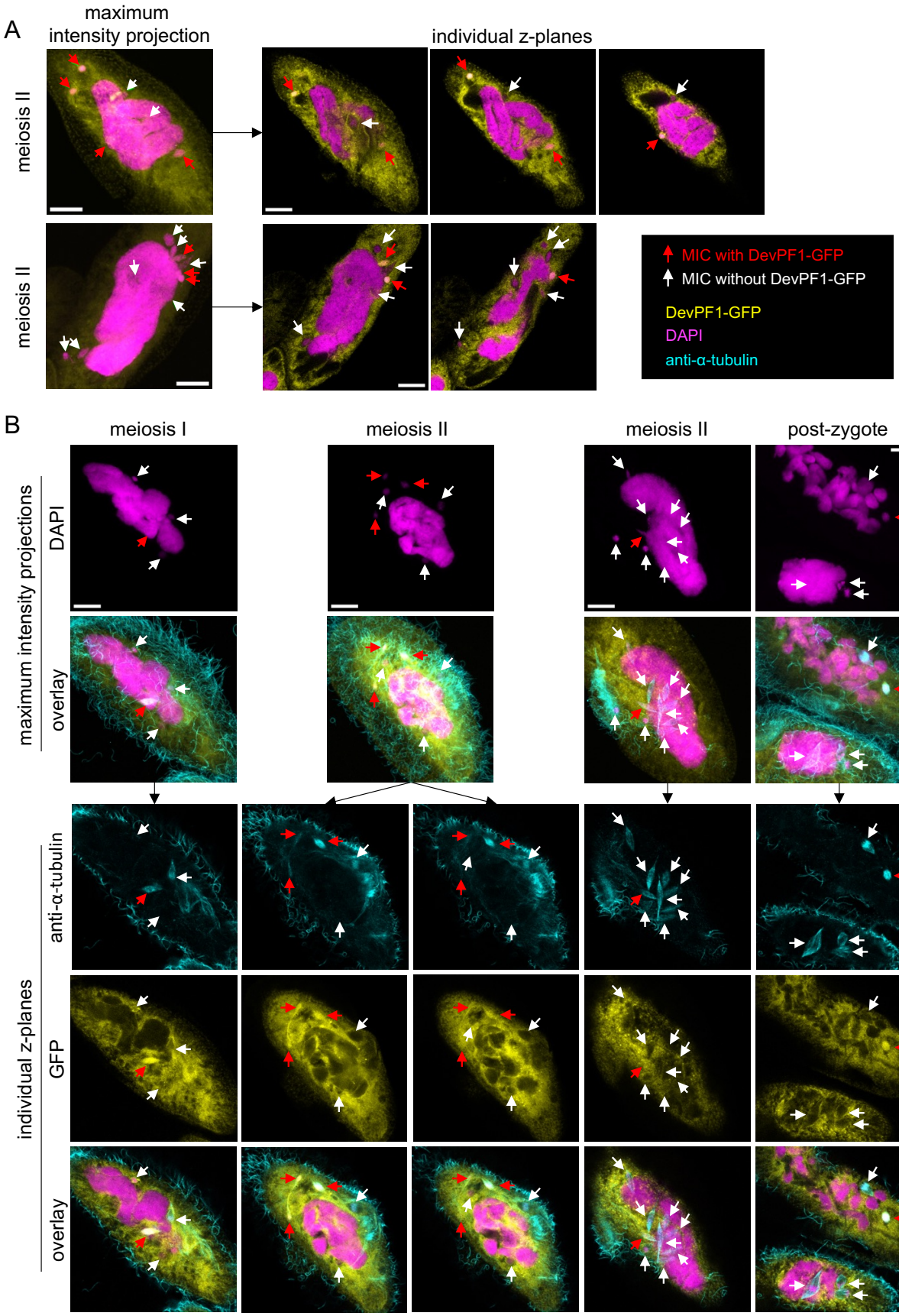


Figure 3: Selective DevPF1-GFP localization in *Paramecium* MICs

(A) Overlay of DAPI (DNA stain; pink) and GFP (yellow) signal in two DevPF1-GFP injected *Paramecium* cells during meiotic stages. Maximum intensity projections (left) and individual z-planes of the same stack (right). (B) DevPF1-GFP localization with visualization of nuclear spindle. DAPI (pink), GFP (yellow) and anti-α-tubulin staining (cyan). Maximum intensity projections (top) for DAPI and overlay (DAPI, GFP and anti-α-tubulin). Individual z-planes of the same stacks (bottom) for anti-α-tubulin, GFP and overlay. (A) and (B): Red arrows: MICs with DevPF1-GFP localization; White arrows: MICs without DevPF1-GFP localization. Scale bar = 10 μm.

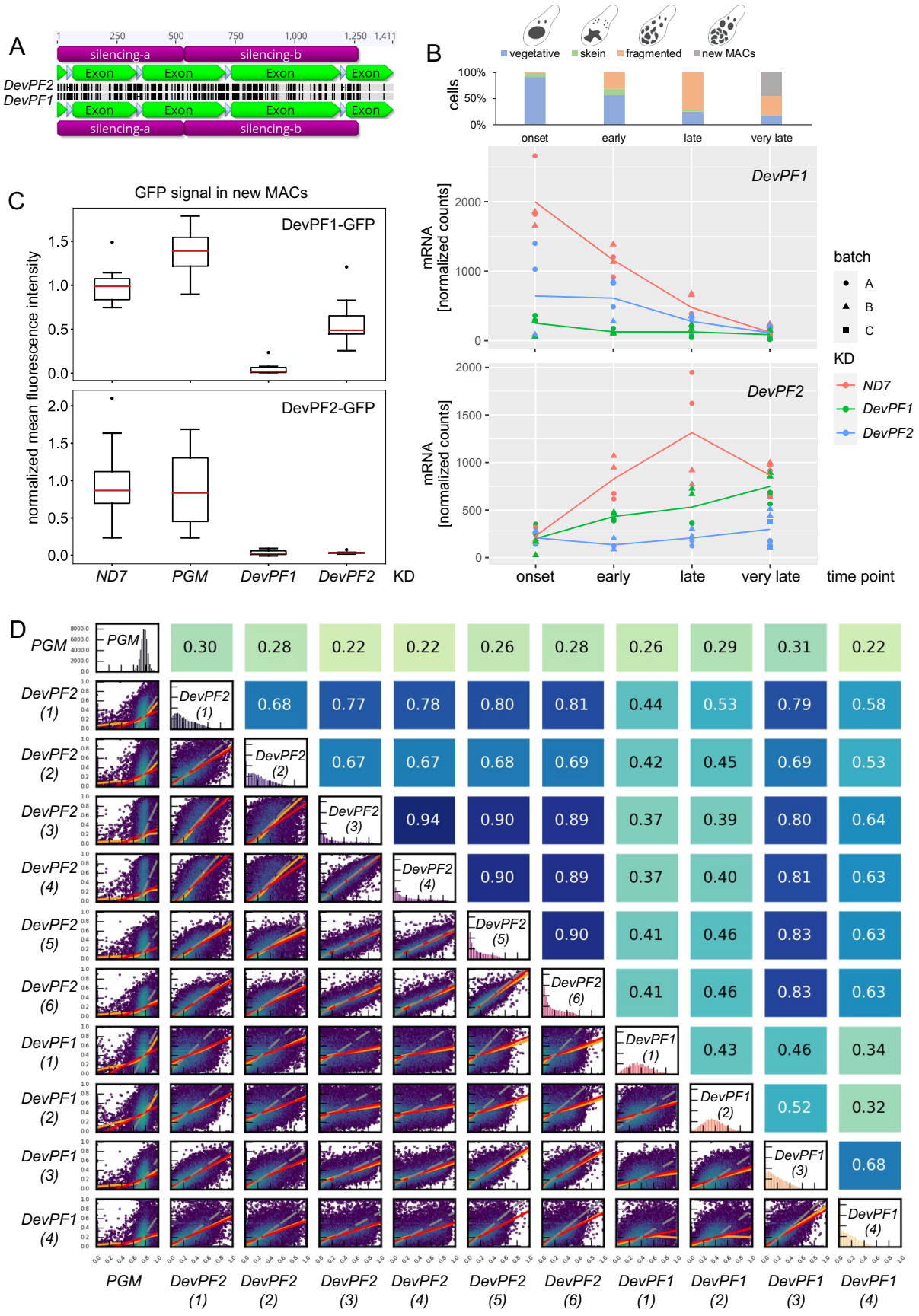
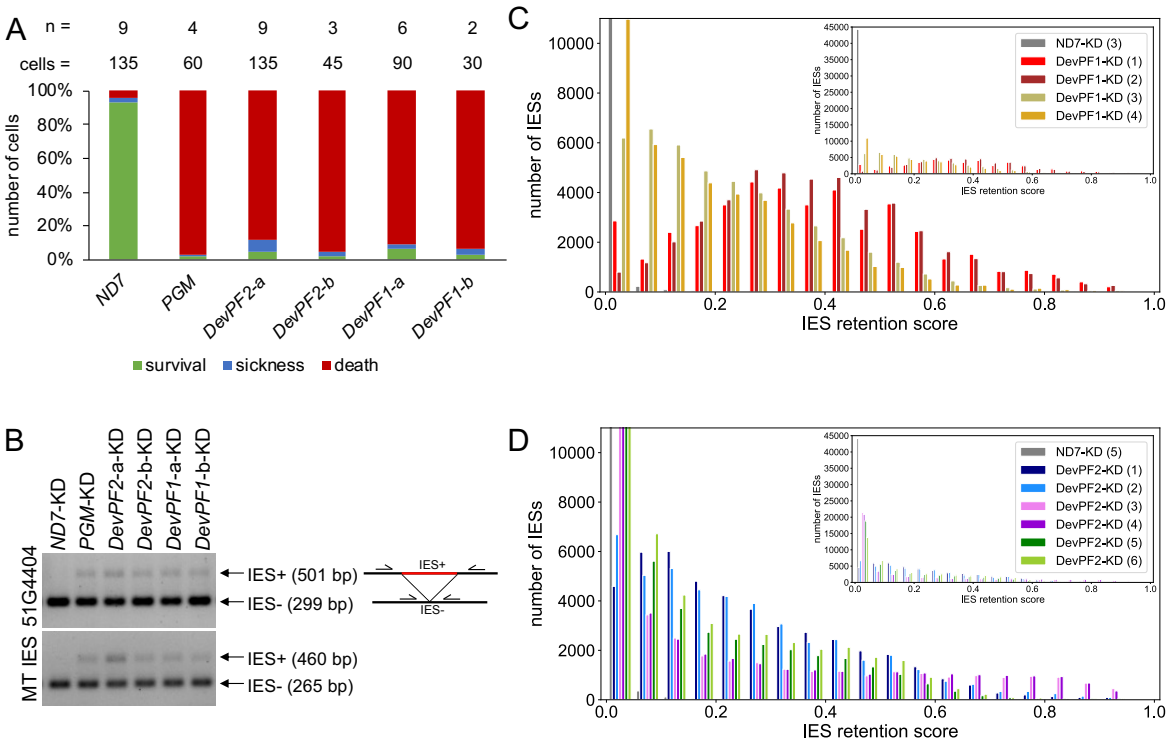


