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

indicates that *DevPF2*-KD replicates were more consistent than the *DevPF1*-KD replicates.

Genes that work closely together are expected to show similar KD effects on IES retention. To identify functionally related genes, *DevPF1*-KD and *DevPF2*-KD IRS data was correlated with published data from other gene KDs (Fig. 5E). *DevPF2*-KD (4) was selected from the *DevPF2* replicates. *DevPF1*-KD (2) and *DevPF1*-KD (4) were selected as representative of the variability observed in the *DevPF1*-KDs. *DevPF2*-KD (4) displayed high correlation with other KDs, such as *TFIIS4* and *DCL2/3/5* (Fig. 5E). Moderate correlation was observed for *DevPF1*-KD (4) with *SPT5m*, whereas *DevPF1*-KD (2) did not correlate well with any of the tested KDs.

Short IESs are proposed to predominantly rely on the excision complex (specifically Pgm (Baudry et al., 2009) and Ku80c (Marmignon et al., 2014)) for removal, while long IESs tend to require additional molecules for excision (Sellis et al., 2021). To determine whether *DevPF1*- and *DevPF2*-KD preferentially affect long IESs, the length distribution of the top 10% of highly retained IESs in each KD was plotted (Fig. S4A,B, Table S1). In comparison to the length distribution of all IESs, *DevPF2*-KD (4) showed an overrepresentation of long IESs, similar to observations in *EZL1*-KD, silencing of the catalytic subunit of the PCR2 complex (Frapporti et al., 2019; Lhuillier-Akakpo et al., 2014), or *DCL2/3/5*-KD, silencing of the scnRNA and iesRNA biogenesis proteins (Lepère et al., 2009; Sandoval et al., 2014), (Fig. S4A). Conversely, the highly retained IESs in *DevPF1*-KD (2) did not show the same overrepresentation and resembled the profile in *PGM*- and *KU80c*-KD, silencing of two members of the excision complex. Again, the replicates of the *DevPF* KDs exhibited variation in the extent of the observed effect (Fig. S4B).

Defects in IES excision not only result in the retention of IESs but can also lead to excision at alternative TA boundaries. So far, alternative excision above background 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).

599 It has been shown that evolutionarily old IESs tend to be short and are excised early in
600 development, independent of additional factors apart from the excision machinery
601 (Sellis et al., 2021). On the other hand, evolutionarily young IESs tend to be long, later
602 excised and dependent on the scnRNA pathway and the deposition of histone
603 modifications in the new MAC for their excision (Sellis et al., 2021; Swart et al., 2014;
604 Zangarelli et al., 2022). In line with this, most gene KDs tested in this study exhibited an
605 overrepresentation of long IESs among their most highly retained IESs, including
606 *DevPF2* (Fig. S4A,B). Only *PGM*-KD, *KU80c*-KD and two of the *DevPF1*-KD replicates
607 showed no preference for long IESs. Pgm and Ku80c are components of the excision
608 machinery and are therefore expected to affect all IESs. While DevPF1 may not be a
609 direct part of the excision machinery, it appears to have a general contribution to IES
610 excision, regardless of the length of the IES. Consequently, we propose that DevPF2
611 contributes to the excision of long IESs, while DevPF1 may serve as a more general
612 factor.
613

614 # Methods

615 ## *Paramecium tetraurelia* cultivation

616 Mating type 7 (MT7) cells from strain 51 of *Paramecium tetraurelia* were grown in
617 Wheat Grass Powder (WGP, Pines International) medium supplemented with 10 mM
618 sodium phosphate buffer (pH 7.3). WGP medium was bacterized with *E.coli* strain
619 HT115 to feed paramecia, and the cultures were maintained either at 27°C or at 18°C
620 according to the standard protocol (Beisson et al., 2010b, 2010c).
621

622 ## Protein localization imaging by fluorescence microscopy

623 Plasmids for microinjection were generated by amplifying the coding and flanking
624 sequences from MT7 genomic DNA and introducing them with the PCR-based method
625 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

861 expressed genes were identified for each time point with a Wald test (false discovery
862 rate (alpha) = 0.1). Differentially expressed genes were filtered with an absolute
863 log2(Fold Change) > 2 (corresponding to a 4-fold change) and an adjusted p-value <
864 0.01. The time point, KD and batch were known sources of variation in the dataset
865 (design = ~ batch + timepoint + KD+ timepoint:KD). All scripts are available from
866 https://github.com/Swart-lab/DevPF_code.
867

868 ## Structure prediction with AlphaFold

869 Protein structures were predicted with AlphaFold2 multimer (Evans et al., 2021; Jumper
870 et al., 2021) using the ColabFold v1.5.2-patch (Mirdita et al., 2022) in Google Colab with
871 default parameters.
872

873 ## Sequence alignment

874 Domains were predicted using InterProScan (Paysan-Lafosse et al., 2023). The
875 nucleotide sequence of DevPF2 and DevPF1 (including introns) were aligned with
876 clustalOmega (Sievers et al., 2011) (version 1.2.3) pairwise sequence alignment tool in
877 Geneious prime (version 2023.2.1) with default parameters (Fig 4A).
878
879 Multiple sequence alignment of PHD domains was done with clustalOmega (version
880 1.2.1) using the MPI bioinformatics toolkit's web interface (Zimmermann et al., 2018)
881 with default parameters.
882

883 ## Manuscript writing

884 Grammar and language refinement were assisted by an AI language model developed
885 by OpenAI (GPT-3.5 architecture) (OpenAI, 2023).
886

