Inter-generational nuclear crosstalk links the control of gene expression to programmed genome rearrangement during the Paramecium sexual cycle

Mélanie Bazin-Gélis, Evangelia Eleftheriou, Coralie Zangarelli, Gaëlle Lelandais, Linda Sperling, Olivier Arnaiz, Mireille Bétermier

Nucleic Acids Research, 2023, 1–15
https://doi.org/10.1093/nar/gkad1006
Genomics

Inter-generational nuclear crosstalk links the control of gene expression to programmed genome rearrangement during the *Paramecium* sexual cycle

Mélanie Bazin-Gélis [1], Evangelia Eleftheriou [1,2], Coralie Zangarelli [1], Gaëlle Lelandais [1], Linda Sperling [1], Olivier Arnaiz [1,*] and Mireille Bétermier [1,*]

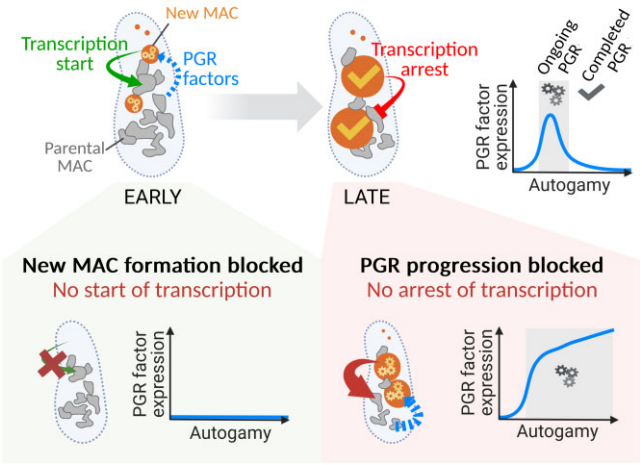
[1]Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France
[2]Institut Pasteur, Université Paris Cité, Inserm U1223, Innate Immunity Unit, Paris, France

[*]To whom correspondence should be addressed. Tel: +33 1 69 82 43 75; Email: olivier.arnaiz@i2bc.paris-saclay.fr
Correspondence may also be addressed to Mireille Bétermier. Email: mireille.betermier@i2bc.paris-saclay.fr

Abstract

Multinucleate cells are found in many eukaryotes, but how multiple nuclei coordinate their functions is still poorly understood. In the cytoplasm of the ciliate *Paramecium tetraurelia*, two micronuclei (MIC) serving sexual reproduction coexist with a somatic macronucleus (MAC) dedicated to gene expression. During sexual processes, the MAC is progressively destroyed while still ensuring transcription, and new MACs develop from copies of the zygotic MIC. Several gene clusters are successively induced and switched off before vegetative growth resumes. Concomitantly, programmed genome rearrangement (PGR) removes transposons and their relics from the new MACs. Development of the new MACs is controlled by the old MAC, since the latter expresses genes involved in PGR, including the *PGM* gene encoding the essential PiggyMac endonuclease that cleaves the ends of eliminated sequences. Using RNA deep sequencing and transcriptome analysis, we show that impairing PGR upregulates key known PGR genes, together with ∼600 other genes possibly also involved in PGR. Among these genes, 42% are no longer induced when no new MACs are formed, including 180 genes that are co-expressed with *PGM* under all tested conditions. We propose that bi-directional crosstalk between the two coexisting generations of MACs links gene expression to the progression of MAC development.

Graphical abstract



Introduction

Eukaryotic cells generally harbor a single nucleus that serves two major functions: gene expression and transmission of genetic information to post-mitotic daughter cells or post-meiotic gametic cells. However, many examples of multinucleate cells have been documented, such as the muscle fiber of animals, the syncytium of filamentous fungi or the early *Drosophila* embryo (1,2). In these giant cells, *in situ* hybridization or single-nucleus RNA sequencing have pointed to heterogeneous transcriptional behavior among individual nuclei or nuclear subtypes occupying different regions of the cytoplasm (3–5). Whether and how the different nuclei of a multinucleate cell coordinate their respective transcriptional activities have been studied in a few cell types, particularly in

Received: April 17, 2023. **Revised:** October 9, 2023. **Editorial Decision:** October 11, 2023. **Accepted:** October 23, 2023

syncytial *Drosophila* embryos (6). In the latter, gradients of maternal mRNA-encoded factors along the antero-posterior axis drive the zygotic activation of spatially distinct combinations of transcription factors, which defines segmental patterning along the embryo (7). Later in development, following cellularization, the identity of each segment is sequentially specified by transcription factors expressed from the *Hox* homeotic gene clusters (8). The regulatory mechanisms at work in other multinucleate cells remain overall poorly understood.

Ciliates are unicellular eukaryotes, in which germline and somatic functions are partitioned between distinct types of nuclei that coexist in the same cytoplasm (9). Gene transcription takes place in the somatic macronucleus (MAC), a highly polyploid nucleus that is essential for cell survival. The diploid germline micronucleus (MIC), although transcriptionally silent, plays a crucial role in sexual processes: it undergoes meiosis and transmits the germline genome to the zygotic nucleus of the next generation. At each sexual cycle, the old (parental) MAC is destroyed, and in the same cell, two new MACs develop from mitotic copies of the zygotic nucleus. MAC development has been extensively studied in ciliate models (10), including *Paramecium* (11). In *Paramecium tetraurelia*, two sexual processes have been described: conjugation between a pair of reactive partners with complementary mating types, and autogamy, a self-fertilization that takes place in a single cell upon starvation. At the start of the sexual cycle, MIC meiosis is accompanied by fragmentation of the old MAC into ∼30 fragments that persist in the cytoplasm and continue to be transcribed throughout new MAC development (12). Gene transcription progressively switches from the old to the new MACs to ensure progeny survival when vegetative growth resumes and old MAC fragments are lost. A large-scale analysis of the *P. tetraurelia* transcriptome during autogamy established that the expression pattern of 17 190 genes (out of 41 533) follows a developmental program, leading to the identification of six clusters of differentially expressed genes (13). An early expression peak (1974 genes) is observed during MIC meiosis (MEI) and old MAC fragmentation (FRG: around the T0 time-point, defined as the time when ∼50% of cells in the population have a fragmented old MAC). Two successive intermediate (2037 genes) and late (468 genes) expression peaks are detected during MAC development: at DEV1 (around T5, with time measured in hours) and DEV2/3 (T10 to T20) stages, respectively. They are followed by the onset of a late induction cluster (3741 genes) at DEV4 (T20 to T30). The gene expression program also includes early and late repression clusters (4536 and 4434 genes, respectively).

In *P. tetraurelia*, MAC development involves successive rounds of genome endoduplication to reach the final ploidy (∼1000 to 1600*n* (14)) of the mature new MAC (12). For the most part, programmed genome rearrangement (PGR) starts early during the DEV1 stage, when the new MAC genome has reached a 32n ploidy, and extends over several consecutive endoduplication rounds (15). In all, PGR removes at least 30% of germline DNA (reviewed in (11)), reducing the haploid genome complexity from 100 to 150 Mb in the MIC (16) to 72 to 76 Mb in the mature MAC (17,18). Eliminated germline sequences include multiple DNA repeats (transposable elements (TE) and minisatellites) that are removed imprecisely, leading to the formation of variable MAC chromosome ends capped by *de novo* telomere addition, and/or to internal deletions (19). In addition, around 45 000 Internal Eliminated Sequences (IES), mostly short, non-coding and originating from ancient TEs, are excised from new MAC chromosomes (16,20) according to a defined temporal program (15). IES ends present a poorly conserved and degenerate consensus sequence, which includes the invariant TA dinucleotide found at IES boundaries with flanking DNA (20,21). Since 47% of genes are interrupted by at least one IES in the MIC, precise IES excision is necessary to assemble functional genes in the new MACs. Thus, PGR not only shapes the structure of MAC chromosomes, but also rearranges the somatic genome to make it suitable for gene expression.

Numerous studies have focused on the search for essential actors involved in PGR. A first line of research has unraveled the role of non-coding RNAs (22–25) in guiding the Ezl1 histone-methyl transferase (26), a component of the *Paramecium* Polycomb repressive complex PRC2 (27,28), to germline-restricted sequences that are eventually eliminated from the developing MAC. Of note, 25% of IESs with stronger nucleotide sequence signals at their ends relative to the general consensus, are excised with high efficiency and precision, independently of non-coding RNAs and heterochromatin marks (15), which suggests that other factors contribute to their recognition. A second line of research has aimed at identifying components of the core IES excision machinery, which are also required for most imprecise DNA elimination (18). A complex of six domesticated PiggyBac transposases, including the catalytically active PiggyMac (Pgm) subunit (29,30) and five Pgm-like (PgmL) partners (31), has been proposed to form the DNA cleavage complex that introduces DNA double-strand breaks (DSB) at IES ends (32). Once IESs are removed, broken excision sites are repaired through non-homologous end joining (NHEJ): following limited processing of free DNA ends, the NHEJ-specific LigaseIV-Xrcc4 complex stitches together the two IES-flanking sequences (33). In addition to playing its expected role in DSB repair, the specialized Ku70/Ku80c heterodimer appears to be part of the DNA cleavage complex itself: it interacts with Pgm, stabilizes its localization in the developing new MACs and is indispensable for DSB introduction at IES ends (34,35).

Most of the known proteins that participate in the epigenetic control or catalysis of PGR are encoded by genes from the early or intermediate expression peaks, which are switched off at the end of autogamy in control conditions (13,36). Previous observations suggested that depleting essential actors of PGR may perturb this gene expression program (34). Indeed, northern blot experiments revealed that *KU80c* and *KU70* mRNAs, which are expressed in the intermediate peak, accumulate continuously until late stages of MAC development in a *PGM* knockdown (KD). Reciprocally, *KU70* and *PGM* mRNAs are upregulated in a *KU80c* KD. To generalize these observations, we studied the *P. tetraurelia* transcriptome during autogamy of cells subjected to three independent KDs of individual genes involved in PGR. We report that impairing the normal course of PGR in the new MACs results in upregulation of 628 genes, among which we found many known PGR genes, suggesting that the list may contain new important candidate genes for PGR. We also show that subsets of genes from the early and intermediate peaks are no longer expressed when new MACs fail to develop. Based on these findings, we propose that crosstalk between the two generations of somatic nuclei that coexist in *Paramecium* cells during

autogamy establishes a regulatory feedback loop that links the transcription program in the old MAC to PGR progression in the new MACs.

Materials and methods

Paramecium strains and culture conditions

Paramecium tetraurelia 51 new (32) or its mutant derivative 51 *nd7-1* (30) were grown at 27°C in a wheat grass infusion medium inoculated with *Klebsiella pneumoniae* (*Kp* medium) and supplemented with 0.8 μg/ml β-sitosterol and 100 μg/ml ampicillin (37).

Microinjection of *GFP* reporter transgenes

All *GFP* reporter transgenes were constructed from the *PGM-GFP* and *GFP-PGM* constructs published in (30). In plasmid p0349, the *GFP* gene was cloned under the control of *PGM* transcription signals (96 bp upstream of the start codon and 54 bp downstream of the stop codon) (Figure 1A). The wildtype promoter of p0349 was then modified using synthetic DNA fragments (from Integrated DNA Technologies) to give plasmids p0364 and p0418, carrying a mutant or an inverted motif in the *PGM* promoter, respectively. All plasmids were linearized by BsaAI (New England Biolabs) and co-injected into the MAC of vegetative *P. tetraurelia* 51 *nd7-1* with a small amount (1:10 mass ratio) of a linearized *ND7*-complementing plasmid, as described in (30). Transformants were screened for trichocyst discharge (trich+) after 2 to 3 cell divisions. Transformed cells were grown in *Kp* medium for 8–10 vegetative divisions and their genomic DNA was extracted using the NucleoSpin Tissue kit (Macherey-Nagel). Transgene injection levels (copy number per haploid genome, cphg) were determined using real-time qPCR (see detailed information in qPCR_MIQE_checklist.xlsx file deposited in Zenodo). Oligonucleotide primers used to amplify *PGM* and *GFP*, with *KU80c* as a genomic reference, are listed in Supplementary Table S1. Since there is no endogenous *GFP* gene in *Paramecium*, we used genomic DNA extracted from cells transformed by the *PGM-GFP* fusion as a standard to quantify *GFP* DNA amounts, with a cphg = 6 based on Southern blot hybridization (30) and qPCR using *PGM* primers. For all *GFP* transformants, we calculated the apparent cphg from the Cp value provided by real-time qPCR:

$$cphg_{apparent} = 2^{(Cp_{genomic\ ref} - Cp_{GFP})}$$

In each experiment a correction factor (*f*) was calculated from the Cp value obtained with the *GFP* standard:

$$cphg_{standard} \times f = 6 \Leftrightarrow f = \frac{6}{cphg_{standard}}$$

The correction factor f was used to correct the value of the *GFP* injection level (cphg) of transformed cells.

$$cphg = cphg_{apparent} \times f$$

Detailed information on the calculation of cphg from Cp values has been deposited as an excel file in Zenodo (qPCR_to_select_cphg.xlsx).