Figure 4: Co-silencing effects observed in *DevPF* knockdowns

(A) Nucleotide identity across *DevPF1* (bottom) and *DevPF2* (top) genes. Screenshot of pairwise sequence alignment in Geneious prime software. Silencing region (violet), exon (green), intron (white), perfect identity (gray) and mismatch/gap (black). Scale in base pairs at the top. (B) mRNA expression levels of *DevPF1* (top) and *DevPF2* (bottom) upon KDs (*ND7* (control), *DevPF1* and *DevPF2*) at different developmental time points (onset, early, late and very late). Lines represent the mean of all replicates for a given KD and time point. The cell stage composition of each time point averaged over all KDs is shown at the top (individual compositions in Fig. S5), along with schematic representations of the considered cell stages. (C) Protein expression upon KD: fluorescence intensities of DevPF1-GFP (top) and DevPF2-GFP (bottom). Red line: median. Whiskers: 1.5 times the interquartile range from the lower or upper quartile. Dots: data points outside the whiskers. Sample size = 10. (D) IES retention score (IRS) correlations between *DevPF1*- and *DevPF2*-KD replicates. Diagonal: IRS distributions of individual KDs. Below diagonal: correlation graphs of pairwise comparisons. Above diagonal: corresponding Spearman correlation coefficients. Red lines: ordinary least-squares (OLS) regression, orange lines: LOWESS, and gray lines: orthogonal distance regression (ODR).



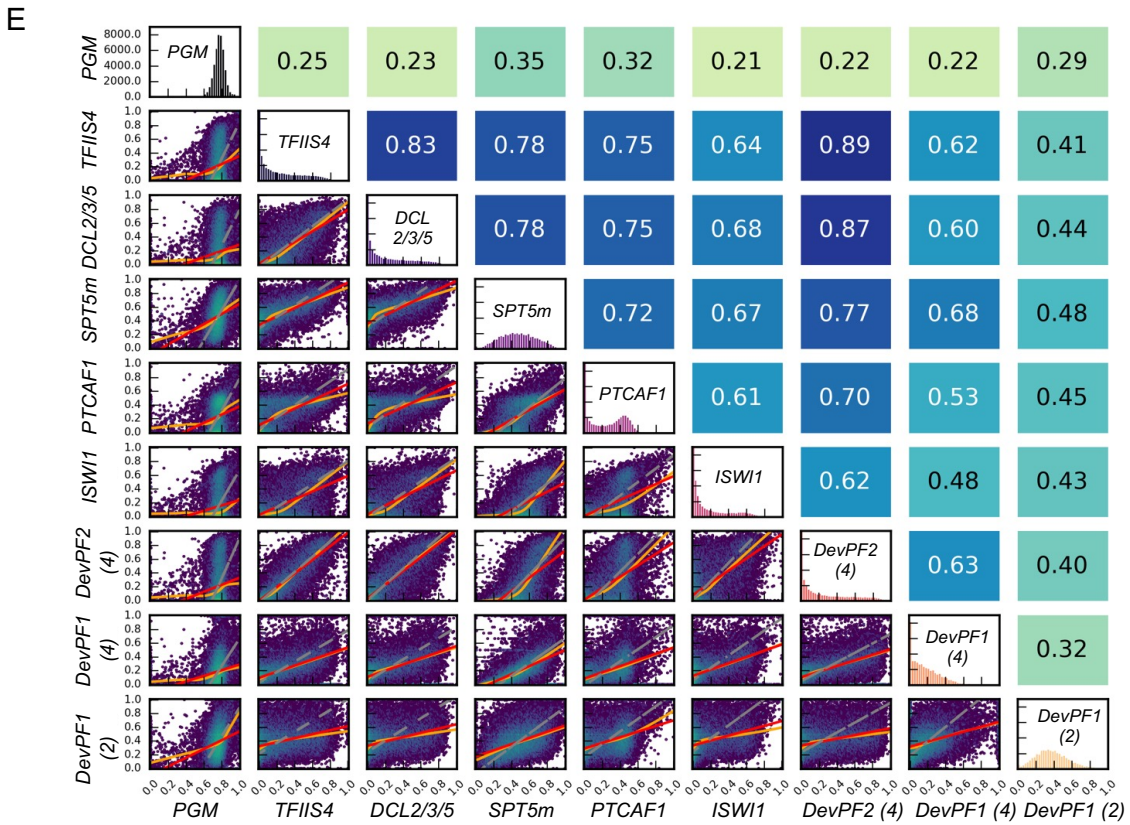


Figure 5: Effects of *DevPF* knockdowns on genome-wide IES retention

(A) Viability of new progeny after KDs (*ND7* (negative control), *PGM* (positive control), *DevPF1* and *DevPF2*) during sexual development. For *DevPF1* and *DevPF2*, two silencing regions were targeted (a and b, see Fig. 4A). The numbers of experiments (n) and cells counted (cells) are indicated at the top. Survival: normal division. Sickness: reduced growth. Death: 3 or less cells after three days. (B) IES retention PCRs for two IESs on genomic DNA isolated from KD cells. (C) and (D): IES retention score (IRS) histograms for *DevPF1* (C) and *DevPF2* (D) KD replicates, indicated in parentheses. (E) IRS correlation between KDs. Diagonal: IRS distributions of individual KDs. Below diagonal: correlation graphs of pairwise comparisons. Above diagonal: corresponding Spearman correlation coefficients. Red lines: ordinary least-squares (OLS) regression, orange lines: LOWESS, and gray lines: orthogonal distance regression (ODR).

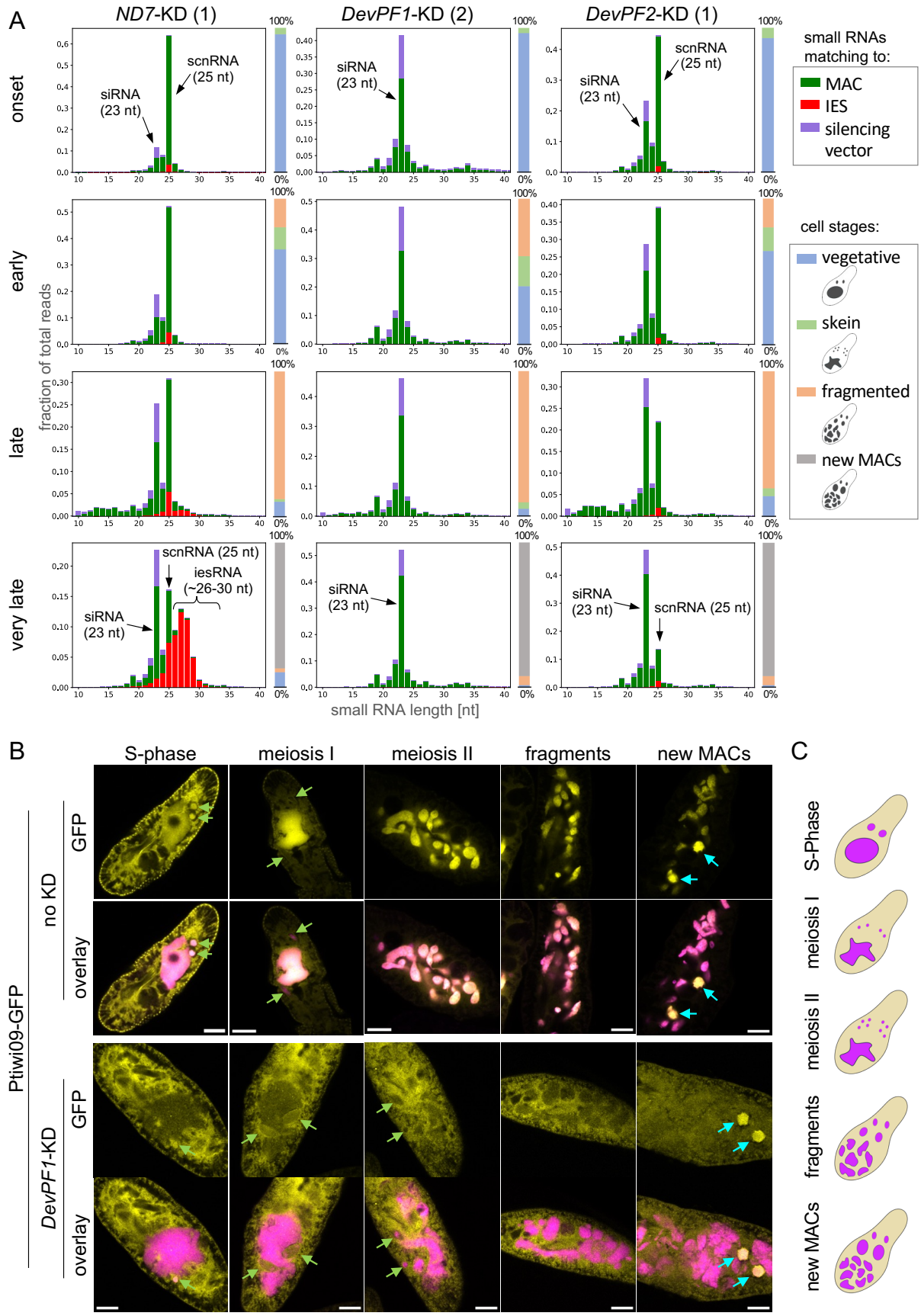
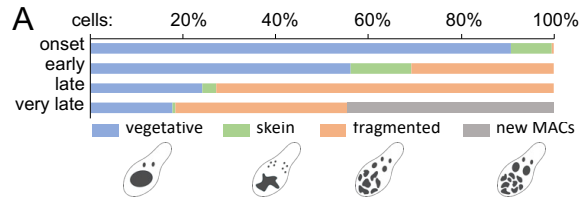


