# Investigation of condensin-related proteins and purification of the DNA excision machinery in *Paramecium tetraurelia*

Inaugural dissertation

of the Faculty of Science,

University of Bern

presented by

Fukai Zhang

from China

Supervisor of the doctoral thesis:

Prof. Dr. Mariusz Nowacki

Institute of Cell Biology

**Investigation of condensin-related proteins and purification of the DNA excision machinery in *Paramecium tetraurelia***

Inaugural dissertation

of the Faculty of Science,

University of Bern

presented by

Fukai Zhang

from China

Supervisor of the doctoral thesis:

Prof. Dr. Mariusz Nowacki

Institute of Cell Biology

Accepted by the Faculty of Science.

Bern, 15.11.2022 The Dean

Prof. Dr.

# 1. Abstract

The spatial organization of chromosomes is vital to life at the cellular level. DNA loops formed by condensin are the crucial structure of the meiotic and mitotic genomes which organize chromatin in interphase and metaphase. The structural maintenance of chromosomes (SMCs) protein family, which is now thought to serve as a DNA motor, is central to this organization. *Paramecium tetraurelia* is a ciliate that performs RNA-mediated widely genome rearrangement, mainly Internal Elimination Sequence (IES) excision, during sexual reproduction. The scanning model, which includes two sets of small RNAs (scanRNA and iesRNA), is widely used to explain this process. While the exact mechanism by which interphase and metaphase genome structure interact with IES excision is unknown to date. Given the importance of SMCs to chromosome structure and the poorly reported condensin in *Paramecium tetraurelia*, we explored the function of condensin SMCs in IES elimination and investigated whether their cooperation displayed different effects on this progress. One of the SMC4s in *Paramecium tetraurelia*, SMC4-2, is a novel player of IES elimination. Knockdown of SMC4-2 is lethal and causes the retention of almost all IESs, while the knockdown of SMC2s and SMC4-1 exhibit lethality to progeny without any sign of IES retention. Most interestingly, combining SMC4-2 RNAi with any of the other SMCs mentioned above abolished the retention in SMC4-2 single silencing. The SMC4-1 and SMC4-2 Co-immunoprecipitation and mass spectrometry analysis indicates a possibility that the interactions between SMC4-1 and SMC4-2 are decided by their concentration ratio of them. Additionally, we developed and optimized a purification method for IES-specific *in vivo* DNA binding complexes. The exact players of IES excision, especial the complexes that perform recognizing, binding and cleaving, are still unclear. This method allows us to purify specific-IES binding excision machinery from the new developing MAC where the old MAC contamination can be excluded. After that, we also tried to analyse the PiggyMac (PGM) interacting proteins by using Flag-HA tagged PGM 3A mutant. We were unable to enrich the PGM 3A-Flag-HA on anti-HA beads. However, we still provide some hints which could be helpful to further study.

# 2. Introduction

## 2.1. The *Paramecium tetraurelia*

*Paramecium tetraurelia* (*P. tetraurelia*) is an abundant, free-living, unicellular organism that is simple to cultivate in a lab. Generally, *P. tetraurelia* lives in various freshwater environments where it feeds on algae, yeast or bacteria[1]. As a very popular studied ciliated which provides a fascinating viewpoint on the fundamental molecular and cellular processes of eukaryotic life[2] due to its location within Alveolate (Figure 1), which is distinct from the models that are most frequently employed[3]. There are several organelles in this foot-shaped cell containing in other eukaryotes, such as ribosomes, mitochondria, endoplasmic reticulum, a Golgi apparatus, a phagosome-lysosome system, an endocytic system of shuttle vesicles, and enveloped nucleus[4]. In the meantime, its huge body (approximately 120 μm long, 50 μm broad) and nuclear size (3 μm to 40 μm) make *P. tetraurelia* an ideal subject of cellular and molecular biology research. The occurrence of two different nuclei inside the same cell is a defining feature of ciliate genetics: the micronucleus (MIC) is diploid and the macronucleus (MAC) is highly polyploid, so-named due to their relative sizes to one other. Two different types of nuclei existing in the cytoplasm are advised as "nuclear dimorphism", and it is a trait shared by most of the ciliates. Another property shared by most ciliates is the extensive genome rearrangements during their sexual development. Both the number of nuclei and how genomic rearrangements are controlled vary greatly amongst ciliate species. In my research material *P. tetraurelia* has one MAC and two MICs in each cell[5].

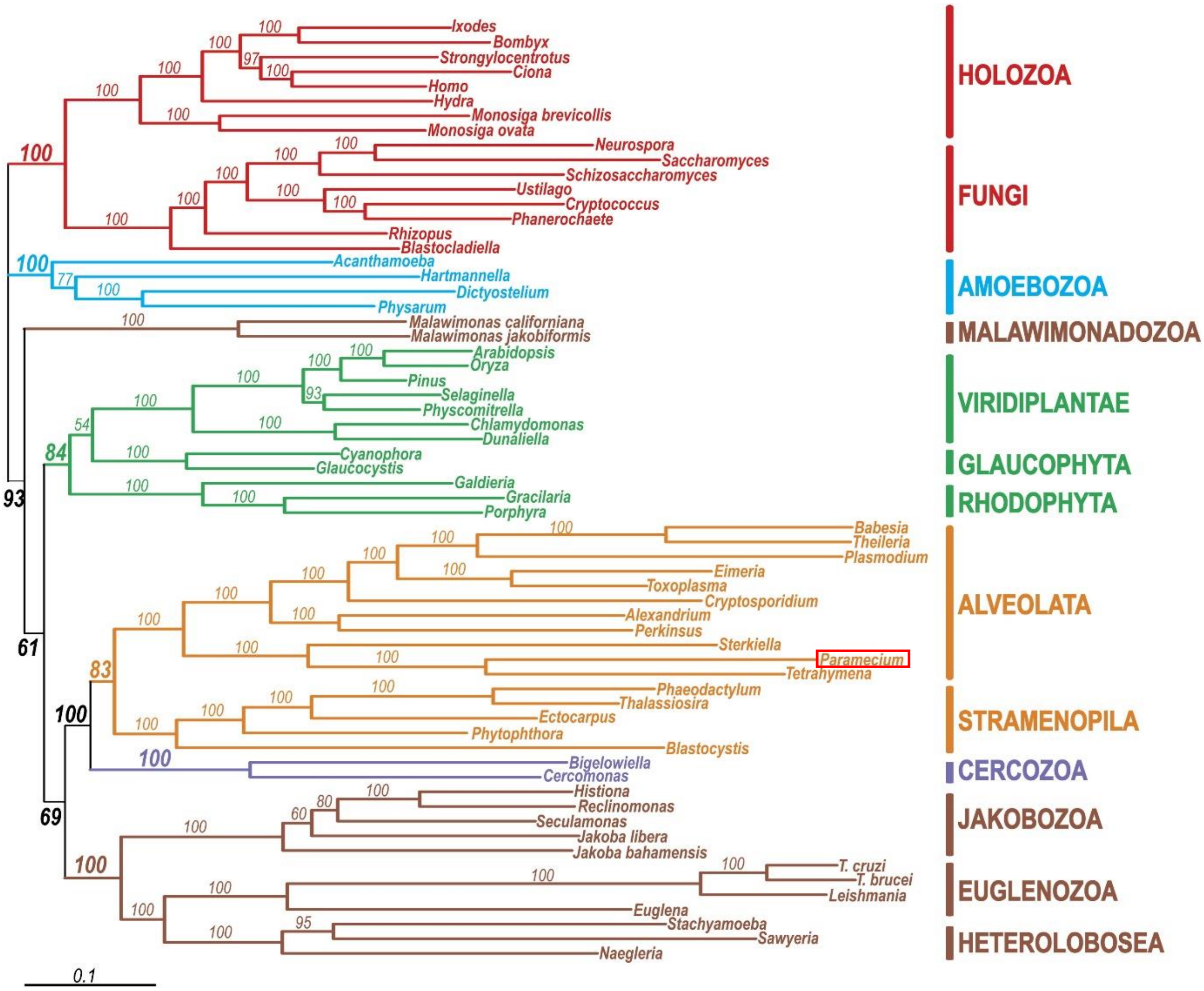


**Figure 1. Phylogenetic tree of Eukaryotes.** Adopted from [6]. The tree, which contains 64 species, is built using 143 concatenated nucleus-encoded proteins (31,604 amino acid positions). The numbers represent RaxML analysis support values (100 repetitions) with the WAG + F +Γ model. For all branches, the posterior probability derived using Bayesian inference with MrBayes is 1.0. The number of amino acid substitutions per site is indicated by the scale bar. A gene fusion was used to root the tree. Ciliates are single-celled organisms that belong to the *Alveolata* superphylum. The tree includes the most widely researched families, *Paramecium*, *Oxytricha* (*Sterkiella*), and *Tetrahymena*. The *Alveolata* also includes parasitic protists such as *Plasmodium*. The ciliate families are further classified as *Oligohymenophora* (*Paramecium*) and *Stichotrichia*.

## 2.2. Life cycle and nuclear development in *Paramecium tetraurelia*

For the study of nuclear function and differentiation[7], the process of meiosis, mitosis, autogamy, and conjugation in *P. tetraurelia* is quite useful. Also in studies of heredity, the ability of Paramecium achieving homozygosity of all loci at once with autogamy is particularly useful. The same is true for the production of heterokaryons by nuclear regeneration. Additionally, the investigator may readily regulate each of these steps. Two diploid MICs fulfil germline functions in one cell body, which is about 3μm in diameter, has around 50 chromosomes and a genome size of ~100 megabases[8]. The MIC is transcriptionally inactive during vegetative growth but can be a source of MAC generation and maturation during development. In addition to this, MIC can also maintain the genome information throughout sexual reproduction by meiosis and mitosis. Without the MAC, *P. tetraurelia* can only go through one or two cycles of fission, but without the MIC, the cells can endure more than one hundred fissions and can even mate[9]. The MAC genome looks like a simplified form of the MIC genome that primarily contains essential sequences for protein synthesis. The maternal MAC disappears during the sexual development of *P. tetraurelia*, and a new one from MIC takes its place[10]. The size of MAC is 35 μm in length and 12 μm in width with highly polyploid up to 800n copies[11]. Due to the elimination of DNA occurring in sexual development, chromosomes in MAC are shredded into approximately 200 linear molecules, 50 kb to 1 Mb in size[12]. The MAC is responsible for gene expression and considered a somatic nucleus. The existence of many nucleoli, which are absent from the MICs, and the fact that amacronucleate cells die without division, provide additional evidence for the gene expression activity in the MAC.

Like other ciliates, *P. tetraurelia* has the capacity for both vegetative growth and sexual development as shown in Figure 2. However, cells in immaturity cannot undergo sexual development like conjugation or autogamy. Additionally, cells can conjugate but cannot go through autogamy during a certain vegetative development phase following autogamy. This period is described as autogamous immaturity. Under normal conditions, like food-rich, less stressful, non-toxic etc., *P. tetraurelia* divides in a vegetative manner every 5 – 6 hours at 27 °C or 12 hours at 18 °C. Vegetative proliferation only occurs by binary fission, in which the MICs reproduce via mitosis. While the process of MAC division is called amitosis due to the absence of centromeres[13], by which the MAC elongates and splits into two halves. The amitotic divisions does not ensure that two halves contain the same number of chromosomes.

After the vegetative division, each daughter cell will regulate its copy number to make it the same as their parental cells, although the mechanism behind is still an open question.

Besides the vegetative growth, *P. tetraurelia* cells must renew their somatic genome by sexual development since they can't divide vegetatively infinitely[14]. However, the presence of immortal ciliates that are incapable of sexual reproduction would imply that cells are essentially eternal and have evolved a mechanism to limit the ability for division when they develop the sexual process. The loss of such a mechanism can be shown by the emergence of strains in fatal species, such as *Tetrahymena pyriformis*, that can divide indefinitely[15]. Conjugation and autogamy are two types of sexual reproduction of *P. tetraurelia* that are usually triggered by moderate starvation (the most employed strategy in the laboratory), environmental disturbance and ageing, which share very similar cytological steps[16]. The main distinction of the two processes is that during conjugation, the genetic material of two compatible mating type cells is exchanged, while autogamy, also referred to as self-fertilization, takes place within a single *P. tetraurelia* cell and forms homozygous cells. While autogamy happens in older cells, conjugation occurs in young cells. Generally, the sexual cycle begins with two meiotic divisions of both MICs in the cell, forming eight haploid nuclei, of which seven degrade quickly[17,18]. The surviving one undergoes mitotic division, leading to two identical haploid nuclei. These two nuclei will be fused in autogamy or exchanged during one type of sexual development, conjugation, between *P. tetraurelia* have suitable mating types that form a zygotic nucleus. In the meantime, the parental MAC becomes fragmented but keeps the ability to transcribe as well as remain in the cells throughout the whole sexual development. The diploid zygote undergoes two further mitotic divisions very shortly after karyogamy. The sister nuclei are pushed to the opposite poles of the cell by the long spindles of the second postzygotic mitosis, with two in the front and two in the posterior area. Two of those nuclei will stay unchanged and generate new MICs in daughter cells. The other two nuclei (anlagen), on the other hand, will create new MACs. This polar localization is required for new MAC development, including enormous DNA synthesis, transcription, and chromosomal rearrangements. This is followed by a specific cell division known as karyonidal division, in which MICs are mitosis-separated into the progeny, whereas the two MACs do not divide and are merely segregated into two daughter cells. The old MAC gradually disappeared throughout the cycle. The fragments, on the other side, stay transcriptionally active for some reason and are processed by subsequent vegetative divisions[18].

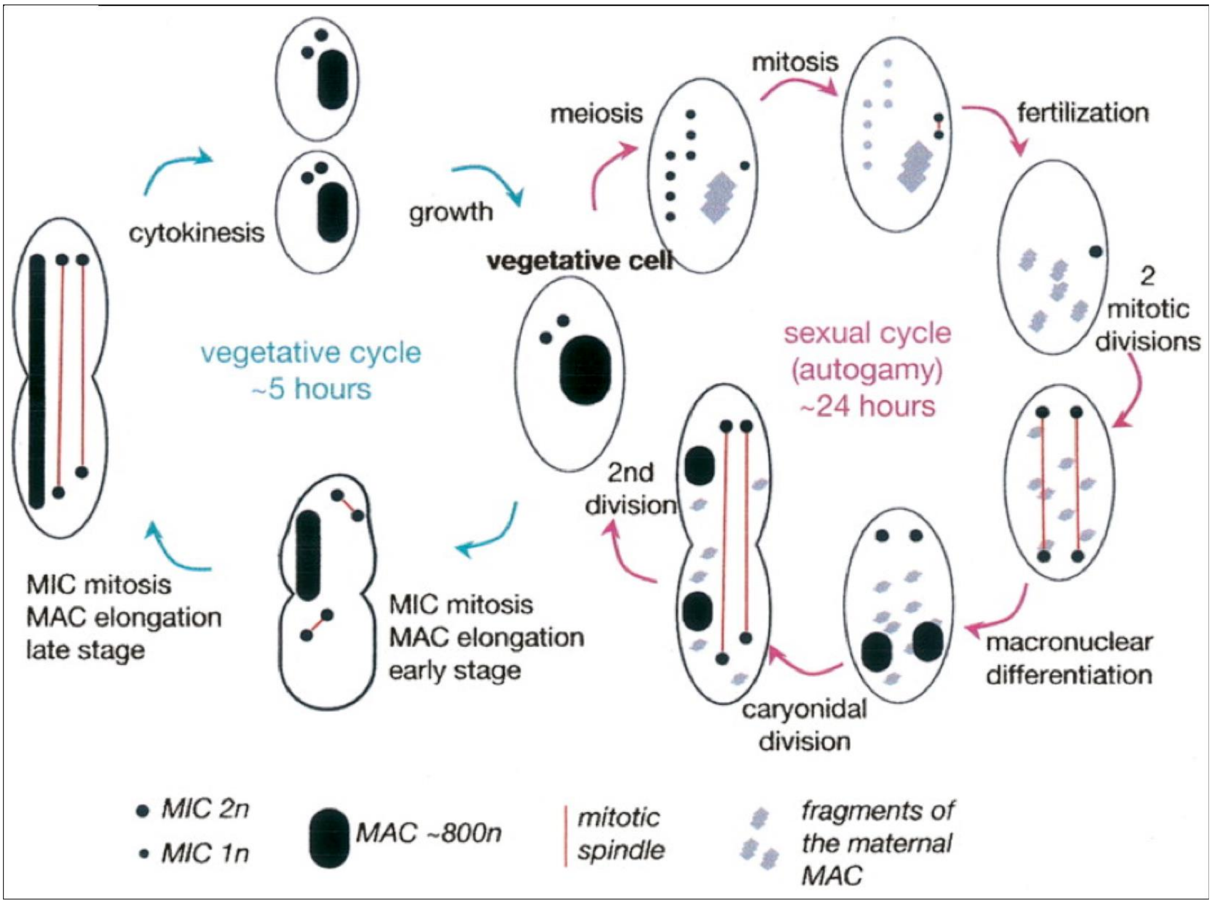


**Figure 2. Vegetative (Left) and sexual (Right) phases of the *Paramecium* life cycle.** Adapted from Beisson J[4]. The vegetative cell cycle (left part) starts from the MIC mitosis and MAC elongation under favourable conditions. The MAC divides amitotically into two daughter cells and the two daughter cells are separated by cytokinesis. The daughter MAC genomes will grow up to the same size as parental MAC in a poorly understood mechanism. Autogamy shown on the right side is one of the sexual cycles in *P. tetraurelia*. This event includes meiosis of each MIC, two parental diploid MICs become eight haploid MICs. In the meantime, old MAC started to fragment and seven out of eight monoploid MICs are degraded, the single surviving one goes through additional mitosis (two haploid MICs) and karyogamy resulting in a new diploid MIC. With two consecutive mitosis, one diploid MIC becomes four diploid MICs in the same cytoplasm. Two of them are chosen to develop into new developing MACs with progressively genome rearrangement, the other two MICs stay as MICs before another round of mitosis. The mechanism of this selection is still unclear. After that, the cell will be divided into two daughter cells with equal distribution of genetic material.

### 2.2.1. Progressive genome rearrangement

Even though the MAC is produced from the MIC, their genomic content and organisation differ significantly. The reason for the differences is that the MAC undergoes dramatic genome rearrangement (Figure 3). The genome rearrangements include DNA amplification leading to high ploidy of MAC (up to 800 n) and the sequences eliminated often involve transposable elements and repeated sequences, like minisatellites. This causes chromosomal fragmentation via internal deletions or *de novo* telomere addition when left regions are reconnected and ligated[12]. Internal eliminated sequences, or IESs, have to be removed precisely from the immature MAC since 80% of the IESs are located within protein-coding regions[19]. As a result, DNA elimination is critical for the formation of a functioning MAC genome. More than 45,000 single-copy, AT-rich IESs have been determined by knocking down the domesticated transposase PiggyMac (PGM), which is required for the removal of IESs during development[20].

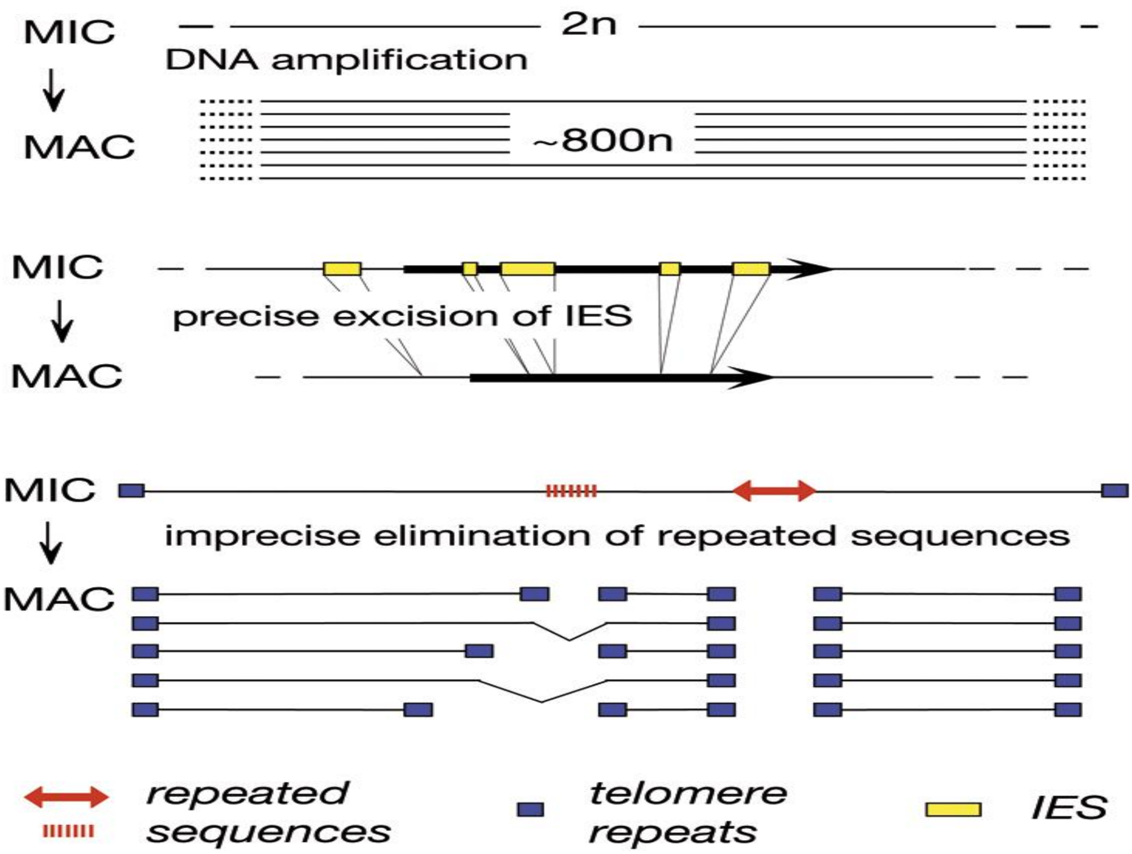


**Figure 3. Genome rearrangement during the development of MAC[4].** Ploidy: In mature MICs, there are only two copies of chromosomes (2n); in mature MAC, the genome is amplified to highly polyploid (up to ~800n). Precise elimination of IESs: Around 45,000 single-copy, AT-rich IESs are eliminated in the MAC genome when developed from MIC. Imprecise elimination of repeated sequences: transposons and minisatellites, repeated sequences, are removed during the MAC's maturation. Fragmentation of chromosomes is formed by these imprecise eliminations and *de novo* telomerization stabilizes the newly formed free chromosome ends.

The length of an IES varies between 26 and 4838 bp, with one-third of IESs being 26-28 bp long. Furthermore, the short IESs found in *P. tetraurelia* have a size distribution (Figure 4) with a periodicity of 10.2bp, which corresponds to a double-stranded DNA helical repeat. While the second peak (expected at 35 bp) as seen in the profile only contains a few IESs, it is so-called the forbidden peak. This size range of IES may result from excision machinery restrictions and evolutionary pressure to IESs outside of certain size ranges[21]. IESs are hypothesised to be evolutionary remains of transposable elements that infiltrated the MIC genome, then degenerated over time by deletions and nucleotide substitutions[22]. Most IESs display lengths below 150 bp and are flanked with 5'-TA-3' repeats, one of them remains non-excised in the developed MAC genome after the IES has been eliminated[23]. Every IES TA repeat is accompanied by the very weak consensus sequence TAYAGYNR, which is supposed to be decisive since any change in the conserved sequences can impede IES excision. The similarity of IES ends to *Tc1/mariner* transposons gives rise to the assumption that IES could be originated the insertion of these mobile elements followed by the missing coding capacity throughout evolution[24]. Besides terminal IES sequences, sequences flanking IESs have been involved in excision in some cases. Excision was abolished in *Paramecium* by removing a piece of its flanking region from one end of the specific IES[25].

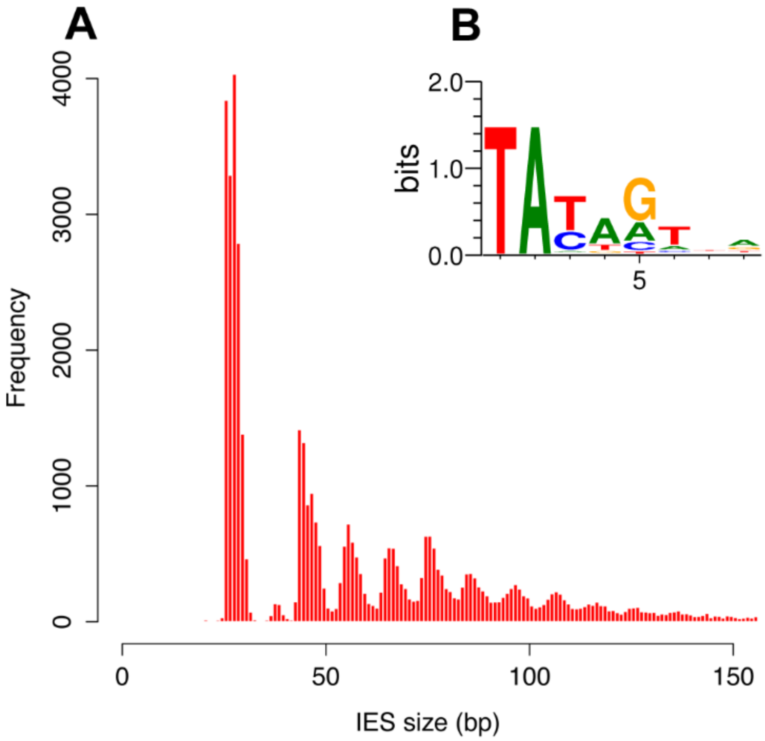

**Figure 4. IES sequence distribution[21].** A) Histogram shows the genome-wide collection of IESs with lengths smaller than 150 bp. B) For the terminals of the genome-wide collection of IESs, a sequence logo depicting sequence preference at each place, was adjusted for a G+C content of 28%.

### 2.2.2. Scanning models in IES excision

The excision machinery's recognition of IESs is not well understood. Excision of these distinct sequences requires a precise match. The IESs' loosely conserved end consensus sequence is deficient in explaining their recognition pattern suggesting that there must be another mechanism behind the precise elimination. Several theories have outlined the presumed roles of small RNAs (sRNAs) in IES excision as well as their participation in epigenetic events[26-30]. This mechanism has been widely studied and accepted in IES excision as a model called "the scanning model" (Figure 5). This model, first proposed in *Tetrahymena thermophilia*, outlines how sRNAs control the process of precise IES excision in a manner comparable to RNA interference (RNAi). In the example of *Paramecium*, two kinds of sRNAs have been found to mediate the cross talk between the MAC and MIC during autogamy: early expressed scan RNAs (scnRNAs) and late expressed IES RNAs (iesRNAs). At the onset of sexual development, in the meiosis of the MICs, the genome is bi-directionally transcribed to form longed double-stranded RNA transcripts[31] which are processed by Dicer-like (Dcl) enzymes named Dcl2 and Dcl3[27]. Recently, it has been reported these *Paramecium* enzymes had cleavage selectivity when generating scnRNA, with Dcl2 keeping a solid size preference for 25 nt and a terminal preference for 5' U and 5' AGA, and Dcl3 exhibiting a terminal preference for 5' UNG. Taken together, functional scnRNAs are 25nt in length and exhibit a 5' UNG preference[32].

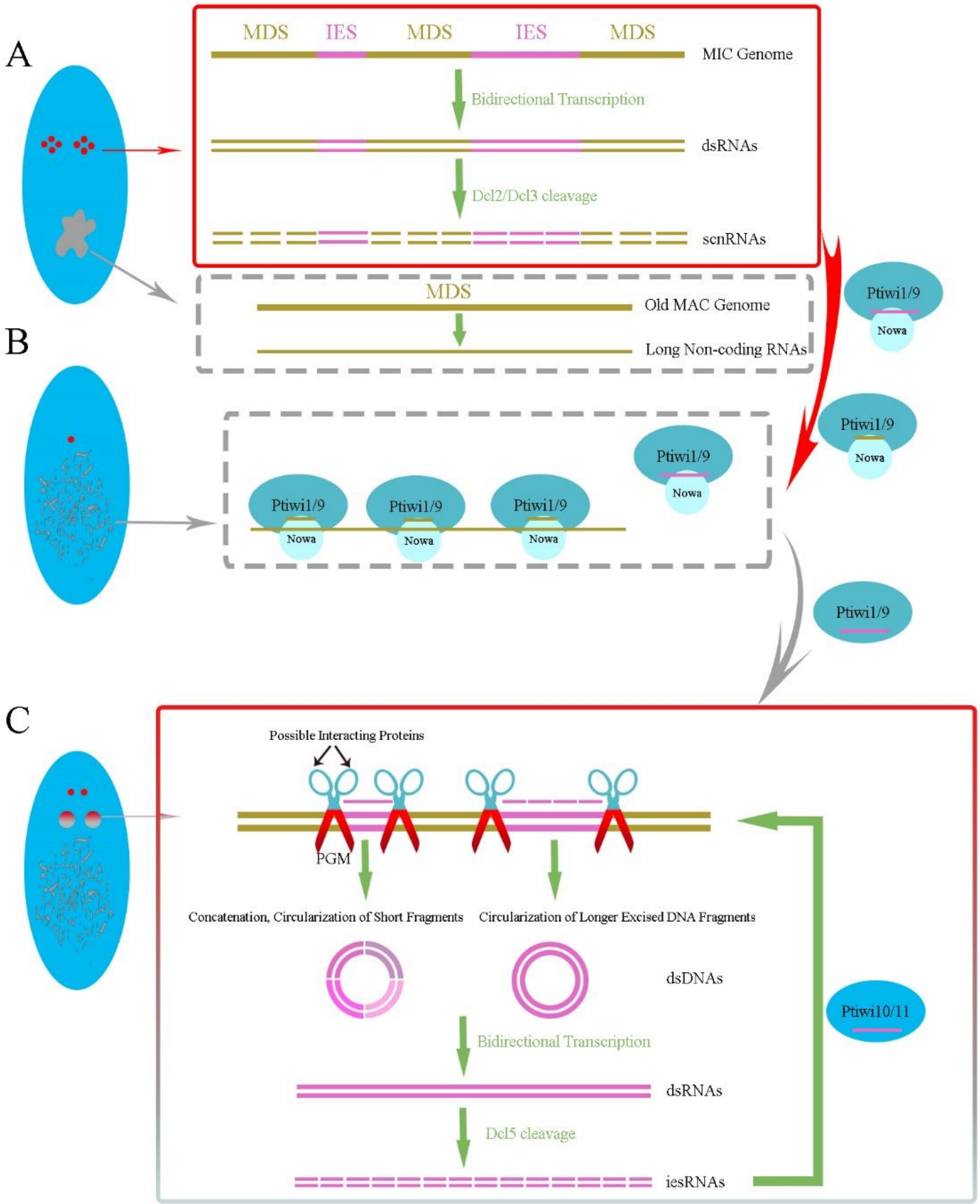


**Figure 5. Scheme elucidating the scanning model in *Paramecium***. A) The meiosis of MIC initiates genome rearrangement. The MIC genome is transcribed bidirectionally, and the dsRNAs are processed into the scnRNAs, which consist of MDS (khaki) and IES (pink) dsRNA, by the Dcl2/3 cleavage. Ptiwi1/9 choose and carry one strand of the scnRNAs, and transport them into the fragmenting maternal MAC. B) This nuclear crosstalk might be mediated by Nowa proteins. The scnRNAs that are unable to match with MAC transcripts are transferred into new developing MACs by Ptiwi1/9. C) The new MACs, being clones of the MICs, include MIC-specific IES sequences, that

must be removed to generate a functioning somatic genome. The DNA regions marked by MIC-specific scnRNAs are supposed to be removed by PGM and co-factors. Shorter excised IESs concatenate then circularise, while longer excised IES circularises directly. The dsRNAs are transcribed from IES circles with a rolling-circle transcription method. Another Dicer-like enzyme, Dcl5, converts these transcripts into the second type of sRNAs: iesRNAs. As the name implies, iesRNAs only map IESs. The iesRNAs are targeted by additional Ptiwi proteins, Ptiwi10/11, as well as anchor more IESs for excision. This procedure is supposed to be a positive feedback loop to ensure all copies of each IES are completely removed.

After being produced in the MICs, scnRNAs are loaded onto specific Piwi proteins, Ptiwi1/9, which remove the matching passenger strand after selecting a guide strand[33] and immigrated into the old MAC. At this time, the old MAC is slowly fragmenting and its genome is transcribed into long non-coding RNAs[34]. The Piwi proteins that transport the scnRNAs are believed to scan those prolonged RNA transcripts with the assistance of Nowa[35]. The scnRNAs that match MDS bind to the transcripts and are inactivated selectively. The scnRNAs only containing germline-specific sequences, on the other hand, cannot match to any sequences from the MAC genome, are those desired scnRNAs. These sets of scnRNAs are carried into the new developing MAC by Ptiwi1/9. This kind of targeting is supposed to be a scan for complementary sequences again. Given that the new developing MAC is derived from MICs, the regions complementary to scnRNAs are IESs and are removed by PGM.

Another kind of short RNA is produced late in the new MAC development from IESs that are going to be removed. These small RNAs, known as iesRNAs, only map IES sequences. They range in size from 26 to 30 nucleotides and have a 5'UAG and 3'CNAU signature, with the characteristic 2nt 3' overhang signature of Dicer/Dicer-like protein activity. Dcl5 is in charge of iesRNA production from a double-stranded RNA precursor, and *DCL5* knockdown entirely prevents its occurrence. Inhibiting IES excision, as with Dcl2/3 co-silencing, significantly reduces iesRNA levels. Thus, iesRNAs are produced from eliminated IESs and function in the positive feedback loop to ensure all copies of each IES removed from the developmental MAC genome[26,32]. Briefly, following excision, these IESs are either joined by alignment of TA overhang when they are longer than 200 bp[36] or similarly ligated with other IESs as long IESs when they are too short for circularisation. These extrachromosomal circles would be the source of the template for iesRNA precursor production. Theoretically, RNA polymerase complexes could interact with IES concatemers, this interaction provides a good chance to generate enough copies of all IESs transcripts in it. As mentioned before, Dcl5 is in charge of iesRNAs production from double-strand iesRNA precursors. This sequence preference, interestingly, matches the ligated IES end conserved sequences of two IESs, or junctions between IESs. According to *in vitro* cleavage experiments, Dcl5 particularly detects these junctions and truncates sRNAs according to the distance between the junction motifs. The elimination would be blocked by mutations in these motifs. Consequently, iesRNA precursors are disassembled back into sRNAs that correspond to individual IESs[27]. Like the chosen scnRNAs, double-strand iesRNA is loaded onto Ptiwi10/11, the second pair of Piwi proteins, which pick a specific strand and carry it to further IES copies in the genome[37]. As IESs are mainly located in the coding regions in the genome, IES excision may be essential for iesRNA generation, as well as for the expression of genes implicated in the pathway. The IES excision in the developmental MAC might be a fundamental method for unblocking genes at the optimal time[37].

There are still serval questions that should be asked in the scanning model. One of them is whether scanning occurs through a DNA: RNA based or RNA: RNA based mechanism. Because both hybrids are found in ciliates by direct and indirect evidence before, a previous member of our lab used Chromatin Associated RNA sequencing (ChAR-seq) to answer the question about DNA: RNA hybrid’s existence in *P. tetraurelia*. Unfortunately, no evidence has been indicated that the sRNA: DNA hybrid presents in the scanning model in our lab. Actually, before answering this question, a prerequisite has to be confirmed whether helicase or similar proteins are included in the RNA-guided IES excision or not. If yes, could it be possible that the IES excision to be coupled with the massive genome amplification in the developmental MAC? If not, then how do the sRNAs recognize double-stranded, highly dense DNA in new MAC? Another question comes from the “cut and repair” mechanism. Following the excision, the ends of the flanking regions are subsequently ligated together in a way similar to the non-homologous end joining DNA repair mechanism. Ligase IV and Ku70/Ku80 have been identified as players to catalyse this repair[19,38]. While the removal of IESs shortens the length of the chromosome when serval shorter IESs or one longer IES (even 4 kb sometime) get truncated. In whatever way of repairing the lesion after excision, the ligation of ends pulls two pieces of chromosome nearer which should be impossible unless there was a mechanism that can close the spatial distance before or during the repairing.

## 2.3. Structural maintenance of chromosomes (SMC)

The genome compaction controls the organization of chromatin and its activity in the nucleus. As shown in Figure 6, in a typical metazoan cell, DNA is condensed up to 10,000 times in length to construct a metaphase chromosome compared to chromatin[39], which means the operating space for functional proteins has been reduced a lot. Since Walther Fleming[40] first defined the dynamic activity of chromosomes during cell division, its duplication and division into two daughter cells are the two essential processes. Chromosomes undergo incremental structural changes that guarantee their segregation with high fidelity, making them more than just passive cargo or information carriers[41]. How this astonishing feat of packing is accomplished and how the chromosomes are reliably segregated into the daughter cells have been central concerns. The SMCs is an abbreviation of structural maintenance of chromosome proteins which have been studied during the last decades. As critical participants in both chromosomal condensation and segregation, the convergence of evidence from genetic and biochemical techniques has been reported[42]. Nearly all investigated living things encompass SMC proteins. They are crucial for mitotic chromosomal dynamics, gene expression regulation, and DNA repair. Their activities in several dimensions of chromosomal activity are conserved in these species[43].

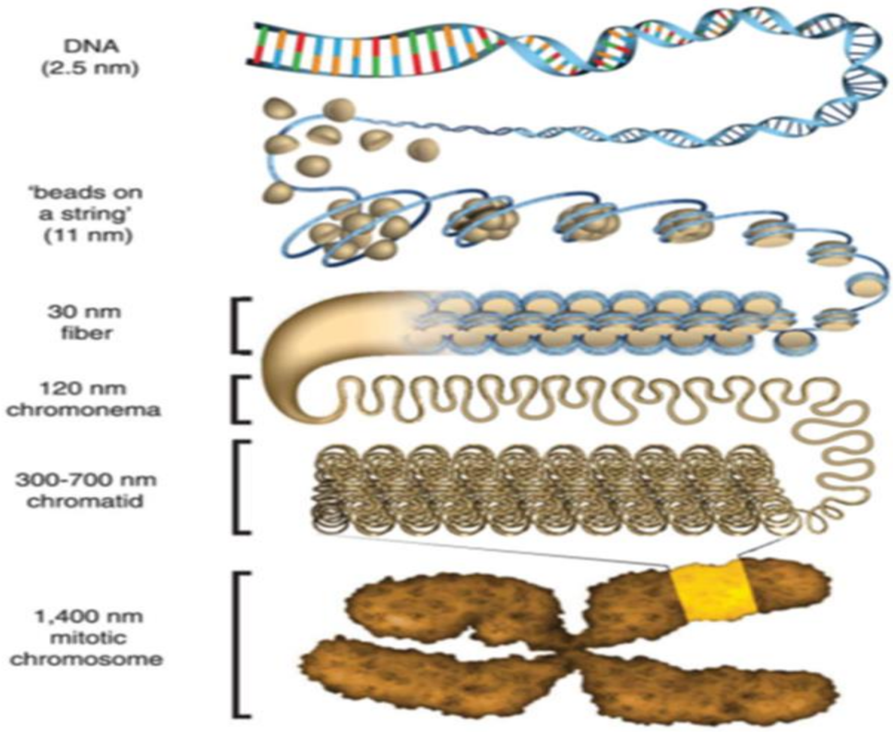

**Figure 6. Hierarchical chromatin-folding model.** Adopted from[44]. DNA is organised into chromatin structures in the nucleus, which controls the activity and inheritance of human genomic DNA. A 147 bp long DNA strand is wrapped around an octamer of histones H2A, H2B, H3, and H4 to form an 11-nm DNA-core nucleosome particle. Each DNA-nucleosome particle is separated by 20 to 75 bp of histone H1-binding DNA. This primary nucleosome polymers further fold into secondary 30 nm fibres. There are two different structural models of the 30 nm fibres which refer to as the zigzag and solenoid fibre models. The solenoid fibre structure is 33 nm in diameter and includes six nucleosomes every 11 nm along the fibre axis. The diameter of the two-start zigzag fibre is 27.2 to 29.9 nm, with five to six nucleosomes per 11 nm. It is assumed that the 30 nm fibre assembles into helically folded 120-nm chromonema, 300 and 700 nm chromatids, and mitotic chromosomes.

Most DNA-based activities are influenced by SMC complexes. Each 1,000–1,300 amino acids long, 110 KDa to 170 KDa SMC protein comprises a core globular hinge domain. Recent research has revealed that the hinge domain is the primary place where these molecules dimerize[45,46]. The hinge interface preferentially binds short single-strand DNA (ssDNA)[47] and can spread structural changes to the head domains via the coiled-coil regions[48,49]. A globular domain containing a Walker A (GxxGxGKS/T) or Walker B motif (XXXXD, where X is any hydrophobic residue) (amino acid consensus sequences present in NTP-binding proteins named after J. E. Walker and co-workers, who initially discovered these motifs[50]) flanks this domain and is bridged by two extended coiled-coils domains. The coiled-coil regions can transmit conformational changes induced by ssDNA interaction of the hinge domain to the head domains which triggers ssDNA-stimulated ATP hydrolysis, involving head-to-head disengagement[51]. The protein is folded at the hinge, allowing the coiled domains to interact in an anti-parallel manner[52]. As a consequence, a functional ATPase domain that resembles an ATP-binding cassette (ABC) domain structurally is formed by joining the amino and carboxyl termini[53,54]. In such a complex, two SMC proteins bind at the hinge region to create long-armed V-shaped dimers[45,55], which are then attached to non-SMC subunits that are distinctive to the complex[56-58]. A member of the kleisin protein family brings together two head domains of the V-shaped SMC protein dimers, allowing SMC complexes to carry on ring-like shapes[59,60].

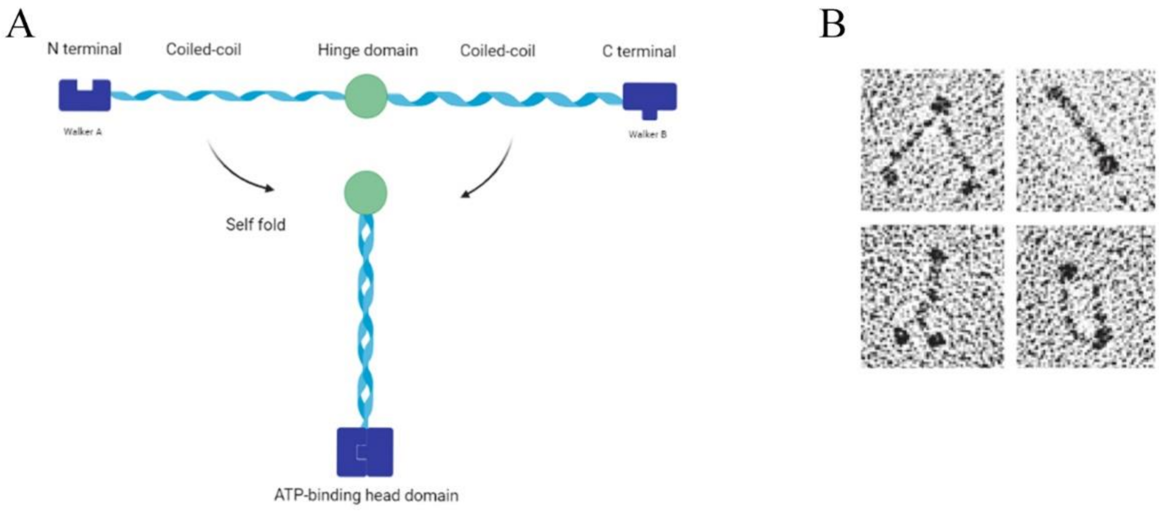


**Figure 7. The architecture of SMC protein and SMC homodimers.** A) A SMC protein's fundamental structure. The SMC self-folds through antiparallel coiled-coil interactions to create a hinge domain end and an ATP-binding head domain end. B) Dimerization is mediated by a hinge-hinge connection between two subunits, which results in a V-shaped molecule. The SMC homodimers display a wide range of conformations[61].

Only one gene of each SMCs has been discovered in each of the prokaryotes examined, and it appears to form homodimers. Based on the sequence homologies, this family can be classified into six subfamilies in eukaryotes[62] as in Figure 8. And the protein appears to form at least three types of heterodimers. These heterodimers involve the SMC1/SMC3 in the cohesin complex, the SMC2/SMC4 in the condensin complex, and the SMC5/SMC6 heterodimer, which forms a core part of a complex participating in DNA repair[63].

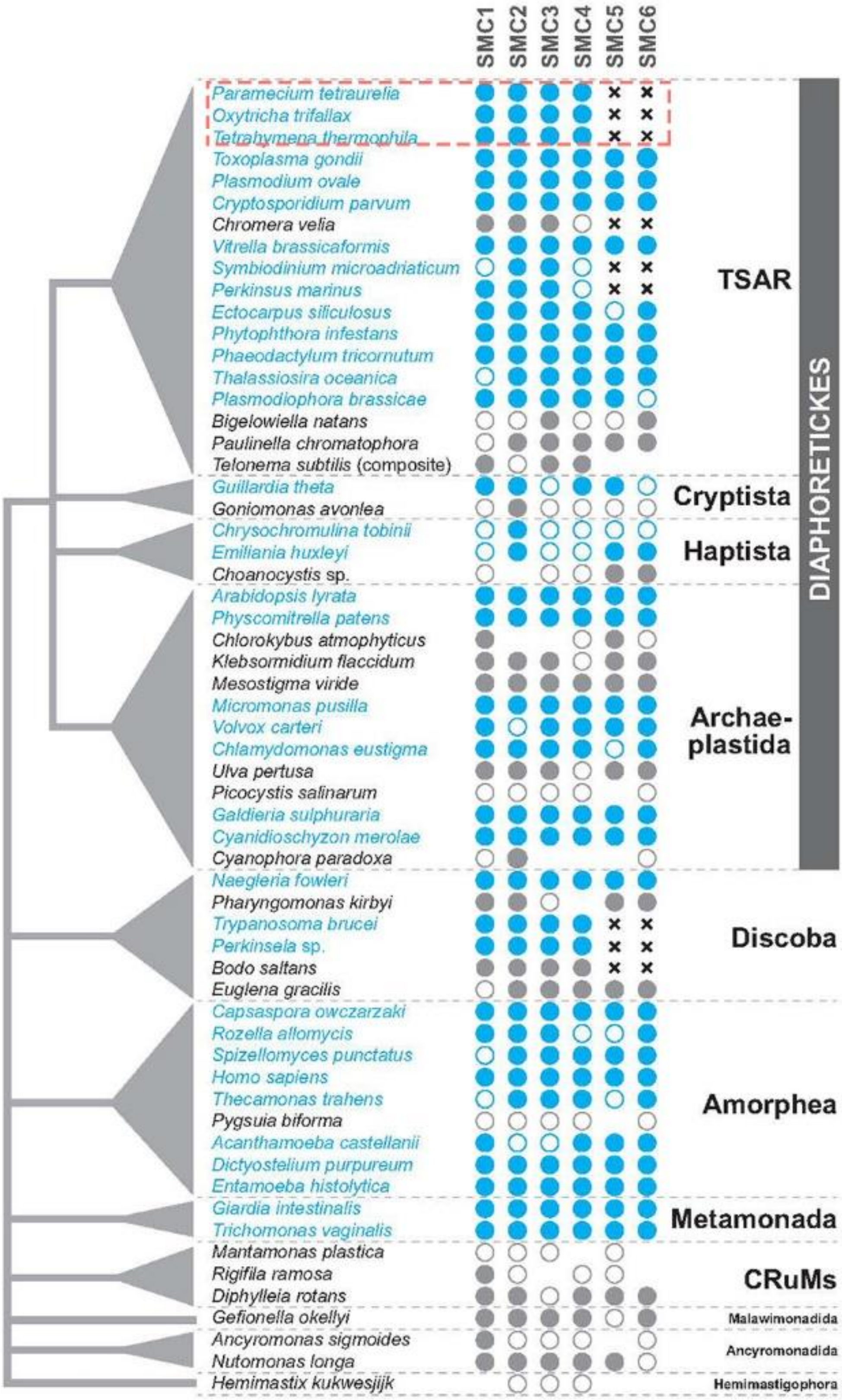


**Figure 8. Exhibition of six SMC subfamilies in eukaryotes.** Adopted from[64]. SMC sequences were found in 59 species, which represent the main taxonomic assemblages in the eukaryotic tree. On the right, symbols of each species present the presence or absence of all these SMCs. The SMC5/SMC6

absent in those eukaryotes (x-marks) were indicated. When a subfamily was not discovered in the relative transcriptome database, no symbol in the relevant column was used to show undetermined.

### 2.3.1. The structure and function of condensin

Observations of mitotic chromosome kinematics in cytology first suggested the presence of molecules that compact chromatids or maintain sister chromatids together. The recovery of mutants deficient in these processes, together with the biochemical identification of mitotic chromosome-associated components, gave rise to the finding of condensin and cohesin[65]. Condensin and cohesin both include the SMC family proteins at their cores, but the two complexes are structurally and functionally distinctive[42,63]. Cohesin holds sister chromatids together until anaphase, before they separate, whereas condensin restructures chromosomes into their compact mitotic configuration[66].

As shown in Figure 9, five subunits comprise the highly conserved condensin that utilises ATP to drive conformational changes in DNA, allowing it to address DNA compaction, architecture, and segregation from bacteria to humans. Higher eukaryotes generally have two condensin complexes (I and II), both of which consist of SMC2 and SMC4 heterodimers as the core part. The main differences between the two condensins are their non-SMC subunits, CAP-D2, CAP-H and CAP-G present in condensin I, while CAP-D3, CAP-H2 and CAP-G2 consist of condensin II[67,68]. Different species have different condensin I/II ratio. The condensin I/II ratio in human HeLa cells is 1:1, but 5:1 in the Xenopus egg, and 10:1 in isolated mitotic chromosomes from chicken DT40 cells[69]. Interestingly, *Caenorhabditis elegans* (*C. elegans*) contains three distinct condensin complexes, condensin I and II, as well as a third one, condensin $I^{DC}$, which exclusively acts in dose compensation[70]. In *T. thermophila*, a specialized condensin (condensin D) that does not appear to function in chromosome segregation has been identified by Howard-Till and co-workers[71]. It is essential for the completion of sexual reproduction and the developmental program of the somatic nucleus including DNA elimination. Condensins I and II also show distinct temporal and spatial localization profiles in higher eukaryotes. Condensin I is predominantly located in the cytoplasm in interphase and engages chromosomes just following the breakdown of the nuclear envelope in prometaphase, on the other hand, condensin II is mostly nuclear localization and attaches chromosomes when condensation occurs in prophase[72,73]. A combination of two functional assays revealed that condensin II interacts with duplicated regions of chromosomes[74]. Mutation or removal of condensin subunits in a diversity of species disturbs appropriate chromosome condensation, resulting in chromosome segregation errors and cell death. Even though the two mitotic condensins are highly comparable, their pattern of localization implies that they may perform unique functions in the chromosomal organisation. Consistent with this assumption, depletion of either condensin I or II results in unique chromosomal defects, whereas depletion of both condensins results in more severe defects[68,75].

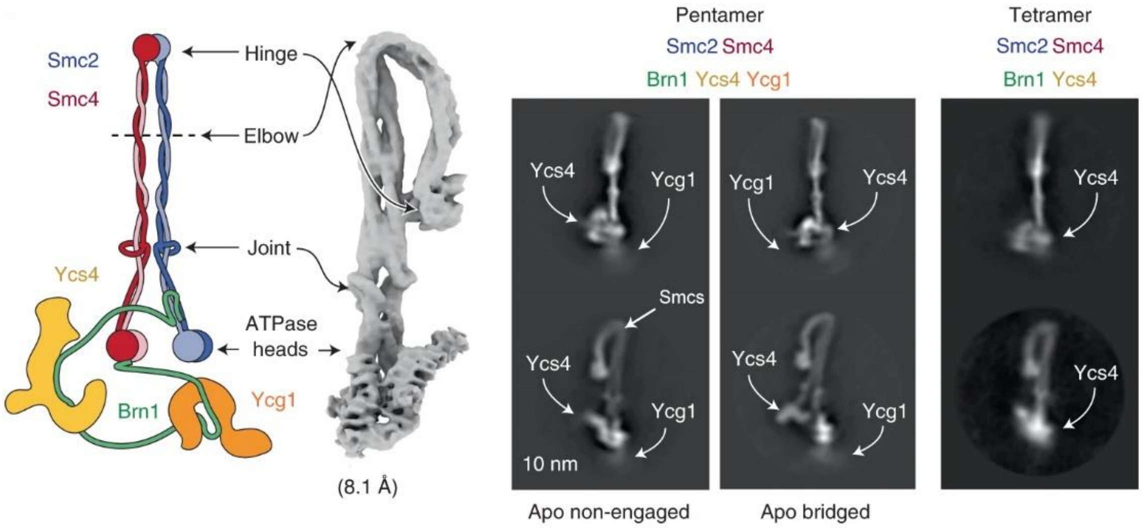


**Figure 9. Cryo-EM structures of the yeast condensin holo complex in the nucleotide-free apo form.** Adopted from[76]. The left part is the schematic model of condensin, and the right part shows the 8.1-Å-resolution 3D map that shows its overall architecture.

Condensin has functional differences across species. Although it was previously thought that, in most organisms, condensin only interacts with chromosomes during the cell cycle when they are condensed (prophase to anaphase), an increasing amount of data have revealed that condensin also plays important functions in controlling the interphase genome[77]. Even though they are widely distributed across the genome, genes that act in comparable processes frequently inhabit similar areas of the nucleus[78]. Condensin associats with DNA replication termination and genome integrity maintenance in the nucleolar organiser territory, which contains ribosomal DNA repeats[79,80]. Condensin both promotes the clustering of distributed loci into subnuclear regions and inhibits homologous connections[81,82]. Condensin II dissolved synapsed chromosomes into individual homologous chromosomes in *Drosophila* ovarian nurse cells[83]. This unpairing activity causes chromosomal compaction during interphase[84]. It has been demonstrated that yeast condensin has a function in interphase chromatin architecture and RNA polymerase III transcribed gene clustering too. Condensin-mediated localization of these genes in the nucleus aids in the three-dimensional organisation of the genome in budding and fission yeast[85]. Mutations in yeast condensin subunits result in tRNA gene stationing deficiencies and partially block tRNA gene-mediated silencing[86], revealing another association between genome organization regulated by condensin and gene expression[87].

As cells advance through the cell cycle, chromosomes undergo substantial changes in structure. During mitosis, chromosomes are compressed into dense arrays of arbitrarily positioned sequential chromatin loops. There is growing evidences that the establishment and extension of DNA loops, also known as loop extrusion[88], is the fundamental mechanism behind SMC complexes' ability to organise DNA. Several putative models for loop extrusion by SMC complexes have been proposed. These models are exhibited in Figure 10. In DNA damage repair, the initial loading of condensin to DNA was thought to occur via the SMC hinge-DNA interactions at the ssDNA region which is generated behind the replicative helicase[89]. Chromosome-conformation capture (Hi-C) on topological domains[90,91] and polymer simulations[92] suggested the formation of such DNA loops, whereas recent *in vitro* single-molecule studies provided experimental evidence of condensin's DNA translocase activity and ability to extrude DNA loops[93]. Also, evidence of loop extrusion by real-time imaging of the formation and processive extension of DNA loops by yeast condensin has been reported in 2019[94]. A direct DNA-binding site in the eukaryotic condensin complex formed by Ycg1$^{Cnd3}$ HEAT-repeat and Brn$^{Cnd2}$ kleisin subunit has been identified which serves as a safety belt that prevents DNA dissociation from the groove[95]. The extrusion can be formed in an ATP hydrolysis-dependent, strictly asymmetric manner by condensin. Recent research results indicate that the active condensin complex transiently entraps the bases of a DNA loop in two distinct chambers. Single-molecule imaging and cryo-electron microscopy point to a possible power-stroke movement at the first chamber, which feeds DNA into the SMC-kleisin ring in response to ATP binding, whereas the second chamber remains upstream of the identical DNA double helix. By removing the stringent separation of the "motor" and "anchor" chambers, condensin is transformed from a one-sided to a bidirectional DNA loop extruder. It is inferred that the directionality of DNA loop extrusion is determined by the direction of two topologically coupled DNA segments during the SMC reaction cycle[96]. These properties mainly mean that condensin makes stable contact with DNA at a binding site of non-SMC subunits and then reels the DNA from only one side.

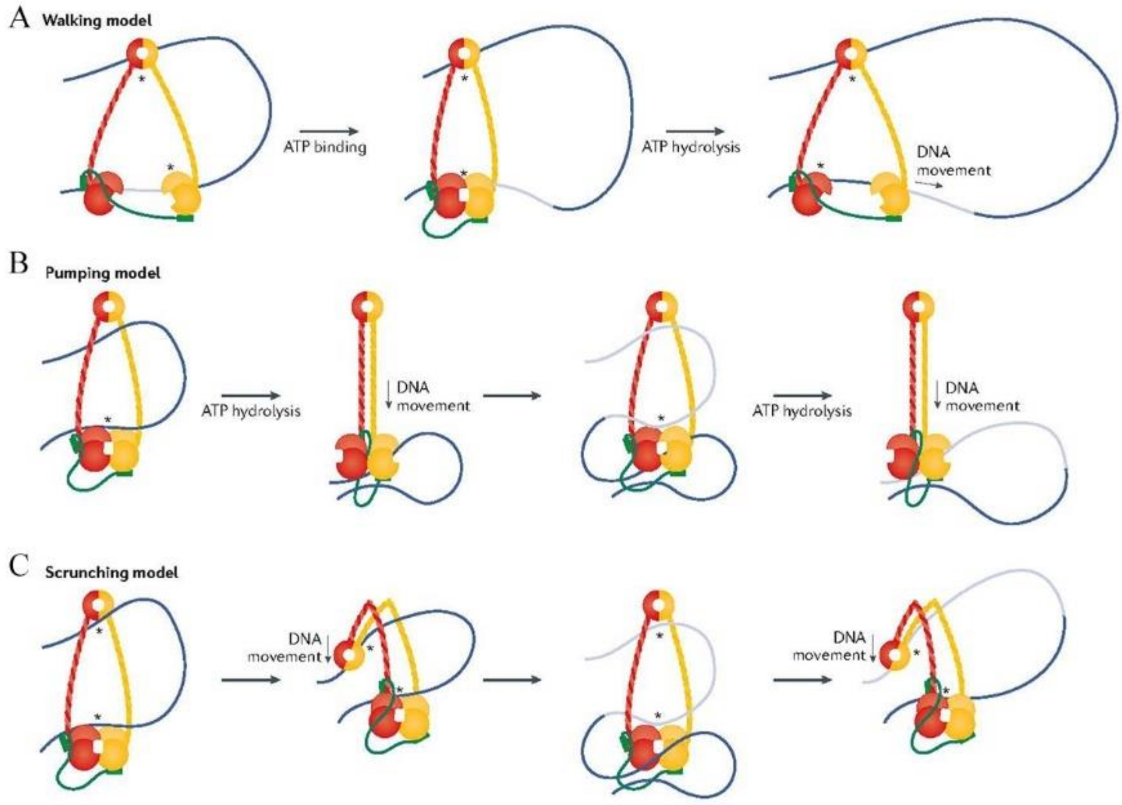


**Figure 10. Putative models of loop extrusion.** Adopted from[97]. A) The walking model. During the SMC ATPase cycle, DNA segments (blue) are passed across SMC ATPase domains. If the hinge domain connects with a distal DNA segment during this process, loop extrusion may occur. If the hinge has a poor affinity for DNA and glides along it, symmetrical loop extrusion may occur. B) The pumping model. The SMC complex encircles a DNA loop and binds to it via the ATPase head domains. ATP hydrolysis causes the SMC coiled coils to align and pushes the DNA loop towards the ATPase heads. If the SMC complex encircles a second DNA loop (grey) after the head engagement, loop extrusion might occur. C) The scrunching model. At the ATPase heads and the hinge, the SMC complex attaches to DNA. If these transitions are connected to the SMC ATPase cycle, a folded conformation may facilitate the handover of a DNA segment from the hinge to the heads or vice versa, resulting in DNA translocation. If one DNA segment is constantly attached throughout this process, asymmetrical loop extrusion may ensue. If DNA binding shifts between the loop's two arms, symmetrical loop extrusion may occur.

Mitotic chromosomes fold as compact arrays of chromatin loops. An emerging model of prometaphase chromosomes is shown in Figure 11. As in chicken cells, the chromatin loops nested in a way of ~400 kb outer loops generated by condensin II split up by ~80 kb inner loops formed by condensin I. Chromosomes shrink as prometaphase progresses due to increased helical winding, with the number of loops per turn rising. Consequently, the size of a helical turn ranges from ~3 Mb (~40 loops) to ~12 Mb (~150 loops)[91].

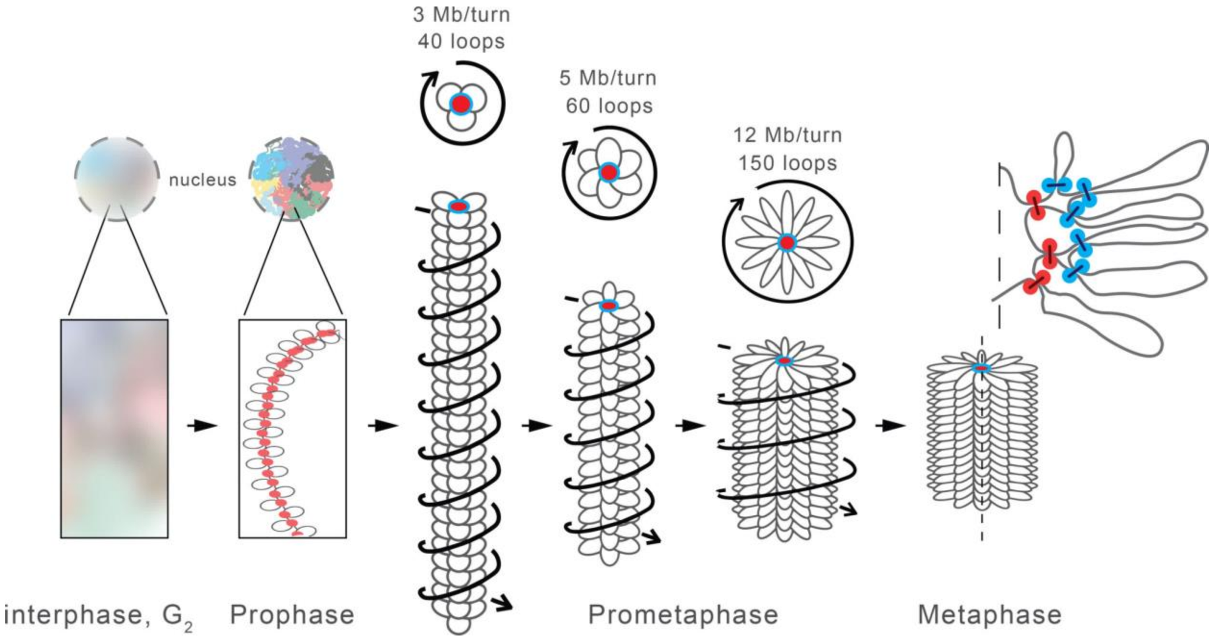


**Figure 11. Schematic model of mitotic chromosome morphogenesis pathway**. Adopted from[91]. Condensin II compacts chromosomes into arrays of successive loops and sister chromatids divide along their length during prophase. In red is the scaffold of condensin II-mediated loop bases. Condensin II-mediated loops become more enormous when the nuclear envelope breaking and the cell enters prometaphase, splitting into smaller 80 kb loops via condensin I. Chromosomes are depicted as loop arrays. The nested configurations of condensin II-mediated loop bases in the centre and condensin I-mediated loop bases in the periphery are depicted in red and blue, respectively. The core scaffold develops a helical configuration during prometaphase, with loops spinning around the scaffold as steps in a "spiral staircase" (the helical path of loops is indicated by arrows). Outer loops develop and the number of loops per turn grows as prometaphase advances, and chromosomes shrink to become the mature mitotic chromosome.

In addition to the loop extrusion, a new evidence shows that condensin might interact with the RNA-DNA hybrid and remove the RNA, followed by the annealing of ssDNA to form dsDNA[89]. Before mitosis, condensin competes with other DNA-bound components and unloads them from the chromosomal region to generate a clean chromosome. The promotion of ssDNA annealing could be required to facilitate their faithful segregation[98]. Studies in *S. pombe* also suggested the involvement of condensin I in DNA repair, which indicates that condensin I is recruited at damage sites of single strand break (SSB) through poly (ADP-ribose) polymerase 1 (PARP-1)[99]. A single condensin molecular moving along DNA can explain the introduction of positive supercoils into closed circular plasmids in the presence of ATP[100-102].