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

1061 Ignarski, M., Singh, A., Swart, E. C., Arambasic, M., Sandoval, P. Y., & Nowacki, M.
1062 (2014). Paramecium tetraurelia chromatin assembly factor-1-like protein PtCAF-1 is
1063 involved in RNA-mediated control of DNA elimination. *Nucleic Acids Research*,
1064 *42*(19), 11952–11964. https://doi.org/10.1093/nar/gku874

1065 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O.,
1066 Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C.,
1067 Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R.,
1068 Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with
1069 AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

1070 Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based
1071 genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature*
1072 *Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

1073 Kinkelin, K., Wozniak, G. G., Rothbart, S. B., Lidschreiber, M., Strahl, B. D., & Cramer,
1074 P. (2013). Structures of RNA polymerase II complexes with Bye1, a chromatin-
1075 binding PHF3/DIDO homologue. *Proceedings of the National Academy of Sciences of*
1076 *the United States of America*, *110*(38), 15277–15282.
1077 https://doi.org/10.1073/pnas.1311010110

1078 Klobutcher, L. A., & Herrick, G. (1995). Consensus inverted terminal repeat sequence of
1079 *Paramecium* IESs: resemblance to termini of Tc1-related and *Euplotes* Tec
1080 transposons. *Nucleic Acids Research*, *23*(11), 2006–2013.
1081 https://doi.org/10.1093/nar/23.11.2006

1082 Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., & Nesvizhskii, A. I.
1083 (2017). MSFragger: ultrafast and comprehensive peptide identification in mass
1084 spectrometry-based proteomics. *Nature Methods*, *14*(5), 513–520.
1085 https://doi.org/10.1038/nmeth.4256

1086 Krueger, F. (2019). *TrimGalore: A wrapper around Cutadapt and FastQC to consistently*
1087 *apply adapter and quality trimming to FastQ files, with extra functionality for RRBS*
1088 *data* (0.6.10) [Computer software]. GitHub.

1089 Lefort-Tran, M., Aufderheide, K., Pouphile, M., Rossignol, M., & Beisson, J. (1981).
1090 Control of exocytotic processes: cytological and physiological studies of trichocyst
1091 mutants in Paramecium tetraurelia. *The Journal of Cell Biology*, *88*(2), 301–311.
1092 https://doi.org/10.1083/jcb.88.2.301

1093 Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdeno-Tárraga, A., Cheng, Y.,
1094 Cleland, I., Faruque, N., Goodgame, N., Gibson, R., Hoad, G., Jang, M., Pakseresht,
1095 N., Plaister, S., Radhakrishnan, R., Reddy, K., Sobhany, S., Ten Hoopen, P.,
1096 Vaughan, R., … Cochrane, G. (2011). The european nucleotide archive. *Nucleic*
1097 *Acids Research*, *39*(Database issue), D28-31. https://doi.org/10.1093/nar/gkq967

1098 Lepère, G., Nowacki, M., Serrano, V., Gout, J.-F., Guglielmi, G., Duharcourt, S., &
1099 Meyer, E. (2009). Silencing-associated and meiosis-specific small RNA pathways in
1100 *Paramecium tetraurelia*. *Nucleic Acids Research*, *37*(3), 903–915.
1101 https://doi.org/10.1093/nar/gkn1018

1102 Lhuillier-Akakpo, M., Frapporti, A., Denby Wilkes, C., Matelot, M., Vervoort, M.,
1103 Sperling, L., & Duharcourt, S. (2014). Local effect of enhancer of zeste-like reveals
1104 cooperation of epigenetic and cis-acting determinants for zygotic genome
1105 rearrangements. *PLoS Genetics*, *10*(9), e1004665.
1106 https://doi.org/10.1371/journal.pgen.1004665

1107 Liu, L., Qin, S., Zhang, J., Ji, P., Shi, Y., & Wu, J. (2012). Solution structure of an
1108 atypical PHD finger in BRPF2 and its interaction with DNA. *Journal of Structural*
1109 *Biology*, *180*(1), 165–173. https://doi.org/10.1016/j.jsb.2012.06.014

1110 Liu, Y., Taverna, S. D., Muratore, T. L., Shabanowitz, J., Hunt, D. F., & Allis, C. D.
1111 (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

1162 Nowacki, M., Zagorski-Ostoja, W., & Meyer, E. (2005). Nowa1p and Nowa2p: novel
1163 putative RNA binding proteins involved in trans-nuclear crosstalk in Paramecium
1164 tetraurelia. *Current Biology*, *15*(18), 1616–1628.
1165 https://doi.org/10.1016/j.cub.2005.07.033

1166 OpenAI. (2023). ChatGPT *[Large language model]* (GPT-3.5) [Computer software].
1167 https://chat.openai.com/chat.

1168 Oppikofer, M., Sagolla, M., Haley, B., Zhang, H.-M., Kummerfeld, S. K., Sudhamsu, J.,
1169 Flynn, E. M., Bai, T., Zhang, J., Ciferri, C., & Cochran, A. G. (2017). Non-canonical
1170 reader modules of BAZ1A promote recovery from DNA damage. *Nature*
1171 *Communications*, *8*(1), 862. https://doi.org/10.1038/s41467-017-00866-0

1172 Pascual, J., Martinez-Yamout, M., Dyson, H. J., & Wright, P. E. (2000). Structure of the
1173 PHD zinc finger from human Williams-Beuren syndrome transcription factor. *Journal*
1174 *of Molecular Biology*, *304*(5), 723–729. https://doi.org/10.1006/jmbi.2000.4308

1175 Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A.,
1176 Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I.,
1177 Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire,
1178 C., … Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, *51*(D1), D418–
1179 D427. https://doi.org/10.1093/nar/gkac993