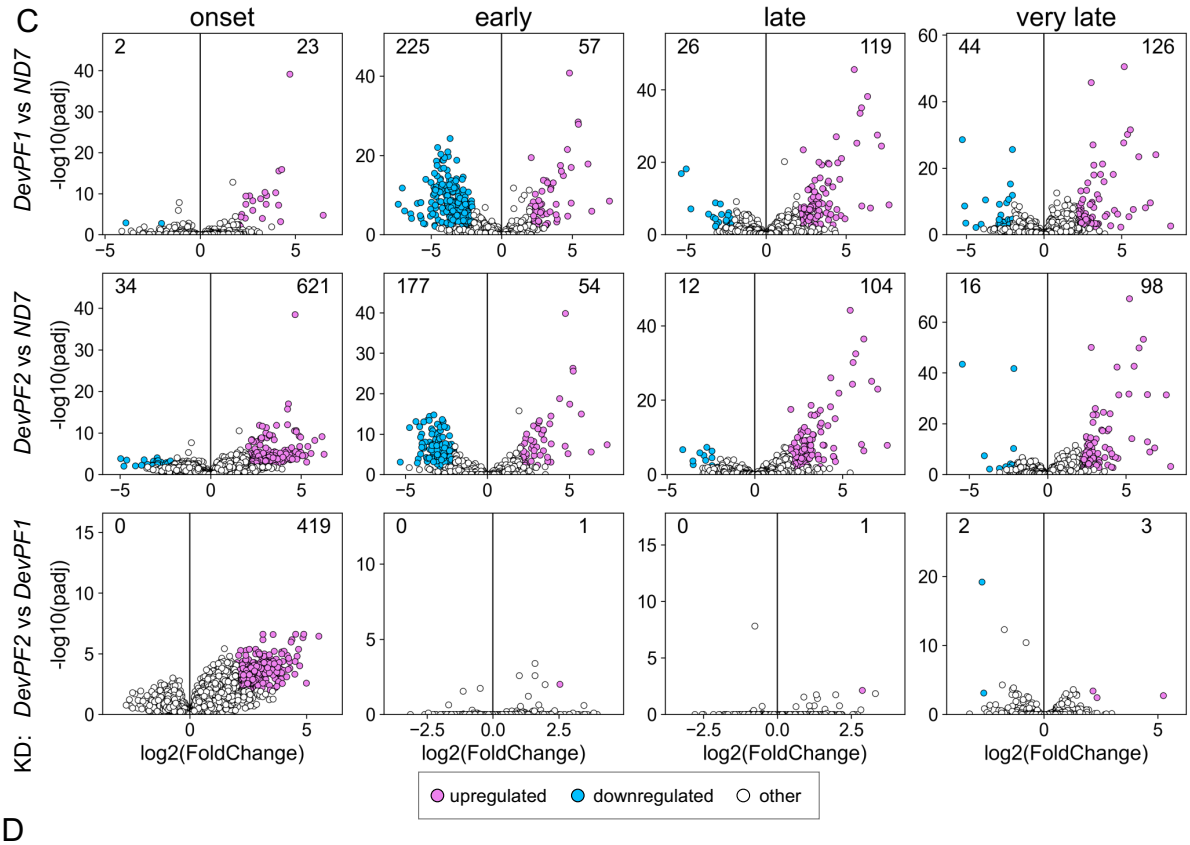
Figure 6: Changes of small RNA populations upon *DevPF* knockdowns

(A) Small RNA populations (10-40 nt) at developmental time points (onset, early, late and very late) in different KDs (*ND7* (control), *DevPF1* and *DevPF2*), mapping to silencing plasmid backbone (vector), MAC or IES sequences. Individual cell stage compositions are indicated by the bar to the right of each diagram, along with schematic representations of the cell stages considered. (B) Ptiwi09-GFP localization at different developmental stages in the context of no (top) and *DevPF1* KD (bottom). DAPI (pink) and GFP (yellow). Individual z-planes for GFP and overlay (DAPI and GFP). Green arrows: MICs. Cyan arrows: new MAC. Scale bar = 10 μm**.** (C) Schematic representation of cell stages in (B).



B

	timepoint	*ND7*-KD	*DevPF1*-KD	*DevPF2*-KD
scnRNAs	onset	+	-	+
	early	+	-	+
	late	+	-	+
	very late	+	-	+
iesRNAs	onset	-	-	-
	early	-	-	-
	late	+	-	-
	very late	+	-	-



D

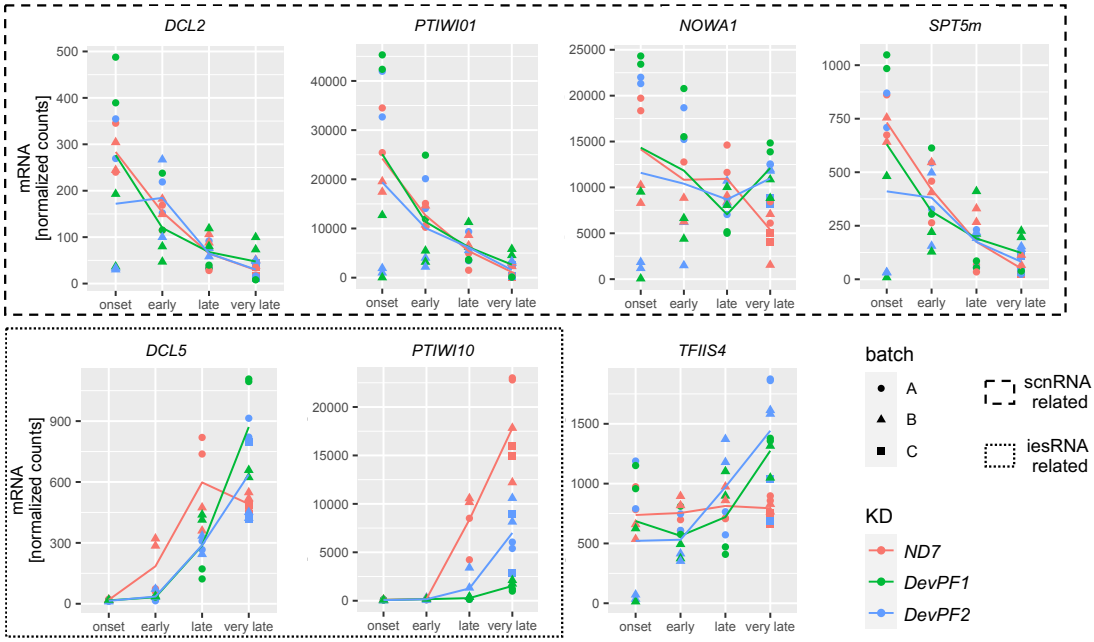


Figure 7: Differential gene expression in *DevPF* knockdowns

(A) Cell stage composition of each time point averaged over all KDs (individual compositions in Fig. S5), along with schematic representations of the considered cell stages. (B) Presence or absence of scnRNAs and iesRNAs in different KDs (*ND7*, *DevPF1* and *DevPF2*) and time points (onset, early, late, very late). (C) Differentially expressed genes in *DevPF1*- (top) or *DevPF2*- (middle) compared to *ND7*-KD or *DevPF1*- compared to *DevPF2*-KD (bottom) at different developmental time points (onset, early, late and very late). Thresholds for up-/downregulation: adjusted p-value < 0.01; |log2(fold change)| > 2. The number of up-/downregulated genes is indicated in each diagram. For all comparisons, 35777 transcripts were analyzed, except for: DevPF1-ND7 onset (33696), DevPF2-ND7 early (35083), and DevPF2-DevPF1 onset (34389). (D) Gene expression levels of selected genes upon KDs (*ND7* (control), *DevPF1* and *DevPF2*) at different developmental time points (onset, early, late and very late). The lines represent the mean of all replicates in a given KD and time point.

Table 1: IES retention scores of IESs at *PTIWI10/11* genes

The genes *PTIWI10* and *PTIWI11* contain IESs in their coding and/or flanking regions, which were proposed to impair their transcription when retained. The IRS values for the three relevant IESs (IDs with prefix IESPGM.PTET51.1) are provided for each KD. Rows are color-coded according to the KDs as shown in the mRNA read count diagrams (i. e. Figs 7D, S7B).