RNAi-mediated gene knockdown (KD) during autogamy

Double-stranded RNA (dsRNA)-induced RNAi was carried out using the feeding procedure (38,39). Transformed or non-injected *Paramecium* cells were grown in *Kp* medium for four vegetative divisions, then transferred to wheat grass infusion medium containing plasmid-free *Escherichia coli* HT115 for 4–5 divisions (40). RNAi was induced as described in (30). Each culture was split and fed for 4–5 divisions with non-induced HT115 containing RNAi plasmids (against target or control genes), then with dsRNA-producing HT115 for 10–15 more divisions before cells starve and start autogamy. Final volumes were 200 ml for medium-scale or 1 l for large-scale experiments. Autogamy was triggered by starvation (41) and the progression of old MAC fragmentation was monitored using DAPI staining of fixed cells (35). In all autogamy time-courses, the T0 time-point is defined as the time when 50% of cells have a fragmented MAC. Three days after autogamy has started, the ability of sexual progeny to develop a functional new MAC was monitored by performing a survival assay from 30 individual autogamous cells, as described (30) (Supplementary Figure S1, see Numerical_data for Figures.xlsx file in Zenodo). Control experiments were performed using plasmid p0ND7c, which targets RNAi against the non-essential *ND7* gene (42). The plasmids used for *PGM*, *KU80c*, *XRCC4* and *CtIPa* + *CtIPb* (= *CtIP)* RNAi were pPBL49HN (29), pL4440-KU80C-2 (34), pXRCC4-R (33), pL4440-PtCtIPa and pL4440-PtCtIPb (43), respectively. Plasmids used to target *DCL2* and *DCL3* were from (23), the sequence of plasmid p0474_L4440_Dcl5 used for *DCL5* RNAi has been deposited in Zenodo.

Cell collection and protein extraction for western blot analysis

50-ml aliquots (~1 to 2 × 10^5 cells) were collected from medium-scale cultures at different autogamy time-points. Cells were washed for 20 min in Dryl's buffer (2 mM sodium citrate, 1 mM NaH_2PO_4, 1 mM Na_2HPO_4, 1 mM $CaCl_2$) before concentration in a final volume of ~40 μl. 20-μl aliquots were flash-frozen in liquid nitrogen and stored at –80°C. For western blots, 20 μl of frozen cells were lysed by a 3-min incubation at 100°C in 5% boiling SDS with 2× final concentration of Protease Inhibitor Cocktail Set 1 (Merck Chemicals). After centrifugation (5 min 15 000g), Laemmli sample buffer (Bio-Rad) with β-mercaptoethanol (354 mM final concentration) was added to the supernatant before a 3-min incubation at 100°C. Electrophoresis was performed in Criterion TGX Stain-free 4–15% precast SDS-Page gels (Bio-Rad) in migration buffer (25 mM Tris-base, 200 mM glycine, 0.1% SDS). Following migration, proteins were electro-transferred to 0.45 μm NC Protran nitrocellulose blotting membranes (Amersham) in transfer buffer (20 mM NaH_2PO_4, 20 mM Na_2HPO_4, pH 6.7).

Pgm was immuno-detected using α-Pgm 2659-GP guinea pig polyclonal antibodies (1:5000 or 1:1000) (30), GFP using the rabbit polyclonal antibody PABG1 (1:2000, ChromoTek), tubulin using the anti-alpha tubulin monoclonal antibody TEU435 (1:10 000 or 1:5000, kindly provided by A. Aubusson-Fleury and A.M. Tassin) (44). Antibodies were diluted in TBST (50 mM Tris–HCl pH 8, 150 mM NaCl, 0.5% Tween) containing 4% low fat milk powder. Primary antibodies were revealed using species-appropriate HRP-conjugated secondary antibodies (anti-rabbit or anti-mouse IgG (1:5000, Promega), or anti-guinea pig IgG (1:5000, Thermo Scientific)). For each antibody, the signal was acquired using the Chemi-Doc Touch Imaging System (Bio-Rad) and quantified with the

Image Lab software (Bio-Rad), and GFP or Pgm signals were normalized relative to alpha tubulin. Unless otherwise specified, the normalized intensity at T6 was set to 1 as a reference to quantify protein amounts in each time-course. All numerical data can be found in Zenodo (Numerical_data for Figures.xlsx file).

RNA sequencing

All RNA preparations used for sequencing were obtained from published autogamy time-course experiments, in which the following genes were knocked down using RNAi: *EZL1* and *ICL7* control (26), *PGM*, *KU80c* & *ND7* control (= ND7_K) (34), *CtIP* (43), *XRCC4* and two independent *ND7* controls (=ND7_X and ND7_L) (33).

RNAseq data from *EZL1* and *ICL7* KDs were published previously (Supplementary Table S2) (27,45). Poly A RNAs extracted from KU80c, PGM and ND7_K time-courses were sequenced from oriented Illumina TruSeq libraries using a HiSeq sequencer (2 × 100 bp paired end), as described (20). For the XRCC4, CtIP, ND7_X and ND7_L time-courses, RNA sequencing was performed on a NextSeq 500 sequencer (2 × 75 bp paired end), as described (27) (Supplementary Table S2).

Reference genomes

The sequencing data were mapped on *P. tetraurelia* strain 51 MAC v1 (ptetraurelia_mac_51.fa) and MAC + IES v1 (ptetraurelia_mac_51_with_ies.fa) reference genomes using Hisat2 (v2.1.0, –rna-strandness FR –min-intronlen 20) (13). Gene annotation v2.0 (ptetraurelia_mac_51_annotation_v2.0.gff3) and IES annotation v1 (internal_eliminated_sequence_PGM_ParTIES.pt_51.gff3) were used in this study (13).

Estimation of IES retention from DNAseq reads in an *XRCC4* KD

For DNA sequencing, developing new MACs were collected at T30 (30 hours after T0) from 500 ml of large-scale cultures of *P. tetraurelia* 51 new grown in *XRCC4* or *ND7* RNAi-inducing media and enriched by centrifugation through a 2.1 M sucrose layer (15,20). Total DNA from flow cytometry-sorted new MACs (Supplementary Figure S2A) was sequenced on a NextSeq 500 sequencer as described (15) (Supplementary Table S2). To monitor the excision status of each IES, we could not use the standard IES retention score (46) because calculating that score relies on counting *de novo* IES– junctions, which are not formed after DSB introduction in Xrcc4-depleted cells (33). IES retention was therefore evaluated by counting the number of reads mapping on each IES left boundary using ParTIES (MIRET module v1.05 default parameters) (46), normalized by total read coverage per Mbp of the MAC + IES genome (CPM, counts per million) (Supplementary Table S3). As a control, a sequencing dataset with 100% IES retention was simulated using ART_Illumina (v2.5.8 -l 80 -ss HS10 -p -qs 90 -qs2 90 -m 250 -s 50) on the MAC + IES genome, with the same sequencing depth as the *XRCC4* KD dataset, and IES boundary coverage was estimated from the simulated data using the method described above (Supplementary Table S3, Supplementary Figure S2B).

Differential expression (DE) analysis

For each RNAseq dataset, raw fragment counts were determined for each gene using htseq-count (v0.11.2, –stranded = yes –mode = intersection-nonempty) on filtered BAM files (samtools v1.3.1) as described (13). Version 4 of R (47) was used for data analyses and to generate images (ggplot2 v3.3.6; VennDiagram v1.7.3) (48,49). Batch effects due to differences in RNA extraction conditions and sequencing protocols (HiSeq or NextSeq) were corrected using ComBat-seq (sva v3.38.0, R package, with no group condition) (50). The relevance of this procedure was verified taking advantage of available technical replicates from the EZL1 and ICL7 time courses that were sequenced twice, using either HiSeq (45) or NextSeq (27) (Supplementary Figure S3, Supplementary Table S4). We then corrected all RNAseq datasets (except for the EZL1 time course, see DESeq2_metadata.xlsx file in Zenodo) as the first step of our analysis workflow. For the ICL7 time course, the time points sequenced both by HiSeq and NextSeq were collapsed, as recommended by DESeq2 for technical replicates, before completing a first round of DESeq2 analysis (DESeq2 v1.30.1, R package) (51). This first round allowed us to obtain normalized counts for all samples, using a blind variance stabilizing transformation (vst) (Part 1 of Supplementary Figure S4, see Expression_data.xlsx file in Zenodo). Identification of differentially expressed genes could not be performed at this stage of the analysis, because 'real' replicates (technical or biological) were available only for control KDs (ND7 and ICL7 time-courses). Hierarchical clustering (derived from a Pearson correlation matrix, stats v4.0.3, R package) and the first two dimensions of principal component analysis (PCA, FactoMiner v2.4, R package) were used to identify groups of similar samples corresponding to 4 autogamy stages for each RNAi condition, in an unsupervised way (45). The resulting groups were manually curated based on cytological observations of autogamous cells (Supplementary Figure S5). Our dataset, therefore, was organized in different types of replicates (Supplementary Figure S6). In each time-course, the different time-points that were attributed to the same autogamy stage were considered as 'pseudo-replicates': this means that, even if they are not rigorously biological replicates, they share a high level of similarity, which can be used for the identification of differentially expressed genes. In addition, for the four biological replicates of the control time courses (ICL7 and ND7), all samples (biological and pseudo-replicates) corresponding to the same autogamy stage were grouped together for DE analysis. The results of DE analysis were extracted using both RNAi conditions and autogamy stages as contrast (Part 2 of Supplementary Figure S4). We considered a gene to be deregulated if its fold-change reached at least 2 compared to controls with an adjusted *P*-value (padj, calculated by DESeq2) lower than 0.05 (Supplementary Table S5, see Deregulation_all_genes.xlsx file in Zenodo).

Identification of a conserved motif

We used STREME 5.4.1 (52) to search for 5 to 20-bp sequence motifs (with a *P*-value < 0.05) that are enriched in the promoters (defined as the 150-bp window upstream of the annotated transcription start site (TSS), see (13)) of the 180 genes that are coregulated with *PGM* (Co-*PGM* genes: intermediate peak genes, upregulated in *XRCC4*, *PGM* and *KU80c* KDs and downregulated in a *CtIP* KD). The

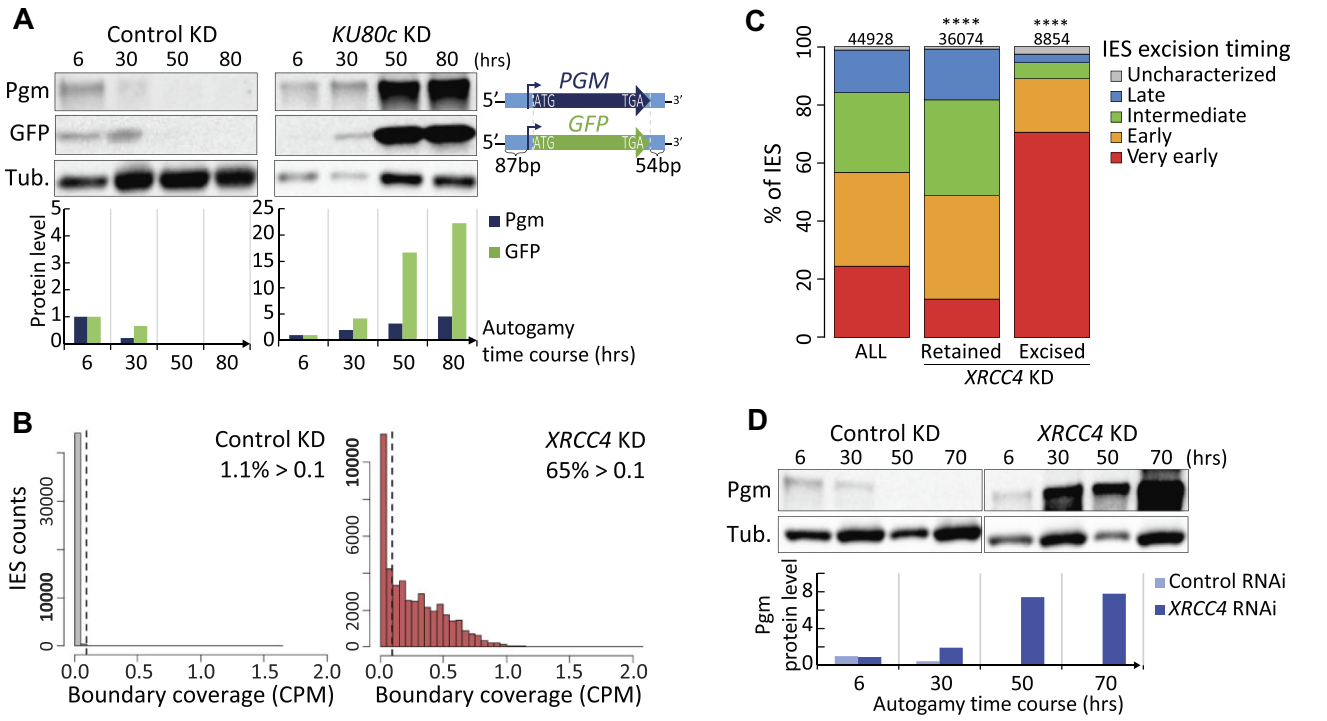


Figure 1. Pgm overexpression in *KU80c* and *XRCC4* KDs. (**A**) Western blot quantification of the levels of endogenous Pgm and transgene-encoded GFP during an autogamy time-course of cells subjected to control or *KU80c* RNAi. The GFP is expressed from a micro-injected transgene (1976 copies per haploid genome (cphg)) under the control of *PGM* transcription signals. (**B**) Evidence for IES retention in an *XRCC4* KD. The histograms represent the number of IESs in each 0.05-CPM (counts per million) interval of normalized IES boundary coverage, following control (gray) or *XRCC4* (red) RNAi. The % in each panel indicates the fraction of IESs with a normalized IES boundary coverage > 0.1 CPM. In the simulated dataset with full IES retention, 96.2% of IESs had a boundary coverage > 0.1 CPM (Supplementary Figure S2B). Note that y-axes are not to the same scale in both panels. (**C**) Repartition among the excision timing groups (15) of the IESs with a boundary coverage >0.1 (Retained) or <0.1 CPM (Excised) in *XRCC4* KD compared to all IESs (ALL) (**** *P*-value < 10^{-200}, chi^2 test). The number of IESs in each category is given above the plots. (**D**) Western blot quantification of endogenous Pgm levels during an autogamy time-course of cells subjected to control or *XRCC4* RNAi. The level of Pgm at T6 in the control RNAi is set to 1 as a reference for all lanes. In panels A and D, time-points are in hours (hrs) following T0. Normalized values (in arbitrary units) are plotted below each western blot.

search was repeated 5 times against different control sets of the same number of randomly-picked gene promoters. The .meme files corresponding to the first hits of each STREME round were aligned and trimmed to the shortest motif size (18 bp) (trim_motifs, universalmotif v1.12.3, R package). A consensus motif was defined by merging the five motifs (merge_motif, universalmotif v1.12.3, R package). For each round, STREME defined an enrichment factor as the fraction of genes with at least one match to the consensus in their promoter. The significance threshold for further screening was based on the mean of the five enrichment factors.

Analysis of the motif in gene promoters

To identify all the genes that carry the motif in their promoter, we used the FIMO 5.4.1 software (53) with default settings. We first extracted the data obtained for the set of 180 Co-*PGM* genes and defined a cut-off p-value that matches the significance threshold calculated from STREME enrichment factors (defined above). We then applied the resulting cut-off to the full list of motifs identified by FIMO.

To compare the position of the motifs in their respective promoter, we considered the position of the first 5′ nucleotide of the motif relative to the TSS and counted the number of motifs found in successive 5-bp windows within gene promoters.

Results

Inhibition of programmed genome rearrangement triggers overexpression of the Pgm endonuclease

Following up on previous experiments (34), we confirmed on western blots that the observed upregulation of *PGM* mRNA in a *KU80c* KD results in accumulation of Pgm at late developmental stages (Figure 1A, Supplementary Figure S1A for survival test). To investigate whether the deregulation takes place at the RNA level (transcriptional or translational) only, or also at the protein level (post-translational), a *GFP* reporter transgene expressed under the control of *PGM* transcription start and termination signals was micro-injected into the MAC of vegetative *P. tetraurelia* cells. GFP levels were monitored on western blots after the onset of autogamy (Figure 1A). We observed that accumulation of GFP in a *KU80c* KD parallels that of endogenous Pgm, indicating that deregulation of *PGM* expression (i) mostly takes place in the old MAC and (ii) is likely not at the protein level. We cannot, however, ascertain whether deregulation is transcriptional or post-transcriptional, since the mRNA produced from the transgene carries both the 5′ and 3′ UTRs of *PGM*.

Pgm and Ku70/Ku80c are essential for DNA cleavage upon initiation of IES excision (29,34). As a consequence, RNAi-mediated silencing of either *PGM* or *KU80c* causes the retention of all IESs (20,35). To examine whether interfering with DSB repair after IES end cleavage, a step that is

necessary for genome assembly in the new MACs, would also upregulate *PGM*, we knocked down *XRCC4*. Previous molecular analyses performed on a handful of IESs established that DSBs introduced at IES ends in an *XRCC4* KD are not repaired, resulting in the persistence of broken ends and a failure to assemble IES- junctions (33). At the genome-wide level, high-throughput DNA sequencing from purified new MACs (Supplementary Figure S2A) revealed that 65% of IESs were retained at the T30 time-point of autogamy (i.e. 30 h after the T0 time-point defined in Materials and Methods) in an *XRCC4* KD, in contrast to a control KD (Figure 1B, Supplementary Table S3). Of note, this set of retained IESs contains a lower proportion than expected of the IESs that are excised the earliest during autogamy (very early IESs) (15) (Figure 1C). Reciprocally, very early excised IESs are overrepresented among the 8854 IESs exhibiting little or no retention. This shows that inhibiting DSB repair after the first IESs are cut out of chromosomes impinges on the normal progression of PGR and results in late IES retention. As shown on western blots, we observed a strong increase in Pgm levels in an *XRCC4* KD relative to control conditions (Figure 1D).

To investigate whether perturbing IES excision at an earlier step than DSB introduction or repair would also deregulate *PGM* expression, we knocked down the *DCL2*, *DCL3* and *DCL5* Dicer-like genes involved in RNA-mediated recognition of eliminated sequences (22,23,26). Previous work from others has established that a *DCL* triple KD leads to the retention of 43% of IESs, and to death of sexual progeny (54). We found that Pgm also accumulates under these conditions at late autogamy stages (Supplementary Figure S7A, B). Thus, disrupting programmed DNA elimination at various stages, including the synthesis of non-coding RNAs responsible for recognizing eliminated sequences (*DCL* triple KD), DNA cleavage, and/or DSB repair (*KU80c* and *XRCC4* KDs), can lead to *PGM* upregulation.

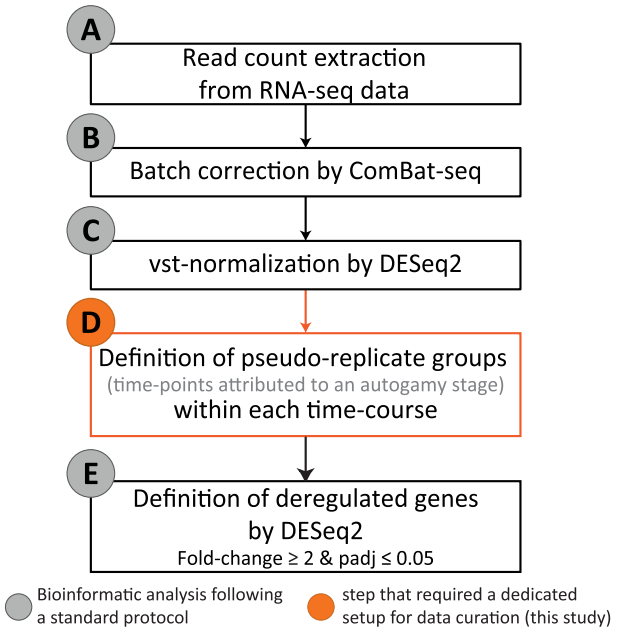


Figure 2. Main steps in RNA-seq data analysis. RNA-seq reads were obtained using Hi-seq or Next-seq sequencing technologies, depending on the collected dataset. Data analysis was divided in five steps: (**A**) read counts were extracted from each dataset using standard protocols (quality control, read mapping and counting); (**B**) batch effects due to the specificities of each sequencing technique were corrected using ComBat-seq (50); (**C**) corrected read counts were normalized using the DESeq2 vst (variance stabilizing transformation) procedure; (**D**) pseudo-replicate groups corresponding to different autogamy stages were defined for each RNAi condition using PCA and hierarchical clustering (Supplementary Figure S6); (**E**) differentially expressed genes were identified using RNAi conditions and autogamy stages as DEseq2 contrasts. Step d (in orange) constitutes a part of the workflow and was tailored for the specific needs of our study. A detailed version of this bioinformatics workflow is displayed in Supplementary Figure S4.

Genome-wide identification of upregulated autogamy genes in *PGM*, *KU80c* and *XRCC4* KDs

The *PGM* gene belongs to a cluster of 2037 differentially expressed genes, called the intermediate expression peak, which are transiently induced during autogamy in control conditions, when PGR takes place in the developing MACs (13). To gain broader insight into how many genes are upregulated at the genome-wide level in the three KDs used in this study to inhibit DNA cleavage or repair during IES excision, we performed a large-scale differential gene expression analysis. We used RNA sequencing data from independent autogamy time-course experiments performed on cells subjected to *KU80c*, *PGM* (34) or *XRCC4* KDs (33) compared to control time-courses. Because of the heterogeneity of the collected datasets, we set up a dedicated bioinformatics workflow to identify differentially expressed genes (Figure 2 and Supplementary Figure S4). We corrected the batch effects and applied principal component analysis and hierarchical clustering to the normalized datasets to group the samples in four developmental stages (Supplementary Figure S6). We observed that each developmental stage (until the LATE stage) included similar autogamy time-points from different conditions, indicating that *KU80c*, *PGM* or *XRCC4* KDs do not affect the progression of autogamy (see also the cytological stages in Supplementary Figure S5). The EARLY, INTERMEDIATE and LATE stages coincide with the time-windows when genes from the early, intermediate and late expression peaks, respectively, reach their maximal expression level (13) (Figure 3A).

To define the sets of genes that are deregulated when PGR is impaired, we focused on the LATE stage, which yielded the largest number of deregulated genes in each KD relative to the controls (Supplementary Table S5). We found that many genes are significantly up- or downregulated at the LATE stage in a *PGM*, *KU80c* or *XRCC4* KD (Figure 3B, Supplementary Table S6A). Fewer upregulated genes were identified in Ku80c- than Pgm-depleted cells, whereas *PGM* and *KU80c* KDs cause similar retention of all IESs (20,35). This is likely due to faster progression of autogamy in this particular *PGM* RNAi experiment (Supplementary Figure S5) leading to a more contrasted difference in transcript levels relative to controls at the LATE stage. In the three KDs, the fold-changes of upregulated genes were dispersed, but the medians of each distribution (2.93, 3.55, 3.35, respectively) were well above the threshold (>2) that we used to define upregulation (Supplementary Figure S7C). In particular, *PGM* mRNA levels increased ∼4.5-fold in *KU80c* and *XRCC4* KDs relative to controls at the LATE stage (Supplementary Figure S7C).

We found that 628 genes are upregulated in all three KDs (Figure 3B, Supplementary Table S6B). We defined this set as 'UP PKX genes' (for 'upregulated genes in *PGM*, *KU80c* and *XRCC4* KDs'). Almost half of the 628 UP PKX genes belong to the early (15%) or intermediate (37%) expression peaks

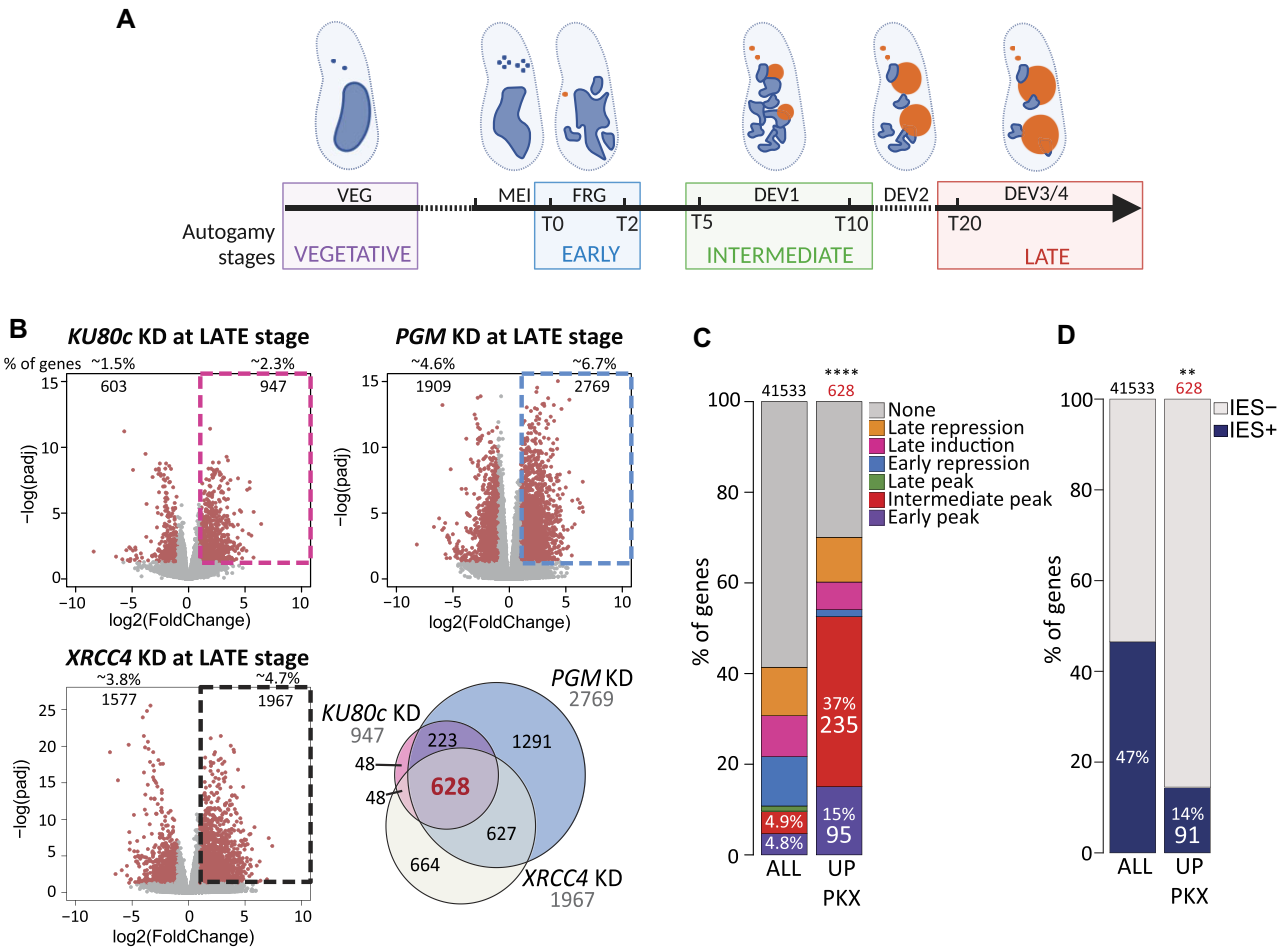


Figure 3. Upregulated genes in *PGM*, *KU80c* and *XRCC4* KDs at the LATE autogamy stage. (**A**) Diagram showing the autogamy stages defined for all time-courses of this study (VEGETATIVE, EARLY, INTERMEDIATE, LATE), and the corresponding stages (VEG, MEI, FRG, DEV1-4) and time-points (in hours) established by Arnaiz *et al.* (13). See Supplementary Figure S6A for details. (Created with BioRender.com). (**B**) Volcano plots showing differentially expressed genes at the LATE stage in *KU80c* (top left), *PGM* (top right) or *XRCC4* (bottom left) KDs relative to the controls. The most deregulated genes are indicated (in red), with their number and percentage relative to all 41533 genes. Upregulated genes are boxed on each plot (P_{adj}: DESeq2 adjusted *P*-value). The Venn diagram shows the intersection of the sets of genes that are upregulated at LATE stages in all three KDs. (**C**) Proportions of genes from each previously identified expression cluster (13) among the different sets of deregulated genes. ALL: all annotated genes; UP PKX: upregulated genes at LATE stage in *PGM*, *KU80c* and *XRCC4* KDs (**** *P*-value $< 10^{-200}$, chi^2 test). (**D**) Percentage of genes without IES (IES–) or carrying at least one IES (IES+) among the two categories described in B (** *P*-value $< 10^{-20}$, chi^2 test).

(13) (Figure 3C), an enrichment that can also be detected in each individual KD (Supplementary Figure S8A–D). We also noted a bias in the germline structure of the LATE upregulated genes, 86% being devoid of any IES compared with 53% for all genes (Figure 3D) and 55% on average for random sets of 628 genes with similar proportions of each expression cluster (Supplementary Figure S8E–H). Despite this bias, we detected an accumulation of intragenic IES+ transcripts at the LATE stage, indicating that transcription from the new MACs takes part in the deregulation (Supplementary Fig S9). The 235 LATE upregulated genes from the intermediate peak (Figure 3C), which represent only ~12% of the expression cluster, include almost all of the genes encoding known components of the DNA cleavage complex (Supplementary Table S6B): *PGML1, PGML2, PGML3a, PGML4a&b, PGML5a&b* (31) and *KU70a* (34). Only *PGM* (29) and*KU80c* (34) are missing since, even if upregulated in the other two KDs, they are not overexpressed in their own respective KD. Our results suggest that mRNAs from genes involved in PGR accumulate (or persist) at LATE stages in *PGM*, *KU80c* or *XRCC4* KDs as a response to defective IES excision in the new MAC.

The developing new MACs control the transcriptional program of a subset of genes throughout MAC development

The shut-off of 330 genes from the early and intermediate expression peaks at the LATE stage of autogamy (95 and 235 genes respectively, see Supplementary Table S6B) seems to critically depend upon the correct completion of PGR in the new developing MACs (Figure 3C), a similar behavior to that of endogenous *PGM* and its reporter *GFP* transgene (Figure 1A and B). This suggests that gene transcription in the old MAC, which drives the production of PGR factors imported in the new MACs, is turned off at the LATE autogamy stage by some signal(s) coming from the new developing MACs, in a regulatory feedback loop.

To investigate whether the new MACs also control gene transcription at earlier stages of autogamy, we examined the production of Pgm in cells depleted in CtIP, an essential protein involved in MIC meiosis (43). In a *CtIP* KD, the formation of the zygotic nucleus is inhibited and no new MAC is formed (Figure 4A, Supplementary Figure S5). In contrast to control conditions, we detected no expression of endogenous

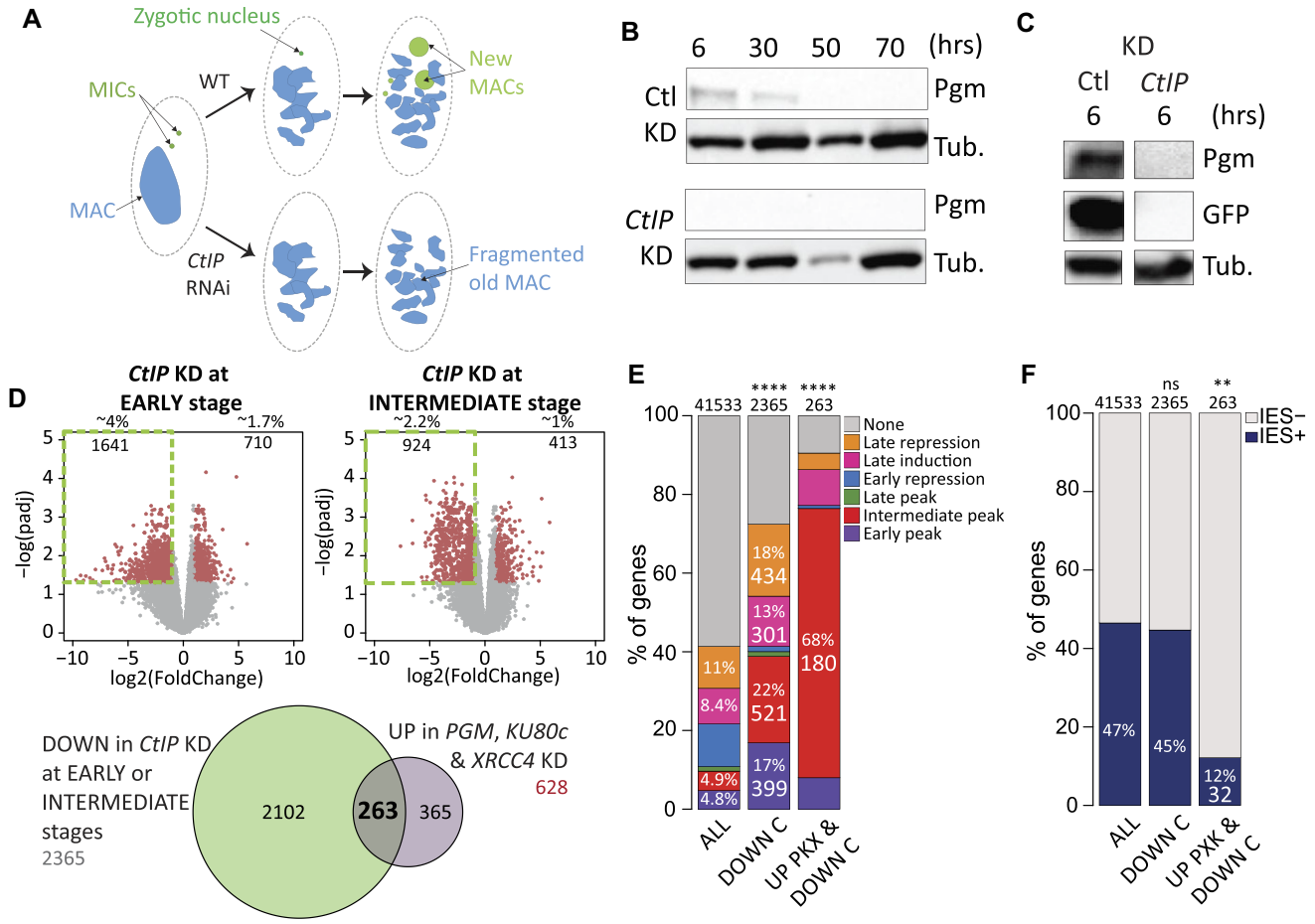


Figure 4. Downregulated genes in a *CtIP* KD at the EARLY or INTERMEDIATE stages. (**A**) *CtIP* RNAi inhibits the formation of the zygotic nucleus and the development of new MACs, while old MAC fragmentation still takes place (43). (**B**) Western blots showing Pgm levels during autogamy time-courses in control (Ctl) or *CtIP* KDs. Time-points are in hours (hrs) following T0. The alpha tubulin (Tub.) was used as a loading control. (**C**) Comparison of the endogenous Pgm and GFP signals on western blots at T6 in an independent experiment using cells harboring the *GFP* reporter transgene (416 cphg), following control or *CtIP* RNAi. The whole membrane is displayed in Supplementary Figure S10A. (**D**) Volcano plots showing differentially expressed genes at the EARLY (left) or INTERMEDIATE (right) stages in a *CtIP* KD relative to the controls (see Figure 3A). Significantly deregulated genes are indicated (in red), with their number and percentage relative to all 41 533 genes. Downregulated genes are boxed (padj: DESeq2 adjusted p-value). The Venn diagram shows the intersection of the sets of genes that are downregulated in a *CtIP* KD at EARLY or INTERMEDIATE stages, and upregulated at LATE stages in all three conditions displayed in Figure 3B. (**E**) Proportions of genes from each previously identified expression cluster (13) among the different sets of deregulated genes. ALL: all annotated genes; DOWN C: downregulated genes at EARLY or INTERMEDIATE stages in *CtIP* KD; UP PKX & DOWN C: intersection of the previously described DOWN C genes and the upregulated genes at LATE stage in *PGM, KU80c* and *XRCC4* KDs (**** *P*-value $< 10^{-200}$, chi^2 test). (**F**) Percentage of genes without IES (IES–) or carrying at least one IES (IES+) among the three categories described in E (** *P*-value $< 10^{-20}$; ns: *P*-value > 0.05 chi^2 test).

Pgm in a *CtIP* KD, whichever autogamy stage was examined (Figure 4B). Following micro-injection of the reporter *GFP* transgene into the old MAC, we observed only a background GFP signal in a *CtIP* KD (Figure 4C and Supplementary Figure S10A, B). Our data show that Pgm is not detectable on western blots when no new MAC develops after the onset of autogamy, suggesting a defect in expression of the *PGM* gene from the old MAC. We therefore extended our study to the whole transcriptome of *CtIP*-silenced cells and performed DE analyses (Figure 2) to identify genes behaving like *PGM*. We searched for genes with a reduced expression level in a *CtIP* KD relative to controls, at the EARLY and INTERMEDIATE autogamy stages (Figure 4D, Supplementary Table S5). This analysis confirmed the reduction of *PGM* expression at the RNA level (Supplementary Figure S10C). More generally, 2365 genes are significantly downregulated at one or the other of these stages (Supplementary Table S6C), with an enrichment in genes from the early or intermediate peaks and from the late induction cluster (Figure 4E and Supplementary Figure S8I), but no significant bias in IES content (Figure 4F and Supplementary Figure S8J). Of note, 263 downregulated genes in a *CtIP* KD at EARLY or INTERMEDIATE stages are also upregulated in *PGM*, *KU80c* and *XRCC4* KDs at LATE stages (Figure 4D, Supplementary Table S6D). Among these, 180 belong to the intermediate peak and thus follow the same expression pattern as *PGM* under all the conditions that we tested in this study (Figure 4E and Supplementary Table S6E). Our data suggest that these 180 genes, hereafter designated as Co-*PGM* genes, are coregulated with *PGM*. Their expression is turned on in the old MAC at the EARLY/INTERMEDIATE stages and remains on until completion of PGR. We propose that regulatory signals originating from the developing new MACs contribute to controlling this expression pattern.

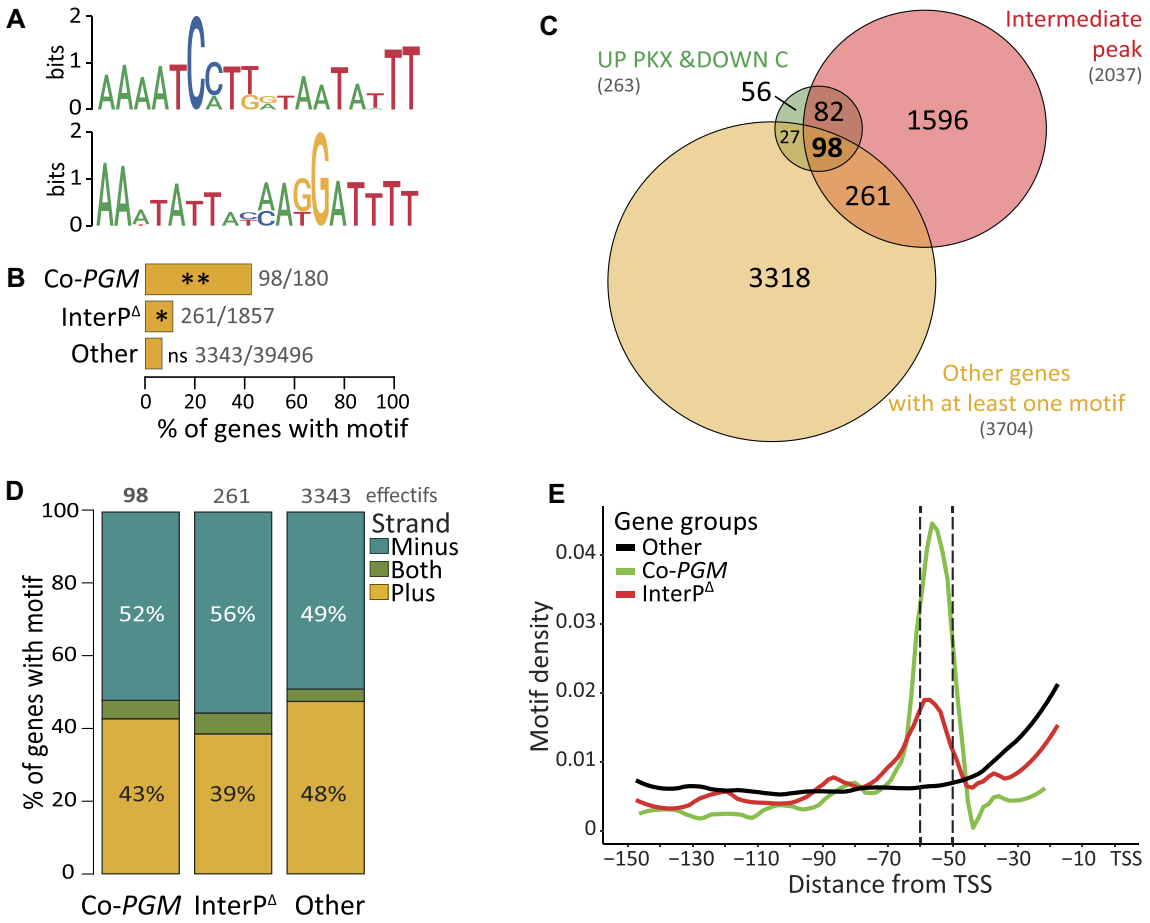


Figure 5. Identification and characterization of an enriched motif in the promoter of Co-*PGM* genes. (**A**) Logo of the 18-bp non-palindromic merged motif (top) and its reverse complement (bottom) identified by STREME in the 150 bp upstream of gene TSSs (Transcription Start Sites). (**B**) The enrichment percentage of genes with at least one motif in their promoter is shown relative to all 41 533 genes (InterP$^{\Delta}$: all intermediate peak genes excluding Co-*PGM* genes; Other: all genes except intermediate peak genes). ** *P*-value < 10^{-20}; * *P*-value < 10^{-10}; ns: *P*-value > 0.05 (chi^2 test). (**C**) Venn diagram showing the intersection of the set of genes that are upregulated genes in *PGM, KU80c* and *XRCC4* KD at LATE stage and downregulated in *CtIP* KD at EARLY or INTERMEDIATE stage (green), with the intermediate peak expression cluster (red) and all genes carrying at least one motif in their promoter (yellow). (**D**) Orientation of the motif in gene promoters for the three gene categories described in (B). (Minus: all motifs are on the minus strand; Plus: all motifs are on the plus strand; Both: the promoter carries at least two motifs in different orientations). (**E**) Motif density as a function of the distance of the 5′ nucleotide from the TSS, for motif-carrying genes from the three categories described in (B). For each category, the number of genes carrying the motif was counted at each position to plot the motif density.

A conserved regulatory motif is enriched in the promoter of co-*PGM* genes

To gain further insight into the mechanisms that may underlie the transcriptional control of Co-*PGM* genes, we used STREME to search for conserved nucleotide sequence motifs in their promoters. We identified a significantly enriched 18-bp non-palindromic consensus motif (5′-AAAATCMTTdWAATAWTT-3′, where M: A or C; d: A, G or T; W: A or T) (Figure 5A), present in either orientation in 98 out of the 180 promoters of Co-*PGM* genes (54%; Figure 5B). At the whole-genome level, a total of 3704 genes carry the motif in one (94% genes) or more (6% genes) copies in their promoter (Figure 5B, Supplementary Table S7): they include the 98 Co-*PGM* genes described above, and 261 other genes from the intermediate expression peak (InterP$^{\Delta}$) that were not deregulated in at least one of the four KDs that we considered in our transcriptome study (Figure 5C). No particular orientation bias was noted for the motif (Figure 5D). However, the position of the motif upstream of the TSS was markedly different for the 98 Co-*PGM* genes, with a significantly higher density in the –60/–50-bp region (Figure 5E).

To address the biological significance of the conserved motif, we focused on the *PGM* gene itself. Indeed, a copy of the motif is present 55 bp upstream of the *PGM* TSS in *P. tetraurelia* and is conserved at the same position in other *Paramecium* species (Figure 6A; Supplementary Figure S11). We mutagenized the motif upstream of our *GFP* reporter transgene and, following micro-injection into the vegetative MAC, we followed on western blots the production of GFP during autogamy of transformed cells harboring either the wild-type or the mutant construct (Figure 6B, Supplementary Figure S12). No GFP production was detected at any stage from the transgene carrying the mutant motif, in the control or the *KU80c* KD, suggesting that the motif is essential for gene transcription. We then tested the effect of changing the orientation of the motif by replacing it with its reverse complement upstream of the *GFP* reporter gene (Figure 6A). In the control, the pattern of GFP production was unchanged during

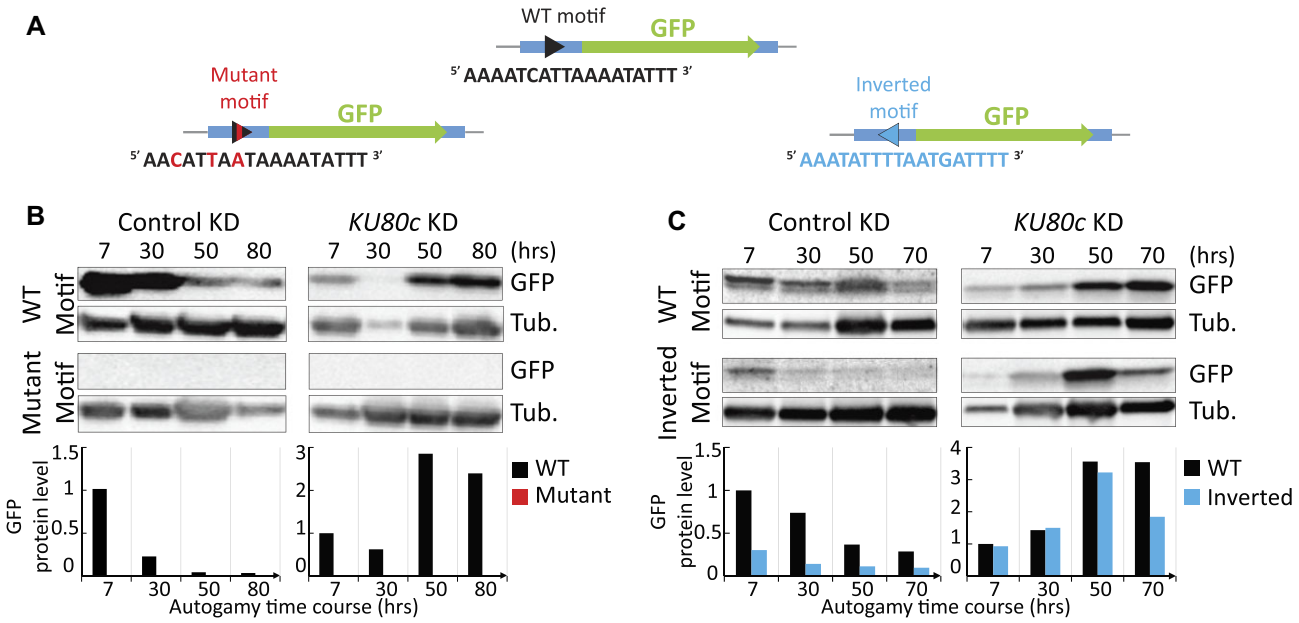


Figure 6. Functional analysis of the motif using *GFP* reporter transgenes. (**A**) Schematic representation of the *GFP* reporter transgenes (see Figure 1A) carrying either the wild-type (WT), mutant or inverted motif in their promoter (see sequences below each diagram). (**B**) Western blot quantification of the levels of endogenous Pgm and the GFP expressed from the reporter constructs (WT: 695 cphg, Mutant: 2524 cphg), during an autogamy time-course of cells subjected to control or *KU80c* RNAi. (**C**) Western blot quantification of the levels of endogenous Pgm and the GFP expressed from the reporter constructs (WT: 204 cphg, Inverted: 188 cphg), during an autogamy time-course of cells subjected to control or *KU80c* RNAi. In (B) and (C), time-points are in hours (hrs) following T0. Normalized values relative to the alpha tubulin (Tub.) signal are plotted below each western blot. In each panel, the level of protein produced at T7 from the WT transgene is arbitrarily set to 1 and used as a reference. Injection levels are indicated (cphg: copies per haploid genome).

autogamy, with a maximum at T7 and a decrease at late stages, even though protein levels were overall lower than with the wild-type construct (Figure 6C, left panels). In the *KU80c* KD, GFP expression from the promoter carrying the inverted motif was upregulated at late autogamy stages, similar to what we observed for the wild-type promoter (Figure 6C, right panel). We conclude from these experiments that the conserved motif is required for *PGM* gene expression, independently of its orientation.

Discussion

Paramecium constitutes an intriguing example of a multinucleate cell, in which structurally distinct nuclei fulfill different functions. In this study, we have focused on a particular stage of the sexual cycle, during which two consecutive generations of somatic nuclei coexist in the same cell. The old (parental) MAC persists in the cytoplasm throughout the sexual cycle and ensures most gene expression, while the new (zygotic) MACs differentiate through a process that involves massive PGR. Previously published conjugation time-course experiments revealed that zygotic transcription starts in the new developing MACs by the time PGR takes place, but the overall contribution of zygotic transcription to total mRNA synthesis remains low (less than 20%), until the first cell division following fertilization (12,41). In autogamous cells, the contribution of zygotic transcription to gene expression and the determination of cell phenotypes may even be lower throughout a 4-day starvation period encompassing our LATE autogamy stage (55). These observations support the notion that the differential expression patterns reported in our transcriptome analysis can, for the most part, be attributed to variations in gene expression from the old MAC. Sequential waves of gene induction and repression were described during autogamy (13). However, what controls the progression of the transcription program during the *Paramecium* sexual cycle is still poorly understood. We show here that abolishing the formation of new MACs (*CtIP* KD) or perturbing the normal course of PGR in the new MACs (*PGM*, *KU80c* or *XRCC4* KDs) induce significant variations of the *P. tetraurelia* transcriptome during autogamy. Our results provide evidence that the new MACs control, at two different levels, the progression of the developmental transcription program in the old MAC.

Developing new MACs control gene expression from the old MAC at EARLY and INTERMEDIATE stages

The first level of *trans*-nuclear control takes place at the EARLY and INTERMEDIATE stages of autogamy. As summarized in Figure 3A, the EARLY stage corresponds to late MIC meiosis and old MAC fragmentation: most of mRNA synthesis, therefore, is essentially carried out in the old MAC. The INTERMEDIATE stage covers the DEV1/2 period of new MAC development, when zygotic transcription starts and IES excision takes place (15): at this stage, the new MACs initiate mRNA production, yet their contribution to overall gene transcription remains minor (see above). Indeed, several published functional complementation experiments using microinjected transgenes confirmed that, consistent with a former study performed during conjugation (12), expression of known PGR genes from the old MAC is sufficient to ensure that functional new MACs develop in the offspring (27,28,30,35,45,56). Conversely, cells harboring a wildtype MIC, but lacking the *PGM* gene in their MAC, fail to un-

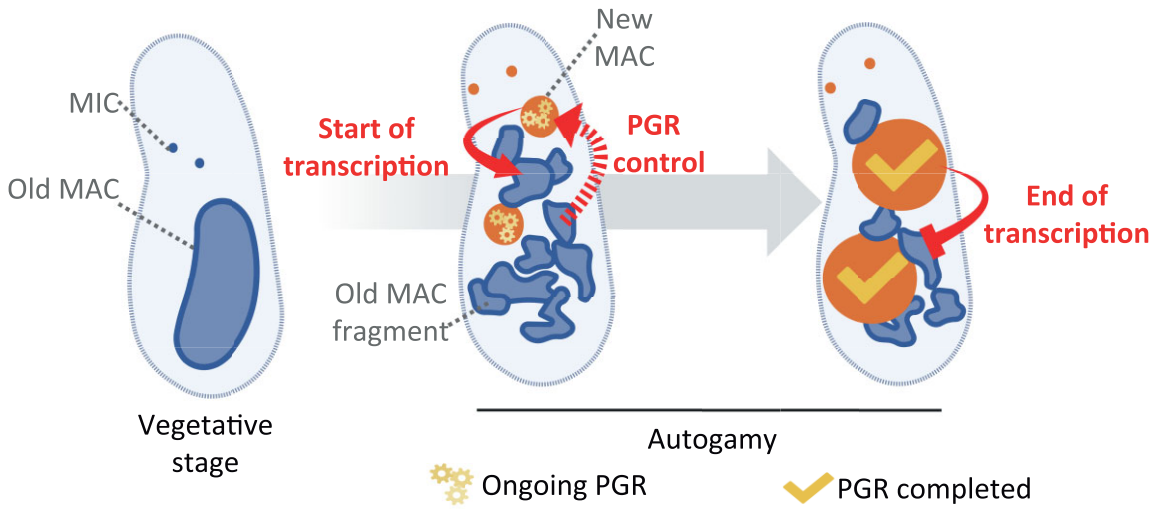


Figure 7. Bi-directional crosstalk between the old and new somatic MACs in *Paramecium*. Upon starvation, vegetative cells (left) become committed to starting their sexual cycle. During autogamy, the MICs undergo meiosis and eventually give rise to the new MICs and MACs (orange), while the parental MAC (blue) is fragmented into ∼30 fragments. Most mRNAs are produced by the old MAC fragments, while programmed genome rearrangement (PGR) takes place in the new MACs. We propose that, at EARLY and INTERMEDIATE stages (middle), the new MACs transmit an inducing signal to old MAC fragments (full arrow) to start the transcription of development-specific genes. This drives the production of proteins that carry out PGR in the new MACs. The old MAC also produces constitutive maternal ncRNAs, which control the determination of sequences that will be eliminated from the new MACs (24). The dotted arrow represents the contribution of the old MAC to the control of PGR in the new developing MACs. In turn, we propose that, at the LATE stage (right), a signal from the new MACs arrests the expression of a subset of genes upon PGR completion (created with BioRender.com).

dergo PGR and, consequently, do not generate functional new MACs after autogamy (30): this underscores that, at least in the case of *PGM*, which has no IES in the non-rearranged genome, transcription from the new MACs alone does not ensure the successful progression of new MAC development. In a *CtIP* KD, the failure to form a zygotic nucleus completely abolishes the development of new MACs, without impinging on old MAC fragmentation (43). Our transcriptome analysis indicates that the absence of new developing MACs reduces the steady-state mRNA levels of 2365 genes at the EARLY and INTERMEDIATE stages (Figure 4), half of which are specifically induced during autogamy under control conditions (13). These include a fraction of genes from the early and intermediate peaks (20% and 26%, respectively) and the late induction cluster (8%). We conclude that the presence of new developing MACs in the cell is necessary to activate part of the transcription program in the old MAC at the EARLY and INTERMEDIATE stages.

The completion of PGR during new MAC development controls gene expression at the LATE autogamy stage

The second level of control exerted by the new MACs depends on PGR outcome. Our work extends previous observations (34) that impeding PGR in the new MACs triggers gene upregulation at the LATE stage, which corresponds to the DEV3/4 stages of autogamy (starting from T20, see Figure 3A). According to published observations (12, 55), the old MAC at this stage is still responsible for at least 80% of total mRNA synthesis in starved autogamous cells. Using three independent KDs interfering with PGR at different steps, we report that 628 so-called UP PKX genes are upregulated at the LATE autogamy stage in *PGM*, *KU80c* and *XRCC4* KDs, instead of being switched off as in the control (Supplementary Table S6B). We did not include EZL1 RNA-seq data in our DE analysis because, as part of the PRC2 complex, Ezl1 is expected to repress gene expression through heterochromatin formation, as established in other species (57,58). We nevertheless compared the set of UP PKX genes with the list of LATE upregulated genes in an *EZL1* KD and found a large overlap between the two sets of genes, with UP PKX genes representing ∼20% of upregulated genes in an *EZL1* KD (Supplementary Figure S13). Thus, two effects of *EZL1* KD on gene expression may be distinguished in our analyses: LATE overexpression of the set of UP PKX genes seems to be a consequence of PGR inhibition (26,45), whereas upregulation of other genes is probably due to loss of Ezl1-mediated chromatin modification, independently of PGR.

Under control conditions, 15% of UP PKX genes are expressed in the early peak and 37% in the intermediate peak, which corresponds to the expression pattern of most known PGR genes (13). In addition to *PGM* and *KU80c*, we found that 8 UP PKX genes encode essential components of the DNA cleavage complex (*PGML1*, *PGML2*, *PGML3a*, *PGML4a&b*, *PGML5a&b* and *KU70a*) (31,34). Ten more UP PKX genes are involved in RNA-mediated targeting of eliminated DNA (*NOWA1* & *NOWA2* (59), *TFIIS4* (25), *PDSG2* (60)), the control of accessibility of DNA sequences to the cleavage complex through chromatin remodeling (*SPT16.1* (56), *ISWI1* (61)) or participate in PGR through still unclear mechanisms (*DIE5a* (62), *mtGa&b* (63), *SPO11* (64)) (Supplementary Figure S13 and Table S6F). Based on these observations, we propose that the set of 628 UP PKX genes may include new candidates, whose function in PGR should be explored further.

Since a *GFP* reporter transgene expressed from the old MAC under the control of *PGM* transcription signals shows a similar expression pattern to endogenous *PGM*, we conclude that LATE upregulation applies to old MAC transcripts. However, we observed that 86% of UP PKX genes are IES-free (Figure 3D), similar to *PGM*, *KU80c* and *XRCC4*, while

74% of LATE downregulated genes carry at least one IES (Supplementary Figure S8E–H). This suggests that deregulation of mRNA production in *PGM*, *KU80c* or *XRCC4* KDs not only takes place in the old MAC, but also applies to the new MACs. A contribution of the new MACs to total transcription is supported by the detection of low levels of IES+ transcripts throughout autogamy in a control time-course experiment, with a peak around T7, which corresponds to the INTERMEDIATE (DEV1) stage, at which IES elimination takes place (Supplementary Figure S9A). In the three KDs, the percentage of IES+ transcripts increases until the LATE autogamy stage, revealing the growing contribution of zygotic transcription over time. We note that genes from the late expression peak and late induction cluster are not particularly enriched among the population of genes producing IES+ mRNA at any autogamy stage (Supplementary Figure S9B). The observed underrepresentation of IES-carrying genes among the set of 628 LATE upregulated UP PKX genes (Figure 3D) may be explained by the selective degradation of IES+ transcripts through Nonsense-Mediated Decay (NMD), because IES retention frequently introduces premature termination codons in mRNAs (20).

Consistent with a previous report (65), we found that *PTIWI10*, a late induced gene involved in non-coding RNA-mediated control of IES excision, is downregulated at LATE stages in *PGM, KU80c* and *XRCC4* KDs (Supplementary Table S6A). *PTIWI10* was proposed to be specifically expressed from the new MACs (65). In the non-rearranged genome, it carries an IES in its promoter and its coding sequence is interrupted by a second, NMD-visible IES. The observed low amounts of *PTIWI10* mRNA in *PGM* and *KU80c* KDs could be explained by promoter inactivation consecutive to IES retention, and/or NMD-mediated degradation of the transcript carrying the internal IES. In the *XRCC4* KD, both IESs, which belong to the very early excised class, are cleaved, but persisting DSBs at their excision sites would inactivate the promoter and abolish the production of full-length *PTIWI10* mRNA.

Taken together, our observations suggest that LATE upregulation also takes place in the new MACs in *PGM, KU80c* and *XRCC4* KDs, as a consequence of IES retention or unrepaired chromosomal DSBs at excision sites. However, IES-containing genes expressed from the new MACs may have been overlooked in our search for UP PKX genes.

Bi-directional crosstalk between two generations of somatic nuclei in a single cell

Several lines of evidence have established that the old MAC controls new MAC development through several pathways. First, maternal non-coding transcripts produced by the old MAC serve as pairing templates for the selection of MIC-restricted small non-coding RNAs that are transferred to the new MACs (24), in which they drive the recognition of eliminated sequences through Ezl1-mediated formation of heterochromatin (27,28). Second, as discussed above, coding transcription in the old MAC ensures the synthesis of key proteins that carry out PGR in the new MACs.