1180 Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S.,
1181 Kamatchinathan, S., Kundu, D. J., Prakash, A., Frericks-Zipper, A., Eisenacher, M.,
1182 Walzer, M., Wang, S., Brazma, A., & Vizcaíno, J. A. (2022). The PRIDE database
1183 resources in 2022: a hub for mass spectrometry-based proteomics evidences.
1184 *Nucleic Acids Research*, *50*(D1), D543–D552. https://doi.org/10.1093/nar/gkab1038

1185 Pinskaya, M., Ghavi-Helm, Y., Mariotte-Labarre, S., Morillon, A., Soutourina, J., &
1186 Werner, M. (2014). PHD and TFIIS-Like domains of the Bye1 transcription factor
1187 determine its multivalent genomic distribution. *Plos One*, *9*(7), e102464.
1188 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

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

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

Zeng, L., Zhang, Q., Li, S., Plotnikov, A. N., Walsh, M. J., & Zhou, M.-M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature*, *466*(7303), 258–262. https://doi.org/10.1038/nature09139

Zhang, J., Yan, G., Tian, M., Ma, Y., Xiong, J., & Miao, W. (2018). A DP-like transcription factor protein interacts with E2fl1 to regulate meiosis in Tetrahymena thermophila. *Cell Cycle*, *17*(5), 634–642. https://doi.org/10.1080/15384101.2018.1431595

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N., & Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*, *430*(15), 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007

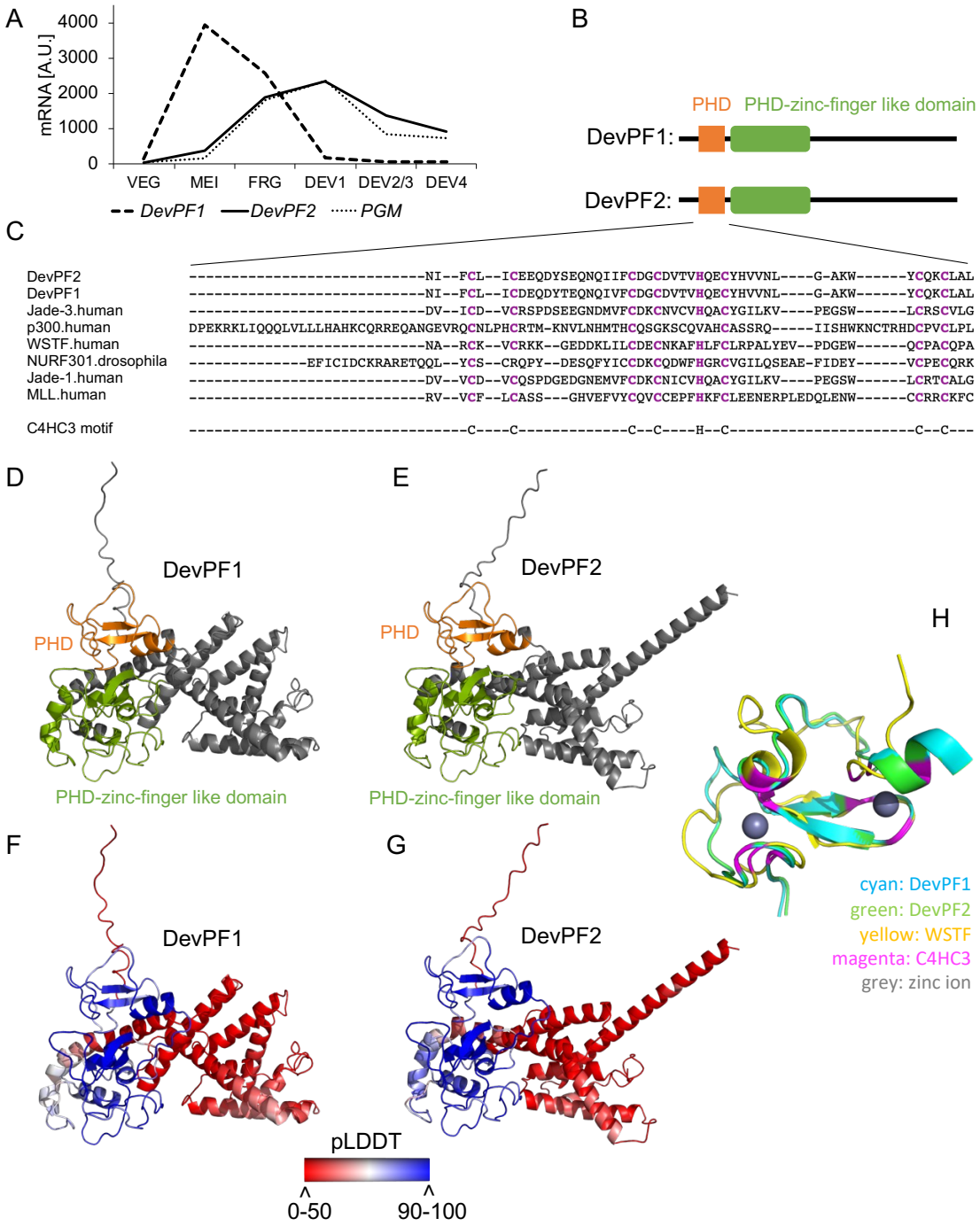


Figure 1: Features of the PHD finger proteins DevPF1 and DevPF2

(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).