KD	Replicate	*PTIWI11*	*PTIWI10*	
		coding region	flanking region	coding region
		IESPGM.PTET51.1.62.345420	IESPGM.PTET51.1.24.407807	IESPGM.PTET51.1.24.408279
ND7	3	0.00	0.00	0.00
	4	0.00	0.00	0.00
	5	0.00	0.00	0.00
DevPF1	1	0.09	0.29	0.11
	2	0.08	0.15	0.06
	3	0.04	0.03	0.06
	4	0.02	0.01	0.01
DevPF2	1	0.10	0.07	0.01
	2	0.02	0.24	0.15
	3	0.00	0.00	0.00
	4	0.00	0.00	0.01
	5	0.03	0.00	0.00
	6	0.01	0.00	0.00

Two paralogous PHD finger proteins participate in *Paramecium tetraurelia*'s natural genome editing

Lilia Häußermann[1], Aditi Singh[1,#], Estienne C. Swart[1,#]

Supplementary information

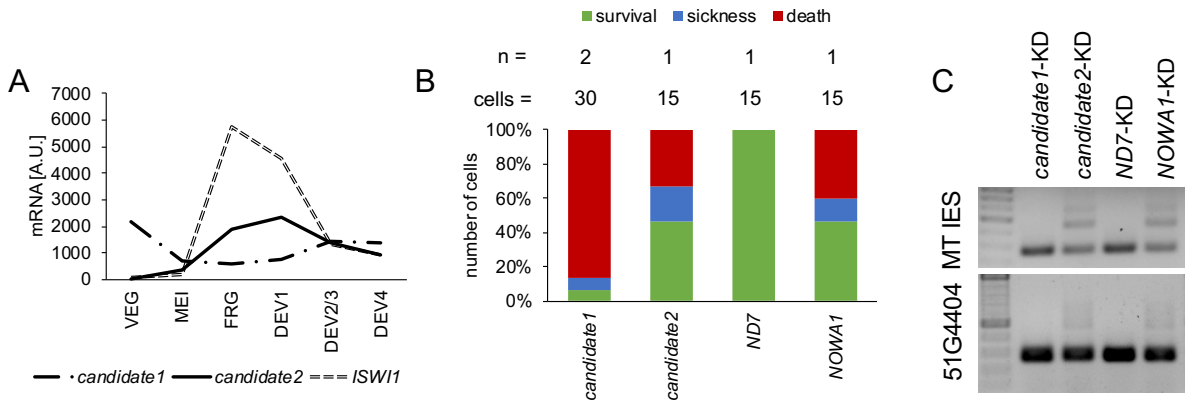


Figure S1: Screening for new candidate genes involved in *Paramecium* genome reorganization

(A) mRNA expression profiles in arbitrary units for candidate 1, candidate 2 and *ISWI1* during various developmental stages: VEG (vegetative growth), MEI (micronuclear meiosis and macronuclear fragmentation), FRG (~50% of the population with fragmented maternal MACs), DEV1 (significant proportion with visible anlagen), DEV2/3 (majority with visible anlagen), DEV4 (majority with visible anlagen). Expression data retrieved from ParameciumDB (Arnaiz *et al*., 2017). (B) Viability of new progeny after knockdowns (*ND7* (negative control) and *NOWA1* (positive control) compared to two candidate genes). Survival: no growth defects. Sickness: reduced division rate. Death: 3 or less cells after three days. The numbers of experiments (n) and cells counted (cells) are indicated at the top. (C) IES retention PCRs for two IESs on genomic DNA isolated from knockdown cells.

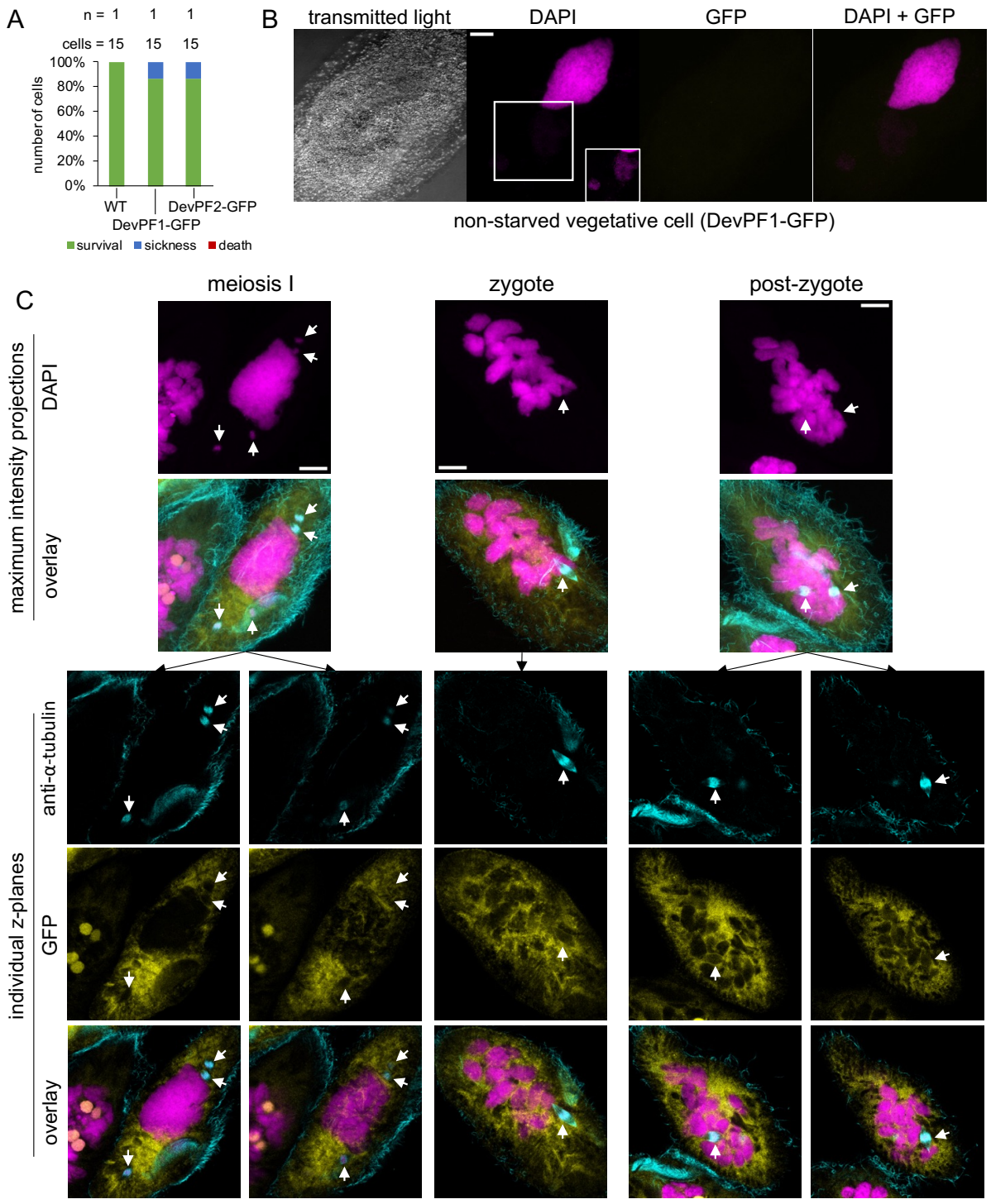
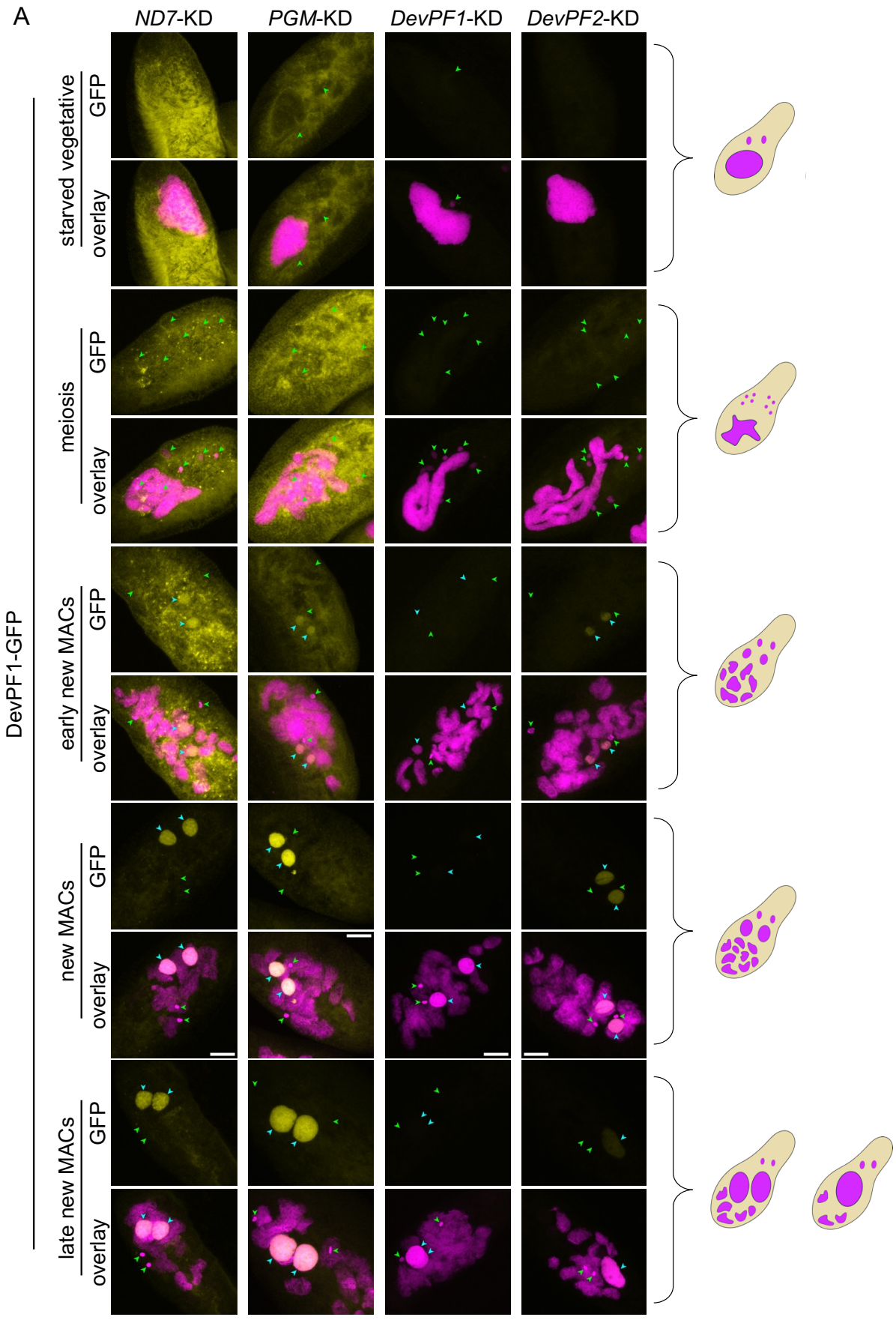


Figure S2: Additional imaging of DevPF1-GFP localization

(A) Viability of cells injected with GFP-fusion proteins compared to non-transformed wild type (WT). Survival: normal division. Sickness: reduced growth. Death: 3 or less cells after three days. The numbers of experiments (n) and cells counted (cells) are indicated at the top. (B) DevPF1-GFP localization in non-starved vegetative cells with bacteria inside food vacuoles. (C) Gametic and post-zygotic nuclei lacking DevPF1-GFP localization. Maximum intensity projections of multiple confocal z-

planes for (B) and (C) and individual confocal planes for (C). (B) and (C): DNA (stained with DAPI) in pink. GFP in yellow. Scale bar = 10 μm.