### 2.3.2. The SMCs in ciliates

As in Fig. 8, SMC1/2/3/4 are present in the genome while SMC5/6 is absent from ciliates. Only a few papers reported some functional analysis of condensin in *Tetrahymena*, and the functions of condensin in *Paramecium*, another ciliate also in the class Oligohymenophorea, have not been investigated till now. At any point throughout the cell cycle, there is no apparent chromosomal condensation in the MAC in *Tetrahymena*. The techniques needed to identify chromosomal condensation in *Saccharomyces cerevisiae*[103], which has chromosomes equivalent in size to *Tetrahymena* MAC chromosomes, are challenging to apply to *Tetrahymena*'s polyploid macronucleus. As mentioned before, the chromosome size in MAC ranges from 50kb to 1Mb, while the loop extrusion makes nested loops with 80kb inner loops from 400kb outer loops in DT40 cells. In extreme cases, condensin could not even extrude even one loop in MAC if the loop size was conserved in all organisms. We have to be careful about imitating those conclusions, especially the loop extrusion, from other organisms to *P. tetraurelia*. Corresponding to the concern, some surprising roles for condensin in *Tetrahymena* amitosis have been reported. First of all, SMC4 was found in both MIC and MAC during the vegetative growth, while the knockout of SMC4 (SMC4 KO) generates an apparent MAC segregation defect. In contrast to condensin depletion in other organisms, SMC4 KO in *Tetrahymena* did not affect the segregation of bulk DNA during the mitosis which could be because of the lack of microtubule elongation inside the MAC[104]. Another study has indicated that condensin works in distinct ways on segregating chromosomes in the MIC and MAC of *Tetrahymena*. Condensin facilitates the condensation and resolution of MIC chromosomes and promotes the spatial distribution and segregation of shorter chromosomes in MAC. They also believed that the loop extrusion could be responsible to separate copied shorter chromosomes in MAC[105]. A development-specific condensin complex, condensin D, has been identified in *Tetrahymena*. Unlike canonical condensin, condensin D is crucial to the completion of sexual reproduction and the development of MAC. This is the first report mentioning that the condensin complex is supposed to promote genome organization in DNA elimination by an indirect way[71].

## 2.4. PiggyMac, a key protein in IES excisions

PiggyMac (PGM) is a domesticated transposase identified in *P. tetraurelia*. Although IES sequences resemble Tc/mariner transposable elements, the excision process is more akin to that seen in PiggyBac Transposons (Figure 12A). Tc/mariner transposons perform a cut and paste transposing which results in a 3 bp overhang. In contrast, the transposition of PiggyBac transposons induces double strand breaks (DSB) that creates 4 bp 5' overhangs, as seen in Paramecium's IESs[24,106]. The PGM is a probable candidate which participates in IES excision since it has high homology with PiggyBac transposase and was shown to possess the catalytic triad, three conserved aspartic acids (D401D491D609). A downstream cysteine-rich (CR) domain and a C-terminal extension are predicted to adopt a coiled-coil (CC) structure which is required for the normal function of an active IES excision complex[107]. The CR domain has been considered to be crucial for PGM’s activity *in vivo*. It can bind two $Zn^{2+}$ and form a cross-brace zinc finger. Additionally, the CR domain can connect with the N-terminal residues of histone H3 *in vitro*[108]. Experiments revealed that the knockdown of PGM resulted in a fatal phenotype and impaired the excision of most IESs. Using a GFP-tagged PGM protein, they also demonstrated that PGM locates in the developmental MAC during the IES excision. As a result, it was suggested that PGM is the endonuclease responsible both for the accurate elimination of IESs and for the imprecise removal of repetitive sequences in the germline[20].

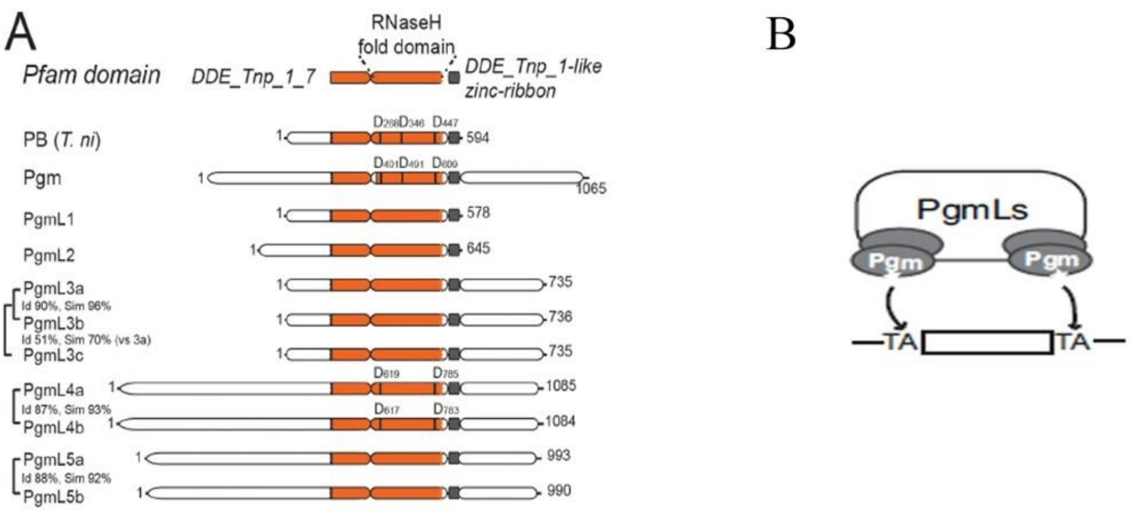


**Figure 12. Domain structure of the PiggyBac transposase (PB) and model of IES excision by PGM/PGMLs complex.** Adopted from[109]. A) The Pfam domain DDE_Tnp_1_7 is shown in the orange domain, at its right side is the RNase H fold (conserved catalytic D residues are indicated by

vertical bars). Id indicates the percentage of sequence identity; Sim presents the percentage of similarity. B) this model is summarized according to previously published data. The active form of PGM is supposed to be a dimer. In the lack of knowledge of the complex's composition, one PGM homodimer is represented at each IES border, with all PGMLs forming a huge bridge structure. PGML subunits are postulated to drive the precise alignment of the PGM active site to those TA cleavage sites in a fully formed complex.

Besides PGM, five groups of distant PGM-like proteins (PgmLs, Figure 12) that are all able to interact with PGM and are essential for its localization and IES excision have been identified in *P. tetraurelia*[109]. A co-factor of PGM in DNA elimination, Spt16-1, a subunit of histone chaperone FACT which is the key to reorganising nucleosomes and ensuring chromatin integrity, facilitating the localization of PGM[110]. The protein Ku80c has also been shown the unique property to license PGM-dependent DNA elimination[111]. The DSB repair at those excised spots is accurate and efficient, ensuring that MAC chromosomes are rebuilt in the correct sequence and open reading frames are regenerated with accuracy[13]. The ligase IV-Xrcc4 complex, an important player of the C-NHEJ pathway[112], was demonstrated to repair DSBs at chromosomal IES elimination positions and during the circularization of eliminated IESs using RNAi-mediated functional investigation of candidate genes[19].

# 3. Aim of the thesis

Aim I:

The first aim of this research was to explore the roles of condensin components in the IES processing in *P. tetraurelia*. So far, no such information has been reported and it is worth noticing that the chromosome structure, especially constructed by condensin, should be an important aspect for the processing of some or all IESs *in vivo*. Hence, to gain a more in-depth look at the relationship between condensin and IES excision, we first analysed the possibility of condensin components participating in IES removal by RNAi. When it was confirmed, additional functional analysis was included to elucidate the roles of condensin in IES excision. The expected results may contribute to supplementing the mechanism of genome rearrangement in *P. tetraurelia*, and provide a new vision into the activity of condensin in other organisms.

Aim II:

In addition to the research work on condensin, we set out to investigate the members of excision machinery in two different ways. The excision machinery is the most interesting complex for people who focus on the mechanism behind the IES excision. Until now, nobody has succeeded in isolating this machinery *in vivo*. So, for the first time, we used biotin-labelled IES complementary oligos to pull down the DNA excision machinery in IES excision. Mass spectrometry was used to analyse the elution to identify the components interacting with IESs. The second way to purify the eliminating complex is mutant PGM Co-immunoprecipitation. Since the PGM is supposed to be the core component of excision machinery according to its functional analysis, mutated PGM with tag in theory should be purified with the IES excision machinery. Also, mass spectrometry analysis is included.

# 4. Results

## 4.1. Part I: Research article: Functional analysis of structural maintenance of chromosomes (SMCs) in internal eliminated sequences (IESs) elimination in *Paramecium tetraurelia*

Author Contributions:

I conducted the design of the experiments, carried out all experiments, wrote the manuscript, produced and formatted the figures.

Sebastian Bechara performed all bioinformatics analysis and create the original version of Figure 2, Figure 4 and Figure 5.

**Functional analysis of structural maintenance of chromosomes (SMCs) in internal eliminated sequences (IESs) elimination in *Paramecium tetraurelia***

Fukai Zhang[1], Sebastian Bechara[1], Mariusz Nowacki[1*]

[1] Institute of Cell Biology, University of Bern, Bern 3012, Switzerland

[*] Corresponding author: Mariusz.nowacki@unibe.ch

Abstract

The structural maintenance of chromosomes (SMCs) is a large family of ATPases that play a variety of roles in the organisation and dynamics of higher-order chromosomes. They are central regulators of chromosome dynamics and the core component of condensin and cohesin. The internal eliminated sequences (IESs) excision is controlled by a lot of different factors that occur between two forms of nuclei in *Paramecium tetraurelia*. During this process, massive genome condensation and dissolution happen with several rounds of meiosis and mitosis where the SMCs play vital roles in organizing chromosomes. While the relationship between IES elimination and SMCs function has not been discovered in *Paramecium tetraurelia* until now. Here we applied RNA interference, genome sequencing, mRNA sequencing, immunofluorescence and mass spectrometry to investigate the roles of SMCs in IES elimination in *Paramecium tetraurelia*. The results indicate that SMC4-2 is a novel factor of IESs excision in *Paramecium tetraurelia*, while SMC4-1 is not. The localization of SMC4s also suggests that SMC4-1 seems like the canonical SMC4 in other organisms, while SMC4-2 is more like an IES-specific SMC4 which only exists in *Paramecium tetraurelia*. Loss of SMC4-1-GFP in old MAC brings up the possibility that decompacted chromosomes help the scanning of scnRNAs. The genesis of iesRNAs is strongly inhibited by the knockdown of SMC4-2. Silencing SMC4-2 with SMC2-1, SMC2-2 or SMC4-1 abolished the retention phenotype is a new finding too. The enrichment of different proteins from SMC4-1 and SMC4-2 suggests a possibility of a competitive relationship between them. These innovations highlight the potential of SMCs participating in IES processing and which in end will lead to a new mechanistic discovery.

## Introduction

*Paramecium tetraurelia* is a widely distributed unicellular eukaryote in fresh water in the highly diverse ciliate phylum. Similar to other ciliates, the nuclear dimorphism makes them a talent model to explore genome dynamics and RNA mediated epigenetic regulation between their two diverse forms of nuclei. The germline micronucleus (MIC) serves as the genetic information stock next generation. The somatic macronucleus (MAC) is responsible for gene expression. Genetic information comparison between MAC and MIC results in large-scale and precise elimination of specified DNA regions from the developmental MAC genome. The packing of meters of linear DNA molecules into compact metaphase chromosomes is an inevitable step in meiosis and mitosis which makes condensin could be important to the genome rearrangement.

This sexual cycle can be triggered by food stress and conjugation which can induce MAC fragmentation and MIC meiosis. The two new developing MACs are derived from two copies of MIC during the sexual cycle after one mitosis and two mitotic divisions of MICs. Endoreplication in the new MAC increases the ploidy from 2 n to 800 n. During this time, massive genome rearrangements occur which reproducibly remove the internal eliminated sequences (IESs). There are almost 45,000 annotated IESs within the new developing MAC. Precise excision is required for the functional maturation of genes containing IESs. No obvious feature has been found to clarify the precise removal of IESs except that most are flanked by TA[1]. Two sets of small RNA defined as scnRNAs (‘scan’ RNAs)[2] and iesRNAs[3] are supposed to be included in IES excision in early and late *P. tetraurelia* development respectively. In brief, during meiosis, the MIC is bi-directionally transcribed, yielding long transcripts that are then processed to yield scnRNAs. "Scanning" for comparable sequences in the parental MAC removes MAC genome-matching scnRNAs while enriching MIC genome-matching ones[2]. Later in the development of the MAC, excised IESs circularize or concatenate to serve as a template for iesRNA precursors. Following that, generated iesRNAs assure the eradication of the majority of IESs[3,4]. While mitotic and meiosis chromosomes are compacted with condensin, how the scanning model works under a compacted chromosome is still unknown.

Condensin is a large protein complex and plays critical roles in chromosome structure and segregation in cells[5]. Most of eukaryotes have two different types of condensin, condensin I and II. Condensin has a conserved dimer structure of two large proteins structural maintenance of chromosomes (SMC), SMC2 and SMC4, involved in organizing the genome by using the energy from ATP hydrolysis[6]. Both SMC2 and SMC4 self-fold to a head-to-end shape by the two ATPase domains. The hinge domain is situated in the heart of this V shape structure, the anti-parallel coiled-coil domains connect ATPase domains and the hinge domain. Dimerized by the hinge domains from SMC2 and SMC4, the core of the condensin forms[7]. In human[8] or other eukaryotes[9,10], condensin I and II regulate chromosome assembly and segregation differently in both meiosis[11] and mitosis[12]. The condensin complex connects to chromatin and progressively extrudes a DNA loop by binding DNA on one side and reeling from another side. Furthermore, chromosomes organized as nested loop arrays winding around a helical ‘spiral staircase’ within a cylindrical chromatid by condensin I and II reveals the conformation of mitotic chromosomes[13]. In addition to condensing chromatin into organized structures, condensin has been indicated such as single strand DNA (ssDNA) binding preference[14], reannealing complementary ssDNA[15], remove ssDNA binding proteins[16] since condensin was first reported.

The relationships between SMCs function and IES elimination in *P. tetraurelia* are supposed to be important. Because we assume that the compacted DNA structure should transiently loosen when an IES inside the DNA loop is being excised. It has been declared that the steric hindrance generated by condensin could impact or inhibit transcription[17]. As mentioned above, massive transcriptions are needed for the IES excision, the hypothesis that IESs cannot be eliminated if the corresponding DNA region was compacted into dense chromosomes could be reasonable. In the present research, we assessed the effects of two homologs of SMC4 (SMC4-1, SMC4-2), and SMC2 (SMC2-1, SMC2-2) in *P. tetraurelia*. Strong IESs elimination failure is observed after RNA interference (RNAi) of SMC4-2. It indicates a direct/indirect effect of condensin proteins on IES excision. SMC location indicates that SMC4-1 participates in the whole development of *P. tetraurelia*, while SMC4-2 mainly exhibits in the late stage. Interacting proteins of SMC4-1 or SMC4-2 have also been investigated to shed the light on the functional complex of SMC4s in *P. tetraurelia*. Therefore, this study contributes to research on SMC4s in *P. tetraurelia* by demonstrating the functional difference between them in IES excision.

## Methods

### *Paramecium* cultivation

All experiments were performed with mating-type 7 of strain 51 of *P. tetraurelia*. *Klebsiella pneumoniae*-infected wheat grass powder medium (WGP; Pines International, Lawrence, KS) with 0.8 mg/L of β-sitosterol was treated to cultivate the cells (Merck). As previously mentioned, cultivation took place at a temperature of 27 °C[18,19].

### Sequence alignment, domain prediction and phylogenetic trees generation

The *Paramecium* SMC4-1 (PTET.51.1.P0410063) and SMC4-2 (PTET.51.1.P0590135) sequences extracted from ParameciumDB (https://paramecium.i2bc.paris-saclay.fr/). Other SMC4 protein sequences were collected from previous reports[20,21]. Multiple sequence alignment was carried out in MAFFT[22] and BLAST tool in ParameciumDB. Conserved domains prediction of SMC4-1 and SMC4-2 were performed with Pfam[23]. Expression pattern builds according to previous report[24]. The phylogenetic tree was generated by using the W-IQ-TREE with the default setting[25].

### Gene silencing of SMC2s and SMC4s

As a subunit of SMC part in condensin, the role of SMC2 in *P. tetraurelia* development needs to be explored at the same time. Similar to SMC4s in *P. tetraurelia*, SMC2-1 (PTET.51.1.G0330075) and SMC2-2 (PTET.51.1.G0450077) are annotated in the database. Knock-down (KD) of SMC2s and SMC4s were reached using RNAi by feeding double-strand RNAs as described before[26]. Candidates' sequences were magnified from wild type MT7 genomic DNA with the primers in Supplementary Table S1. Next, either single or dual sequences were inserted inside of the two reversed T7 promoters in the L4440 vector[27]. The plasmids were expanded in the feeding cells: HT115 (DE3) *Escherichia coli*. An empty L4440 vector without insertion was the negative control. Additionally, a PiggyMac (PGM)

RNAi plasmid from our lab was chosen as a positive control[28]. Cross-silencing of other *P. tetraurelia* genes, according to RNAi off-target tests performed using the ParameciumDB[29] tool (https://paramecium.i2bc.paris-saclay.fr/cgi/tool/rnai_off_target), is unlikely. Around 200 cells/ml of *P. tetraurelia* cells were transplanted into the silencing medium. A total of 14 single cells that had completed sexual reproduction in the silencing media were then isolated and added to a freshly bacterized medium to assess the survival of the progeny following autogamy. Three days following their separation, cells were checked and recorded into three groups based on the observed phenotype (healthy, weak and death).

### DNA extraction, IES PCR and Illumina Sequencing

Total DNA from 100 mL of each postautogamous culture was extracted with the GenElute Mammalian Genomic DNA Miniprep Kit (G1N70-1KT, Sigma-Aldrich). IES retention PCR was analyzed with genomic DNA and certain primers as previously described[3]. For deep sequencing, DNA of developmental MAC from 400 mL cells was extracted as previously described[30]. According to established Illumina techniques, a 150-cycle paired-end Illumina TruSeq DNA library was created and sequenced at the University of Bern's NGS platform.

### Calculation of IES retention scores (IRSs) in genome-wide and correlation matrix

The IRSs were estimated using ParTIES[31]. The number of reads associated with the eliminated IESs with just the MAC IES junction is denoted as IES, whereas the number of reads which accommodates the IES region is denoted as IES+. Only read pairs that were mapped were counted. Each read was tallied just once to prevent excessive counting brought on by paralogous matches. To prevent length biases brought on by IES length variance, reads were exclusively counted at IES ends. Then, an IRS is determined as follows: IRS = $IES^+/(IES^+ + IES^-)$. Correlations were estimated with the Pearson method.

### Total RNA extraction, mRNA sequencing and small RNA (sRNA) analysis

Total RNA was obtained from 200 mL of *P. tetraurelia* at 4 hours after 100% fragmentation. This timing is determined according to the strong GFP signal of SMC4-2 on the western blot. TRI reagent (Sigma-Aldrich) extraction was done following the suggested protocol. The mRNA library was produced based on standard Illumina protocols and sequenced at the NGS platform at the University of Bern. For sRNA sequencing, total RNA was sequenced by Fasteris SA (Geneva, Switzerland). The sRNAs were categorized into several size sets (15 – 35 nts), then aligned with HiSat2 (version 2.1.0) using default parameters[32]. Those reads mapped were sorted to OES, IES and MAC sequences, the mitochondrial, DNA from *Klebsiella pneumoniae* and the vector backbone.

## The GFP and mCherry fusion constructs, microinjection and localization

The SMC4-1 or SMC4-2-GFP fusion construct under the endogenous regulatory sequences respectively contained MAC sequences upstream of the ATG and downstream of the TGA. The optimized GFP coding sequence[34] was inserted ahead of the stop codon. Before performing the microinjection, all plasmids containing the fusion transgene were digested with the AhdI (R0584, New England BioLabs) or SapI (R0569L New England BioLabs) to linearize them. The products were filtered through 0.22 µm Ultrafree® MC GV filter (UFC30GV0S, Millipore), and enriched with pure ethanol. DNA was dissolved using DNase-free dd$H_2O$ to a final concentration of ~ 5.5 µg/uL. Finally, linearized DNA was microinjected into vegetative cell MACs[35]. SMC4-1-mCherry constructs as described above. Positive injections were picked up by checking green and red signals under a microscope. A positive single clone was expanded to high density (3000 cells/ml). At various life cycle phases, small samples were taken and counterstained with 4′,6-diamidino-2-phenylindole (DAPI). Fluorescence microscopy was used to detect GFP localization (Leica AF6000 system).

## Flag-HA fusion construct, microinjection, immunoprecipitation, mass spectrometry

Fusion construct and microinjection were performed as described above. The positive injection was confirmed by Dot Blot[33,36]. Immunoprecipitation was conducted as described before[37,38]. Non-crosslinking was performed because the IP of SMC4-2 under crosslinking did not work at a pH below 10.4. In detail, 400 mL cells were harvested at 4 hours after 100% fragmentation, pellets were resuspended in 2 mL fresh lysis buffer (50 mM Tris pH 8.8, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 1% Triton X-100, 1× protease inhibitor complete tablet (Roche), and 10% glycerol) and sonicated until complete lysis. Lysates were spin down at 13,000× g, 4 °C for 30 min. Beads were washed with 1 mL IP buffer (10 mM Tris pH 8.8, 150 mM NaCl, 0.01% NP-40, 1 mM $MgCl_2$, 1× protease inhibitor and 5% glycerol) for three times before incubation. A total of 1 mL of the supernatant was mixed with 50 μL of Anti-HA affinity resin (Roche) at 4 °C while rotating O/N. Another 1 mL supernatant was frozen in liquid nitrogen and then store at -80 °C before use. After incubation, beads were washed using 1 mL IP buffer five times. Washed beads were resuspended in 50 μL IP buffer, boiled with 25 μL 5× SDS loading buffer at 95 °C, after cooling down on the ice, and immediately used for western blot and mass spectrometry analysis at the University of Bern.

## Results

### Identification of the *P. tetraurelia* SMC4s homolog

In Fig. 1a, the domain organization of SMC4s in *P. tetraurelia* is shown. A Pfam domain search predicted that domains are conserved in these two SMC4s while the sequence blast indicates only 37.36% identity between them. It highlights the question that why *P. tetraurelia* needs two SMC4s. In Figure 1b, we showed the phylogenetic tree of SMC4s from 22 species with maximum likelihood analysis, SMC4-1 in *P. tetraurelia* locates in an independent branch of ciliate SMC4, while SMC4-2 has a closer localization with *Tetrahymena*, a close relative with *Paramecium*. To compare the difference between SMC4-1 and SMC4-2, authors made an expression curve according to the ParameciumDB as shown in Fig. 1c. It is indicated that SMC4-1 has a long-term expression in the whole development stage of *P. tetraurelia*, while SMC4-2 starts expression after MIC meiosis which makes it a potential candidate of IES excision related proteins. Fig. 1c also presents a cross of expression and alterations in the expression's prominent position between SMC4s which gives the idea that specific SMC4 is needed in the distinct development stages and SMC4s may posse different functions referred to the given stage.

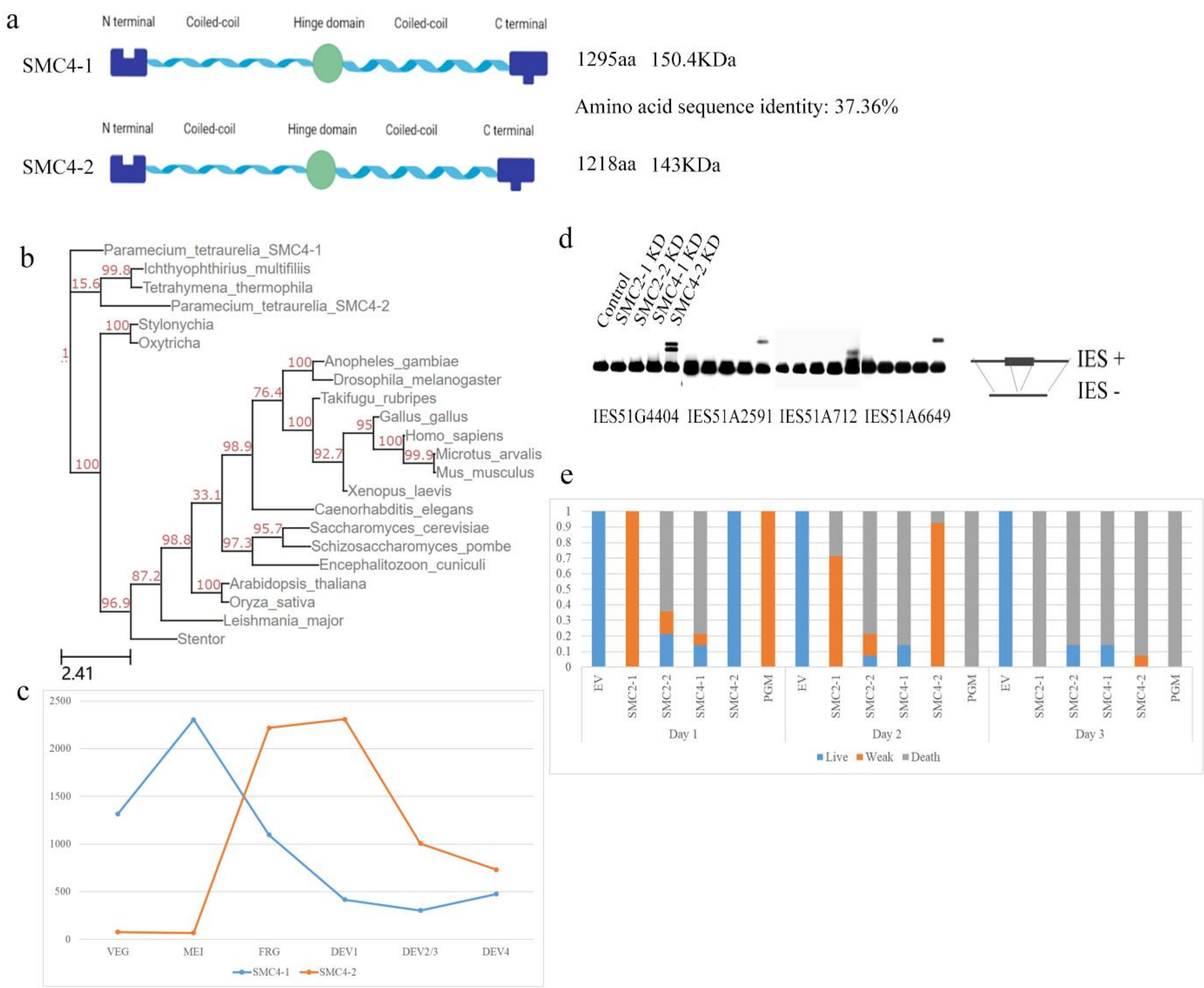


**Figure 1. Structure and function comparison between SMC4-1 and SMC4-2 in *P. tetraurelia*.** (a) Similarity and differences between SMC4-1 and SMC4-2 in *P. tetraurelia*. All information obtains from ParameciumDB, amino acid sequence identity gains from the BLAST tool on the website. Conserved domain/motifs in different colours. (b) Maximum Likelihood tree of SMC4s in ciliates and randomly selected organisms. (c) Time-course of gene expression for *P. tetraurelia* SMC4-1 and SMC4-2. The curve in blue represents the expression of SMC4-1, and the orange one indicates the expression of SMC4-2. The Y axis means the mean expression level. VEG: vegetative cells; MEI: beginning of MAC fragmentation and MIC meiosis; FRG: population when ~50% of cells have fragmented MACs; DEV1: earliest stage when a significant proportion of cells has visible MAC anlagen; DEV2/3: majority of cells with MAC anlagen; DEV4: majority of cells with MAC anlage. (d) Effect of EV, SMC2-1, SMC2-2, SMC4-1, SMC4-2 knockdowns on IES excision by IES PCRs. The excised form is shown as (IES-) and the unexcised form is shown as (IES+). The IES- form is always detectable due to the presence of the parental MAC in the sample. The IES+ is only present in IES retention in the newly developing MAC. (e) Effect of EV, SMC2-1, SMC2-2, SMC4-1, SMC4-2 and PGM KD on cell survival. The healthy, weak and death cells are shown in different colours. Blue: percentage of healthy cells (cell growing at a normal rate); orange: percentage of sick cells (altered

number of divisions or behaviour); grey: death cells. PGM KD was the positive control. The empty L4440 vector as the negative control. For each bar n = 14 cells.

## SMC4-2 is essential for IES excisions in *Paramecium*

To figure out what the functions of SMC4s are in IES excision, RNAi was used to make a knockdown (KD) of SMC4s expressions at the mRNA level. The IES PCR results obtained from the preliminary analysis of KD are shown in Fig. 1d where the retention of IES excision is shown by two bands, the upper one is the DNA region that contains IES when elimination fails, and the lower bands are regions that succeed in IES excisions. From here we can tell that SMC4-2 KD provokes all tested IESs retention while there was no evidence that SMC4-1 KD influences that. As another core part of condensin, SMC2-1 and SMC2-2 KD IES PCR had been shown here too. Same with SMC4-1 KD, no IES retention was found. A survival test is a common way to determine the influence on the next generation after KD and whether all cells are under RNAi in *P. tetraurelia*. As shown in Fig. 1e, over 80% death rate has been shown in all SMCs KD which means SMCs play important roles in *P. tetraurelia*.

To gain a genome-wide perspective on how IESs are affected by SMC4-2 KD, developing MAC was isolated from the autogamy cells after RNAi and starvation, high-throughput sequencing analysis was performed, the IES retention score (IRS) distribution result shown in Fig. 2a. The most striking result to emerge from the data is that 99% (43889 out of 43900) IESs excision are affected by SMC4-2 KD, the peak of retention score is around 0.4 which indicates that SMC4-2 may participate in IES excision in an unknown way. This result is beyond our expectations and only PGM KD shows a similar but stronger right-skewed distribution (Fig. 2b) compare to all other previously examined factors of excision machinery. The results of the correlational analysis are summarized in Fig. 2c which illustrates some of the main characteristics of retention pattern after these factors' KD. It is apparent from this table that very few correlations between SMC4-2 KD and any other KD. More efforts are needed before we can confirm that SMC4-2 engage in IES processing in a new way.

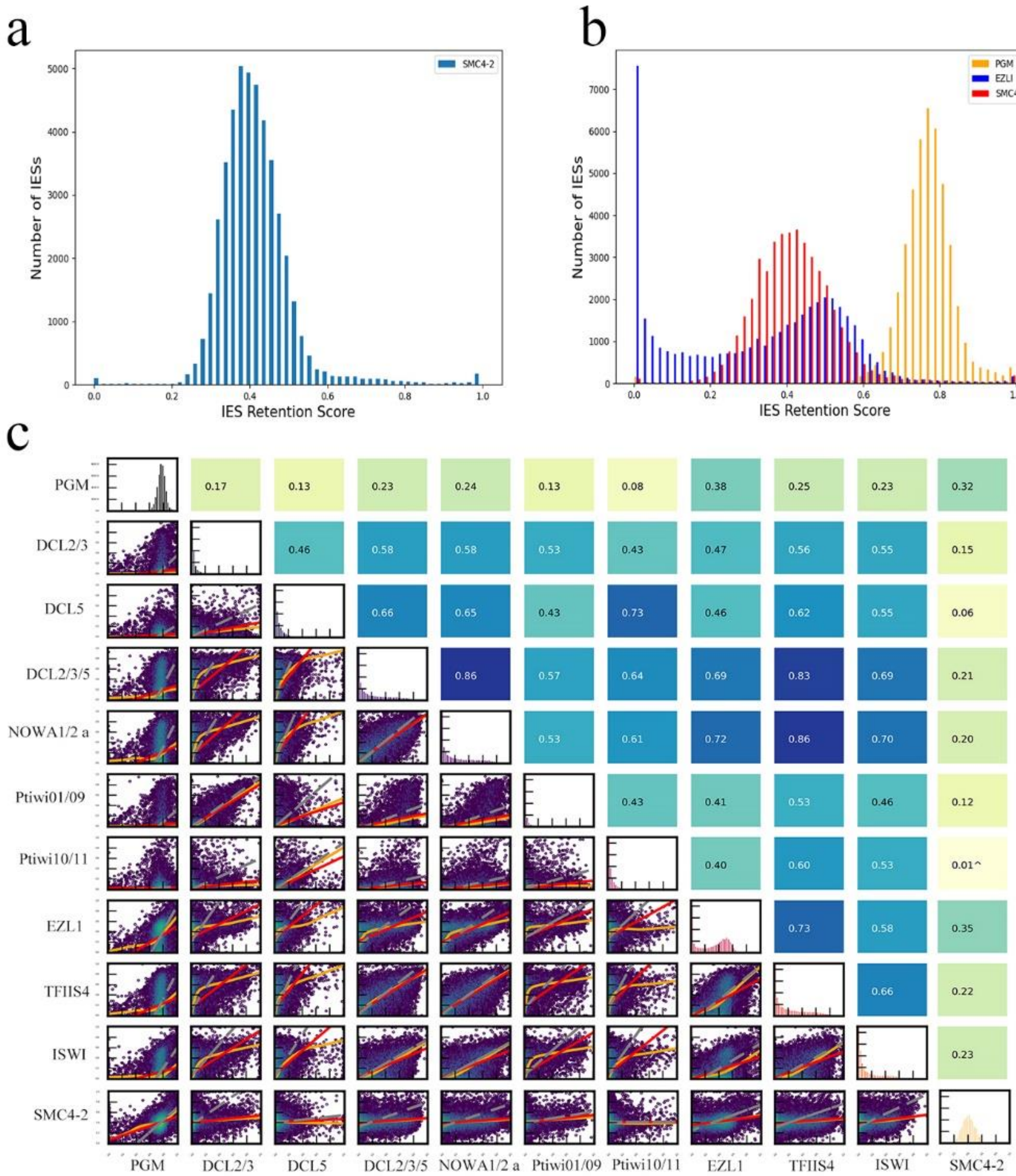


**Figure 2. IES retention pattern analysis.** (a) Retention score distribution was determined by sequencing DNA extracted from a cell fraction enriched in new MACs after SMC4-2 KD. Y axis means the number of IESs. X axis means the percentage of single unexcised IES among its copies. (b) Integrated IES retention distribution of PGM (in orange), EZL1 (in blue) and SMC4-2 (in red) KDs. (c) Correlation in IES retention. The retention score of PGM, DCL2/3, DCL5, DCL2/3/5, NOWA1/2a, Ptiwi01/09, Ptiwi10/11, EZL1, TFIIS4, ISWI are from our lab, SMC4-2 retention score is from this study. The correlation coefficients of Pearson are given at the right part of the graph correspondingly.

Two SMC4s localization shows diverse patterns in the cell cycle

The next section of the survey was concerned with why SMC4-2 are such important to IES excision. Authors first constructed an SMC4-2-GFP vector with 300bp up- and downstream of the SMC4-2 coding sequence then linearized the vector and microinjected it into vegetative MAC. The localization of SMC4-2 in Fig. 3a highlights that SMC4-2 uniquely locates in newly developing MACs although the green signal loses in the post autogamy stage. Here we speculate that SMC4-2 exhibits the possibility to process IESs excision spatially and temporally. While another problem rises from here, if SMC4-2 was also a component of condensin, its absence in vegetative, early and post-autogamy is unacceptable to chromosome structure. As we know from the expression pattern, there is still another SMC4—SMC4-1 in *P. tetraurelia* and it has a long-term expression throughout the whole life cycle. So, it’s reasonable that SMC4-1 serve as the “condensin SMC4” in *P. tetraurelia*. To figure out how the localization of SMC4-1 in the development of the *P. tetraurelia* looks, a linearized SMC4-1-GFP vector was injected into vegetative MAC in the same way as SMC4-2-GFP did. In Fig. 3b, the green signal represents SMC4-1 exclusively locates in nuclei including MICs through life, MACs except for the one that is going to be, as well as already fragmented MACs (old MACs). Considering the roles of condensin SMC4 in the dynamics of chromosomes and the scnRNA and iesRNA model, authors hypothesise that the dissociation of SMC4-1 in old MACs may be the premise of small RNAs from MICs matching to the long non-coding transcripts in old MAC genome. In Supplementary Figure S1, authors made a double injection by mixing equal amounts of SMC4-1-mCherry (red) and SMC4-2-GFP (green) into the vegetative MACs to check the co-localization of SMC4s in the development of *P. tetraurelia*. There was a significant positive correlation between them, the red and green signals can be visualized in newly developing MAC at a late time point. The single most interesting doubt to emerge from the localization was why two SMC4s are needed in the newly developing MACs if SMC4-1 and SMC4-2 served the same. If not, then what different functions they had on chromosome structure and IES excision.

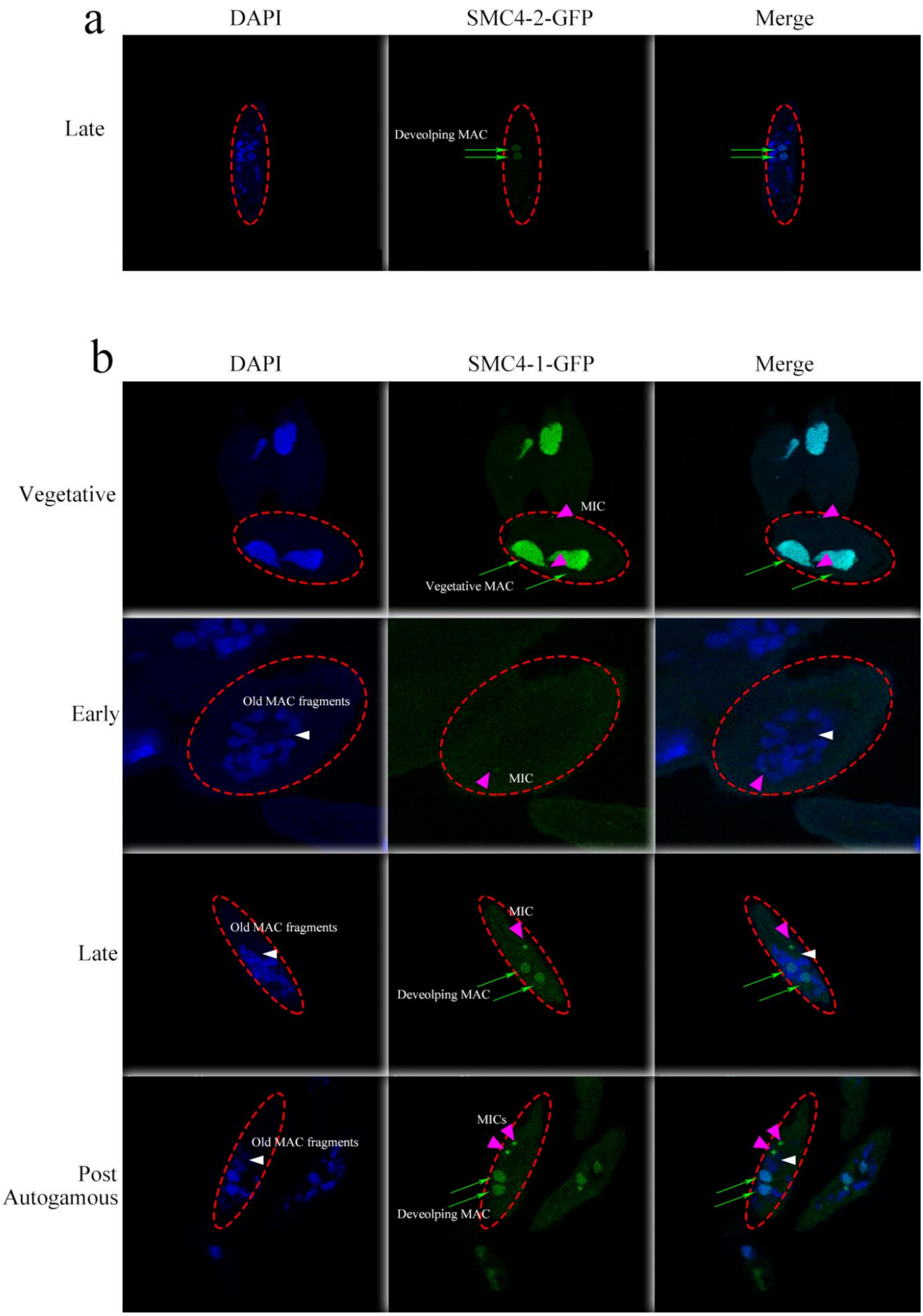


**Figure 3. Localization of SMC4-1 and SMC4-2 tagged with GFP.** (a) SMC4-2 tagged with GFP was exclusively localized in the macronuclear. (b) SMC4-1 tagged with GFP has a wide distribution

including vegetative macronuclear and micronuclear, new developing macronuclear and micronuclear, not in macronuclear which is going to be fragmented and old macronuclear fragments. The red dotted line represents the *P. tetraurelia* cell body. DAPI staining in blue represents DNA. Green signals represent SMC4-2 in a and SMC4-1 in b. Green arrows represent developing MAC. Pink arrowheads point to MICs, and arrowheads in white represent old MAC fragments.

## Knockdown of SMC4-2 disrupts genes expression at a late time point in *P. tetraurelia*

To know how SMC4-2 KD affect the IES excision and what kind of process SMC4-2 participate in *P. tetraurelia*, we performed mRNA sequencing of RNA samples from the late time of EV control and SMC4-2 KD. From Fig. 4a we can tell, SMC4-2 KD disrupts the expression of quite a long of genes. The downregulation of PTIWI06, PTIWI07, PTIWI10 and PTIWI11 highlighted on the plot could be because IESs exist in their promoters and coding regions. When we narrow to the top 50 most differentially expressed genes (DEGs, where the variance is the highest between samples) in Fig. 4b, it is significant that SMC4-2 KD can upregulate gene expression in a late time point including PGML5 and other PGMLs (not in the top 50, but still significant), the co-factor of PGM in IES excision[39]. In Fig. 4c, we provide a GO enrichment map. Enriched RNA-related GO terms indicate SMC4-2 could participate in RNA processing pathway. For the first time, SMC4-2 was reported to regulate the gene expression in *P. tetraurelia*. It is supposed that SMC4-2 plays downregulating roles in the development of *P. tetraurelia*, when disrupts the expression of SMC4-2, the depression from it would be released, and then a large number of genes could be expressed.

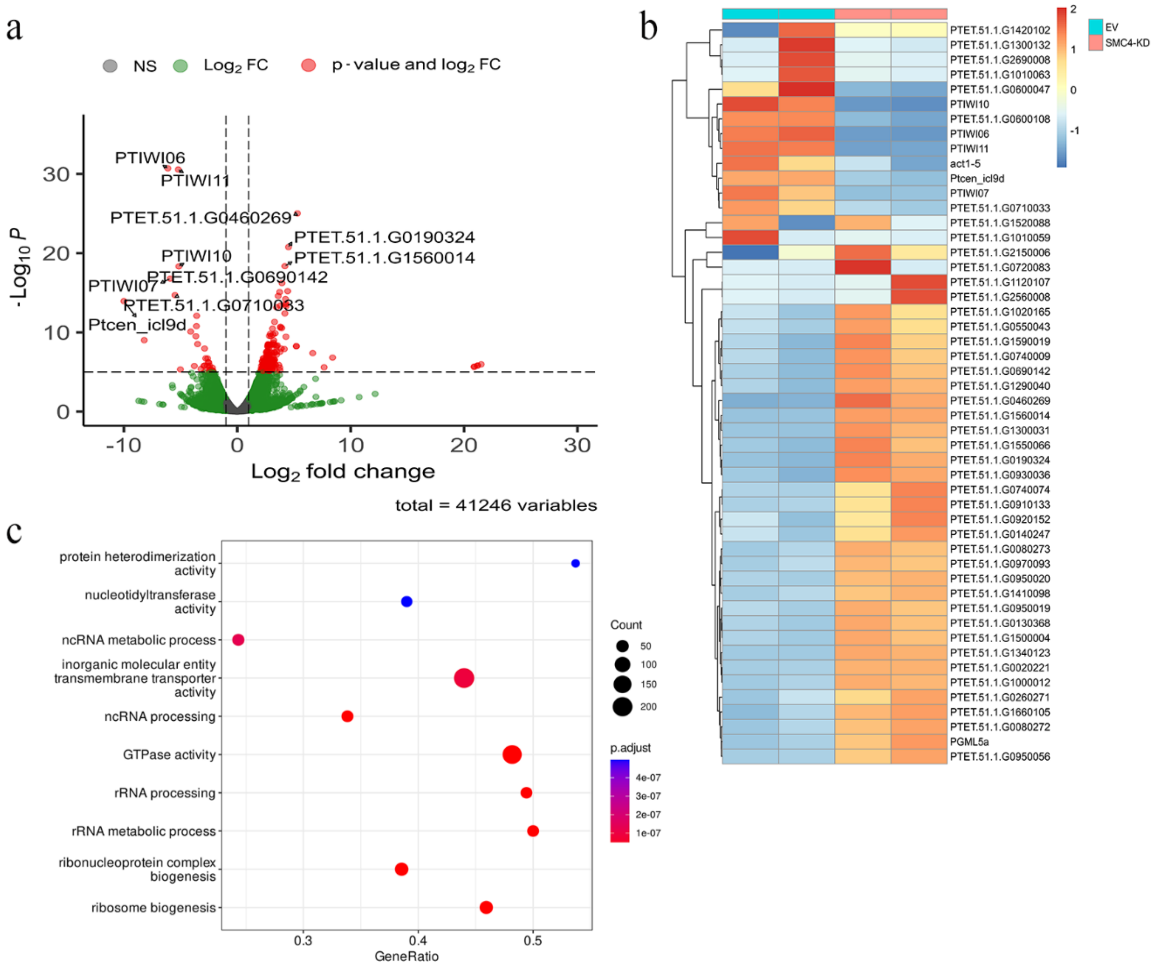


**Figure 4. Differential expression and GO enrichment in SMC4-2 KD compared to EV.** (a) Volcano plots showing the differentially expressed genes detected in SMC4-2 KD compare to the control. The most statistically significant genes are shown toward the top, with upregulated genes on right and downregulated proteins on left. Grey dots mean no significant difference, and green dots represent significance only in fold change. The X-axis represents $\log_2$ (fold change) values and Y-axis represents $-\log_{10}$ (pval) values. (b) The top 50 most differentially expressed genes in SMC4-2 KD compare to the control. Light blue on the top represents two replicates of EV, the orange means two replicates of SMC4-2 KD. The red (upregulated) and blue (downregulated) colour and intensity of the boxes represent changes in gene expression. (c) GO terms enrichment map. The x-axis shows the Gene ratio, the percentage of total DEGs in the given GO term; the y-axis corresponds to GO terms. Count represents the number of genes enriched in a GO term. The colour of the dot represents the p adjust value.

SMC4-2 knockdown wipes out the production of iesRNAs at a late time point

To know how SMC4-2 engages in most IESs excision, a small RNA analysis was performed to detect whether scnRNA or iesRNA are affected by the RNAi of SMC4-2. Since we already know that SMC4-2 exclusively locates in newly developing MAC and scnRNA is produced during the early meiosis in *P. tetraurelia* [2,3], it's unlikely the RNAi of SMC4-2 can disrupt the production of scnRNA. So, authors only focused on the genesis of iesRNAs which are 26-30 nts long and generated in the newly developing MAC[3,4,37]. To assess the production of iesRNA in a high resolution, small RNA sequencing was performed and the detailed information of detected small RNAs are clustered into several groups. All these RNAs are shown in the bar chart (Fig. 5a, b) with a size strategy. Compared with EV, it is apparent from here that extremely few iesRNAs can be recognized from 26-30 nts. In general, SMC4-2 KD disrupts IES excision may be partly because of abolishing the genesis of iesRNAs.

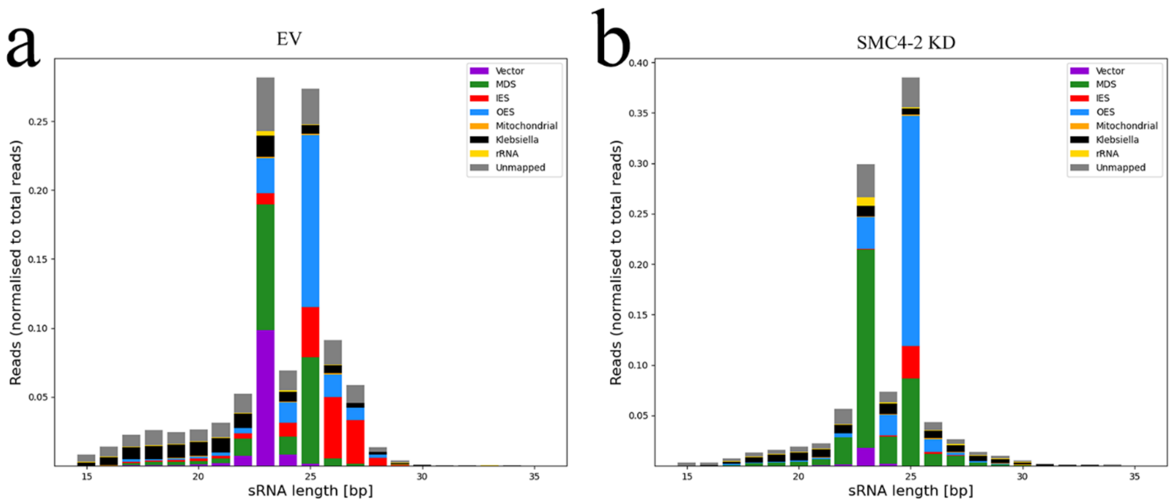


**Figure 5. Small RNA sequencing in EV and SMC4-2 KD.** (a, b) Histograms of small RNAs classified by length. Both panels show the distribution at a late time point (specifically 4 hours after 100% fragments). The reads map to MDS is shown in green or IESs in red. A significant absence of 26-30 bp, in which the iesRNAs are, is observed in SMC4-2 KD.

Knockdown of SMC2-1/SMC2-1/SMC4-1 getting rid of the IES retention after SMC4-2 KD

In theory, SMC4 functions in cells in proteins complex way like condensin. So, if SMC4-2 manipulated IES excision in condensin form, the disruption of the remaining part of condensin should be robust or at least retain the IES retention phenotype under SMC4-2 KD. Therefore, the authors conducted double silencing of SMC4-2 with SMC2-1, SMC2-2 or SMC4-1. Surprisingly, as in Fig. 6a, the IES retention phenotype has disappeared in all these double silencings. The survival test in Fig. 6b represents that at least 80% of cells have been affected after feeding with a double silencing medium. To get a convincing result, in addition to mixing two cultures of a single RNAi medium, the authors also constructed 3 silencing vectors which contain corresponding regions to double silencing, by this way, differential production of double strand RNA (dsRNA) to each gene should be excluded. Still, the IES retention phenotype induced by SMC4-2 silencing is no longer detectable.

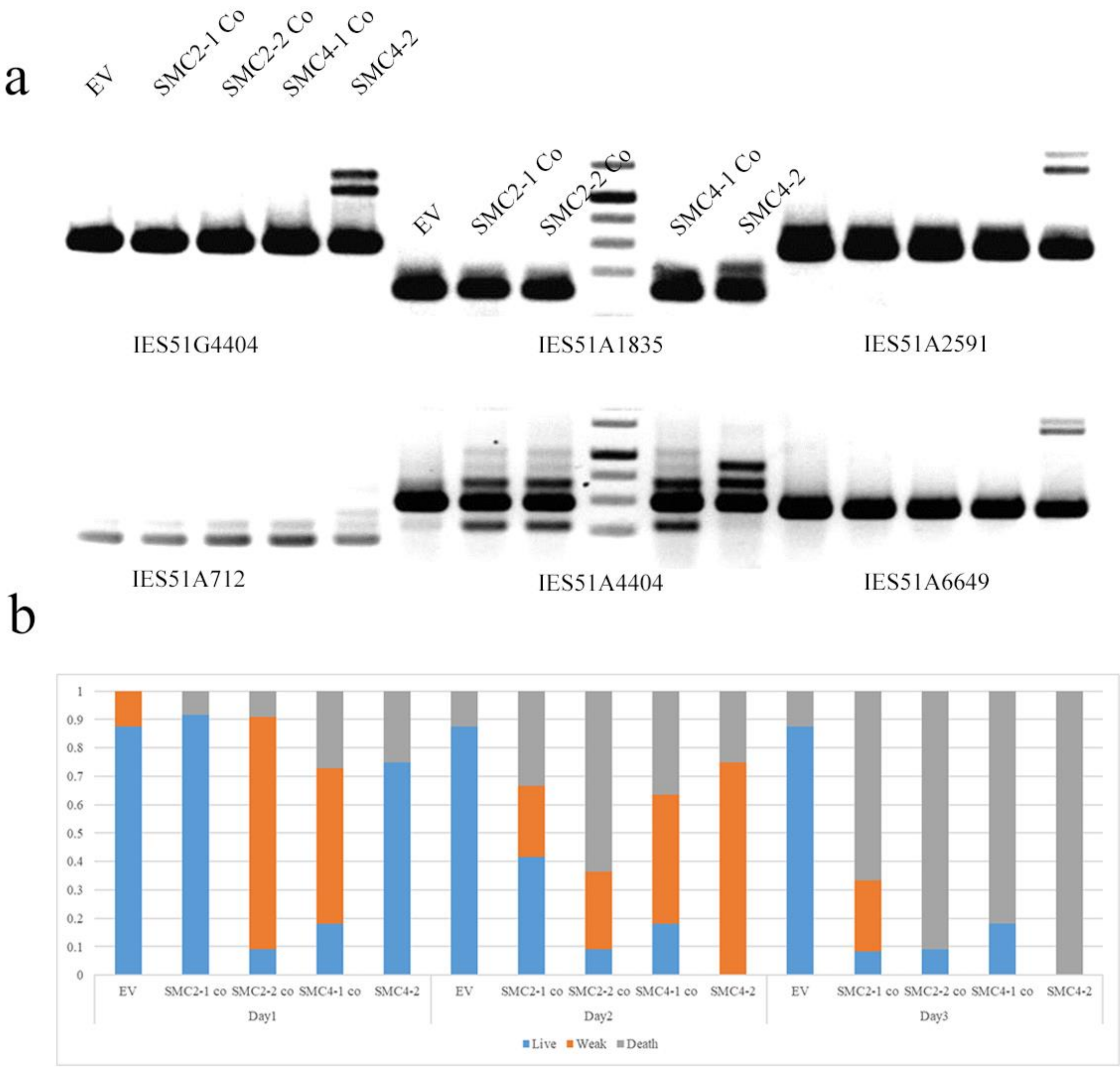


**Figure 6. Co-silencing of SMC4-2 with SMC2-1, SMC2-2 or SMC4-1.** (a) IES PCR of co-silencing. IES retention was tested by PCR using primers flanking IESs sequences. (b) Survival test of co-silencing.

Co-immunoprecipitation identify distinct interacting proteins between SMC4-1 and SMC4-2

To compare the difference between SMC4-1 and SMC4-2, authors applied co-immunoprecipitation coupling mass spectrometry to analyse the interacting proteins of Flag-HA tagged SMC4s at the same time point. As shown in Fig. 7a, SMC4-2-Flag-HA and SMC4-1-Flag-HA have been detected by anti-HA primary antibodies in cell lysate and bound solution. After analysed by Shotgun Liquid Chromatography Tandem Mass spectrometry (LC-MS/MS) with non-labelled samples, PGM or other IES critical proteins are undetectable in any of these interacting proteins is beyond our expectation. While the more surprising observation is the enrichment difference of SMC2s in these two SMC4s. the Venn diagrams (Fig. 7b and Fig. 7c) illustrate the relationships of differential enrichment and unique proteins among SMC4-1, SMC4-2 and WT interactions. In Fig. 7b, the grey arrow points to the SMC2-1 means it highly accumulated in both SMC4-1 and SMC4-2 enrichment compared to WT, but there is no different enrichment between SMC4-1 and SMC4-2 which indicates SMC4-1 and SMC4-2 may bind the same amount of SMC2-1 in this time point. The red arrow point to SMC4-2 and SMC2-2 means these two proteins not only upregulated in SMC4-1 and SMC4-2 compared to WT but also enriched in SMC4-2 compared to SMC4-1. The blue arrow pointing to SMC4-1 in Fig. 7c indicates that SMC4-1 is a unique protein which can only be detected in the SMC4-1 interacting dataset. Taken together, these results provide completely new insights into the diverse elements of SMC4s interactions in the development of *P. tetraurelia*.

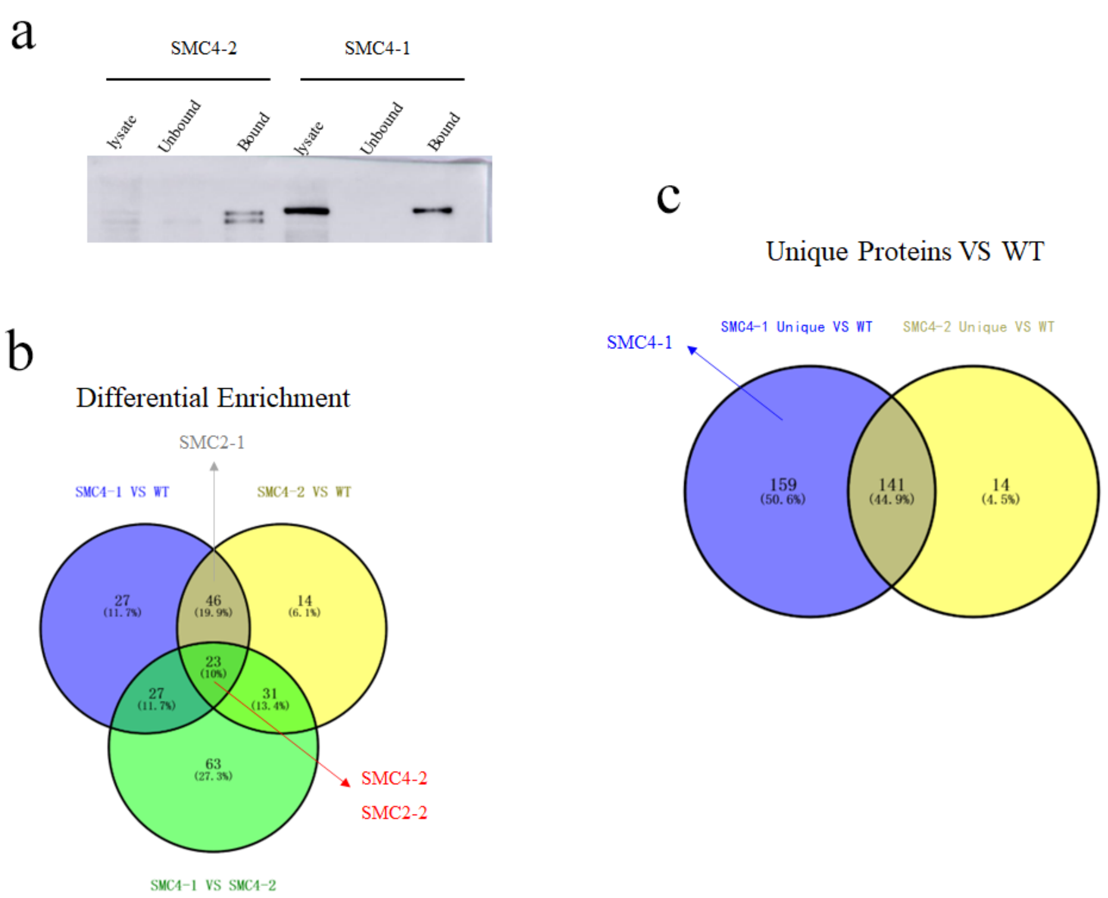


**Figure 7. Co-immunoprecipitation of Flag-HA tagged SMC4-1 and SMC4-2 at a late time point.** (a) Western blot of Flag-HA tagged SMC4-2 and SMC4-1 pulled down with anti-HA antibodies. Lysate: cell lysate; Unbound: supernatant by spinning down anti-HA beads-lysate mixture after overnight incubation; Bound: enriched target proteins on beads after five times washing. The molecular weight of SMC4-1 ~150KD, molecular weight of SMC4-2 ~143KD. (b, c) Venn diagram of differential enrichment and unique proteins vs wild type. (b) Venn diagrams showing the overlap in differential enrichment of SMC4-1 vs WT, SMC4-2 vs WT and SMC4-1 vs SMC4-2. The grey arrow represents non-regulated in SMC4-1 vs SMC4-2. The red arrow means upregulated in SMC4-2 compared to SMC4-1. (c) Venn diagram shows the overlap in unique proteins of SMC4-1 vs WT and SMC4-2 vs WT. The blue arrow indicates unique protein only can be seen in SMC4-1 enrichment.

## Discussion

Structural maintenance of chromosomes (SMC) complexes are the primary chromosome organizers in all living organisms[5,40]. These protein complexes play central roles in DNA replication[41], chromosome condensation and segregation[42], gene expression[6], and DNA damage repair[43]. In reviewing the literature, no data was found on the association between SMC4 and IES excision in *P. tetraurelia* even expand to ciliates. This is the first report that discovered and explored the relationship between SMC4-2 and IES processing. The important finding here is SMC4-2 KD shows a very strong IES retention phenotype, while other members of the *P. tetraurelia* SMC family including SMC2s and SMC4-1 don't. From here, authors suspect that SMC4-2 must have some unique function in addition to serving as the potential condensin component. Because if not, the knockdown of SMC2s and SMC4-1 should have similar results on IES excision. But the mechanism behind SMC4-2 regulating IES excision is still unknown, and from the structural prediction of SMC4-2, nothing significant characteristics can be presumed as the possible reason for this regulation. So, more indirect regulations have been taken into consideration.

The second interesting result from this work is the dynamic changes of SMC4s localization in the life cycle. As mentioned in the results part, SMC4-2 is localized specifically in newly developing MACs while SMC4-1 shows a more complex and remarkable positing strategy in both MICs and MACs. The observed green signal in newly developing MAC of SMC4-2-GFP injected cells could be attributed to the possibility that SMC4-2 processes IES excision or is involved in that at least. This study did not find a significant SMC4-2 localization in any other nucleus. This finding suggests that SMC4-2 is not the fully functional condensin-SMC4 in *P. tetraurelia*. Hence, SMC4-1-GFP injection was performed in vegetative cells and the GFP signal was observed from the vegetative stage to the post-autogamy cells. Interestingly, SMC4-1 has localization in all nuclei including MICs and vegetative and newly developing MACs. These results are consistent with data obtained in *Tetrahymena* which suggests that Smc4p may be involved in amitotic, mitotic and nuclear division[21]. One unanticipated finding was that the SMC4-1 signal was not seen in early MACs and old MAC fragments which suggests a dynamic dissociation of SMC4-1 from the genome as cells into autogamy. It's still unknown what would happen after SMC4 released from the chromosome and what the chromosome looks like in *P. tetraurelia*. The most significant progress here is scnRNAs produced from MICs and matured in old MAC before guiding IES excision[44]. Hence, it conceivably we hypothesized that the dissociation of SMC4-1 in old MAC exposes a highly accessible DNA structure to DNA binding proteins[45], this would be a good opportunity for scnRNAs to scan their target DNA regions. The co-localization results confirm the temporal and spatial association between SMC4-1 and SMC4-2 at a late time point. Either SMC4-1 and SMC4-2 have some overlap in function or it’s just a coincidence needs further study.

The mRNA sequencing results indicate that only several PTIWI proteins were affected by SMC4-2 KD while other important proteins like PGM and Dcls (except upregulated Dcl5) were not included in the DEGs. The downregulation of PTIWI6/7/10/11 could be because of IESs’ insertion in their promoter and coding regions. The IESs retention introduced by SMC4-2 KD disrupts the transcription and translation of these genes. This result reminds us of the possibility that IESs retention under SMC4-2 KD is not because of the downregulation of DNA excision machinery. Another interesting result under SMC4-2 KD is the upregulation of such amount of genes which reminds us that SMC4-2 could be a repressor of gene expression in the development of *P. tetraurelia*. Further study needs to be done to explore the real role of SMC4-2 in *P. tetraurelia*.

Another interesting result from the sRNA sequencing is that the population of the iesRNA was markedly depressed. For the first time, SMC4-2 was supposed to potentially manipulates the yield of specific sRNAs in ciliate. Here we assume that the possibility of SMC4-2 KD clearing away iesRNAs could be because SMC4-2 inhibits the early IES excision which is the template of iesRNA transcription. Further study is needed to explore the detailed mechanism behind SMC4-2 affect iesRNA production in *P. tetraurelia*.