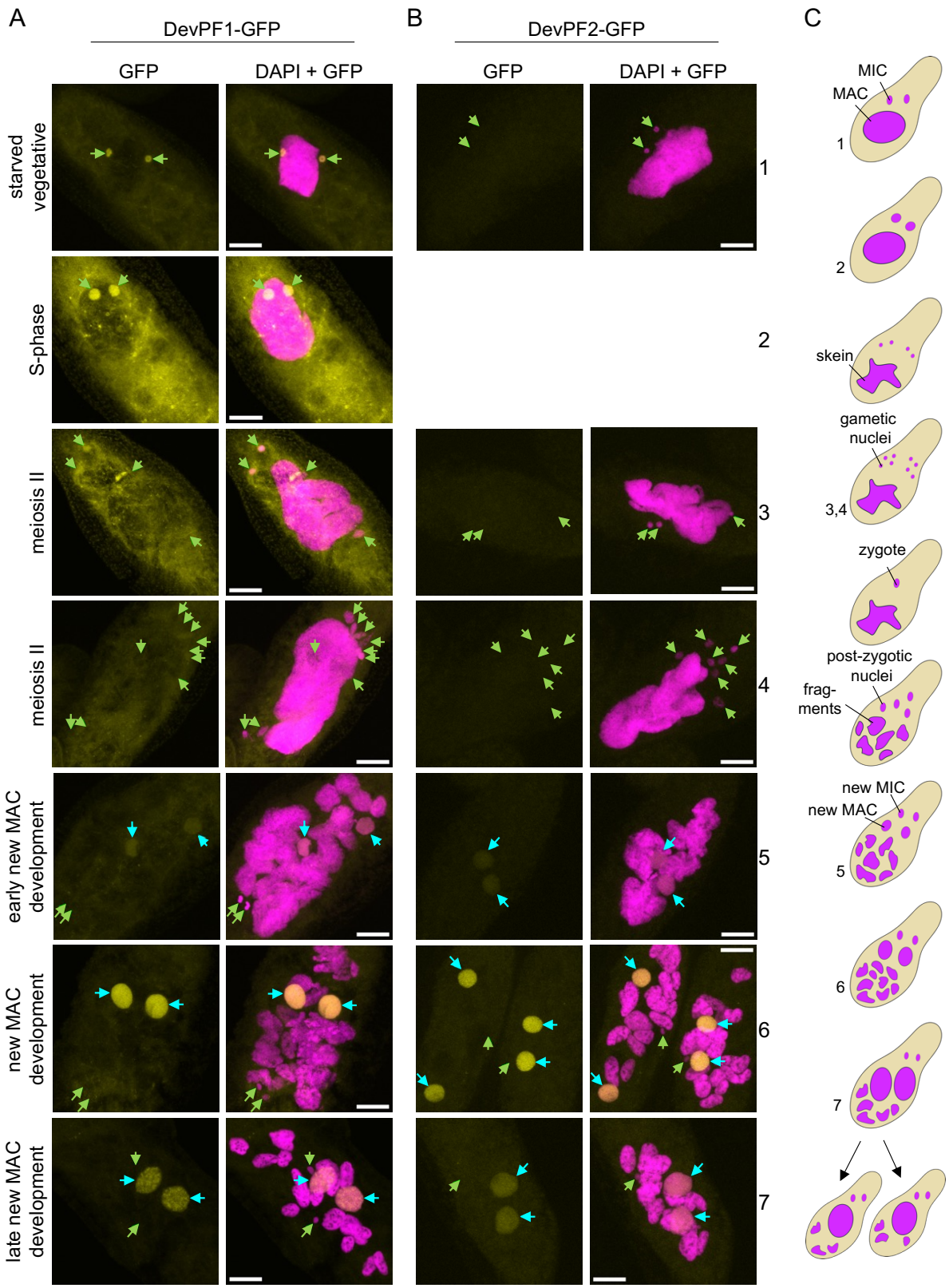


Figure 2: Subcellular localization of DevPF-GFP proteins

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.