A
ND7-KD
PGM-KD
DevPF1-KD
DevPF2-KD
DevPF1-GFP
starved vegetative
meiosis
early new MACs
new MACs
late new MACs
GFP
overlay

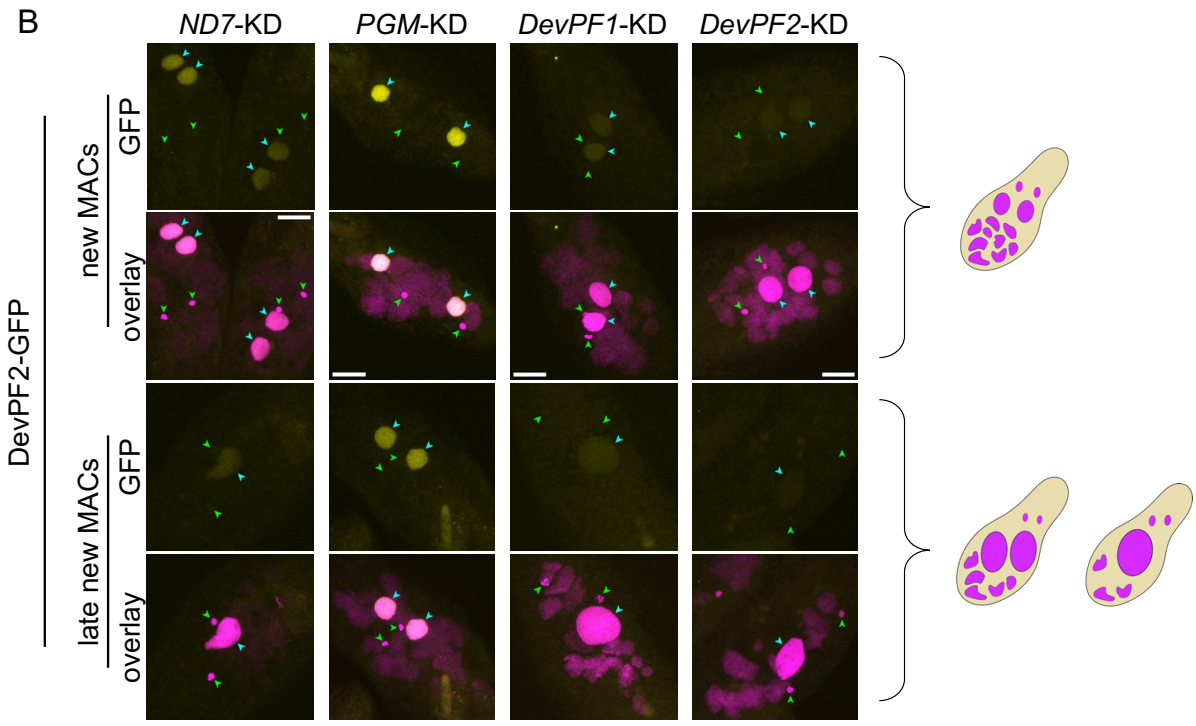


Figure S3: Localization of DevPF-GFP proteins upon gene knockdowns

DevPF1-GFP (A) and DevPF2-GFP (B) localization upon *ND7*, *PGM*, *DevPF1* and *DevPF2* knockdown. GFP signal (yellow) and DNA stained with DAPI (pink). Scale bar = 10 µm. Schematic representation of the corresponding cell stage (right).

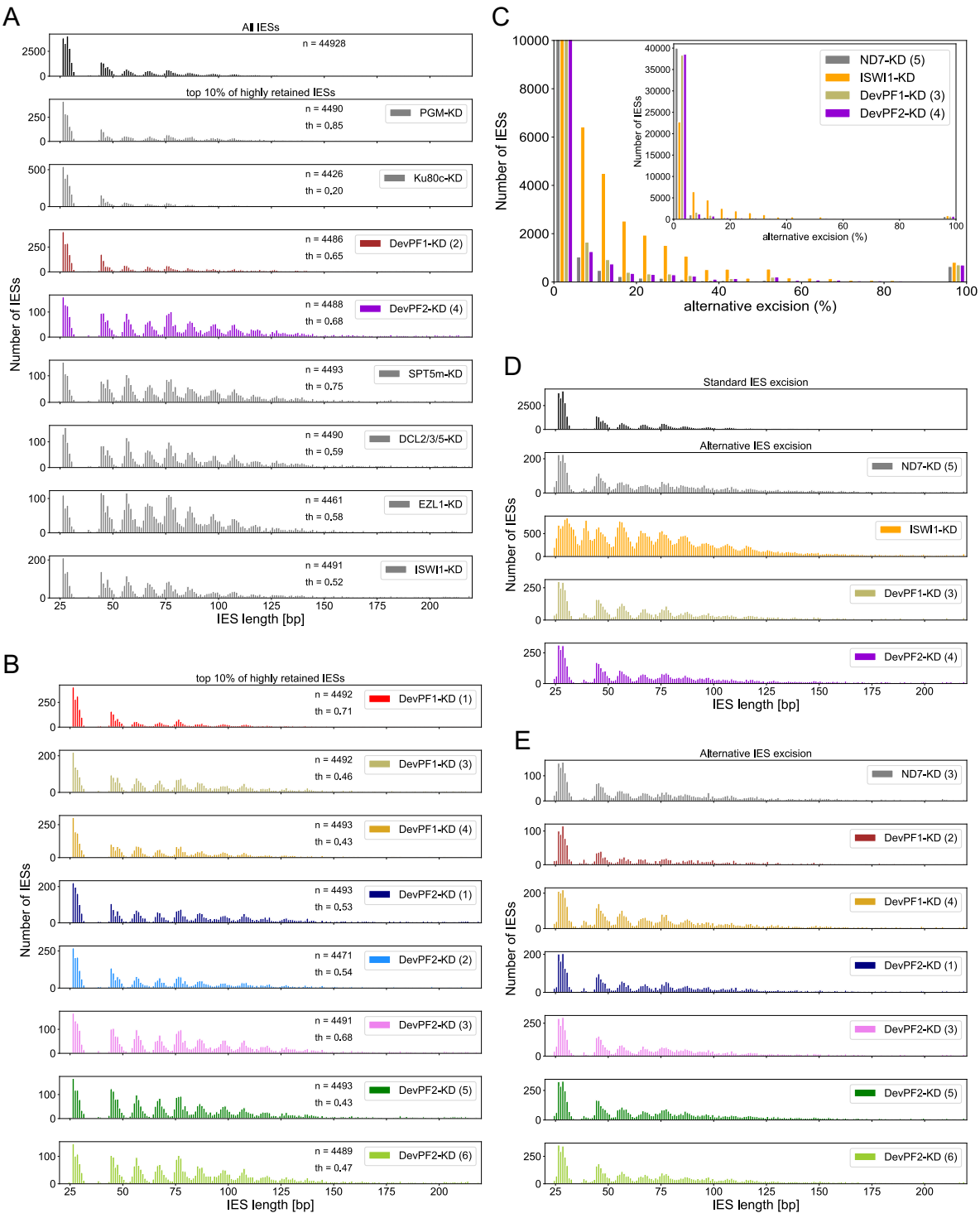


Figure S4: Analysis of length distributions of standard and alternatively excised IESs upon gene knockdown

(A) and (B): Length distribution of the top 10% most highly retained IESs in different knockdowns (A) and *DevPF* replicates (B). Indicated are the number of IESs (n) and the IRS threshold (th). (C) Histogram showing the fraction of alternative excision (%) for each IES in various knockdowns. *ND7:* negative control, *ISWI1*: positive control.

(D) and (E): Length distribution of alternatively excised IES in different knockdowns (D) and *DevPF* replicates (E).