Here, we provide evidence that, reciprocally, the new developing MACs control the onset and progression of part of the gene expression program in the old MAC throughout the successive stages of autogamy (Figure 7). The mechanistic details of this *trans*-nuclear control are still unclear and may involve different pathways at the EARLY or LATE stages, including the production of development-specific regulatory molecules (non-coding RNAs or zygotic transcription factors) by the new MACs. We have nevertheless identified a subset of 180 UP PKX genes that are coregulated with *PGM* under all tested conditions (Co-*PGM* genes, Figure 4E): these genes are specifically induced in the intermediate expression peak, their induction depends on the presence of new developing MACs in the cell and their mRNA levels drop down after the successful completion of PGR. Future studies will unravel the mechanism(s) that underlie the control of gene expression by the new MACs. One attractive hypothesis postulates that the production of activating factors is triggered by the developing new MACs to turn on gene transcription in the old MAC at the EARLY or INTERMEDIATE stages. Our finding that 98 out of 180 Co-*PGM* genes carry a conserved motif in their promoter, like *PGM* itself, suggests that genes from this subset are activated at the INTERMEDIATE stage through the sequence-specific binding of a transcription factor to the motif, while the expression of Co-*PGM* genes lacking the motif is likely governed by a different regulatory pathway. The observed LATE upregulation of Co-*PGM* genes when PGR does not proceed normally may indicate that the activator(s) is expressed from a non-rearranged locus in the new MACs and/or that a repressor produced after PGR completion is necessary to switch off gene expression.

As shown by published transcriptome studies in *Paramecium* (13,36) and *Tetrahymena* (66,67), MAC development in ciliates involves a coordinated gene expression program, with the sequential induction and repression of distinct gene clusters. Similar to the *Drosophila* embryo, the clusters of differentially expressed genes may be regulated by various combinations of transcription activators or repressors. The progression of developmental programs is often regulated by transcription factor networks, consisting of successive feed-forward and feedback regulatory loops (68). In particular, interesting comparisons can be drawn between MAC development in *Paramecium* and B-lymphocyte differentiation in mammals - two processes involving the induction of programmed DNA double-strand breaks and their subsequent repair by the NHEJ pathway (69). In preB cells, the temporal expression of *RAG* recombinase genes is controlled by a series of transcription activators and repressors to ensure that the rearrangement of immunoglobulin genes is restricted to a specific developmental stage (70,68). In *Paramecium*, similar transcription factor networks may regulate the timely expression of *PGM* and its co-regulated genes. Other regulatory mechanisms may control the timing of PGR in ciliates, such as programmed mRNA degradation. Little is known, however, about the ciliate transcription factors and molecular pathways that drive the progression of the gene expression program.

We report here that the extended period of time, during which the old and new MACs co-exist in the cytoplasm in ciliates, especially in *Paramecium*, may add more layers of complexity to transcription regulation during development: (i) zygotic transcription initiates in the new MACs while gene expression still massively originates from the old MAC and (ii) PGR in the new MACs is concomitant with the start of zygotic transcription. This raises the possibility that some regulatory factors are produced by the new MACs during development, perhaps from the unrearranged or partially rear-

ranged genome. An exhaustive understanding of the contribution of PGR to the regulation of zygotic gene expression in *Paramecium* will need the complete annotation of the developing new MAC genome, which should take the intermediate steps of programmed DNA elimination into account (see (15) for further discussion). Of note, numerous metazoans were also shown to eliminate portions of germline DNA from their somatic precursor lines during embryonic development, including germline-specific genes (71,72). To our knowledge, whether eliminated genes also contribute to the developmental program and control somatic differentiation in these organisms has not been established.

The existence of bi-directional cross-communication between the two successive generations of somatic nuclei throughout the *Paramecium* sexual cycle is reminiscent of the mother-embryo crosstalk that has been reported during mammalian development (see (73) for a review). During pregnancy, maternal hormones, signaling factors and small RNAs transported by extracellular vesicles control the development of the embryo. Reciprocally, regulatory factors and miRNAs produced by the embryo are also packaged in extracellular vesicles and modulate the maternal response. In both systems, the parental and zygotic somatic lines appear to mutually control each other through the production of coding and non-coding RNAs, proteins and other regulatory factors. Another parallel may be drawn with bacterial sporulation, during which a cascade of σ transcription factors are produced alternatively by the mother cell or by the daughter forespore to control the transcription of developmental genes in the other cell (74). Interestingly, in *Bacillus subtilis*, the expression of one of the sporulation σ factors is not only controlled at the transcription level, but also through the programmed elimination of an internal DNA fragment (75). Compared with these two systems, a notable difference in *Paramecium* is that inter-generational exchanges are intracellular. How a continuous dialogue is established at the molecular level between the old and the new developing MACs will require further investigation. We propose that inter-nuclear crosstalk allows gene transcription in the old MAC and genome rearrangements in the new MACs to proceed in a concerted manner to give rise to viable sexual progeny.

Data availability

The gene data underlying this article are available in ParameciumDB at https://paramecium.i2bc.paris-saclay.fr/ under the following accession codes: *CtIPa* (PTET.51.1.G0650078), *CtIPb* (PTET.51.1.G0980137), *EZL1* (PTET.51.1.G1740049), *ICL7a* (PTET.51.1.G0700039), *KU80c* (PTET.51.1.G1140146), *ND7* (PTET.51.1.G0050374), *PGM* (PTET.51.1.G0490162) and *XRCC4* (PTET.51.1.G1110086) (36). The sequencing data underlying this article are available in the European Nucleotide Archive at https://www.ebi.ac.uk/ena/browser/home under accession number PRJEB60900. The plasmid maps, raw images of all western blots, statistical data, scripts (Supplemental Codes) and numerical data underlying this article are available in Zenodo at https://doi.org/10.5281/zenodo.7835183.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We would like to thank Sandra Duharcourt and Eric Meyer for fruitful discussions, sharing data ahead of publication and providing helpful advice and support as members of MBG's thesis advisory committee. We are grateful to Anne Aubusson-Fleury and Anne-Marie Tassin for the generous gift of TEU435 antibodies. We thank Joël Acker, Julien Bischerour and Valerio Vitali for critical reading of the manuscript, all members of the Bétermier lab for stimulating discussions, Pascaline Tirand and Cindy Mathon for technical assistance, and Tim Kamara and Camille Poitrenaud for help in plasmid construction during their short-term internships.

Funding

Intramural funding from the Centre National de la Recherche Scientifique (CNRS) and grants from the Agence Nationale de la Recherche ['PIGGYPACK' ANR-14-CE10-0005 to M.B., L.S., O.A., 'LaMarque' ANR-18-CE12-0005-02 and 'CURE' ANR-21-CE12-0019-01 to M.B., 'POLYCHROME' ANR-19-CE12-0015 to O.A.]; Fondation pour la Recherche Médicale [FRM EQU202103012766 to M.B.]; M.B.G. was supported by a PhD fellowship from Paris-Saclay University [Ecole Doctorale no. 577 SDSV] and by a training grant from the Department of Genome Biology of I2BC to attend the 'DU-Omique' course in bioinformatics (Université Paris-Cité); we acknowledge the sequencing and bioinformatics expertise of the I2BC high-throughput sequencing facility, supported by France Génomique [funded by the French National Program 'Investissement d'Avenir' ANR-10-INBS-09]; fluorescence-assisted nuclear sorting experiments benefited from the expertise of the Imagerie-Gif core facility, supported by the Agence Nationale de la Recherche [ANR-11-EQPX-0029/Morphoscope, ANR-10-INBS-04/FranceBioImaging, ANR-11-IDEX-0003-02/ Saclay Plant Sciences]. Funding for open access charge: Fondation pour la Recherche Médicale [FRM EQU202103012766].

Conflict of interest statement

None declared.

References

1. Padilla,J.R., Ferreira,L.M. and Folker,E.S. (2022) Nuclear movement in multinucleated cells. *Development*, **149**, dev200749.
2. Deshpande,O. and Telley,I.A. (2021) Nuclear positioning during development: pushing, pulling and flowing. *Semin. Cell Dev. Biol.*, **120**, 10–21.
3. Dundon,S.E.R., Chang,S.-S., Kumar,A., Occhipinti,P., Shroff,H., Roper,M. and Gladfelter,A.S. (2016) Clustered nuclei maintain autonomy and nucleocytoplasmic ratio control in a syncytium. *Mol. Biol. Cell*, **27**, 2000–2007.
4. Kim,M., Franke,V., Brandt,B., Lowenstein,E.D., Schöwel,V., Spuler,S., Akalin,A. and Birchmeier,C. (2020) Single-nucleus transcriptomics reveals functional compartmentalization in syncytial skeletal muscle cells. *Nat. Commun.*, **11**, 6375.
5. Petrany,M.J., Swoboda,C.O., Sun,C., Chetal,K., Chen,X., Weirauch,M.T., Salomonis,N. and Millay,D.P. (2020) Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. *Nat. Commun.*, **11**, 6374.
6. Ahmad,K. and Henikoff,S. (2022) The Drosophila embryo as a tabula rasa for the epigenome. *F1000Prime Rep*, **11**, 40.

7. Harrison,M.M. and Eisen,M.B. (2015) Transcriptional activation of the zygotic genome in Drosophila. *Curr. Top. Dev. Biol.*, **113**, 85–112.
8. Duboule,D. (2022) The (unusual) heuristic value of Hox gene clusters; a matter of time?*Dev. Biol.*, **484**, 75–87.
9. Prescott,D.M. (1994) The DNA of ciliated protozoa. *Microbiol. Rev.*, **58**, 233–267.
10. Drews,F., Boenigk,J. and Simon,M. (2022) *Paramecium* epigenetics in development and proliferation. *J. Eukaryot. Microbiol.*, **69**, e12914.
11. Bétermier,M. and Duharcourt,S. (2014) Programmed Rearrangement in Ciliates. *Paramecium. Microbiol. Spectr.*, **2**, MDNA–0035–2014.
12. Berger,J.D. (1973) Nuclear differentiation and nucleic acid synthesis in well-fed exconjugants of *Paramecium aurelia*. *Chromosoma*, **42**, 247–268.
13. Arnaiz,O., Dijk,E.V., Bétermier,M., Lhuillier-Akakpo,M., Vanssay,A.d., Duharcourt,S., Sallet,E., Gouzy,J. and Sperling,L. (2017) Improved methods and resources for paramecium genomics: transcription units, gene annotation and gene expression. *Bmc Genomics [Electronic Resource]*, **18**, 483.
14. Preer,J.R. (1976) Quantitative predictions of random segregation models of the ciliate macronucleus. *Genet. Res.*, **27**, 227–238.
15. Zangarelli,C., Arnaiz,O., Bourge,M., Gorrichon,K., Jaszczyszyn,Y., Mathy,N., Escoriza,L., Bétermier,M. and Régnier,V. (2022) Developmental timing of programmed DNA elimination in *Paramecium tetraurelia* recapitulates germline transposon evolutionary dynamics. *Genome Res.*, **32**, 2028–2042.
16. Sellis,D., Guérin,F., Arnaiz,O., Pett,W., Lerat,E., Boggetto,N., Krenek,S., Berendonk,T., Couloux,A., Aury,J.-M., *et al.* (2021) Massive colonization of protein-coding exons by selfish genetic elements in *Paramecium* germline genomes. *PLoS Biol.*, **19**, e3001309.
17. Aury,J.-M., Jaillon,O., Duret,L., Noel,B., Jubin,C., Porcel,B.M., Ségurens,B., Daubin,V., Anthouard,V., Aiach,N., *et al.* (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, **444**, 171–178.
18. Guérin,F., Arnaiz,O., Boggetto,N., Wilkes,C.D., Meyer,E., Sperling,L. and Duharcourt,S. (2017) Flow cytometry sorting of nuclei enables the first global characterization of *Paramecium* germline DNA and transposable elements. *Bmc Genomics [Electronic Resource]*, **18**, 327.
19. Le Mouël,A., Butler,A., Caron,F. and Meyer,E. (2003) Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia. *Eukaryot. Cell*, **2**, 1076–1090.
20. Arnaiz,O., Mathy,N., Baudry,C., Malinsky,S., Aury,J.-M.M., Wilkes,C.D., Garnier,O., Labadie,K., Lauderdale,B.E., Mouël,A.L., *et al.* (2012) The Paramecium germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet.*, **8**, e1002984.
21. Klobutcher,L.A. and Herrick,G. (1995) Consensus inverted terminal repeat sequence of *Paramecium* IESs: resemblance to termini of *Tc1*-related and *Euplotes Tec* transposons. *Nucleic Acids Res.*, **23**, 2006–2013.
22. Sandoval,P.Y., Swart,E.C., Arambasic,M. and Nowacki,M. (2014) Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting. *Dev. Cell*, **28**, 174–188.
23. Lepère,G., Nowacki,M., Serrano,V., Gout,J.-F., Guglielmi,G., Duharcourt,S. and Meyer,E. (2009) Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucl. Acids Res*, **37**, 903–915.
24. Lepère,G., Bétermier,M., Meyer,E. and Duharcourt,S. (2008) Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev.*, **22**, 1501–1512.
25. Maliszewska-Olejniczak,K., Gruchota,J., Gromadka,R., Denby Wilkes,C., Arnaiz,O., Mathy,N., Duharcourt,S., Bétermier,M. and Nowak,J.K. (2015) TFIIS-Dependent Non-coding Transcription Regulates Developmental Genome Rearrangements. *PLoS Genet.*, **11**, e1005383.
26. Lhuillier-Akakpo,M., Frapporti,A., Denby Wilkes,C., Matelot,M., Vervoort,M., Sperling,L. and Duharcourt,S. (2014) Local effect of enhancer of zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements. *PLoS Genet.*, **10**, e1004665.
27. Miró-Pina,C., Charmant,O., Kawaguchi,T., Holoch,D., Michaud,A., Cohen,I., Humbert,A., Jaszczyszyn,Y., Chevreux,G., Maestro,L.D., *et al.* (2022) *Paramecium* Polycomb repressive complex 2 physically interacts with the small RNA-binding PIWI protein to repress transposable elements. *Dev. Cell*, **57**, 1037–1052.
28. Wang,C., Solberg,T., Maurer-Alcalá,X.X., Swart,E.C., Gao,F. and Nowacki,M. (2022) A small RNA-guided PRC2 complex eliminates DNA as an extreme form of transposon silencing. *Cell Rep.*, **40**, 111263.
29. Baudry,C., Malinsky,S., Restituito,M., Kapusta,A., Rosa,S., Meyer,E. and Bétermier,M. (2009) PiggyMac, a domesticated *piggyBac* transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev.*, **23**, 2478–2483.
30. Dubois,E., Mathy,N., Régnier,V., Bischerour,J., Baudry,C., Trouslard,R. and Bétermier,M. (2017) Multimerization properties of PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements. *Nucleic Acids Res.*, **45**, 3204–3216.
31. Bischerour,J., Bhullar,S., Denby Wilkes,C., Régnier,V., Mathy,N., Dubois,E., Singh,A., Swart,E., Arnaiz,O., Sperling,L., *et al.* (2018) Six domesticated PiggyBac transposases together carry out programmed DNA elimination in Paramecium. *eLife*, **7**, e37927.
32. Gratias,A. and Bétermier,M. (2003) Processing of double-strand breaks is involved in the precise excision of *Paramecium* internal eliminated sequences. *Mol. Cell. Biol.*, **23**, 7152–7162.
33. Kapusta,A., Matsuda,A., Marmignon,A., Ku,M., Silve,A., Meyer,E., Forney,J.D., Malinsky,S. and Bétermier,M. (2011) Highly precise and developmentally programmed genome assembly in *Paramecium* requires ligase IV–dependent end joining. *PLoS Genet.*, **7**, e1002049.
34. Marmignon,A., Bischerour,J., Silve,A., Fojcik,C., Dubois,E., Arnaiz,O., Kapusta,A., Malinsky,S. and Bétermier,M. (2014) Ku-mediated coupling of DNA cleavage and repair during programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *PLoS Genet.*, **10**, e1004552.
35. Abello,A., Régnier,V., Arnaiz,O., Bars,R.L., Bétermier,M. and Bischerour,J. (2020) Functional diversification of *Paramecium* Ku80 paralogs safeguards genome integrity during precise programmed DNA elimination. *PLoS Genet.*, **16**, e1008723.
36. Arnaiz,O., Meyer,E. and Sperling,L. (2019) ParameciumDB 2019: integrating genomic data across the genus for functional and evolutionary biology. *Nucleic Acids Res.*, **48**, 599–605.
37. Beisson,J., Betermier,M., Bre,M.H., Cohen,J., Duharcourt,S., Duret,L., Kung,C., Malinsky,S., Meyer,E., Preer,J.R. Jr, *et al.* (2010) Mass culture of *Paramecium tetraurelia*. *Cold Spring Harb. Protoc.*, **2010**, pdb prot5362.
38. Beisson,J., Bétermier,M., Bré,M.-H., Cohen,J., Duharcourt,S., Duret,L., Kung,C., Malinsky,S., Meyer,E., Preer,J.R., *et al.* (2010) Silencing Specific *Paramecium tetraurelia* Genes by Feeding Double-Stranded RNA. *Cold Spring Harb. Protoc.*, **2010**, pdb.prot5363.
39. Galvani,A. and Sperling,L. (2002) RNA interference by feeding in *Paramecium*. *Trends Genet.*, **18**, 11–12.
40. Timmons,L., Court,D.L. and Fire,A. (2001) Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene*, **263**, 103–112.
41. Bétermier,M., Duharcourt,S., Seitz,H. and Meyer,E. (2000) Timing of developmentally programmed excision and circularization of *Paramecium* internal eliminated sequences. *Mol. Cell. Biol.*, **20**, 1553–1561.

42. Garnier,O., Serrano,V., Duharcourt,S. and Meyer,E. (2004) RNA-mediated Programming of Developmental Genome Rearrangements in *Paramecium tetraurelia*. *Mol. Cell. Biol.*, **24**, 7370–7379.
43. Godau,J., Ferretti,L.P., Trenner,A., Dubois,E., von Aesch,C., Marmignon,A., Simon,L., Kapusta,A., Guérois,R., Bétermier,M., *et al.* (2019) Identification of a miniature Sae2/Ctp1/CtIP ortholog from *Paramecium tetraurelia* required for sexual reproduction and DNA double-strand break repair. *DNA Repair (Amst.)*, **77**, 96–108.
44. Callen,A.M., Adoutte,A., Andrew,J.M., Baroin-Tourancheau,A., Bré,M.H., Ruiz,P.C., Clérot,J.C., Delgado,P., Fleury,A. and Jeanmaire-Wolf,R. (1994) Isolation and characterization of libraries of monoclonal antibodies directed against various forms of tubulin in *Paramecium*. *Biol. Cell*, **81**, 95–119.
45. Frapporti,A., Miró Pina,C., Arnaiz,O., Holoch,D., Kawaguchi,T., Humbert,A., Eleftheriou,E., Lombard,B., Loew,D., Sperling,L., *et al.* (2019) The Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in *Paramecium*. *Nat. Commun.*, **10**, 2710.
46. Denby Wilkes,C., Arnaiz,O. and Sperling,L. (2015) ParTIES: a toolbox for *Paramecium* interspersed DNA elimination studies. *Bioinformatics*, **32**, 599–601.
47. Core Team,R. (2021) R: a language and environment for statistical computing. In: *Foundation for Statistical Computing*. Vienna.
48. Chen,H. and Boutros,P.C. (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinf.*, **12**, 35.
49. Wickham,H. (2016) In: *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, NY.
50. Zhang,Y., Parmigiani,G. and Johnson,W.E. (2020) ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform*, **2**, lqaa078.
51. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
52. Bailey,T.L. (2021) STREME: accurate and versatile sequence motif discovery. *Bioinformatics*, **37**, 2834–2840.
53. Grant,C.E., Bailey,T.L. and Noble,W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017.
54. Swart,E.C., Wilkes,D., C.,S., P.Y.,H., C.,S., A.,F., D.I.,A., M.,I. and Nowacki,M. (2017) Identification and analysis of functional associations among natural eukaryotic genome editing components [version 1; peer review: 1 approved, 1 approved with reservations]. *F1000Research*, **6**, 1374.
55. Berger,J.D. (1976) Gene expression and phenotypic change in Paramecium tetraurelia exconjugants. *Genet. Res.*, **27**, 123–134.
56. de Vanssay,A., Touzeau,A., Arnaiz,O., Frapporti,A., Phipps,J. and Duharcourt,S. (2020) The *Paramecium* histone chaperone Spt16-1 is required for Pgm endonuclease function in programmed genome rearrangements. *PLoS Genet.*, **16**, e1008949.
57. Margueron,R. and Reinberg,D. (2011) The Polycomb complex PRC2 and its mark in life. *Nature*, **469**, 343–349.
58. Blackledge,N.P. and Klose,R.J. (2021) The molecular principles of gene regulation by Polycomb repressive complexes. *Nat. Rev. Mol. Cell Biol.*, **22**, 815–833.
59. Nowacki,M., Zagorski-Ostoja,W. and Meyer,E. (2005) Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in *Paramecium tetraurelia*. *Curr. Biol. CB*, **15**, 1616–1628.
60. Arambasic,M., Sandoval,P.Y., Hoehener,C., Singh,A., Swart,E.C. and Nowacki,M. (2014) Pdsg1 and Pdsg2, novel proteins involved in developmental genome remodelling in *Paramecium*. *PLoS One*, **9**, e112899.
61. Singh,A., Maurer-Alcalá,X.X., Solberg,T., Häußermann,L., Gisler,S., Ignarski,M., Swart,E.C. and Nowacki,M. (2022) Chromatin remodeling is required for sRNA-guided DNA elimination in *Paramecium*. *EMBO J.*, **41**, e111839.
62. Matsuda,A., Shieh,A.W.-Y., Chalker,D.L. and Forney,J.D. (2010) The Conjugation-Specific Die5 Protein Is Required for Development of the Somatic Nucleus in both *Paramecium* and *Tetrahymena*. *Eukaryot. Cell*, **9**, 1087–1099.
63. Bhullar,S., Denby Wilkes,C., Arnaiz,O., Nowacki,M., Sperling,L. and Meyer,E. (2018) A mating-type mutagenesis screen identifies a zinc-finger protein required for specific DNA excision events in *Paramecium*. *Nucleic Acids Res.*, **46**, 9550–9562.
64. Rzeszutek,I., Swart,E.C., Pabian-Jewuła,S., Russo,A. and Nowacki,M. (2022) Early developmental, meiosis-specific proteins — Spo11, Msh4-1, and Msh5 — Affect subsequent genome reorganization in *Paramecium tetraurelia*. *Biochim. Biophys. Acta BBA - Mol. Cell Res.*, **1869**, 119239.
65. Furrer,D.I., Swart,E.C., Kraft,M.F., Sandoval,P.Y. and Nowacki,M. (2017) Two sets of Piwi proteins are involved in distinct sRNA pathways leading to elimination of germline-specific DNA. *Cell Rep.*, **20**, 505–520.
66. Xiong,J., Lu,X., Zhou,Z., Chang,Y., Yuan,D., Tian,M., Zhou,Z., Wang,L., Fu,C., Orias,E., *et al.* (2012) Transcriptome analysis of the model protozoan, Tetrahymena thermophila, using Deep RNA sequencing. *PLoS One*, **7**, e30630.
67. Xiong,J., Lu,Y., Feng,J., Yuan,D., Tian,M., Chang,Y., Fu,C., Wang,G., Zeng,H. and Miao,W. (2013) Tetrahymena functional genomics database (TetraFGD): an integrated resource for Tetrahymena functional genomics. *Database J. Biol. Databases Curation*, **2013**, bat008.
68. Sigvardsson,M. (2023) Transcription factor networks link B-lymphocyte development and malignant transformation in leukemia. *Genes Dev.*, **37**, 703–723.
69. Bétermier,M., Borde,V. and de Villartay,J.-P. (2020) Coupling DNA damage and repair: an essential safeguard during programmed DNA double-strand breaks?*Trends Cell Biol.*, **30**, 87–96.
70. Grawunder,U., Leu,T.M., Schatz,D.G., Werner,A., Rolink,A.G., Melchers,F. and Winkler,T.H. (1995) Down-regulation of RAG1 and RAG2 gene expression in preB cells after functional immunoglobulin heavy chain rearrangement. *Immunity*, **3**, 601–608.
71. Wang,J. and Davis,R.E. (2014) Programmed DNA elimination in multicellular organisms. *Curr. Opin. Genet. Dev.*, **27**, 26–34.
72. Estrem,B. and Wang,J. (2023) Programmed DNA elimination in the parasitic nematode Ascaris. *PLoS Pathog.*, **19**, e1011087.
73. Idelevich,A. and Vilella,F. (2020) Mother and embryo cross-communication. *Genes*, **11**, 376.
74. Errington,J. (2003) Regulation of endospore formation in *Bacillus subtilis*. *Nat. Rev. Microbiol.*, **1**, 117–126.
75. Stragier,P., Kunkel,B., Kroos,L. and Losick,R. (1989) Chromosomal rearrangement generating a composite gene for a developmental transcription factor. *Science*, **243**, 507–512.

Received: April 17, 2023. **Revised:** October 9, 2023. **Editorial Decision:** October 11, 2023. **Accepted:** October 23, 2023