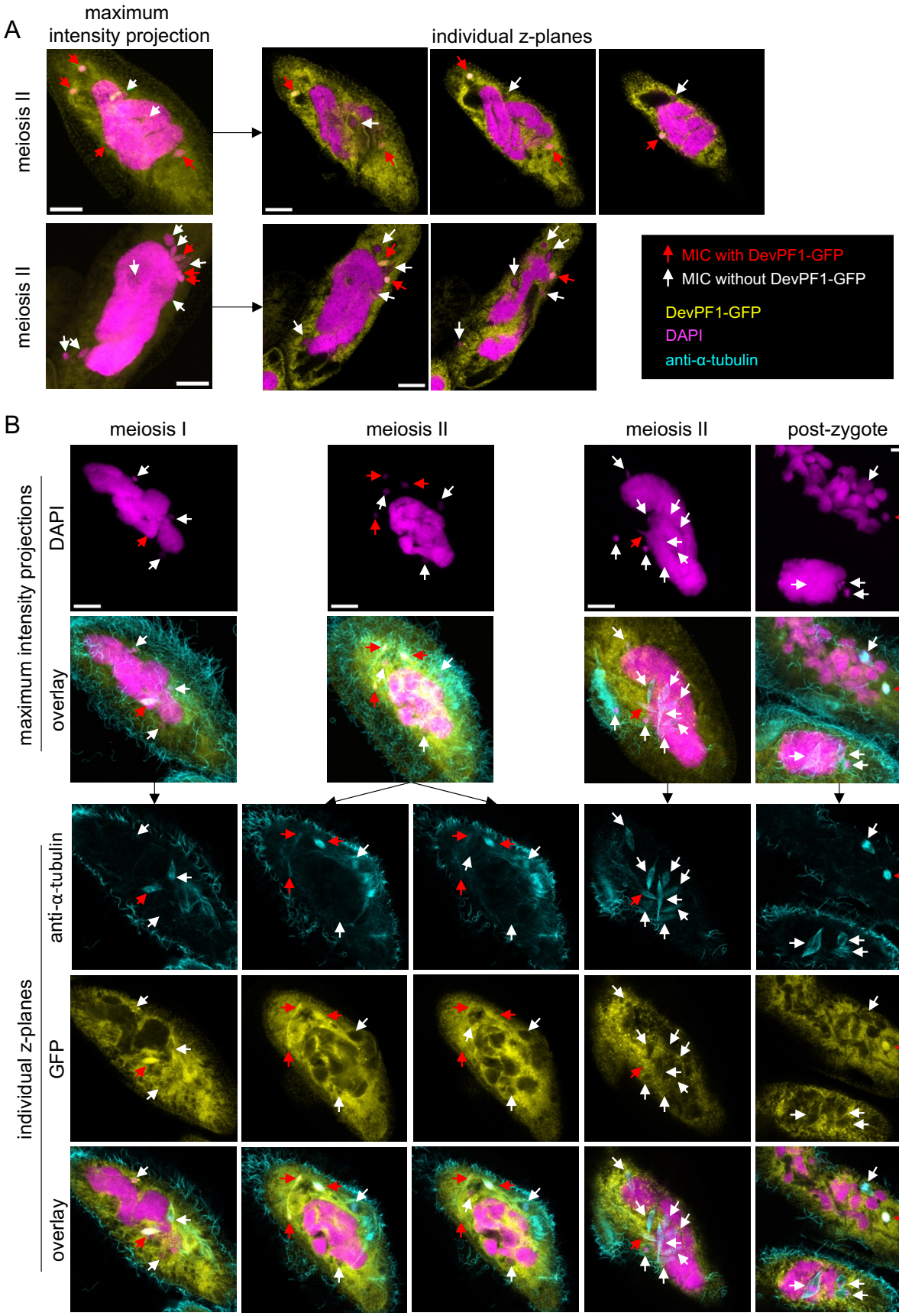


Figure 3: Selective DevPF1-GFP localization in *Paramecium* MICs

(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.