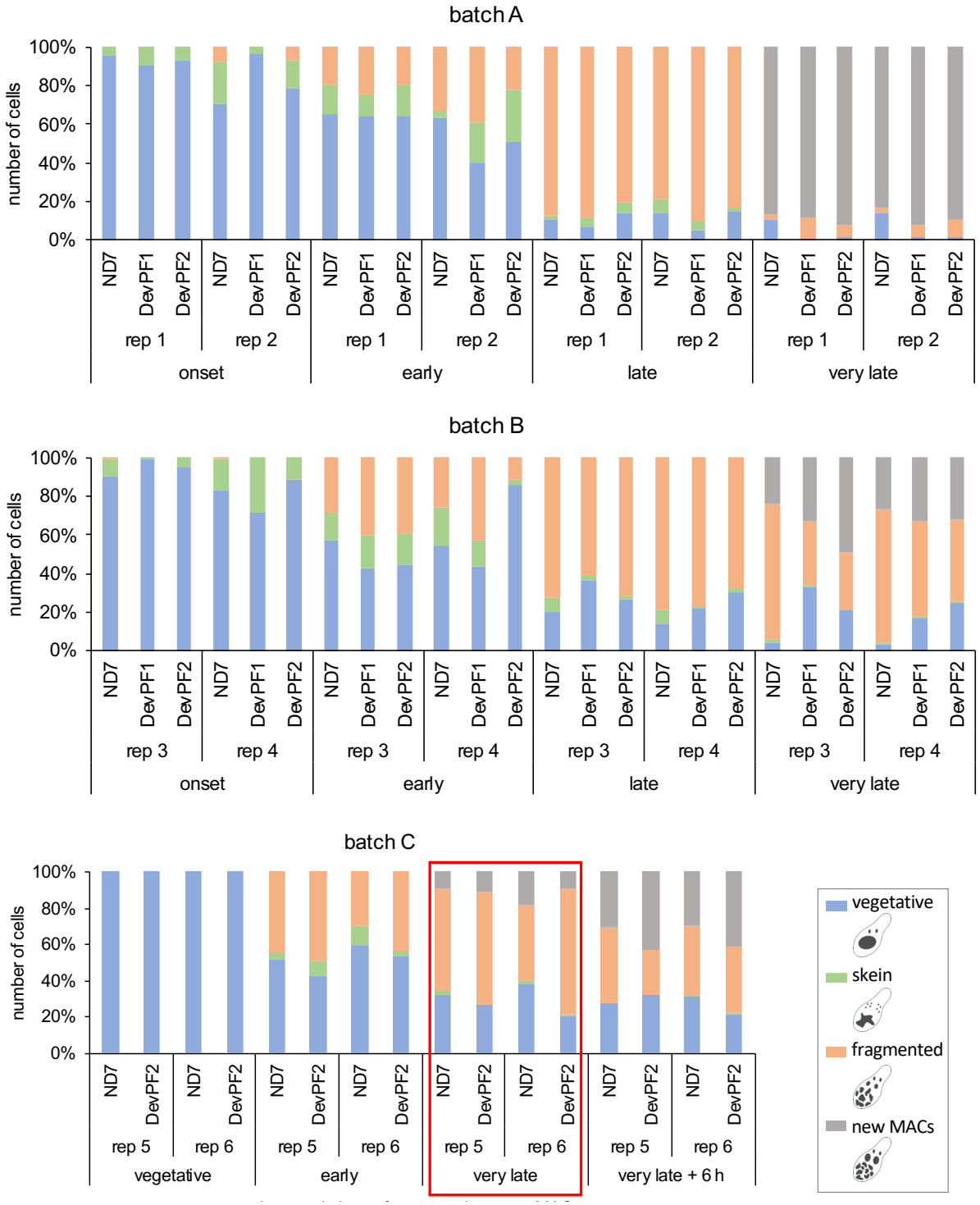


Figure S5: Cell staging for gene knockdown time course experiments

Cell staging of 6 time series conducted in 3 batches. A and B included *ND7*-KD, *DevPF1*-KD and *DevPF2*-KD, while C included *ND7*-KD and *DevPF2*-KD. All replicates and time points of batch A and B and “very late” time point of batch C were used for mRNA differential expression analysis. Cells with visible new MACs were considered in the “very late” and “very late + 6 h” time points. The schematic representation of the considered cell stages is provided.

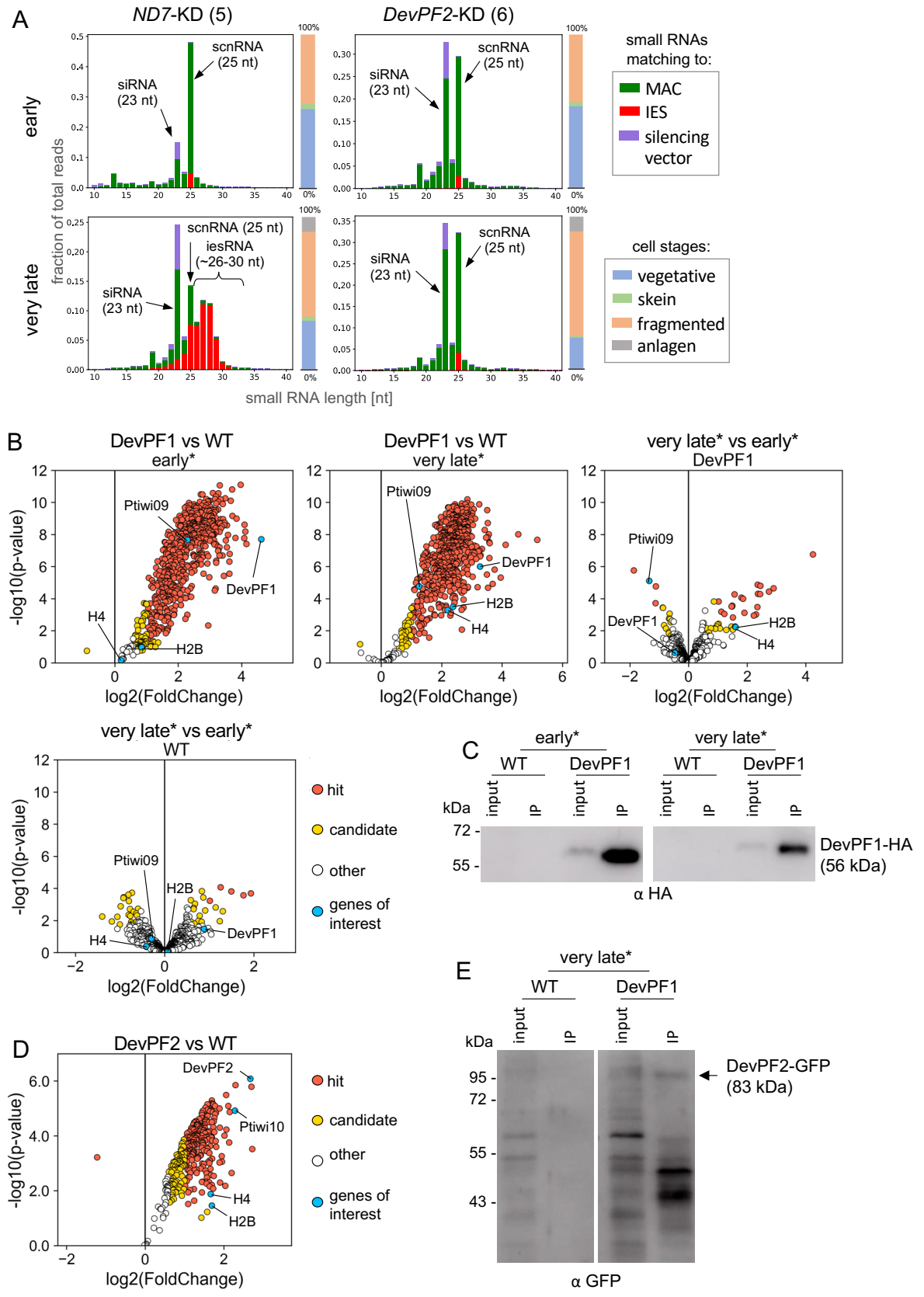


Figure S6: Small RNA populations and potential DevPF interaction partners

(A) Small RNA populations (10-40 nt) at “early” and “very late” time points in different knockdowns (*ND7* (control) and *DevPF2*) mapping to silencing plasmid backbone

(vector), MAC or IES sequences. Individual cell stage compositions are indicated in the bar to the right of each diagram, along with schematic representations of the cell stages considered. (B) and (D): Genes identified by mass spectrometry in DevPF1-HA (B) or DevPF2-GFP (D) immunoprecipitations (IP). Non-transformed wildtype cells (WT) as control. “hit”: false discovery rate (FDR) smaller 5% and a fold-change (FC) of at least 100%; “candidate”: FDR below 20% and a FC of at least 50%. Early (about 30% fragmentation) in (B) and very late (100% fragmentation + visible new MACs) time points in (B) and (D) were collected. For DevPF1-IP (B), multiple comparisons are shown. Above each subgraph the comparison between samples is indicated, with the common condition just below. Genes of interest are labeled with one name in cases where the detected peptides cannot be unambiguously distinguished. (C) and (E): Western blots of input and IP fraction from IPs performed on DevPF1-HA (HA-affinity IP) (E) and DevPF2-GFP (GFP-affinity IP) (F) with wildtype (WT) as control. (*): These time points are not represented in the cell stage compositions in Fig S5.

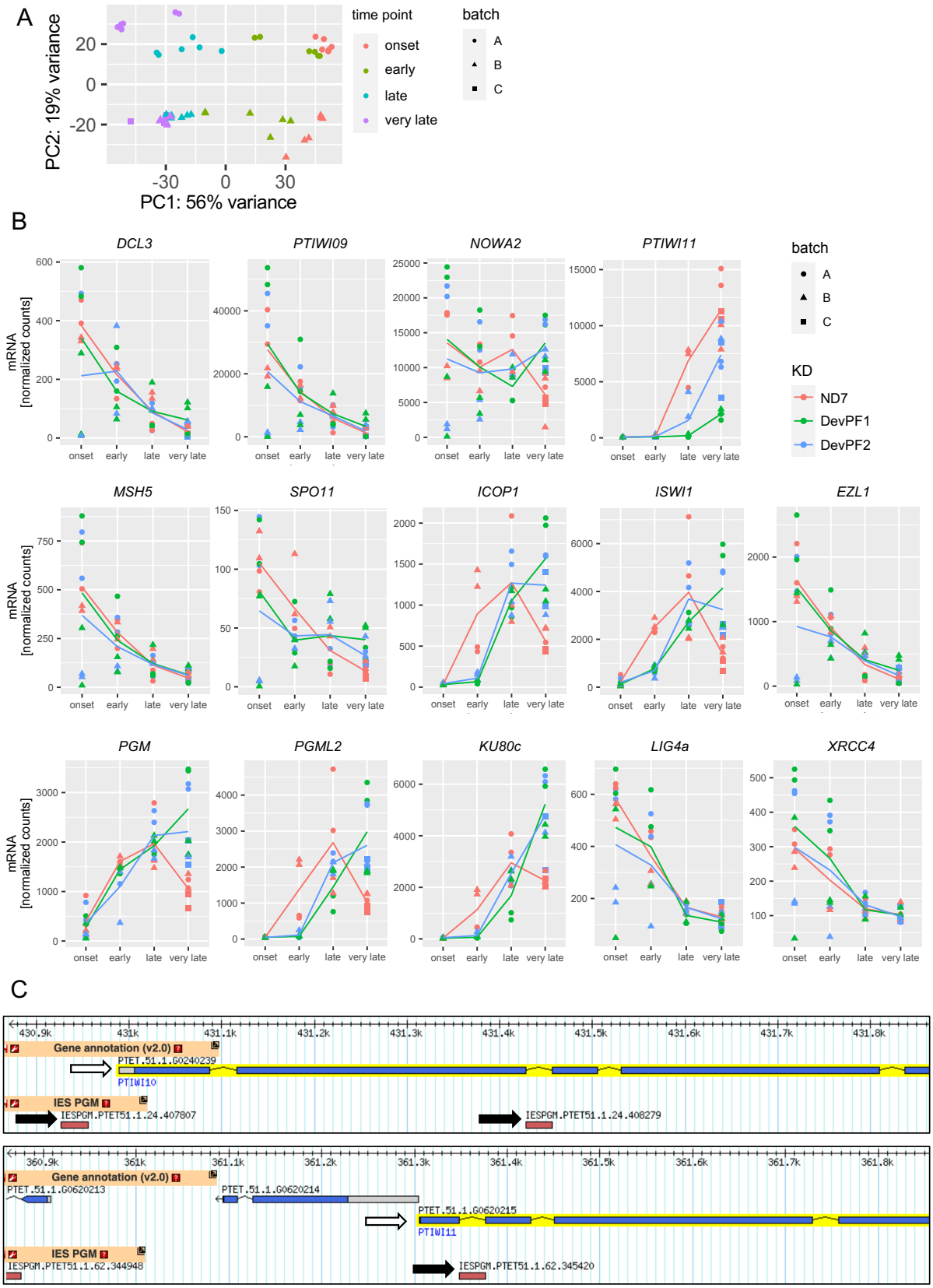

A
PC2: 19% variance
PC1: 56% variance
time point
onset
early
late
very late
batch
A
B
C
B
DCL3
PTIWI09
NOWA2
PTIWI11
MSH5
SPO11
ICOP1
ISWI1
EZL1
PGM
PGML2
KU80c
LIG4a
XRCC4
mRNA [normalized counts]
onset
early
late
very late
KD
ND7
DevPF1
DevPF2
C
Gene annotation (v2.0)
PTET.51.1.G0240239
PTIWI10
IES PGM
IESPGM.PTET51.1.24.407807
IESPGM.PTET51.1.24.408279
PTET.51.1.G0620213
PTET.51.1.G0620214
PTET.51.1.G0620215
PTIWI11
IESPGM.PTET51.1.62.344948
IESPGM.PTET51.1.62.345420