The co-silencing results here indicate an outcome regard of restoring IES elimination under SMC4-2 KD by introducing a knockdown of SMC2-1, SMC2-2 or SMC4-1. The double silencing should robust the IES retention phenotype or retain it if SMC4-2 formed condensin with SMC2s in *P. tetraurelia*. If SMC4-2 had nothing to do with SMC4-1 and SMC2s, these double silencing should exhibit combing phenotypes of single silencing such as IES retention. In this study, no IES retention can be observed under double silencing. This is the first report that mention the IES retention phenotype can be restored when introducing another knockdown. No related report can be learned from to explain this phenotype. The LC-MS/MS results give us detailed information about how SMC4s interact with other proteins at the late time point. It is somewhat surprising that no PGM or other IES-related proteins were noted in the SMC4-2 interacting dataset. Not surprisingly, no such proteins were found in SMC4-1 interactions. Although there is no direct evidence to prove that SMC4-2 processes IES excision, the truth that SMC4-2 KD affect IES excision and iesRNA production is still clear. In addition to these results, more interactions were found in differential expression and unique datasets in these two analyses. As mentioned before, when performing Co-IP by SMC4-1-Flag-HA as bait, all SMC4s and SMC2s can be captured at this time point, while only SMC2s were exhibited when using SMC4-2-Flag-HA as bait, SMC4-1 was not presented. The relationship between SMC4-1 and SMC4-2 has to be reconsidered. If they interacted with each other, the absence of SMC4-1 in the SMC4-2 pull-down is unacceptable since SMC4-2 can be enriched by SMC4-1. In these two experiments, the only thing that changed was the concentration of them because of the injection which introduced exceeding protein compared to their endogenous amount. Under this circumstance, the ratio between SMC4-1 and SMC4-2 at a late time point has been either flipped or enhanced which could be the reason why the SMC4-1 and SMC4-2 interaction was lost. This observation may support the hypothesis that SMC4-2 may potentially replace the position of SMC4-1 in condensin that interacts with SMC2s in a competitive binding way. This assumption is based on the fact that SMC4-1 and SMC4-2 share conserved domains and the potential to form condensin complexes with SMC2. Similar to the endogenous ratio, when introducing a large amount of SMC4-2 into a late time point, it may compete with endogenetic SMC4-1 on binding with SMC2s, the SMC4-1 dissociated from chromosomes but not degraded according to the continuous fluorescence. The ‘SMC4-2 plus SMC2s’ condensin complex could be the late condensin form that may participate in IESs excision as SMC4-2 was supposed to be the important co-factor of this process but SMC2s are not in this study. Contrary to the native ratio, as injecting SMC4-1 into MACs, the ratio at the late time point between SMC4-1 and SMC4-2 was flipped, SMC4-2 was still able to occupy SMC4-1’s position while a high amount of SMC4-1 could inhibit this process that may stable the interaction with SMC4-2. For sure, this assumption needs to be interpreted with caution. Here we just bring up one possibility which can fit all results got from the study. Readers should judge the model carefully.

## Conclusion

This research aimed to study the effect of SMCs in processing IES excision in *P. tetraurelia*. The most interesting finding here is that SMC4-2 play important role in manipulating IESs and iesRNAs production. To our knowledge, this is the first report of the interaction between SMC4-2 and IES excision in *P. tetraurelia*. The second major finding was that SMC4-1 degraded in old MAC potential to be the initial step of IES processing which is also the first report of SMC4-1 dissociation at an early time point. The research has also shown that it's very likely that SMC2s and SMC4-1 take participate in SMC4-2 handling IES excision which has not been explored before. The last but not the least finding is that we described the interactions between SMC2s and SMC4s and the absence of SMC4-1 was reported in the SMC4-2 Co-IP results. Combining all results in this study, the authors proposed a competitive model which makes an explanation for each part of these findings. The current data highlight the importance of SMC4-2 in the genome rearrangement in *P. tetraurelia* beyond the classic definition of SMC4 in the maintenance of chromosomes in other organisms[46,47]. This project is the first comprehensive investigation of condensin SMCs' independent or dependent influences on IES excision which should contribute to existing knowledge of IES processing by providing newly potential co-factors from a completely new direction. The generalizability of these results is subject to certain limitations. Despite this, the study certainly adds a novel angle at which to look at this process. The precise mechanism of SMC4-2 in this progress remains to be elucidated. Further work is still needed to establish whether the competitive model is true or not. In general, this study provides a new foundation for additional investigation which may in the end lead to an exciting mechanistic discovery.

Reference


1 Bhullar, S. *et al.* A mating-type mutagenesis screen identifies a zinc-finger protein required for specific DNA excision events in Paramecium. *Nucleic acids research* **46**, 9550-9562, doi:10.1093/nar/gky772 (2018).

2 Lepere, G. *et al.* Silencing-associated and meiosis-specific small RNA pathways in Paramecium tetraurelia. *Nucleic acids research* **37**, 903-915, doi:10.1093/nar/gkn1018 (2009).

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

4 Allen, S. E. *et al.* Circular Concatemers of Ultra-Short DNA Segments Produce Regulatory RNAs. *Cell* **168**, 990-999 e997, doi:10.1016/j.cell.2017.02.020 (2017).

5 Hirano, T. Condensin-Based Chromosome Organization from Bacteria to Vertebrates. *Cell* **164**, 847-857, doi:10.1016/j.cell.2016.01.033 (2016).

6 Wood, A. J., Severson, A. F. & Meyer, B. J. Condensin and cohesin complexity: the expanding repertoire of functions. *Nature reviews. Genetics* **11**, 391-404, doi:10.1038/nrg2794 (2010).

7 Anderson, D. E., Losada, A., Erickson, H. P. & Hirano, T. Condensin and cohesin display different arm conformations with characteristic hinge angles. *The Journal of cell biology* **156**, 419-424, doi:10.1083/jcb.200111002 (2002).

8 Ono, T., Fang, Y., Spector, D. L. & Hirano, T. Spatial and temporal regulation of Condensins I and II in mitotic chromosome assembly in human cells. *Molecular biology of the cell* **15**, 3296-3308, doi:10.1091/mbc.e04-03-0242 (2004).

9 Shintomi, K. & Hirano, T. The relative ratio of condensin I to II determines chromosome shapes. *Genes & development* **25**, 1464-1469, doi:10.1101/gad.2060311 (2011).

10 Chan, R. C., Severson, A. F. & Meyer, B. J. Condensin restructures chromosomes in preparation for meiotic divisions. *The Journal of cell biology* **167**, 613-625, doi:10.1083/jcb.200408061 (2004).

11 Nishide, K. & Hirano, T. Overlapping and non-overlapping functions of condensins I and II in neural stem cell divisions. *PLoS genetics* **10**, e1004847, doi:10.1371/journal.pgen.1004847 (2014).

12 Green, L. C. *et al.* Contrasting roles of condensin I and condensin II in mitotic chromosome formation. *Journal of cell science* **125**, 1591-1604, doi:10.1242/jcs.097790 (2012).

13 Gibcus, J. H. *et al.* A pathway for mitotic chromosome formation. *Science* **359**, doi:10.1126/science.aao6135 (2018).

14 Sakai, A., Hizume, K., Sutani, T., Takeyasu, K. & Yanagida, M. Condensin but not cohesin SMC heterodimer induces DNA reannealing through protein-protein assembly. *The EMBO journal* **22**, 2764-2775, doi:10.1093/emboj/cdg247 (2003).

15 Sutani, T. & Yanagida, M. DNA renaturation activity of the SMC complex implicated in chromosome condensation. *Nature* **388**, 798-801, doi:10.1038/42062 (1997).

16 Akai, Y. *et al.* Opposing role of condensin hinge against replication protein A in mitosis and interphase through promoting DNA annealing. *Open biology* **1**, 110023, doi:10.1098/rsob.110023 (2011).

17 Festuccia, N., Gonzalez, I., Owens, N. & Navarro, P. Mitotic bookmarking in development and stem cells. *Development* **144**, 3633-3645, doi:10.1242/dev.146522 (2017).
18 Beisson, J. *et al.* Mass culture of Paramecium tetraurelia. *Cold Spring Harbor protocols* **2010**, pdb prot5362, doi:10.1101/pdb.prot5362 (2010).
19 Beisson, J. *et al.* Maintaining clonal Paramecium tetraurelia cell lines of controlled age through daily reisolation. *Cold Spring Harbor protocols* **2010**, pdb prot5361, doi:10.1101/pdb.prot5361 (2010).
20 Cobbe, N. & Heck, M. M. The evolution of SMC proteins: phylogenetic analysis and structural implications. *Molecular biology and evolution* **21**, 332-347, doi:10.1093/molbev/msh023 (2004).
21 Cervantes, M. D., Coyne, R. S., Xi, X. & Yao, M. C. The condensin complex is essential for amitotic segregation of bulk chromosomes, but not nucleoli, in the ciliate Tetrahymena thermophila. *Molecular and cellular biology* **26**, 4690-4700, doi:10.1128/MCB.02315-05 (2006).
22 Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics* **20**, 1160-1166, doi:10.1093/bib/bbx108 (2019).
23 J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L.L. Sonnhammer, S.C.E. Tosatto, L. Paladin, S. Raj, L.J. Richardson, R.D. Finn, A. Bateman. *Nucleic Acids Research* doi: 10.1093/nar/gkaa913 (2020).
24 Arnaiz, O. *et al.* Improved methods and resources for paramecium genomics: transcription units, gene annotation and gene expression. *BMC genomics* **18**, 483, doi:10.1186/s12864-017-3887-z (2017).
25 Trifinopoulos, J., Nguyen, L. T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic acids research* **44**, W232-235, doi:10.1093/nar/gkw256 (2016).
26 Beisson, J. *et al.* Paramecium tetraurelia: the renaissance of an early unicellular model. *Cold Spring Harbor protocols* **2010**, pdb emo140, doi:10.1101/pdb.emo140 (2010).
27 Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811, doi:10.1038/35888 (1998).
28 Baudry, C. *et al.* PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. *Genes & development* **23**, 2478-2483, doi:10.1101/gad.547309 (2009).
29 Arnaiz, O. & Sperling, L. ParameciumDB in 2011: new tools and new data for functional and comparative genomics of the model ciliate Paramecium tetraurelia. *Nucleic acids research* **39**, D632-636, doi:10.1093/nar/gkq918 (2011).
30 Arnaiz, O. *et al.* The Paramecium germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS genetics* **8**, e1002984, doi:10.1371/journal.pgen.1002984 (2012).
31 Denby Wilkes, C., Arnaiz, O. & Sperling, L. ParTIES: a toolbox for Paramecium interspersed DNA elimination studies. *Bioinformatics* **32**, 599-601, doi:10.1093/bioinformatics/btv691 (2016).
32 Bechara, S., Kabbani, L., Maurer Alcala, X. & Nowacki, M. Identification of novel, functional long non-coding RNAs involved in programmed, large scale genome rearrangements. *Rna*, doi:10.1261/rna.079134.122 (2022).
33 Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology*, 37(8), 907–915. doi.org:10.1038/s41587-019-0201-4 (2019).

34 Nowacki, M., Zagorski-Ostoja, W. & Meyer, E. Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in Paramecium tetraurelia. *Current biology : CB* **15**, 1616-1628, doi:10.1016/j.cub.2005.07.033 (2005).
35 Beisson, J. *et al.* DNA microinjection into the macronucleus of paramecium. *Cold Spring Harbor protocols* **2010**, pdb prot5364, doi:10.1101/pdb.prot5364 (2010).
36 Furrer, D. I., Swart, E. C., Kraft, M. F., Sandoval, P. Y. & Nowacki, M. Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA. *Cell reports* **20**, 505-520, doi:10.1016/j.celrep.2017.06.050 (2017).
37 Hoehener, C., Hug, I. & Nowacki, M. Dicer-like Enzymes with Sequence Cleavage Preferences. *Cell* **173**, 234-247 e237, doi:10.1016/j.cell.2018.02.029 (2018).
38 Reuter, M. *et al.* Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nature structural & molecular biology* **16**, 639-646, doi:10.1038/nsmb.1615 (2009).
39 Bischerour, J. *et al.* Six domesticated PiggyBac transposases together carry out programmed DNA elimination in Paramecium. *eLife* **7**, doi:10.7554/eLife.37927 (2018).
40 Uhlmann, F. SMC complexes: from DNA to chromosomes. *Nature reviews. Molecular cell biology* **17**, 399-412, doi:10.1038/nrm.2016.30 (2016).
41 Terakawa, T. *et al.* The condensin complex is a mechanochemical motor that translocates along DNA. *Science* **358**, 672-676, doi:10.1126/science.aan6516 (2017).
42 Hirano, T., Kobayashi, R. & Hirano, M. Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a Xenopus homolog of the Drosophila Barren protein. *Cell* **89**, 511-521, doi:10.1016/s0092-8674(00)80233-0 (1997).
43 De Piccoli, G., Torres-Rosell, J. & Aragon, L. The unnamed complex: what do we know about Smc5-Smc6? *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **17**, 251-263, doi:10.1007/s10577-008-9016-8 (2009).
44 Michelini, F. *et al.* From "Cellular" RNA to "Smart" RNA: Multiple Roles of RNA in Genome Stability and Beyond. *Chemical reviews* **118**, 4365-4403, doi:10.1021/acs.chemrev.7b00487 (2018).
45 Thattikota, Y. *et al.* Cdc48/VCP Promotes Chromosome Morphogenesis by Releasing Condensin from Self-Entrapment in Chromatin. *Molecular cell* **69**, 664-676 e665, doi:10.1016/j.molcel.2018.01.030 (2018).
46 Hirano, T. At the heart of the chromosome: SMC proteins in action. *Nature reviews. Molecular cell biology* **7**, 311-322, doi:10.1038/nrm1909 (2006).
47 Burmann, F. *et al.* An asymmetric SMC-kleisin bridge in prokaryotic condensin. *Nature structural & molecular biology* **20**, 371-379, doi:10.1038/nsmb.2488 (2013).

Supplementary Material

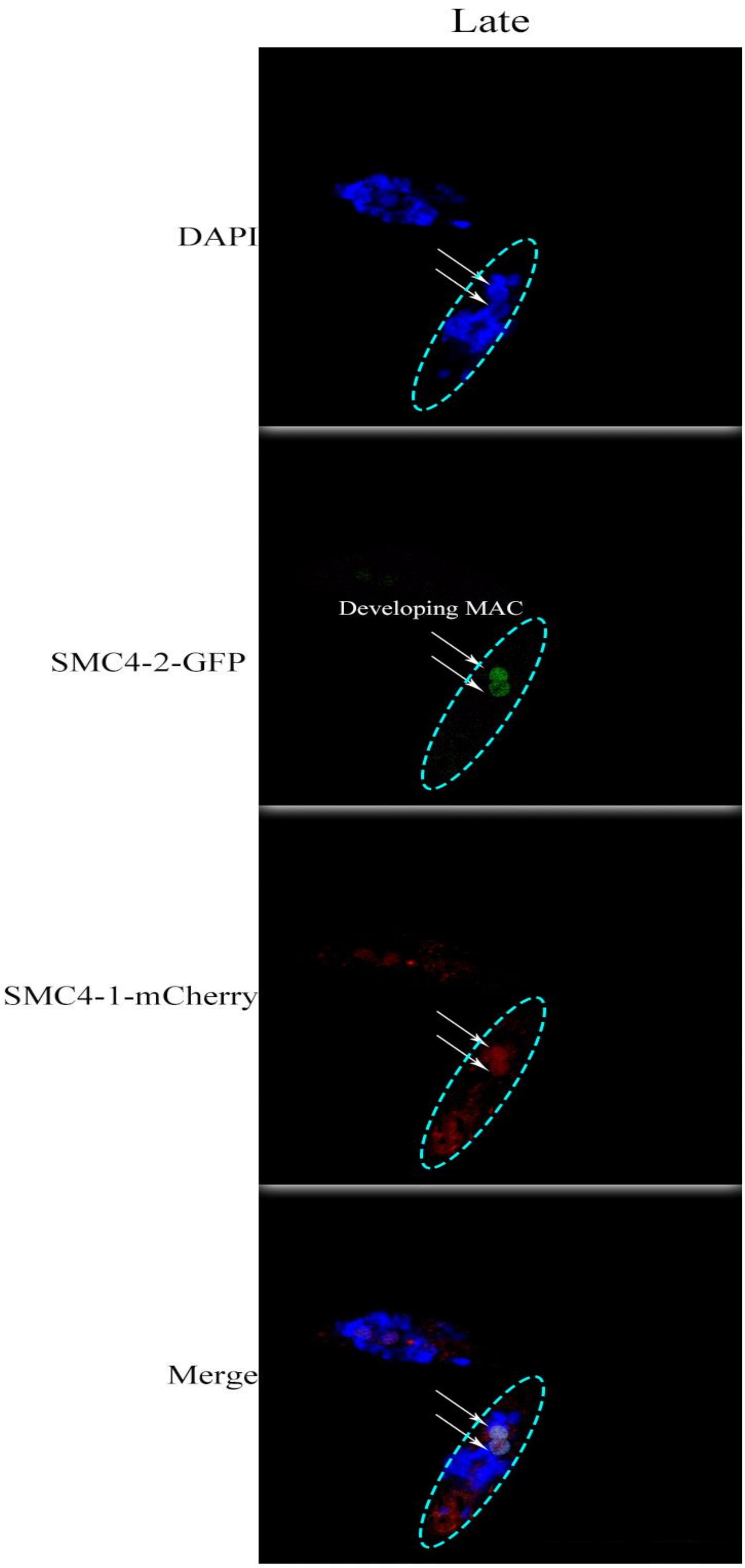

**Figure S1. Co-localization of SMC4-1 tagged with mCherry and SMC4-2 tagged with GFP at late time point.** DAPI represents DNA in blue, green signal represents SMC4-2 tagged with GFP, red signal represents SMC4-1 tagged with mCherry.

| Oligo name | Sequence 5'-3' |
|---|---|
| SMC4-1 +f R | TCAACATGAATAATTAATTTTAAATTCTTCATTCTT |
| SMC4-1 +f F | AAAGAATTAACTAGTATTTTGCAGAAAAAATATTAAG |
| SMC4-1 +fv 3 F | TGAAATGATTTTAAAATATTCATATTCCATTATTG |
| SMC4-1 +fv 3 R | TCTATTTTAATTAACGTTTTCATTGTTTTC |
| SMC4-1 +fv F | CAGACTGAAAGTGAAGAGCCAT |
| SMC4-1 +fv R | CATTTTATGTAAATATATTTAATATTTTTAAATTAACTTTTGG |
| SMC4-2 +F R | GAATTAATTCAAGATCTAAATATTCGTAACA |
| SMC4-2 +F F | ATGCGCAATTTAATTTGAGCAATATT |
| SMC4-2 +fv 3 F | TGAAATTTATATTAAATTATAATCAATTAATTTTATC |
| SMC4-2 +fv 3 R | AACTTCTAATTAGATGACTTCTGTTGAA |
| SMC4-2 +fv F | ATTAAGGAAGTTATTTTGGAGAATTTTAAATC |
| SMC4-2 +fv R | CATATTTTCTATCATTTAATTTTCTACTCAAA |
| SMC2-1 sil F | ATATGAGCTCATGTGGATCAAAGAAATCATTATCGA |
| SMC2-1 sil R | TACCAGGAGACAAAAAGAAACTATAGACTCGAGTAAA |
| SMC2-2 sil F | ATATGAGCTCAGAACAACTGAATAGAGAGATTACACA |
| SMC2-2 sil R | AAAAGAATTCTAATCATTAAAAGACAAAAAGTTGCTCGAGTAAA |
| SMC4-1 sil F | ATATGAGCTCATGCAGACTGAAAGTGAAGAGC |
| SMC4-1 sil R | TTTACTCGAGTGTTTTGTTATGATGTCAGATTGTTCACT |
| SMC4-2 sil F | ATATGAGCTCATGATTAAGGAAGTTATTTTGGAGAATTTT |
| SMC4-2 sil R | TTTACTCGAGTATCTACCTTATTTTATTTTTCTGGAATCA |
| SMC2-1 Co-sil F | ATATCCCGGGATGTGGATCAAAGAAATCATTATCGA |
| SMC2-2 Co-sil F | ATATCCCGGGAGAACAACTGAATAGAGAGATTACACA |
| SMC4-1 Co-sil F | ATATCCCGGGATGCAGACTGAAAGTGAAGAGC |
| SMC4-2 Co-sil-R | TTTACCCGGGTATCTACCTTATTTTATTTTTCTGGAATCA |

**Table S1. Oligo sequences used for the study.**

## 4.2. Part II: Purification of IES excision machinery by affinity capture and mass spectrometry from *Paramecium tetraurelia*.

The elimination of IESs can be regulated at many different levels, such as sequence preference on DNA, and the interaction with small RNAs and proteins. Protein is the main executor of life activity, proteins bind to distinct sites on the DNA, which recognize and cleavage the corresponding IESs. From previous studies, PGM, a domesticated transposase, has been supposed to cleavage IESs during development[20]. Despite being such an important protein in this process, nobody succeeds in identifying the protein complexes that interact within/around the IES regions *in vivo* which makes the mechanism behind the IES excision still unclear. Here, we propose an *in vivo* method, adopted from Pooja Murarka[140], for the identification of IES binding excision machinery complexes in *P. tetraurelia* where the 3 tandem IESs-protein complexes formed *in vivo* are crosslinked by formaldehyde. In the autogamy or conjugation of *P. tetraurelia*, the development of new MAC is always accompanied by fragmented old MAC which makes any attempt to purify new MAC-specific stuff like the excision machinery always heavily contaminated by nonspecific fractions. The highlight point of this method is the idea that we use IES regions as the hot point to purify the excision machinery which is only present in new developing MAC. So, we can exclude the old MAC contamination. These complexes are further purified and the DNA is sequenced, bound proteins are identified, and potential binding RNA is kept for analysis. The method is schematically described in Figure 13. This method provides a new way to isolate *in situ* excision machinery based on complementary oligos purification and mass spectrometry analysis in the development of *Paramecium*.

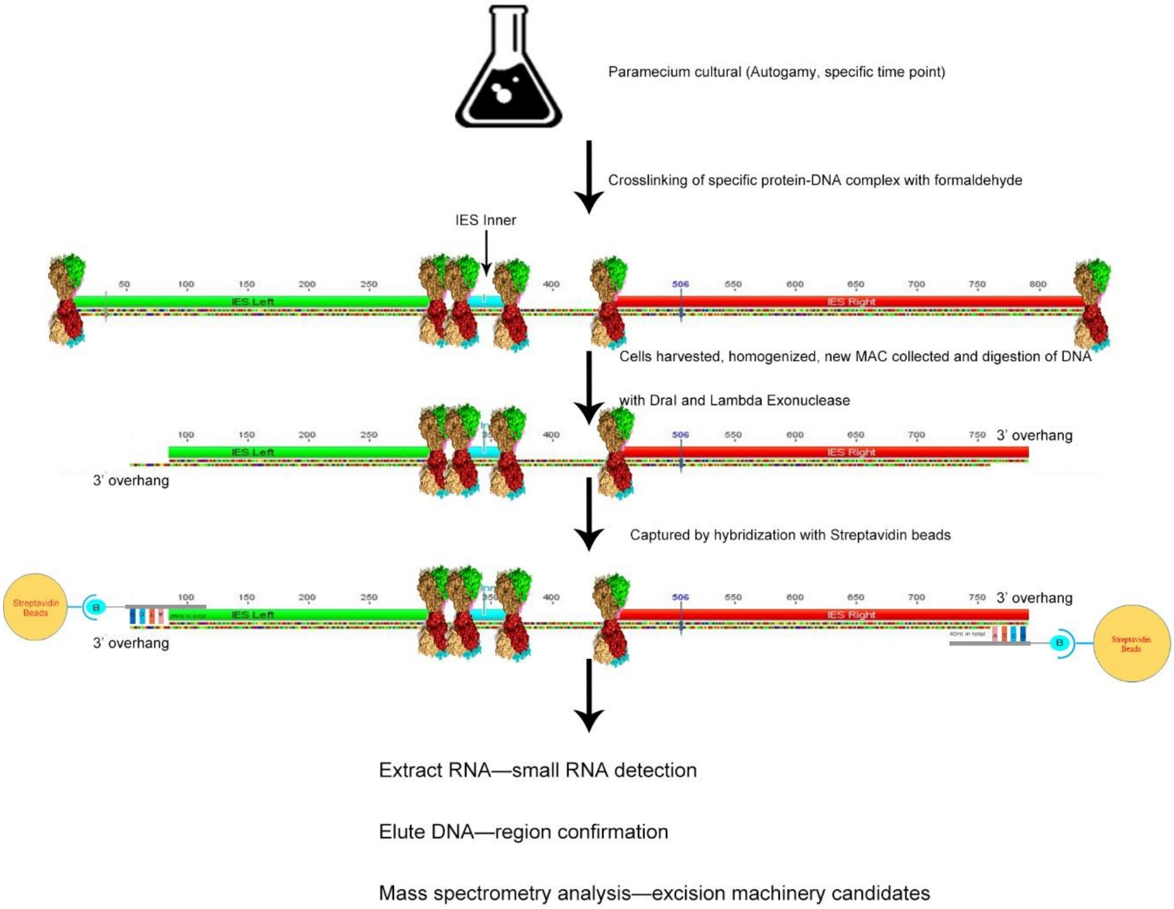


**Figure 13. Schematic diagram presenting the experimental procedure followed.** A method for isolation and identification of excision machinery (protein complex supposed to be attached at the end of each IES when cutting). By cross-linking the excision machinery to IESs with formaldehyde when cells are in the timing of IESs elimination, such protein-DNA complexes are fixed in vivo. The new MAC is then subjected to digestion with restriction enzymes to generate overhangs, which further binds to corresponding biotinylated oligos attached to streptavidin beads. The purified complex then proceeds accordingly.

A major problem in isolating the excision machinery from paramecium is that nobody for sure knows what protein is in the complex. Even the PGM is putative to be the executor of IES cleavage just because of some indirect evidences[20,107-109]. The chromatin immunoprecipitation (ChIP) assay is another extensively used approach for detecting DNA binding protein[141]. The ChIP assay is used to profile DNA-binding proteins throughout the whole genome. Aside from cost and availability, ChIP has technological limitations. Because this approach requires specific antibodies, it cannot be utilised when the target proteins are unknown[142]. Therefore, to gain a more comprehensive and precise insight into the intact excision machinery, accurately capturing the moment of IESs elimination by the complex has to be performed from DNA or RNA level. In this method, we decided to achieve the goal by performing on the DNA level since RNA is relatively unstable compared to DNA.

The first step of this experiment is to determine the target region in the genome which contains 3 tandem IESs regions and the length should be limited. There are several concerns with these specific requirements. Firstly, these 3 tandem IESs should include two longer IESs at the outside of the inner one (less than 100 bp) because this method needs restriction enzyme digestion on the outer IESs. The two outer IESs need a restriction site >100 bp from the innermost IES end (restriction sites should not cut inside of the outer restriction cut sites). This treatment shortens the IES and the excision machinery out of the restriction sites will be lost. Hence, longer IESs are needed to be sure there is enough IES residual left. Secondly, the possibility that short IES (26-28bp) employ different mechanisms of elimination compared to longer one reminds us that involving different length IESs would be a better way to have a comprehensive insight into the components of excision machinery. Thirdly, the priming site should be specific to these two outer IES ends (priming sites exposed by exonuclease digestion). Fourthly, since various proteins can bind with DNA in a bunch of biology progress, a negative control has to be included to exclude the IES-unspecific binding proteins. This negative control is picked up according to several principles like similar total length, same restriction sites, and close existence time with the target region (3 tandem IESs). Bioinformatic analysis was conducted by Victor Mason, detailed information regarding all candidates can be found in the appendix. To determine when the target region and negative control are present in new developing MAC of *Paramecium*, genomic DNA from different time point (100% fragments to 10 hours after 100% fragments, 2 hours interval) has been extracted and PCRs with specific primers were performed respectively as shown in Figure 14. The negative control refers to IESPGM.PTET51.1.23.362798 (798-, if not specified) and target region referring to IESPGM.PTET51.1.18.134377 (377+, if not specified) were chosen because both of them can be detected at the same time point. In addition to this, they have the same restriction site as Dra I which generate similar length of products. After sequencing the PCR products, we confirmed that 10 hours after 100% fragments is a reasonable time point for the experiment.

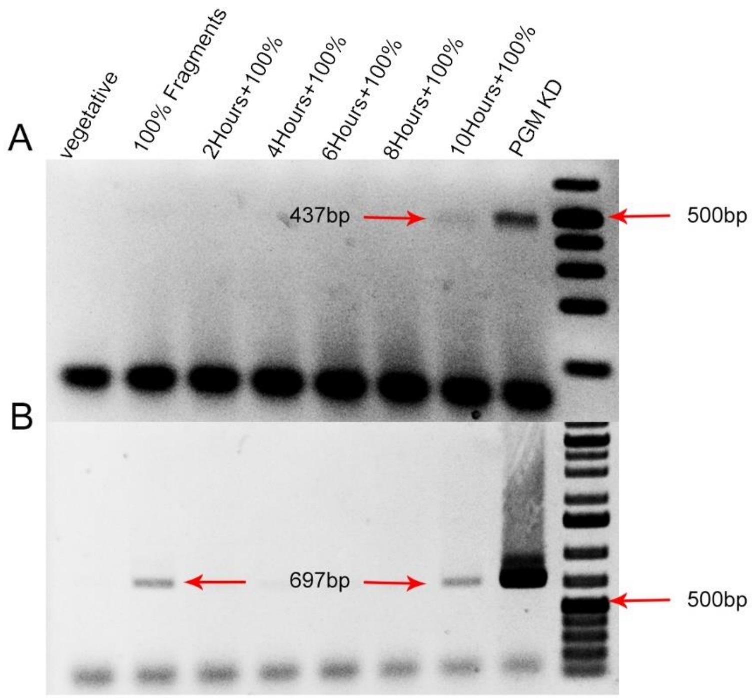


**Figure 14. PCR determines the presence of the 798- (697bp) and 377+ (437bp) target regions.** Genomic DNA of different time points were taken from vegetative (negative) to 10 hours + 100% fragments, as well as genomic DNA from PGM knock down as a positive control. Oligos which were used to pull down target regions were taken as primers without biotin modification. A) The 377+ region can only be detected at 10 hours + 100% fragments and positive control. B) The 798- region presented at 100% fragments, 10 hours + 100% fragments and positive control. The overlap timing of 377+ and 798- is 10 hours + 100% fragments.

The second step is the purification of new developing MAC without breaking the nuclear envelope. Since the MIC genome contains the IES sequence, it would be contamination to the specific binding of primer. So, we have to exclude the MIC genome from the purifying step. Due to the IES elimination is being believed to happen in developmental MAC, and we need the DNA-protein interactions not dissolved before elution, any intensive treatment that would potentially disrupt the interactions are forbidden including heating and a high amount of detergent. We modified a purification of the developing MAC method from O. Arnaiz[21] to isolate the intact developing MAC. This modification gives us the opportunity to keep the interactions in nuclei stable after ultracentrifuge.

The third step is nuclear membrane lysis and chromatin release. A high-quality chromatin preparation is very important to DNA-protein interaction capture. This step is critical to the successful enriching of the target region with binding proteins, since we need to lysis the nuclei membrane but not disrupt the DNA-protein interactions in nuclei. Commonly, 0.1% sodium dodecyl sulfate (SDS) is the concentration used in nuclear lysis, and luckily, according to H. Belaghzal's report[143], nuclear proteins that are not cross-linked to DNA can be eliminated, and the chromatin would be opened for better and more homogenous digestion treated with 0.1% SDS which means cross-linked DNA-proteins will not be disrupted when treated with 0.1% SDS. An additional step to open the chromatin by incubating at 65℃ for 10 minutes was included.

Detail operations including restriction enzyme digestion, Biotin labelled oligo incubation, Dynabeads™ MyOne™ magnetic beads purification and separation of different components are presented in the 'Method' part.

Methods

1. *Paramecium* strain 51, mating type 7 were used. Cultivation and autogamy were performed as previously described[122].

2. 100 mL of cells was used to extract DNA at each time point (100% fragments to 10 hours after 100% fragments, 2 hours intervals). PCR confirmation of 377+ and 798- were done with GoTaq polymerase (M3001, Promega) using primers without Biotin.

3. Collect cells at 10 hours + 100% in 15ml tubes with 20ul 50X protease inhibitor. Aliquots all into eight 15ml tubes, 4 for non-cross linking with two replicates for 798- and 377+, and 4 for the cross-linking group.

4. For cross-linking: 1 ml/sample formaldehyde added into the cell, shaking at room temperature (RT) 10min, gently invert the tube every 1-2 min.

   For non-cross linking, to step 6

5. Quench the cross-link with 100ul 1.25M Glycine, mix gently and incubate at RT 5min.

6. Fill up a few ml pre-cold PBS, gently resuspend, then fill the rest of the tube with PBS and invert the tube to mix, 4℃, 279rcf for 2min twice. Remove as much supernatant as possible.

7. Resuspend in 2.5 volumes of Lysis Buffer 1 and transfer half of them to the middle size homogenizer; incubate 5min on ice. In the meantime, pre-cool the centrifuge

8. Homogenise cells and check with DAPI staining frequently until no intact cell can be found.

9. Transfer lysate to a 15 mL tube, fill up with Wash Buffer, mix by inverting, centrifuge at 1000 RCF, 4 ℃ for 3 min, and remove supernatant.

10. Resuspend the pellet with 2 mL Wash Buffer, fill up to 15 mL, invert mixing, and spin down as in step 9.

    *Using a cut tip to avoid breaking MACs.*

11. Resuspend lysate in 3 volumes of Sucrose Buffer.

12. Transfer 3 mL of Sucrose Buffer to an ultracentrifuge tube (Beckman Coulter 344060 14×95 mm). Slowly add the cell lysate on top, then fill up with Wash Buffer (gently, avoid disrupting the layer formation).

13. Balance the samples and spin down for 1 h, 4˚C, 35,000 RPM with rotor: SW40Ti

14. Discard the supernatant with a Pasteur pipette, and then wash the pellet twice by resuspending it in 500 µL of ice-cold 1× ThermoPol® Reaction Buffer (10× attached, diluted to 1X). Then transfer all to a 1.5 mL RNase-free tube, centrifuging for 5 min at 2,500 ×g, 4˚C.

15. Remove the supernatant, resuspend in 1× ThermoPol® Reaction Buffer, and check the MACs under a microscope by DAPI staining.

*Save 18 µL of lysate for the chromatin integrity control:*

*Take 18 µl of lysate and add 32µL with 1x ThermoPol® Reaction Buffer and 10µL of Proteinase K (10 mg/mL). Incubate at 65 °C for 30 min. Purify DNA by single phenol-chloroform extraction without ethanol precipitation. Check the quality of the sample by running it on 0.75% agarose gel. The sample is good if DNA is either stuck in the well or runs as a single high band (>23 kb) (2ul DNA+8ul $H_2O$+2ul Loading Dye).*

16. Add 40 µL of 1% SDS to each sample and mix carefully by pipetting until MAC staining becomes transparent. Avoid making bubbles.

17. Metal bath for 10 min at 65 °C to dissolve chromatin. Then put tubes on ice immediately.

18. Quench the SDS with 44.4 µL of 10% Triton X-100. Mix gently by pipetting and avoid making bubbles.

19. Add 9.4 µL of 10× ThermoPol® Reaction Buffer to have a final 1× ThermoPol® Reaction Buffer).

20. Add 50 U Dra I, mix gently, and incubate at 37 °C for 5 - 15 min. Add 50 U Lambda exonuclease to the mixture and incubate for 15 s at 37 ℃. Incubate at 75 ˚C for 10 min to inactivate.

*(Optional) ASSESS THE DIGESTION*

*Mix 10 µL products with 40 µl of 1x ThermoPol® Reaction Buffer and 10 uL of Proteinase K (10 mg/ml). incubate at 65 °C, 30 min, purify DNA by phenol-chloroform extraction like mentioned before and run 0.75% gel electrophoresis.*

21. During the incubation, resuspend the Dynabeads™ MyOne™ magnetic beads in the vial by vortex for >30 sec.

22. Transfer 100ul beads to an RNase-free tube.

23. Add 1 mL of 1X Binding & Washing Buffer and resuspend.

24. Keep the tube on a magnet for 1 min and remove the supernatant.

25. Remove the tube from the magnet and resuspend the washed beads in 1ml W&B buffer.

26. Repeat steps 24-25 twice, for total of 3 washes.

27. Resuspend washed beads in 100 µL 2× Binding & Washing Buffer to a final concentration of beads at 5 µg/µL.

28. After inactivation, immediately add 5.52 µL of each Biotin-labelled primer (primer concentration at 1 µM) into the sample, mix well and incubate at 49˚C for 10 min.

29. Mix incubation and resuspended beads, and add 94.25 µL 5M NaCl to adjust NaCl concentration to 1 M. Incubate at 4˚C O/N on rotation.

30. Separate the coated beads with a magnet for 2 min, and keep the supernatant in a new RNase-free tube.

*Optional (check the pull-down efficiency):*

*Aliquots 50 µL mix with 10ul SDS loading dye and boil at 95˚C for 15-20 min and store at -20 ˚C.*

*Another 50 µL store at -20 ˚C for 377+ PCR detection.*

*Freeze left in liquid nitrogen, and store at -80 ˚C*

31. Resuspend beads with 1 mL 1X Binding & Washing Buffer and separate the beads with a magnet for 2 min, discard the supernatant.

32. Repeat step 31 twice.

33. Resuspend in 150 µL RNase free water and transfer 50 µL into a new 1.5 mL tube, add 12.5 µL 5× SDS loading buffer, boil at 95 ˚C for 15-20 min; Add another 100 µL

RNase free water to the leftover and transfer 100 µL in another new RNase-free tube, freeze in liquid nitrogen for RNA extraction later.

DNA Extraction from the beads

34. Wash the DNA-coated Dynabeads in 50 µL 1× SSC with magnet separation.

35. Resuspend the beads in another 50 µL of 1× SSC, and incubate at 95 °C for 5 minutes.

36. Immediately put the tube on the magnet, stand for 1 – 2 minutes, and transfer the supernatant to a nuclease-free tube, the supernatant contains a non-biotinylated DNA strand.

37. Performing PCR to check the existence of the 377+ regions.

RNA extraction from the beads

38. Add 1 mL TRI Reagent® to the frozen sample and vortex thoroughly, transfer all solution into a 15 mL tube and add another 2 mL TRI into the 15 mL tube.

39. Put 15mL tube on rotation at RT for 5min.

40. Add 0.6 mL chloroform into the sample, vortex for at least 15 sec, then let the sample stand for 5 min at RT.

41. Spin down at max. speed for 15 min at 4 ˚C.

42. Carefully transfer the upper colourless phase into a 15 mL tube and aliquots 750 µL to each 1.5 mL tube.

43. Add 500 µL isopropanol to each tube, mix by inverting and stand at RT for 10 min.

44. Spin down at 1,200 RCF for 10 min at 4 ˚C.

45. Discard the supernatant, and wash the pellet with 1 mL of 75% ethanol.

46. Spin down at 7,500 RCF for 5 min at 4 ˚C.

47. Discard supernatant, and wash with 1 mL 75% ethanol once.

48. Discard supernatant, centrifuge for 30 sec, and discard the supernatant.

49. Make a tent with RNase Zap cleaned alumina, leave the lid opening, and air dry the pellet under 37 ˚C for 30 min.

50. Resuspend with 30 µL nuclease-free dd$H_2O$, and incubate at 60 ˚C for 15 min to dissolve the pellet.

Silver staining to check the enriched proteins

51. A 10% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gel run as usual.

52. Fix gel in 100 mL fixative solution overnight at RT with shaking.

53. Wash the gel 20 min with 50% Ethanol on shaking at RT, twice.

54. Treat the gel exactly for 1 min with 98ml Sodium thiosulphate (3-4 flakes in 100ml MilliQ water).

55. Wash the gel 20 seconds with MilliQ water three times.

56. Treat the gel with $AgNO_3$ for 30 min.

57. Wash the gel 20 seconds with MilliQ water for three times.

58. Develop the gel with 100ml developing solution until the bands reached the desired intensity.

59. Stop the reaction with 100ml Stop Solution

60. Scan the gel warped with glossy film.

Mass spectrometry preparation and analysis

Two replicates of each group were sent for mass spectrometry at the Mass Spectrometry and Proteomics facility (PMSCF, Bern, Switzerland). A 12 % SDS-PAGE gel was prepared and run 0.6cm into the resolving gel, the apparatus was disassembled and the stacking gel was removed. The gel was stained with InstantBlue (Expedeon) until the bands became visible and the band of interest was then cut out and cut into 6 small cubes of about 1 $mm^2$ and the cubes were put in an Eppendorf tube. The cubes were covered with 100 µL of 20% ethanol, and stored at 4 °C before sending.

Protein identification and statistics were performed by Manfred Heller at the PMSCF. Shotgun Liquid Chromatography Tandem Mass spectrometry (LC-MS/MS) with non-labelled samples was performed and the peptide identification was made with EasyProt software and processed with MaxQuant software for estimation of protein abundances[144,145]. For the analysis, the LFQ and Top3 algorithms built into MaxQuant software were used for the label-free protein quantification.

Only proteins exclusively present in the 377+ group and not 798- samples, as well as deemed significant in both Top3 and LFQ were analysed. To filter the results, the IDs were converted into UniProt Accession numbers and entered into the PANTHER and AgBase databases to retrieve Gene Ontology terms and protein names[146-148]. Known contaminants such as ribosomal proteins and proteins not present in the same cellular compartment as the bait was excluded. The remaining proteins were manually searched for in the ParameciumDB to find out potentially interested proteins.

Recipes

Lysis buffer 1

| | Amount | Final concentration |
|---|---|---|
| $H_2O$ | 4.915ml | |
| Sucrose | 0.43g | 0.25M |
| 2M $MgCl_2$ | 25ul | 10mM |
| 1M Tris pH6.8 | 50ul | 10mM |
| NP-40 | 10ul | 0.2% |
| Roche protease inhibitor | 100ul | 1X |

Wash buffer

| | Amount | Final concentration |
|---|---|---|
| $H_2O$ | 197ml | |
| Sucrose | 17.2g | 0.25M |
| 2M $MgCl_2$ | 1ml | 10mM |
| 1mM Tris pH7.4 | 2ml | 10mM |

Sucrose buffer

| | Amount | Final concentration |
|---|---|---|
| $H_2O$ | 14.775ml | |
| Sucrose | 10.77g (little by little) | 2.1M |
| 2M $MgCl_2$ | 75ul | 10mM |
| 1M Tris pH7.4 | 150ul | 10mM |

2X Binding & Washing Buffer

| | Amount | Final concentration |
|---|---|---|
| 1M Tris-HCl pH 9 | 100ul | 10mM |
| 0.5M EDTA | 20ul | 1mM |
| 2M NaCl | 9.88ml | 2M |

Fixative Solution

| | Amount | Final concentration |
|---|---|---|
| MilliQ water | 39.95ml | |
| Methanol | 50ml | 50% |
| Acetic Acid | 10ml | 10% |
| Formaldehyde | 50ul | 50ul |

Developing Solution

| | Amount |
|---|---|
| MilliQ water | Up to 100ml |
| $Na_2CO_3$ | 6g |
| Sodium Thiosulphate | 2ml |
| Formaldehyde | 50ul |

Stop Solution

| | Amount |
|---|---|
| MilliQ water | 95ml |
| Acetic Acid | 5ml |

Results

To assess the potential timing for the purification, PCR using 377+ and 798- primers without Biotin were conducted at each time point separately. The results were shown in Figure 14. As shown, only at 10 hours +100% fragments can both 377+ and 798- be seen which means it's a possible time window for pulling down interacting proteins *in vivo*. To confirm the corresponding bands are the desired region we expected, the PCR products were analysed and confirmed by sanger sequencing in Microsynth company (Switzerland) using corresponding primers.

To assure the time point was also correct when performing the pull-down assay. Total DNA was extracted from the aliquot of culture and checked the presence of the target region with PCR. As seen in Figure 15, both target regions can be observed in the harvested cells. To confirm the target DNA was enriched on the beads, another PCR was done with the eluted DNA and the results were shown in Figure 15. Taken together, it can be assured that desired DNA regions were enriched on beads.

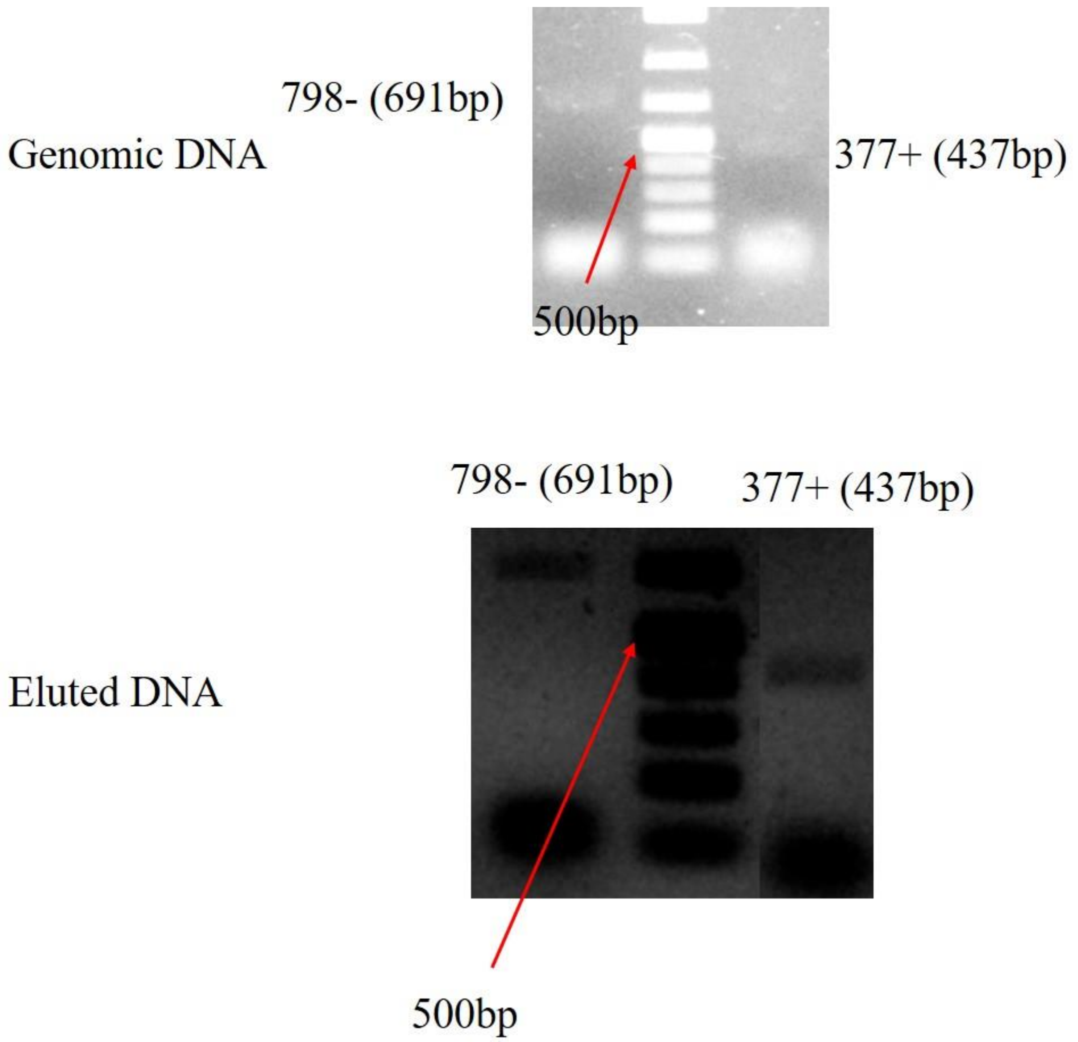


**Figure 15. PCR determines the presence of the 798- (697bp) and 377+ (437bp) from genomic DNA and elution.** DNA from cells and elution of 10 hours + 100% fragments were extracted.

To normalize the loading amount of each sample for mass spectrometry, we assumed that the number of proteins should be the same within the group if we could balance the concentration of DNA on beads. After eluting the DNA from the beads and measuring the concentration of each sample, samples corresponding to the same amount of DNA were loaded onto the SDS gel to check the enriched proteins (as shown in Figure 16) and prepare for mass spectrometry analysis.

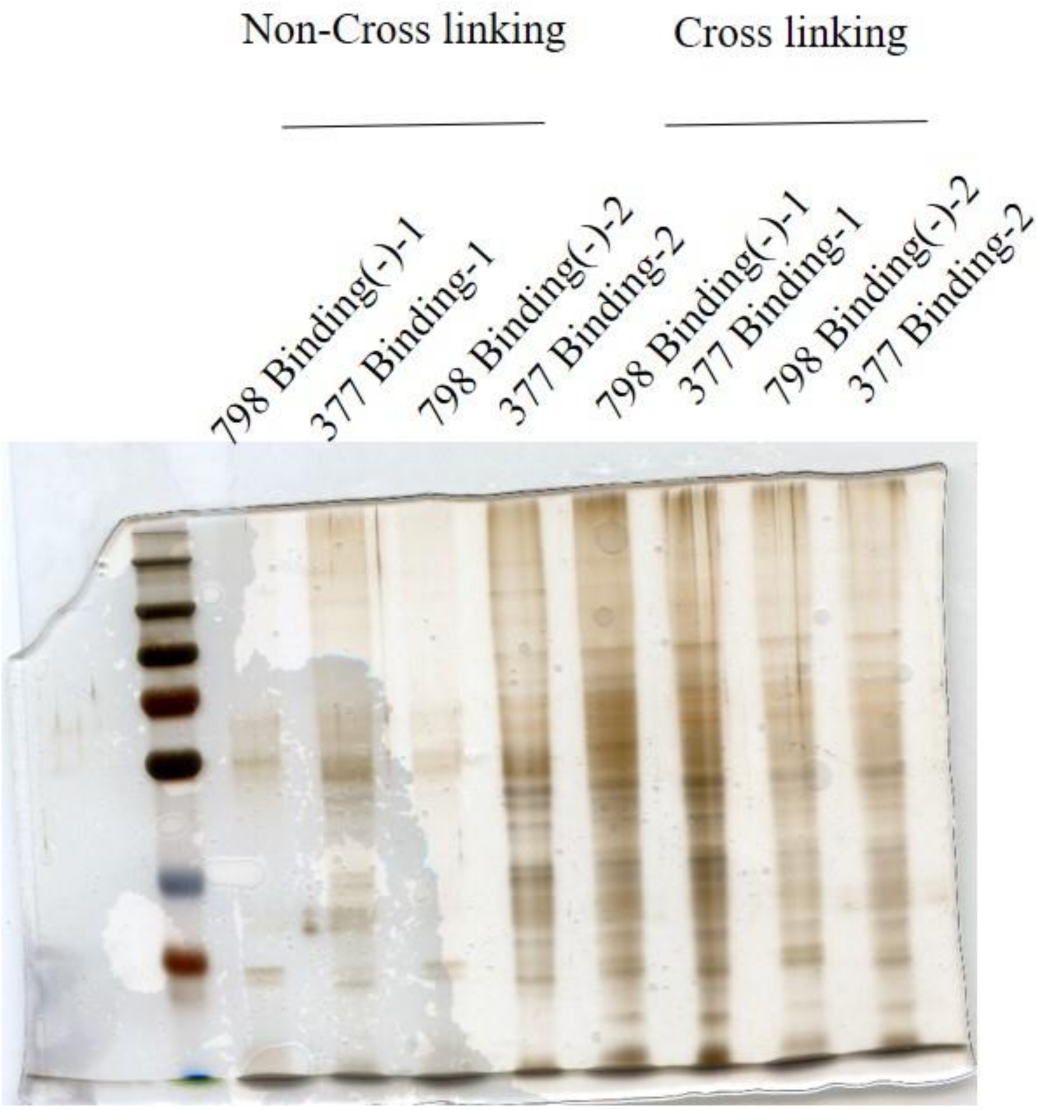


**Figure 16. Silver staining of the enriched complex.** All loading samples were normalized according to the eluted DNA amount. Both non-cross-linking and cross-linking are included. A 10% SDS-PAGE gel was used. Silver staining was performed.

All enriched proteins that were only presented in 377+ groups were shown in Table 1. However, none of the known key players of *P. tetraurelia* genome rearrangements, like the PTIWI proteins, PGM, NOWA proteins, EZL1, etc. was found by mass spectrometry. Several assumptions can explain this phenomenon. The first one is that the excision machinery which may contain PGM transiently binds on its target region and detach immediately after introducing cleavage on the IES boundary. This potential ‘cut and go’ mechanism indicates the chance of the excision machinery enrichment is extremely low which means it’s possible that we only pull down the DNA region without the excision machinery. Although we have already confirmed the time point, we took is a good one on the DNA level, it's still possible that 10 hours + 100% fragments is too late for IES elimination. In general, key players of IES elimination have relatively early expression profiles, like PGM, PTIWIs, DCLs, NOWAs, etc. are highly expressed from MEI to DEV1 which are all earlier than 10 hours + 100% fragments. In the meantime, M. Bétermier etc. has already reported that excision was shown to start 12 to 14 h after conjugation and was essentially completed at 16 h, while conjugation and autogamy share very similar cytological steps[16]. We did not compare the 16 hours after conjugation with 10 hours + 100% fragments experimentally, but according to our experiences, 16 hours after conjugation is much earlier than 10 hours + 100% fragments which means by that time, IES excision should have been finished and the excision machinery should not work anymore. The above possibilities remind us that other time points should be taken into consideration for enriching the excision machinery.

**Table 1. List of unique peptides found in 377+, not 798- mass spectrometry.**

| | Synonyms | Description | iTop3 log2FC 377NC - 798NC |
|---|---|---|---|
| PTET.51.1.P0010534 | TMP2d, TMP2D, TMP_T2d, | Coiled coil domain | 2.534444058 |
| PTET.51.1.P0030180 | | Ammonium transporter AmtB-like domain | 3.166234709 |
| PTET.51.1.P0040316 | | Coiled coil domain | 3.517747077 |
| PTET.51.1.P0060088 | EPI31 | Protein of unknown function DUF2816 | 5.789938319 |
| PTET.51.1.P0070055 | | Coiled coil domain | 4.692241051 |
| PTET.51.1.P0070167 | | Ammonium transporter AmtB-like domain | 3.11997365 |
| PTET.51.1.P0090210 | TMP21j, PTETG900001001 | Coiled coil domain | 5.785905363 |
| PTET.51.1.P0110059 | | Coiled coil domain | 5.277202169 |
| PTET.51.1.P0110160 | | Ammonium transporter AmtB-like domain | 3.155710977 |
| PTET.51.1.P0110226 | PTETG1100007001, SFA12d | Coiled coil domain | 4.091869323 |
| PTET.51.1.P0110247 | | Coiled coil domain | 4.184921652 |
| PTET.51.1.P0130214 | EPI21 | Protein of unknown function DUF2816 | 3.40344546 |
| PTET.51.1.P0130289 | | Coiled coil domain | 2.254467887 |
| PTET.51.1.P0130338 | EPI17 | Protein of unknown function DUF2816 | 5.822698301 |
| PTET.51.1.P0160035 | EPI26 | Protein of unknown function DUF2816 | 6.81661733 |
| PTET.51.1.P0170089 | EPI36, KdA2 | Protein of unknown function DUF2816 | 5.370628316 |
| PTET.51.1.P0170267 | EPI45 | Protein of unknown function DUF2816 | 5.733245774 |
| PTET.51.1.P0180254 | | | 3.938814947 |
| PTET.51.1.P0180262 | TMP1f, T1f, TMP_T1f, | Coiled coil domain | 4.015563305 |
| PTET.51.1.P0190295 | EPI16 | Protein of unknown function DUF2816 | 2.254481094 |

| | | | |
|---|---|---|---|
| PTET.51.1.P0200169 | EPI19 | Protein of unknown function DUF2816 | 2.106901089 |
| PTET.51.1.P0210006 | | Coiled coil domain | 3.343525572 |
| PTET.51.1.P0230038 | | | 3.963123772 |
| PTET.51.1.P0230185 | TMP21c, PTETG2300009001 | Coiled coil domain | 3.503985457 |
| PTET.51.1.P0260023 | | Coiled coil domain | 3.296661406 |
| PTET.51.1.P0260169 | TMP21h, PTETG2600001001 | Coiled coil domain | 2.927813739 |
| PTET.51.1.P0270102 | EPI46, KdA10 | Protein of unknown function DUF2816 | 8.374352378 |
| PTET.51.1.P0310127 | | Histone-fold | 1.129961656 |
| PTET.51.1.P0330153 | EPI13 | Protein of unknown function DUF2816 | 6.145864103 |
| PTET.51.1.P0350305 | PTETG3500005001 | Coiled coil domain | 2.804309603 |
| PTET.51.1.P0360200 | TMP42f, PTETG3600005001 | Coiled coil domain | 2.748648714 |
| PTET.51.1.P0380195 | | Glyceraldehyde/Erythrose phosphate dehydrogenase family | 3.236344749 |
| PTET.51.1.P0390135 | TMP1i | Coiled coil domain | 3.683297633 |
| PTET.51.1.P0440082 | TMP31e | Coiled coil domain | 4.324150318 |
| PTET.51.1.P0460182 | TMP1d, T1d, TMP_T1d | Coiled coil domain | 4.696906944 |
| PTET.51.1.P0500133 | EPI29 | Protein of unknown function DUF2816 | 6.840888567 |
| PTET.51.1.P0510026 | PTETG5100005001 | Coiled coil domain | 1.542966515 |
| PTET.51.1.P0520250 | EPI34 | Protein of unknown function DUF2816 | 6.095102095 |
| PTET.51.1.P0540235 | | Coiled coil domain | 3.349462273 |
| PTET.51.1.P0580083 | | Fibrillarin | 1.736609963 |
| PTET.51.1.P0590095 | | Coiled coil domain | 3.994218106 |
| PTET.51.1.P0590139 | | Coiled coil domain | 0.83176968 |
| PTET.51.1.P0600043 | EPI35 | Protein of unknown function DUF2816 | 3.149456236 |
| PTET.51.1.P0620035 | | Coiled coil domain | 4.04062303 |
| PTET.51.1.P0660064 | | | 1.115913082 |
| PTET.51.1.P0660264 | | Coiled coil domain | 4.393618351 |
| PTET.51.1.P0710131 | alphaPT4, tub_alphaPT4, | Tubulin | 1.598686896 |
| PTET.51.1.P0710219 | EPI39 | Protein of unknown function DUF2816 | 3.248001092 |
| PTET.51.1.P0720201 | | Coiled coil domain | 3.82938176 |
| PTET.51.1.P0760096 | EPI2 | Protein of unknown function DUF2816 | 4.194158052 |
| PTET.51.1.P0790177 | | Coiled coil domain | 2.738097151 |
| PTET.51.1.P0870174 | | Histone H2B | 3.361690641 |
| PTET.51.1.P0890143 | EPI32 | Protein of unknown function DUF2816 | 2.76999129 |
| PTET.51.1.P0900049 | EPI41 | Protein of unknown function DUF2816 | 5.405962689 |
| PTET.51.1.P0940194 | | Ammonium transporter | 4.716764313 |
| PTET.51.1.P0980135 | T1-b, TMP1b | Coiled coil domain | 2.802661429 |

| | | | |
|---|---|---|---|
| PTET.51.1.P0980151 | TMP42q, PTETG9800005001 | Coiled coil domain | 4.046552875 |
| PTET.51.1.P0990148 | EPI49 | Protein of unknown function DUF2816 | 3.926971376 |
| PTET.51.1.P1120164 | | Coiled coil domain | 2.988272234 |
| PTET.51.1.P1170022 | | Histone-fold | 3.085289639 |
| PTET.51.1.P1190042 | | Ammonium transporter AmtB-like domain | 4.363070913 |
| PTET.51.1.P1190122 | EPI33 | Protein of unknown function DUF2816 | 3.312833769 |
| PTET.51.1.P1210094 | TMP4b, T4c, TMP_T4c, | Coiled coil domain | 3.043418246 |
| PTET.51.1.P1210124 | TMP32b | Coiled coil domain | 2.059115894 |
| PTET.51.1.P1370064 | H3P3 | Histone H3/CENP-A | 2.870751576 |
| PTET.51.1.P1370127 | PtRPB1, RPB1 | RNA polymerase Rpb1, domain 5 | 1.463019759 |
| PTET.51.1.P1400098 | EPI51 | Protein of unknown function DUF2816 | 3.338539768 |
| PTET.51.1.P1470130 | TMP21e | Coiled coil domain | 2.883304001 |
| PTET.51.1.P1470169 | PTETG14700002001 | Coiled coil domain | 3.910248714 |
| PTET.51.1.P1500107 | PTETG15000003001 | Coiled coil domain | 2.566474837 |
| PTET.51.1.P1630088 | beta-PT2, bPT2, SU3, tub_betaPT2 | Tubulin | 3.837856426 |
| PTET.51.1.P1710041 | EPI6 | Protein of unknown function DUF2816 | 7.081169883 |
| PTET.51.1.P1730070 | | EF-hand domain pair | 6.162553257 |
| PTET.51.1.P1770034 | | | 5.329433056 |
| PTET.51.1.P5560009 | | Histone H2A | 5.877785018 |

## 4.3. Part III: Expression of dominant negative form of PiggyMac in *Paramecium tetraurelia*

Nobody has succeeded in enriching PGM interacting proteins *in P. tetraurelia*. As mentioned above, the excision machinery potentially works in a 'cut and go' way in *P. tetraurelia* which indicates the interaction between excision machinery and the IES regions is transient. Our previous method does not have the potential to capture such short-time interaction. Hence, we conducted an expression of dominant negative PGM[107] (PGM 3A) tagged with Flag-HA in *P. tetraurelia* and performed Co-immunoprecipitation at the late time point. Since the three conserved aspartic acids are supposed to constitute a functional catalytic triad[20], the PGM 3A harbours D to A substitution at these positions has a dominant-negative effect in *P. tetraurelia*. Using this mutant, we assumed that the presumed 'cut and go' way would be switched to 'bind but stay', which gives us a great chance to purify the PGM interacting proteins without considering the transient binding. By this method, we were trying to answer the question that why this purification is still not reported and it should fill the gap if we could succeed.

First of all, the PGM coding gene was amplified from the genomic DNA and inserted into the pGEM-T easy vector by TA cloning. After that, the Flag-HA-coding fragment which recoded according to the codon preference of *Paramecium* was inserted before the stop codon of the PGM 3A gene. In this construct, the PGM 3A-Flag-HA coding sequence joints upstream and downstream non-coding sequences for its expression and regulation. Then, the positive plasmid was extracted and linearized by Ahd I (New England Biolabs). The purified fragment was microinjected into the vegetative MAC and positive injection strains were confirmed by doing Dot Blot. The last step is enrichment. Different time points were taken for detecting the HA signal by western blot as shown in Figure 17. In the meantime, we performed IES PCR and survival test to check the influence of PGM 3A mutation on the IES elimination and survival of its progeny. From Figure 18 A and B, we can tell all chosen IESs' excision were inhibited which is consistent with the dominant-negative effect on sexual progeny survival. It supports the idea that dominant-negative protein not only coexists with endogenous protein but also exhibits an impaired function out of the normal one.

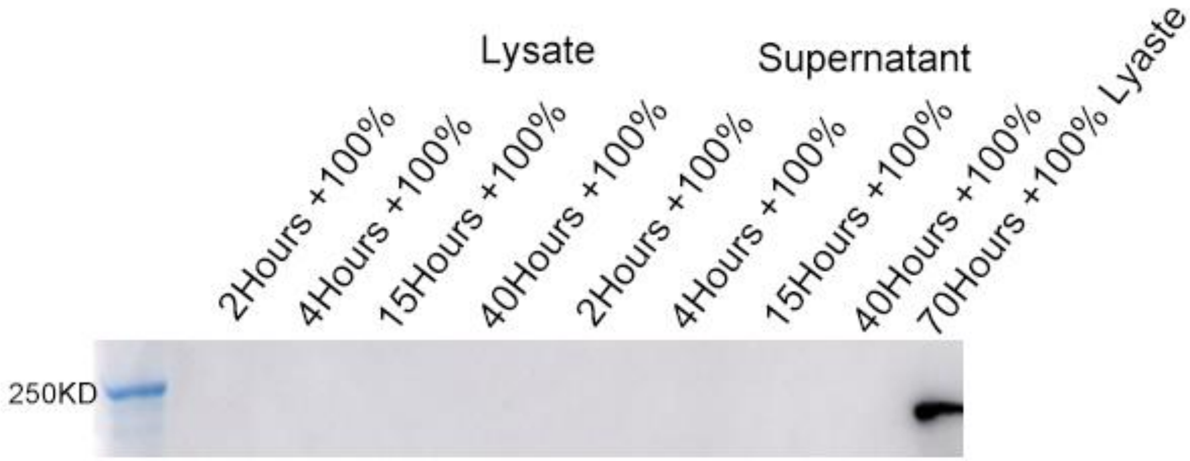


**Figure 17. Western blot analysis with the anti-HA antibody of PGM-3A-Flag-HA injected cells from different time points.** Transformed cells were cultured to corresponding time points, and cell lysate and supernatant were taken equally to prepare the samples for western blot. At the 70hours + 100% fragments, only cell lysate was taken.

There are several points that should be looked at carefully. We tried a lot of different conditions before the band observed on the membrane although it is a bit too late for IES excision. Therefore, we concluded some hints for future study. The first one we want to emphasize is the pre-treatment with 5% SDS under 100 ℃ for 3 min before boiled with 5X SDS loading buffer as previously described[109]. Western blot never returns a positive band in PGM 3A-Flag-HA without this pre-treatment. An additional 5% SDS (final concentration) supply reminds us the possibility of why there is no report regarding the *in vivo* purification using PGM as bait. It looks like the PGM is not disrupted even boiled with 2% SDS at 95 ℃. We assumed that the HA tag may be folded inside of the PGM structure which makes the Anti-HA primary antibody fail on binding. If our assumption was true, then the Co-IP using Anti-HA beads would not enrich the PGM 3A-Flag-HA from the lysate either. To test this possibility, we introduced a codon optimized GFP sequence between PGM and Flag-HA in the pGEM-T easy vector. This insertion could in theory increase the spatial distance between PGM and Flag-HA after being expressed in *P. tetraurelia* MAC, which could make the HA tag detectable. Another benefit of introducing GFP is the determination of timing and localization[107] of PGM 3A in the development of *P. tetraurelia*. In Figure 19, PGM 3A-GFP-Flag-HA displayed a homogeneous pattern without significant foci suggesting that PGM foci are associated with its normal function, especially with IES elimination as shown in Figure 18 A and B.

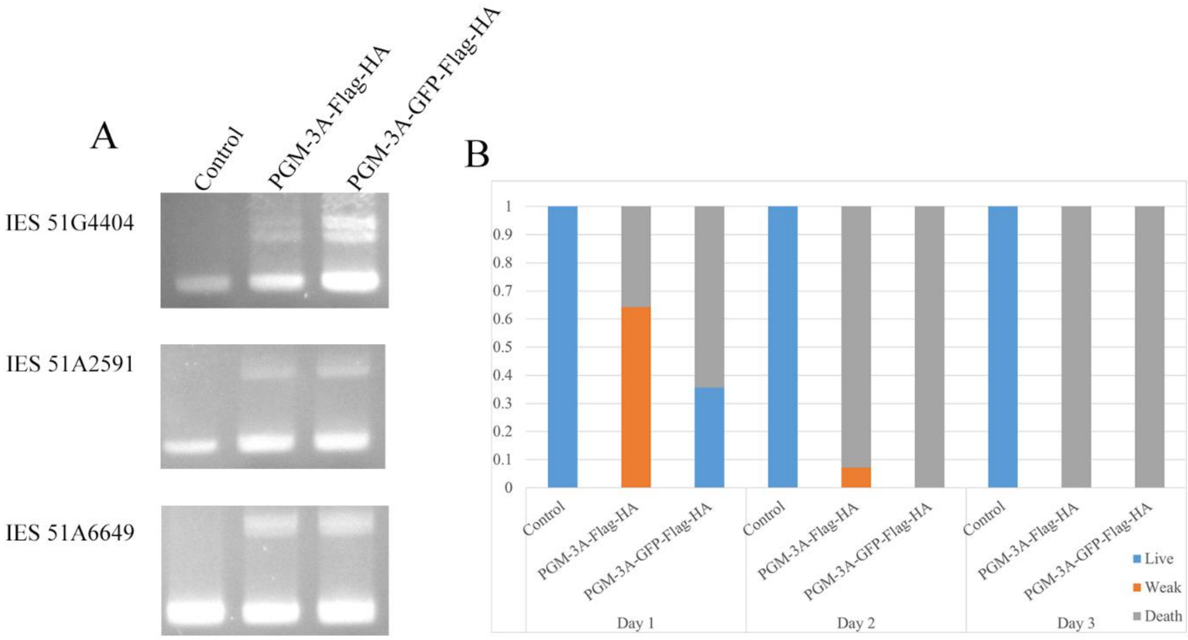


**Figure 18. Effect of PGM-3A-Flag-HA and PGM-3A-GFP-Flag-HA injection on IES excision and progeny survival.** A) The IES retention PCR results. Genomic DNA was extracted from vegetative cells, PGM-3A-Flag-HA and PGM-3A-GFP-Flag-HA injected cells at the post autogamy stage. The retention was indicated by upper bands. B) The survival test. A total of 14 cells were taken from each group. Y axis presents the percentage of live, weak or death cells. X axis presents each group in 3 days.

The second one is the time point. According to the expression pattern of PGM in *P. tetraurelia*, there should be a large amount of PGM expressed in each time point we took for the western blot and Co-IP. According to the results, the PGM 3A-Flag-HA signal can only be seen at a postautogamous time point (around 70 hours + 100% fragments). As we mentioned above, 10 hours +100% fragmentation is too late to purify the excision machinery, the signal we got at such a late time point (70 hours +100% fragmentation) could just be because of the accumulation same as the previous report[107]. This accumulation could be useless because of the possibility that interacting proteins disassembled either excision happened or not.

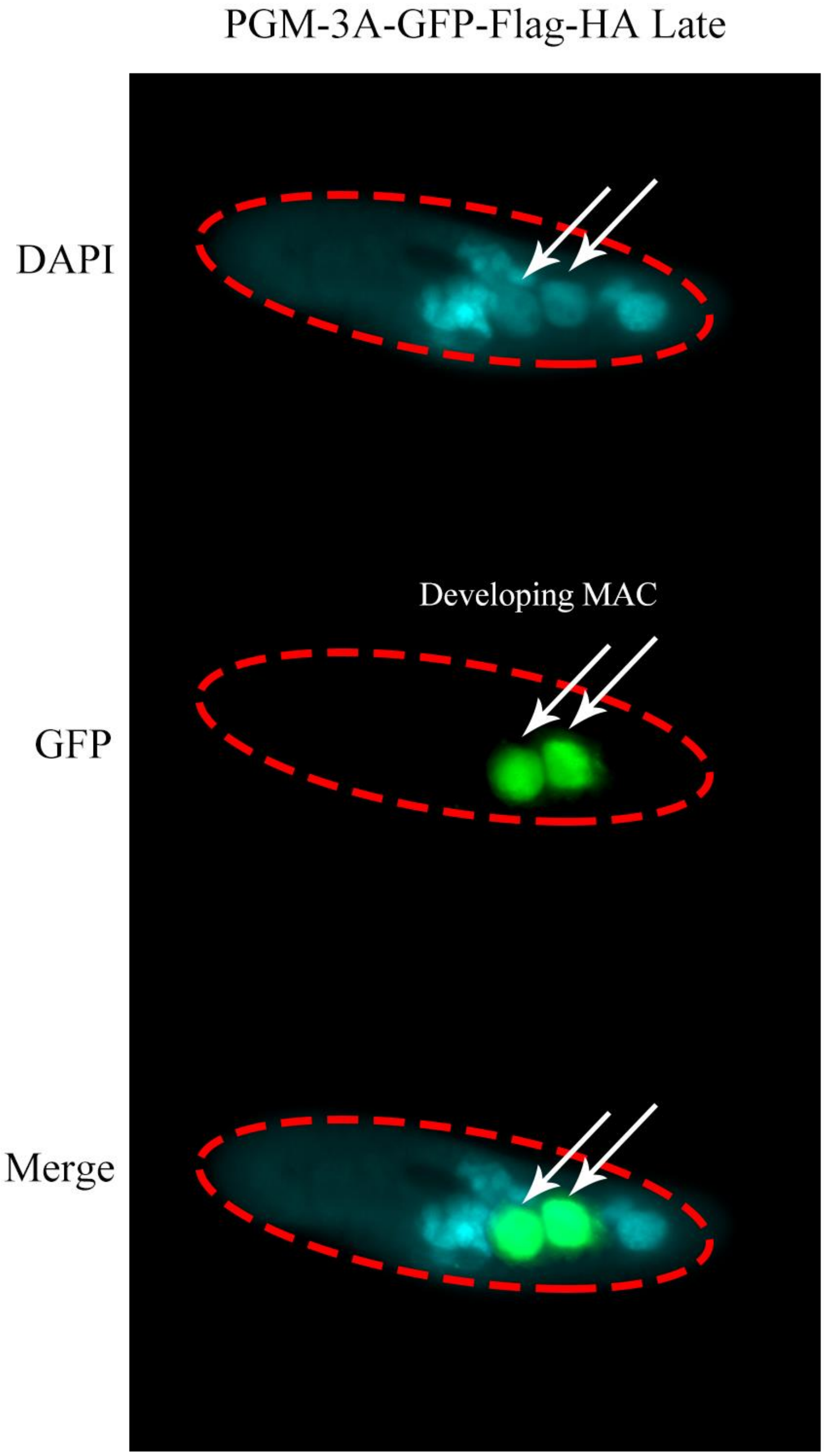


**Figure 19. PGM-3A-GFP-Flag-HA localization in *Paramecium tetraurelia*.** Localization of PGM-3A-GFP-Flag-HA in the developing MAC. The top panel indicates DAPI staining in blue, which

indicates DNA. The panel in middle shows the GFP signal in green. The panel at the bottom represents a merge of GFP and DAPI. White arrows point new developing MACs.

Another hint is about the transferring condition. Although we did not test other transferring conditions on 70 hours +100% fragments sample, we do try different transferring time on other time points and none of them succeeded in imaging the HA signal if we ignored the time point issue. So, the only condition that returns a clear HA signal on the membrane after western blot is 4 hours of transferring under constant 110V on ice. Be careful, with long time transferring, keeping the solution at a low temperature is crucial.

## 5. Conclusions and perspective

*Paramecium tetraurelia* is characterized by the co-existence of two types of nuclei in one cytoplasm: a diploid micronucleus (MIC) and a highly polyploid macronucleus (MAC). The development of the MAC raising from the MIC and the genome rearrangement in this progress is an excellent model to understand the mechanism behind DNA elimination and recombination in other organisms. The mechanism of genome rearrangement in *P. tetraurelia* has been thoroughly investigated in recent years, and numerous novel components involved in IES elimination have been found. With processing IESs, new developing MAC generated from MIC go through several rounds of meiosis and mitosis where condensin is supposed to be essential to the faithful genome structure and separation. However, whether condensin participated in IES elimination or not is still unclear.

In all species, structural maintenance of chromosomes (SMCs) proteins are essential for effective chromosomal transmission during replication and genomic segregation. Condensin core subunits (SMCs) are conserved across all eukaryotic species investigated so far. In addition to the SMC subunits, condensin also contains a distinct set of non-SMC subunits. Although SMCs in ciliates are close relatives to those in other organisms, little is known of their detailed functions in *P. tetraurelia*. Furthermore, the investigation of *P. tetraurelia* SMCs may reveal a new aspect of the function of SMCs. To explore whether condensin participated in IES processing, we focused on the behaviour of SMC subunits of condensin in IES elimination. In this study, we identified one of two SMC4s in *P. tetraurelia*, SMC4-2, is necessary for proper genome rearrangement and progeny survival. The specific localization of SMC4-2 in new developing MAC provides the spatial possibility for participating in IESs elimination. The results from IES PCR, genome sequencing and survival test confirm that SMC4-2 is needed in the processing of IESs in the development of *P. tetraurelia* in some unknown mechanism. One potential explanation for this finding is that SMC4-2 can manipulate the production of iesRNAs, one group of small RNAs which are crucial to the positive feedback of IES elimination in scanning model, according to the small RNAs sequencing analysis. Distinct from the performance of SMC4-2 in *P. tetraurelia*, SMC4-1 expresses and behaves in a relatively early stage which seems like be able to get the genome ready for producing scnRNAs. This hypothesis comes from the localization of SMC4-1 in all stage of nuclei except the MAC that is going to be fragments. Interestingly, although no IES retention phenotype has been found in the RNAi of SMC2-1, SMC2-2 and SMC4-1, co-silencing of these genes with SMC4-2 separately abolished the IES retention phenotype in single silencing of SMC4-2. These phenotypes remind us that SMC2s and SMC4-1 could participate in SMC4-2 processing IES elimination in unknown mechanisms. By performing *in vivo* Co-immunoprecipitation and mass spectrometry analysis, we found that the interactions between SMC4-1 and SMC4-2 changed accordingly. Using SMC4-1 as bait, SMC4-2 can be enriched as an interacting protein of SMC4-1, while SMC4-1 cannot be detected using SMC4-2 as bait *in vivo*. We proposed a competitive model in that the competitive binding of DNA by SMC4-1 and SMC4-2 changes according to the ratio of concentration between them. This competition blocks or unlocks the IES accessibility to the excision machinery. Due to time constraints, we did not do further research to test the model. Therefore, this hypothesis has to be looked at carefully.

Programmed DNA elimination is a significant developmental genome modification in which certain DNA sequences, up to 90% of the genome in some cases, are removed from somatic lineages. Programmed DNA elimination in animals has sparked a lot of discussion curiosity and speculation since its discovery in 1887. The single-cell ciliates including *P. tetraurelia* are the best-studied instances of programmed DNA elimination in eukaryotes. While the mechanism behind this massive genome rearrangement remains poorly understood. Although a scanning model has been widely accepted to explain how two sets of small RNAs guide the IES excisions, the real players (also refer to as an excision machinery) of IES elimination are still unclear. In this part of my study, we developed an improved method to purify the excision machinery by affinity capture and mass spectrometry. The main idea of this attempt is to capture potential DNA excision machinery binding on the IES region by oligos complementary to the IES. By using bioinformatic analysis, we picked up several candidates for both target region and negative control as well as confirmed the existence time of these regions. With the Biotin labelled oligos complementary to the target region or negative control, we successfully developed, optimized and practiced this 3-IESs capturing method on *P. tetraurelia*. We were able to collect enriched interacting data from specific DNA regions at a specific time point during genome rearrangement. However, no key players of this progress like PGM, PTIWI, NOWA, or even the SMC4-2 we identified above, have been found in the results. Although the target region and control DNA regions have been confirmed in both input and output, the MS results still need to be looked at carefully. We provided several possibilities for these scenarios: I) the timing of IES elimination could be earlier than the time point we chose for the capturing. It has been reported before that IES excision was shown to start 12 to 14 h after conjugation and was essentially completed at 16 h which means the time point (10 hours + 100% fragments) we took could be too late to capture the DNA excision machinery since the IES excision has been done. II) it also could be because the excision machinery behaves in a ‘cut and go’ way, which means its binding on DNA is transient and the time window for capturing the excision machinery-DNA complex is too narrow.

The PGM is a crucial candidate in IES excision which has high homology with PiggyBac transposase. It should be a good candidate to study the mechanism behind IES excision in *P. tetraurelia*. By introducing Flag-HA tagged PGM with three mutants on its conserved aspartic acids, we try to do the *in vivo* co-immunoprecipitation to enrich the PGM complex during the development of *P. tetraurelia*. By introducing these mutants, those PGMs were supposed to be not able to cut the IES regions, hence the DNA excision machinery where PGM should be involved would not dissociate from the IES region. We assume it would provide a good chance to enrich the DNA excision machinery by the IES-specific pull down method we developed. Although we failed on detecting any HA signal from PGM-3A on anti-HA beads enrichment, we do provide some hints for further study. At least to PGM, *P. tetraurelia* can exhibit the impaired PGM effect even the endogenous PGM is there, which makes further functional research easier. Because researchers do not need to knock down or knock out the endogenous proteins before introducing the manipulated one.

# 6. Methods and Materials

## 6.1.*Paramecium* strain and cultivation

The strain 51, mating type 7 of *Paramecium tetraurelia*, was used. Cultivation and autogamy were performed at 27 °C as described[4].

## 6.2.Ethanol fixation and staining for *Paramecium*

Collect and wash samples at specific time point as previously described, add 5 mL ice-cold 70 % ethanol to the cell pellet for fixation. Transfer all into a 15 ml tube. Store at 4 °C until further implementation.

Transfer 150 µL fixed cell on glass slide, air dry it at RT.

Wash with PBS pH 7.4 3 times in a chamber.

Air dry the slide.

Add one drop staining solution and cover with cover glass, keep in the dark.

## 6.3.Cloning

All gene fragments were amplified using primers listed and Phusion® High-Fidelity DNA Polymerase or Q5® High-Fidelity DNA Polymerase and cloned into pGEM T easy vector or L4440 vector.

PCR products purification was performed always.

Gibson assembly

| | Volume |
|---|---|
| Vector | 50-100 ng |
| Insert | 5-fold molar excess of vector* |
| GA mix | 5 μL |
| $H_2O$ | To 10 μL |

**: calculate with NEBioCalculator (https://nebiocalculator.neb.com/#!/ligation)*

50 ℃ 30 min

65 ℃ 20 min

Immediately use or keep at 4 ℃ for short time.

Dialysis the ligation product for 20-40 min.

Mix with 50 μL competent cell (10 μL cell mix with 190 μL dd$H_2O$), incubate on ice for 30 min.

Electroporation and resuspend in 500 μL TB, shaking at 37 ℃ before plating on selective medium.

**6.4.Silencing specific gene by dsRNAs feeding**

Protocol adapted from[4,149].

Bacterial pre-culture: Incubate *E.coli* feeding bacteria (strain HT115) accommodating related DNA sequence in 5 mL LB selecting medium with 0.1 mg/mL ampicillin and 0.0125 mg/mL tetracycline O/N shaking at 37 ℃.

Dilute the pre-culture ~1:100 in WGP (900ul pre-culture in 90ml WGP) containing 0.1mg/ml ampicillin and incubate O/N shaking at 37 ℃ until the $OD_{600}$ > 0.16.

Dilute the culture to $OD_{600}$= 0.04 with WGP containing 0.1mg/ml ampicillin to 200ml, shaking at 37 ℃ until $OD_{600}$> 0.07 (around 80min).

Add IPTG to a final concentration of 0.4 mM, incubate at least 4 hours or O/N shaking at 37 ℃

Cooling the medium to 27 °C by water bath and supply 0.8 mg/l of β-sitosterol.

Add mature *P. tetraurelia* cells to the culture at a concentration of 200 cells/mL.

Keep the culture at 27 °C until specific time point.

### 6.5.Survival test

Prepare depression slides with 600 μL 0.2× WGP containing0.8 mg/l β-sitosterol in each compartment.

Pick single cell into each compartment under microscope.

Monitor the growth of the cells with marked as healthy (normal dividing), weak (dividing slowly or not dividing) or dead for at least 3 days (27 °C).

### 6.6.Genomic DNA extraction of *P. tetraurelia* cells

100 mL of post-autogamous or vegetative cells was collected and washed as previously described. Genomic DNA was extracted using the GenElute Mammalian Genomic DNA MiniPrep Kit (Sigma).

### 6.7.IES retention PCR

Chosen IESs were checked for retention on knockdown of specific genes. GoTaq G2 DNA polymerase (Promega) and standard primers (primers list) were used. A total 12.5 ng of genomic DNA serve as input of each PCR.

### 6.8.Microinjection linearized DNA fragments into the Macronucleus

Protocol adapted from[4].

DNA preparation:

Plasmid DNA form 100 mL culture was extracted using the QIAGEN Plasmid Plus Midi kit.

Linearize 100 μg of plasmid by AhdI or SapI.

Equal volume of Phenol : chloroform was added into the digestion product.

Vortex for 50sec, spin down at max. speed for 5min at 4 ℃, DNA will be separated into the upper layer.

Carefully transfer only upper layer into the Ultra-free-MC-GV centrifugal filters (Millipore) and spin down at max. speed for 4 min.

Add 1/10 volume of 3 M NaOAc (pH 5.2) and 2.5 volume of ice-cold pure ethanol, flick to mix, incubate for at least 2 hours to O/N at -20 ℃.

Centrifuge at max. speed for 10 min at 4 °C and wash DNA pellet with 1 mL ice-cold 75% ethanol.

Centrifuge at max. speed for 5min at 4 °C once more, discard ethanol as much as possible.

Cover the tube with tinfoil and air dry the pellet at 37 °C.

Resuspend in 5 μL nuclease free water, measure the concentration and adjust it to 5.5 μg/μL

Cells preparation

Young cells (6-8 generations) recovered from post autogamous wildtype cells were transferred into slides containing 600 μL 2mg/mL Volvic-BSA solution.

Washing cells by transferring them to another slide containing Volvic-BSA solution.

Make of Volvic-BSA droplets in immersion oil and pick one cell into each droplet under microscope.

Injection:

Remove the Volvic-BSA solution to immobilize the cell in the droplet.

Inject prepared DNA with injection needle into the MAC and add the Volvic-BSA solution back to the droplet.

Transfer all cells into 600 μL 0.2× WGP media to wash the DNA outside of the cells.

Transfer the injected cells back into slides containing 0.2× WGP media separately and culture at 27℃.

### 6.9.Dot Blot

For the positive control, add 10pg, 50pg, 100pg of injected plasmid into 400 μL dd$H_2O$ in 1.5 mL tubes.

Transfer 400 μL of injected culture (cell density ~1,000 cells/mL) into 1.5mL Eppendorf tube.

Add 50 μL of 0.5 M EDTA (pH 8) and 50 μL of 4 M NaOH to samples including control.

*Cells can be stored at -80 ℃ after EDTA adding, in that case, mix NaOH before using.*

Incubate all samples at 68 ℃ for 30min.

*In this 30min, prepare the Dot Blot equipment, whatman paper and Hybond-XL membrane (GE Healthcare Life Science). Add 20 mL Church buffer into the incubating bottle.*

Briefly soaking membrane and paper in 0.4M NaOH before using, set the equipment.

Spin down all samples at max. speed for 1 min, chill them on ice.

Run all samples through the membrane by vacuum.

Incubate the membrane in 0.4 M NaOH on shaker for 15 min.

Wash the membrane with 2× SSC buffer for 15 min by shaking.

Dry the membrane with whatman paper, put it into the bottle containing Church buffer, incubate at 60 ℃ for 2 hours.

Heat the radioactive probe at 100 ℃ for 10 min.

Discard the Church buffer and incubate the membrane with radioactive probe at 55 ℃ for O/N.

The radioactive probe can be re-used, wash the membrane twice with 2× SSC containing 0.1% SDS at 55 ℃ for 10-15 min.

Dry the membrane carefully, wrap in saran foil.

Expose the membrane on the IP screen for more than 2 hours.

Labeling radioactive probes by Klenow enzyme

Order [α-$P^{32}$] ATP, 2 Mbq = 5 μl.

DNA preparation:

DNA fragments used for labelling with GoTag G2 DNA Polymerase (Promega) were amplified by PCR.

Purify the PCR products, dilute to 25 ng/μL.

Mix 25 ng DNA with 20 μL dd$H_2O$, denature by boiling for 5 min.

Put samples on ice immediately.

Add: 1 μL 500 μM TTP

1 μL 500 μM CTP

1 μL 500 μM GTP

20 μL 2.5X RadPrime buffer

Add 5 μL (~2 MBq [α-$P^{32}$] ATP) and 1 μL Klenow enzyme in C-lab.

Mix carefully and quick spin down 15-30 sec.

Incubate at 37 °C for 10 min.

Add 5 μL Stop buffer to stop the reaction.

Add 10 mL Church buffer and store at -20 °C.

### 6.10. Macronuclear DNA isolation

Protocol adapted from[21].

Day 1

Pre-cool the centrifuge, put all buffers on ice.

Collect and wash the cell as previously described.

Resuspend the pellet in 2.5 volumes of lysis buffer 1, transfer into pre-cold homogenizer on ice. Incubate 5 min.

Lyse mixture on ice by homogenizing. Check with DAPI frequently (10μL cells + 1μL DAPI). Stop cracking until only intact MACs are visible by DAPI staining.

Transfer all to a 15 mL tube. Mix with 14 mL of pre-cold wash buffer by inverting.

Spin down at 4 ℃, 1000RCF for 3 min, discard supernatant.

Resuspend the pellet with 2 mL wash buffer using cut-tip pipette, repeat the wash step above.

Resuspend the pellet in 3 volumes of pre-cold sucrose buffer. The volume should not over 2mL. Avoid DNA shearing.

Transfer 3 mL sucrose buffer to ultracentrifugation tube. Add the lysate on top of sucrose buffer. Add wash buffer on the top of lysate.

Centrifuge for 1 hour at 4 °C, 35,000 RPM, rotor: SW40Ti.

Discard supernatant thoroughly.

Resuspend the MAC pellet in 500 μL of 10 mM Tris pH 7.4, 10 mM $MgCl_2$.

Transfer all into a 15 mL tube, mix with 3 volumes of lysis Buffer 2.

Incubate O/N at 55˚C.

Day 2:

Phenol extraction: add 1 volume phenol : chloroform : IAA (25:24:1) to the sample.

Shake the sample a few times and spin down at max. speed for 10 min at 4 ℃.

Pre-soak the Slide-a-Lyzer Dialysis cassette in ice-cold dialysis buffer (10 mM Tris-HCl pH 8, 1 mM EDTA pH 8).

Transfer aqueous products into the cassette with a syringe through the cap.

*Remove the excess air after loading the sample by pressing the membrane.*

Dialysis at 4 ℃ in a large beaker by slowly rotating using a magnetic stirrer.

Replace buffer after 2 and 4 hours, then incubate O/N.

Day 3:

Recover the DNA from the cassette with a syringe.

Concentrate the DNA using the Amicon Ultra 2 mL filters.

Fill the device with sample and centrifuge at 4,000 RPM for 10 min.

Invert the device, centrifuge at 1,000 RPM for 2 min to harvest concentrated DNA.

Measure DNA concentration. Load 1 μg on gel to check the DNA integrity.

Buffers

| Lysis buffer 1 | Amount | Final concentration |
|---|---|---|
| $H_2O$ | 4.915 mL | |
| Sucrose | 0.43 g | 0.25 M |
| 2 M $MgCl_2$ | 25 μL | 10 mM |
| 1 M Tris pH 6.8 | 50 μL | 10 mM |
| NP-40 | 10 μL | 0.2% |

| Wash buffer | Amount | Final concentration |
|---|---|---|
| $H_2O$ | 59.1 mL | |
| Sucrose | 5.16 g | 0.25 M |
| 2 M $MgCl_2$ | 300 μL | 10 mM |
| 1 M Tris pH 7.4 | 600 μL | 10 mM |

| Sucrose buffer | Amount | Final concentration |
|---|---|---|
| $H_2O$ | 9.85 mL | |
| Sucrose | 7.18 g | 2.1 M |
| 2 M $MgCl_2$ | 50 µL | 10 mM |
| 1 M Tris pH 7.4 | 100 µL | 10 mM |

| Lysis buffer 2 | Amount | Final concentration |
|---|---|---|
| 0.5 M EDTA | 4.25 mL | ~0.5 M |
| 10% SDS | 500 µL | 1% |
| N-lauryl sarcosine | 0.05 g | 1% |
| 20 mg/mL proteinase K | 250 µL | 1 mg/mL |

| Dialysis buffer | Amount | Final concentration |
|---|---|---|
| 1 M Tris pH 8 | 10 mL | 10 mM |
| 0.5 M EDTA | 2 mL | 1 mM |
| $H_2O$ | To 1 L | |

## 6.11. Total RNA extraction with TRI

Frozen collected cell pellet in liquid nitrogen immediately by drops and stored at -80 ℃.

Add 6 mL of TRI to the frozen pellet on dry ice.

Vortex the pellet until completely dissolved one by one.

Incubate 5min at RT on rotator.

Add 1.2 mL chloroform and vortex for 15 sec.

Stand 5 min at RT.

Centrifuge at max. speed for 15 min at 4 ℃.

Carefully transfer the top phase into a 15 mL falcon tube, make aliquots of 750 µL in RNase free Eppendorf tubes.

Add 500 µL isopropanol to each aliquot, mix by inverting and stand at RT for 10 min.

Centrifuge at 12,000 RCF, 4 °C for 10 min to pellet the RNA.

Remove the supernatant, resuspend with 1 mL 75% ethanol.

Centrifuge at 7,500 RCF, 4 °C for 5 min.

Repeat wash with another 1 mL 75% ethanol and centrifuge.

Remove supernatant thoroughly.

Air dry the RNA pellet at 37 °C covered with a tinfoil tent.

Add 30 µL nuclease free dd$H_2O$ into the RNA pellet, incubate at 60 °C for 15 min to dissolve RNA.

## 6.12. Analysis of sRNA in denaturing gel

Gel preparation

Add 5 mL of autoclaved water into an autoclaved 100 ml screw cap bottle (Type A is best for stirring)

Add 16 ml Acrylamide (30% Acrylamide/ Bis Solution 19:1) with sterile pipette

Add 12.6 g Urea (pellet not powder)

Add 3 mL 10× TBE

Dissolve it with a magnetic stir bar (approx. 1 h)

16% polyacrylamide-7M Urea gel

| | |
|---|---|
| 30% Acrylamide/ Bis Solution 19:1 | 16 mL |
| Urea-7M (60.1 g/ mol) | 12.6 g |
| 10× TBE | 3 ml |
| $ddH_2O$ autoclaved | Up to 30 mL |

Chamber preparation

Wash entire equipment with hot water, $ddH_2O$ and 100% ethanol carefully.

Air dry all parts before assemble the system.

Pour the gel

Pass gel solution through a 0.22 μm filter into a autoclaved 100 ml screw cap bottle with a 50 mL syringe.

Chill on ice over 5 min.

Add 150 μL 10% APS and 30 μL TEMED, mix gently.

Chill on ice over 5 min.

Pour the gel with 50 mL syringe, avoid bubbles, insert the comb in immediately.

Let the gel polymerize for 1-2 hours.

RNA 5' end labelling

| Master mix | 1 reaction |
|---|---|
| 10× Buffer B (exchange buffer) | 0.6 µL |
| PEG 6000 (24%) | 1.2 µL |
| RNase inhibitor | 0.6 µL |
| $\gamma$-$P^{32}$-ATP (0.4 Mbq, add in C-lab) | 1 µL |
| PNK | 0.6 µL |

Add 4 µL of master mix to each 2 µL RNA sample in C-lab.

Incubate at 37 ℃ for 30 min.

Mix 12 µL RNA loading dye (100 µL 6× DNA loading dye + 900 µL formamide) with each sample.

Heat samples to 80 ℃ for 10 min, put on ice afterwards.

Add 0.5× TBE buffer to the assembled chamber.

Remove combs and clean slots with 0.5× TBE buffer (use syringe).

Pre-run for more than 30 min at 25 mA (fix).

Wash slots again before loading.

Load the samples and ladder, better separate samples by loading dye. Fill empty slots with loading dye.

Run the gel at 25 mA (fix) until the blue is at half of the gel (30-40 min).

Disassemble the frame and cover the gel (still attached to one of the plates) with saran foil.

Put phosphor screen on top and expose for 1-2 hours or O/N.

### 6.13. Co-immunoprecipitation

Adapted from[150].

Collect and wash the cells as previously described. Transfer the cell pellet (~0.5 mL) into 15 mL Falcon tubes.

Non-crosslinking: go the wash step or put samples at 4 °C before washing.

Cross linking:

Add 1 mL of 1% PFA and shake for 10 minutes at RT.

Add 100 μL of 1.25 M Glycine, incubate for 5 min to quench on shaker at RT.

Wash samples twice with PBS (pH 7) at 279 RCF, 4 °C for 2 min.

Gently resuspend with a few mL of PBS, then fill it up and invert the tube to mix.

Discard supernatant. Add 2 mL of ice-cold lysis buffer to the pellet and resuspend thoroughly.

Sonicate with 30 % amplitude, 10 sec on, 1min rest, 15 cycles.

*Stain 10 μL product with DAPI, check under microscope to make sure the sonication is efficient.*

Transfer the sample to 2 ml tubes and centrifuge at 13, 000 RCF, 4 °C for 30 min.

Transfer the supernatant to two new RNase free 1.5 ml Eppendorf tubes (1 ml to each).

Freeze all components in liquid nitrogen, store at -80 °C. Keep one tube on ice when perform further enrichment immediately.

Transfer 50 μL beads mixture of each IP sample into nuclease free tubes.

Add 1 mL IP buffer to the beads, mix gently by inverting.

Centrifuge at 500 RCF, 4 °C for 2 min.

Remove supernatant gently. Repeat wash twice.

Add 1 mL of the lysate to cleaned beads, incubate at 4 °C O/N on rotating.

Centrifuge at 500 RCF, 4 °C for 2 min.

Transfer the supernatant as unbound in a nuclease free tube on ice for further analysis.

Mix gently with 1 ml of IP buffer by inverting, do not pipetting.

Repeat this wash for another 4 times.

Resuspend beads gently in 1 mL of ice-cold IP buffer, transfer all to another nuclease free tube.

Centrifuge at 500 RCF, 4 °C for 2 min, discard supernatant (leave about 50 μl).

*For PGM included samples, first add equal volume of 10% SDS and boil at 100 °C for 3 min. Add 12.5 μL 5× loading dye and boil at 95 °C for 10-20 min. store at -20 °C.*

Buffers

| Lysis buffer | 10 mL | Final concentration |
|---|---|---|
| 1 M Tris HCl pH 8.8 *(modify accordingly)* | 500 μL | 50 mM |
| 2 M NaCl | 750 μL | 150 mM |
| 2 M $MgCl_2$ | 25 μL | 5 mM |
| 100% Triton X-100 | 100 μL | 1% |
| 50× protease inhibitor *(EDTA free)* | 200 μL | 1× |
| 80% Glycerol | 1.25 mL | 10% |
| DTT | 10 μL | 1 mM |
| Milli-Q water | 7.165 mL | |

| IP buffer (W/O NP-40) | 200 mL | Final concentration |
|---|---|---|
| 1 M Tris HCl pH 8.8 *(modify accordingly)* | 2 mL | 10 mM |
| 2 M NaCl | 15 mL | 150 mM |
| 2 M $MgCl_2$ | 100 μL | 1 mM |
| 80% Glycerol | 12.5 mL | 5% |
| Milli-Q water | 170.4 mL | |

Store at RT.

*Add 1 μL NP-40 to 10 mL IP buffer to make the complete IP buffer before usage.*

### 6.14. SDS-PAGE

Gel preparation:

Clean glass plates with hot water and pure ethanol, air dry on tissue paper.

Prepare the gel according to the Protocol (Modified from Harlow and Lane 1988).

**TABLE A8-9** Solutions for Preparing Resolving Gels for Tris-glycine SDS-Polyacrylamide Gel Electrophoresis

| ⇓ Components / Gel Volume ⇒ | Volume (ml) of Components Required to Cast Gels of Indicated Volumes and Concentrations | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 5 ml | 10 ml | 15 ml | 20 ml | 25 ml | 30 ml | 40 ml | 50 ml |
| **6% gel** | | | | | | | | |
| $H_2O$ | 2.6 | 5.3 | 7.9 | 10.6 | 13.2 | 15.9 | 21.2 | 26.5 |
| 30% acrylamide mix <!> | 1.0 | 2.0 | 3.0 | 4.0 | 5.0 | 6.0 | 8.0 | 10.0 |
| 1.5 M Tris (pH 8.8) | 1.3 | 2.5 | 3.8 | 5.0 | 6.3 | 7.5 | 10.0 | 12.5 |
| 10% SDS | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 |
| 10% ammonium persulfate <!> | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 |
| TEMED <!> | 0.004 | 0.008 | 0.012 | 0.016 | 0.02 | 0.024 | 0.032 | 0.04 |
| **8% gel** | | | | | | | | |
| $H_2O$ | 2.3 | 4.6 | 6.9 | 9.3 | 11.5 | 13.9 | 18.5 | 23.2 |
| 30% acrylamide mix <!> | 1.3 | 2.7 | 4.0 | 5.3 | 6.7 | 8.0 | 10.7 | 13.3 |
| 1.5 M Tris (pH 8.8) | 1.3 | 2.5 | 3.8 | 5.0 | 6.3 | 7.5 | 10.0 | 12.5 |
| 10% SDS | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 |
| 10% ammonium persulfate <!> | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 |
| TEMED <!> | 0.003 | 0.006 | 0.009 | 0.012 | 0.015 | 0.018 | 0.024 | 0.03 |
| **10% gel** | | | | | | | | |
| $H_2O$ | 1.9 | 4.0 | 5.9 | 7.9 | 9.9 | 11.9 | 15.9 | 19.8 |
| 30% acrylamide mix <!> | 1.7 | 3.3 | 5.0 | 6.7 | 8.3 | 10.0 | 13.3 | 16.7 |
| 1.5 M Tris (pH 8.8) | 1.3 | 2.5 | 3.8 | 5.0 | 6.3 | 7.5 | 10.0 | 12.5 |
| 10% SDS | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 |
| 10% ammonium persulfate <!> | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 |
| TEMED <!> | 0.002 | 0.004 | 0.006 | 0.008 | 0.01 | 0.012 | 0.016 | 0.02 |
| **12% gel** | | | | | | | | |
| $H_2O$ | 1.6 | 3.3 | 4.9 | 6.6 | 8.2 | 9.9 | 13.2 | 16.5 |
| 30% acrylamide mix <!> | 2.0 | 4.0 | 6.0 | 8.0 | 10.0 | 12.0 | 16.0 | 20.0 |
| 1.5 M Tris (pH 8.8) | 1.3 | 2.5 | 3.8 | 5.0 | 6.3 | 7.5 | 10.0 | 12.5 |
| 10% SDS | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 |
| 10% ammonium persulfate <!> | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 |
| TEMED <!> | 0.002 | 0.004 | 0.006 | 0.008 | 0.01 | 0.012 | 0.016 | 0.02 |
| **15% gel** | | | | | | | | |
| $H_2O$ | 1.1 | 2.3 | 3.4 | 4.6 | 5.7 | 6.9 | 9.2 | 11.5 |
| 30% acrylamide mix <!> | 2.5 | 5.0 | 7.5 | 10.0 | 12.5 | 15.0 | 20.0 | 25.0 |
| 1.5 M Tris (pH 8.8) | 1.3 | 2.5 | 3.8 | 5.0 | 6.3 | 7.5 | 10.0 | 12.5 |
| 10% SDS | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 |
| 10% ammonium persulfate <!> | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 |
| TEMED <!> | 0.002 | 0.004 | 0.006 | 0.008 | 0.01 | 0.012 | 0.016 | 0.02 |

Modified from Harlow and Lane (1988).

**TABLE A8-10** Solutions for Preparing 5% Stacking Gels for Tris-glycine SDS-polyacrylamide Gel Electrophoresis

| ⇓ Components / Gel Volume ⇒ | Volume (ml) of Components Required to Cast Gels of Indicated Volumes | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 ml | 2 ml | 3 ml | 4 ml | 5 ml | 6 ml | 8 ml | 10 ml |
| $H_2O$ | 0.68 | 1.4 | 2.1 | 2.7 | 3.4 | 4.1 | 5.5 | 6.8 |
| 30% acrylamide mix <!> | 0.17 | 0.33 | 0.5 | 0.67 | 0.83 | 1.0 | 1.3 | 1.7 |
| 1.0 M Tris (pH 6.8) | 0.13 | 0.25 | 0.38 | 0.5 | 0.63 | 0.75 | 1.0 | 1.25 |
| 10% SDS | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.08 | 0.1 |
| 10% ammonium persulfate <!> | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.08 | 0.1 |
| TEMED <!> | 0.001 | 0.002 | 0.003 | 0.004 | 0.005 | 0.006 | 0.008 | 0.01 |

Modified from Harlow and Lane (1988).

Gel running:

Run at 80 V (~30 min), switch the voltage to 110V as soon as the dye reaches the resolving gel.

Buffers

| 10× SDS running buffer | 1 L | Final concentration |
|---|---|---|
| Tris Base | 30.2 g | 0.25M |
| Glycine | 144 g | 1.92 M |
| SDS | 10 g | 1% |
| $ddH_2O$ | To 1 L | |

*Dilute 1:10 before usage.*

| 5× SDS loading buffer | 10 mL |
|---|---|
| 10% SDS | 2 mL |
| 0.5 M Tris-HCl pH 6.8 | 1.2 mL |
| 50% Glycerol | 4.8 mL |
| Bromphenol blue | Tiny |
| β-Mecaptoethanol | 500 µL |
| ddH2O | To 10 mL |

*Aliquot to 1 mL/tube, store at -20 °C.*

## 6.15. Silver staining

Fix the gel from SDS-PAGE in 100 mL Fixative solution at RT O/N.

Wash with 50% ethanol 20 min twice.

Treat the gel for exactly 1 min with 98 mL Sodium Thiosulphate (3-4 flakes in 100 mL Milli-Q water).

Wash with $ddH_2O$ 20 sec for 3 times.

Treat the gel with $AgNO_3$ solution for 30 min. ($AgNO_3$ can be reused)

Wash with $ddH_2O$ 20 sec for 3 times.

Develop the gel with 100 mL Developing solution until bands clear.

Stop the reaction with 100 mL Stop solution and scan the gel.

Solutions

| Fixative solution *(mix in order)* | 100 mL | Final concentration |
|---|---|---|
| Milli-Q water | 39.95 mL | |
| Methanol | 50 mL | 50% |
| Acetic Acid | 10 mL | 10% |
| Formaldehyde | 50 µL | |

| Developing solution | 100 mL |
|---|---|
| $Na_2CO3$ | 6 g |
| Sodium Thiosulphate solution | 2 mL |
| Formaldehyde | 50 µL |
| Milli-Q water | To 100 mL |

| Stop solution | 100 mL |
|---|---|
| Milli-Q water | 95 mL |
| Acetic acid | 5 mL |

### 6.16. Western blot

Run the SDS-PAGE as usual.

Wash gel, resins the gel, sponges, whatman paper and membrane in ice-cold working transfer buffer.

Assemble the cassette from the black side


*Using 15 mL tube to remove the bubbles.*

Setup the cassette in correct direction in the chamber.

Put an ice package into the chamber, fill it up with ice-cold working transfer buffer. Put the chamber in ice bath if necessary

Transfer for at least 2 hours under 110 V. Time and voltage can be adjusted according to different proteins.

Disassemble the equipment, block the membrane with PBST containing 5% milk for 30 min-1 h at RT.

Incubate with primary Antibody (HA-Tag (C29F4) Rabbit mAb #3724) 1:1000 in blocking solution 3 - 4 h at RT or O/N at 4 °C.

Wash 3× 5 min with PBST at RT.

Incubate secondary antibody (1 μL in 10 mL blocking solution).

Shaking at RT for 1 hour.

Wash 3× 5 min with PBST at RT.

Develop the membrane with Luminata Crescendo Western HRP Substrate (Millipore), expose membrane with Amersham Imager 600.

### 6.17. N-6× His-SUMO-PGM Expression

Cloning the N-6× His-SUMO-PGM plasmid into BL21 strain cells.

Pick up single positive clone into 15 mL LB culture containing 15 μL kanamycin, incubate at 37 °C, O/N.

Dilute the pre-cultural to 1 : 100 with LB, supplying with 1 : 1000 kanamycin.

Incubate the dilution at 37 °C on shaker until the $OD_{600}$ is 0.4.

Add IPTG to final concentration 0.2 mM, incubate over 6 hours.

Collect all culture by centrifuge at 4200 RPM, 4 °C for 1 hour.

Remove the supernatant, resuspend the pellet with 0.9% NaCl.

Centrifuge at max. speed, 4 °C for 15 min.

Frozen the pellet in liquid nitrogen, store in -80 °C freezer.

### 6.18. N-6× His-SUMO-PGM purification

Add 25 ml lysis buffer into the pellet from expression, keep on ice 30min, and dissolve by pipetting.

Sonication: 30% amplitude, 10 sec robust, 1 min rest, total robust 6 min until transparent.

Take 100 μL lysate mix with 100 μL 10% SDS, boil at 100 °C for 3 min.

Add 50 μL 5× loading buffer, boil at 95 °C for 10min (lysate).

Wash the glass tube with hot water, dd$H_2O$, and 100% ethanol before usage.

Transfer lysate after sonication into two 30ml glass tubes, 15ml each, balance by scale. Fit tube into rubber tube, spin down lysates at 11000 RPM (~14000 ×g), 4 °C for 45min.

Well mix the beads (50% of mixture), transfer 2 mL beads mixture into a 50 mL falcon tube.

Wash beads with 20 mL wash buffer and spin down 5min at 500 ×g, 4°C, 3times.

Take 100 μL supernatant and pellet which resuspended with same volume of input, mix with 100 μL 10% SDS, boil at 100 °C for 3 min separately.

Add 50 μL 5× loading buffer, boil at 95 °C for 10min (supernatant and pellet).

Add all supernatant (~30 mL) from above to gently resuspend beads, incubate in 4 °C room for 30 min-1 hour to bind PGM by rotating.

Transfer all mixture into a 20 mL Chromatography Columns, collect flow through as unbound.

Wash the column three times with 25 mL washing buffer, collect the first wash for western blot (Wash).

Aliquot SUMO protease 4ul/tube keep in -20 °C.

Resuspend beads in column with 1 mL working buffer (2 mL in total), transfer to 2 mL tube and incubate in 4 °C room on rotating for 40 min or O/N.

Wash the Chromatography Columns with 20 mL dd$H_2O$ 10times, keep column filled with dd$H_2O$ at 4 °C for reusing.

Collect the flow through which is the protein needed. Take 6 μL in PCR tube, measure concentration with Nanodrop Protein 280 (standard BSA control 1 mg/mL).

Take 100 μL protein and beads which diluted into same volume with input, mix with 100 μL 10% SDS, boil at 100 °C for 3 min separately.

Add 50 μL 5× loading buffer, boil at 95 °C for 10min (elution and beads).

Aliquot protein in 50-100 μL/PCR tube. Two tubes should be free of glycerol, frozen in liquid nitrogen immediately. Add equal 80% glycerol into other tubes, frozen in liquid nitrogen, store according further requirements.

Buffers

| Lysis buffer | 50 mL | Final concentration |
|---|---|---|
| 1M Tris HCl pH 9 | 2.5 mL | 50 mM |
| 5M NaCl | 1.5 mL | 150 mM |
| 80% Glycerol | 6.25 mL | 10% |
| β- mercaptoethanol | 35 μL | 10 mM |
| 50× protease inhibitor EDTA-free | 1 mL | 1× |
| Lysozyme | 50 mg | 1 mg/mL |
| dd$H_2O$ | 38.715 | |

| Wash buffer | 300 mL | Final concentration |
|---|---|---|
| 1M Tris HCl pH 9 | 15 mL | 50 mM |
| 5M NaCl | 30 mL | 500 mM |
| 80% Glycerol | 37.5 mL | 10% |
| DTT | 300 µL | 1 mM |
| Imidazole | 408.48 mg | 20 mM |
| dd$H_2O$ | 217.29 mL | |

| Working buffer | 2 mL | Final concentration |
|---|---|---|
| 1M Tris HCl pH 8 | 40 µL | 20 mM |
| 5M NaCl | 200 µL | 500 mM |
| SUMO protease | 4 µL | |
| dd$H_2O$ | 1756 µL | |

### 6.19. Denaturing urea polyacrylamide gel electrophoresis

Dissolve the gel solution at 37 °C water bath in a 50 mL tube, shake frequently until complete dissolve.

During this time, clean SDS-PAGE equipment with hot water, dd$H_2O$ and 100% ethanol, ary dry them.

Filter the solution into a new 50 mL falcon tube through 0.22 μm.

Chill the solution on ice.

Add 75 μL 10% APS and 15 μL TEMED, mix gently.

Chill on ice for another 5min.

Pour solution into the gap and insert the comb, lay the equipment on a tip box and put a bottle on the top of cover glass to remove the gap between comb and gel.

Using 200 μL pipette to clean the hole.

Pre-running at 150 V, 20 min.

Mix 5 μL DNA or RNA with 10 μL loading dye.

Boil at 95 °C 2 min, immediately put them on ice for at least 5 min before loading.

Fill the equipment with 1× TBE solution.

Load all samples, fill the gap with dye, run at 150 V for about 1 hour.

Ethidium bromide (EB) staining and imaging.

| | |
|---|---|
| Gel solution | 15 mL |
| 10 TBE | 1.5 mL |
| 40% acrylamide 19:1 | 3 mL |
| Urea | 6.3 g |
| ddH2O | 3.5 mL |

## 6.20. Primer List

| Oligo name | Sequence 5'-3' |
|---|---|
| SMC4-1 +f R | TCAACATGAATAATTAATTTTAAATTCTTCATTCTT |
| SMC4-1 +f F | AAAGAATTAACTAGTATTTTGCAGAAAAAATATTAAG |
| SMC4-1 +fv 3 F | TGAAATGATTTTAAAATATTCATATTCCATTATTG |
| SMC4-1 +fv 3 R | TCTATTTTAATTAACGTTTTCATTGTTTTC |
| SMC4-1 +fv F | CAGACTGAAAGTGAAGAGCCAT |
| SMC4-1 +fv R | CATTTTATGTAAATATATTTAATATTTTTAAATTAACTTTTGG |
| SMC4-2 +F R | GAATTAATTCAAGATCTAAATATTCGTAACA |
| SMC4-2 +F F | ATGCGCAATTTAATTTGAGCAATATT |
| SMC4-2 +fv 3 F | TGAAATTTATATTAAATTATAATCAATTAATTTTATC |
| SMC4-2 +fv 3 R | AACTTCTAATTAGATGACTTCTGTTGAA |
| SMC4-2 +fv F | ATTAAGGAAGTTATTTTGGAGAATTTTAAATC |
| SMC4-2 +fv R | CATATTTTCTATCATTTAATTTTCTACTCAAA |
| SMC2-1 sil F | ATATGAGCTCATGTGGATCAAAGAAATCATTATCGA |
| SMC2-1 sil R | TACCAGGAGACAAAAAGAAACTATAGACTCGAGTAAA |
| SMC2-2 sil F | ATATGAGCTCAGAACAACTGAATAGAGAGATTACACA |
| SMC2-2 sil R | AAAAGAATTCTAATCATTAAAAGACAAAAAGTTGCTCGAGTAAA |
| SMC4-1 sil F | ATATGAGCTCATGCAGACTGAAAGTGAAGAGC |
| SMC4-1 sil R | TTTACTCGAGTGTTTTGTTATGATGTCAGATTGTTCACT |
| SMC4-2 sil F | ATATGAGCTCATGATTAAGGAAGTTATTTTGGAGAATTTT |
| SMC4-2 sil R | TTTACTCGAGTATCTACCTTATTTTATTTTTCTGGAATCA |
| SMC2-1 Co-sil F | ATATCCCGGGATGTGGATCAAAGAAATCATTATCGA |
| SMC2-2 Co-sil F | ATATCCCGGGAGAACAACTGAATAGAGAGATTACACA |
| SMC4-1 Co-sil F | ATATCCCGGGATGCAGACTGAAAGTGAAGAGC |
| SMC4-2 Co-sil-R | TTTACCCGGGTATCTACCTTATTTTATTTTTCTGGAATCA |
| 377 F | AAAGATAAATATGAAATTTGTTGGTGAATTTTAAT |
| 377 R | AAATTAACTTGGATTATGAATTCAAATATAAAT |
| 798 F | AAAATTAAGTTTCTGGATAATATATGATATGGTTA |
| 798 R | AAATAACCTATCCAAAATAACCTAACAGAA |

| IES | | Retention score | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Classical IES name | IES name in reference MAC+ IES assembly | *Maternal control score* | *EV* | *PGM*-KD | *CAF1*-KD | *SDCP*-KD | *NOWA1*-KD | *DCL2/3*-KD | *DCL5*-KD | *DCL/2/3/5*-KD | *PT01/09*-KD | *PT10/11*-KD |
| 51G11 | IESPGM.PTET51.1.51.451201 | | 0.03 | **0.80** | **0.5** | **0.44** | **0.53** | 0.05 | **0.20** | **0.56** | 0.02 | **0.20** |
| 51G1413 | IESPGM.PTET51.1.51.452624 | 0.00 | 0.01 | **0.78** | **0.35** | 0.12 | 0.1 | 0.01 | 0.02 | 0.04 | 0.00 | 0.00 |
| 51G1832 | IESPGM.PTET51.1.51.453043 | 0.00 | 0.00 | **0.73** | 0.04 | 0.04 | 0.02 | 0.00 | 0.00 | 0.02 | 0.01 | 0.01 |
| 51G2832 | IESPGM.PTET51.1.51.454043 | **0.87** | 0.00 | **0.77** | **0.6** | 0.2 | **0.59** | **0.23** | 0.00 | **0.71** | **0.41** | 0.00 |
| 51G4404 | IESPGM.PTET51.1.51.455615 | **0.97** | 0.00 | **0.84** | **0.57** | **0.48** | **0.81** | **0.60** | 0.01 | **0.75** | **0.77** | 0.01 |
| 51G6447 | IESPGM.PTET51.1.51.457658 | | 0.00 | **0.78** | 0.05 | 0.09 | 0.05 | 0.04 | 0.00 | 0 | 0.01 | 0.01 |
| 51A712 | IESPGM.PTET51.1.106.281631 | **0.34** | 0.00 | **0.78** | **0.44** | **0.64** | **0.75** | 0.07 | 0.06 | **0.71** | **0.18** | 0.08 |
| 51A1835 | IESPGM.PTET51.1.106.284157 | 0.00 | 0.00 | **0.80** | 0.02 | 0.08 | 0.03 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 |
| 51A2591 | IESPGM.PTET51.1.106.284913 | **0.99** | 0.00 | **0.91** | **0.63** | **0.48** | **0.81** | **0.54** | 0.00 | **0.81** | **0.66** | 0.00 |
| 51A4404 | IESPGM.PTET51.1.106.286750 | 0.00 | 0.00 | **0.83** | 0 | **0.23** | 0 | 0.00 | 0.00 | 0 | 0.00 | 0.00 |
| 51A4578 | IESPGM.PTET51.1.106.286924 | 0.00 | 0.00 | **0.78** | **0.41** | **0.24** | **0.20** | 0.05 | 0.00 | 0.08 | 0.03 | 0.00 |
| 51A6435 | IESPGM.PTET51.1.106.288781 | 0.00 | 0.00 | **0.77** | 0.01 | 0.04 | 0.01 | 0.00 | 0.00 | 0 | 0.00 | 0.00 |
| 51A6649 | IESPGM.PTET51.1.106.288995 | **0.58** | 0.00 | **0.81** | **0.53** | **0.60** | **0.57** | **0.56** | 0.01 | **0.74** | **0.48** | 0.00 |
| Dcl5d-01 | IESPGM.PTET51.1.8.257314 | | 0.02 | **0.75** | **0.39** | 0.23 | **0.77** | 0.03 | **0.50** | **0.73** | 0.07 | **0.25** |
| Dcl5d-02 | IESPGM.PTET51.1.16.506105 | | 0.00 | **0.95** | **0.61** | **0.65** | **0.47** | 0.02 | **0.38** | **0.73** | 0.05 | **0.51** |
| Dcl5d-03 | IESPGM.PTET51.1.39.60104 | | 0.00 | **0.74** | 0.13 | 0.09 | **0.31** | 0.00 | **0.37** | 0.6 | 0.00 | **0.25** |
| Dcl5d-04 | IESPGM.PTET51.1.168.68297 | | 0.00 | **0.69** | 0.04 | 0.03 | 0.46 | 0.05 | **0.37** | **0.81** | 0.06 | **0.22** |

| IES | Primer | | Position | | Band Size | | Annealing Temperature |
|---|---|---|---|---|---|---|---|
| Classical IES name | Forward Primer | Reverse Primer | Scaffold | Start | IES+ | IES- | Tm |
| 51G11 | ATCATAAGATTGATATCTTCTCCCTTCTCC | ACTTGCTACTAAAGCAAGAAACATTGAGAG | 51 | 478642 | 296 | 253 | 54 |
| 51G1413 | GAAGCTGCTTGTGTTAAGAATTCTACTGG | GCATCCAGCACTAGTTGAATTTACTGTAC | 51 | 480108 | 220 | 168 | 62 |
| 51G1832 | CTATAACTCTTGAAGCTGCTTGTAATATG | TTGTCAATGAGCCATTAACAGTTGCTGGAT | 51 | 480579 | 217 | 187 | 62 |
| 51G2832 | GCTATAACTCTTGAAGCTGCTTGTAATATG | TTGTCAATGAGCCATTAACAGTTGCTGGAT | 51 | 481609 | 393 | 164 | 60 |
| 51G4404 | CTGTTGCTACACATTGTGCATATGTTACT | GCTGTAAGATTAACATTGAGCATGATCAAG | 51 | 483410 | 501 | 299 | 60 |
| 51G6447 | AATGCATCAAATGTAGTAACTACTCCTGCT | AATTTGTAAAGTATCCAGCGCAGGCAG | 51 | 485675 | 137 | 109 | 57 |
| 51A712 | TTTGTCAAAAAGACATGTATCAAAATGCAG | TAGAATACTAAGAGATTCAATACAACAAAC | 106 | 196699 | 234 | 157 | 57 |
| 51A1835 | TAATGTATTGATAAGGCTTGCTCTACAGCC | ATCTAACATCCTTGAATAGTTACTGATCC | 106 | 299405 | 166 | 138 | 60 |
| 51A2591 | ATGTGTTTGGACTGGATTGGCATGTAGAAG | GATGTAGCATAACATTTATCAACAATCCAT | 106 | 300189 | 639 | 269 | 60 |
| 51A4404 | TGGAATAGTGCTGCATCACCAGCTGCTTGC | CCAGTTATTGAACTGCAACTTACTGCAGTG | 106 | 302396 | 355 | 278 | 57 |
| 51A4578 | CCTGCAGTAAGTTGCAGTTCAATAACTGG | TGTAGTCTTAAAATCTTAGCATGTTGTACC | 106 | 302647 | 1007 | 125 | 57 |
| 51A6435 | CAAATTGTGTCACTAGAGGTACATGTTTCC | GCGACATCAATAGTAACAGCTGAGCATGAG | 106 | 305407 | 305 | 277 | 57 |
| 51A6649 | ACTGCACCTCTAACTTTAACAAGCGAAGCA | CAGCAGTACATCCAGCTCTCTAAGTTTAGC | 106 | 305649 | 628 | 258 | 57 |
| MT | GGTGTTTATATCTTAATTGTTGACCCTCAC | CCATCTATACTCCATTCTTTATCTTAATTCAT | | | 460 | 265 | 55 |
| Dcl5d-01 | CCAGTCTTTATAACTCCAAATATACTAATGTTAATTGCC | CTTGCTGGTTGAATATCAATTGAAAAATCTTGATG | 8 | 269999 | 296 | 249 | 55 |
| Dcl5d-02 | TTTTCATACTCATCCTCACCCTACTCCC | ATAATATAAACTTTGAAGCCCCTGAAGCCG | 16 | 525921 | 520 | 424 | 55 |
| Dcl5d-03 | TTCATCGTTCTCAAAATGGATGCCC | ATCAATATGAATTTTATTTGTTTGTTTATAAGCGTCTGG | 39 | 64383 | 235 | 163 | 55 |
| Dcl5d-04 | TTAAGGTACTTCTTCATCCATAGTCAGCC | AACAATATGAGTTAGAAATTTAAAATCTTGGAGGAATCC | 168 | 70859 | 278 | 220 | 55 |

# 7. References


1 Prescott, D. M. The DNA of ciliated protozoa. *Microbiological reviews* **58**, 233-267, doi:10.1128/mr.58.2.233-267.1994 (1994).

2 Gao, F. *et al.* Systematic studies on ciliates (Alveolata, Ciliophora) in China: Progress and achievements based on molecular information. *European journal of protistology* **61**, 409-423, doi:10.1016/j.ejop.2017.04.009 (2017).

3 Philippe, H., Germot, A. & Moreira, D. The new phylogeny of eukaryotes. *Current opinion in genetics & development* **10**, 596-601, doi:10.1016/s0959-437x(00)00137-4 (2000).

4 Beisson, J. *et al.* Paramecium tetraurelia: the renaissance of an early unicellular model. *Cold Spring Harbor protocols* **2010**, pdb emo140, doi:10.1101/pdb.emo140 (2010).

5 Cummings, D. J. Studies on Macronuclear DNA from Paramecium aurelia. *Chromosoma* **53**, 191-208, doi:10.1007/BF00329171 (1975).

6 Rodriguez-Ezpeleta, N. *et al.* Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Current biology : CB* **17**, 1420-1425, doi:10.1016/j.cub.2007.07.036 (2007).

7 Betermier, M. & Duharcourt, S. Programmed Rearrangement in Ciliates: Paramecium. *Microbiology spectrum* **2**, doi:10.1128/microbiolspec.MDNA3-0035-2014 (2014).

8 Katz, L. A. Evolution of nuclear dualism in ciliates: a reanalysis in light of recent molecular data. *International journal of systematic and evolutionary microbiology* **51**, 1587-1592, doi:10.1099/00207713-51-4-1587 (2001).

9 Hedgpeth, J. W. The Biology of Paramecium. *Science* **118**, 727-727, doi:doi:10.1126/science.118.3076.727.b (1953).

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

11 Wagtendonk, W. J. v. A Ciliate: The Anatomy of Paramecium aurelia *Science* **170**, 155-156, doi:doi:10.1126/science.170.3954.155.b (1970).

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

13 Dubois, E. *et al.* Transposon Invasion of the Paramecium Germline Genome Countered by a Domesticated PiggyBac Transposase and the NHEJ Pathway. *International journal of evolutionary biology* **2012**, 436196, doi:10.1155/2012/436196 (2012).

14 Terue Harumoto, B. Paramecium: Genetics and Epigenetics. *The Quarterly Review of Biology* **84**, 201-202, doi:10.1086/603492 (2009).

15 Morgens, D. W. & Cavalcanti, A. R. Amitotic chromosome loss predicts distinct patterns of senescence and non-senescence in ciliates. *Protist* **166**, 224-233, doi:10.1016/j.protis.2015.03.002 (2015).

16 Chalker, D. L., Meyer, E. & Mochizuki, K. Epigenetics of ciliates. *Cold Spring Harbor perspectives in biology* **5**, a017764, doi:10.1101/cshperspect.a017764 (2013).

17 Betermier, M. Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate Paramecium. *Research in microbiology* **155**, 399-408, doi:10.1016/j.resmic.2004.01.017 (2004).

18 Meyer, E. & Garnier, O. Non-Mendelian inheritance and homology-dependent effects in ciliates. *Advances in genetics* **46**, 305-337, doi:10.1016/s0065-2660(02)46011-7 (2002).

19 Kapusta, A. *et al.* Highly precise and developmentally programmed genome assembly in Paramecium requires ligase IV-dependent end joining. *PLoS genetics* **7**, e1002049, doi:10.1371/journal.pgen.1002049 (2011).

20 Baudry, C. *et al.* PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. *Genes & development* **23**, 2478-2483, doi:10.1101/gad.547309 (2009).

21 Arnaiz, O. *et al.* The Paramecium germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS genetics* **8**, e1002984, doi:10.1371/journal.pgen.1002984 (2012).

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

23 Steele, C J et al. “Developmentally excised sequences in micronuclear DNA of Paramecium.” Proceedings of the National Academy of Sciences of the United States of America vol. 91,6: 2255-9. doi:10.1073/pnas.91.6.2255 (1994)

24 Plasterk, R. H., Izsvak, Z. & Ivics, Z. Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends in genetics : TIG* **15**, 326-332, doi:10.1016/s0168-9525(99)01777-1 (1999).

25 Dubrana, K., Le Mouel, A. & Amar, L. Deletion endpoint allele-specificity in the developmentally regulated elimination of an internal sequence (IES) in Paramecium. *Nucleic acids research* **25**, 2448-2454, doi:10.1093/nar/25.12.2448 (1997).

26 Allen, S. E. *et al.* Circular Concatemers of Ultra-Short DNA Segments Produce Regulatory RNAs. *Cell* **168**, 990-999 e997, doi:10.1016/j.cell.2017.02.020 (2017).

27 Hoehener, C., Hug, I. & Nowacki, M. Dicer-like Enzymes with Sequence Cleavage Preferences. *Cell* **173**, 234-247 e237, doi:10.1016/j.cell.2018.02.029 (2018).

28 Lepere, G. *et al.* Silencing-associated and meiosis-specific small RNA pathways in Paramecium tetraurelia. *Nucleic acids research* **37**, 903-915, doi:10.1093/nar/gkn1018 (2009).

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

30 Mochizuki, K. & Gorovsky, M. A. Small RNAs in genome rearrangement in Tetrahymena. *Current opinion in genetics & development* **14**, 181-187, doi:10.1016/j.gde.2004.01.004 (2004).

31 Chalker, D. L. & Yao, M. C. Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila. *Genes & development* **15**, 1287-1298, doi:10.1101/gad.884601 (2001).