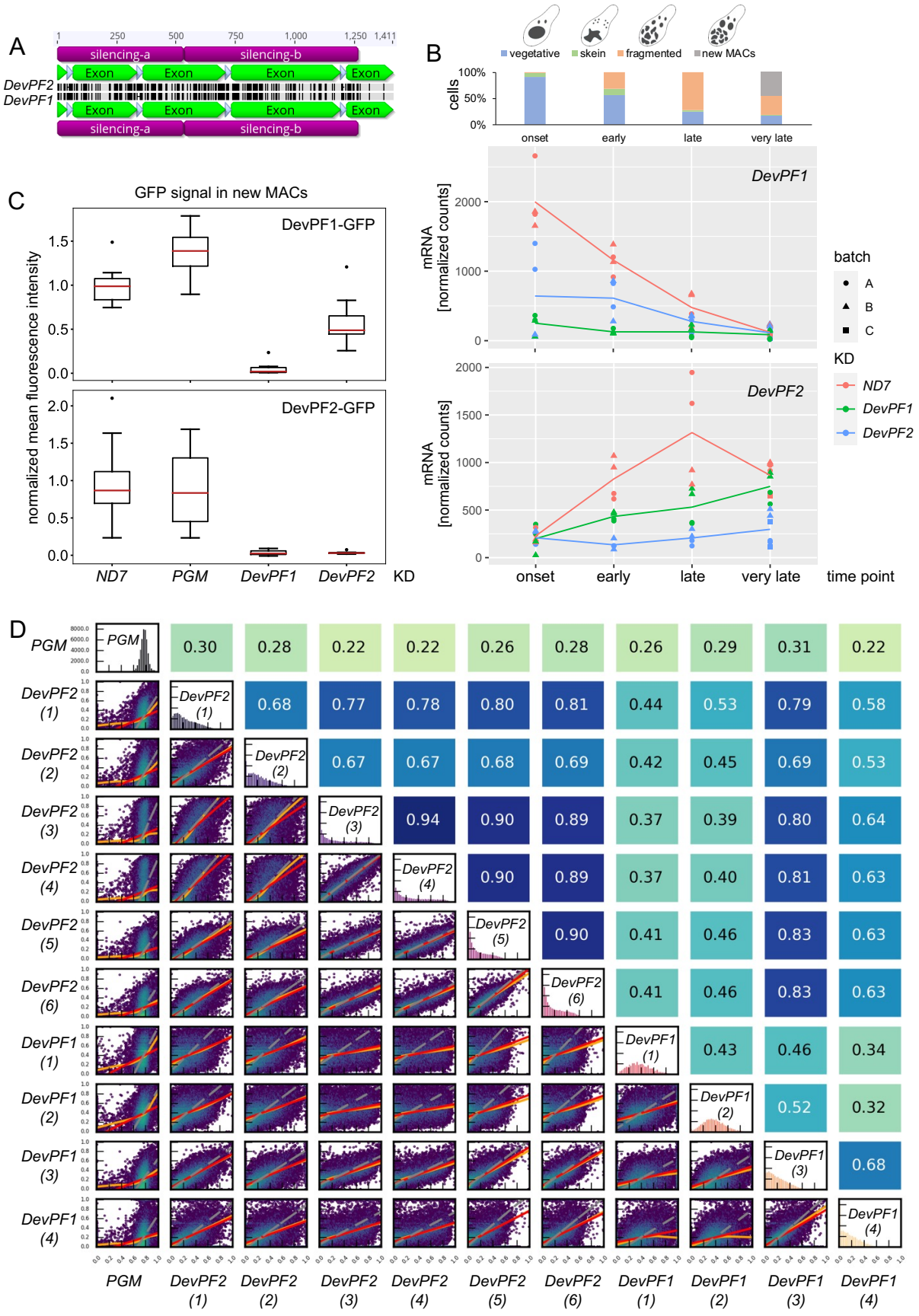
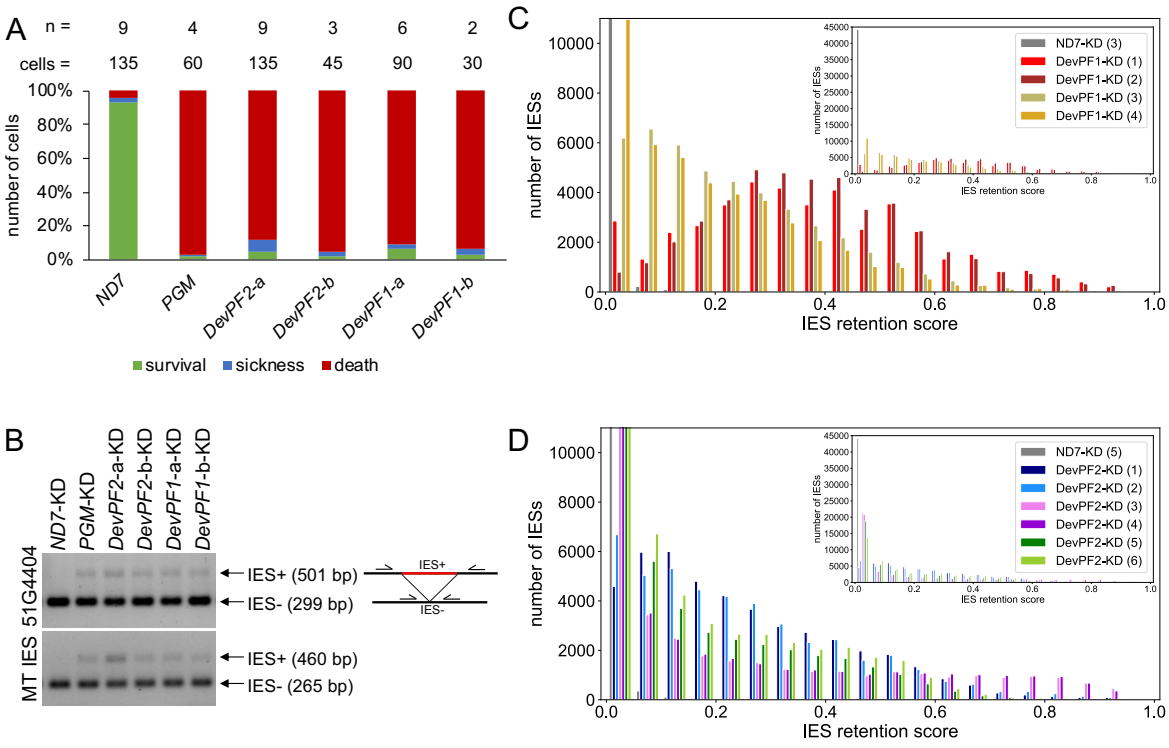


Figure 4: Co-silencing effects observed in *DevPF* knockdowns

(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).



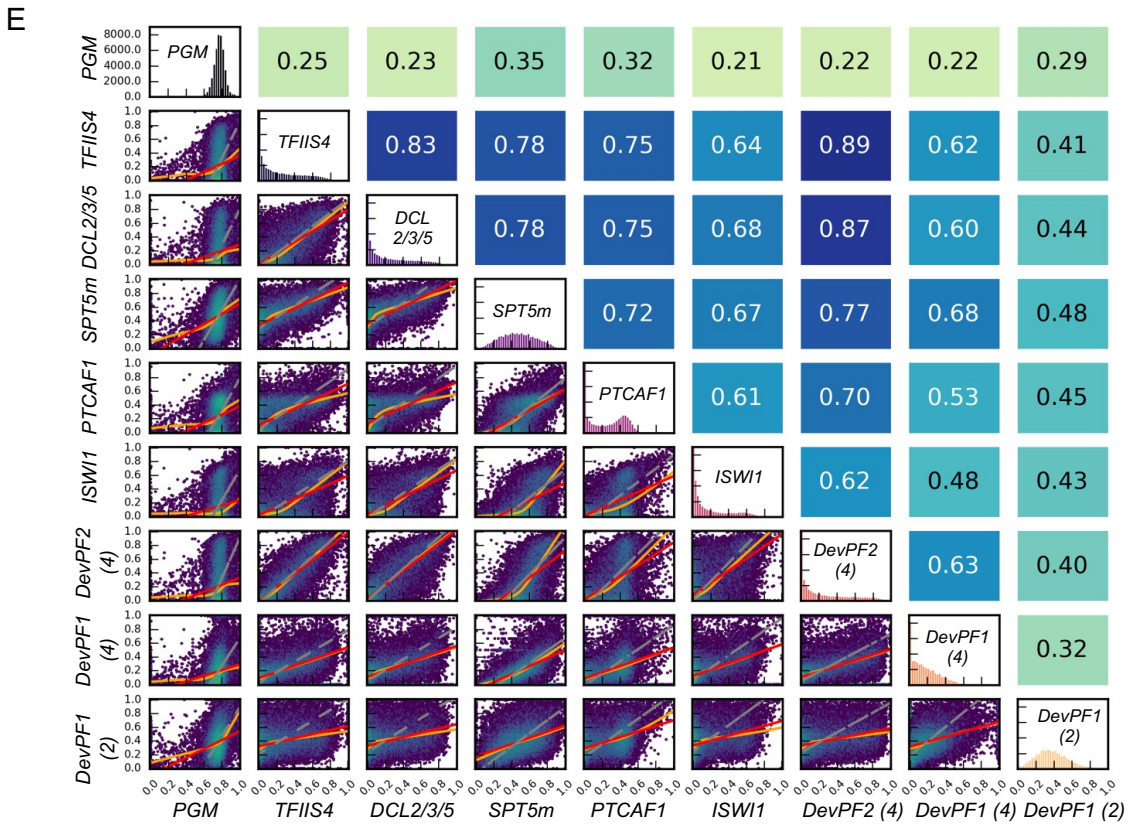


Figure 5: Effects of *DevPF* knockdowns on genome-wide IES retention

(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).

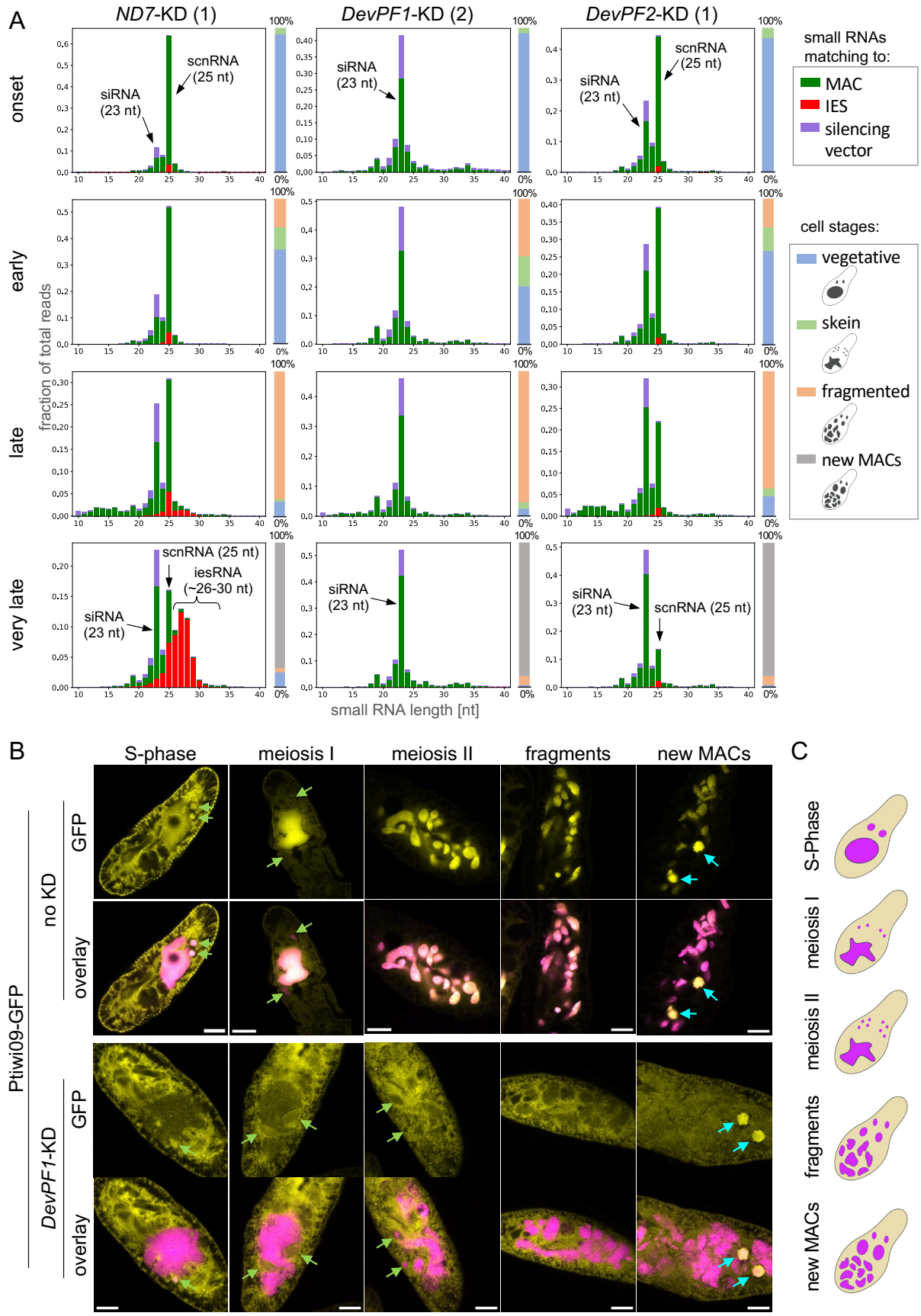
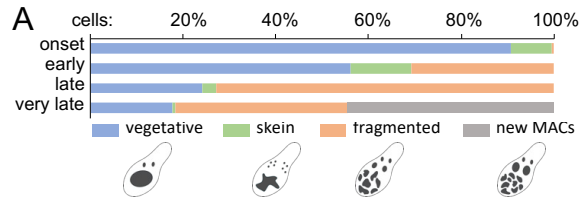


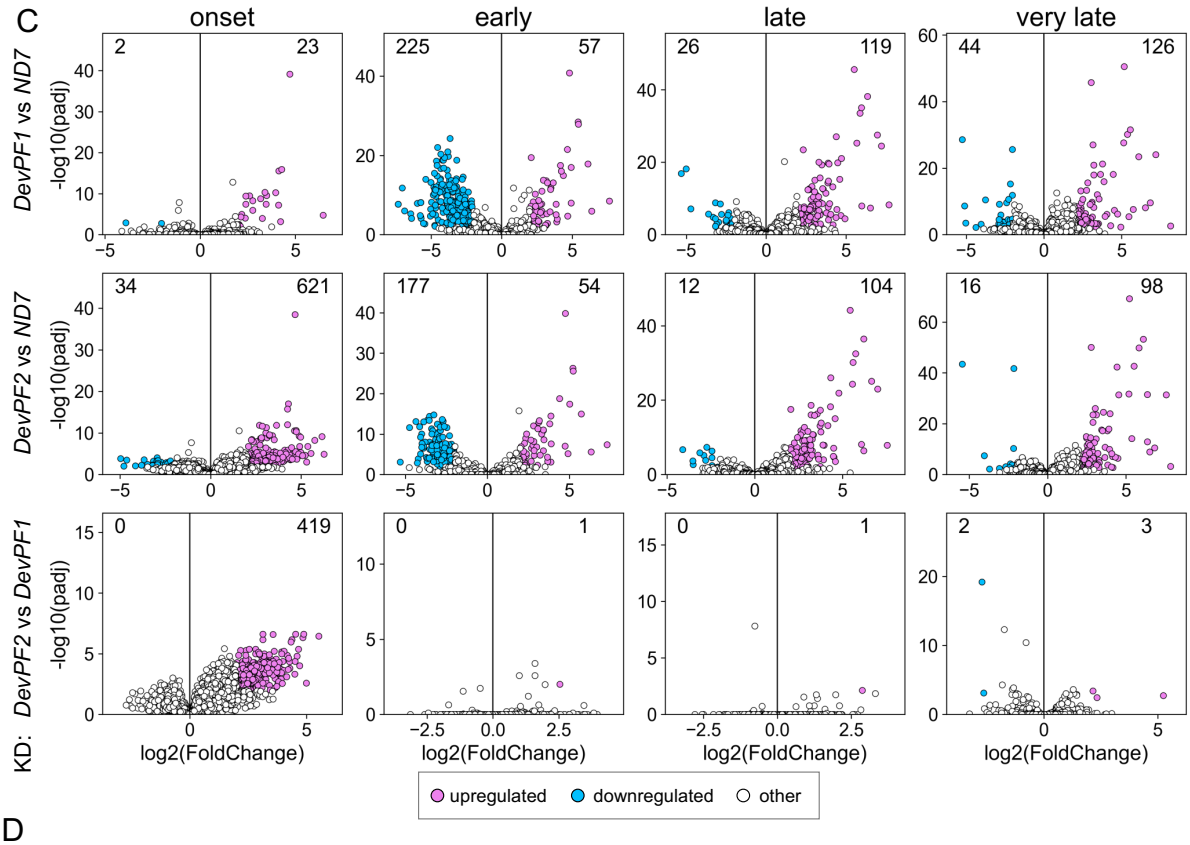
Figure 6: Changes of small RNA populations upon *DevPF* knockdowns

(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

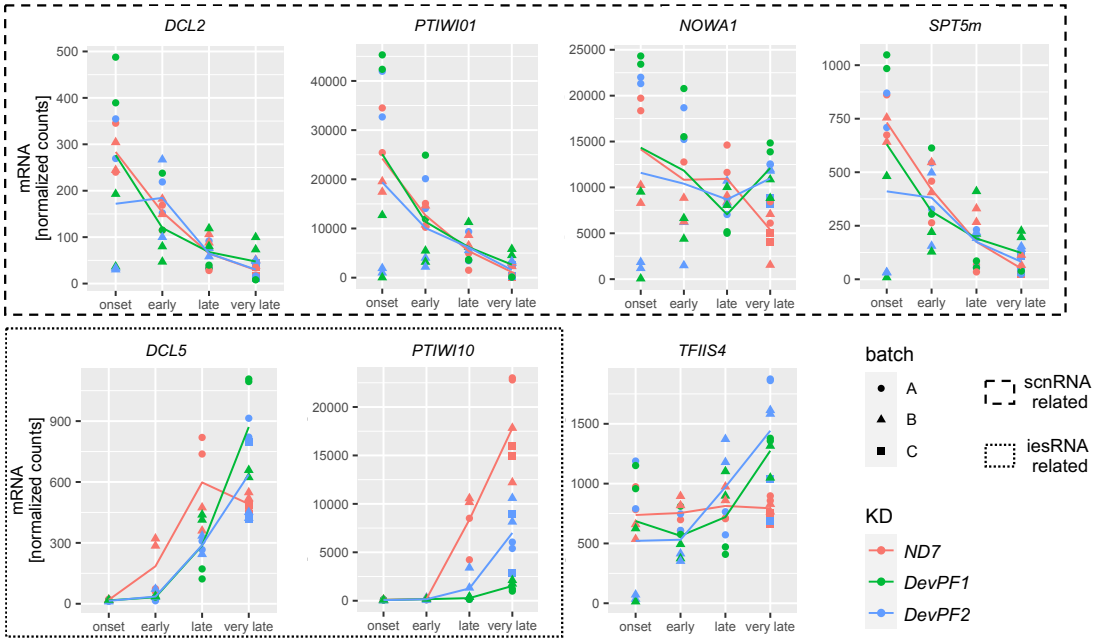


Figure 7: Differential gene expression in *DevPF* knockdowns

(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