Figure S7: Differential expression of genes involved in *Paramecium* genome reorganization upon *DevPF* knockdowns

(A) Principal Component Analysis (PCA) for mRNA-seq samples. (B) Gene expression levels of selected genes upon knockdowns (*ND7* (control), *DevPF1* and *DevPF2*) at developmental time points (onset, early, late and very late). The lines represent the mean of all replicates in a given knockdown and time point. (C) Screenshot from GBrowse tool on ParameciumDB showing IESs in *PTIWI10* (top) and *PTIWI11* (bottom) flanking/coding regions. White arrow: start of the gene. Black arrow: IES.

Table S1: Lengths of the most highly retained IESs

For each KD, the median and mean length [bp] of the top 10% most highly retained IESs are provided.

Knockdown	Median [bp]	Mean [bp]
all IESs	50	79.1
PGM-KD	67	130.0
KU80c-KD	45	86.8
DevPF1-KD (2)	66	143.9
DevPF2-KD (4)	84.5	134.8
SPT5m-KD	85	179.4
DCL2/3/5-KD	85	153.2
EZL1-KD	77	145.7
ISWI1-KD	77	148.6
DevPF1-KD (1)	62	124.9
DevPF1-KD (3)	85	184.1
DevPF1-KD (4)	75	160.4
DevPF2-KD (1)	86	166.9
DevPF2-KD (2)	76	141.1
DevPF2-KD (3)	81	130.0
DevPF2-KD (5)	79	133.9
DevPF2-KD (6)	85	143.6

Table S2: Statistics alternative excision

For each KD, the percentage of alternative excision events was determined for each IES.The median and mean values for all IESs are provided. Replicates are indicated in parentheses. Rep = replicates.

Knockdown	Rep	median [%]	mean [%]
ND7	3	0	2.72
	4	0	2.43
ISWI1	-	4.35	10.88
DevPF1	2	0	2.53
	3	0	3.82
	4	0	3.49
DevPF2	1	0	3.54
	3	0	3.85
	4	0	3.96
	5	0	3.52
	6	0	3.53

Table S3: Protein abundance in DevPF1-IP

Selected proteins identified by mass spectrometry in DevPF1- and DevPF2-IP. Comparison: the compared samples and, if applicable, the common condition. Protein: Ambiguous hits are labeled with one name. “hit”: false discovery rate (FDR) smaller 5% and a fold-change (FC) of at least 100%; “candidate”: FDR below 20% and a FC of at least 50%.

Comparison	Protein	LogFC	Log(p)	Hit annotation
DevPF1 vs WT (early)	DevPF1	4.61	7.69	hit
	Ptiwi01	2.29	7.67	hit
	Histone 2B	0.84	0.97	candidate
	Histone 4	0.20	0.16	no hit
DevPF1 vs WT (late)	DevPF1	3.26	6.00	hit
	Ptiwi01	1.25	4.77	hit
	Histone 2B	2.37	3.49	hit
	Histone 4	2.20	3.29	hit
late vs early (DevPF1)	DevPF1	-0.47	0.64	no hit
	Ptiwi01	-1.35	5.11	hit
	Histone 2B	1.60	2.23	candidate
	Histone 4	1.59	2.26	candidate
late vs early (WT)	DevPF1	0.89	1.48	no hit
	Ptiwi01	-0.30	0.88	no hit
	Histone 2B	0.07	0.05	no hit
	Histone 4	-0.42	0.40	no hit
DevPF2 vs WT	DevPF2	2.67	6.08	hit
	Ptiwi10	2.28	4.92	hit
	Histone 2B	1.70	1.45	hit
	Histone 4	1.67	1.88	hit

Table S4: Differential expression of selected genes (p-value)

Adjusted p-values for differential expression (thresholds: |log2(Fold Change)| > 2; adjusted p-value < 0.01) of selected genes in different knockdown comparisons and time points. Magenta = upregulation; Blue = downregulation.

	DevPF1-KD vs *ND7*-KD				*DevPF2*-KD vs *ND7*-KD				*DevPF2*-KD vs *DevPF1*-KD			
Gene	onset	early	late	very late	onset	early	late	very late	onset	early	late	very late
DCL2	9.97E-01	1.00E+00	9.80E-01	6.60E-01	6.25E-01	1.00E+00	9.87E-01	9.85E-01	6.09E-01	1.00E+00	1.00E+00	1.00E+00
DCL3	9.96E-01	1.00E+00	9.99E-01	3.93E-01	6.85E-01	1.00E+00	9.97E-01	8.19E-01	7.26E-01	1.00E+00	1.00E+00	1.00E+00
DCL5	9.74E-01	9.87E-07	1.23E-01	1.97E-01	6.23E-01	7.36E-06	3.02E-01	6.29E-01	8.40E-01	1.00E+00	1.00E+00	1.00E+00
DevPF1	1.14E-02	6.41E-04	1.15E-01	5.81E-01	1.54E-01	1.00E+00	8.51E-01	9.61E-01	1.80E-01	8.07E-01	1.00E+00	1.00E+00
DevPF2	9.83E-01	6.44E-01	4.61E-02	6.63E-01	8.10E-01	2.36E-07	4.51E-08	5.54E-04	9.71E-01	2.25E-01	1.00E+00	4.75E-01
EZL1	9.97E-01	1.00E+00	9.20E-01	4.47E-01	6.19E-01	1.00E+00	9.53E-01	7.08E-01	6.53E-01	1.00E+00	1.00E+00	1.00E+00
GTSF1	9.93E-01	1.00E+00	9.16E-01	5.10E-01	7.25E-01	1.00E+00	9.64E-01	7.54E-01	8.26E-01	1.00E+00	1.00E+00	1.00E+00
ICOP1	9.94E-01	1.87E-20	8.06E-01	2.52E-03	6.28E-01	3.39E-13	9.97E-01	5.26E-03	5.83E-01	1.00E+00	1.00E+00	1.00E+00
ICOP2	8.98E-01	5.32E-25	9.70E-01	1.28E-01	3.22E-01	6.03E-14	7.81E-01	1.34E-02	8.69E-01	1.00E+00	1.00E+00	1.00E+00
ISWI1	8.16E-01	1.08E-03	7.46E-01	1.48E-02	9.13E-01	1.07E-04	9.90E-01	9.36E-03	3.43E-01	1.00E+00	1.00E+00	1.00E+00
KU80c	9.09E-01	1.44E-19	2.89E-01	5.67E-02	6.37E-01	8.69E-11	9.14E-01	4.30E-02	2.32E-01	1.00E+00	1.00E+00	1.00E+00
LIG4a	9.37E-01	1.00E+00	8.45E-01	8.04E-01	3.82E-01	1.00E+00	9.96E-01	9.92E-01	7.68E-01	1.00E+00	1.00E+00	1.00E+00
MSH4a	9.84E-01	1.00E+00	9.13E-01	7.58E-01	4.85E-01	1.00E+00	9.62E-01	9.82E-01	6.50E-01	1.00E+00	1.00E+00	1.00E+00
MSH4b	9.96E-01	1.00E+00	6.50E-01	4.96E-01	4.89E-01	1.00E+00	9.26E-01	7.73E-01	5.35E-01	1.00E+00	1.00E+00	1.00E+00
MSH5	9.94E-01	1.00E+00	9.75E-01	7.96E-01	7.16E-01	1.00E+00	9.94E-01	8.26E-01	7.84E-01	1.00E+00	1.00E+00	1.00E+00
ND7	1.65E-13	3.43E-20	3.25E-24	1.69E-46	2.89E-11	1.62E-16	2.85E-18	7.18E-51	6.89E-01	1.00E+00	1.00E+00	1.00E+00
NOWA1	9.87E-01	1.00E+00	8.57E-01	3.23E-01	6.00E-01	1.00E+00	9.62E-01	3.13E-01	7.21E-01	1.00E+00	1.00E+00	1.00E+00
NOWA2	9.87E-01	1.00E+00	7.92E-01	3.16E-01	5.68E-01	1.00E+00	9.60E-01	2.75E-01	7.02E-01	1.00E+00	1.00E+00	1.00E+00
PDSG2	8.53E-01	1.93E-13	6.82E-01	3.83E-02	3.18E-01	1.54E-12	9.64E-01	1.07E-01	9.93E-01	1.00E+00	1.00E+00	1.00E+00
PGM	9.31E-01	1.00E+00	9.71E-01	7.76E-02	8.30E-01	1.00E+00	9.61E-01	8.70E-02	7.76E-01	1.00E+00	1.00E+00	1.00E+00
PGML1	8.64E-01	1.65E-12	8.66E-01	9.71E-03	4.68E-01	2.62E-08	9.16E-01	3.97E-02	8.99E-01	1.00E+00	1.00E+00	1.00E+00
PGML2	9.96E-01	1.46E-19	3.47E-01	1.30E-02	8.92E-01	2.71E-14	8.99E-01	5.60E-03	9.36E-01	1.00E+00	1.00E+00	1.00E+00
PGML3a	9.49E-01	1.47E-12	9.96E-02	6.76E-02	6.66E-01	4.12E-08	9.49E-01	1.78E-02	9.95E-01	1.00E+00	1.00E+00	1.00E+00
PGML3b	8.98E-01	1.11E-06	4.02E-01	3.70E-01	9.87E-01	7.64E-03	9.11E-01	1.09E-01	5.15E-01	1.00E+00	1.00E+00	1.00E+00
PGML3c	9.87E-01	4.72E-01	4.14E-02	2.65E-01	5.12E-01	5.41E-01	4.73E-01	2.46E-01	6.44E-01	1.00E+00	1.00E+00	1.00E+00
PGML4a	9.58E-01	2.15E-10	7.33E-01	2.46E-04	8.05E-01	3.26E-08	8.67E-01	3.57E-02	9.23E-01	1.00E+00	1.00E+00	1.00E+00
PGML4b	9.20E-01	2.13E-08	9.15E-01	3.77E-03	8.67E-01	5.82E-07	8.67E-01	7.23E-03	7.08E-01	1.00E+00	1.00E+00	1.00E+00
PGML5a	9.50E-01	6.89E-09	9.28E-01	5.96E-03	5.70E-01	4.48E-06	9.26E-01	2.19E-02	9.04E-01	1.00E+00	1.00E+00	1.00E+00
PGML5b	9.94E-01	2.06E-17	4.86E-01	1.25E-01	9.25E-01	2.17E-09	9.14E-01	3.09E-01	9.89E-01	1.00E+00	1.00E+00	1.00E+00
PTCAF1	9.99E-01	1.00E+00	9.79E-01	4.52E-01	6.71E-01	1.00E+00	9.60E-01	6.73E-01	6.92E-01	1.00E+00	1.00E+00	1.00E+00
PTIWI01	9.97E-01	1.00E+00	9.65E-01	5.85E-01	8.81E-01	1.00E+00	9.79E-01	7.11E-01	8.68E-01	1.00E+00	1.00E+00	1.00E+00
PTIWI09	9.96E-01	1.00E+00	9.48E-01	5.78E-01	8.72E-01	1.00E+00	9.82E-01	7.56E-01	8.43E-01	1.00E+00	1.00E+00	1.00E+00
PTIWI10	9.77E-01	1.00E+00	6.29E-19	3.04E-11	5.72E-01	1.00E+00	7.47E-06	2.73E-02	7.70E-01	1.00E+00	1.73E-01	3.46E-03
PTIWI11	9.87E-01	1.00E+00	1.33E-17	1.96E-04	9.58E-01	1.00E+00	3.54E-03	4.99E-01	9.41E-01	1.00E+00	7.50E-03	2.11E-01
SPO11	9.77E-01	1.00E+00	8.32E-01	2.05E-01	6.36E-01	1.00E+00	9.02E-01	4.19E-01	8.36E-01	1.00E+00	1.00E+00	1.00E+00
SPT5m	9.96E-01	1.00E+00	9.66E-01	4.37E-01	6.67E-01	1.00E+00	9.81E-01	6.80E-01	7.24E-01	1.00E+00	1.00E+00	1.00E+00
TFIIS4	9.87E-01	1.00E+00	9.54E-01	6.85E-01	5.42E-01	1.00E+00	9.37E-01	4.74E-01	6.76E-01	1.00E+00	1.00E+00	1.00E+00