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

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

34 Lepere, G., Betermier, M., Meyer, E. & Duharcourt, S. Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in Paramecium tetraurelia. *Genes & development* **22**, 1501-1512, doi:10.1101/gad.473008 (2008).
35 Nowacki, M., Zagorski-Ostoja, W. & Meyer, E. Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in Paramecium tetraurelia. *Current biology : CB* **15**, 1616-1628, doi:10.1016/j.cub.2005.07.033 (2005).
36 Betermier, M., Duharcourt, S., Seitz, H. & Meyer, E. Timing of developmentally programmed excision and circularization of Paramecium internal eliminated sequences. *Molecular and cellular biology* **20**, 1553-1561, doi:10.1128/MCB.20.5.1553-1561.2000 (2000).
37 Furrer, D. I., Swart, E. C., Kraft, M. F., Sandoval, P. Y. & Nowacki, M. Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA. *Cell reports* **20**, 505-520, doi:10.1016/j.celrep.2017.06.050 (2017).
38 Marmignon, A. *et al.* Ku-mediated coupling of DNA cleavage and repair during programmed genome rearrangements in the ciliate Paramecium tetraurelia. *PLoS genetics* **10**, e1004552, doi:10.1371/journal.pgen.1004552 (2014).
39 Chen, D. *et al.* Condensed mitotic chromatin is accessible to transcription factors and chromatin structural proteins. *The Journal of cell biology* **168**, 41-54, doi:10.1083/jcb.200407182 (2005).
40 Paweletz, N. Walther Flemming: pioneer of mitosis research. *Nature reviews. Molecular cell biology* **2**, 72-75, doi:10.1038/35048077 (2001).
41 Hirano, T. Chromosome cohesion, condensation, and separation. *Annual review of biochemistry* **69**, 115-144, doi:10.1146/annurev.biochem.69.1.115 (2000).
42 Cobbe, N. & Heck, M. M. Review: SMCs in the world of chromosome biology- from prokaryotes to higher eukaryotes. *Journal of structural biology* **129**, 123-143, doi:10.1006/jsbi.2000.4255 (2000).
43 Cobbe, N. & Heck, M. M. The evolution of SMC proteins: phylogenetic analysis and structural implications. *Molecular biology and evolution* **21**, 332-347, doi:10.1093/molbev/msh023 (2004).
44 Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, doi:10.1126/science.aag0025 (2017).
45 Haering, C. H., Lowe, J., Hochwagen, A. & Nasmyth, K. Molecular architecture of SMC proteins and the yeast cohesin complex. *Molecular cell* **9**, 773-788, doi:10.1016/s1097-2765(02)00515-4 (2002).
46 Hopfner, K. P. *et al.* The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. *Nature* **418**, 562-566, doi:10.1038/nature00922 (2002).
47 Griese, J. J., Witte, G. & Hopfner, K. P. Structure and DNA binding activity of the mouse condensin hinge domain highlight common and diverse features of SMC proteins. *Nucleic acids research* **38**, 3454-3465, doi:10.1093/nar/gkq038 (2010).
48 Hirano, T. At the heart of the chromosome: SMC proteins in action. *Nature reviews. Molecular cell biology* **7**, 311-322, doi:10.1038/nrm1909 (2006).
49 Barysz, H. *et al.* Three-dimensional topology of the SMC2/SMC4 subcomplex from chicken condensin I revealed by cross-linking and molecular modelling. *Open biology* **5**, 150005, doi:10.1098/rsob.150005 (2015).
50 Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO journal* **1**, 945-951, doi:10.1002/j.1460-2075.1982.tb01276.x (1982).

51 Hirano, M. & Hirano, T. Opening closed arms: long-distance activation of SMC ATPase by hinge-DNA interactions. *Molecular cell* **21**, 175-186, doi:10.1016/j.molcel.2005.11.026 (2006).
52 Jeppsson, K., Kanno, T., Shirahige, K. & Sjogren, C. The maintenance of chromosome structure: positioning and functioning of SMC complexes. *Nature reviews. Molecular cell biology* **15**, 601-614, doi:10.1038/nrm3857 (2014).
53 Hopfner, K. P. *et al.* Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* **101**, 789-800, doi:10.1016/s0092-8674(00)80890-9 (2000).
54 Lowe, J., Cordell, S. C. & van den Ent, F. Crystal structure of the SMC head domain: an ABC ATPase with 900 residues antiparallel coiled-coil inserted. *Journal of molecular biology* **306**, 25-35, doi:10.1006/jmbi.2000.4379 (2001).
55 Hirano, M. & Hirano, T. Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. *The EMBO journal* **21**, 5733-5744, doi:10.1093/emboj/cdf575 (2002).
56 Fousteri, M. I. & Lehmann, A. R. A novel SMC protein complex in Schizosaccharomyces pombe contains the Rad18 DNA repair protein. *The EMBO journal* **19**, 1691-1702, doi:10.1093/emboj/19.7.1691 (2000).
57 Hirano, T., Kobayashi, R. & Hirano, M. Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a Xenopus homolog of the Drosophila Barren protein. *Cell* **89**, 511-521, doi:10.1016/s0092-8674(00)80233-0 (1997).
58 Yamazoe, M. *et al.* Complex formation of MukB, MukE and MukF proteins involved in chromosome partitioning in Escherichia coli. *The EMBO journal* **18**, 5873-5884, doi:10.1093/emboj/18.21.5873 (1999).
59 Schleiffer, A. *et al.* Kleisins: a superfamily of bacterial and eukaryotic SMC protein partners. *Molecular cell* **11**, 571-575, doi:10.1016/s1097-2765(03)00108-4 (2003).
60 Gruber, S., Haering, C. H. & Nasmyth, K. Chromosomal cohesin forms a ring. *Cell* **112**, 765-777, doi:10.1016/s0092-8674(03)00162-4 (2003).
61 Melby, T. E., Ciampaglio, C. N., Briscoe, G. & Erickson, H. P. The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. *The Journal of cell biology* **142**, 1595-1604, doi:10.1083/jcb.142.6.1595 (1998).
62 Jessberger, R., Frei, C. & Gasser, S. M. Chromosome dynamics: the SMC protein family. *Current opinion in genetics & development* **8**, 254-259, doi:10.1016/s0959-437x(98)80149-4 (1998).
63 Hirano, T. The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. *Genes & development* **16**, 399-414, doi:10.1101/gad.955102 (2002).
64 Yoshinaga, M. & Inagaki, Y. Ubiquity and Origins of Structural Maintenance of Chromosomes (SMC) Proteins in Eukaryotes. *Genome biology and evolution* **13**, doi:10.1093/gbe/evab256 (2021).
65 Jessberger, R. The many functions of SMC proteins in chromosome dynamics. *Nature reviews. Molecular cell biology* **3**, 767-778, doi:10.1038/nrm930 (2002).
66 Hagstrom, K. A. & Meyer, B. J. Condensin and cohesin: more than chromosome compactor and glue. *Nature reviews. Genetics* **4**, 520-534, doi:10.1038/nrg1110 (2003).
67 Ono, T., Fang, Y., Spector, D. L. & Hirano, T. Spatial and temporal regulation of Condensins I and II in mitotic chromosome assembly in human cells. *Molecular biology of the cell* **15**, 3296-3308, doi:10.1091/mbc.e04-03-0242 (2004).

68 Ono, T. *et al.* Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. *Cell* **115**, 109-121, doi:10.1016/s0092-8674(03)00724-4 (2003).
69 Ohta, S. *et al.* The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell* **142**, 810-821, doi:10.1016/j.cell.2010.07.047 (2010).
70 Csankovszki, G. *et al.* Three distinct condensin complexes control C. elegans chromosome dynamics. *Current biology : CB* **19**, 9-19, doi:10.1016/j.cub.2008.12.006 (2009).
71 Howard-Till, R., Tian, M. & Loidl, J. A specialized condensin complex participates in somatic nuclear maturation in Tetrahymena thermophila. *Molecular biology of the cell* **30**, 1326-1338, doi:10.1091/mbc.E18-08-0487 (2019).
72 Collette, K. S., Petty, E. L., Golenberg, N., Bembenek, J. N. & Csankovszki, G. Different roles for Aurora B in condensin targeting during mitosis and meiosis. *Journal of cell science* **124**, 3684-3694, doi:10.1242/jcs.088336 (2011).
73 Shintomi, K. & Hirano, T. The relative ratio of condensin I to II determines chromosome shapes. *Genes & development* **25**, 1464-1469, doi:10.1101/gad.2060311 (2011).
74 Hirano, T. Chromosome Dynamics during Mitosis. *Cold Spring Harbor perspectives in biology* **7**, doi:10.1101/cshperspect.a015792 (2015).
75 Hirano, T. & Mitchison, T. J. A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. *Cell* **79**, 449-458, doi:10.1016/0092-8674(94)90254-2 (1994).
76 Lee, B. G. *et al.* Cryo-EM structures of holo condensin reveal a subunit flip-flop mechanism. *Nature structural & molecular biology* **27**, 743-751, doi:10.1038/s41594-020-0457-x (2020).
77 Kalitsis, P., Zhang, T., Marshall, K. M., Nielsen, C. F. & Hudson, D. F. Condensin, master organizer of the genome. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **25**, 61-76, doi:10.1007/s10577-017-9553-0 (2017).
78 Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature genetics* **42**, 53-61, doi:10.1038/ng.496 (2010).
79 Tsang, C. K., Wei, Y. & Zheng, X. F. Compacting DNA during the interphase: condensin maintains rDNA integrity. *Cell cycle* **6**, 2213-2218, doi:10.4161/cc.6.18.4733 (2007).
80 Kobayashi, T. Strategies to maintain the stability of the ribosomal RNA gene repeats--collaboration of recombination, cohesion, and condensation. *Genes & genetic systems* **81**, 155-161, doi:10.1266/ggs.81.155 (2006).
81 Wood, A. J., Severson, A. F. & Meyer, B. J. Condensin and cohesin complexity: the expanding repertoire of functions. *Nature reviews. Genetics* **11**, 391-404, doi:10.1038/nrg2794 (2010).
82 Gerlich, D., Hirota, T., Koch, B., Peters, J. M. & Ellenberg, J. Condensin I stabilizes chromosomes mechanically through a dynamic interaction in live cells. *Current biology : CB* **16**, 333-344, doi:10.1016/j.cub.2005.12.040 (2006).
83 Hartl, T. A., Smith, H. F. & Bosco, G. Chromosome alignment and transvection are antagonized by condensin II. *Science* **322**, 1384-1387, doi:10.1126/science.1164216 (2008).

84 Joyce, E. F., Williams, B. R., Xie, T. & Wu, C. T. Identification of genes that promote or antagonize somatic homolog pairing using a high-throughput FISH-based screen. *PLoS genetics* **8**, e1002667, doi:10.1371/journal.pgen.1002667 (2012).

85 Iwasaki, O., Tanaka, A., Tanizawa, H., Grewal, S. I. & Noma, K. Centromeric localization of dispersed Pol III genes in fission yeast. *Molecular biology of the cell* **21**, 254-265, doi:10.1091/mbc.E09-09-0790 (2010).

86 Haeusler, R. A., Pratt-Hyatt, M., Good, P. D., Gipson, T. A. & Engelke, D. R. Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes. *Genes & development* **22**, 2204-2214, doi:10.1101/gad.1675908 (2008).

87 Lau, A. C. & Csankovszki, G. Condensin-mediated chromosome organization and gene regulation. *Frontiers in genetics* **5**, 473, doi:10.3389/fgene.2014.00473 (2014).

88 Nasmyth, K. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annual review of genetics* **35**, 673-745, doi:10.1146/annurev.genet.35.102401.091334 (2001).

89 Akai, Y. *et al.* Opposing role of condensin hinge against replication protein A in mitosis and interphase through promoting DNA annealing. *Open biology* **1**, 110023, doi:10.1098/rsob.110023 (2011).

90 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).

91 Gibcus, J. H. *et al.* A pathway for mitotic chromosome formation. *Science* **359**, doi:10.1126/science.aao6135 (2018).

92 Alipour, E. & Marko, J. F. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic acids research* **40**, 11202-11212, doi:10.1093/nar/gks925 (2012).

93 Terakawa, T. *et al.* The condensin complex is a mechanochemical motor that translocates along DNA. *Science* **358**, 672-676, doi:10.1126/science.aan6516 (2017).

94 Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102-105, doi:10.1126/science.aar7831 (2018).

95 Kschonsak, M. *et al.* Structural Basis for a Safety-Belt Mechanism That Anchors Condensin to Chromosomes. *Cell* **171**, 588-600 e524, doi:10.1016/j.cell.2017.09.008 (2017).

96 Shaltiel, I. A. *et al.* A hold-and-feed mechanism drives directional DNA loop extrusion by condensin. *Science* **376**, 1087-1094, doi:10.1126/science.abm4012 (2022).

97 Davidson, I. F. & Peters, J. M. Genome folding through loop extrusion by SMC complexes. *Nature reviews. Molecular cell biology* **22**, 445-464, doi:10.1038/s41580-021-00349-7 (2021).

98 Yanagida, M. Clearing the way for mitosis: is cohesin a target? *Nature reviews. Molecular cell biology* **10**, 489-496, doi:10.1038/nrm2712 (2009).

99 Heale, J. T. *et al.* Condensin I interacts with the PARP-1-XRCC1 complex and functions in DNA single-strand break repair. *Molecular cell* **21**, 837-848, doi:10.1016/j.molcel.2006.01.036 (2006).

100 Kimura, K. & Hirano, T. ATP-dependent positive supercoiling of DNA by 13S condensin: a biochemical implication for chromosome condensation. *Cell* **90**, 625-634, doi:10.1016/s0092-8674(00)80524-3 (1997).

101 Tsao, Y. P., Wu, H. Y. & Liu, L. F. Transcription-driven supercoiling of DNA: direct biochemical evidence from in vitro studies. *Cell* **56**, 111-118, doi:10.1016/0092-8674(89)90989-6 (1989).

102 Bazett-Jones, D. P., Kimura, K. & Hirano, T. Efficient supercoiling of DNA by a single condensin complex as revealed by electron spectroscopic imaging. *Molecular cell* **9**, 1183-1190, doi:10.1016/s1097-2765(02)00546-4 (2002).
103 Guacci, V., Hogan, E. & Koshland, D. Chromosome condensation and sister chromatid pairing in budding yeast. *The Journal of cell biology* **125**, 517-530, doi:10.1083/jcb.125.3.517 (1994).
104 Cervantes, M. D., Coyne, R. S., Xi, X. & Yao, M. C. The condensin complex is essential for amitotic segregation of bulk chromosomes, but not nucleoli, in the ciliate Tetrahymena thermophila. *Molecular and cellular biology* **26**, 4690-4700, doi:10.1128/MCB.02315-05 (2006).
105 Howard-Till, R. & Loidl, J. Condensins promote chromosome individualization and segregation during mitosis, meiosis, and amitosis in Tetrahymena thermophila. *Molecular biology of the cell* **29**, 466-478, doi:10.1091/mbc.E17-07-0451 (2018).
106 Mitra, R., Fain-Thornton, J. & Craig, N. L. piggyBac can bypass DNA synthesis during cut and paste transposition. *The EMBO journal* **27**, 1097-1109, doi:10.1038/emboj.2008.41 (2008).
107 Dubois, E. *et al.* Multimerization properties of PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements. *Nucleic acids research* **45**, 3204-3216, doi:10.1093/nar/gkw1359 (2017).
108 Guerineau, M. *et al.* The unusual structure of the PiggyMac cysteine-rich domain reveals zinc finger diversity in PiggyBac-related transposases. *Mobile DNA* **12**, 12, doi:10.1186/s13100-021-00240-4 (2021).
109 Bischerour, J. *et al.* Six domesticated PiggyBac transposases together carry out programmed DNA elimination in Paramecium. *eLife* **7**, doi:10.7554/eLife.37927 (2018).
110 de Vanssay, A. *et al.* The Paramecium histone chaperone Spt16-1 is required for Pgm endonuclease function in programmed genome rearrangements. *PLoS genetics* **16**, e1008949, doi:10.1371/journal.pgen.1008949 (2020).
111 Abello, A. *et al.* Functional diversification of Paramecium Ku80 paralogs safeguards genome integrity during precise programmed DNA elimination. *PLoS genetics* **16**, e1008723, doi:10.1371/journal.pgen.1008723 (2020).
112 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181-211, doi:10.1146/annurev.biochem.052308.093131 (2010).
113 Bhullar, S. *et al.* A mating-type mutagenesis screen identifies a zinc-finger protein required for specific DNA excision events in Paramecium. *Nucleic acids research* **46**, 9550-9562, doi:10.1093/nar/gky772 (2018).
114 Hirano, T. Condensin-Based Chromosome Organization from Bacteria to Vertebrates. *Cell* **164**, 847-857, doi:10.1016/j.cell.2016.01.033 (2016).
115 Anderson, D. E., Losada, A., Erickson, H. P. & Hirano, T. Condensin and cohesin display different arm conformations with characteristic hinge angles. *The Journal of cell biology* **156**, 419-424, doi:10.1083/jcb.200111002 (2002).
116 Chan, R. C., Severson, A. F. & Meyer, B. J. Condensin restructures chromosomes in preparation for meiotic divisions. *The Journal of cell biology* **167**, 613-625, doi:10.1083/jcb.200408061 (2004).
117 Nishide, K. & Hirano, T. Overlapping and non-overlapping functions of condensins I and II in neural stem cell divisions. *PLoS genetics* **10**, e1004847, doi:10.1371/journal.pgen.1004847 (2014).

118 Green, L. C. *et al.* Contrasting roles of condensin I and condensin II in mitotic chromosome formation. *Journal of cell science* **125**, 1591-1604, doi:10.1242/jcs.097790 (2012).
119 Sakai, A., Hizume, K., Sutani, T., Takeyasu, K. & Yanagida, M. Condensin but not cohesin SMC heterodimer induces DNA reannealing through protein-protein assembly. *The EMBO journal* **22**, 2764-2775, doi:10.1093/emboj/cdg247 (2003).
120 Sutani, T. & Yanagida, M. DNA renaturation activity of the SMC complex implicated in chromosome condensation. *Nature* **388**, 798-801, doi:10.1038/42062 (1997).
121 Festuccia, N., Gonzalez, I., Owens, N. & Navarro, P. Mitotic bookmarking in development and stem cells. *Development* **144**, 3633-3645, doi:10.1242/dev.146522 (2017).
122 Beisson, J. *et al.* Mass culture of Paramecium tetraurelia. *Cold Spring Harbor protocols* **2010**, pdb prot5362, doi:10.1101/pdb.prot5362 (2010).
123 Beisson, J. *et al.* Maintaining clonal Paramecium tetraurelia cell lines of controlled age through daily reisolation. *Cold Spring Harbor protocols* **2010**, pdb prot5361, doi:10.1101/pdb.prot5361 (2010).
124 Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics* **20**, 1160-1166, doi:10.1093/bib/bbx108 (2019).
125 Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic acids research* **39**, D225-229, doi:10.1093/nar/gkq1189 (2011).
126 Arnaiz, O. *et al.* Improved methods and resources for paramecium genomics: transcription units, gene annotation and gene expression. *BMC genomics* **18**, 483, doi:10.1186/s12864-017-3887-z (2017).
127 Trifinopoulos, J., Nguyen, L. T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic acids research* **44**, W232-235, doi:10.1093/nar/gkw256 (2016).
128 Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811, doi:10.1038/35888 (1998).
129 Arnaiz, O. & Sperling, L. ParameciumDB in 2011: new tools and new data for functional and comparative genomics of the model ciliate Paramecium tetraurelia. *Nucleic acids research* **39**, D632-636, doi:10.1093/nar/gkq918 (2011).
130 Denby Wilkes, C., Arnaiz, O. & Sperling, L. ParTIES: a toolbox for Paramecium interspersed DNA elimination studies. *Bioinformatics* **32**, 599-601, doi:10.1093/bioinformatics/btv691 (2016).
131 Bechara, S., Kabbani, L., Maurer Alcala, X. & Nowacki, M. Identification of novel, functional long non-coding RNAs involved in programmed, large scale genome rearrangements. *Rna*, doi:10.1261/rna.079134.122 (2022).
132 Rzeszutek, I., Swart, E. C., Pabian-Jewula, S., Russo, A. & Nowacki, M. Early developmental, meiosis-specific proteins - Spo11, Msh4-1, and Msh5 - Affect subsequent genome reorganization in Paramecium tetraurelia. *Biochimica et biophysica acta. Molecular cell research* **1869**, 119239, doi:10.1016/j.bbamcr.2022.119239 (2022).
133 Beisson, J. *et al.* DNA microinjection into the macronucleus of paramecium. *Cold Spring Harbor protocols* **2010**, pdb prot5364, doi:10.1101/pdb.prot5364 (2010).
134 Reuter, M. *et al.* Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nature structural & molecular biology* **16**, 639-646, doi:10.1038/nsmb.1615 (2009).

135 Uhlmann, F. SMC complexes: from DNA to chromosomes. *Nature reviews. Molecular cell biology* **17**, 399-412, doi:10.1038/nrm.2016.30 (2016).
136 De Piccoli, G., Torres-Rosell, J. & Aragon, L. The unnamed complex: what do we know about Smc5-Smc6? *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **17**, 251-263, doi:10.1007/s10577-008-9016-8 (2009).
137 Michelini, F. *et al.* From "Cellular" RNA to "Smart" RNA: Multiple Roles of RNA in Genome Stability and Beyond. *Chemical reviews* **118**, 4365-4403, doi:10.1021/acs.chemrev.7b00487 (2018).
138 Thattikota, Y. *et al.* Cdc48/VCP Promotes Chromosome Morphogenesis by Releasing Condensin from Self-Entrapment in Chromatin. *Molecular cell* **69**, 664-676 e665, doi:10.1016/j.molcel.2018.01.030 (2018).
139 Burmann, F. *et al.* An asymmetric SMC-kleisin bridge in prokaryotic condensin. *Nature structural & molecular biology* **20**, 371-379, doi:10.1038/nsmb.2488 (2013).
140 Murarka, P. & Srivastava, P. An improved method for the isolation and identification of unknown proteins that bind to known DNA sequences by affinity capture and mass spectrometry. *PLoS One* **13**, e0202602, doi:10.1371/journal.pone.0202602 (2018).
141 Gilmour, D. S. & Lis, J. T. In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster. *Molecular and cellular biology* **5**, 2009-2018, doi:10.1128/mcb.5.8.2009-2018.1985 (1985).
142 Gade, P. & Kalvakolanu, D. V. Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity. *Methods in molecular biology* **809**, 85-104, doi:10.1007/978-1-61779-376-9_6 (2012).
143 Belaghzal, H., Dekker, J. & Gibcus, J. H. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* **123**, 56-65, doi:10.1016/j.ymeth.2017.04.004 (2017).
144 Gluck, F. *et al.* EasyProt--an easy-to-use graphical platform for proteomics data analysis. *Journal of proteomics* **79**, 146-160, doi:10.1016/j.jprot.2012.12.012 (2013).
145 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).
146 Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nature protocols* **8**, 1551-1566, doi:10.1038/nprot.2013.092 (2013).
147 McCarthy, F. M. *et al.* AgBase: a unified resource for functional analysis in agriculture. *Nucleic acids research* **35**, D599-603, doi:10.1093/nar/gkl936 (2007).
148 Thomas, P. D. *et al.* PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic acids research* **31**, 334-341, doi:10.1093/nar/gkg115 (2003).
149 Galvani, A. & Sperling, L. RNA interference by feeding in Paramecium. *Trends in genetics : TIG* **18**, 11-12, doi:10.1016/s0168-9525(01)02548-3 (2002).
150 Cora, E. *et al.* The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. *Rna* **20**, 773-781, doi:10.1261/rna.044701.114 (2014).

# 8. Appendix

Supplementary Table S1. Differential expression genes between SMC4-2 and EV KD at late time point.

| | baseMean | log2FoldChange | lfcSE | stat | pvalue | padj |
|---|---|---|---|---|---|---|
| PTET.51.1.G5220002 | 40.71833 | 21.51369 | 4.408693 | 4.879833 | 1.06E-06 | 0.000284 |
| PTET.51.1.G0080326 | 32.65949 | 21.2026 | 4.409016 | 4.808918 | 1.52E-06 | 0.000358 |
| PTET.51.1.G0210128 | 32.02327 | 21.17525 | 4.409048 | 4.802681 | 1.57E-06 | 0.000363 |
| PTET.51.1.G2560008 | 26.69836 | 20.92597 | 4.409379 | 4.745786 | 2.08E-06 | 0.000443 |
| PTET.51.1.G5840001 | 25.66103 | 20.87199 | 4.40946 | 4.733457 | 2.21E-06 | 0.000465 |
| PTET.51.1.G0120376 | 32.64916 | 8.401617 | 1.600445 | 5.249551 | 1.52E-07 | 6.30E-05 |
| PTET.51.1.G4230002 | 19.04111 | 7.654903 | 1.626836 | 4.705393 | 2.53E-06 | 0.000521 |
| PTET.51.1.G0950057 | 56.83275 | 6.652865 | 1.21276 | 5.485721 | 4.12E-08 | 1.97E-05 |
| PTET.51.1.G0460269 | 685.057 | 5.30066 | 0.505258 | 10.49099 | 9.50E-26 | 1.06E-21 |
| PTET.51.1.G0680082 | 56.55277 | 5.202766 | 0.892612 | 5.828701 | 5.59E-09 | 3.67E-06 |
| PTET.51.1.G1780066 | 56.55277 | 5.202766 | 0.892612 | 5.828701 | 5.59E-09 | 3.67E-06 |
| PTET.51.1.G0190324 | 656.2619 | 4.513589 | 0.473992 | 9.522507 | 1.69E-21 | 1.42E-17 |
| PTET.51.1.G0950056 | 168.0462 | 4.440877 | 0.549309 | 8.084484 | 6.24E-16 | 2.32E-12 |
| PTET.51.1.G0910133 | 184.057 | 4.315784 | 0.572903 | 7.533182 | 4.95E-14 | 9.21E-11 |
| PTET.51.1.G1290040 | 214.7117 | 4.310907 | 0.56506 | 7.629118 | 2.36E-14 | 5.28E-11 |
| PTET.51.1.G0760010 | 130.2862 | 4.293149 | 0.684386 | 6.272989 | 3.54E-10 | 3.39E-07 |
| PTET.51.1.G0740074 | 223.7873 | 4.251348 | 0.559518 | 7.598226 | 3.00E-14 | 6.28E-11 |
| PGML5a | 2031.005 | 4.249252 | 0.544829 | 7.799236 | 6.23E-15 | 1.60E-11 |
| PTET.51.1.G1560014 | 523.7589 | 4.194562 | 0.469851 | 8.927437 | 4.36E-19 | 2.62E-15 |
| PTET.51.1.G1410098 | 134.8058 | 4.194174 | 0.57794 | 7.257112 | 3.95E-13 | 6.62E-10 |
| PTET.51.1.G0700198 | 84.95626 | 4.149045 | 0.728964 | 5.691697 | 1.26E-08 | 7.26E-06 |
| PTET.51.1.G0690142 | 4122.739 | 3.936999 | 0.470598 | 8.36594 | 5.96E-17 | 2.50E-13 |
| PTET.51.1.G0920152 | 346.0847 | 3.923691 | 0.589339 | 6.65778 | 2.78E-11 | 3.58E-08 |
| PTET.51.1.G0370120 | 70.66986 | 3.877675 | 0.678655 | 5.71376 | 1.11E-08 | 6.49E-06 |
| PTET.51.1.G0930036 | 1387.979 | 3.808416 | 0.567336 | 6.71281 | 1.91E-11 | 2.66E-08 |
| PTET.51.1.G1660105 | 339.0518 | 3.804513 | 0.570187 | 6.672397 | 2.52E-11 | 3.37E-08 |
| PTET.51.1.G1550066 | 3913.099 | 3.788689 | 0.501772 | 7.550612 | 4.33E-14 | 8.53E-11 |
| PTET.51.1.G0020221 | 178.2637 | 3.774387 | 0.679003 | 5.558722 | 2.72E-08 | 1.52E-05 |
| PTET.51.1.G1760029 | 33.51372 | 3.766921 | 0.833364 | 4.520137 | 6.18E-06 | 0.001072 |
| PTET.51.1.G1240104 | 30.24863 | 3.744827 | 0.80014 | 4.680214 | 2.87E-06 | 0.000582 |
| PTET.51.1.G0080272 | 861.277 | 3.732785 | 0.4637 | 8.049995 | 8.28E-16 | 2.77E-12 |

| | | | | | | |
|---|---|---|---|---|---|---|
| PTET.51.1.G0360075 | 38.99482 | 3.722471 | 0.880888 | 4.225816 | 2.38E-05 | 0.003097 |
| PTET.51.1.G2080006 | 104.4516 | 3.659177 | 0.755907 | 4.840776 | 1.29E-06 | 0.000323 |
| PTET.51.1.G0130368 | 705.9818 | 3.612385 | 0.456568 | 7.912037 | 2.53E-15 | 7.07E-12 |
| PTET.51.1.G1300031 | 1316.057 | 3.535777 | 0.602317 | 5.870294 | 4.35E-09 | 3.10E-06 |
| PTET.51.1.G1390004 | 70.81666 | 3.514586 | 0.858046 | 4.096035 | 4.20E-05 | 0.004939 |
| PTET.51.1.G0890016 | 81.87751 | 3.460816 | 0.639799 | 5.40922 | 6.33E-08 | 2.94E-05 |
| PTET.51.1.G0230275 | 113.3842 | 3.459901 | 0.591881 | 5.845599 | 5.05E-09 | 3.45E-06 |
| PTET.51.1.G1270118 | 30.75336 | 3.459795 | 0.805711 | 4.294087 | 1.75E-05 | 0.002437 |
| PTET.51.1.G0740032 | 29.85139 | 3.458307 | 0.779958 | 4.433964 | 9.25E-06 | 0.001434 |
| PTET.51.1.G0970093 | 909.4569 | 3.455094 | 0.461216 | 7.491267 | 6.82E-14 | 1.20E-10 |
| PTET.51.1.G0350183 | 224.8513 | 3.45189 | 0.622481 | 5.545376 | 2.93E-08 | 1.58E-05 |
| PTET.51.1.G1570125 | 222.3059 | 3.438796 | 0.843702 | 4.075842 | 4.58E-05 | 0.005281 |
| PTET.51.1.G1790053 | 141.6275 | 3.416503 | 0.589884 | 5.791819 | 6.96E-09 | 4.32E-06 |
| PTET.51.1.G1190058 | 132.3896 | 3.387229 | 0.700372 | 4.836327 | 1.32E-06 | 0.000326 |
| PTET.51.1.G0140247 | 454.6272 | 3.340586 | 0.51995 | 6.424821 | 1.32E-10 | 1.52E-07 |
| PTET.51.1.G1030082 | 35.09834 | 3.303252 | 0.722917 | 4.569336 | 4.89E-06 | 0.000891 |
| PTET.51.1.G1590019 | 2632.748 | 3.28974 | 0.671642 | 4.89806 | 9.68E-07 | 0.000265 |
| PTET.51.1.G0860044 | 105.1833 | 3.288363 | 0.641295 | 5.127694 | 2.93E-07 | 0.000105 |
| PTIWI08 | 805.5734 | 3.285098 | 0.475292 | 6.911751 | 4.79E-12 | 7.29E-09 |
| PTET.51.1.G1090151 | 68.56817 | 3.283213 | 0.642023 | 5.113857 | 3.16E-07 | 0.00011 |
| PTET.51.1.G1000012 | 582.4541 | 3.280811 | 0.59635 | 5.501489 | 3.77E-08 | 1.83E-05 |
| PTET.51.1.G0210235 | 338.4711 | 3.27581 | 0.591792 | 5.535412 | 3.10E-08 | 1.65E-05 |
| PTET.51.1.G0260271 | 197.0646 | 3.269804 | 0.637142 | 5.13199 | 2.87E-07 | 0.000104 |
| PTET.51.1.G0150061 | 230.1202 | 3.268416 | 0.553312 | 5.907001 | 3.48E-09 | 2.60E-06 |
| PTET.51.1.G1340123 | 1197.373 | 3.252477 | 0.682486 | 4.765629 | 1.88E-06 | 0.000409 |
| PTET.51.1.G0860006 | 250.6767 | 3.213096 | 0.526024 | 6.108273 | 1.01E-09 | 9.12E-07 |
| PTET.51.1.G1300051 | 65.29121 | 3.200415 | 0.633153 | 5.054726 | 4.31E-07 | 0.000142 |
| PGML5b | 840.5487 | 3.191825 | 0.531636 | 6.003776 | 1.93E-09 | 1.70E-06 |
| PTET.51.1.G0740009 | 5603.293 | 3.187419 | 0.633586 | 5.030757 | 4.89E-07 | 0.000152 |
| PTET.51.1.G0050231 | 68.92674 | 3.164081 | 0.595836 | 5.310322 | 1.09E-07 | 4.70E-05 |
| PTET.51.1.G0550043 | 899.7438 | 3.147149 | 0.493723 | 6.374325 | 1.84E-10 | 1.99E-07 |
| PTET.51.1.G0080273 | 538.5461 | 3.142084 | 0.621969 | 5.051834 | 4.38E-07 | 0.000142 |
| PTET.51.1.G0950020 | 519.1471 | 3.12217 | 0.654368 | 4.771275 | 1.83E-06 | 0.000403 |
| PTET.51.1.G0180198 | 1381.869 | 3.110291 | 0.468988 | 6.631923 | 3.31E-11 | 4.11E-08 |
| PTET.51.1.G0360066 | 77.95532 | 3.094268 | 0.639124 | 4.841424 | 1.29E-06 | 0.000323 |
| PTET.51.1.G0200051 | 701.4204 | 3.092402 | 0.520896 | 5.936696 | 2.91E-09 | 2.50E-06 |

| | | | | | | |
|---|---|---|---|---|---|---|
| PTET.51.1.G0860046 | 336.6854 | 3.069771 | 0.556678 | 5.514444 | 3.50E-08 | 1.75E-05 |
| PTET.51.1.G0560201 | 54.09134 | 3.068474 | 0.664301 | 4.619102 | 3.85E-06 | 0.000742 |
| PTET.51.1.G1500004 | 833.7234 | 3.055053 | 0.672208 | 4.544802 | 5.50E-06 | 0.000974 |
| PTET.51.1.G1410092 | 213.4016 | 3.031674 | 0.623238 | 4.864391 | 1.15E-06 | 0.000298 |
| PGML3a | 514.2986 | 3.019893 | 0.582971 | 5.18018 | 2.22E-07 | 8.44E-05 |
| PTET.51.1.G1500021 | 228.1698 | 3.019433 | 0.704807 | 4.284054 | 1.84E-05 | 0.002498 |
| PTET.51.1.G1400105 | 160.2507 | 3.008773 | 0.654409 | 4.597696 | 4.27E-06 | 0.000795 |
| PTET.51.1.G2630002 | 288.132 | 2.999124 | 0.596295 | 5.029597 | 4.92E-07 | 0.000152 |
| PIE2 | 280.2876 | 2.978457 | 0.554192 | 5.374415 | 7.68E-08 | 3.48E-05 |
| PTET.51.1.G0120140 | 1265.478 | 2.969349 | 0.471013 | 6.304177 | 2.90E-10 | 2.94E-07 |
| PTET.51.1.G3820001 | 31.06187 | 2.966221 | 0.689143 | 4.304215 | 1.68E-05 | 0.002358 |
| PTET.51.1.G1020165 | 2146.4 | 2.951828 | 0.534945 | 5.518002 | 3.43E-08 | 1.74E-05 |
| PTET.51.1.G0070281 | 90.18804 | 2.946778 | 0.624989 | 4.714924 | 2.42E-06 | 0.000503 |
| PTET.51.1.G1620005 | 211.9253 | 2.945819 | 0.498813 | 5.905655 | 3.51E-09 | 2.60E-06 |
| PTMB.203, PTETG100132001 | 92.07212 | 2.925432 | 0.605446 | 4.831861 | 1.35E-06 | 0.000328 |
| PTET.51.1.G0270188 | 30.78304 | 2.924083 | 0.719234 | 4.065552 | 4.79E-05 | 0.00544 |
| PTET.51.1.G0030402 | 324.3273 | 2.922524 | 0.568282 | 5.142733 | 2.71E-07 | 9.96E-05 |
| PTET.51.1.G0650036 | 552.0582 | 2.910121 | 0.454679 | 6.40038 | 1.55E-10 | 1.73E-07 |
| PTET.51.1.G0070261 | 152.9081 | 2.874669 | 0.563474 | 5.101686 | 3.37E-07 | 0.000114 |
| PTET.51.1.G1770066 | 33.27577 | 2.852072 | 0.693217 | 4.114252 | 3.88E-05 | 0.004629 |
| PTET.51.1.G0120377 | 1037.544 | 2.85097 | 0.547239 | 5.209731 | 1.89E-07 | 7.62E-05 |
| PTET.51.1.G1300067 | 329.2038 | 2.846041 | 0.549652 | 5.177901 | 2.24E-07 | 8.44E-05 |
| sAG_51g | 61.73219 | 2.843252 | 0.634583 | 4.480507 | 7.45E-06 | 0.001225 |
| PTET.51.1.G1530121 | 725.1364 | 2.841345 | 0.48061 | 5.911958 | 3.38E-09 | 2.60E-06 |
| ku70-1 | 766.8925 | 2.837117 | 0.594759 | 4.7702 | 1.84E-06 | 0.000403 |
| PTET.51.1.G0260286 | 1532.243 | 2.832844 | 0.449056 | 6.308436 | 2.82E-10 | 2.94E-07 |
| PTET.51.1.G0950019 | 1311.486 | 2.831603 | 0.69686 | 4.063376 | 4.84E-05 | 0.005472 |
| PTET.51.1.G2860001 | 334.7067 | 2.830554 | 0.533893 | 5.301728 | 1.15E-07 | 4.86E-05 |
| PTET.51.1.G1520043 | 5367.705 | 2.827908 | 0.489797 | 5.773638 | 7.76E-09 | 4.72E-06 |
| PTET.51.1.G0810201 | 203.6296 | 2.820272 | 0.519775 | 5.425949 | 5.76E-08 | 2.72E-05 |
| PGML4a | 128.8772 | 2.819598 | 0.579477 | 4.86576 | 1.14E-06 | 0.000298 |
| PTET.51.1.G0280071 | 156.729 | 2.815108 | 0.578502 | 4.866205 | 1.14E-06 | 0.000298 |
| PTET.51.1.G0750201 | 87.73424 | 2.809246 | 0.610999 | 4.597792 | 4.27E-06 | 0.000795 |
| PTET.51.1.G0970158 | 351.81 | 2.805533 | 0.720953 | 3.891421 | 9.97E-05 | 0.009933 |
| PTET.51.1.G0400261 | 1200.415 | 2.798471 | 0.474077 | 5.902992 | 3.57E-09 | 2.60E-06 |

| | | | | | | |
|---|---|---|---|---|---|---|
| PTET.51.1.G1670111 | 547.2172 | 2.793773 | 0.481728 | 5.799487 | 6.65E-09 | 4.21E-06 |
| PTET.51.1.G0820079 | 53.43555 | 2.787327 | 0.608543 | 4.58033 | 4.64E-06 | 0.000854 |
| PGML2 | 372.3613 | 2.783207 | 0.517235 | 5.380929 | 7.41E-08 | 3.40E-05 |
| SETb1 | 329.1162 | 2.782301 | 0.620196 | 4.486161 | 7.25E-06 | 0.001222 |
| PTMB.205, PTETG100130001 | 113.2494 | 2.780495 | 0.603566 | 4.606776 | 4.09E-06 | 0.00077 |
| PTET.51.1.G1590093 | 91.18207 | 2.778736 | 0.610625 | 4.550639 | 5.35E-06 | 0.000953 |
| PGML3c | 54.86999 | 2.775097 | 0.604637 | 4.589692 | 4.44E-06 | 0.000821 |
| PTET.51.1.G1410101 | 96.85685 | 2.773006 | 0.609259 | 4.551444 | 5.33E-06 | 0.000953 |
| PTET.51.1.G1440089 | 5924.881 | 2.769054 | 0.501817 | 5.518052 | 3.43E-08 | 1.74E-05 |
| PTET.51.1.G0620172 | 346.0612 | 2.754072 | 0.62407 | 4.413082 | 1.02E-05 | 0.001544 |
| PTET.51.1.G0160234 | 101.0379 | 2.749016 | 0.570449 | 4.819035 | 1.44E-06 | 0.000345 |
| PTET.51.1.G1550035 | 208.6115 | 2.740887 | 0.5253 | 5.217754 | 1.81E-07 | 7.40E-05 |
| PTET.51.1.G0920155 | 131.1203 | 2.726684 | 0.554091 | 4.921005 | 8.61E-07 | 0.000241 |
| PTET.51.1.G0020207 | 51.01128 | 2.718359 | 0.599136 | 4.537133 | 5.70E-06 | 0.001005 |
| PTET.51.1.G1580082 | 958.9166 | 2.715114 | 0.45831 | 5.924185 | 3.14E-09 | 2.56E-06 |
| PTET.51.1.G0060408 | 1794.571 | 2.713553 | 0.551472 | 4.920569 | 8.63E-07 | 0.000241 |
| PTET.51.1.G0970059 | 429.1529 | 2.707744 | 0.543015 | 4.986499 | 6.15E-07 | 0.000184 |
| PTET.51.1.G1410093 | 1167.943 | 2.699896 | 0.456568 | 5.91346 | 3.35E-09 | 2.60E-06 |
| PTET.51.1.G0020442 | 196.284 | 2.698962 | 0.533732 | 5.056776 | 4.26E-07 | 0.000141 |
| NOWA2 | 2567.751 | 2.691485 | 0.558237 | 4.821399 | 1.43E-06 | 0.000343 |
| PTET.51.1.G0440127 | 766.8217 | 2.682755 | 0.462616 | 5.799103 | 6.67E-09 | 4.21E-06 |
| PTET.51.1.G0910001 | 53.78835 | 2.679645 | 0.654026 | 4.097151 | 4.18E-05 | 0.004932 |
| PTET.51.1.G0050297 | 419.5044 | 2.67641 | 0.457695 | 5.847585 | 4.99E-09 | 3.45E-06 |
| PTET.51.1.G0120047 | 63.96908 | 2.654335 | 0.576064 | 4.607709 | 4.07E-06 | 0.00077 |
| PTET.51.1.G0040217 | 127.0865 | 2.649768 | 0.51921 | 5.103464 | 3.33E-07 | 0.000114 |
| PTET.51.1.G0510207 | 46.75858 | 2.645398 | 0.666315 | 3.97019 | 7.18E-05 | 0.007587 |
| PTET.51.1.G0440203 | 583.3799 | 2.640862 | 0.471728 | 5.598276 | 2.16E-08 | 1.23E-05 |
| PTET.51.1.G0040036 | 1592.522 | 2.640707 | 0.494962 | 5.335167 | 9.55E-08 | 4.21E-05 |
| PTET.51.1.G1280098 | 204.5025 | 2.639252 | 0.479603 | 5.50299 | 3.73E-08 | 1.83E-05 |
| PTET.51.1.G1190050 | 493.1198 | 2.637954 | 0.639295 | 4.126348 | 3.69E-05 | 0.004479 |
| PTET.51.1.G1070131 | 218.6531 | 2.634468 | 0.534803 | 4.926054 | 8.39E-07 | 0.000238 |
| PtRPB1 | 2460.999 | 2.630865 | 0.625569 | 4.205554 | 2.60E-05 | 0.003355 |
| PTET.51.1.G0120198 | 649.9158 | 2.627062 | 0.50006 | 5.25349 | 1.49E-07 | 6.25E-05 |
| PTET.51.1.G6710001 | 134.7625 | 2.611914 | 0.581985 | 4.487937 | 7.19E-06 | 0.001222 |
| PTET.51.1.G0570172 | 251.7891 | 2.611164 | 0.524902 | 4.974572 | 6.54E-07 | 0.000193 |

| | | | | | | |
|---|---|---|---|---|---|---|
| PTET.51.1.G2740007 | 768.0912 | 2.604055 | 0.489847 | 5.31606 | 1.06E-07 | 4.61E-05 |
| PTET.51.1.G0580153 | 225.031 | 2.602068 | 0.523338 | 4.972063 | 6.62E-07 | 0.000193 |
| PTET.51.1.G0320111 | 95.68941 | 2.596914 | 0.594487 | 4.368325 | 1.25E-05 | 0.001847 |
| PTET.51.1.G0140243 | 508.2914 | 2.593007 | 0.560574 | 4.625627 | 3.73E-06 | 0.000727 |
| PTET.51.1.G0110239 | 6606.55 | 2.582082 | 0.46509 | 5.551787 | 2.83E-08 | 1.55E-05 |
| PTET.51.1.G0490010 | 125.5337 | 2.580456 | 0.513584 | 5.024412 | 5.05E-07 | 0.000155 |
| PTET.51.1.G1210169 | 23468.07 | 2.579541 | 0.616348 | 4.185204 | 2.85E-05 | 0.003628 |
| PTET.51.1.G0500162 | 328.0773 | 2.577497 | 0.502433 | 5.130031 | 2.90E-07 | 0.000104 |
| PTET.51.1.G0400276 | 622.208 | 2.564517 | 0.477437 | 5.37142 | 7.81E-08 | 3.49E-05 |
| PTET.51.1.G0100167 | 326.9162 | 2.556505 | 0.462731 | 5.524819 | 3.30E-08 | 1.73E-05 |
| PTET.51.1.G0370143 | 61.06252 | 2.544095 | 0.618104 | 4.115964 | 3.86E-05 | 0.004627 |
| PTET.51.1.G0110132 | 1969.304 | 2.519988 | 0.524264 | 4.806716 | 1.53E-06 | 0.000359 |
| PTET.51.1.G1500049 | 43.70777 | 2.500593 | 0.634021 | 3.944026 | 8.01E-05 | 0.008257 |
| PTET.51.1.G0730142 | 111.9686 | 2.497814 | 0.628727 | 3.972813 | 7.10E-05 | 0.007528 |
| PTET.51.1.G0730048 | 45.69381 | 2.497235 | 0.634579 | 3.935264 | 8.31E-05 | 0.008511 |
| NOWA1 | 1244.285 | 2.496632 | 0.55738 | 4.479223 | 7.49E-06 | 0.001225 |
| PTET.51.1.G1130106 | 344.9909 | 2.483521 | 0.480796 | 5.165441 | 2.40E-07 | 8.93E-05 |
| PTET.51.1.G0970003 | 486.3809 | 2.475304 | 0.483251 | 5.122193 | 3.02E-07 | 0.000106 |
| PTET.51.1.G0680057 | 40.00062 | 2.473119 | 0.629147 | 3.930911 | 8.46E-05 | 0.00864 |
| PTET.51.1.G1250162 | 97.15832 | 2.468204 | 0.531193 | 4.646526 | 3.38E-06 | 0.000673 |
| PTET.51.1.G0610198 | 194.0407 | 2.465366 | 0.550048 | 4.482091 | 7.39E-06 | 0.001225 |
| PTET.51.1.G0400181 | 1623.94 | 2.455199 | 0.59472 | 4.128329 | 3.65E-05 | 0.004466 |
| DCL5 | 391.1537 | 2.454135 | 0.547896 | 4.479197 | 7.49E-06 | 0.001225 |
| PTET.51.1.G0040071 | 116.5924 | 2.453062 | 0.515181 | 4.761555 | 1.92E-06 | 0.000412 |
| PTET.51.1.G0940013 | 201.6209 | 2.445154 | 0.480517 | 5.088595 | 3.61E-07 | 0.000121 |
| PTET.51.1.G0530225 | 218.4427 | 2.444252 | 0.552322 | 4.425412 | 9.63E-06 | 0.001479 |
| PTET.51.1.G1180014 | 351.1036 | 2.442894 | 0.527688 | 4.629431 | 3.67E-06 | 0.000718 |
| PTET.51.1.G0370136 | 74.22967 | 2.442789 | 0.574356 | 4.253094 | 2.11E-05 | 0.002791 |
| PTET.51.1.G0440244 | 872.8997 | 2.437151 | 0.615563 | 3.959224 | 7.52E-05 | 0.007837 |
| PTET.51.1.G1390096 | 305.243 | 2.435998 | 0.538761 | 4.521486 | 6.14E-06 | 0.001071 |
| PTET.51.1.G0280060 | 116.5327 | 2.433382 | 0.57365 | 4.24193 | 2.22E-05 | 0.00291 |
| PTET.51.1.G1130040 | 301.94 | 2.428765 | 0.499168 | 4.865631 | 1.14E-06 | 0.000298 |
| PTET.51.1.G0380064 | 269.0768 | 2.422327 | 0.584566 | 4.143807 | 3.42E-05 | 0.00419 |
| PTET.51.1.G1670070 | 603.9813 | 2.420244 | 0.487359 | 4.966034 | 6.83E-07 | 0.000197 |
| PTET.51.1.G2550004 | 406.8497 | 2.42016 | 0.579875 | 4.173592 | 3.00E-05 | 0.003775 |
| PTET.51.1.G0580061 | 650.3978 | 2.41614 | 0.493432 | 4.8966 | 9.75E-07 | 0.000265 |

| | | | | | | |
|---|---|---|---|---|---|---|
| PTET.51.1.G1660029 | 5517.256 | 2.414408 | 0.582067 | 4.147993 | 3.35E-05 | 0.00413 |
| PTET.51.1.G0430272 | 130.2183 | 2.40415 | 0.50487 | 4.76192 | 1.92E-06 | 0.000412 |
| PTET.51.1.G0120094 | 187.4943 | 2.388572 | 0.482202 | 4.953471 | 7.29E-07 | 0.000209 |
| PTET.51.1.G0070139 | 1168.461 | 2.387628 | 0.552254 | 4.323425 | 1.54E-05 | 0.002171 |
| H3P4 | 9110.24 | 2.382971 | 0.477626 | 4.989194 | 6.06E-07 | 0.000183 |
| PTET.51.1.G0070416 | 146.5691 | 2.37141 | 0.582101 | 4.073882 | 4.62E-05 | 0.005303 |
| PTET.51.1.G1250095 | 328.0021 | 2.371047 | 0.470652 | 5.037791 | 4.71E-07 | 0.000149 |
| PTET.51.1.G0230221 | 206.7634 | 2.350453 | 0.496496 | 4.734083 | 2.20E-06 | 0.000465 |
| PTET.51.1.G0630147 | 1865.088 | 2.339471 | 0.52224 | 4.47969 | 7.48E-06 | 0.001225 |
| PTET.51.1.G0230023 | 559.1881 | 2.330033 | 0.456544 | 5.103634 | 3.33E-07 | 0.000114 |
| PTET.51.1.G1180067 | 269.0999 | 2.325907 | 0.478994 | 4.85582 | 1.20E-06 | 0.000309 |
| PTMB.219, PTETG100122001 | 1213.072 | 2.317893 | 0.445815 | 5.199227 | 2.00E-07 | 7.70E-05 |
| epsilon-51D | 66.14468 | 2.312644 | 0.58419 | 3.958722 | 7.54E-05 | 0.007837 |
| PTET.51.1.G0660068 | 290.3745 | 2.31078 | 0.490484 | 4.711228 | 2.46E-06 | 0.000509 |
| PTET.51.1.G0230328 | 717.4595 | 2.310656 | 0.477241 | 4.841701 | 1.29E-06 | 0.000323 |
| PTET.51.1.G1530110 | 372.8879 | 2.310105 | 0.483309 | 4.779769 | 1.75E-06 | 0.000389 |
| PTET.51.1.G2090002 | 443.2323 | 2.300989 | 0.476078 | 4.833215 | 1.34E-06 | 0.000328 |
| PTET.51.1.G0350166 | 117.0893 | 2.29998 | 0.565598 | 4.066455 | 4.77E-05 | 0.005437 |
| PTET.51.1.G0730130 | 487.829 | 2.299647 | 0.492079 | 4.673325 | 2.96E-06 | 0.000598 |
| PtKu80-3 | 1586.418 | 2.294639 | 0.512947 | 4.473442 | 7.70E-06 | 0.001245 |
| PTET.51.1.G1290041 | 211.0183 | 2.291167 | 0.494021 | 4.637795 | 3.52E-06 | 0.000694 |
| PTET.51.1.G0660063 | 211.2992 | 2.288728 | 0.485106 | 4.717992 | 2.38E-06 | 0.000499 |
| PTET.51.1.G0410123 | 2478.354 | 2.284122 | 0.467316 | 4.887751 | 1.02E-06 | 0.000275 |
| PTET.51.1.G1560117 | 418.6633 | 2.28406 | 0.518736 | 4.403129 | 1.07E-05 | 0.00161 |
| PTET.51.1.G0730088 | 122.5067 | 2.279981 | 0.547195 | 4.166669 | 3.09E-05 | 0.003834 |
| PTET.51.1.G0550130 | 4679.534 | 2.26818 | 0.435724 | 5.205542 | 1.93E-07 | 7.62E-05 |
| PTET.51.1.G0360089 | 198.4496 | 2.267642 | 0.509254 | 4.452871 | 8.47E-06 | 0.001338 |
| PTET.51.1.G0120071 | 729.7377 | 2.26284 | 0.46792 | 4.835954 | 1.33E-06 | 0.000326 |
| PTET.51.1.G0900199 | 641.9858 | 2.262316 | 0.472568 | 4.787284 | 1.69E-06 | 0.000383 |
| PGML1 | 496.6463 | 2.255236 | 0.50389 | 4.475652 | 7.62E-06 | 0.001238 |
| PTET.51.1.G0720026 | 277.3071 | 2.253309 | 0.460172 | 4.896661 | 9.75E-07 | 0.000265 |
| PTET.51.1.G0320284 | 228.7547 | 2.252531 | 0.504406 | 4.465711 | 7.98E-06 | 0.001279 |
| PTET.51.1.G0430206 | 471.3678 | 2.252493 | 0.446703 | 5.042483 | 4.60E-07 | 0.000147 |
| PTET.51.1.G0430205 | 282.4239 | 2.248925 | 0.484349 | 4.643195 | 3.43E-06 | 0.00068 |
| PTET.51.1.G0590028 | 72.74106 | 2.242661 | 0.565833 | 3.96347 | 7.39E-05 | 0.007779 |

| | | | | | | |
|---|---|---|---|---|---|---|
| PTET.51.1.G0460076 | 916.3596 | 2.242076 | 0.537508 | 4.171243 | 3.03E-05 | 0.003795 |
| PTET.51.1.G1040164 | 136.3426 | 2.240305 | 0.562013 | 3.986217 | 6.71E-05 | 0.00716 |
| PTET.51.1.G0030168 | 341.4735 | 2.232553 | 0.550964 | 4.052086 | 5.08E-05 | 0.005667 |
| PTET.51.1.G0370037 | 494.6388 | 2.230852 | 0.459642 | 4.853458 | 1.21E-06 | 0.00031 |
| PTET.51.1.G0990129 | 440.0647 | 2.2296 | 0.50411 | 4.422849 | 9.74E-06 | 0.00149 |
| PTET.51.1.G0350295 | 849.363 | 2.224485 | 0.564676 | 3.939399 | 8.17E-05 | 0.008392 |
| PTET.51.1.G1080182 | 546.8664 | 2.212443 | 0.493962 | 4.478976 | 7.50E-06 | 0.001225 |
| PTET.51.1.G1000101 | 351.9914 | 2.211109 | 0.535955 | 4.125552 | 3.70E-05 | 0.004479 |
| PTET.51.1.G0480017 | 223.4175 | 2.192723 | 0.511163 | 4.289675 | 1.79E-05 | 0.002456 |
| PTET.51.1.G0900069 | 245.2465 | 2.189541 | 0.536429 | 4.081694 | 4.47E-05 | 0.005181 |
| PTET.51.1.G0950081 | 2961.64 | 2.18286 | 0.43299 | 5.041365 | 4.62E-07 | 0.000147 |
| PTET.51.1.G0030393 | 1942.252 | 2.180625 | 0.523033 | 4.169195 | 3.06E-05 | 0.003806 |
| PTET.51.1.G0540171 | 359.3771 | 2.178693 | 0.542121 | 4.018832 | 5.85E-05 | 0.006422 |
| PTET.51.1.G1830010 | 180.3891 | 2.17165 | 0.553591 | 3.92284 | 8.75E-05 | 0.008881 |
| PTET.51.1.G0380169 | 560.3811 | 2.170141 | 0.471084 | 4.606692 | 4.09E-06 | 0.00077 |
| PTET.51.1.G1300152 | 250.7813 | 2.169912 | 0.535997 | 4.048368 | 5.16E-05 | 0.005738 |
| PTET.51.1.G0120304 | 501.0776 | 2.168201 | 0.518348 | 4.182908 | 2.88E-05 | 0.003651 |
| PTET.51.1.G0320285 | 388.4793 | 2.164238 | 0.474714 | 4.559039 | 5.14E-06 | 0.000925 |
| PTET.51.1.G1460041 | 939.697 | 2.161561 | 0.486287 | 4.445031 | 8.79E-06 | 0.001375 |
| PTET.51.1.G1720081 | 385.1844 | 2.160682 | 0.5038 | 4.288766 | 1.80E-05 | 0.002456 |
| PTET.51.1.G0180124 | 191.9209 | 2.158871 | 0.484721 | 4.453839 | 8.43E-06 | 0.001338 |
| PTET.51.1.G1650089 | 282.0665 | 2.152482 | 0.534061 | 4.030406 | 5.57E-05 | 0.006134 |
| PTET.51.1.G1560077 | 187.3277 | 2.139719 | 0.492396 | 4.345526 | 1.39E-05 | 0.002006 |
| EZL4 | 82.91756 | 2.138917 | 0.54733 | 3.907912 | 9.31E-05 | 0.009391 |
| PTET.51.1.G1150123 | 614.2262 | 2.137212 | 0.472973 | 4.518674 | 6.22E-06 | 0.001074 |
| PTET.51.1.G0650095 | 1181.983 | 2.136757 | 0.479234 | 4.458697 | 8.25E-06 | 0.001315 |
| PTET.51.1.G1170173 | 136.9032 | 2.134201 | 0.540341 | 3.949732 | 7.82E-05 | 0.008087 |
| PTET.51.1.G0750138 | 1195.485 | 2.133276 | 0.444233 | 4.802155 | 1.57E-06 | 0.000363 |
| PTET.51.1.G0160206 | 189.8574 | 2.132465 | 0.542708 | 3.929306 | 8.52E-05 | 0.008672 |
| PTET.51.1.G0710091 | 270.7601 | 2.131505 | 0.505028 | 4.220569 | 2.44E-05 | 0.003151 |
| PTET.51.1.G0080258 | 363.179 | 2.130887 | 0.489716 | 4.351269 | 1.35E-05 | 0.001962 |
| PTET.51.1.G0800198 | 7233.238 | 2.119321 | 0.521089 | 4.067096 | 4.76E-05 | 0.005437 |
| PTET.51.1.G0930087 | 653.1718 | 2.119229 | 0.474088 | 4.470113 | 7.82E-06 | 0.001259 |
| PTET.51.1.G0680147 | 231.4977 | 2.118286 | 0.482961 | 4.386038 | 1.15E-05 | 0.001718 |
| PTET.51.1.G0560236 | 219.31 | 2.108102 | 0.532219 | 3.960969 | 7.46E-05 | 0.007812 |
| PTET.51.1.G0120102 | 228.5704 | 2.108008 | 0.517222 | 4.075638 | 4.59E-05 | 0.005281 |

| | | | | | | |
|---|---|---|---|---|---|---|
| PTET.51.1.G0630048 | 366.4367 | 2.103531 | 0.525748 | 4.001027 | 6.31E-05 | 0.006858 |
| PTET.51.1.G0710159 | 304.3996 | 2.0932 | 0.487561 | 4.293209 | 1.76E-05 | 0.002437 |
| PTET.51.1.G0560087 | 471.2201 | 2.091788 | 0.45269 | 4.620791 | 3.82E-06 | 0.00074 |
| PTET.51.1.G1120122 | 258.908 | 2.080446 | 0.47328 | 4.395801 | 1.10E-05 | 0.00165 |
| PTET.51.1.G1670032 | 192813.1 | 2.074888 | 0.511053 | 4.060025 | 4.91E-05 | 0.005514 |
| PTET.51.1.G0730049 | 329.3651 | 2.073011 | 0.485438 | 4.27039 | 1.95E-05 | 0.002614 |
| PTET.51.1.G0490131 | 118.5305 | 2.072974 | 0.502517 | 4.125178 | 3.70E-05 | 0.004479 |
| PTET.51.1.G0910136 | 264.4539 | 2.069992 | 0.477622 | 4.333953 | 1.46E-05 | 0.002078 |
| mtGb | 747.1749 | 2.052756 | 0.499064 | 4.113213 | 3.90E-05 | 0.004634 |
| PTET.51.1.G0090337 | 1038.254 | 2.046414 | 0.50969 | 4.015018 | 5.94E-05 | 0.006506 |
| PTET.51.1.G0930132 | 249.2437 | 2.042952 | 0.478296 | 4.271313 | 1.94E-05 | 0.002614 |
| PTET.51.1.G0740177 | 1565.875 | 2.036402 | 0.469124 | 4.340861 | 1.42E-05 | 0.00204 |
| PTET.51.1.G1900001 | 274.9885 | 2.035641 | 0.47904 | 4.249415 | 2.14E-05 | 0.002826 |
| PTET.51.1.G0980174 | 226.7236 | 2.029533 | 0.483448 | 4.198038 | 2.69E-05 | 0.003455 |
| PTET.51.1.G0070099 | 1279.256 | 2.024488 | 0.456012 | 4.439553 | 9.01E-06 | 0.001404 |
| PTET.51.1.G0480186 | 335.04 | 2.021889 | 0.478517 | 4.225322 | 2.39E-05 | 0.003097 |
| PTET.51.1.G1640109 | 567.4071 | 2.020518 | 0.497368 | 4.062423 | 4.86E-05 | 0.005476 |
| PTET.51.1.G0230268 | 868.5775 | 2.016391 | 0.471363 | 4.277786 | 1.89E-05 | 0.002549 |
| PTET.51.1.G1210101 | 138.9781 | 2.007073 | 0.503484 | 3.986373 | 6.71E-05 | 0.00716 |
| PTET.51.1.G0160137 | 121.1726 | -2.01259 | 0.482558 | -4.17067 | 3.04E-05 | 0.003795 |
| PTET.51.1.G1330058 | 413.0646 | -2.05203 | 0.47826 | -4.29061 | 1.78E-05 | 0.002456 |
| PTET.51.1.G1470167 | 398.0095 | -2.05435 | 0.474002 | -4.33404 | 1.46E-05 | 0.002078 |
| PTET.51.1.G0260144 | 429.584 | -2.08087 | 0.47325 | -4.39697 | 1.10E-05 | 0.001649 |
| PTET.51.1.G0620217 | 127.9542 | -2.1171 | 0.488349 | -4.33522 | 1.46E-05 | 0.002078 |
| PTET.51.1.G0250252 | 247.7278 | -2.12006 | 0.476252 | -4.45156 | 8.52E-06 | 0.00134 |
| PTET.51.1.G0800029 | 234.6636 | -2.15313 | 0.527023 | -4.08546 | 4.40E-05 | 0.005115 |
| PTET.51.1.G1070064 | 1072.941 | -2.16781 | 0.479905 | -4.51718 | 6.27E-06 | 0.001076 |
| PTET.51.1.G0230099 | 205.8165 | -2.18512 | 0.48712 | -4.4858 | 7.26E-06 | 0.001222 |
| PTET.51.1.G0090176 | 771.705 | -2.18894 | 0.468603 | -4.6712 | 2.99E-06 | 0.000601 |
| PTET.51.1.G0850039 | 401.906 | -2.2259 | 0.570228 | -3.90352 | 9.48E-05 | 0.009506 |
| PTET.51.1.G0170227 | 236.0226 | -2.2955 | 0.506151 | -4.5352 | 5.75E-06 | 0.001009 |
| PTET.51.1.G1160096 | 275.9696 | -2.30286 | 0.521351 | -4.41711 | 1.00E-05 | 0.001523 |
| PTET.51.1.G0970001 | 51.83007 | -2.37503 | 0.587607 | -4.04186 | 5.30E-05 | 0.00588 |
| PTET.51.1.G0590129 | 84.6369 | -2.38701 | 0.523011 | -4.56397 | 5.02E-06 | 0.000909 |
| PTET.51.1.G0080386 | 59.27146 | -2.38891 | 0.546799 | -4.3689 | 1.25E-05 | 0.001847 |
| PTET.51.1.G0580014 | 354.508 | -2.41508 | 0.578287 | -4.17626 | 2.96E-05 | 0.003745 |

| | | | | | | |
|---|---|---|---|---|---|---|
| PTET.51.1.G0880185 | 129.606 | -2.43207 | 0.507082 | -4.79619 | 1.62E-06 | 0.000371 |
| PTET.51.1.G0930099 | 56.63102 | -2.46537 | 0.548413 | -4.49547 | 6.94E-06 | 0.001186 |
| PTET.51.1.G0240243 | 636.4707 | -2.48265 | 0.528552 | -4.69708 | 2.64E-06 | 0.000539 |
| PTET.51.1.G1650093 | 255.3306 | -2.51486 | 0.505706 | -4.97296 | 6.59E-07 | 0.000193 |
| PTET.51.1.G1650045 | 45.56399 | -2.51798 | 0.58523 | -4.30254 | 1.69E-05 | 0.002366 |
| PTET.51.1.G0750147 | 133.806 | -2.56041 | 0.600603 | -4.26306 | 2.02E-05 | 0.00268 |
| PTET.51.1.G0970152 | 180.4474 | -2.56528 | 0.512942 | -5.00111 | 5.70E-07 | 0.000174 |
| PTET.51.1.G0520074 | 156.2173 | -2.59246 | 0.63399 | -4.08912 | 4.33E-05 | 0.005053 |
| PTET.51.1.G0760223 | 99.27008 | -2.62027 | 0.503869 | -5.2003 | 1.99E-07 | 7.70E-05 |
| PTET.51.1.G0870018 | 50.07805 | -2.6512 | 0.664547 | -3.98949 | 6.62E-05 | 0.007131 |
| PTET.51.1.G1540003 | 76.45525 | -2.65376 | 0.553512 | -4.79441 | 1.63E-06 | 0.000372 |
| PTET.51.1.G0820146 | 96.50806 | -2.7874 | 0.665356 | -4.18934 | 2.80E-05 | 0.003576 |
| PTET.51.1.G0290167 | 769.3246 | -2.79062 | 0.535886 | -5.2075 | 1.91E-07 | 7.62E-05 |
| PTET.51.1.G0160050 | 53.28775 | -2.84703 | 0.652296 | -4.36464 | 1.27E-05 | 0.00187 |
| PTET.51.1.G0020281 | 145.9975 | -2.88207 | 0.504211 | -5.716 | 1.09E-08 | 6.49E-06 |
| PTET.51.1.G0010175 | 145.4726 | -2.88557 | 0.599856 | -4.81043 | 1.51E-06 | 0.000358 |
| PTET.51.1.G0050186 | 60.00411 | -2.89247 | 0.653542 | -4.42584 | 9.61E-06 | 0.001479 |
| PTET.51.1.G0020367 | 36.78415 | -2.95233 | 0.617151 | -4.7838 | 1.72E-06 | 0.000387 |
| PTET.51.1.G0600108 | 2013.863 | -3.00222 | 0.752746 | -3.98835 | 6.65E-05 | 0.007142 |
| PTET.51.1.G1210173 | 22.04244 | -3.07927 | 0.719033 | -4.28252 | 1.85E-05 | 0.002506 |
| PTET.51.1.G0690122 | 28.35534 | -3.12044 | 0.789109 | -3.95438 | 7.67E-05 | 0.007956 |
| PTET.51.1.G1270004 | 74.27108 | -3.21308 | 0.697039 | -4.60961 | 4.03E-06 | 0.00077 |
| PTET.51.1.G1410005 | 25.50615 | -3.21889 | 0.760336 | -4.23351 | 2.30E-05 | 0.00301 |
| PTET.51.1.G4980001 | 54.2862 | -3.50393 | 0.590704 | -5.93178 | 3.00E-09 | 2.51E-06 |
| PTET.51.1.G0650058 | 574.3621 | -3.58073 | 0.500155 | -7.15925 | 8.11E-13 | 1.29E-09 |
| PTET.51.1.G0690117 | 1159.077 | -3.59992 | 0.534368 | -6.73678 | 1.62E-11 | 2.36E-08 |
| PTET.51.1.G0760177 | 87.57607 | -3.65167 | 0.579859 | -6.2975 | 3.02E-10 | 2.98E-07 |
| PTET.51.1.G0230294 | 19.93125 | -3.79325 | 0.793167 | -4.78241 | 1.73E-06 | 0.000387 |
| PTET.51.1.G0120225 | 16.4522 | -4.04168 | 0.926857 | -4.36063 | 1.30E-05 | 0.001897 |
| PTET.51.1.G0400056 | 86.44225 | -4.10707 | 0.631697 | -6.50165 | 7.94E-11 | 9.50E-08 |
| PTET.51.1.G1540004 | 14.93137 | -5.021 | 1.097103 | -4.5766 | 4.73E-06 | 0.000865 |
| PTIWI10 | 2443.73 | -5.15723 | 0.578201 | -8.91943 | 4.69E-19 | 2.62E-15 |
| PTIWI11 | 1734.302 | -5.21111 | 0.447931 | -11.6337 | 2.78E-31 | 4.65E-27 |
| PTET.51.1.G0710033 | 81.41969 | -5.49578 | 0.692059 | -7.9412 | 2.00E-15 | 6.10E-12 |
| PTIWI07 | 126.6251 | -5.92237 | 0.695432 | -8.5161 | 1.65E-17 | 7.89E-14 |
| PTIWI06 | 998.904 | -6.12354 | 0.525005 | -11.6638 | 1.95E-31 | 4.65E-27 |

| | | | | | | |
|---|---|---|---|---|---|---|
| PTET.51.1.G4270003 | 24.03736 | -8.21054 | 1.342599 | -6.11541 | 9.63E-10 | 8.96E-07 |
| Ptcen_icl9d | 83.32795 | -10.0037 | 1.295224 | -7.7235 | 1.13E-14 | 2.71E-11 |

Supplementary Table S2. Differential expression of proteins identified by mass spectrometry from SMC4-1 and SMC4-2 Co-IP experiments at late time point.

| Leading_Protein | Synonyms | Description | Score | OnOff WT vs SMC4_1 | OnOff WT vs SMC4_2 | OnOff SMC4_1 vs SMC4_2 | Top3 log2FC SMC4_1 - WT | Top3 log2FC SMC4_2 - WT | Top3 log2FC SMC4_2 - SMC4_1 |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0010184 | PTMB.323, PTETG100077001 | Nucleic acid-binding, OB-fold | 17.419 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0010254 | PTMB.266 | Surfeit locus 1/Shy1 | 31.621 | SMC4_1 | SMC4_2 | | | | -2.452939601 |
| PTET.51.1.P0010279 | PTMB.248c, PTETG100038001 | P-loop containing nucleoside triphosphate hydrolase | 61.028 | | | | 2.107835054 | 1.104759802 | -1.003075252 |
| PTET.51.1.P0010297 | PTMB.233 | Rossmann-like alpha/beta/alpha sandwich fold | 31.245 | | | | -1.636188966 | -0.252085296 | 1.384103669 |
| PTET.51.1.P0010326 | PTMB.208 | UvrABC system protein C | 36.981 | SMC4_1 | SMC4_2 | | | | -0.178349541 |
| PTET.51.1.P0010348 | mag3, PTMB.190c | G surface protein, allelic form 156 | 21.388 | | | | 1.989942861 | 3.963007169 | 1.973064308 |
| PTET.51.1.P0010379 | PTMB.166c, PTETG100158001 | tRNA methyltransferase complex GCD14 subunit | 51.496 | | | | 1.955001852 | 2.71942523 | 0.764423378 |
| PTET.51.1.P0010479 | PTMB.81, PTETG100101001 | 2-oxoglutarate dehydrogenase E1 component | 55.85 | | WT | SMC4_1 | 2.891235338 | | |
| PTET.51.1.P0010562 | PTMB.14c, PTETG100074001 | GrpE | 18.65 | SMC4_1 | SMC4_2 | | | | -1.193046312 |
| PTET.51.1.P0010572 | PTETG100018001, PTMB.08c | Fibrinogen alpha/beta chain family | 18.024 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0020102 | | DNA/RNA-binding protein Alba-like | 23.593 | | | | -1.561043039 | -1.818658142 | -0.257615102 |
| PTET.51.1.P0020184 | | Calcineurin-like phosphoesterase | 55.323 | | | | 1.608248666 | 1.454068232 | -0.154180435 |
| PTET.51.1.P0020273 | | Carbon-nitrogen hydrolase | 18.822 | SMC4_1 | SMC4_2 | | | | 0.717268812 |
| PTET.51.1.P0020350 | | TLDc domain | 32.144 | SMC4_1 | SMC4_2 | | | | -3.875236638 |
| PTET.51.1.P0020475 | | V-type ATP synthase subunit E | 14.041 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0030003 | | Isopenicillin N synthase-like | 93.124 | | | | 0.446955028 | 0.516986776 | 0.070031748 |
| PTET.51.1.P0030017 | | Ulp1 protease family, C-terminal catalytic domain | 26.117 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0030065 | | Nucleoside diphosphate kinase B | 25.866 | | | | -1.460048189 | 0.317893291 | 1.777941479 |
| PTET.51.1.P0030131 | | Protein kinase domain | 38.952 | | | | 0.390049091 | -1.110009578 | -1.500058669 |
| PTET.51.1.P0030193 | | Eukaryotic initiation factor 3, gamma subunit | 11.329 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0030205 | | NADP transhydrogenase, beta subunit | 40.143 | | | | 1.561873229 | 0.381939853 | -1.179933376 |
| PTET.51.1.P0030248 | | S-adenosyl-L-methionine-dependent methyltransferase-like | 33.344 | SMC4_1 | SMC4_2 | | | | -1.196441278 |
| PTET.51.1.P0030254 | | | 13.629 | SMC4_1 | SMC4_2 | | | | -2.672132695 |
| PTET.51.1.P0030381 | | Protein transport protein Sec23B | 26.467 | SMC4_1 | SMC4_2 | | | | 0.331888012 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0030393 | PTETG300037001 | Transcription elongation factor Spt6 | 106.78 | | | | 0.280491455 | -0.734596941 | -1.015088396 |
| PTET.51.1.P0040135 | | Peptidase M16 inactive domain | 76.731 | | | | 2.842187056 | 3.161357131 | 0.319170074 |
| PTET.51.1.P0040137 | | Nascent polypeptide-associated complex subunit alpha | 34.396 | | | | 0.434630979 | 0.206621515 | -0.228009464 |
| PTET.51.1.P0040233 | PTETG400003001 | Biotin synthase | 35.47 | SMC4_1 | SMC4_2 | | | | -0.38740496 |
| PTET.51.1.P0040314 | | Alpha/beta hydrolase family | 12.944 | SMC4_1 | SMC4_2 | | | | -1.433886732 |
| PTET.51.1.P0040332 | | tRNA synthetases class I (W and Y) | 79.117 | | | | 0.180198757 | 0.483183437 | 0.30298468 |
| PTET.51.1.P0040350 | | Peptidase C1-like family | 12.816 | WT | WT | | | | |
| PTET.51.1.P0050055 | | Aldehyde dehydrogenase family | 30.504 | SMC4_1 | SMC4_2 | | | | -2.158637257 |
| PTET.51.1.P0050154 | | Cyclic nucleotide-binding domain | 23.366 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0050179 | | WD domain, G-beta repeat | 73.8 | | | | 0.156462689 | 0.550973586 | 0.394510897 |
| PTET.51.1.P0050419 | | Coiled coil domain | 11.82 | WT | WT | | | | |
| PTET.51.1.P0060018 | | Nop domain | 18.445 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0060034 | PTIWI06 | Piwi-like protein 1 | 24.925 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0060144 | PTETG600028001 | Chloramphenicol acetyltransferase-like domain | 41.118 | | WT | SMC4_1 | -2.324672857 | | |
| PTET.51.1.P0060199 | | YTH domain-containing protein ECT3 | 13.86 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0060205 | | High mobility group nucleosome-binding domain-containing protein 5 | 12.133 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0060411 | | Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform | 19.92 | | | | 1.162876546 | -0.115876638 | -1.278753184 |
| PTET.51.1.P0070001 | | Isopenicillin N synthase-like | 59.826 | | | | 0.233963125 | -0.22944554 | -0.463408665 |
| PTET.51.1.P0070020 | PTETG700003001 | Ubiquitin-conjugating enzyme/RWD-like | 24.448 | | | | 0.25256362 | -1.30899206 | -1.56155568 |
| PTET.51.1.P0070092 | | Glycylpeptide N-tetradecanoyltransferase 2 | 44.888 | | | | -0.72925668 | -2.15732847 | -1.42807179 |
| PTET.51.1.P0070145 | | Coiled coil domain | 11.286 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0070153 | | Ribosome biogenesis protein BRX1 | 18.028 | WT | | SMC4_2 | | 1.584180061 | |
| PTET.51.1.P0070340 | | S-adenosylmethionine synthetase | 13.921 | | | | 1.255192661 | 0.841024053 | -0.414168608 |
| PTET.51.1.P0070404 | | Mitochondrial substrate/solute carrier | 54.414 | | | | 6.556738303 | 4.561035413 | -1.99570289 |
| PTET.51.1.P0080015 | | Glycosyl hydrolases family 25 | 25.693 | SMC4_1 | SMC4_2 | | | | -1.489272596 |
| PTET.51.1.P0080040 | | Pyrroline-5-carboxylate reductase 1, mitochondrial | 40.169 | | | | 1.791039032 | 1.270707694 | -0.520331339 |
| PTET.51.1.P0080054 | | Metallo-dependent phosphatase-like | 20.565 | | | | 0.432367747 | -0.150787319 | -0.583155066 |
| PTET.51.1.P0080095 | | Prohibitin | 30.648 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0080169 | | Putative purine permease 20 | 11.397 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0080174 | | Chaperonin Cpn60/TCP-1 | 82.126 | | | | -0.383829878 | -0.300297405 | 0.083532473 |
| PTET.51.1.P0080218 | | Citrate synthase, C-terminal domain | 14.147 | SMC4_1 | SMC4_2 | | | | 1.233861029 |
| PTET.51.1.P0080260 | PTETG800018001 | Ubiquitin-conjugating enzyme | 23.617 | | | | 1.009567552 | 1.537666678 | 0.528099126 |
| PTET.51.1.P0080295 | | Succinate dehydrogenase | 73.57 | | | | 1.889751223 | 3.001087723 | 1.1113365 |
| PTET.51.1.P0080397 | PTETG800045001 | 2-oxoglutarate dehydrogenase, mitochondrial | 12.771 | SMC4_1 | | SMC4_1 | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0080438 | | Lipoyl synthase, chloroplastic | 19.087 | SMC4_1 | SMC4_2 | | | | -1.061753821 |
| PTET.51.1.P0090053 | | 40S ribosomal protein SA | 13.042 | | | | -0.448734959 | 0.981609103 | 1.430344062 |
| PTET.51.1.P0090086 | | ATPase, V1/A1 complex, subunit E | 34.584 | | | | 1.521022788 | 1.639750351 | 0.118727563 |
| PTET.51.1.P0090113 | | Translation elongation factor IF5A | 47.885 | | | | -0.034150444 | 0.07714448 | 0.111294923 |
| PTET.51.1.P0090195 | | Alpha/Beta hydrolase fold | 17.436 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0090213 | PTETG900022001 | Spliceosome-associated protein 130 B | 154.53 | | | | -0.808181324 | 0.795713227 | 1.603894551 |
| PTET.51.1.P0090252 | | Structural maintenance of chromosomes protein 1 | 18.356 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P0090255 | | Vacuolar protein sorting-associated protein 26 | 11.81 | | WT | SMC4_1 | -0.155609699 | | |
| PTET.51.1.P0100002 | | Replication protein A 70 kDa DNA-binding subunit | 30.691 | | | | 0.590355933 | -0.61280584 | -1.203161774 |
| PTET.51.1.P0100009 | | Putative F-box protein At3g21170 | 251.07 | | | | -0.130721056 | -1.629018838 | -1.498297783 |
| PTET.51.1.P0100011 | | Coiled coil domain | 323.31 | | | | -0.405088547 | -1.22126825 | -0.816179703 |
| PTET.51.1.P0100050 | | Endoplasmic reticulum oxidoreductin 1 | 11.35 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0100097 | PTETG1000030001 | Adenosylhomocysteinase | 100.08 | | | | 0.042660057 | 2.851257011 | 2.808596954 |
| PTET.51.1.P0100106 | | Ribosomal protein L3 | 36.793 | | | | 0.702407944 | 1.131395661 | 0.428987717 |
| PTET.51.1.P0100128 | | Regulator of chromosome condensation 1/beta-lactamase-inhibitor protein II | 77.806 | SMC4_1 | SMC4_2 | | | | -1.606315422 |
| PTET.51.1.P0100202 | | Probable enoyl-CoA hydratase echA8 | 25.29 | | | | -1.126928877 | -0.101745669 | 1.025183208 |
| PTET.51.1.P0100252 | PTETG1000006001 | Cytochrome c oxidase, subunit Vb | 18.668 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0100278 | PTETG1000002001 | Enolase | 123.51 | | | | -0.337512013 | -1.189446027 | -0.851934013 |
| PTET.51.1.P0100395 | | MORN motif | 13.544 | | WT | SMC4_1 | 0.759480328 | | |
| PTET.51.1.P0110022 | | Persulfide dioxygenase ETHE1, mitochondrial | 30.063 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0110028 | | RuvB-like | 18.679 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0110036 | | Armadillo-like helical | 112.23 | | | | -0.014427818 | -0.041542256 | -0.027114438 |
| PTET.51.1.P0110039 | | 60S ribosomal protein L10 | 62.998 | | | | 4.552670661 | 3.599627459 | -0.953043202 |
| PTET.51.1.P0110119 | | Nucleotide-binding, alpha-beta plait | 12.509 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0110137 | PTETG1100001001 | Thioredoxin domain-containing protein 3 homolog | 11.612 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0110162 | | Calmodulin binding protein-like | 12.439 | | WT | SMC4_1 | -1.903779238 | | |
| PTET.51.1.P0110239 | | Uncharacterized protein YfdF | 12.606 | SMC4_1 | SMC4_2 | | | | -0.792907888 |
| PTET.51.1.P0110278 | | Succinyl-CoA ligase, alpha subunit | 51.232 | | | | 0.352119309 | 0.335368303 | -0.016751006 |
| PTET.51.1.P0110289 | PIE2, PDSG2 | Elongation factor Ts | 46.202 | | | | 1.134453695 | -1.622791075 | -2.75724477 |
| PTET.51.1.P0110362 | | Citrate synthase-like | 69.304 | | | | 1.493807884 | 1.834998943 | 0.341191059 |
| PTET.51.1.P0110390 | | Protein kinase domain | 11.475 | SMC4_1 | SMC4_2 | | | | 0.822604204 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0110429 | | ATPase, F1 complex, delta/epsilon subunit, N-terminal | 13.179 | | | | -0.746699668 | -1.093507466 | -0.346807798 |
| PTET.51.1.P0120057 | | Protein kinase domain | 17.373 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0120086 | | T-complex protein 1 subunit beta | 131.92 | | | | -0.109328783 | -0.044042031 | 0.065286752 |
| PTET.51.1.P0120113 | | Alpha/beta hydrolase domain-containing protein 11 | 18.855 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0120119 | | Protein max | 13.123 | SMC4_1 | SMC4_2 | | | | 0.505966603 |
| PTET.51.1.P0120205 | | Macroglobulin domain MG4 | 243.43 | SMC4_1 | SMC4_2 | | | | 0.988791886 |
| PTET.51.1.P0120274 | | DEAD-box ATP-dependent RNA helicase 15 | 45.209 | | | | 2.911888405 | 2.891835029 | -0.020053376 |
| PTET.51.1.P0120277 | | Mitochondrial 2-oxoglutarate/malate carrier protein | 72.134 | | | | 0.849335348 | -0.442263356 | -1.291598705 |
| PTET.51.1.P0120293 | | Molybdopterin oxidoreductase | 25.308 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0120294 | | NAC domain | 19.81 | | | | -1.476503445 | -1.720539396 | -0.244035951 |
| PTET.51.1.P0120377 | | PHD-zinc-finger like domain | 12.382 | | WT | SMC4_1 | 1.96136759 | | |
| PTET.51.1.P0130008 | | Protein STIP1 homolog | 18.659 | SMC4_1 | SMC4_2 | | | | 0.502132381 |
| PTET.51.1.P0130113 | | MT-A70 | 11.673 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0130197 | | Adenylate kinase | 24.213 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0130217 | PTETG1300015001 | HEAT repeats | 43.859 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0130319 | | Protein piccolo | 17.033 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0130320 | | Mannose-binding lectin | 20.42 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0140021 | | Citrate synthase-like | 38.823 | | WT | SMC4_1 | 1.336657789 | | |
| PTET.51.1.P0140093 | | Mating-type locus allele B6 protein | 18.697 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0140342 | arp2-1, PTETG1400002001 | Actin-related protein | 46.625 | | | | -0.111035807 | 0.912499468 | 1.023535275 |
| PTET.51.1.P0150014 | | V-type ATP synthase subunit I | 13.147 | | WT | SMC4_1 | -2.497668145 | | |
| PTET.51.1.P0150066 | | ATP-NAD kinase-like domain | 19.914 | | | | 0.00894269 | 0.147573698 | 0.138631008 |
| PTET.51.1.P0150089 | | Serine--tRNA ligase, cytoplasmic | 70.172 | SMC4_1 | SMC4_2 | | | | 0.252865885 |
| PTET.51.1.P0150105 | | DNA/RNA-binding protein Alba-like | 11.446 | | | | -2.650666526 | -3.797327417 | -1.146660891 |
| PTET.51.1.P0150119 | | Ribosomal protein L6e | 17.404 | WT | | SMC4_2 | | 2.871930347 | |
| PTET.51.1.P0150223 | | Probable disease resistance protein At4g19520 | 25.131 | | | | 0.372610564 | -1.985499086 | -2.35810965 |
| PTET.51.1.P0150275 | | Type III restriction enzyme, res subunit | 11.555 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0150343 | | ClpP/crotonase-like domain | 49.709 | | | | -0.838646188 | -0.501055949 | 0.337590239 |
| PTET.51.1.P0160023 | | Ribosomal protein S10 | 14.546 | | | | 1.526874572 | 1.913528156 | 0.386653584 |
| PTET.51.1.P0160032 | | Ribosomal protein L30, ferredoxin-like fold domain | 33.107 | | | | -0.563672687 | -0.520203266 | 0.04346942 |
| PTET.51.1.P0160123 | | WD domain, G-beta repeat | 12.564 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0160139 | | Metallopeptidase family M24 | 11.914 | | | | -0.64291707 | 0.231691677 | 0.874608747 |
| PTET.51.1.P0160163 | | Phospholipid-transporting ATPase IG | 13.741 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0160240 | | Ultraviolet-B receptor UVR8 | 21.819 | SMC4_1 | SMC4_2 | | | | -2.031172549 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0160244 | | Protein disulfide-isomerase A6 | 30.408 | | WT | SMC4_1 | 1.886979141 | | |
| PTET.51.1.P0160330 | | Putative mitochondrial 2-oxoglutarate/malate carrier protein | 28.014 | SMC4_1 | SMC4_2 | | | | -2.866191739 |
| PTET.51.1.P0160340 | | Coiled-coil domain-containing protein 18 | 72.105 | | | | -0.967408531 | -0.119216034 | 0.848192497 |
| PTET.51.1.P0160357 | | Coiled coil domain | 323.31 | | | | -1.0774546 | -1.536090578 | -0.458635978 |
| PTET.51.1.P0170077 | NOWA1 | Coiled coil domain | 19.672 | SMC4_1 | SMC4_2 | | | | -1.274407887 |
| PTET.51.1.P0170151 | PTETG1700002001 | GDP dissociation inhibitor | 33.656 | | | | 1.383357078 | 2.849518574 | 1.466161497 |
| PTET.51.1.P0170184 | | Polyribonucleotide nucleotidyltransferase | 32.942 | | | | 1.684678508 | 0.124522539 | -1.560155969 |
| PTET.51.1.P0170225 | | Baculoviral IAP repeat-containing protein 6 | 39.148 | | | | 2.493843364 | 1.625314876 | -0.868528489 |
| PTET.51.1.P0170257 | | Protein of unknown function (DUF541) | 189.41 | | | | 0.667158036 | 0.363964075 | -0.303193961 |
| PTET.51.1.P0190251 | | Rhodanese-like domain | 17.345 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0200201 | KdB2 | Cysteine synthase 2 | 82.37 | | | | -0.648948092 | 2.767185277 | 3.416133369 |
| PTET.51.1.P0200315 | | 50S ribosomal protein L11 | 17.121 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0200334 | | Methylmalonate-semialdehyde dehydrogenase | 15.433 | | | | 0.729894429 | -0.419585787 | -1.149480216 |
| PTET.51.1.P0210034 | | Ribosomal protein L10/acidic P0 | 22.042 | SMC4_1 | SMC4_2 | | | | -0.33495953 |
| PTET.51.1.P0210041 | | D-isomer specific 2-hydroxyacid dehydrogenase, catalytic domain | 62.109 | | | | 1.590870709 | 1.766809321 | 0.175938612 |
| PTET.51.1.P0210175 | | AMP-binding enzyme | 37.426 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0210203 | | Ribosomal protein S7e | 13.419 | SMC4_1 | SMC4_2 | | | | -0.200620291 |
| PTET.51.1.P0210280 | PTETG2100008001 | Leishmanolysin-like peptidase | 138.25 | | | | 0.979620444 | -0.05456776 | -1.034188204 |
| PTET.51.1.P0210302 | | DNA-directed RNA polymerase subunit alpha | 44.1 | | | | 0.074318978 | -1.847364452 | -1.92168343 |
| PTET.51.1.P0220003 | | Gamma carbonic anhydrase 3, mitochondrial | 12.902 | | WT | SMC4_1 | 1.479151769 | | |
| PTET.51.1.P0220057 | | Protein phosphatase 2C (PP2C)-like domain | 26.306 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0220116 | | T-complex protein 1 subunit zeta | 74.619 | | | | -0.879052388 | -1.692869172 | -0.813816784 |
| PTET.51.1.P0220177 | PTETG2200025001 | Insulin-like growth factor binding protein, N-terminal | 18.965 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0220197 | | Elongation factor 1-gamma 2 | 26.289 | | | | -1.253582662 | -1.050572976 | 0.203009686 |
| PTET.51.1.P0220247 | | Methionine--tRNA ligase, cytoplasmic | 18.269 | SMC4_1 | SMC4_2 | | | | -0.412404028 |
| PTET.51.1.P0220269 | PPN2 | Metallo-dependent phosphatase-like | 78.135 | | | | 1.192614105 | 0.276595123 | -0.916018982 |
| PTET.51.1.P0220270 | | Mitochondrial carrier protein | 19.657 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0230160 | | Peptidase M16 inactive domain | 11.737 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0230214 | | AMP-binding enzyme | 27.086 | SMC4_1 | SMC4_2 | | | | 0.182906893 |
| PTET.51.1.P0230219 | | Paramecium surface antigen domain | 38.651 | SMC4_1 | SMC4_2 | | | | 3.824277118 |
| PTET.51.1.P0230221 | | DNA-directed RNA polymerase subunit beta | 20.079 | SMC4_1 | SMC4_2 | | | | -0.654581522 |
| PTET.51.1.P0230249 | | Chaperonin Cpn60/TCP-1 | 19.551 | SMC4_1 | SMC4_2 | | | | -1.216301179 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0230261 | PTETG2300007001 | Intraflagellar transport protein 74 homolog | 11.875 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0230286 | | Isocitrate/isopropylmalate dehydrogenase | 116.93 | | | | -0.631028527 | -0.149864854 | 0.481163673 |
| PTET.51.1.P0250050 | | Plectin/S10, N-terminal | 24.205 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0250100 | | Calcium-binding mitochondrial carrier protein SCaMC-1 | 49.545 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0250101 | | small GTPase Rab1 family profile. | 25.154 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0250236 | | Coiled coil domain | 104.09 | | | | 0.957156873 | 1.123070216 | 0.165913343 |
| PTET.51.1.P0250319 | | Ribosomal protein L32 | 11.239 | SMC4_1 | SMC4_2 | | | | -1.019550834 |
| PTET.51.1.P0260039 | | Ribosomal protein S2, eukaryotic/archaeal | 135.05 | | | | -0.649888062 | 1.229495822 | 1.879383884 |
| PTET.51.1.P0260174 | PTETG2600021001 | Spliceosome-associated protein 130 B | 25.664 | SMC4_1 | SMC4_2 | | | | 3.270369529 |
| PTET.51.1.P0270023 | | DNA polymerase epsilon subunit 2 | 11.846 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0270129 | | UvrABC system protein C | 82.496 | | | | 0.189550989 | 0.83097285 | 0.641421861 |
| PTET.51.1.P0270186 | | DnaJ protein homolog | 18.166 | SMC4_1 | SMC4_2 | | | | -2.499273707 |
| PTET.51.1.P0270209 | PTETG2700003001 | P-loop containing nucleoside triphosphate hydrolase | 287.66 | | | | 0.658647673 | 0.780559098 | 0.121911426 |
| PTET.51.1.P0270210 | | Uridine kinase-like protein 3 | 20.385 | | | | 1.087676047 | 0.854083649 | -0.233592398 |
| PTET.51.1.P0270239 | | Enoyl-CoA hydratase, mitochondrial | 46.129 | | | | 0.326029743 | 0.553293501 | 0.227263758 |
| PTET.51.1.P0280112 | | Porphobilinogen synthase | 11.91 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P0280283 | | Eukaryotic translation initiation factor 3 subunit I | 18.716 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0290214 | | Protein kinase-like domain | 11.654 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0300021 | | Fumarate lyase family | 198.11 | | | | -0.139729929 | -1.496342407 | -1.356612478 |
| PTET.51.1.P0300158 | | Leishmanolysin-like peptidase | 18.824 | SMC4_1 | SMC4_2 | | | | 0.350665289 |
| PTET.51.1.P0300222 | PTETG3000005001 | Parkin co-regulated protein | 17.969 | SMC4_1 | SMC4_2 | | | | -0.076671389 |
| PTET.51.1.P0300231 | | Calcium and integrin-binding family member 2 (Fragment) | 11.863 | SMC4_1 | SMC4_2 | | | | -0.72900787 |
| PTET.51.1.P0300272 | | Clathrin heavy chain 1 | 13.381 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0300291 | | Coiled coil domain | 71.174 | | | | 0.422381266 | 0.45692988 | 0.034548614 |
| PTET.51.1.P0310063 | | Proteasome component (PCI) domain | 11.76 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0310118 | | Aspartate/other aminotransferase | 50.932 | | | | -0.109647509 | -0.952719278 | -0.843071769 |
| PTET.51.1.P0310186 | | Peptidase C1A | 62.116 | | | | 0.43354445 | -2.896053969 | -3.329598419 |
| PTET.51.1.P0310192 | | 60S ribosomal protein L6E | 37.123 | | | | 0.58052149 | 0.313217362 | -0.267304128 |
| PTET.51.1.P0310216 | | Mitochondrial carrier protein | 13.157 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0320031 | | Carbonic anhydrase 2 | 11.286 | | | | -0.544950612 | -0.177351666 | 0.367598946 |
| PTET.51.1.P0320061 | | Tetratricopeptide repeat protein 7A | 56.745 | | | | 2.522862006 | 0.791100914 | -1.731761092 |
| PTET.51.1.P0320109 | | Paramecium surface antigen domain | 258 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P0320151 | act1_3 | Actin-related protein | 266.05 | | | | 2.351626061 | 2.722879407 | 0.371253346 |
| PTET.51.1.P0320296 | | Aldehyde dehydrogenase family | 200.25 | | | | 0.002931382 | -1.027449721 | -1.030381103 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0330075 | PtSMC2-1, SMC2 | P-loop containing nucleoside triphosphate hydrolase | 312.21 | | | | 6.06350107 | 6.558768196 | 0.495267126 |
| PTET.51.1.P0330130 | | ATP synthase subunit a | 12.069 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0330220 | Hsp70Pt01, PTETG3300001001 | Heat shock 70 kDa protein 5 | 57.713 | | | | 2.200872247 | 1.996549596 | -0.204322651 |
| PTET.51.1.P0330225 | | ATP synthase subunit alpha, mitochondrial | 225.99 | | | | 0.350177958 | -0.213219196 | -0.563397154 |
| PTET.51.1.P0330227 | | Pyruvate, phosphate dikinase | 11.033 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0330250 | | 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 4 | 18.119 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P0330264 | | Thiamin diphosphate-binding fold | 19.138 | | | | 0.720461985 | -0.628300069 | -1.348762054 |
| PTET.51.1.P0340161 | PTETG3400005001 | Aldehyde/histidinol dehydrogenase | 19.861 | | WT | SMC4_1 | 1.093478881 | | |
| PTET.51.1.P0340162 | PTETG3400002001 | EF-hand domain pair | 74.555 | | | | 0.032398142 | -0.074491662 | -0.106889804 |
| PTET.51.1.P0340170 | | Enoyl-CoA hydratase/isomerase | 20.492 | | | | -1.026576652 | -2.758310423 | -1.731733771 |
| PTET.51.1.P0340171 | | Armadillo-like helical | 40.073 | | | | 0.608533219 | -0.848699802 | -1.457233021 |
| PTET.51.1.P0340233 | | ATP-dependent 6-phosphofructokinase 7 | 12.872 | | WT | SMC4_1 | 1.154499788 | | |
| PTET.51.1.P0340238 | | Ribosomal protein S6e | 25.404 | SMC4_1 | SMC4_2 | | | | 0.302843197 |
| PTET.51.1.P0350129 | | Isocitrate and isopropylmalate dehydrogenases family | 23.909 | SMC4_1 | SMC4_2 | | | | -2.204176974 |
| PTET.51.1.P0350184 | | Tubby C-terminal-like domain | 27.508 | | | | -1.228503834 | -3.531520044 | -2.303016209 |
| PTET.51.1.P0350280 | | Sensory histidine kinase/phosphatase NtrB | 62.898 | | | | 0.994220215 | 1.083863865 | 0.08964365 |
| PTET.51.1.P0360019 | | Isoleucine--tRNA ligase, cytoplasmic | 19.017 | SMC4_1 | SMC4_2 | | | | 0.275004612 |
| PTET.51.1.P0360034 | | Delta-1-pyrroline-5-carboxylate dehydrogenase 12A1, mitochondrial | 11.487 | | | | 0.011311024 | -0.263889232 | -0.275200256 |
| PTET.51.1.P0360064 | | Acyl-CoA dehydrogenase/oxidase C-terminal | 11.996 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0360133 | | Coiled coil domain | 12.653 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0360184 | | GroEL-like equatorial domain | 12.133 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0360228 | | AIR synthase-related protein, C-terminal domain | 12.142 | WT | | SMC4_2 | | -0.257230755 | |
| PTET.51.1.P0370010 | | Cytochrome b-c1 complex subunit Rieske-2, mitochondrial | 43.425 | | | | 0.968822761 | 0.525989773 | -0.442832988 |
| PTET.51.1.P0370172 | | Isocitrate and isopropylmalate dehydrogenases family | 86.542 | | | | 1.957771629 | -1.381926136 | -3.339697765 |
| PTET.51.1.P0370298 | | Superoxide dismutase | 19.22 | | | | -0.911859892 | -3.801110611 | -2.889250719 |
| PTET.51.1.P0380053 | | High frequency lysogenization protein HflD homolog | 99.343 | SMC4_1 | SMC4_2 | | | | -1.152086789 |
| PTET.51.1.P0380073 | PGML2 | Transposase IS4 | 17.832 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0380195 | GAPDH | Glyceraldehyde/Erythrose phosphate dehydrogenase family | 162.46 | | | | 2.265369893 | 1.99806137 | -0.267308523 |
| PTET.51.1.P0380223 | | Chromo (CHRromatin Organisation MOdifier) domain | 55.977 | SMC4_1 | SMC4_2 | | | | -0.741023612 |
| PTET.51.1.P0380286 | | Pentatricopeptide repeat-containing protein At3g29230 | 76.354 | SMC4_1 | SMC4_2 | | | | -1.02924861 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0390035 | PTETG3900002001 | Cytochrome c oxidase subunit 1+2 | 12.741 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P0400049 | Hsp70Pt07, PTETG4000011001 | Heat shock protein 70 family | 180.93 | | | | 1.36269449 | 1.892984613 | 0.530290123 |
| PTET.51.1.P0400094 | | o-succinylbenzoate synthase | 53.822 | | | | 0.628905735 | 0.170428294 | -0.458477441 |
| PTET.51.1.P0400246 | | Coiled coil domain | 25.082 | SMC4_1 | SMC4_2 | | | | -1.62934145 |
| PTET.51.1.P0400271 | | Protein kinase-like domain | 86.474 | | | | 0.493840028 | -0.102421581 | -0.596261609 |
| PTET.51.1.P0410006 | | Serine/threonine-protein phosphatase 6 catalytic subunit | 13.451 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0410041 | | T-complex protein 1 subunit epsilon | 67.486 | | | | -0.43536455 | -0.160353763 | 0.275010786 |
| PTET.51.1.P0410063 | PtSMC4-1, SMC4 | P-loop containing nucleoside triphosphate hydrolase | 227.5 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0410099 | | 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase | 24.129 | | | | 0.169590236 | -1.143449914 | -1.313040149 |
| PTET.51.1.P0410101 | | Prephenate dehydrogenase | 30.759 | SMC4_1 | SMC4_2 | | | | -1.566692993 |
| PTET.51.1.P0410180 | | NADH dehydrogenase ubiquinone Fe-S protein 4, mitochondrial | 24.597 | SMC4_1 | SMC4_2 | | | | -1.409215414 |
| PTET.51.1.P0410229 | | Agglutinin alpha chain | 12.322 | WT | WT | | | | |
| PTET.51.1.P0420093 | vATPase_B2, PTETG4200001001 | P-loop containing nucleoside triphosphate hydrolase | 83.715 | | | | 0.94134053 | 0.638551971 | -0.302788559 |
| PTET.51.1.P0420119 | | Ribosomal protein L24e-related | 51.697 | | | | 2.365595125 | 0.065761182 | -2.299833943 |
| PTET.51.1.P0420242 | | NADH dehydrogenase | 67.064 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0430034 | | Aminoacyl-tRNA synthetase, class II (D/K/N)-like | 99.66 | | | | 4.751919598 | 4.584256565 | -0.167663034 |
| PTET.51.1.P0430169 | | Dehydrogenase E1 component | 18.867 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0430270 | | 60S acidic ribosomal protein P0 | 26.064 | SMC4_1 | SMC4_2 | | | | 0.009100119 |
| PTET.51.1.P0440040 | | UPF0302 protein YpiB | 14.147 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0450025 | | WD40/YVTN repeat-like-containing domain | 13.023 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0450054 | | ATP-dependent helicase/nuclease subunit A | 65.617 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0450076 | | ARP2/3 complex, 16kDa subunit (p16-Arc) | 32.831 | | | | -0.213054494 | 0.128919141 | 0.341973635 |
| PTET.51.1.P0450077 | PtSMC2-2, SMC2, PTETG4500016001 | Structural maintenance of chromosomes protein 2-1 | 26.079 | SMC4_1 | SMC4_2 | | | | 0.194467034 |
| PTET.51.1.P0450094 | | Glutathione synthetase | 71.076 | | | | -0.244490257 | 2.087092866 | 2.331583123 |
| PTET.51.1.P0450140 | PTETG4500010001 | Ribosomal protein S15 | 35.385 | | | | -0.747229521 | -0.287214182 | 0.46001534 |
| PTET.51.1.P0450180 | | DNA mismatch repair protein MSH7 | 47.961 | | WT | SMC4_1 | 1.596631365 | | |
| PTET.51.1.P0450187 | | Probable cinnamyl alcohol dehydrogenase 2 | 13.45 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P0450229 | | Mitochondrial-processing peptidase subunit beta | 30.851 | SMC4_1 | SMC4_2 | | | | 1.066887234 |
| PTET.51.1.P0450230 | | NAD(P)-binding domain | 93.916 | | | | 0.658565356 | 0.583027761 | -0.075537594 |
| PTET.51.1.P0460020 | | TCP-1/cpn60 chaperonin family | 84.724 | | | | -0.723045446 | -0.593793734 | 0.129251713 |
| PTET.51.1.P0460022 | | Thioredoxin | 19.246 | | WT | SMC4_1 | 0.811512224 | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0460036 | | Leucine-rich repeat, SDS22-like subfamily | 31.763 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0460041 | | Chaperonin Cpn60/TCP-1 | 28.153 | SMC4_1 | SMC4_2 | | | | -2.728400173 |
| PTET.51.1.P0460076 | | P-loop containing nucleoside triphosphate hydrolase | 38.887 | SMC4_1 | SMC4_2 | | | | -3.03607072 |
| PTET.51.1.P0460136 | | Inorganic polyphosphate/ATP-NAD kinase | 64.351 | | | | 4.259209253 | 5.065061732 | 0.805852479 |
| PTET.51.1.P0460194 | PTETG4600006001 | Phosphoenolpyruvate carboxykinase | 53.39 | | | | 1.421388448 | 1.814765312 | 0.393376863 |
| PTET.51.1.P0470012 | | Ribosomal L30 N-terminal domain | 12.157 | | | | 0.410636505 | 0.461470135 | 0.050833631 |
| PTET.51.1.P0470151 | | Trimeric LpxA-like | 33.512 | | WT | SMC4_1 | 2.635407607 | | |
| PTET.51.1.P0470179 | | Coiled coil domain | 18.36 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0470190 | | Arsenite methyltransferase | 47.969 | SMC4_1 | SMC4_2 | | | | -0.687532832 |
| PTET.51.1.P0480006 | PTETG4800001001 | Guanine nucleotide-binding protein subunit beta-like protein | 323.31 | | | | -0.46234253 | -1.344036171 | -0.881693641 |
| PTET.51.1.P0480063 | | Phosphofructokinase domain | 103.1 | | | | 1.062554201 | -1.575487736 | -2.638041938 |
| PTET.51.1.P0480070 | | Protein male abnormal 21 | 11.956 | WT | | SMC4_2 | | 1.481110931 | |
| PTET.51.1.P0480074 | | NADH-ubiquinone oxidoreductase, 21kDa subunit, N-terminal | 11.47 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0480132 | PTETG4800002001 | EF-hand domain pair | 18.41 | | | | 1.023181836 | -0.67694217 | -1.700124006 |
| PTET.51.1.P0480261 | | Eukaryotic porin/Tom40 | 12.829 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0490112 | | Calpain family cysteine protease | 25.128 | | WT | SMC4_1 | 0.908134307 | | |
| PTET.51.1.P0490213 | | Gamma-crystallin-related | 44.968 | | | | 1.862257539 | 2.019732382 | 0.157474843 |
| PTET.51.1.P0490229 | | Replication protein 1a | 18.536 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0500041 | | Coiled coil domain | 117.83 | SMC4_1 | SMC4_2 | | | | -1.277202834 |
| PTET.51.1.P0500099 | | 40S ribosomal protein S15a | 21.291 | | | | -0.816570125 | -0.459560199 | 0.357009926 |
| PTET.51.1.P0500184 | GapC | Glyceraldehyde/Erythrose phosphate dehydrogenase family | 20.008 | SMC4_1 | SMC4_2 | | | | -1.359295245 |
| PTET.51.1.P0500227 | | Serine/threonine-protein kinase ATG1b | 11.724 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0510037 | | Ribosomal L28e protein family | 12.482 | SMC4_1 | SMC4_2 | | | | 1.425011988 |
| PTET.51.1.P0510055 | | Ribosomal protein S3Ae | 18.499 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0510092 | | Proteasome component ECM29/Translational activator GCN1 | 11.148 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0510120 | | Ribosomal protein L6 | 55.558 | | | | 4.959577879 | 5.269939715 | 0.310361836 |
| PTET.51.1.P0510134 | PTETG5100001001, AK2 | Arginine kinase | 39.629 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P0510147 | sAG_51C | Paramecium surface antigen domain | 11.685 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P0510203 | | Cyclophilin-like domain | 13.69 | | | | -0.342323353 | -0.522614588 | -0.180291235 |
| PTET.51.1.P0520133 | | Elongation factor Tu | 33.276 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0520161 | | Cyclophilin-like domain | 12.743 | | WT | SMC4_1 | 1.087725893 | | |
| PTET.51.1.P0520173 | KdC1 | Porin domain | 112.39 | | | | 2.133198112 | 2.417501713 | 0.2843036 |
| PTET.51.1.P0530017 | | Armadillo-type fold | 19.237 | | | | | | -1.554547074 |
| PTET.51.1.P0530030 | | P-loop containing nucleoside triphosphate hydrolase | 196.08 | | | | 0.005650187 | -0.374573867 | -0.380224053 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0530059 | | Skp1 family, tetramerisation domain | 24.501 | | WT | SMC4_1 | 1.370356504 | | |
| PTET.51.1.P0530127 | | Chaperonin Cpn60/TCP-1 | 12.708 | SMC4_1 | SMC4_2 | | | | -1.555673662 |
| PTET.51.1.P0530165 | | alpha/beta hydrolase fold | 73.402 | SMC4_1 | SMC4_2 | | | | -2.066899206 |
| PTET.51.1.P0530218 | | RAC serine/threonine-protein kinase | 17.797 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0530254 | | TCP-1/cpn60 chaperonin family | 85.122 | | | | 0.314480304 | 0.490167007 | 0.175686703 |
| PTET.51.1.P0530256 | | Succinyl-CoA synthetase, beta subunit | 52.085 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0540213 | | Pyruvate/Phosphoenolpyruvate kinase-like domain | 24.34 | | WT | SMC4_1 | 0.563096121 | | |
| PTET.51.1.P0550098 | | von Willebrand factor type A domain | 11.557 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0550100 | | Tropomyosin like | 11.464 | | WT | SMC4_1 | 0.257315153 | | |
| PTET.51.1.P0550159 | | Protein-tyrosine phosphatase-like | 27.101 | SMC4_1 | SMC4_2 | | | | 1.570436882 |
| PTET.51.1.P0550219 | | Malate dehydrogenase | 116.81 | | | | 0.364104663 | -0.119465331 | -0.483569993 |
| PTET.51.1.P0560051 | | Peptidase C1A | 32.697 | | | | -0.913941133 | -3.393888661 | -2.479947528 |
| PTET.51.1.P0560226 | PTETG5600001001 | Cytochrome c oxidase subunit 1+2 | 48.751 | SMC4_1 | SMC4_2 | | | | 3.681874706 |
| PTET.51.1.P0570051 | PGML5a | Transposase IS4 | 19.074 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0570068 | | Luc7-like protein 3 | 48.671 | SMC4_1 | SMC4_2 | | | | 0.270719889 |
| PTET.51.1.P0570074 | | TNF receptor-associated protein 1 homolog, mitochondrial | 34.733 | SMC4_1 | SMC4_2 | | | | -0.649767307 |
| PTET.51.1.P0570163 | | Aconitase/isopropylmalate dehydratase | 24.336 | | WT | SMC4_1 | 1.909683374 | | |
| PTET.51.1.P0580041 | | RuvB-like | 13.861 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0580083 | | Fibrillarin | 56.081 | SMC4_1 | SMC4_2 | | | | -1.053108113 |
| PTET.51.1.P0580112 | | Saccharopine dehydrogenase C-terminal domain | 65.224 | | | | 0.896053335 | -0.232619529 | -1.128672864 |
| PTET.51.1.P0580147 | | Superoxide dismutase, copper/zinc binding domain | 25.876 | | | | 0.490457282 | 1.504697606 | 1.014240324 |
| PTET.51.1.P0590040 | PTETG5900041001 | Ribosomal L27e protein family | 18.356 | SMC4_1 | SMC4_2 | | | | 1.195422691 |
| PTET.51.1.P0590042 | | 60S ribosomal protein L6E | 24.225 | | | | 2.424388281 | 2.142912527 | -0.281475754 |
| PTET.51.1.P0590059 | | Ribosomal protein L22/L17, eukaryotic/archaeal | 13.811 | SMC4_1 | SMC4_2 | | | | 0.807889943 |
| PTET.51.1.P0590135 | PtSMC4-2, SMC4 | RecF/RecN/SMC, N-terminal | 243.16 | SMC4_1 | SMC4_2 | | | | 5.696556993 |
| PTET.51.1.P0590146 | DRPC1 | Dynamin central region | 11.664 | SMC4_1 | SMC4_2 | | | | 0.062880742 |
| PTET.51.1.P0590225 | | Ribosomal protein L14b/L23e | 124.16 | | | | 0.665724657 | 0.259025419 | -0.406699238 |
| PTET.51.1.P0600091 | | Ribosomal protein S9/S16 | 27.773 | | | | 3.001018729 | 2.4794536 | -0.521565129 |
| PTET.51.1.P0600104 | KdC2 | Uncharacterized killer plasmid pGKL-2 helicase | 11.171 | SMC4_1 | SMC4_2 | | | | 0.814148221 |
| PTET.51.1.P0600114 | | Cyclophilin-like domain | 36.814 | | | | 0.674016279 | -0.462850037 | -1.136866316 |
| PTET.51.1.P0600152 | | Prephenate dehydrogenase | 19.692 | | | | | -0.622636357 | |
| PTET.51.1.P0600185 | | 60S ribosomal protein L14 | 12.315 | SMC4_1 | SMC4_2 | | | | 0.502231667 |
| PTET.51.1.P0600189 | | Pyruvate dehydrogenase E1 component subunit beta | 17.805 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0610066 | PTETG6100008001 | AAA domain | 13.307 | | WT | SMC4_1 | 0.629025916 | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0610077 | PTETG6100010001 | Chloramphenicol acetyltransferase-like domain | 25.011 | | | | -0.08085226 | -2.956501353 | -2.875649093 |
| PTET.51.1.P0610150 | | Glycine cleavage system P protein | 26.278 | SMC4_1 | SMC4_2 | | | | -1.899402178 |
| PTET.51.1.P0610190 | | SecY translocase | 24.893 | | | | 1.848137737 | 0.709782674 | -1.138355063 |
| PTET.51.1.P0610217 | | Kinesin-like protein | 11.038 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0610254 | | Probable mitochondrial import receptor subunit TOM40-2 | 27.925 | SMC4_1 | SMC4_2 | | | | -2.119280759 |
| PTET.51.1.P0620102 | | Staphylococcal nuclease (SNase-like), OB-fold | 17.655 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0620182 | | Creatine kinase M-type | 19.196 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0620230 | | Coiled coil domain | 11.5 | WT | WT | | | | |
| PTET.51.1.P0620246 | | Peptidase M1, alanine aminopeptidase/leukotriene A4 hydrolase | 65.156 | SMC4_1 | SMC4_2 | | | | 1.022904209 |
| PTET.51.1.P0620250 | | Ribosomal protein L31e | 17.692 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0630027 | can-A5-2, PTETG6300001001 | ATP synthase delta (OSCP) subunit | 50.113 | | | | 1.477651728 | 0.616067503 | -0.861584225 |
| PTET.51.1.P0630039 | | 2-oxoglutarate dehydrogenase E1 component | 57.803 | | WT | SMC4_1 | 1.999695673 | | |
| PTET.51.1.P0630139 | | Fructose-1-6-bisphosphatase, N-terminal domain | 11.475 | WT | WT | | | | |
| PTET.51.1.P0640022 | PtSERCA1, SERCA, SERCA1 | Calcium-transporting ATPase 2, endoplasmic reticulum-type | 92.398 | | | | 1.906902643 | -0.057684191 | -1.964586834 |
| PTET.51.1.P0650012 | | Thioredoxin | 49.994 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0650034 | STO7, PtSto7, STO | Stomatin-like protein 2, mitochondrial | 12.2 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P0660034 | PTETG6600006001 | Ubiquitin-like modifier-activating enzyme 1 | 30.16 | SMC4_1 | SMC4_2 | | | | 0.384354978 |
| PTET.51.1.P0660075 | | P-loop containing nucleoside triphosphate hydrolase | 83.436 | | | | 0.645849523 | 0.096721825 | -0.549127698 |
| PTET.51.1.P0660088 | | Ribosomal protein L22e | 48.162 | | | | 2.140610805 | 2.142047378 | 0.001436573 |
| PTET.51.1.P0660146 | | Uncharacterized protein YggE | 44.475 | | | | 0.840886608 | 0.871096426 | 0.030209819 |
| PTET.51.1.P0660147 | | Protein of unknown function DUF541 | 49.105 | | | | 0.376209149 | 0.306949418 | -0.069259731 |
| PTET.51.1.P0670034 | PTETG6700004001 | Ubiquitin-conjugating enzyme/RWD-like | 17.071 | | | | 0.347719802 | -0.667080408 | -1.01480021 |
| PTET.51.1.P0670042 | | Thiol protease SEN102 | 13.361 | SMC4_1 | SMC4_2 | | | | 2.037950184 |
| PTET.51.1.P0670074 | | RNA recognition motif domain, eukaryote | 23.026 | | | | -3.35645567 | -3.500869393 | -0.144413722 |
| PTET.51.1.P0670172 | | Ornithine aminotransferase | 100.44 | | | | 0.534834953 | 0.944850725 | 0.410015772 |
| PTET.51.1.P0670236 | | Major vault protein, N-terminal | 67.527 | SMC4_1 | SMC4_2 | | | | -0.201303798 |
| PTET.51.1.P0690129 | | Ribosomal protein S5, C-terminal domain | 72.4 | | | | 1.604653985 | 2.079791979 | 0.475137993 |
| PTET.51.1.P0700060 | | Insulin-like growth factor binding protein, N-terminal | 39.256 | SMC4_1 | SMC4_2 | | | | -0.744665076 |
| PTET.51.1.P0700131 | | Chaperonin Cpn60/TCP-1 | 82.651 | | | | -0.329424262 | -0.410698235 | -0.081273972 |
| PTET.51.1.P0700138 | | Protein piccolo | 17.095 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0710112 | PTIWI01, PTETG7100004001 | Ribonuclease H-like domain | 17.432 | SMC4_1 | | SMC4_1 | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0710131 | alphaPT4, tub_alphaPT4, PTETG7100002001 | Tubulin alpha chain | 323.31 | | | | 0.457990877 | 1.588300465 | 1.130309588 |
| PTET.51.1.P0710133 | | Ribosomal L22e protein family | 12.684 | | | | 2.129112835 | 1.711487111 | -0.417625724 |
| PTET.51.1.P0710140 | | Fe-S cluster assembly protein DRE2 | 13.87 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0710220 | | dITP/XTP pyrophosphatase | 19.327 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0720056 | | Ribosomal protein L18e/L15P | 17.482 | SMC4_1 | SMC4_2 | | | | -1.508538399 |
| PTET.51.1.P0720102 | | Coiled coil domain | 30.208 | | | | 0.898986084 | 3.501253089 | 2.602267005 |
| PTET.51.1.P0720184 | | Cse1 | 24.803 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0730164 | | KH domain | 185.65 | | | | 2.368489155 | 1.742818213 | -0.625670943 |
| PTET.51.1.P0740114 | | Zinc finger CCCH domain-containing protein 13 | 12.068 | | WT | SMC4_1 | 0.403114662 | | |
| PTET.51.1.P0740177 | | Protein of unknown function DUF3546 | 12.886 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0740192 | | HAD-like domain | 28.31 | | | | 2.557731101 | 2.591874429 | 0.034143328 |
| PTET.51.1.P0750047 | | Nephrocystin-3 | 38.474 | | WT | SMC4_1 | 0.375292 | | |
| PTET.51.1.P0750166 | | Protein kinase domain | 14.303 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0750175 | | NAD(P)-binding domain | 12.768 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P0750181 | | MmgE/PrpD | 89.878 | | | | 1.215722326 | 1.516434711 | 0.300712386 |
| PTET.51.1.P0760045 | | Phosphofructokinase domain | 38.084 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0760195 | | Ornithine aminotransferase, mitochondrial | 11.905 | SMC4_1 | SMC4_2 | | | | -0.251839819 |
| PTET.51.1.P0760207 | | dTDP-4-dehydrorhamnose reductase | 11.511 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0770007 | | Putative DNA-binding domain | 24.534 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0770142 | | Glycine cleavage system T protein | 13.142 | | | | -0.57496509 | -0.562851342 | 0.012113747 |
| PTET.51.1.P0770143 | | CID domain | 11.874 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0770145 | | V-ATPase subunit C | 20.772 | SMC4_1 | SMC4_2 | | | | -0.473311621 |
| PTET.51.1.P0770146 | | 3-ketoacyl-CoA thiolase B, peroxisomal | 135.56 | | | | 0.059419468 | 0.770768025 | 0.711348557 |
| PTET.51.1.P0770174 | | Aspartate/other aminotransferase | 17.79 | | | | 0.215110468 | 1.574331696 | 1.359221227 |
| PTET.51.1.P0770205 | | Pyridine nucleotide disulphide reductase class-I signature | 13.056 | | WT | SMC4_1 | 3.371753647 | | |
| PTET.51.1.P0780038 | | ATPase, F1 complex, gamma subunit domain | 12.269 | | | | 2.525764859 | 1.480945077 | -1.044819782 |
| PTET.51.1.P0780047 | | Phosphoglycerate kinase | 53.096 | | | | -0.221356355 | -0.310192096 | -0.088835741 |
| PTET.51.1.P0780060 | | WD40-repeat-containing domain | 11.894 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0780129 | | Uridylate kinase | 18.668 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0790053 | | Heat shock protein 104 | 68.078 | | | | 1.728400086 | 1.496874204 | -0.231525881 |
| PTET.51.1.P0800043 | PTETG8000005001 | Protein kinase-like domain | 23.083 | | WT | SMC4_1 | -1.817091039 | | |
| PTET.51.1.P0800091 | | Coiled coil domain | 37.057 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0800119 | | Peptidase C1A, papain C-terminal | 13.802 | SMC4_1 | SMC4_2 | | | | 1.279702687 |
| PTET.51.1.P0800141 | | 60S ribosomal protein L4 | 90.397 | | | | 2.54424805 | 3.109300998 | 0.565052947 |
| PTET.51.1.P0800155 | | Endoplasmin | 54.805 | | | | 0.307310857 | -0.227896294 | -0.535207151 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0800198 | | Chromo (CHRromatin Organisation MOdifier) domain | 12.289 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0810081 | | Pseudouridine synthase I, TruA | 18.339 | | | | -0.774260861 | -0.936102595 | -0.161841734 |
| PTET.51.1.P0810085 | | Thiolase, N-terminal domain | 11.294 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P0810095 | rab_B78 | Intraflagellar transport protein 22 homolog | 13.161 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0810192 | | Phosphoribosyltransferase-like | 33.115 | SMC4_1 | SMC4_2 | | | | 0.789853101 |
| PTET.51.1.P0820022 | | Tetratricopeptide-like helical domain | 119.42 | | | | 1.182600901 | 1.566934307 | 0.384333406 |
| PTET.51.1.P0820041 | | Actin-related protein 2/3 complex subunit 4 | 24.108 | | | | 1.098900301 | 1.718865598 | 0.619965297 |
| PTET.51.1.P0820051 | | PCI/PINT associated module | 24.145 | SMC4_1 | SMC4_2 | | | | 0.5328822 |
| PTET.51.1.P0820072 | | Ribosomal protein L5 eukaryotic/L18 archaeal | 50.117 | | | | 3.248641374 | 4.097227152 | 0.848585778 |
| PTET.51.1.P0820086 | | Ribosomal protein L18a/LX | 33.11 | | | | 0.06055101 | -0.183125824 | -0.243676834 |
| PTET.51.1.P0820101 | PTETG8200012001 | Ubiquitin-related domain | 16.221 | | | | 0.639921203 | 1.586432709 | 0.946511506 |
| PTET.51.1.P0820112 | rab_A30 | P-loop containing nucleoside triphosphate hydrolase | 12.229 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0820161 | | Ribosomal protein S4/S9 N-terminal domain | 25.139 | SMC4_1 | SMC4_2 | | | | 1.29765986 |
| PTET.51.1.P0820162 | | High-osmolarity-induced transcription protein 1 | 12.699 | | | | -0.664629979 | -0.248592501 | 0.416037478 |
| PTET.51.1.P0820178 | | P-loop containing nucleoside triphosphate hydrolase | 11.348 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0830027 | | Cilia- and flagella-associated protein 57 | 17.311 | | WT | SMC4_1 | 1.05023004 | | |
| PTET.51.1.P0830052 | | Probable 3-hydroxybutyryl-CoA dehydrogenase | 46.687 | | | | 1.058295185 | 0.685062806 | -0.373232378 |
| PTET.51.1.P0830133 | | Mitochondrial carrier protein | 47.981 | SMC4_1 | SMC4_2 | | | | -1.341912592 |
| PTET.51.1.P0840196 | | Serine/threonine-protein kinase HT1 | 18.41 | | WT | SMC4_1 | 2.01207079 | | |
| PTET.51.1.P0850007 | | Heat shock protein 70 family | 196.95 | | | | 2.51638606 | 3.648594762 | 1.132208703 |
| PTET.51.1.P0850107 | | Clathrin adaptor, mu subunit | 76.334 | | | | -0.229347848 | -0.207111229 | 0.022236619 |
| PTET.51.1.P0850132 | | Histidine protein kinase 1 | 43.636 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0850133 | act1-5 | Actin-related protein | 24.665 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0850169 | | Cysteine-rich with EGF-like domain protein 1 | 18.53 | | | | -0.541387637 | 0.087977144 | 0.62936478 |
| PTET.51.1.P0860027 | | Protein NLRC5 | 53.957 | | | | -1.194150362 | 1.252859145 | 2.447009507 |
| PTET.51.1.P0860064 | | Coiled coil domain | 12.037 | | WT | SMC4_1 | -1.460038451 | | |
| PTET.51.1.P0860086 | | Tubby C-terminal-like domain | 41.528 | | | | -0.277263929 | -0.431123688 | -0.153859759 |
| PTET.51.1.P0860204 | | Redoxin | 11.313 | | | | -0.145466249 | 0.065661712 | 0.211127961 |
| PTET.51.1.P0870133 | | Alpha-protein kinase vwkA | 18.364 | | WT | SMC4_1 | | | |
| PTET.51.1.P0870152 | | Acyl-coenzyme A thioesterase 9, mitochondrial | 36.809 | SMC4_1 | SMC4_2 | | | | 1.777348704 |
| PTET.51.1.P0870166 | | Trichocyst matrix protein T1-B | 11.421 | SMC4_1 | | SMC4_1 | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P0870174 | | Histone H2B type 1-M | 24.358 | | | | 2.051335896 | 2.068718942 | 0.017383046 |
| PTET.51.1.P0880068 | | H-type lectin domain | 323.31 | | | | -1.03534828 | -1.059324892 | -0.023976612 |
| PTET.51.1.P0880113 | | Protein phosphatase 2A, regulatory subunit PR55 | 12.358 | SMC4_1 | SMC4_2 | | | | -1.184663319 |
| PTET.51.1.P0890039 | | Reticulocyte-binding protein 2 homolog a | 31.403 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0900010 | | Pyridine nucleotide disulphide reductase class-I signature | 112.59 | | | | 2.183429211 | 1.431544828 | -0.751884383 |
| PTET.51.1.P0900182 | | Ribosomal protein L7Ae/L30e/S12e/Gadd45 family | 18.74 | SMC4_1 | SMC4_2 | | | | -1.287875511 |
| PTET.51.1.P0900212 | | Isocitrate/isopropylmalate dehydrogenase | 82.406 | | | | 3.109820213 | 2.110750642 | -0.999069571 |
| PTET.51.1.P0910161 | PTETG9100014001 | Thiolase-like | 37.306 | | | | -1.419875227 | -3.596261095 | -2.176385867 |
| PTET.51.1.P0920028 | | Transmembrane protein 120B-A | 18.749 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0920068 | PTETG9200004001 | 40S ribosomal protein S12 | 24.439 | | | | 1.771498908 | 4.083573 | 2.312074092 |
| PTET.51.1.P0920087 | | Pyridine nucleotide-disulphide oxidoreductase, FAD/NAD(P)-binding domain | 26.19 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0920172 | | Glutathione S-transferase, C-terminal-like | 25.513 | | | | 0.18462559 | 2.51529703 | 2.330671439 |
| PTET.51.1.P0940124 | | Rcd1 | 12.578 | | WT | SMC4_1 | 0.950297874 | | |
| PTET.51.1.P0940159 | | Fructose-bisphosphate aldolase class-I, eukaryotic-type | 17.417 | | | | 0.468315033 | 0.48864932 | 0.020334287 |
| PTET.51.1.P0950040 | | Coiled coil domain | 32.226 | | | | -0.236895856 | -3.596998175 | -3.360102319 |
| PTET.51.1.P0950122 | KdC3 | Histone-lysine N-methyltransferase SETD7 | 92.957 | | | | 1.342372 | -0.491711431 | -1.83408343 |
| PTET.51.1.P0950124 | | JAB1/MPN/MOV34 metalloenzyme domain | 13.17 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0960133 | | Nucleotide-binding, alpha-beta plait | 11.951 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0970177 | | Pyridoxal 5'-phosphate synthase subunit PdxT | 12.228 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0980013 | | 60S ribosomal protein L18 | 21.605 | | | | 1.471434067 | 2.169883154 | 0.698449088 |
| PTET.51.1.P0980025 | | Ribosomal protein S17e | 16.767 | SMC4_1 | SMC4_2 | | | | -1.670851477 |
| PTET.51.1.P0980041 | | E3 ubiquitin-protein ligase Hakai | 18.056 | | | | 2.271677753 | 0.327816346 | -1.943861407 |
| PTET.51.1.P0980107 | | 50S ribosome-binding GTPase | 60.173 | | | | 1.409304793 | 5.111311386 | 3.702006593 |
| PTET.51.1.P0980124 | | Lysine-specific demethylase 6A | 11.712 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P0980168 | | 40S ribosomal protein SA | 94.167 | | | | -0.786320836 | 0.841717292 | 1.628038127 |
| PTET.51.1.P0980184 | PTETG9800002001 | 14-3-3 protein | 30.322 | | WT | SMC4_1 | -0.215862996 | | |
| PTET.51.1.P0980191 | | 50S ribosomal protein L9 | 24.793 | SMC4_1 | SMC4_2 | | | | 0.330378004 |
| PTET.51.1.P1000056 | | Na(+)/H(+) antiporter 2 | 75.107 | SMC4_1 | SMC4_2 | | | | 0.725208741 |
| PTET.51.1.P1010044 | | Protein kinase-like domain | 44.23 | | WT | SMC4_1 | 0.639295127 | | |
| PTET.51.1.P1010047 | | TCP-1/cpn60 chaperonin family | 122.5 | | | | 0.309760411 | -0.683942043 | -0.993702455 |
| PTET.51.1.P1010085 | HSP90-1, PTETG10100003001 | Heat shock protein 81-3 | 182.89 | | | | -0.051692721 | 0.061835307 | 0.113528028 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P1010086 | | Ribosomal protein S7e | 12.609 | SMC4_1 | SMC4_2 | | | | 0.447002848 |
| PTET.51.1.P1010109 | | AMP-dependent synthetase/ligase | 36.165 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1010158 | | T-complex protein 1 subunit delta | 162.25 | | | | -0.126914271 | 0.018558121 | 0.145472392 |
| PTET.51.1.P1020022 | | ATP-grasp domain | 51.459 | | | | 0.191320437 | 0.226518855 | 0.035198418 |
| PTET.51.1.P1020045 | Hsp70Pt08 | Luminal-binding protein 5 | 13.453 | SMC4_1 | SMC4_2 | | | | 2.441626658 |
| PTET.51.1.P1020084 | | 40S ribosomal protein S12 | 12.99 | SMC4_1 | SMC4_2 | | | | -0.463552511 |
| PTET.51.1.P1020090 | | Enoyl-CoA hydratase/isomerase | 12.054 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1020102 | | Protein phosphatase 2A, regulatory subunit PR55 | 95.561 | | | | 0.100893338 | 0.438603113 | 0.337709775 |
| PTET.51.1.P1020121 | PTETG10200002001 | WD40-repeat-containing domain | 12.364 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1050110 | | Malate dehydrogenase | 28.51 | | | | 1.475140728 | 0.851578813 | -0.623561915 |
| PTET.51.1.P1050126 | PTETG10500002001 | Glutathione S-transferase Mu 4 | 24.399 | SMC4_1 | SMC4_2 | | | | 0.157341576 |
| PTET.51.1.P1050161 | | Glutaredoxin | 19.563 | SMC4_1 | SMC4_2 | | | | 2.327755898 |
| PTET.51.1.P1060046 | | WD40/YVTN repeat-like-containing domain | 77.382 | | | | 0.581379236 | 1.332501327 | 0.751122091 |
| PTET.51.1.P1060077 | | Profilin-1A | 14.246 | SMC4_1 | SMC4_2 | | | | 1.997624453 |
| PTET.51.1.P1060094 | | Phenylalanyl-tRNA synthetase | 29.413 | | WT | SMC4_1 | 2.029097809 | | |
| PTET.51.1.P1060137 | | Superoxide dismutase | 21.378 | | WT | SMC4_1 | 2.344131336 | | |
| PTET.51.1.P1060141 | | Major vault protein beta | 58.25 | SMC4_1 | SMC4_2 | | | | 0.081582513 |
| PTET.51.1.P1060180 | sAG_51A, PTETG10600003001 | G surface protein, allelic form 168 | 323.31 | | | | 0.022167241 | 6.482447954 | 6.460280713 |
| PTET.51.1.P1070023 | | B30.2/SPRY domain | 19.443 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1070046 | | P-loop containing nucleoside triphosphate hydrolase | 112.09 | | WT | SMC4_1 | 0.197453322 | | |
| PTET.51.1.P1070063 | | Profilin | 19.3 | | | | 1.187765191 | 1.840672946 | 0.652907754 |
| PTET.51.1.P1080068 | | HSP20-like chaperone | 26.622 | SMC4_1 | SMC4_2 | | | | 0.486858772 |
| PTET.51.1.P1080081 | | Putative amino-acid ABC transporter-binding protein PatH | 28.61 | | WT | SMC4_1 | 1.570228787 | | |
| PTET.51.1.P1080104 | | F-box-like/WD repeat-containing protein TBL1X | 89.49 | | | | -0.09690444 | 1.844742054 | 1.941646494 |
| PTET.51.1.P1080183 | | Peptidase C2, calpain, catalytic domain | 12.52 | SMC4_1 | SMC4_2 | | | | -0.299752334 |
| PTET.51.1.P1080192 | rab_C103 | RanBP1 domain | 20.196 | | | | -0.425457957 | -0.084103902 | 0.341354055 |
| PTET.51.1.P1100030 | | ADP-ribosylation factor family | 21.865 | | | | 0.674859696 | -0.727734637 | -1.402594333 |
| PTET.51.1.P1110038 | | Homoserine O-acetyltransferase | 35.16 | SMC4_1 | SMC4_2 | | | | -2.127433883 |
| PTET.51.1.P1110100 | | Transketolase | 55.305 | | | | 0.694456376 | -1.405371008 | -2.099827384 |
| PTET.51.1.P1110112 | PCM1 | NAC domain-containing protein 68 | 25.237 | SMC4_1 | SMC4_2 | | | | 5.291532417 |
| PTET.51.1.P1110113 | PCM1 | Gamma-crystallin D | 323.31 | | | | -0.517525484 | 0.588232921 | 1.105758405 |
| PTET.51.1.P1110150 | | Ribosomal protein L23 | 13.369 | SMC4_1 | SMC4_2 | | | | -0.822986529 |
| PTET.51.1.P1120047 | | Vacuolar protein sorting-associated protein 35, Vps35 | 24.377 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1120052 | | Ribosomal protein L44e | 19.5 | | | | 1.258718406 | 0.647902934 | -0.610815472 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P1120149 | | MmgE/PrpD family | 25.535 | | | | 1.338002123 | -0.279697861 | -1.617699984 |
| PTET.51.1.P1120155 | | 40S ribosomal protein S15-B | 29.115 | | | | 3.524331024 | 2.100154352 | -1.424176672 |
| PTET.51.1.P1130041 | | Glutathione S-transferase, C-terminal-like | 17.954 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1150054 | TEF1, PTETG11500003001 | Elongation factor Tu GTP binding domain | 156.07 | | | | 0.982599106 | 1.329231722 | 0.346632616 |
| PTET.51.1.P1150082 | | Mitochondrial substrate/solute carrier | 26.345 | | | | 2.332730903 | 0.356298569 | -1.976432334 |
| PTET.51.1.P1150112 | | Mitochondrial acidic protein MAM33 | 16.856 | | | | 0.869871225 | -1.02218374 | -1.892054966 |
| PTET.51.1.P1150131 | | Pyridine nucleotide-disulphide oxidoreductase, FAD/NAD(P)-binding domain | 151.71 | | | | 1.830143446 | -0.615779252 | -2.445922698 |
| PTET.51.1.P1160067 | | Protein kinase domain | 11.517 | | | | 0.929878052 | 1.227154242 | 0.297276191 |
| PTET.51.1.P1160120 | | 15-cis-zeta-carotene isomerase, chloroplastic | 20.727 | | | | 1.149910628 | 0.352210727 | -0.797699901 |
| PTET.51.1.P1170020 | | Ribosomal proteins L26 eukaryotic, L24P archaeal | 12.371 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1170022 | | Histone-fold | 24.943 | WT | WT | | | | |
| PTET.51.1.P1170122 | | Mitochondrial phosphate carrier protein 3, mitochondrial | 63.927 | | | | 3.27310856 | 1.871290831 | -1.401817729 |
| PTET.51.1.P1170129 | | SURF1 family | 21.674 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1170142 | | WD domain, G-beta repeat | 11.841 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1180044 | | Transaldolase | 12.623 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1180048 | | P-loop containing nucleoside triphosphate hydrolase | 40.078 | SMC4_1 | SMC4_2 | | | | -0.085526657 |
| PTET.51.1.P1180049 | | ATPase, F1 complex, gamma subunit domain | 23.67 | | | | 3.839332296 | 2.673869644 | -1.165462652 |
| PTET.51.1.P1180090 | | Thioredoxin-like fold | 22.037 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1180110 | | Heat shock protein 70 family | 201.3 | | | | 0.931212543 | 0.466951899 | -0.464260645 |
| PTET.51.1.P1180158 | | Calpain family cysteine protease | 22.393 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1190018 | | Thioredoxin-like fold | 25.785 | SMC4_1 | SMC4_2 | | | | 1.195004824 |
| PTET.51.1.P1190028 | | Ribosomal protein S13-like, H2TH | 19.053 | SMC4_1 | SMC4_2 | | | | -2.300146215 |
| PTET.51.1.P1190058 | | WD40-repeat-containing domain | 34.919 | | | | 0.07788074 | 0.731801254 | 0.653920513 |
| PTET.51.1.P1200007 | | Sec63 domain | 17.646 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1210120 | rab_A10 | Ras-related protein Rab-1A | 12.524 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1220029 | | Pyridine nucleotide-disulphide oxidoreductase | 93.949 | SMC4_1 | SMC4_2 | | | | -2.971455419 |
| PTET.51.1.P1220083 | | Serine hydroxymethyltransferase 2, mitochondrial | 13.624 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1220095 | | NAD(P)-binding domain | 118.86 | | | | 0.585026759 | -0.165789367 | -0.750816126 |
| PTET.51.1.P1220101 | | 26S proteasome, regulatory subunit Rpn7 | 18.411 | | | | 2.803607409 | 1.667982254 | -1.135625156 |
| PTET.51.1.P1220116 | | Pyridoxal phosphate-dependent transferase | 13.082 | | WT | SMC4_1 | -0.203272801 | | |
| PTET.51.1.P1230083 | | Cingulin-like protein 1 | 35.152 | | | | 2.290168973 | 1.566806791 | -0.723362182 |
| PTET.51.1.P1230130 | | Nucleoside phosphatase GDA1/CD39 | 73.737 | SMC4_1 | SMC4_2 | | | | -0.24888278 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P1240009 | | Succinyl-CoA synthetase, beta subunit | 12.619 | | | | 0.864675047 | 0.374056346 | -0.490618701 |
| PTET.51.1.P1240016 | | Nuclear pore complex protein GP210 | 14.025 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1240020 | | 26S proteasome regulatory complex, non-ATPase subcomplex, Rpn1 subunit | 30.79 | SMC4_1 | SMC4_2 | | | | -1.807403992 |
| PTET.51.1.P1240024 | | Protein arginine N-methyltransferase 1.1 | 45.073 | | | | -1.506895769 | -1.802640004 | -0.295744234 |
| PTET.51.1.P1240046 | | H-type lectin domain | 16.791 | | | | -0.959342905 | -1.129090743 | -0.169747837 |
| PTET.51.1.P1240098 | | Adenylate kinase | 26.531 | | | | 1.917139291 | -0.876237301 | -2.793376592 |
| PTET.51.1.P1250020 | | Protein disulfide-isomerase | 13.065 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1250118 | | Ubiquitin-conjugating enzyme/RWD-like | 133.41 | | WT | SMC4_1 | 2.316623531 | | |
| PTET.51.1.P1250136 | | Ribosomal protein L11/L12 | 26.78 | | | | 2.907491599 | 2.166520982 | -0.740970617 |
| PTET.51.1.P1260091 | | Netrin receptor UNC5D | 48.601 | SMC4_1 | SMC4_2 | | | | -0.954801324 |
| PTET.51.1.P1270007 | | WD40-repeat-containing domain | 13.339 | | WT | SMC4_1 | -0.685506805 | | |
| PTET.51.1.P1270143 | PTETG12700004001 | Serine/threonine-protein kinase Nek3 | 16.61 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1280037 | | Iron sulphur-containing domain, CDGSH-type, subfamily | 24.96 | | | | 0.484430327 | -0.811496155 | -1.295926482 |
| PTET.51.1.P1280060 | | Coenzyme A transferase | 19.639 | SMC4_1 | SMC4_2 | | | | -0.165252255 |
| PTET.51.1.P1290088 | | Ribonuclease P protein subunit p25-like protein | 12.392 | | | | 0.631763337 | 0.726406933 | 0.094643596 |
| PTET.51.1.P1290090 | PTETG12900002001 | Ubiquitin-conjugating enzyme/RWD-like | 12.842 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P1300003 | | Proteasome subunit Rpn10 | 19.785 | SMC4_1 | SMC4_2 | | | | 2.142926869 |
| PTET.51.1.P1300018 | | Glutathione S-transferase, C-terminal domain | 36.052 | | | | 0.372621272 | 1.261651879 | 0.889030606 |
| PTET.51.1.P1300068 | | Elongation factor G-1, chloroplastic | 87.093 | SMC4_1 | SMC4_2 | | | | -1.902561375 |
| PTET.51.1.P1300125 | | Coiled coil domain | 12.374 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1300163 | | Nucleic acid-binding, OB-fold | 24.228 | | | | -2.269195043 | -1.488746203 | 0.78044884 |
| PTET.51.1.P1310018 | | Beta/Gamma crystallin | 323.31 | | | | -0.637974026 | -1.170518936 | -0.53254491 |
| PTET.51.1.P1310046 | | Ribosomal protein L13 | 19.794 | | | | -0.22092009 | 0.740273385 | 0.961193475 |
| PTET.51.1.P1320133 | | 40S ribosomal protein S4 | 99.809 | | | | 0.375382546 | 0.887916365 | 0.512533819 |
| PTET.51.1.P1330007 | | Ribonuclease T2-like | 34.433 | | | | 1.160025841 | 0.826409547 | -0.333616293 |
| PTET.51.1.P1330041 | pp2r-1, PTETG13300003001 | Calcineurin-like phosphoesterase | 19.466 | | | | 0.135928447 | -0.589197925 | -0.725126372 |
| PTET.51.1.P1330115 | | haloacid dehalogenase-like hydrolase | 19.574 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1330129 | | Capsid protein | 19.376 | | WT | SMC4_1 | 1.633443182 | | |
| PTET.51.1.P1330138 | | Mercuric resistance operon regulatory protein | 23.79 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1340014 | | Mitochondrial carrier protein | 13.184 | SMC4_1 | SMC4_2 | | | | -1.398398284 |
| PTET.51.1.P1340068 | | BTB/POZ fold | 39.8 | | | | 0.613566915 | -0.757155826 | -1.370722741 |
| PTET.51.1.P1350053 | PTETG13500002001 | Intraflagellar transport protein 172 homolog | 60.24 | | WT | SMC4_1 | -0.720985075 | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P1350121 | | Ribosomal protein L18 | 18.22 | SMC4_1 | SMC4_2 | | | | 2.906453069 |
| PTET.51.1.P1360074 | PTETG13600002001 | Ubiquitin-conjugating enzyme | 11.824 | SMC4_1 | SMC4_2 | | | | 1.399597445 |
| PTET.51.1.P1360118 | | HAD-like domain | 12.223 | SMC4_1 | SMC4_2 | | | | |
| PTET.51.1.P1370083 | | Adenylate kinase | 11.25 | SMC4_1 | SMC4_2 | | | | 0.021812029 |
| PTET.51.1.P1390019 | | Insulin-like growth factor binding protein, N-terminal | 46.094 | | | | -1.240494258 | -1.433062785 | -0.192568527 |
| PTET.51.1.P1390033 | | 26S proteasome non-ATPase regulatory subunit 11 | 11.658 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1400024 | can-A4-2, PTETG14000006001 | ATPase, OSCP/delta subunit | 13.619 | | | | -2.160970389 | -1.733024893 | 0.427945497 |
| PTET.51.1.P1400027 | | DNA double-strand break repair Rad50 ATPase | 11.03 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1400050 | | Uncharacterized protein C458.02c | 52.702 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1400060 | | Ribosomal protein L15e | 31.687 | | | | 0.095325499 | 1.498916489 | 1.40359099 |
| PTET.51.1.P1400095 | PTETG14000016001 | V-type ATP synthase alpha chain | 30.454 | | | | 0.235037571 | -0.389344583 | -0.624382154 |
| PTET.51.1.P1400107 | | Probable mitochondrial 2-oxoglutarate/malate carrier protein | 46.069 | SMC4_1 | SMC4_2 | | | | -1.193078776 |
| PTET.51.1.P1420051 | PCM2, PTETG14200002001 | Insulin-like growth factor binding protein, N-terminal | 239.05 | | | | -1.265557647 | -0.632218445 | 0.633339202 |
| PTET.51.1.P1420091 | PTETG14200003001 | ARP2/3 complex, 34kDa subunit (p34-Arc) | 57.471 | | | | 0.244356598 | 1.530006146 | 1.285649547 |
| PTET.51.1.P1430120 | | Alanine--tRNA ligase | 18.107 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1440017 | | Oligoribonuclease | 161.77 | | WT | SMC4_1 | 0.884989695 | | |
| PTET.51.1.P1440071 | PtCenBP1, CenBP1 | Coiled coil domain | 12.244 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1440105 | ran_B14 | Ran GTPase | 73.916 | | | | 2.478509648 | 2.695605156 | 0.217095508 |
| PTET.51.1.P1440125 | | 2-acylglycerol O-acyltransferase 1 | 35.927 | SMC4_1 | SMC4_2 | | | | -1.177335223 |
| PTET.51.1.P1440137 | | Oxygen-dependent choline dehydrogenase | 18.957 | | | | 0.047823335 | -0.174734402 | -0.222557737 |
| PTET.51.1.P1450009 | | 50S ribosome-binding GTPase | 11.335 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1450044 | | Chromo (CHRromatin Organisation MOdifier) domain | 22.105 | SMC4_1 | SMC4_2 | | | | -3.344759907 |
| PTET.51.1.P1460073 | | Lipid-A-disaccharide synthetase | 61.174 | SMC4_1 | SMC4_2 | | | | -2.969791857 |
| PTET.51.1.P1460081 | | Accumulation-associated protein | 85.927 | | | | 3.821062816 | 1.982678154 | -1.838384663 |
| PTET.51.1.P1470096 | | Prohibitin | 18.631 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1470170 | | Protein translocase subunit SecA | 35.519 | | | | 1.825159106 | 0.813759709 | -1.011399397 |
| PTET.51.1.P1480062 | | 54S ribosomal protein L17, mitochondrial | 11.878 | SMC4_1 | SMC4_2 | | | | 1.211895697 |
| PTET.51.1.P1480099 | | Thymidine kinase | 19.165 | | | | 0.431361615 | 0.690825846 | 0.259464231 |
| PTET.51.1.P1490073 | | Succinate dehydrogenase | 59.133 | | | | 4.666455542 | 4.285994996 | -0.380460546 |
| PTET.51.1.P1490078 | | Fibrillarin | 17.918 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1490080 | | C3a anaphylatoxin chemotactic receptor | 20.303 | SMC4_1 | SMC4_2 | | | | -2.434963222 |
| PTET.51.1.P1490087 | | Branched-chain-amino-acid aminotransferase | 26.418 | | WT | SMC4_1 | 2.770852094 | | |
| PTET.51.1.P1500083 | | Ribosomal protein S3Ae | 86.55 | | | | 0.559399878 | 0.181437586 | -0.377962292 |
| PTET.51.1.P1500100 | | 40S ribosomal protein S4 | 29.476 | SMC4_1 | SMC4_2 | | | | 0.453406053 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P1510028 | PTETG15100006001 | Peptidase C48, SUMO/Sentrin/Ubl1 | 11.626 | SMC4_1 | SMC4_2 | | | | 0.494294568 |
| PTET.51.1.P1510031 | | 60S ribosomal protein L10a-3 | 68.376 | | | | 2.479402854 | 3.037658986 | 0.558256132 |
| PTET.51.1.P1510079 | | Ribulose bisphosphate carboxylase small chain 3B, chloroplastic | 18.178 | | | | 0.327038715 | -1.078034119 | -1.405072834 |
| PTET.51.1.P1520081 | | Protein kinase domain | 17.984 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1530028 | | tRNA dimethylallyltransferase | 26.258 | SMC4_1 | SMC4_2 | | | | -0.671453806 |
| PTET.51.1.P1530088 | | Ribosomal protein S26e | 12.72 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1540048 | | ssDNA-binding transcriptional regulator | 22.987 | SMC4_1 | SMC4_2 | | | | -0.306716605 |
| PTET.51.1.P1540072 | | K(+) efflux antiporter 1, chloroplastic | 323.31 | | | | -0.440158101 | -1.827741941 | -1.38758384 |
| PTET.51.1.P1550057 | | ATP-dependent DNA helicase hus2/rqh1 | 57.515 | | | | 2.826145337 | 2.62925964 | -0.196885698 |
| PTET.51.1.P1550131 | | Feline leukemia virus subgroup C receptor-related protein 1 | 16.231 | | | | 1.051504885 | -0.578287127 | -1.629792012 |
| PTET.51.1.P1560087 | | Ribosomal protein S12/S23 | 12.105 | SMC4_1 | SMC4_2 | | | | -0.997217404 |
| PTET.51.1.P1570010 | | 50S ribosome-binding GTPase | 33.631 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1570012 | | Cyclin-dependent kinase 3 | 24.144 | | | | -0.616313758 | 3.139649555 | 3.755963312 |
| PTET.51.1.P1570072 | | UBA-like | 11.648 | | WT | SMC4_1 | -1.805547921 | | |
| PTET.51.1.P1570128 | | ADP,ATP carrier protein | 173.12 | | | | 1.282219882 | 0.233600387 | -1.048619495 |
| PTET.51.1.P1580009 | | Thioredoxin-like fold | 48.547 | | | | 0.903770202 | 0.493540044 | -0.410230158 |
| PTET.51.1.P1580012 | | Ribosomal Proteins L2, C-terminal domain | 33.415 | SMC4_1 | SMC4_2 | | | | 0.28592987 |
| PTET.51.1.P1580031 | | Ribosomal protein S17 | 33.724 | | | | 0.002319576 | 0.350455343 | 0.348135767 |
| PTET.51.1.P1580060 | | Palmitoyl-acyl carrier protein thioesterase, chloroplastic | 12.935 | SMC4_1 | SMC4_2 | | | | -1.964528414 |
| PTET.51.1.P1590096 | | Fez1 | 11.146 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1590123 | alpha-51D, sAG_alpha51D | Paramecium surface antigen | 89.221 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1600056 | | Poly(ADP-ribose) polymerase, catalytic domain | 17.166 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1600099 | | Protein kinase-like domain | 32.104 | SMC4_1 | SMC4_2 | | | | -1.541738597 |
| PTET.51.1.P1610030 | | Thiamin diphosphate-binding fold | 20.535 | SMC4_1 | SMC4_2 | | | | -0.044620492 |
| PTET.51.1.P1620066 | | CTD small phosphatase-like protein 2 | 11.817 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1630038 | | Probable mitochondrial transport protein fsf1 | 18.25 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1630088 | beta-PT2, bPT2, SU3, tub_betaPT2 | Tubulin | 323.31 | | | | 1.248976646 | 1.519761843 | 0.270785197 |
| PTET.51.1.P1630095 | | Respiratory-chain NADH dehydrogenase 51 Kd subunit | 24.358 | SMC4_1 | SMC4_2 | | | | -0.410413817 |
| PTET.51.1.P1640075 | | Protein phosphatase 1 regulatory subunit 3B | 18.766 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1640126 | | SF-assemblin/beta giardin | 18.857 | | | | -0.750119447 | 0.83535631 | 1.585475758 |
| PTET.51.1.P1650072 | | Ubiquinol-cytochrome C reductase hinge domain | 12.432 | | | | -0.15242399 | 1.892996119 | 2.045420109 |
| PTET.51.1.P1650092 | | Ribosomal protein L6 | 18.464 | SMC4_1 | SMC4_2 | | | | -1.157853725 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P1660054 | | Protein kinase-like domain | 56.732 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1660069 | | 60S ribosomal protein L21-B | 20.026 | SMC4_1 | SMC4_2 | | | | 0.31336729 |
| PTET.51.1.P1660073 | | Heat-inducible transcription repressor HrcA | 14.394 | | | | -0.421504265 | -0.672641633 | -0.251137368 |
| PTET.51.1.P1670032 | | Eukaryotic translation initiation factor 3 subunit I | 21.048 | SMC4_1 | SMC4_2 | | | | 0.806905964 |
| PTET.51.1.P1670039 | | Chalcone--flavonone isomerase | 46.469 | | | | 3.250695806 | 2.537188173 | -0.713507633 |
| PTET.51.1.P1680035 | | Trimeric LpxA-like | 39.857 | | | | 1.240003353 | -1.896899339 | -3.136902692 |
| PTET.51.1.P1680046 | | Virilizer, N-terminal | 57.565 | | | | 3.5619662 | 3.399349516 | -0.162616683 |
| PTET.51.1.P1680080 | | Ribosomal protein L28e | 19.239 | SMC4_1 | SMC4_2 | | | | -0.656113316 |
| PTET.51.1.P1680081 | | Ribosomal protein S4e | 24.207 | | | | 0.415201384 | -0.22081778 | -0.636019164 |
| PTET.51.1.P1700045 | | Zonadhesin | 323.31 | | | | 2.455499603 | 1.142054361 | -1.313445242 |
| PTET.51.1.P1700072 | | Sulfide:quinone oxidoreductase, mitochondrial | 55.399 | SMC4_1 | SMC4_2 | | | | -3.355770293 |
| PTET.51.1.P1710013 | | Alcohol dehydrogenase superfamily, zinc-type | 17.303 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P1710019 | | ClpP/crotonase-like domain | 135.52 | | | | 0.385598398 | 0.264768192 | -0.120830205 |
| PTET.51.1.P1710064 | | 40S ribosomal protein S5 | 31.335 | SMC4_1 | SMC4_2 | | | | -1.992516361 |
| PTET.51.1.P1730008 | | Probable butyrate kinase | 18.541 | SMC4_1 | SMC4_2 | | | | -1.105734912 |
| PTET.51.1.P1740041 | | Dynein heavy chain-like protein PF11_0240 | 11.894 | | WT | SMC4_1 | -0.498329767 | | |
| PTET.51.1.P1750015 | | Ribosomal protein L10/acidic P0 | 65.449 | | | | 1.716416174 | 1.89241648 | 0.176000305 |
| PTET.51.1.P1750049 | | Coiled coil domain | 16.837 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1750074 | | EF-hand domain pair | 17.706 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1780068 | | PCI domain | 11.547 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1790015 | PTETG17900001001 | P-loop containing nucleoside triphosphate hydrolase | 50.163 | | | | 2.56757276 | 2.268188729 | -0.299384031 |
| PTET.51.1.P1790028 | | Polynucleotide 3'-phosphatase ZDP | 13.979 | | | | 0.532951296 | -1.810747817 | -2.343699113 |
| PTET.51.1.P1820010 | | Clathrin heavy chain 1 | 238.04 | | | | 2.865940145 | 1.197758278 | -1.668181867 |
| PTET.51.1.P1820013 | | PKHD-type hydroxylase AZC_3753 | 28.618 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1820041 | | Thioredoxin-like fold | 25.016 | | SMC4_2 | SMC4_2 | | | |
| PTET.51.1.P1840023 | | Coiled coil domain | 25.332 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P1840047 | PTETG18400001001 | Protein kinase domain | 323.31 | | | | 4.538036906 | 1.1679505 | -3.370086406 |
| PTET.51.1.P1850001 | | DNA double-strand break repair Rad50 ATPase | 12.299 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P2080011 | | Lipoamide Acyltransferase | 19.841 | | WT | SMC4_1 | -2.207487422 | | |
| PTET.51.1.P2150003 | | 6-phosphogluconate dehydrogenase, NADP-binding | 37.23 | | | | 0.445928526 | -0.978808082 | -1.424736608 |
| PTET.51.1.P2410004 | | Protein kinase-like domain | 29.443 | | | | -0.341831096 | 0.957945197 | 1.299776293 |
| PTET.51.1.P2530004 | | Probable xyloglucan galactosyltransferase GT19 | 12.129 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P2550005 | | rRNA methyltransferase 1, mitochondrial | 36.057 | | | | 0.572601156 | -0.434777814 | -1.007378971 |
| PTET.51.1.P2710001 | | 40S ribosomal protein S14 | 11.721 | SMC4_1 | SMC4_2 | | | | -0.327344848 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| PTET.51.1.P2710005 | CAP domain | 11.386 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P2740007 | Homeodomain-like | 30.069 | SMC4_1 | | SMC4_1 | | | |
| PTET.51.1.P3030001 | Molybdopterin synthase catalytic subunit 2 | 44.243 | | | | 2.484049115 | -0.81146107 | -3.295510185 |
| PTET.51.1.P3490004 | P-loop containing nucleoside triphosphate hydrolase | 64.813 | | | | 3.15236338 | 2.550851555 | -0.601511825 |
| PTET.51.1.P4420003 | Cytochrome C1 family | 74.259 | | | | 2.448505491 | -0.540811163 | -2.989316654 |
| PTET.51.1.P5040001 | Ribosomal protein S8e | 14.449 | | | | -0.28599972 | 1.575060577 | 1.861060297 |
| PTET.51.1.P5560008 | Arginine-tRNA-protein transferase, C-terminal | 25.684 | | WT | SMC4_1 | 0.786201325 | | |
| PTET.51.1.P5560045 | Maintenance of mitochondrial structure and function | 12.939 | SMC4_1 | SMC4_2 | | | | -0.323164633 |

## Supplementary sequences list

BgIII specific 3-IES candidates

**ID 390**

Left IES ID=IESPGM.PTET51.1.60.189390
sequence=TACAGATCTTAACTTAATTTTTAATATTTTTAATTTTGATCGCTAAAACATATTATATTTTTAATTGAAAAAGAGATTCTAATGATTCCATAAGTATTGAAAATAGATTCGAATAATTGAAAAAGAAAAAAGAAATCATGTTTATTG

Middle IES ID=IESPGM.PTET51.1.60.189879
sequence=TATAGCTAATAACATATGTTTTGTAATTAATGTCGCGTTTAAAAATAATTTGAGATCAATAAGTTGAATTTTACTG

Right IES ID=IESPGM.PTET51.1.60.190332
sequence=TACATTCAATTAAGTAAAATTTACTGATGATCCTTGATATTTAAAATTACAAAATAATACCATTAATGAAAAGGCGAAAGAATAATATTTATTTTTTCCATTTTTTTGAATATAATAAAAAATCAAAATATATTTCAAGATCTG

Region sequence

TAACTTAATTTTTAATATTTTTAATTTTGATCGCTAAAACATATTATATTTTTAATTGAAAAAGAGATTCTAATGATTCCATAAGTATTGAAAATAGATTCGAATAATTGAAAAAGAAAAAAGAAATCATGTTTATTGTATTGTTTGATTGGCTAGAATTGCTTCCTTCGAGTTGTTTCTTATTGGACTAAGTTTTGAGGTATTGTCACTGCTTTCTTGATTATAAAATCAAATTCTTGTTAGAATAGAGTATCATTTTATTGAATGATGCAAATGGAGAATCGACTTATCTAAACTATAAAATTTATGATGAAGAACAGTATAATAAAACTATATTGAAGGCATAATCCTTCTTAGCATTTATGAGGATCTATGCCTCTTAGATCGTCCAATAATGTGGAGATTAATTTTTATCAGCCATAGATAAATTATTATAATAAAACAATTTTGATATCATTGAAAAAAATAAGGATTATACAATTTTAGAATTGCCTCTAAAAGTTGCTTAATTATTTAAGAGTGGAAATAAAGAAAATACATAATTAGCAAATACTTTGATGGTTGATTTCAATAAGAAAATGAAATAATTAAAAAACTCTGATTTTGATGTTTCTGTGTTAGAAGAGGTATAGCTAATAACATATGTTTTGTAATTAATGTCGCGTTTAAAAATAATTTGAGATCAATAAGTTGAATTTTACTGTAGGAGCTAATCTTGGATATGATATTTTAGCAGATCCAGTAGAATTAATGGAAATGAAAAATTTACTAAAGATAAGAATTAAAAATAAGGCATATTCACTTGAAACATATGAAAAATATTTTGATTATCAAATGAAACTAGTCTCCTTGATTAGACATTATGATGAAGAATTGTCTATGTAATTAATTTTTGAAAAGATACTATCTGACGATCCATAAAATCTCAAAGTTCAGATTAATTATGATTGGCTTAAGTTTTCAGAAAATAAAATCTCACAATCACAATTGATTGAAATATATTAAAAATATCTCCCTGCCCCACCAAAGATTGAACAAATCATTTACAAAAGAATAGGAACACTAAAATTAATAAATAATTAAATAAATGAAGCAAAAGCTTGTTTTGAAAAAATATAAGGTGATGATTACGAGGCCAGTTTGGCGCTTGGTATGATACATTCAATTAAGTAAAATTTACTGATGATCCTTGATATTTAAAATTACAAAATAATACCATTAATGAAAAGGCGAAAGAATAATATTTATTTTTTCCATTTTTTTGAATATAATAAAAAATCAAAATATATTTCA

LEFT_PRIMER: TGAAAAAGAGATTCTAATGATTCCATAAGT

LEFT_PRIMER_START_LENGTH: (56, 30)

RIGHT_PRIMER: TTCGCCTTTTCATTAATGGTATTATTTTGT

RIGHT_PRIMER_START_LENGTH: (1233, 30)

PCR Product size = 1178

**ID 75**

Left IES ID=IESPGM.PTET51.1.86.181075

sequence=TACAGTGGTTTTATACAAATATTTATTTATAAAAATATAATCTCATATTTGAAAACAACATAAATTAATTAATGATAATGAAAATAAGATCTATAAATATTGGGGAATTCTAACTACCTTGGCATCACTTTAACTGATAAATCTCATCTTTCTAACTAGAGTAGTGAAATAAAGGAAGAATTAATTATTTTTCAAATTGACTTAATGGGATTTCTTTTAAAATGAATAGCAAAAAAGAATTATTTGACTTACTCAAATTTAGTTATCTAAAACAACTCACTTTAAAAACTCCAATAGAGTAATCTGATAAAACTATAAAGAAGTCCATGAAGATGATAACCAGAGTGCCAAAATCAACACCAAATTGTATTATTGATATACTAATCGAAATTAATAGAAATAATTAGGTTCCTAAGTGCCTTATAAATTTAAAGAGATAATTGAAATCATTGAGGGGAAAACGAAGAAAAGCAGATCTGTCAAGATGAAGATACAAGTGGAACAAGCGATAACAGATATCAATGAATAATTTATTGTATGTTTAAGCTCATTAACAGACATAATAATTATAAGATTCAATAACCAAAATGATTGGTTATATAAATTAATTGTTCTAACAAAAATATAGTATATTATATTTGAATTTCCCAATCCCTATTTTATTCTATATTATAAATTTAGGGGTTTTCCTCCTTTAAATTGGTTGTTTTATCCCTTTTATATTAGGTAAATTAATCTTATACATTTCTAATTGGAATATTATATAAGATTATACAAATACAAATCTTTTGATATAATTGTATGAAAACACCTTAGAGAAACTTTTCTTAAATTTAATTAAATAGAGAAAGATGACAAGATTTAAAGTAATTTAAAAAAAATTTAGCTTATTTTATCAAATAAATAAACTATAAGCAAAACTGATGATTTAGAAACAAAAAAATATTATTTACTTAAAAAAGTAACAATTTTAAGTAGCCTCAACTG

Middle IES ID=IESPGM.PTET51.1.86.182566 sequence=TATAGTCTGTTCTTAATATTCAAATATTCATAATAAAGTAATATCGACTG

Right IES ID=IESPGM.PTET51.1.86.182821
sequence=TATGTTTCTTGCAAAAAAGAATTTATGGAGTATCAATTATTTAATTATAAAATGATTATGATGAAATTCCTCAAAATTTGATATCTTATAAATAGAACTACACATATATAAGATAATTGAAAAACAATGAATTATTTATAAATTTCCATCTATATGTCTCTTCCTTCTAGGAAAATCTTATATGCAAGATCTA

Region sequence

GTCAAGATGAAGATACAAGTGGAACAAGCGATAACAGATATCAATGAATAATTTATTGTATGTTTAAGCTCATTAACAGACATAATAATTATAAGATTCAATAACCAAAATGATTGGTTATATAAATTAATTGTTCTAACAAAAATATAGTATATTATATTTGAATTTCCCAATCCCTATTTTATTCTATATTATAAATTTAGGGGTTTTCCTCCTTTAAATTGGTTGTTTTATCCCTTTTATATTAGGTAAATTAATCTTATACATTTCTAATTGGAATATTATATAAGATTATACAAATACAAATCTTTTGATATAATTGTATGAAAACACCTTAGAGAAACTTTTCTTAAATTTAATTAAATAGAGAAAGATGACAAGATTTAAAGTAATTTAAAAAAAATTTAGCTTATTTTATCAAATAAATAAACTATAAGCAAAACTGATGATTTAGAAACAAAAAAATATTATTTACTTAAAAAAGTAACAATTTTAAGTAGCCTCAACTGTATCAATCAATTTATTTTAAAATTCTTATGAACACTACATTTTCAAAACAGGCTTCCTTGCATTATCAATTACTTTGACATCCTTAATAAAGCCTTGTATCTAAAACTATATACCATAATACTTTGAATTATTCAGCAATTTTATACAAAGAGATCCTTTCTTATTCATTTGTTTCTATTACAATATCAATATCAGAGTCAAGTAAATTTAACCCCGTATAGCAAGAACCGTATAAAAAGGCACGAGATTTAGAAGATACAGTTTGAACAATCTAATAATTGAATGAGAATAGATCATACTTCATCAACCATCTACACAATTTTGTCAATAATTGGCCTTTATTCTTCCAATAATTTATTCAGCTCTTGAGAGAAGGTATTGATTTCAAAAGATAATTACCTAAATTTACTAATATGGTTACTTAACAATGTCATCTAAGGTAGTTCTGTGTTAAGGTAAAGTAACGATGGTTTCAGAGATCAATTAATCATAAAATGGTTGATTTATATAATTATGCTCTGGAATTTATGATTCCGAGTCCTTCTTTTCTTTGTTTTCTGTTTGGATTGCTTGATCATTGAATACTTTGGGTGGTTTA

TGAAAAATTTCACAACAATCCTCATATTTGGATGACTTCTGCGAATTTTAATTATTTTCAACTAATTCTTCTTACTTTCTTTTTCTATTTGGG
GATTTTTGAAATTGAATTTCTTCTTCAATAAGTTAGTTTCTTGTTTCTTGATTGTAAGCAACATCTAATGGTGTTTGATGTTGTTAATTTTTT
TTTTTATAGAGATTTTCTTTTTGACAAATACTTTTTATAATTTATAAATTGCCCAATCTAGATGCTACATGTAACGCTGTGTCTCCATTTTCAT
CCTGTCCAGAAGCCCCATACTATTCAATTATCTTTTGTGCCGTTTTAAAATTGCCTTTACGCAGTGATTTGATTAAGTGTTAAGGCTTTTCA
AGGGAGGTGAGTTGAAGCAATGTTTCCACAATATCCCAATTATCTTATTCTGATTAAATCGCGATATCAAGTGCTAGTTTAAATTATTTTTT
CTCAAACTTTTCTTTGTATTTTATAATTTCATTTAATATGCTAATTTTGCCGCTTTTTGCTGCCTAGATCAAAGCAGTGTTTCCTTGATTATC
CTTATAATTGAAATCAATTTATTATCTTGATATTTTTGGAAAATATTTACTTTAAAATTTATTAAATGCATTATAAGTGTTCTTCACTATCTCTA
GTATAATAGATGCATTCAACTAAGGCTTATGAAAAATTCATTTATGCTGTTTAATTTACTCGTATATAATCTTCTAATTAAACATATTCACCTA
AATTTCAAATGTAGTGGTACCTTGATTAGTAAATTCGACTATTTTTTATTAGTGAATAGATTATTATTACATATCGTGATTTTTAATAATTTAT
TTAATACATTAACAAAAATTCTCATTATAATAAAACAGTCCATATATACAGATATAGTCTGTTCTTAATATTCAAATATTCATAATAAAGTAAT
ATCGACTGTAATAATAAATTTAGTAAAAACAAAAACTGTTTTATCAAATATGGATTGAATCAAAGTGGCTAAGGTGATTCTTATAAAATAA
GTGTCTTTTGCCCGTTTAGCAATATGTGTATTAATTCAAATAGTCTCATGTTTCCATCGAGTTTCTACGTTTTCACTCTGCCCACTAAGATA
GATAAATAAGGATGTAGAAACTTCCTTTTTTAATTAGTTTTATTAATTCTATCCCAATACGCTTGTGAAATTAGGTTTGATATGTTTCTTGCA
AAAAAGAATTTATGGAGTATCAATTATTTAATTATAAAATGATTATGATGAAATTCCTCAAAATTTGATATCTTATAAATAGAACTACACATA
TATAAGATAATTGAAAAACAATGAATTATTTATAAATTTCCATCTATATGTCTCTTCCTTCTAGGAAAATCTTATATGCA

LEFT_PRIMER: GGTTTTCCTCCTTTAAATTGGTTGTTTTAT

LEFT_PRIMER_START_LENGTH: (204, 30)

RIGHT_PRIMER: CCTAGAAGGAAGAGACATATAGATGGAAAT

RIGHT_PRIMER_START_LENGTH: (2474, 30)

PCR Product size = 2271

DraI specific 3-IES candidates

**ID 377**

Left IES ID=IESPGM.PTET51.1.18.134377
sequence=TAAAGCAGGCCCATTTTATCTCCCATAAATTTCCTAAACTGGAAATTCATTTAACATGTATACACTTGGGTGAATTTACCTTTTG
ACAAATTAATCATTTCATATTTGAAACACAAAGAATAATTTCTTATTATTTAAAGATAAATATGAAATTTGTTGGTGAATTTTAATATTTAGA
AAATCATAGTAAACATTTTGCTTAAAATATTGTTAATTCAATGAATAAGAACTGTATATATTATAATTTTTACATTTAATAATCAAAATACGA
AATATTAATTGTAAATTCACTCAAAATGTATTTTAATTG

Middle IES ID=IESPGM.PTET51.1.18.134388

sequence=TACACTCTTATATATTATAAATAAGTTAATTTTATAATTATATATGG

Right IES ID=IESPGM.PTET51.1.18.134467
sequence=TACAGTCGATCTTATATTATTAAACATTTTTGAGTTGAACTTTTATTCATCTGATTATTTTGACAATTTACCAAATCCTCTTTTCAT
CTACTAATTTATATTTGAATTCATAATCCAAGTTAATTTAAAATTATAGGTGATATGTTAATGAATGGGTTTGTATGATACTTAATGTACATTG
ATTAATCCATTATCTTTTATGCTGTTTAAGAAAAATATGTCTTAAATCAAAAATTATTTTGCAGCCTTAATTAAATCAATTACATCATAAATT
ACATTTCGATAAATTCATTTCGAAAAATATTTTCTATATAGATTTTAGATTATAACAATAAATAAATTTGGATTATCAATTTACAATTTCTAAT
TGAAAAATGTTTAATAATATAAGCTCTTCTG

Region sequence

GATAAATATGAAATTTGTTGGTGAATTTTAATATTTAGAAAATCATAGTAAACATTTTGCTTAAAATATTGTTAATTCAATGAATAAGAACT
GTATATATTATAATTTTTACATTTAATAATCAAAATACGAAATATTAATTGTAAATTCACTCAAAATGTATTTTAATTGTAGTTATGAACTACA
CTCTTATATATTATAAATAAGTTAATTTTATAATTATATATGGTATCATCTACAAATACTGAACTAAACCTTATAAAGATATTAATATGAATTTT
AGATTTATTTTTGAAGAAATCTTTAGGTACAGTCGATCTTATATTATTAAACATTTTTGAGTTGAACTTTTATTCATCTGATTATTTTGACAA
TTTACCAAATCCTCTTTTCATCTACTAATTTATATTTGAATTCATAATCCAAGTTAA

LEFT_PRIMER: ACGAAATATTAATTGTAAATTCACTCAAAATGT

LEFT_PRIMER_START_LENGTH: (128, 33)

RIGHT_PRIMER: GATGAAAAGAGGATTTGGTAAATTGTCAAA

RIGHT_PRIMER_START_LENGTH: (395, 30)

PCR Product size = 268

PciI specific 3-IES candidates

**ID 144**

Left IES ID=IESPGM.PTET51.1.6.380144
sequence=TACATTCACAATTTCATTCAATAATTTTTTTAATAAAAAGAGTTTATTATCTCTTTTATATGAAATTTAGTTTCATAATTATACAA
GAATTATGTAAAATTAAGAAAATAGGTAATAATGCCATTAATTATTTAATGAAAATAAAGCAAGAAAAACTTAATTTAAATTTGAACATCT
CAAAAAAATGACAATATTTTCTCATAAAAAACCATGCAAAAATTTGCTCTTCGAAAATAATATATAAATAATTAATGTCATTTAAGAGATTA
GAAGATTCCTCTAAACTTTAGATGAATCTTATTTCACTAATTAAGATTTTAATTTTGATTTACGTATGTCATTGTTTATTAATGACATGTAATT
GATCTTAAGTGACATATTTGCAACTATTTTGTATAAGAAAGAAATCTTTAAATTTTTTGAATTATTTATTTAAACATTTTTTAAATATCTTCAT
AAATTTAATAAAATATTTAGAAAATATAGTATAAAATATTAAAAAAAAAAATAATAAAACAATTAAAGAAAAACATATAAATTATAACTTTCA
ATATTTTTCATATAAAATTAATGAAAATTTACTAAAAAATAAATGTTGATTTTGCTA

Middle IES ID=IESPGM.PTET51.1.6.381684

sequence=TATAGAAGATAAGACAGAAATGAGCTA

Right IES ID=IESPGM.PTET51.1.6.382023
sequence=TATTGCTTGATTAATCACTCAGTTTGGATTACTCGATAATAAACTATCAGTTTATCTAAGCTAATTGTAAATATAAAATCCATATTA
GTTTAATAATTTATTTTAGTTTAAGAGATAAAATTGAACATGTAAAATATATAAAATCTGAAAAACTTATTAGTATATAAACCTAAAAATCTC
ATTTGGAATTGGTTGACAA

Region sequence

AATTGATCTTAAGTGACATATTTGCAACTATTTTGTATAAGAAAGAAATCTTTAAATTTTTTGAATTATTTATTTAAACATTTTTTAAATATCT
TCATAAATTTAATAAAATATTTAGAAAATATAGTATAAAATATTAAAAAAAAAAATAATAAAACAATTAAAGAAAAACATATAAATTATAACT
TTCAATATTTTTCATATAAAATTAATGAAAATTTACTAAAAAATAAATGTTGATTTTGCTATATTAGTGGATAGCAAAATAAGCCTGAATCG
CATTGCATATTCAATTAAAATCATCTTTTTCCAACATATTTAATCAAGAAAATAATATTGAGGGTGATTTAACAAAGATATTAAAGCTTTTT
AAGTATTGAATTCATATACGAATTCAATACTTAATATTATTTCTTAATAGGAATCTTTTTAATAAAGATGCAGATCTTAGATTGATCCATAATT
TCAATGACATAGCTTAAGAAAATTCATATAATTTAAATCTTGATAAGTATTAGAAAATGCTTTATTAAGCTTGGATACCTTTTTCAATTTAAT
AGTTTGAATATTTATTTTCTTTAGATTTATCAAAAAATAAATTATAAATTAATATTAAATTATTTTGATAATTATTTATAACAATAATTACTCAA
TCATTTTTTGTTCATTGTCCCTCCATATTTTACTTTTCCTAATAAAAACTCACAAACAGCCACAGGTGCTTTTTCTTTCATTTCATCCAATG
AAATCCACTTTTAATCAGTTGAATCTTCGAAACACACTAAAAAATCTACTCTTCCATTATCGTGTCTAACTTGTTTGACTTAACAAGGTTT
ACTTTTTCTTGGCTTATGAGGACCATTTCCTGTATTATTGTTAAGCACATTATCAACATCTTTCTATATTTACTATTGTTTTGGTTTATCTGAA
ATTTTTTCTTGAGGCTTTTCAATAATTTTTTCAGTTTTCTCAACTTTATCCACTTTGTCATTTGACCTTTTTTCTTTTTTTTGTTCCTCTAATT
TTTTCCCATTATTGTTTTGATTGATAATTTAAGTGTCTTGTTTTATATGTTTTTGAGGTTTAGGTTCTTGAGGGTCATGCCCTAATTGTTTTA
ATTCATAATCATGGACTAGAACCAGCATTTCAGGAGTTAATTGAGTCATGGGTTCAATTGATACTGAGTTGTTACTCCATCTGAACTTATA
CATGATGGCTTCAGGTTCAATCTTTTTCTTCAATAATTCAACAGGATGTTTTGAATTCATAATAGATAAATTAATTATTAAGAGAAGGCTCT
ATCAATGTTTATGTAGGACAATTATTGCAATGGTTTATCAAAATATTATATAACATTTATTGTTTAATTTTAGTTTGAGTTTCATAAAAAAAA
TACACAATATTTGAAATAATAAATAATTTTATAATTTCTTTTTTCACTTCTATAATCTGGTTATAATATTTGTTTATTTTATTTATAATCTTTAAA
AAATTGAATAATTTAAATATAAAATGGCAAATATACATAGATTGGGTGATAATAATGATTAGTAAAATAATTATCAGAACTTGGGGGGTAAT
GGAAGACAACAGATGATTTCTATATTCGGTAAGCAAATTCATTTAAAGATATCAAGGAGGGATGGATGCAGACCCAAGAGGTGAGAATT
TTTTTGATATGTTGAAAAAATCATTCTGTCCAAGACTTAAACTTATTTCATTTACTACAATAGTCTCAGCATTGATTATAATTCTATACATCA
CAATGTTAGGAGTTGGAGGAATAAATATAGAAGATAAGACAGAAATGAGCTATAAACATGATCGATTTTTATCAGTTTATGAGAAAACCT
TGAATGATTTTGGTGGAAATGATCCAGATGATGTTAAATCCAACTATGAAATTTTCAGATGGGTGACATCCTTACTCTTAGTTGGAGACTT
CTATAATCTAATACTTGCGATCTTTATGATATTAATATGCTATTCGATATTAGAAGCCACATAAGGATTGAATTTAACATTAATAGTGTTTTTT
GGGGCAGGAGCATGTGGTGCCCTATTTGGAGATTTGTGCAATATTTGCAAATACAGAACTTACTCTGAAACCTTTTCTTGCATATACGCTT
GTGTAGGGTTCTTGATAGGAGTATATATTGCTTGATTAATCACTCAGTTTGGATTACTCGATAATAAACTATCAGTTTATCTAAGCTAATTGT
AAATATAAAATCCATATTAGTTTAATAATTTATTTTAGTTTAAGAGATAAAATTGA

LEFT_PRIMER: TCTTAAGTGACATATTTGCAACTATTTTGT

LEFT_PRIMER_START_LENGTH: (6, 30)

RIGHT_PRIMER: AGTTTATTATCGAGTAATCCAAACTGAGTG

RIGHT_PRIMER_START_LENGTH: (2197, 30)

PCR Product size = 2192

PsiI-v2 specific 3-IES candidates

**ID 799**

Left IES ID=IESPGM.PTET51.1.33.359799
sequence=TATTGTTTGATATAAATTTAATTTTAAATCAGAATAATTTGTTACTTTTCAAATTTACTTTTTTAAAATTAAATAATTAAAAATTCC
AATTTCTAAATATTTGGAATGAAATTAAATTTTTCAATCTATATTTTTAAATTCTACTTTCAAATAAGCCAAAAAATTTTTACTTAATTTTATA
GCTTTACTTGAAGAATAGCAAACTAAAAAAATAAGAATAAGATCTTCAGTTAATACAATTTACTCAATTATAAAATTTATTAACTATCTTAA
TTTACTCTACTTTATCAAAAACAATTTTTAATGATAAATTTATAGATATATAGGTATAATTATAGATTGATTGAATTAACAGAATGATTTAATT
AATGGATGTTATTCCCTTAAATTGATTCAATAATAAAATAATAATTAGCATTCCTAAAATAATTAGCAACCAATTAAGATTTGCATTATTATAT
ATATCAAAATTAAATTATATATTCTTCTCTTATTAACAA

Middle ID=IESPGM.PTET51.1.33.361301
sequence=TATAGTCTAAAATTTTAAAATATTATAGAATTAGATTTCTTAGATAATAAATCATAAAATGTCATCGATTCAACTA

Right IES ID=IESPGM.PTET51.1.33.361772
sequence=TATAGATGCTGCACAAAGGAAATATGTAGTATTTAAATTTATATCTAAATTAATAATTGGAGAACCATAGTGCTAGTTTTTTAAGT
AAAACAAAAGAGTTAATTAAAATTTAAATTGGTTTATAATAAATTTCATGATAGAATAGAAAGACTATTTAAAATTGATTAAATAATTTATT
AATCATATTCTTTACCAATTATAAGTATTAATTTTTTCATCTCAAAATAACCATCTAGCAAATTCTTCTCTGCAAACACTA

Region sequence

AATTTATTAACTATCTTAATTTACTCTACTTTATCAAAAACAATTTTTAATGATAAATTTATAGATATATAGGTATAATTATAGATTGATTGAAT
TAACAGAATGATTTAATTAATGGATGTTATTCCCTTAAATTGATTCAATAATAAAATAATAATTAGCATTCCTAAAATAATTAGCAACCAATT
AAGATTTGCATTATTATATATATCAAAATTAAATTATATATTCTTCTCTTATTAACAATAGATTAATTAAAGAAATGCAAAATCATTGAGATAA
TATTTTCAGAGGACTTTATTAGAAATTTAGCTTACAAAGCTAAAATTATTAATGCCTATAAAAATGAAATTGTTACCCATCCAGGTTTATAC
ATCGTCATTTAAGGTGGAATGAAAGTATAAATATAAGCTGAATATGAAACTAAATAAATATTAAAAGTAATTTCAATATATATAATTTAGAAT
GGGGATTATTTTGGTCTTATTGAAATGATCTTAAATAAAAGTTAAAATTTATTTCTAAAATCTATAAAAGAAGAAAGTCTTCTAATTTATAT
TTCAGGTGATGATTTTCATCATAAAATAAAAGAGTTTACGGAGGATTATGAAAAAATGAGATTTATACATGATCAATTGCTGTTCAGTTCA
AGTACAACGAAAATCAAATAGAAATGCTATTTTTGTTAATAATATCATGTTCCTTCATAATGTAAAATAATAAATTTCAAACCTGAATTTGA
TTTTGAGAAATTGAAATTGGATGAATAAAATGATAGAAGATATTTTTAAAGAGCCAATAAGAAGCACATTTTTTAAATAAGAAATAATTCA
GAATCATCTAAAAGTTCAAATGAATTTGAAGAGTCCTTTGAGATCGATCCAGTCAGTTCAGAAATGCAAAATGAAAGATAAATCGGTAA
GAATATTTTAGGATCCATTTCTTAAATGATTGTGTGCAATTCAAAAATATCCCAAGAAGGATTCCCATAATTACTTTAAACTTAGAGTTTTA
AGGTATAAGAAAGTTAAACGTAACAAAATCAGGTAAATACAACTTAAGGATGGAGAGTTTTAGACTTAGACAATAGAAATATTTGCATAG
ACAATTTAAGGAATTTTTAGTTTTACGATCCTGTGTATAACATAGAAACTATAATCCGAAAAGTAAATAAATTTTTAGAACATTGAACTTTT
ATTATTCATGATTATTCAACATGATCAGCAACTACATCTACTGGGGCTTCACAAACACCTCCAAAGTATAAGTTGAGTCTGTATCCATTCTT
CCATGCATAACAAGTTCTGTGTTTAATTGATTGTCTTTCGAGTAAATCTCCATTGTCATTCAATAATTCATAAATTGTAGAATCAGCCAAAA
AGATCAATCGATATTAATACCAAACTTCAGTTCTAATCTTTGTAGTGAAAATTCTTATCAGTTGATCAGGATGTTCATAGGGAACTACACT
ATTGTCATAAGCGTAGGCTGCAATTTCTACTAAATCTTATTCTTGGCAATTTGGAACTACGTATAATCCTTGTTCATTATATTTCAAACATGA
TTGGGCTCTTCTGAAGACAAATCTATCGCTATCTTTATGATGGAGTGATCCAAAACATCTAGAATAGCCCCATAATTTATTCCAAGAACTC
ATGCATCTCTATTAGTTTTCATCATGAGGAACAACATAAGTTGCTCCAGTGACACCTAACTTGATTCTTATTTGAGAGATTTCTGAGTTCT
TTTATAGTCTAAAATTTTAAAATATTATAGAATTAGATTTCTTAGATAATAAATCATAAAATGTCATCGATTCAACTATAATAAACAGGTAAT
GTGGGAGTTTCATATTGATCACCTTTTTTGATTGTCAATGTGTACTCTGCTTAATGAGTAGATGCAATCAAAATATATAACAATAATATCATT
AAAATTTATGAGTGATTAATTGATCATTAAAAGGAGGAGAGGTAATACCGTAGAGACCACGAATTAAATTATCACTCTAAATTTGTCAGA
AGCTAATTCCACATTGTTATCACTTAAGAAAGAGATATAAAAGATAAGTATTATTTTTGGAATAATCCATTAGCTAGAATCAAGCCATTAAG
ATAGTGGTTAACAAGCGAAGATTTAAAAACTGTGTATGATGATAATAGCAAACCTTTGAATTAATGTGGATTTAAGAGTGGATAAGTCTT
GGTTGTAAAGGACTTAGGACCTTAGATGCTTTGGATTACAGTATTTTATGCAGAATATGTGGGTCCAATCTTGATGTTTGTTTTGTTGTACT
ATAGATGCTGCACAAAGGAAATATGTAGTATTTAAATTTATATCTAAATTAATAATTGGAGAACCATAGTGCTAGTTTTTTAAGTAAAACAA

AAGAGTTAATTAAAATTTAAATTGGT

LEFT_PRIMER: AGCATTCCTAAAATAATTAGCAACCAATTA

LEFT_PRIMER_START_LENGTH: (159, 30)

RIGHT_PRIMER: TAAATACTACATATTTCCTTTGTGCAGCAT

RIGHT_PRIMER_START_LENGTH: (2329, 30)

PCR Product size = 2171

**ID 655**

Left IES ID=IESPGM.PTET51.1.88.149655
sequence=TATTGTAGATTAATCATCTCTACTCAATTTTAAATTTAATTTGTAACAAGAAATCTGATTTAAATTTATTTCTTAGTTTAGCATTAG
TATCTTAAATAATAGATTGATAAAGATTAAAAAAATTATAATTTTATTCAACTTAATGAAGAGGTTTGTTATTTAGTTATTAACTTATTCTGT
AATCAATTTGCTGTAATAATTTCTGTTTCCATTTAAAATATTCAAATTTTAAACAATTAAATTATTAAAATTGTTTTATATATTTCACAA

Middle IES ID=IESPGM.PTET51.1.88.149727
sequence=TACACTCAATATGAAGTGAATACAATTTAGATTATCAAATAAAACTCATTTTTATTG

Right IES ID=IESPGM.PTET51.1.88.151525
sequence=TATAGCTCTTACACAAAAGATTTATTCAATTTTATAGTAAATTGATATGAAGAAGATTTTTTAATAAAAATATAAGTATCGATAAA
TTTAATTATTATATCAAGGATCAAAAATTAATTATTATTATTCTCTTTAACATATAGTCAACTCAATCAATCATATTGCAATTATTTTTCCAATC
TGTTATCATTTACTAATTAGGATTAGAAATATTTCTTAAATTTATTAAGAGATTTCTAAAAATTTTATAAGTATTAAAATATATTTTTGTATTTC
TGAGTTGGAATTGAAGGACAAATGTAGATTAATTTTGGGATCAGGAAAGTTTAGAATAATTGATACTTCTTTTTCTAGCTAAGTTTTTAAG
AGATCCATAACAACTTTCCAACTAAAAGGTTTAAAGTTTGCTCTACTTTATCTAGCTATAAATTATTTTTCATTGTTATTCACATTTTTCTTC
AAGGTAATTTATTTTCACCATAGAGATGCTTTGTACATTATCTTATTAACAACATTTATGATTCTGAATAATAAGATAAGAGTTAATTAGAA
AATAAAATTTATACAGTAAACAATCAATGACAAGCATATTAATATTGAGTTTTCACAAATGAAAGAAATTATTATCTATTTATTTTGAAAAT
GAAATTACTATTTTCTAAATGGTACTACATAGATGATCTATGCAAGTGCTA

Region sequence

TTTTATTCAACTTAATGAAGAGGTTTGTTATTTAGTTATTAACTTATTCTGTAATCAATTTGCTGTAATAATTTCTGTTTCCATTTAAAATATT
CAAATTTTAAACAATTAAATTATTAAAATTGTTTTATATATTTCACAATACCAATCTACAGATTTATCACTAATTTCAAATGATGTTTATGGG
ATAACTTCTCCTTAATCAAAAGTCTGTTACACTCAATATGAAGTGAATACAATTTAGATTATCAAATAAAACTCATTTTTATTGTATTTGATA
AAATATGATGAAATTTTATTTTTGTTGTCATCTTGTGTATGACTTGATATCAATTTCACTAGAAAATAATATTTACCTTCATCAAGTTATAGA
ATCGATTGAATATGATCTTTCGATGAATTCATCACTTAGATTTCAAATATTTATTTGGCAAGATTAGTTTCATTTTTTGTTCCTTATACTAATA
TAGATCCAGAGCAAGTAAATACTTCCAAGAAGAGTTTTGATTTTTATTACAAAAAGACTTCATAATAGTTTGGTTATCCCATCTTAGTTAA
CTAATATTGTGATATGCTATCAATTAAAGCATTTAACTCACCATTTGTAATTATTACTCCAATAAAAGTCTAATTATGTAAGGCCTTGATGGA
AATTTCATACAAACCAAATTAATCATACAGGTGATATGAGATAGCTTTCGGGTATTTATTAAAATAATTAATAGGTTTGATGGAAGTATGAG
TAATATCAAAATATAGATTTTCAATGGATTGATGAATATTTTAGATTTGTAGAGTTACAAACTTCTATTGATTATTAGATGTTTGAGTGACTG
ATTGATATTAGAGTTTTATAGTCTCCTGAGCTTTTATTTAATAAGTCATCATCTGACCCAAATGAATAGTGTAAGTCTATTTCTCAATTCTAA
CTGCACTGATCATAAAGACAGAAACAGGTACCAATGATTCAACTTTTATTGTAATCAATGATTCTAAGGTTTTCTTTTTTTCGAATAGAAT
TAAGGTGTCCTATTAAATTCTGGATGTCAGATCTGCTTCATCATTATGCAGAAGTTCTTAACCTTCCTTCAAATATAATCTAATGTCCCCGC
TTATCAATGAGACACTGATTTGAACATCAACATTTGAATAAAAGCGATATTGAGAAGACTCTGACTCACTCAAAACTTCTCTCCATTGAA
GATTGTCGTATAGGGTAGTCTAATACTTTTGACCACTAAATTCAATCAAACCTTTACTATTCCCTTAGACGGCTGTAATTTCAATTGAATAA
ACGCAACCCACTTTTTCGCAAGTCTTCACTTTAAATGATTTTTGTTTATCAACAATTTCATCAATAAACAAATGATCATTAATACTAATAAT
AGTTTCATCATCAATAGACTCCAATGCTTTTGCCCTGACTTACAATTTACCTTGGAACGCAAACACAATGATTTATATTCCTTATCATTTG
GATCATCTCTAAAAATGTAATAATGAGGTATGTCTTCATGGAATAGAACATCAAATGGATAGCCAAGGAAAGCATTGACAAATTTTGCAT
CGTCAGTATTGTAATTAATTTAAAATTCACATTCTTCTTCCCCAAACACGCCTATGATATAATCATCCTTTGCCAAATATCGATTATTCACCT
CTTATAGAATGTAAGTTGTTCTAAAAATATGTGACCACTATTGATTAAATTCAGATAGGCTGGGTCTTTAGTCTTTCCCTTTAGATATTCTA
GCATTAACAAATTCACCACCTCCTTTGACATGAATATTAACTTGTATCGACTTAACATCTTATGTTAAACTGAGTTTGTAATATTTGGTTTA
ACCTTTTTAAATAACGTCGATAATAGGGTCTGATAGAGTCAATAGTTTAACTTCTTCATAGACACCAATCGTATAGGTGTCATTTTAATAAT
GAGGCTCTATAGGTTAAACACGGATGTATAAAACCTAATACTTCTCAGTTTCAGGGATTGTAAAACTTTAAAATTAGGTGAAATAGGAAT
GTGCTTTGAATGTTTCATATCTCAGATCAGAATTATCTTTGTTATAGCTCTTACACAAAAGATTTATTCAATTTTATAGTAAATTGATATGAA
GAAGATTTTTTAATAAAAATATAAGTATCGATAAATTTAATTATTATATCAAGGATCAAAAATTAATTATTATTATTCTCTTTAACATATAGTC
AACTCAATCAATCATATTGCAATTATTTTTCCAATCTGTTATCATTTACTAATTAGGATTAGAAATATTTCTTAAATTTATTAAGAGATTTCTA
AAAATT

LEFT_PRIMER: TCAATTTGCTGTAATAATTTCTGTTTCCAT

LEFT_PRIMER_START_LENGTH: (54, 30)

RIGHT_PRIMER: AATTGCAATATGATTGATTGAGTTGACTAT

RIGHT_PRIMER_START_LENGTH: (2237, 30)

PCR Product size = 2184

SspI specific 3-IES candidates

**ID 795**

Left IES ID=IESPGM.PTET51.1.9.406795
sequence=TATTATTAATTTATATTTCAAACGCTTTTAATTTTAAATATTATTTAGTCATGAAATCATAAGTAATTAATAAATCCTCAAAAAAAC
AACAAAATCCCATAACATTTAACGTTTATACAAATAAATCCAATAATTTTTTAATTTGATAA

Middle IES ID=IESPGM.PTET51.1.9.406958

sequence=TATATAATTAAATACTTATAAATACAT

Right IES ID=IESPGM.PTET51.1.9.407393
sequence=TATAGTTGATTCACAGATTATTTTTATTTTAATTCCTTAAATTTTTATAAATTACTTAAGATATTAGGGTATTTAAATAAATCAGTTT
GTGCTTGAGTAGTAAAAGGTGATAAATATTCAGATTTCTTATTAAAATATATCTTAATTCAAGTAAGAAAAAATATCAATTTTAAAAATAATTA
ATTATTCGAATTTCTATTGAATTGACTA

Region sequence

ATTTAGTCATGAAATCATAAGTAATTAATAAATCCTCAAAAAAACAACAAAATCCCATAACATTTAACGTTTATACAAATAAATCCAATAA
TTTTTTAATTTGATAATAGCTACGCTAAGCTTCATACAAGTAAATCAGCTTCCAAAAAGAAGCCTAATTTAATGGAATGTAAAAACTAAAT
TTAATGTACAAAAGAAACAAAAGAGATAAAATAAAATGATATTAAAAAGATTTAATCAGGGAACAAAAAACAGAGTGAAACTACAGAT
TATATAATTAAATACTTATAAATACATTATACATTCAGACCGAAAAAAAATGAAGTGCCAATAGATAAGTAAAATAGAGCAAGGAAGAATT
AAAAATAGAAGTAGAGTATTGATCTTTGTGTCAATACAGATTTGAAGGAAATTATTCACATTATTAACGACTTAGATTTATTTAATAATAGA
TTAAAAAAGAAGTCATGATGAGAATTAATTTTGAATATAGACTATTTATTTATTAAATGACAAATCATTATTTAATTTACTTAAAAGCCTATT
ATAAATATCATTTATGAAGTAATAATTCTTTTTATTTGAGTATTGCATATTCTGAAACAATCTATTTCAACCATTTCCAAACCTATTTTAACTG
AGGTGGCGCTGGGAATTGCACTTTAGAATGAAAGCTCATAAACATTTAAATTAGAGATTCAATGAATTTCAATTCAGAAGATCTCTTTGT
ACTATAGTTGATTCACAGATTATTTTTATTTTAATTCCTTAAATTTTTATAAATTACTTAAGATATTAGGGTATTTAAATAAATCAGTTTGTGC
TTGAGTAGTAAAAGGTGAT

LEFT_PRIMER: AACAACAAAATCCCATAACATTTAACGTTT

LEFT_PRIMER_START_LENGTH: (42, 30)

RIGHT_PRIMER: TCAAGCACAAACTGATTTATTTAAATACCC

RIGHT_PRIMER_START_LENGTH: (827, 30)

PCR Product size = 786

**ID 453**

Left IES ID=IESPGM.PTET51.1.52.211453
sequence=TACAGCTGTAAGAGATTTCCTTCATTTTTGAATAGAACAAAAATTCAATATTGGATAGAGTAAGTTACTAATACTAATCTAAGTT
ATAATTAAAATTTTACTTCTAGTAAATAAAATAGACAGTTGAAATCACTAAAAATCCTTTAAGAATGTCTA

Middle IES ID=IESPGM.PTET51.1.52.212381

sequence=TACACTACAAATAGTTCACATTCTTAATTTAAATAATTATAACAG

Right IES ID=IESPGM.PTET51.1.52.212637
sequence=TATAACTTTAGCAAACAATATATATATTTTGTCTAGTAGTTTCTATTAATTTAATTAGAAAATAAATAGCTAATTAATTAAAATATA
ATTTATGGTGAAAAAGCAGAAAAACCTGTTTAAATATTAAATATTAGCCCAGCATTTAAATTATTTATAAATTTATGATTTTCATTATATAAC
TTCTGGAAATATTCAGCTTGTGAGAACAA

Region sequence

GGATAGAGTAAGTTACTAATACTAATCTAAGTTATAATTAAAATTTTACTTCTAGTAAATAAAATAGACAGTTGAAATCACTAAAAATCCTT
TAAGAATGTCTATACTATTCTCCATTATAAATCTTATTATTTATAGGAATTTCAACAAAATAAACTTCTTCGTAATCTGGTTTTTCATACTCAT
CTTTAGATAAATCATTGTCTAAATACTCTTCAAAATCAAGAGATGAATCTGTTTTTGCTAAAAATGATAATCCATTATCAATCAATTTATATT
TAGAAATTTCAACAAATCTCTACATTGCCCTAAAACATGTTTAGGCATACAAATAAGAAACTAAAGGGTTTACATCTTAATCTAAATCAAC
TCCATAATCTTCTTCAATTTCTGATCTAGTTTCCATTTCTTTCATCATAAGTTTTCTTCCCCCTTCTTCAGTTAAATTAATTTCAAGTTTTTCT
AAATAAGCCTCCTTTATAACAGCAAACAAGACTAAATTATAAGGAATGTTGCCCTAAGCATCGCATTTGAGTTACATATCCATAATTTTGA
AAAGTGTATTCCTTATCAAATCTGCTTGTATATTAAATTTAAGTTGATTTATTTCTTCTTAAGTGTGTTACGGATTGATGTTTTTCCAATCTT
CTACACCCTTTTACACGGATTTGTTAATGTCTAAACCCAATGGTTTTGGTAATTAAAGAACTAGATCAGTCAATTTATTTTAACTAATAATT
GCTCCATCAGAGCTAAAACGAGTCCAAATATCAAGCATTTTGTTTTAGTACTCATCAAATCTATTGAGTGGATTGTCCTCATGAAAATAAT
TTATGTCATATTGATCTAATACAATAAGGATAAATAGATTAAGCATAACAAAATTTTAAATCAAAATGAAGACAACCCAAAATAAAGCATT
AAAATTTGAACCATTATTCATCGTATCATACATAATTCTATTCCAATCCAACCCTGTTTAAGAAAGGAACAACAAAATGAAAGAATGATGA
AAATCTGAGAAGTTGTTAGTACACTACAAATAGTTCACATTCTTAATTTAAATAATTATAACAGTATCAGATATAATTTTCCCCTTCTTAATA
TCTGAAAACAAAAATACTCCCAAAACTGAAAAAATTGAATGGATCAAGAAGTATAAAGCAGCCGCATTTATGATTGCAGGAATTGAAAA
AACTGCAGTATCTATTAGTTTTTTGAGCCCCTTAAATTATTTGACTAATCTTAATAATCTTATAGCTCTTAAAATTCTAAACCCTTTTCCTAT
AGATAATCCACTCTCATTTGTTACATAATCAATAATGGATATAACTATAACTTTAGCAAACAATATATATATTTTGTCTAGTAGTTTCTATTAA
TTTAATTAGAAAATAAATAGCTAATTAATTAAAATATAATTTATGGTGAAAAAGCAGAAAAACCTGTTTA

LEFT_PRIMER: ACAGTTGAAATCACTAAAAATCCTTTAAGA

LEFT_PRIMER_START_LENGTH: (67, 30)

RIGHT_PRIMER: CAGGTTTTTCTGCTTTTTCACCATAAATTA

RIGHT_PRIMER_START_LENGTH: (1447, 30)

PCR Product size = 1381

XbaI specific 3-IES candidates

**ID 913**

Left IES ID=IESPGM.PTET51.1.106.284913
sequence=TATTCTAAAATTAACAAAAATACATTATTTTACTTTTTATAAATATTTGCTTTGAATTGTGAAATAATTCAATATTTGAAATAATAT
AAGAAAGTAGGAAAAATTTAAAAAAGAGTATGTTATAGTGATTATTAAAATACTTGATTCATAAGTTTAAAAGCTTTGGCTAATAGTCATA
GAAAAGGTTACTTAATAAAATTAATTTCATTTCAAAATCATATATCTAACTATAAGTTCTAGATGTATAATATTGAGATTTATATCTTTTTTCT
CAAATTCAGCTTGCCTATTTTTGTTACCTATTATTTTACAAAACTAATACCGATTAAAATTATTGAATAATCCATCATTTTATTTCTTAACTTC
TCAA

Middle IES ID=IESPGM.PTET51.1.106.286750
sequence=TATTGTGGTGTTATTTTAAAAAGTAATATATTTTTTAAAATATTTGGTTTTCTCATAAAAAAAACTTAAACCTACCA

Right IES ID=IESPGM.PTET51.1.106.286924
sequence=TACAGCTAACGTAAGAAATTTTATTGTAAATATATTTTCAGCTAGATTTATTAGAAAAGAATATGAATTATCTGGAACCCTTTAAT
GAGTTAACTAATGAATTCTCATTTAGATTACATATTACGGTGCTTTTAAGTTATTAAATTCGAGACTTCCATTATCAAGACATCTTCTAATAA
TTTTTTAAATAGTCTTTAGAATTCTTTGAGACTAACAACCATTTTTTTTTAATTATAATCTCCTTTCTTATCATTTGAGAATAACACTGACAT

TAAAGATTCTAGAAAGCAAGGAAAATATTAATACCCTGAATTTTATCCTAATTCAAATCTTTGAATTATTGGATGAAAAACACGAGATTAC
TTTGAATTGAAAATTTGATTTTTTGTTTAAAATTAAACTAGTTTTATTTAAATGGAATGTTTACACCACTATGAAAGAAATTAGCGTTAACA
GTAATGATATTTTCAGTTAACAACTCTTGAAAAAAATAGTCTAATCCAAATACGCTATCAGATTTTTCACCTTCAGAGATGATCATTTTTAAA
TTAAATATCATAATCATTATCTTTGTTTTTAAAAACATGATGGAATTAATTAAATGAAACTCTAAGGATCTGTTGATCAACTAGAATTAATA
GAATTACAAAAATAATAAAATTTAGCAAAAAAAGTAAGAAAACATAATATCTTTTTAAATTCCAAAAAGTTTTGAATATCTTTTGAGTTTA
CAAATCTACTTTTATAAAATGATCTTGCATTTAATATTTCTAAATAATCAATCAAAAAGGAAATCTACTATTTGAATGTTAAAAAATTCATTC
CATAAAACCAAACTTAATTTGTGAATTAATTCAATTCAAAAAATTTCATTTCCTAGACTG

Region sequence

TGTATAATATTGAGATTTATATCTTTTTTCTCAAATTCAGCTTGCCTATTTTTGTTACCTATTATTTTACAAAACTAATACCGATTAAAATTAT
TGAATAATCCATCATTTTATTTCTTAACTTCTCAATACCACGATCAGCTAATTGTACAGACTATATGACACAAGCTTAATGCCATAAAACAC
TCACAAATTTAACAGCAAATGATGATTGTAAATGGATTGTTGATAAATGTTATGCTACATCTAGCTTTGCTTCTGGAGCTTGCACATCTTTC
AAAGGAACTTAAACAATGTGCTAAGGATATAGAGTTGGATGCACAAATACTAATTCAGCAACTTCTTCTACTGCTTGTACTTTGGATTGT
ACATTAAAAACTGGAACGGGCTTAGCATTTTCAGATTGCTAAGCTGTAGATACTACATGCTCAGTAAATAGTACTGGAACTGGATGCATA
GCTATATAATCTGCCTGCACAGGATATGGATAAACAGCTGCTAATTGCTTTAGATCTACAGCTGGATTATGTGCTATGAATGCTGCCTCTCC
AGCAACCTGCTAAGCAGTAACTTAGGCATCTGAATGTGTATTAGTGACTGGAAAAACTGGTTTAGATCATGCAAAATGTTAAGCTTATCA
TACCTCATGCACTTCTTTAAATGATGGAACTGGATGTTAAGAGTTTAAGGCAACTTGTCCTGCTTTCACTGGAGACGCTGCACATTGCAC
TGCCTCATAATAAGGTAAATGCTACTTGAGTGGTTCCGATTGCGTACGTTTCTCTACTTGTGCTGCAATTTCTGGAATTGGATTAACTGAT
GCTAAATGCGCAGGATATAATGCAGATTGCACTGTTAATGCTGCTGGCACTGCATGCTAAGAGTAAAAGGCTACATGTGCTGTGTACTTA
ACTCAAGATTCTTGCACTACATCAAAAGATACAGCAACAGCCGATAAATGTGCTTGGAGTGGAACAGCTTGCGTTGCAGTAACAACTGT
TGCAACTTAGTGTGCATTTGTTACAGGGTCTGGTTTGGATGATACATAATGTGCAACTTATAATGCTGGTTGTGTTGCTAATGCTACAGGA
ACAGCTTGTTAAGAGAAAAAGGCAGCATGTACAGACTATACAACTTCTACAGCTTGTGCAACATCTACAGCTGCCAAATGCTATTGGAA
TAGTACTACTTCACCAGCTGCTTGCATTTCTATTACTACAGTGGCAACTGATTGTTAATTAGTTTTGGGATCTGGTTTAGATGATGCTAAAT
GTGGAGCTTACCTAACAGGCTGTGTCGCACTTTCAACCGGAGCGGGTTGTTAAGAGAAAAAGGCCACTTGTGCAGCTTACACAACCAC
TACCGCATGTGGAACATCAACTGCTGCCAAATGCTATTGGAATAGTGCTGCATCACCAGCTGCTTGCATTTCAATTACTACAGCTTCGAG
CGATTGTTAATTAGTTTTGGGATCTGGTTTAGATGATACTAAATGTGGAGCTTACCTTACAGATTGTGTTGCACTTTCTACCGGAGCAGGT
TGTTAAGAGAAAAAGGCCACTTGTGCAGGTTATACAACCTCTACCGCATGTGGAACATCAACAGCTGCCAAATGCTATTGGAATAGTAC
TGCTTGTATTCAAATTTCTACAGTTGGAACTGATTGTTAATAAGTAACTGGAACAGGTTTAGATGATGCTAAATGCATAGCTTATAACGCT
GGTTGTGTCGCAAATTCAACAGGAACAGCTTGTTAAGAGAAAAAGGCAGCATGTACAGACTATACAACTTCTACAGCTTGTGGAACATC
AACTGCTGCCAAATGCTATTGGAATAGTGCTGCATCACCAGCTGCTTGCATTCAAATATCCACAGTTGCAACTGATTGTTAATTAGTATTG
GGATCAGGTTTGAATGATAGTAAATGTTCAGCTTACAACGCAGCTTGCACATCTCTAGTTGATGGTATTGTGGTGTTATTTTAAAAAGTAA
TATATTTTTTAAAATATTTGGTTTTCTCATAAAAAAAACTTAAACCTACCATACTGCTTGTTAAGAGGAAAAAGCAAACTGTAAAGATTAC
ACCACACAAAATAAATGTATTTCAACTTCTAGTGTAACTTGTATTTGGTTTGAAAATGCTTGTTATTCTATCACTGCAGTAAGTTGCAGTT
CAATAACTGGAACTTCTCTTGATCATACAAAATGCCAAGCCTATACAGCTAACGTAAGAAATTTTATTGTAAATATATTTTCAGCTAGATTT
ATTAGAAAAGAATATGAATTATCTGGAACCCTTTAATGAGTTAACTAATGAATTCTCATTTAGATTACATATTACGGTGCTTTTAAGTTATTA
AATTCGAGACTTCCATTATCAAGACATCTTCTAATAATTTTTTAAATAGTCTTTAGAATTCTTTGAGACTAACAACCATTTTTTTTTAATTAT
AATCTCCTTTCTTATCATTTGAGAATAACACTGACATTAAAGAT

LEFT_PRIMER: TCAGCTTGCCTATTTTTGTTACCTATTATT

LEFT_PRIMER_START_LENGTH: (36, 30)

RIGHT_PRIMER: CATTAGTTAACTCATTAAAGGGTTCCAGAT

RIGHT_PRIMER_START_LENGTH: (2315, 30)

PCR Product size = 2280

DraI specific inside IES

**ID 798**

IES ID=IESPGM.PTET51.1.23.362798
sequence=TATTGTCAAATTAAAAATACAACATTATTGTGACAATGAATACTGTTGAATAATTGGGTGCTATAAAAGCTTTAATAGAATTGTA
TCATCAATTGAAATACATTCCTACTATTCTTGTCAATATCAGTAATTTTATAGTTTAGTTGATAAACTATTATAATCAGGTATTTTAGCATAAA
TAATCGTAACTATGTCATCAGCGAACTGAAAACTTTAACATCAATAATTAAGTGAGCTTTTTAGTTCATTAAGAAACTCTTCAACTTAAAT
ATTATATAGGGTTGGAAAAAGTGGTGATCCTTGAGGAACACCATTAAAAAAGTAGTATTTTAATTTTCATAAATTAAGAGGCAGATTTAG
AGTAAGTTGGAATGCAATGGAAGTCTATTTCAGGTGGATCTAAGATTACTTTTTATTCTAATTATATTATAAAATATTTCTATTAACAGTATT
AACAGTATTAATAAAATCAGTTTTTATATATTTAGTAGTGTAATTATTTATTAGATTTAATAAATCGGTTATGTTTATTTAGTTGATGAGGAGG
GAATAAAACCAATTTAATCTTTAGAAAGTCGCTCAATCAAATAGCTTAGGCAGGAATCTGAATTCTAGTAATTTGTACAAAGCGCTAAGA
ATAGTAATAGGTCTCAAATAAAGTTTTTTAGGGATATTAAGGAAAGCTTTATTTAATGGTATTAATCTCGCTTTTCAACATAATTAGTTTGTT
TCAATAGTGGAGTTTTTCCAAATATTAAGAAATAAAGCAGAGTTCTTTGTTTGTTTAATGAAAGTATCAGAAGTTCCATCATATGAAATTG
CTTTGTTAGAAGACATATTCTATTACAGTTTTTAAAATTCTATATTGGTGATAAAGAGTCGGAAGTTTTACGGATTAAACGATGATCATTGA
AAGATGCCATCTAGTTGGTTTAAACAATTGTCAATTGCCTTCTTACCAATTTGTATATTTTTTTCATCATCTAAATAATAAGATAATATTCCT
CCATCCCTTTTATTCGGCTCATGGTAATAACTTTTGTTAAGAAAATTATTGAATTTTATTTATTATTCTGAAATACTATTAAGAGCTTCTCCT
ATAACGATTGTTTATTATAATCATTGTTTTCTTGAATTTTGGCACTATTCTTCTTTTAATCATCTGGATGATCTTTTGCTTTCTGATGTTATTC
TAAATAATTTTATAATAGCATGTAAAATCAGTAAATTTAGATGATGCAGTTTTTTAAGATTTAGAGCATTCTCAATATCAGTTTAAAATTAA
GTTTCTGGATAATATATGATATGGTTAGTTTTACTTTCAATTTTTTTCAAAATTTATATCTACTAAAAGTATTTTGTGTTTAGGAATAGCTTAA
GGAAAAAAGCAAAAGGAGGACTAATAGAGTTAAAATCTCCGCAAATTATTATTCTTTGATGAGTTGTTTTAGTTTAATTCAAATTTATTAG
AATGTTTGAAACTAATAGGTGAAACATTCCAAGAGATTATCGAAATGGATATTATGATTTTGAATTATGACACTTTTGGTTTATACTGCTGA
GACTTTAAGATAGTTTCATATCCTTAAACTCATTAAATAAGTTATCCATTTAAGGTATGTGATCTGTCTTTAGGATAATTTAATAGTATTTTAT
TTTTTTCTAACATTTTATTAATTCTTTAGGTTAAGGAGCGAGGTCCATACTTCATACTTGAAGTATTTTTTCTTTTTTGCATGTATTTTATTAT
TGATTAGTTGAGTTTTTCAGCTAAGCACTAAACTATCTGATAAAAATAACCTTATTAAACAATTTCAGTGCATTTTAGTCATTTCTAAGAGT
TGCCTCTTGAAGGATTGTTATTTATACAGTCCTATACTCAGAATGATCTTGAAGAATTTGAAGGATATTAGTTTGTTCTGATGGATGTTTTA
TTGGTTCTGTTAGGTTATTTTGGATAGGTTATTTAAATTGAATTAAATCATCATTTATAAGTTACCCTTGCCGGTTAACTTTATCTTCAAGAA
GTGGTAACTTTGAAAGGTTTTCATTTATTGTAGAACTTACACTATTAACAAATTGTATAATCCAAAATGTTAAATTCACTATAATATTTTAA
GTATGCTTATCTGTGATGTAGTTGGTTTAAGAATTAAGCAGGGAAAGCTTTAGATACTTATAAATTCTTGAAAAAATTCTTTATAGTACAAT
CTTAATAAACAAATGAAAATATGGAATATTCTTCTTCCAACATAATCATTCATAGCTTAACGTTATAATGAATTATCCTTAGGTATAAGAATA
CAAATACCTAGCGATATTAATCAGTTTCAAATGAAAATTAAATAGCACATTCTTAATCTAAAAATTAATCATTATTGAGCTAAAATGGATAC
CTCAACATCTTTCTGGAAATGATAGCTTCTTCATTGGTAAAATTACATTATTTCCCATATATTATAATCTTTAAGTGCAATTCCTTTCTAAAG
CTATTGAATATCTTCAAAGAACTAGAAAAGATCTATAAGGCCATTTCAAAAGGTAGTAGGTTTTTCCAAAATGTCATCGAAACAAATTAA
TTACCTATTAATTCCCTAACTTTTGACTTTAATTAAAAAAGCCTACTCGAAAACTATTAATAAGAAGTAATTTATACAAATAACTTAAATTC

CATCATGAAATATGAATTTCCAGAAAATTAAATGCAAACCAAAAGTATAAGCTAAATCAATTTTGATTACCTTCCTAATAATTTTATAACTT
TATTCAAATTATATCCTCAATATCTGTAAAGAGCACCAGTGTAATATCCTTCCAACATATTATGAATTATCATTCAAAATATCATGCACCGAA
ATAAAATTATCTTAGAAATAAAATTAAGAAAGAGCTTAGAAATTTTATTTCCATAATTGAACAATTATGATTTGATAGTATTCGTGTGTTAA
TAAATTTAATTTTTAAACCATAAAATAATTTAGAAACCGATATCTAATATTAATAAAATTCTAAATTTTTTTATGTAAAATAGAGAAAGCTCA
ATTTAATTGAAACCTAGCATAAATGATCTAAGAAATCTTTCTTAATCATTCAAGATCAATTAATCTTCTGATGAAATTGATTCTAAGGATGG
CTACTATTTTAATTTGAAATAATTTTATTTAAGGTATTAGGACTTAAAAAAAAGATAACTTACAATTTCTATAAGATTCTTTAAAAGACTTA
AAAAATCAAATTATTGATTTATAAAATTCAAAATTAATTCCATTTGTATAATCTGGACAAATTTAAAATATGGCTAATTTATTTCAAAAAAC
AAAATATAATCTTAAAAACACAAAAATATTCAGAAACCTTGACAA

IES length = 3362

length of target RE framgent 697

left_RE_position: 373696

right_RE_position: 374393

Target sequence

TTTAAAATTAAGTTTCTGGATAATATATGATATGGTTAGTTTTACTTTCAATTTTTTTCAAAATTTATATCTACTAAAAGTATTTTGTGTTTAG
GAATAGCTTAAGGAAAAAAGCAAAAGGAGGACTAATAGAGTTAAAATCTCCGCAAATTATTATTCTTTGATGAGTTGTTTTAGTTTAATT
CAAATTTATTAGAATGTTTGAAACTAATAGGTGAAACATTCCAAGAGATTATCGAAATGGATATTATGATTTTGAATTATGACACTTTTGGT
TTATACTGCTGAGACTTTAAGATAGTTTCATATCCTTAAACTCATTAAATAAGTTATCCATTTAAGGTATGTGATCTGTCTTTAGGATAATTT
AATAGTATTTTATTTTTTTCTAACATTTTATTAATTCTTTAGGTTAAGGAGCGAGGTCCATACTTCATACTTGAAGTATTTTTTCTTTTTTGCA
TGTATTTTATTATTGATTAGTTGAGTTTTTCAGCTAAGCACTAAACTATCTGATAAAAAATAACCTTATTAAACAATTTCAGTGCATTTTAGTC
ATTTCTAAGAGTTGCCTCTTGAAGGATTGTTATTTATACAGTCCTATACTCAGAATGATCTTGAAGAATTTGAAGGATATTAGTTTGTTCTG
ATGGATGTTTTATTGGTTCTGTTAGGTTATTTTGGATAGGTTATTTAAA

**ID 437**

IES ID=IESPGM.PTET51.1.155.26437
sequence=TATTATTTATGAAATAAAAATCTAGGTTTTTTATATGAAACTTGAAAAACAAAATTTTTGTTCAACATGATAGATTATTTTAAAA
GTTATTTATCATTTTTCAAAACTCTTTCTTTAGGAATTATTTATTTAAAAATGGAACAATGGATGCAATACATTTTTTCAATTAACATTCCAG
AAATTAATTAAAAATTAATGAACATACTTATTAATATTATTTTAACTCATGATTTTCGATTGCTAATAAAAACTGTAAGGTTATTCCACATAA
TTTATATAGGATTGAGTTATGGAGAATAAAGAAACAAAAACCTTATAGTAACTACTTTTAACAGTTTTTGTACCGATTTAAAATCTTTGAA
ATCATTTGAGTTTCCACGTATCTACGATTTTTAATTATGTCCAAGAGCAAATTCATCCAAAAACAAAATACAAAGTTGAATTATTCAACGG
TATAGACTAGGTAAACTATGGAGTATTAATATTTCTTTTCTAGTAACTACTTTGTGTTTCCTCTCTTATCTTTTTATAAGGCAATTTGGAACT
AGCATTACAATATTTTCTAATCAAAATTTGAAACTATCTTTGTCATATTGAATGTTCAGATTATTTAAAATGCTCTAAAATTCTTTGGTAAGA
TTTCGCTACAATAAAATTATGTAAGAGGGATCTATTTGATTGGTTTCTTTAACCATTTTGAATGGATTTCGTTCAAGTTAAAACAATTTAAT
TAATCTTATTTCATTTCTTCATTTAGAATTATACACTTCACTTAAGATGGTTTTGTTTTTAAGATGAAAGCTATAGAATATGCAGTGAATCCA
TTTTTCAGAGTTTGACCACTATGTTCAGATTTTGTAAAAATTCAAAACTAAACTCTTTTTTTACCAAAATAATTGGGTTAATCATAAATATT
TAGAGTCGATAAAAATATAAAATAGGCTAAGTTCGGTATTTTGAGAGGAATTAAACTAATTTATTGATTTTATTTTACTAAAAATAGAGATA
GGTTATATTAGGATGGGCTGCTCACAATTTTTTTTTATGATATAAATCCATATTCTATTTTTCAACCACTAAATGAAATATTTTTATAAGAATG
ATAATTTGAGTAATAGAAAGGTATATTAGTAATTTAAACTTCTCATTGACAACATCAAAGAATTTCCATCTAAATACAATTACAATTGAACT
AAATTTGGAATGAAATGTTAATAATAATTGCTCTAAAAATTAAAAAAAAGGTTTTGGTTTCGTAAGTATATTTGTAAGGGTAGTAATAACA
GTATTGTAAATAGGAAAATAACATTATGGAATAATTAGACCTAACATTCGCTAGAACAATCCAAGGAAAGGTAGTATAAACTCTCTTAGA
CTCCATTTGTTCCTCAGGAAAACCATATAGTAAGAAAAATCTCCTGTTTACTACCTTTACCAAATCAGGCTAAAGAAAATAAACGATTGA
AGAATAACTCATGAAATTGTTTACATTATTTAATCAGCCAAGCCTCAAGCAAGCTTTCCAAAATATAAAAAGAATGGCAACTCTTAATCCT
TCAAAAAAGGAAGCTGGTATCTTTAGATGCTTCAAGGATGATCAAGATAATATAAATACTCATCCAGATACTGAATAAAATGTAAGACAA
GAATTAGACAGAATCTCGCAGAGAAGTGTATATCAGCCTCTTACCCATACTAATTTGCCCTCTATATGGAATGAAGAGATTCAGAAGCTAT
AAAAGCTCATGTCAAAAGATAAAGCTATTACGATGGATGGTATATCAGACAATTTTATAAGGATCAAAACATCTGGACATTTTAAAGATCT
TTGGAATGATAACACTGGCTAAATAAATCAATATATTGATCAAGCTCGTGTGGTCCCACTAAATAAGGTATATCAAGAAATACCCTCAAAA
TAGCAATTCAGACCTATCACAAAACTTAGCCCATTTTTCAAATTTTTAGAATTAAGATTTTTAAACAAGACTCAAAGATTTTGAGAGCAG
CGCCTAAAATATGAACAAAATGGCTTCATGTTAATAGAGGAACATAAGTAAATATAGAGAAATTGGTTGCTTACTTAGGTAGTGTGAATA
GAAAGGAGAGGAAATGTGTAATTTTTATAGATTTTCAAATGCATATTGCATCATGATAACAAACAAGTTGTTTTAAATTATGGAAGAAATA
AGTATAATACTTGACGAATTAACATTTTCAAAAGGATTACACTCATTACTTCATTATTAAGACCCTAGGAAAGAAAAAGAATATGGTGGTT
TAAAAATGGAGTCGTTTAAGTATCCCTATTATCACCATCACTACTCAATATTTACTTGGATACCTTTTTATAAAATTTATCTTCAAAACTGAA
TCAGGAGCTGAACGTTAGAGCTTATGCAGGTGATCTAGCTTTTATTGTAGACGACAAAGCTGCGAAGAATTATAATTATCTTGGAATCAA
TTTAACAGATAAAGCTTCTACATCTAACTAGATTAACAAAATAAAAGGAAGAATTAATTATATCTCCAATTAACTCAAATGGATCTTCTGA
AAAATGGATATTAAGCAAAAAATTATCCTCTGGAAATTTACATAAGGCCACTTATCACCTATTCGAATTTAGCTATTTAAACAGCCATTAA
AAAACTCCAAAGAATGATATAATAAAACTATAAAGAAAATCTCTAAAGACGATAACTAGAATGCCAAAATCTACACCAAATCCAATCATT
GATATATTAATCAAAATTAACAGGTATAGCCAGGTTCCTAACAACTGGGAAAAATATGCAATATTCATAAATGCAAGATGAATGATTCACA
TATAATTGAATAAGATTTTGGGCTTAAATTTGAAGAATTATTGGTAATCCCCCAGGGGAGAATGTAGAAAAGCAAAGCTATCAAGATGAA
AAAACGAATAGAACAAGCGATAACAGAAATTAATCAAAATATCAGAGAATGAAATATTAATGTGTGTTTAGGTCCCAAAAAGGACATAA
TAATAATTATAATAATAAATAGGAAAATAGTAAATATAAAGATTCTAAATTTATATGAAAGGATGTTAATGTTGTTTTAAATATTGTAATAAG
ATGGTTAGATAATAACTTCGTAATTTAAAATTGAGATAAAGATTATTATGTAATCAATCAGACTTAGCCCTTTCAAATAATCTAAATCTCGG
ATAATTATTTCTTTACTAA

IES length = 3211

length of target RE framgent 689

left_RE_position: 28326

right_RE_position: 29015

Target sequence

TTTAAACTTCTCATTGACAACATCAAAGAATTTCCATCTAAATACAATTACAATTGAACTAAATTTGGAATGAAATGTTAATAATAATTGCT
CTAAAAATTAAAAAAAAGGTTTTGGTTTCGTAAGTATATTTGTAAGGGTAGTAATAACAGTATTGTAAATAGGAAAATAACATTATGGAAT
AATTAGACCTAACATTCGCTAGAACAATCCAAGGAAAGGTAGTATAAACTCTCTTAGACTCCATTTGTTCCTCAGGAAAACCATATAGTA
AGAAAAATCTCCTGTTTACTACCTTTACCAAATCAGGCTAAAGAAAATAAACGATTGAAGAATAACTCATGAAATTGTTTACATTATTTAA
TCAGCCAAGCCTCAAGCAAGCTTTCCAAAATATAAAAAGAATGGCAACTCTTAATCCTTCAAAAAAGGAAGCTGGTATCTTTAGATGCT
TCAAGGATGATCAAGATAATATAAATACTCATCCAGATACTGAATAAAATGTAAGACAAGAATTAGACAGAATCTCGCAGAGAAGTGTAT

ATCAGCCTCTTACCCATACTAATTTGCCCTCTATATGGAATGAAGAGATTCAGAAGCTATAAAAGCTCATGTCAAAAGATAAAGCTATTAC
GATGGATGGTATATCAGACAATTTTATAAGGATCAAAACATCTGGACATTTTAAA

**ID 670**

IES ID=IESPGM.PTET51.1.174.130670
sequence=TATACTCAAATACAATTTTCATCTACAACTTCAGACATCCTAATTTTTTTATTATAATATAAATATATTTCTATATCTAATCACAATC
AAAAACACACAAACACATAATTTAATTTTTATAATGTGAAACTATAAATAAAAATAAAAGGCTTATTATATTGAGAGTTGGAGGGATGAAT
TAAGGAGCGAATGATTAACTCCCATTCGATTCCAAAATTCGAAACTCTGCTTTGACTAAAATCTCCTTCAAAGTACTAAAGAATTCCAGA
CTAACAAACTGCGATTGCTACGTTATTGCTATTCAAAAGGAGCGGAAACTTTTTGCACATTTACTTGTGAAAATTTATATTAAAAACTTGA
ATTTATATAATAGCATCAGATGTATGATTGTAAAACCTAACTATGAATTTAAGTTTGAAGTAGAGATGTTTGAACTTGGGTTGATGTATAAT
GGAGTTAAGCACAACATGTATATTACAAATTCTAAGGCAGTTATTGCCACTTGCTTAAGTTAAGGGCTTAAATGGAGGAAGAAATAAATA
ATAAATAAATGATAATTGATTCAGAATTTATTATTATCATTTGATGGAGGAATTATTCAGGACTAAGTTAAAATTATTCACTTCATTATTTTG
CTAAAAGCAGCTTTAAGGTTTCATTCAGAAATTCTCCAAATACCTATTGATTTACATTTCAATTAACCCAGAATCAAATACTAAACCTATC
GCAACCGTCAGTATCAATTACAAATTTACCTTTTTAAGGTGTTCGATTCTATATACAACAATAGACACCTATCTGGTTTGCCAACTTCATCC
AGTTTTAAAGCTAAATGGATCCACATAGACTAATTTTTAATCCGAGCTGAGAAGATTAGTTAATTATTTTGCAAATTCAAATCTTCTTGTTA
AAGTACATTTTTTTTGAGGATTAAATTAACAAAATAAGTAAGTGATAAGTATACAACCCAAAACCATTAAAAACTATTGATTACTATGAAA
ATTCACAATAATTATTAATATGTTTGATTTCAAATGAGCAATATTATCTTAAACTACAAATTTTTTCAATCCATAATTTGACCTATTCATCAA
ATGAAAAAGCAAATTATTGACACATAGAATTAGAAAACTCTAAAACTTACTAATATTGATATTTAGCCCTTCAAATTGATTTATTAAATATC
CTGTTAAAATTTAGATTTTATAGAAATTCCTATTTCTTAAGCAATTTACTTTGGTCCCTTTGTTAGTCGTTAATCTTTGCATTCCTTCCCTTAT
CACTATCAAATATTATGTGTATCATCAAAAGGAATCTATTTGTGTATTTACTGATGAAATTTTTTTTTATACAATTAGGTTTTAATAAAATTAT
TGACAGACATCATAGCATGCAAGTATCACAACATAGTTTAAGTTTAAGGCCTGCTATAGCAAATAAAATTATTACAGAGAATAAACTCATT
AAATTATGAGTAAGATTAAATGAAAAAGTATATATTTTGTTAATTTTTGGAAATGTCTTGAATTTAAAAATAATCCAATAAAAATGTATGTT
ATAATTTATATTAAAAAATATTAATTAATTTTTTTGATATATTCTTTATTATGTTTATTTTTAATTTGCTATAAAAGATTTTTTTTTTCAATATAA
ATTTTAGTTTTGAATGCTAAATAAATTGAGTCAGGTTTCCATTTTACAAGCAACCAAACAAACGGTTCTGTTGAATCATTTTATTATGGAT
TAATCTTAGCGGGATATTATTTGCCCGCATTCCATATGTTTGGAGAATACGATAATATTTTCTCAGAGTTGCTAAAAAAAAATCTCTTGCTCA
GAATCAAAAATGTCCTAACAGGATAATTGTTTCAGCAACTAAAGAGGTATTTTACACAATAATTTAAGAATCTAAATGACTATATTTTGGC
ATAGCAGATGAAATCTACACTCATCAAATGCGGGCTATCAATATTCTTAAAGCGATGGAGCTGAATCATTCTTAATTCAAGAATCAAAGG
GACTAGATTAATCAAATCTAAAGCAAGATGACCTAATGCAAAGATTTTACAATCAAAATTTTCAATCGTTATTTAGATACAAAGGTTGGA
AGAAATAAAAGTGAGATATTTTTGCAAAAAATGAAGGAGCAAGTTGAGAAATTTAAATAAGATGAAACAAAAGCAAGTAGTCTCCCAA
AGATCTCACTTAAGCAAAAAACTTACTTAAAGTAGCAAAGATAACTAAGAGTTGGATCAAATAAATTAAGGAATTTGAGAAGAAGTATA
ATAAGCGAAGTCTTTTTAAAGAAGAAATTGAATAATTATTAATAATAAAAACAATAAGATCTAATCTTATGGATAATTTTAATTGCTAAATT
TCTGCAACCTTTTTTCTGTATTCTCTACTCAAAGAAGTTTTTCATCTTCGTTACAATCTTTATAATAGTGCTTCAAAATATCGATATTCATCC
TGTTCGCTTATGAATTCTTTAATGTATTTTTAGCATTTCCATTTTTCAGCAATTCTCTTTCTTCAAAATAAAATTCTTGCTAGTTATTTTCTTT
ATCATTTTTCATCTTAGTTTTAGGTGCTTGGTTTATGATTTGAAATTTTTTTGGTTTTATTTTATTAGAGGTTTGTTAAAACACACTAACTTT
TTTCATAAACTTTTTGTATTCCTTAGTTGATTTTTGCTAGTCAATCTATAACTCTGAGTCCAAATTTTTCATCAATTGTTCACTAAAATTAAA
ACGATGGCATTGATGTTCAATTCCCCTAACAATACTTTCTAAGTACATTGTCAAGTTATGGGGAAGTTGGGATATATTTTCAGATAATTTGA
AAATAATAATGAAAATAAAATAATACACCTACATGAAATAATTTGCTATAAAATGTGCGATTTTCCTTTCTCATCTAATGCAAAATAGATTA
TTACCAGAAATATATAAATAATGTGCATAAAAATTACTAAAAATGTTAGCTAAAATATGTTTTAACTA

IES length = 3001

length of target RE framgent 712

left_RE_position: 138050

right_RE_position: 138762

Target sequence

TTTAAAGCTAAATGGATCCACATAGACTAATTTTTAATCCGAGCTGAGAAGATTAGTTAATTATTTTGCAAATTCAAATCTTCTTGTTAAAG
TACATTTTTTTTGAGGATTAAATTAACAAAATAAGTAAGTGATAAGTATACAACCCAAAACCATTAAAAACTATTGATTACTATGAAAATT
CACAATAATTATTAATATGTTTGATTTCAAATGAGCAATATTATCTTAAACTACAAATTTTTTCAATCCATAATTTGACCTATTCATCAAATG
AAAAAGCAAATTATTGACACATAGAATTAGAAAACTCTAAAACTTACTAATATTGATATTTAGCCCTTCAAATTGATTTATTAAATATCCTG
TTAAAATTTAGATTTTATAGAAATTCCTATTTCTTAAGCAATTTACTTTGGTCCCTTTGTTAGTCGTTAATCTTTGCATTCCTTCCCTTATCA
CTATCAAATATTATGTGTATCATCAAAAGGAATCTATTTGTGTATTTACTGATGAAATTTTTTTTTATACAATTAGGTTTTAATAAAATTATTG
ACAGACATCATAGCATGCAAGTATCACAACATAGTTTAAGTTTAAGGCCTGCTATAGCAAATAAAATTATTACAGAGAATAAACTCATTAA
ATTATGAGTAAGATTAAATGAAAAAGTATATATTTTGTTAATTTTTGGAAATGTCTTGAATTTAAA

**ID 181**

IES ID=IESPGM.PTET51.1.196.12181
sequence=TATTATCAATTCAAAATTCATCATTTTATTTTTGAAATGCAGAAGGTTCAAATAATTTTCCAACTGATTTAGACCAAATTTATTTA
CCCTAATATTTGAGTTCCTGAGATAGCTTTCAAGAATAAACTACATTCATTTTATTCTTTAAAGAAGACAAATTTCCCTCTTTCCTTCATGG
TAACTTTTTTAAAAATCCCTTAAAAACCCTCAGCCATTTGTTTAACAATTATTTCTAATTAATCATCTCTTTAAATTGTTTATGAATTTAAGG
GTTTGAACTTAGGAAAATCGTAAGGAATTTAACTCAAGTAGTGATCCAAATAATAATTTTAGAGCAATTCATGGAAATTCTACAAATGTTT
TTTATCAGATACAAATAAAGAGGATTAGTTGTAGATTTATTAAAATTTCAAAAGGGACAACCTCATTATCTGAGAAGAGTGGTATGATAGG
TGAATAATTACAGATTTGCAACTAATCAACTTAACAACAACTAAATGAATCAGTGATGATAGAATAAATAGTTTAGCCAGTTTAACAGATA
GGAAATGGGCTATGTTAGAAATATTGATCAAATACAATAATATTCCTAACAAAATCTTAGCATTAATTGCAAAGTTGTGTATGCTTAGGTTG
AATCATAAGGCCATGCAATACCTGATGTAACAAAGAGGAAACGCTGAAAAATACAAATATAAATGATGTATTCAAGCACAAGCAGCAGA
TGAAAATAAAGAAAAGAACAATTGTTGAATAGATCATATTACTTGAGGTATCTATCAACAAGTACATGTTGTACTCTAAATCTAGTATTAA
TTTCGTCTGCTCAAGGTTAACAGGTTTAAGATCCAGAAAAGTAAACGATGCCCATATATGGACGTGCATTAATTTTAAAATAGGGATGCA
AAACAGCATATTTGTAAGTTACATAATTTTCTTGTTTTGCTAGATACTTGTAGATTAAGAACAAAAAAAAACATGTAAAACATTGATCTTGA
TATCTAAAGATAATATACAAAAGGACTGTGCTAAGAGGCTTCTAATTTCCATTCTTAAATTTATACCCAGATTGTGGACAAGGATTTCAAA
TAGTTTGCCTTTTGTTGATACATATATTAACTTAATTCTTATATAATTAAATGTATGGAAGTATAGAAGATTATAAAGTTGAACAACCAGCAG
ATAAATTAGAAATAAATGGCTGATGCATTCAAATAGTTCTAAATTAACCATGATATTACACTTATTTAGGTGGTTAAGTTACTTAGAACACA
ACCAACATCACTAAAGCTTATATTACACTATTATGGAACATTGATAAATTCTAAGCATATGAAATTTTAAGCATAAGGGGCGATATATAGGA
CCTTATATGCAACAAAAATAAAAGCCCCATGAGAATATATATAGAGAAATAAGATGAACAAAGACTCCAAGCAGAAGTAATGTTCATTAT
CTATAAAAATGAACAAAATTCTTCATCATCACTAACTAACAGGATTAATTGCAAGTTAATAAAAGTTGCATATTGTTTTGGTAATATTTTTT
TTTAGCAAACCGCACATACTATTATGATCACCTTGAATGTATAGTGATGAATAATACGAAAAGTATTACACTAATACTATCAATCCTATGTAT
TAAAGTGAAGCATTAGACAGCAGATTAGGAAGAAGTTATGGAGGTTTTGACATGGGTAGTTTTTGAACAACAACTTTAAAAAGTTGAA
GGCTTGCATATCTAATTTTTAGGTTCATTTTGTTGCATGACATGTTTTAGAGTTGGCATAGCAACATTAAATTTCTCTCTGATTAAAACAGA
TTATGATTTGTAACTACTTCTTGAATCAACTTCAACTGCACACGAAGTGTTGGAACTGTTAAATTAATCATAAGAGTGGTTAACACTGTTA

ATTTAGTTTTACAAGTTTCAGGTATTTAAAAACAATACTAGAGCAATTAGTGCAAGTAAACTATCTAACTATACTTGTAACTGTGTGGTTC
TCTAAGATAGTTCCACAATTATAATAGAGGTAACAATTCAAGTAAGTATTGTAGAGACAAAATAAATTATATTATCGCAGTCCTCCTAACA
GGACATGCTTGACACTTTCCAAATTACATAAGTAGTAAAATATAATTCGTAACTGAAATATGAGTTTCTTTAGTCTTCTATAGATTAAGAAG
GCTTACATAGTCTGAACGTTTATGCGGATGTTGAGAAGTAGCCACTGTTTATTCAGTAGTATAAGGGTTATAAGAATATCAAGGAACGTAA
ATAATTAGATTGATTATCTAAGAACAATTCAATAAGGGTTGATGAGATACCAGAGGAAGACATTAGTCTGAACATCCATAACGTTTTAAGA
ATTAATGCAAGTAGTGCTGTGGAGTAAATATCGATATCGAACGAAATGATTGTCATTAAGTAGAAAGTATTAGCATCATATTAATAGTAAA
GCACTCAATGCACTATAATATTATAAGAGTAGTTTTAAAGGTATATCGTATTACTATATAAATAGCATGACTCCAATATTGCCTAATCTGTTA
ACTATCATAAAATGTAGATACATTCATTTACCTAATAATGAGGCTACTGTATTTTAAGTTTGTCTATTCTCAATTCACTTGTAGATATTTAAG
TTACCGTTAATTACAATTAACTCCATTAAGAGACAAATAATCCTTGCAACATAGCAACTATTAGAAAATTATGTGGATATACTGTTGGAGC
TTTACATTTAATTCAAATTAGAAATTATATTACTTTAGTTTTAGGGTTTCAAGGAGTTGAATGTTGTCAACAAACAAACAAAGTGCATGTC
ATAGATTAGAAAGTATATTTTTCAATTCATGTTAAAACAAGATGGACTAAACCAAATTAAGACATACAGAACCTTCATGCATCAAACCTTC
AAACACTATGAGTTGTACAGCAAATATAGAAACAAGAAGAACATGTTTCACCAGCCCTCTAGGTTGTCCTAGTAAAGTAAGAATTGGAT
AATAAGTAGACTATTAACCAGCCTGTAATGCAACATTTGAAGTGCACTTCTCAAATGTAAAAGTAAGGCTGGCATCATAGAATCCTATTG
ATTGAGTGTTCCAATTTTAAAATAATTTTTATGCATCTATTTGCAGGTTCGTTACTCCTATATTGTTTACAACTTTAGGCTGTGTAAATGGCA
GTTCAGTTTTACTGTTAAAACAAAAGGATGTGAAATATCCTACACCCGAGGGAAAATCATTGTGCATATAATATAAAGCCTCATTGAGTCT
GAATAGTATTGAAAGGAGAAAATTACAGGAGGCAGTTTCAATAGGTTTGCTAAAACTGATTTTCATAACTCACTGAGAATGTCCAATATC
AGAATGAAATTTAAGATTTTTTTGAAAAAAAAGTTACCAAAGCCTCTGAATAGCTACAAAATTAATATTTTAAAAGGTTAGTTGATCATTTG
AACAAGTTTGGAAAAAATGTAATATACAA

IES length = 3489

length of target RE framgent 831

left_RE_position: 13230

right_RE_position: 14061

Target sequence

TTTAAAATAGGGATGCAAAACAGCATATTTGTAAGTTACATAATTTTCTTGTTTTGCTAGATACTTGTAGATTAAGAACAAAAAAAACATG
TAAAACATTGATCTTGATATCTAAAGATAATATACAAAAGGACTGTGCTAAGAGGCTTCTAATTTCCATTCTTAAATTTATACCCAGATTGT
GGACAAGGATTTCAAATAGTTTGCCTTTTGTTGATACATATATTAACTTAATTCTTATATAATTAAATGTATGGAAGTATAGAAGATTATAAA
GTTGAACAACCAGCAGATAAATTAGAAATAAATGGCTGATGCATTCAAATAGTTCTAAATTAACCATGATATTACACTTATTTAGGTGGTT
AAGTTACTTAGAACACAACCAACATCACTAAAGCTTATATTACACTATTATGGAACATTGATAAATTCTAAGCATATGAAATTTTAAGCATA
AGGGGCGATATATAGGACCTTATATGCAACAAAAATAAAAGCCCCATGAGAATATATATAGAGAAATAAGATGAACAAAGACTCCAAGCA
GAAGTAATGTTCATTATCTATAAAAATGAACAAAATTCTTCATCATCACTAACTAACAGGATTAATTGCAAGTTAATAAAAGTTGCATATT
GTTTTGGTAATATTTTTTTTTAGCAAACCGCACATACTATTATGATCACCTTGAATGTATAGTGATGAATAATACGAAAAGTATTACACTAAT
ACTATCAATCCTATGTATTAAAGTGAAGCATTAGACAGCAGATTAGGAAGAAGTTATGGAGGTTTTGACATGGGTAGTTTTTGAACAACA
ACTTTAAA

**ID 191**

IES ID=IESPGM.PTET51.1.579.2191
sequence=TATTATTTAAGACCATAAATTCAAAAGTAATGATAAAATATTAGACTATGGAATCTTATATTTCATTAATATATATCACAATTATTAG
GTAATTCGTAGAACATTGATATTAGAGGTATCTAAATATTAAAAAATACTAAGGAAGTTCATCAATAGGAAGGATGAAAATTTATTTGTAG
AGATTATTTGAAGAATAATGTTGTCCTATTTAACACCCCATAAACAGATAAGTGTATCACCGCTACCAGATACTATTGTAGTATTATAAAGG
CTGATTAAACTAAATTTTATGATCACCAAATTTGCCTTTTTGTTGAGACGTTTTTACTTCCTTACCATAAGAGAAGTGTATGACAATCAGA
ATCCTCAAAAAAAAAGAAGAGATTACTCGAACAAATACATTCTTTCTTTTAAAAATTCAAAACTTACAAACATTTTCCTTCTTTTATAATT
CAAAACTTTAATCATCCTTGATTATCCTTTATTAGTGATGATAATGAAATTTGATCAATTCTTTTCAACTTCTTCTTACACACTTTCAACAA
GTGAAAATAGCAACTCTACTACAATTTCTTCAAATACAAAAATAGGATGTGTTTTTGTTGCATTATTTTTATATCTTTTTATCAAGCATAATA
TTCTGACAGCTTATTTAATTACAAAACCTTCATTAAACATATACAACAACAACTTAATTTATAATCTTTCTTTTTCCAAACTTCTTGATTTTA
ATCAATCGAAAAAACCACAGAATTTTGTAGATCACAGCATATCTAGTTAAATAGACATGCTTTCACGCTTATATTAAGAATACTTTTCTTAC
TGTGGTAAAATGCTGAAGAATATATGAATTAATTTTGATAAATTCTTTTGAATTTCTGAATCAAAAAAATAGCTTTTAGTTCGATTTTTTTC
GTAACTATAAAATTATATAAGCTTCAAGCTTATGGATCATGCCTAAGCTTTAACAATTAATTACTTAAACTTCTACAACTACTGCTAATAAA
TGTTTATGAGGTAAATTATCTAAATATAGCCAATGAATAAAAACTATTTTGAGCATTTAGACTTTGATTAAAGACACTTTAAAAAATGAAA
GGCTTTATTGATATACTGCCAAACAAATTTAATAAACTAACTCAGATGACGCGCAAAATTTGTAATTAACATAATTTTATTTTGTCATTTAA
ACATATTATGGTATGCCATTCAAAATTCTTGGTCTAAATTATAAGTACGTTTTAAGGATGACTATTACAATAAAAAATAAATTAGAAGAAAA
TAGAAGAGTTAATAAATAATAAAAAGAGTTTATGAATTTGATATTTGATTAATATGAGTTAATTTTAATATTATATGGATTCATTACCTCCCC
ATTTTGGATTTCTTAGGAATGAGATTAAATTTGTTTAAGTATTAGCCTTCAAATCTTTTAACAAAGATTCAATCTTATTCTCTTGAAACAAG
ATTCTATTGATTAATTATACTTAAATTTCGAAATATACTTATAAACCTTAGCAATTATCTTTTTATATTTGATTAGGTTGAGACTCATAAAGTT
AGATATATATAGAACCCTAAAGAAGAAACATGAAGTGTACCTACCAGAAATATTCAATACCTCAAAATTTAAGAGCCTGAAATTATAAGT
GTGTGAACATCTATATATACATCAATACTTAAGTTAGAGACTTTTATTAAATTTATTGAAACTACAAAAAGATATATTAAAGTACAAATTTTA
TGAATATTATTAAATGCTAAGGATTTGAATATTCTATTTAAACTGATGCTCTAATAGTTTTTAATGATTAGATAAAGTAAAATTGTTAGTTAC
TTTGCCTTGGATATAAATGTTTATTCAGTTAAATAAAGTGGCTTAAAAAATAAAATATTCTTAATTCAACTCAAATTAAAATTAAATCATGA
AAAAATCAAAACAATTTACTTGTTAATAAGTTGTTGTTGTTGTGCAAATCTTACATAAAGCAGCTTCTAATCTTCTTATTAAAATATGAGC
GTTAAGCTCTTCTACTAAATCGATTGATTGATAAACATTCTTCTGATTATTGCATAATCAAAAAAAGTAAACTAATTTAAGCATCTTTGTAT
TTTAGTTAAATATATGTTATATATATTTGATGTAGATGCTAAGTAGTACTTACAAAGGCATAATACATATTCAAATGATTTAACAATATACTAA
GGAACCTGAACTGAGTCAGAAGGATGAAAGTATGTATGTATCACATTCATCACACCAGTATTAAATTTAGAATTGTTGGGATCCTTGCAG
AATACATCTTTCGCTTAATCACCAGCATTTCAATATGTCTTGAATAGTTCTTTACACTCTAAAAAACTAAGAAACAACCTTTCAATATATGA
TAGCTATTCAGCTTAGCCAATTGTGCAAGAAGTATATCATCCTATTCTAAAGTGCCATCAGCATCTAATGGTGCACCACGTTGTGCTTTCT
TTCTCGGTTGCAAGAGCCTACATTCTAGTAGCAACTAGATGTACAGGAATTATATTTAGTTAAAAAATATAAAGCATTATTACCAATTAATA
GTTCAAGTGGTTAGATATGAAACATCCACAATTACAATAAAATATGCTTTTACTTCTAAGCTGATCAATTATGTTTTACAATTGGTCAGAAT
TAGCCACCTAAGCATACACATTTACAAAGTTTAACTACATTCTATAGTTTGAGCCTAATTATTGAACGATTTTAATTGGGCAACACAATTTA
GCACATGATGAACATTAAATAAAAACGGCAGGGAATATTGGAGTAGTAAATTATTTTAAAACACCAGCTCATACTGCTACTCGAATCAAA
AAATATCTGAAATAATTCTCATTTCACTTTGCCTGATACATTTAAGATTGCTGGAATATTGGTGAAATCCATATAATCAATTATTGCAATAAC
TAAGGCTAATAAAATTCTAACCAAAAATTATTATTAAGGAAGTTATTAAAAAGTCTAAAACAAGAGACAATGAATATGATATTTCTTTCGT
TACATTAGTTACAGGTTACATAATGTGATAGTGCATTCGTGATCAGGTATGTGCTTAATCATATTCTAATTATTTACCCTATATTCATGATATT
TGATGGCTAAGCACTTGAGGAGCTTTAATCTTCATCACAAGAATTAAGCGAAGGATATATTTTATCTTAAAGTAATACTATAAACACTATA
ATTTATTATTACATGTTGAATAATGTAGACAGAATCCATGAACAAAATTTTAATATAACGGCCATCATTAACTGAATTAAATTGAACATATTT
AATACGCCATAATATTTTAAGGTAGAAGAATCATAATTAGTAACGAAAGTTACATCACTGAACTTTTATTTCATCAAAGTGGAAGAAAAAT
TGAAACTCATTTCAATCTAGACTACCTGCCTAATTGAATTGACATGATTGTACATCAGTAGAACTATTACTCCAGTTGTACAAAAGGTCCA
AACAGAATTGAAAGAATGACAAACATTAGTGTTTGCAGTAACAACACCAATCAAAAAAGCTACTGCATGTTAATTGATACGTACTGTGC

CAATTTCACATTTTTAGCTCTAAATAAAATTGGACATCAGTATAACTATTAGTAACAAAATTAATGCAGCAAATTGTATTGTGCTGTAGAAT
ATCATCTCTCTCTCAAATATTTTTGGGAATTCATTATTCTAAATGGAAAATATTTGAATGAGCATATTATCTTTTTTTACTAATTTTCTTTCCC
ATTTTTATAATCTTAACTTTGTTTTTGTGAGAGTTCCTCCTTATTTTAATTCTCCGCTTAATATTCTTCTGAAAGTATTAATTCTGTCAGATTG
ATTTGCTAATCATTTCATCGATTATTTCCTAAAATTAAATTTGTATGATTTTTCATTATTTAAACATTTGATAGGAGAAGTTCATCTAATAAAT
GAATTTAATTGATTAAATAATAAAATACAAAATCTCAATTTTGGCTATAATTTATTATTTTCCTTATTATTTTTTTCTTTCATATACATTTAAAA
ACATAAATAGTATTCTCCATTTGTAAATAGGCTTCAACTTCCTAATGCTGCTTTATTGTAGCTTCAGTTGGGGTGTTTCTTAGAAGGATAGA
AATGTATTTTTTTAGTTCCAATTAATTCAAATCTCTTGTTTTAGTTTAAATTTAAGTTTTCAAATTTAATCTTAAATGTTAAAATTTTTTTGTT
TTAATGGCAAAATTTTAATTGATTCTTTGTTTTGTTTTTGTAATGACACACATGATTATGGTGCAAAAAATGCAACATTTCATTAAATGAAA
AACCTTTCCTGATCTAAAAATGATATTTATAATTTATCATCAGTATCATCATAACTGAGAACTTACAGTGTACATTACTCAAATATGATAATC
TAAATAATTTTCGTACATCAGCAATCTGTAAAGTGTTTTGTAAATTTCCATTTTTAATTCTTTTAATTCTTCCCATAATGATATTCATCACATT
CTGAAATTTGATACAAATGTGAGAATATCTTATCTTCTTCAAAGATTTGAGAATGATGTTGTTTGAGATGTTTACAAAATGTATATATTTTT
GCAAAGTTACGATCTCTTGGATTACACTTTGAATTAAAACATTTGTTGTTCTTTCCCAACAAAAAGTACTTTACAAACCTTAAATGCTATG
ATTAAATGTGATCAATCAATAAACTCAAATTGTCAAATTTTATCTTACATAAATAACATTGCAAATATTATTATGTGATTTGTGCCAATAACA
GACCTTCATTTTCTAATTCAATCAAGTTTAATATAAACAACATAACTTTATTAATAAAGTTTTAGAATTTAATTTCTTCTTCTTCTTCTTCATA
TTAATCTTCATCATTATTATTTAAAAATTTCATTAAATTTTACATAATTTATTTTGAATCGAAAAAGTTATCAAAGTACTATTTCTATTAATCC
ATTCTCTACACTTTTTATTCTTATTCTTCTTCTTCTTCTTCTTTTTTTTTTTACGATTAATTCCCTCAACTTTTCCAAACATAATCTAATCATCTC
TTCGACAACAATTTCATTACCTTTTGAAATCCACAATTTATCAAATACTACTCCTCCACATTCCAATAATTTAAATTTTTTCAAAATGGATT
TTAAATTAAACAATTATTTATAATATTATAACTTGTCATAACTGTCTAATACATTTAATTACTTCTTCCAAACTTTCATCAAAATTATTATTATA
ATTCTCTAAGTGATCTTAATTATCCAAAATGTATTCATAATGTCTGATGATATTGA

IES length = 5314

length of target RE framgent 1113

left_RE_position: 4999

right_RE_position: 6112

Target sequence

TTTAAAACACCAGCTCATACTGCTACTCGAATCAAAAAATATCTGAAATAATTCTCATTTCACTTTGCCTGATACATTTAAGATTGCTGGA
ATATTGGTGAAATCCATATAATCAATTATTGCAATAACTAAGGCTAATAAAATTCTAACCAAAAATTATTATTAAGGAAGTTATTAAAAAGT
CTAAAACAAGAGACAATGAATATGATATTTCTTTCGTTACATTAGTTACAGGTTACATAATGTGATAGTGCATTCGTGATCAGGTATGTGCT
TAATCATATTCTAATTATTTACCCTATATTCATGATATTTGATGGCTAAGCACTTGAGGAGCTTTAATCTTCATCACAAGAATTAAGCGAAG
GATATATTTTATCTTAAAGTAATACTATAAACACTATAATTTATTATTACATGTTGAATAATGTAGACAGAATCCATGAACAAAATTTTAATAT
AACGGCCATCATTAACTGAATTAAATTGAACATATTTAATACGCCATAATATTTTAAGGTAGAAGAATCATAATTAGTAACGAAAGTTACAT
CACTGAACTTTTATTTCATCAAAGTGGAAGAAAAATTGAAACTCATTTCAATCTAGACTACCTGCCTAATTGAATTGACATGATTGTACAT
CAGTAGAACTATTACTCCAGTTGTACAAAAGGTCCAAACAGAATTGAAAGAATGACAAACATTAGTGTTTGCAGTAACAACACCAATCA
AAAAAGCTACTGCATGTTAATTGATACGTACTGTGCCAATTTCACATTTTTAGCTCTAAATAAAATTGGACATCAGTATAACTATTAGTAAC
AAAATTAATGCAGCAAATTGTATTGTGCTGTAGAATATCATCTCTCTCTCAAATATTTTTGGGAATTCATTATTCTAAATGGAAAATATTTG
AATGAGCATATTATCTTTTTTTACTAATTTTCTTTCCCATTTTTATAATCTTAACTTTGTTTTTGTGAGAGTTCCTCCTTATTTTAATTCTCCG
CTTAATATTCTTCTGAAAGTATTAATTCTGTCAGATTGATTTGCTAATCATTTCATCGATTATTTCCTAAAATTAAATTTGTATGATTTTTCAT
TATTTAAA

## 9. Acknowledgements

First of all, I would like to thank Mariusz Nowacki from the bottom of my heart for saving my research career and for taking me in this lab and giving me this constructive subject when I quitted from my last lab. So, I can happily carry out my favourite scientific work in a lab that is so harmonious. Mariusz is also very open and supportive of the direction of my research and provided very constructive guidance throughout the research process. I also thank my thesis committee for their efforts on supporting and evaluating.

I thank the members of the Nowacki lab for all the discussions and support. I would like to thank the former colleague Chundi Wang for helping me to quickly open up completely new territory when I first stepped into this lab. I would also like to thank Therese Solberg, for all her constructive comments. I would like to thank Dr. Sebastian Bechara for his efforts on bioinformatic analysis on SMC4 project. Thanks Dr. Ryuma Matsubara for his support on the PGM projects. I want to thank Dr. Victor Mason, for his efforts in bioinformatic analysis in other projects. Thanks to everyone in this lab for your kindness, support and understanding.

I would like to thanks China Scholarship Council for the chance they provide that I can study abroad.

Below, I would like to thank those people who support me in thousands of miles away in Chinese.

首先，我要感谢我的父母，感谢他们无私的抚养我长大，不遗余力的支持我追求想要的生活，感谢他们在身后默默的支持我。谨以此论文献给我的父母。其次，我要感谢我的朋友们，感谢他们在身边亦或千里之外的鼓励，支持和帮助。

自 2018 年 11 月 30 日到达德国开始，眨眼间四年已经在身后流淌而过。回首一幕幕，刚出国时的兴奋，进入第一个实验室的紧张，改派时的忐忑，到新实验室的重新振作，探讨问题时的激动，加班甚至通宵取时间点的疲惫，上班路上时隐时现的雪山的震撼，第一次攀岩，第一次漂流，第一次真正意义上的远距离徒步，第一次尝试瑞士餐厅的终身难忘，第一次见识圣诞舞会的好奇，还有数不尽的感触无法一一描述。总之，四年跌宕起伏的博士生涯告一段落，无所谓好与坏，对得起自己的付出，接下来我要砥砺前行。

# 10. Declaration of Consent

## Declaration of consent

on the basis of Article 18 of the PromR Phil.-nat. 19

Name/First Name: Zhang Fukai

Registration Number: 19-125-343

Study program: PhD in Molecular Life Science

Bachelor ☐ Master ☐ Dissertation ☑

Title of the thesis: Investigation of condensin related structural maintenance of chromosomes and Piggmac in Paramecium tetrauelia

Supervisor: Prof. Dr. Mariusz Nowacki

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 paragraph 1 litera r of the University Act of September 5th, 1996 and Article 69 of the University Statute of June 7th, 2011 is authorized to revoke the doctoral degree awarded on the basis of this thesis.
For the purposes of evaluation and verification of compliance with the declaration of originality and the regulations governing plagiarism, I hereby grant the University of Bern the right to process my personal data and to perform the acts of use this requires, in particular, to reproduce the written thesis and to store it permanently in a database, and to use said database, or to make said database available, to enable comparison with theses submitted by others.

Bern 26.09.2022

Place/Date

Signature Fukai Zhang