Table S5: Differential expression of selected genes (fold change)

Log2 transformed fold change for differential expression (thresholds: |log2(Fold Change)| > 2; adjusted p-value < 0.01) of selected genes in different knockdown comparisons and time points. Magenta = upregulation; Blue = downregulation.

	DevPF1-KD vs *ND7*-KD				*DevPF2*-KD vs *ND7*-KD				*DevPF2*-KD vs *DevPF1*-KD			
Gene	onset	early	late	very late	onset	early	late	very late	onset	early	late	very late
DCL2	0.04	-0.28	0.07	0.60	-0.61	0.27	0.08	0.04	-0.65	0.56	0.00	-0.55
DCL3	-0.07	-0.31	0.01	1.31	-0.67	0.09	0.02	0.49	-0.60	0.40	0.01	-0.82
DCL5	-0.23	-2.50	-1.15	0.82	-0.43	-2.37	-1.09	0.40	-0.20	0.12	0.07	-0.41
DevPF1	-2.94	-3.13	-1.98	-0.70	-1.44	-0.88	-0.75	-0.10	1.50	2.25	1.23	0.60
DevPF2	-0.16	-0.92	-1.38	-0.35	-0.19	-2.61	-2.74	-1.57	-0.03	-1.69	-1.36	-1.22
EZL1	-0.04	-0.01	0.29	1.02	-0.71	-0.21	0.29	0.63	-0.66	-0.20	0.00	-0.39
GTSF1	-0.17	0.03	0.31	0.96	-0.53	-0.27	0.22	0.57	-0.36	-0.29	-0.09	-0.39
ICOP1	-0.06	-3.78	-0.29	1.43	0.35	-3.05	-0.01	1.18	0.41	0.74	0.28	-0.25
ICOP2	-0.48	-3.65	-0.05	0.72	-0.61	-2.70	0.44	0.92	-0.13	0.95	0.49	0.21
ISWI1	-0.72	-1.69	-0.39	1.27	-0.08	-1.91	-0.03	1.18	0.64	-0.22	0.35	-0.09
KU80c	-0.51	-4.13	-0.93	1.13	0.38	-3.08	-0.28	1.07	0.89	1.05	0.65	-0.05
LIG4a	-0.37	0.10	-0.28	-0.23	-0.60	-0.20	-0.01	-0.01	-0.24	-0.30	0.26	0.21
MSH4a	-0.23	0.01	0.26	0.43	-0.80	-0.27	0.20	-0.05	-0.56	-0.28	-0.06	-0.47
MSH4b	-0.03	0.10	0.68	0.63	-0.62	-0.09	0.30	0.34	-0.58	-0.20	-0.38	-0.29
MSH5	-0.10	-0.24	0.10	0.41	-0.49	-0.45	0.03	0.38	-0.39	-0.21	-0.06	-0.03
ND7	1.71	2.09	2.33	3.03	1.57	1.94	2.05	2.78	-0.14	-0.15	-0.28	-0.25
NOWA1	-0.17	-0.01	-0.40	1.05	-0.62	-0.25	-0.19	1.09	-0.44	-0.24	0.21	0.04
NOWA2	-0.18	-0.20	-0.52	0.99	-0.62	-0.34	-0.19	1.07	-0.44	-0.14	0.33	0.08
PDSG2	-0.97	-4.21	-0.61	1.48	-0.96	-4.13	0.14	1.14	0.01	0.08	0.74	-0.34
PGM	-0.42	-0.15	0.06	1.11	-0.18	-0.62	0.13	1.01	0.24	-0.47	0.06	-0.10
PGML1	-0.74	-3.48	0.25	1.52	-0.60	-2.90	0.28	1.15	0.13	0.59	0.03	-0.36
PGML2	-0.04	-4.19	-0.88	1.41	-0.12	-3.63	-0.33	1.35	-0.08	0.56	0.55	-0.07
PGML3a	-0.35	-3.38	-1.23	1.12	-0.36	-2.75	-0.17	1.21	-0.01	0.62	1.06	0.10
PGML3b	1.11	-3.96	-1.10	0.83	0.03	-2.45	-0.40	1.26	-1.08	1.51	0.71	0.43
PGML3c	0.13	-1.04	-1.30	0.68	0.53	-1.07	-0.87	0.70	0.40	-0.03	0.43	0.02
PGML4a	-0.31	-3.09	0.44	1.92	-0.21	-2.80	0.43	1.13	0.10	0.29	-0.02	-0.78
PGML4b	-0.39	-2.32	0.14	1.33	-0.12	-2.15	0.35	1.10	0.27	0.17	0.21	-0.23
PGML5a	-0.36	-3.08	-0.15	1.68	-0.48	-2.60	0.26	1.31	-0.12	0.48	0.42	-0.37
PGML5b	-0.08	-4.31	-0.77	1.03	-0.09	-3.16	-0.30	0.75	-0.01	1.15	0.47	-0.28
PTCAF1	-0.02	-0.22	0.08	0.94	-0.56	0.01	0.23	0.65	-0.54	0.23	0.15	-0.29
PTIWI01	0.06	-0.15	0.20	1.09	-0.31	-0.32	0.19	0.86	-0.37	-0.17	-0.01	-0.23
PTIWI09	0.11	-0.01	0.30	1.26	-0.38	-0.33	0.18	0.85	-0.50	-0.33	-0.12	-0.41
PTIWI10	-0.23	-0.11	-5.01	-3.76	-0.54	-0.18	-2.91	-1.43	-0.30	-0.07	2.10	2.33
PTIWI11	-0.15	-0.72	-5.32	-2.61	-0.06	-0.18	-2.45	-0.73	0.09	0.54	2.87	1.88
SPO11	-0.31	-0.69	0.49	1.41	-0.59	-0.58	0.54	1.04	-0.29	0.11	0.04	-0.37
SPT5m	-0.10	-0.27	0.14	1.10	-0.65	-0.07	0.13	0.73	-0.56	0.20	-0.02	-0.37
TFIIS4	-0.17	-0.49	-0.13	0.47	-0.64	-0.58	0.29	0.79	-0.47	-0.10	0.42	0.32

Table S6: Primers used in IES Retention PCRs

IES	Primer sequence (5’ to 3’ orientation)
MT Locus F	GGTGTTTATATCTTAATTGTTGACCCTCAC
MT Locus R	CCATCTATACTCCATTCTTTATCTTAATTCAT
51G4404 F	CTGTTGCTACACATTGTGCATATGTTACT
51G4404 R	GCTGTAAGATTAACATTGAGCATGATCAAG

Table S7: Adapter sequences used for trimming

IES	Illumina adapter sequence (5’ to 3’ orientation)
DNA-seq (read 1)	AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
DNA-seq (read 2)	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
mRNA-seq (read 1)	AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC
mRNA-seq (read 2)	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT