



HAL
open science

Etude du complexe impliqué dans l'élimination programmée d'ADN pendant la différenciation du noyau somatique chez *Paramecium tetraurelia*

Mélanie Bazin-Gelis

► **To cite this version:**

Mélanie Bazin-Gelis. Etude du complexe impliqué dans l'élimination programmée d'ADN pendant la différenciation du noyau somatique chez *Paramecium tetraurelia*. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université Paris-Saclay, 2023. Français. NNT : 2023UPASL003 . tel-04368449

HAL Id: tel-04368449

<https://theses.hal.science/tel-04368449>

Submitted on 1 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Étude du complexe impliqué dans
l'élimination programmée d'ADN pendant
la différenciation du noyau somatique
chez *Paramecium tetraurelia*

*Study of the complex involved in programmed DNA
elimination during the differentiation of the somatic nucleus in
Paramecium tetraurelia*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°577, Structure et dynamique des systèmes vivants (SDSV)
Spécialité de doctorat : Biologie moléculaire et cellulaire
Graduate School : Sciences de la vie et santé. Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **Integrative Biology of the
Cell (I2BC)** (Université Paris-Saclay, CEA, CNRS),
sous la direction de **Mireille BETERMIER**, directrice de recherche CNRS

Thèse soutenue à Paris-Saclay, le 6 février 2023, par

Mélanie BAZIN-GELIS

Composition du Jury

Membres du jury avec voix délibérative

Gaëlle LELANDAIS Professeure, Université Paris-Saclay	Présidente
Pascale LESAGE Directrice de recherche, IRSL	Rapporteur & Examinatrice
Laurent DURET Directeur de recherche, LBBE	Rapporteur & Examineur
Émilie BRASSET Professeure, Université Clermont Auvergne	Examinatrice
Cyril DENBY-WILKES Chercheur, CEA	Examineur
Éric MEYER Directeur de recherche, IBENS	Examineur

Titre : Étude du complexe impliqué dans l'élimination programmée d'ADN pendant la différenciation du noyau somatique chez *Paramecium tetraurelia*

Mots clés : Ciliés, réarrangements programmés d'ADN, transcriptomique, transposase domestiquée

Résumé : Chez de nombreux eucaryotes, le développement de la lignée somatique nécessite un réarrangement du génome. Chez l'unicellulaire *P. tetraurelia*, dont le dimorphisme nucléaire permet la séparation des génomes germinale et somatique dans deux types de noyaux (respectivement appelés MIC et MAC), cela se traduit par l'élimination programmée de ~50% des séquences germinales. Une partie de ces séquences doivent être précisément excisées pour restaurer l'intégrité des gènes exprimés à partir du génome somatique. Des cassures double-brin sont introduites à leurs bornes par une transposase PiggyBac domestiquée, PiggyMac (Pgm). L'hétérodimère Ku70/Ku80 est requis non seulement pour la réparation des cassures, mais également pour leur introduction. La ligation finale est réalisée par le complexe Ligase 4-Xrcc4.

Pendant ma thèse, j'ai étudié la régulation des gènes codant pour les sous-unités du complexe d'excision. Des analyses moléculaires ont révélé

que certaines sous-unités connues du complexe sont fortement surexprimées lorsque l'élimination d'ADN est perturbée par l'inactivation de *PGM*, *KU80c* ou *XRCC4*. Une analyse bioinformatique globale du transcriptome m'a permis de généraliser cette observation et d'identifier 628 gènes également surexprimés. Cette liste, qui contient la plupart des partenaires déjà connus de Pgm, devrait permettre d'identifier de nouveaux acteurs des réarrangements.

Mes analyses transcriptomiques, combinées à l'étude d'un rapporteur GFP introduit dans l'ancien MAC, suggèrent que les réarrangements du génome dans les nouveaux MAC régulent l'expression des gènes à partir de l'ancien MAC. L'induction de 42% de ces gènes dépend de la présence des nouveaux MAC. Il y aurait donc un dialogue entre les deux générations de MAC qui co-existent dans les cellules pendant le cycle sexuel.

Title: Study of the complex involved in programmed DNA elimination during the differentiation of the somatic nucleus in *Paramecium tetraurelia*

Keywords: ciliates, programmed DNA rearrangements, transcriptomics, domesticated transposase

Abstract: In many eucaryotes, the development of the somatic line requires programmed genome rearrangements. In the unicellular microorganism *P. tetraurelia*, whose nuclear dimorphism splits the germinal and somatic genomes in two types of nuclei (respectively called MIC and MAC), this consists in the programmed elimination of ~50% of germinal sequences. Part of these sequences must be excised precisely in order to restore gene integrity and enable their expression from the somatic genome. Double-strand breaks are introduced at their extremities by a domesticated PiggyBac transposase, PiggyMac (Pgm). The Ku70/Ku80 heterodimer is not only needed for the repair of DNA breaks but also for their introduction. The final ligation is carried out by the Ligase 4-Xrcc4 complex.

During my thesis, I studied the regulation of genes coding for the subunits of the excision

complex. Molecular analyses revealed that some known subunits of the complex are strongly overexpressed when DNA elimination is impaired by a *PGM*, *KU80c* or *XRCC4* knockdown. A global analysis of the transcriptome allowed me to extend this observation and identify 628 overexpressed genes under these conditions. This list contains most of the known partners of Pgm and should open the way to the identification of new actors involved in genome rearrangements.

My transcriptomics data, combined with the study of a GFP reporter gene introduced in the old MAC, suggest that genome rearrangements in the new MAC regulate the transient expression of genes from the old MAC. The induction of 42% of these genes depends on the presence of the new MAC. The data reported here point to a cross-talk between the two generations of MAC that co-exist in cells during their development.

Université Paris-Saclay
École doctorale 577 : Structure et dynamique des systèmes vivants
Institut de Biologie Intégrative de la Cellule

Thèse de Doctorat

Spécialité « Sciences de la vie et de la santé »

présentée par

Mélanie BAZIN-GELIS

Étude du complexe impliqué dans l'élimination programmée d'ADN pendant la différenciation du noyau somatique chez *Paramecium tetraurelia*

Soutenue le 06/02/2023 devant le jury composé de

Gaëlle Lelandais	Présidente
Pascale Lesage	Rapporteur & Examinatrice
Laurent Duret	Rapporteur & Examinateur
Émilie Brassat	Examinatrice
Cyril Deby Wilkes	Examinateur
Éric Meyer	Examinateur
Mireille Bétermier	Directrice de thèse

Remerciements

Aïe ! c'est la partie la plus difficile à écrire de ma thèse car elle convoque des sentiments comme la gratitude et l'admiration, que j'ai du mal à exprimer à l'égard de toutes les personnes qui m'ont entourée tout au long de ma formation et sans qui cette thèse n'aurait pas vu le jour. Certains de passage et d'autres comme des points de repère immuables. Cet exercice est finalement pour moi une bonne occasion d'écrire ce qui m'est difficile à dire, j'espère ne pas être trop maladroite.

Je me rends compte que nombreuses sont les personnes qui ont joué un rôle dans l'histoire de cette thèse. Cela commence dès mon plus jeune âge et c'est pourquoi je commencerais par remercier mes parents pour avoir toujours encouragé ma curiosité en répondant patiemment (ou me donnant les moyens de trouver les réponses) aux innombrables « pourquoi ? » et « comment ? » qui me traversaient l'esprit. Puis, beaucoup plus tard, les différentes équipes dans lesquelles j'ai eu la chance de faire mes stages. Je tiens à remercier tout spécialement Jacques Montagne qui, lors de mon premier stage, m'a transmis sa passion pour la recherche scientifique. C'est le souvenir de cette aura joyeuse et passionnée qui m'a confortée dans mon choix de vouloir faire de la recherche en laboratoire.

Faire une thèse est comme une histoire qui commence lorsque l'on rejoint son équipe. La mienne a donc commencé lorsque j'ai rejoint l'équipe « Réarrangements programmés du génome » pour mes stages de master et, plus particulièrement, lors de la préparation du concours de l'école doctorale. Cela était pour moi la première fois, enfin, que je pouvais me projeter sur un sujet plus complet sur lequel je travaillerai plus que quelques semaines. Je tenais donc à commencer en remerciant tous les membres de l'équipe pour leur investissement dès la préparation du concours de l'école doctorale. J'en profite également pour remercier les autres équipes et étudiants qui font répéter les candidats, car leur aide est précieuse dans la préparation de ce concours.

Remerciements

Travailler dans cette équipe m'a permis de rencontrer des personnalités très variées qui se complètent. Cela donne des réunions de laboratoire souvent très animées et intenses, car chacun a une opinion bien à lui. Mais confronter son avis à celui des autres permet de mieux construire sa pensée et de consolider son argumentaire. Il en ressort souvent de bonnes idées. Ces réunions ont été très formatrices et je suis très heureuse d'avoir eu ces échanges avec vous. Je vous suis également très reconnaissante pour ces petits échanges informels et le temps passé à expliquer et montrer les protocoles. Grâce à vous, j'ai pu acquérir non seulement de solides bases au niveau expérimental, mais surtout apprendre à réfléchir plus en chercheur qu'en étudiante. De manière un peu plus personnelle, je souhaite remercier Julien pour son honnêteté et sa franchise à toute épreuve, qui pousse à se dépasser, et Vinciane pour sa rigueur, qui fait mesure d'exemple et oblige à prendre en compte ce qui semblerait n'être qu'un détail. Sans oublier Marc pour le temps consacré à me montrer et m'expliquer les différents protocoles, Linda et plus récemment Valerio pour les discussions très intéressantes. Je voudrais aussi remercier Coralie toujours très volontaire et impliquée pour aider dans les manip' et pour les discussions plus techniques qui apportent un angle nouveau aux différents sujets. Un grand merci à Olivier qui a un peu été mon encadrant de l'ombre et m'a beaucoup appris pour toute la partie bio-informatique de mon projet de thèse.

Je voudrais tout particulièrement remercier Mireille, ma directrice de thèse. Merci pour tout ce temps que tu as consacré à me former pendant mes années de thèse, mais également lors des stages. Après toutes ces années à apprendre auprès de toi, tu es devenue pour moi un mentor à la paillasse et au labo de manière générale. Je retiendrais principalement cette bienveillance qui te caractérise et ton implication sans borne dans tout ce que tu fais. Mais également l'entrain avec lequel tu m'as toujours proposé ton aide pour les expériences sur les week-ends ou très (trop ?) tôt le matin lors des périodes de manip' intenses, dites « à la Mélanie ». Tu m'as aussi été d'une grande aide pour arriver à avancer malgré la pandémie du covid-19 qui nous a tous tenus éloignés de la paillasse et qui nous a forcés à réorienter ma thèse. Là encore, tu as su me donner les clés pour réussir en me proposant la formation DU-omique. Cette formation a été très bénéfique pour mon projet et le sera sans aucun doute à l'avenir. J'en profite pour te remercier pour le soutien que tu m'apportes dans ma projection de « l'après-

thèse » (même si je suis consciente que je ne te facilite pas trop la tâche pour ça). Je ne pense pas que ces quelques mots pourront rendre compte de tout ce que tu as pu m'apporter aussi bien scientifiquement qu'humainement. Sache simplement que je te suis très reconnaissante pour tout ce que j'ai pu apprendre à tes côtés pour ton soutien et ta bienveillance.

Mais une thèse ne se limite souvent pas à une équipe. Ma thèse n'aurait en effet pas été la même sans tous ces cafés et déjeuners partagés avec nos voisins de couloir : Anne-Marie, Michel, Khaled, Pierrick et Logan. Au-delà du « couloir paramécie », je souhaite remercier Carl Mann pour son aide sur la mise en place du Cut&Tag pleine de bienveillance et de générosité. Je voudrais également remercier les professeurs du DU-omique : Gaëlle Lelandais, Pierre Poulain et Bertrand Cosson ; ainsi que toute la promo 2020-2021 (même si on ne s'est pas beaucoup vus avec tout le distanciel imposé par la pandémie).

Enfin, il me tient tout particulièrement à cœur de remercier les personnes qui travaillent dans l'ombre, mais sans qui tout serait plus compliqué. À commencer par les « filles de la laverie » Cindy et Pascaline dont le travail nous simplifie les manip' au quotidien. Également merci à Artur que j'ai peu vu, mais qui nous a beaucoup facilité la vie pour les commandes (notamment en janvier) ainsi que Christine et Judith dont la disponibilité et la réactivité ont toujours été très appréciées pour les questions plus administratives.

Je remercie également les membres de mon comité de thèse : Éric, Sandra et Gaëlle, pour les discussions sur mes résultats lors de mes comités de thèse et les précieux conseils qu'ils ont pu me donner. Merci tout particulièrement à Gaëlle et Éric dont l'implication a dépassé les simples réunions annuelles. Gaëlle par son aide très précieuse (et sa patience) pour les analyses statistiques que j'ai eu à faire (pour lesquelles je suis toujours un peu perdue). Éric pour les nombreuses discussions pendant les réunions paramécie et par mail.

Un grand merci aussi à mes rapporteurs Pascale Lesage et Laurent Duret pour les corrections proposées sur mon manuscrit de thèse (malgré les petits problèmes administratifs). Merci également à l'ensemble de mon jury de thèse pour avoir accepté de m'écouter lors de ma dernière présentation en tant qu'étudiante. Un grand merci également pour la discussion très intéressante qui

Remerciements

a suivi et pour tous vos commentaires sur ma thèse qui m'ont vraiment touchée. Je suis très heureuse d'avoir pu partager ce moment avec vous.

Cependant, la thèse n'est pas qu'une histoire de laboratoire et de sciences. C'est pourquoi je souhaite également remercier mes parents et ma sœur pour avoir toujours été là pour me soutenir lors de cette thèse (et avant !) et pour avoir rendu les périodes de confinement liées à la pandémie du covid-19 presque agréables. Les mots me manquent pour remercier Mathieu, qui partage ma vie depuis le début de cette dernière année de thèse et qui a pour moi été un soutien indéfectible au quotidien. Cette dernière année n'a sûrement pas été la plus simple avec la pression de la fin de thèse et les désillusions de 4^{ème} année, mais vous m'avez tous permis de la traverser dans les meilleures conditions morales possibles.

Au-delà de la famille, je tiens à remercier toute ma petite bande d'amis qui, même par écran interposé, ont été présents pour moi tout au long de ces années. Ces moments partagés ont été comme des bulles d'air, m'obligeant à rentrer du labo le soir ou, plus récemment, à lever la tête de mon manuscrit de thèse. Un grand merci à toutes les CL que j'ai traversées pendant ces années sur FF et surtout un grand merci à Ayla, Luti, DD, Fima, Jinow, Hito, Shushu, Haro, Kao, Esphé, Suna, Mili, Darki, et tous ceux que j'oublie, pour ces heures de raid et de vocal !

Ces remerciements sont plus longs que ce que j'avais envisagé, mais je suis sûre que j'oublie encore des personnes et je m'en voudrai quand je m'en souviendrai. Donc à tous ceux que j'ai cités et tous ceux que j'ai oubliés, mais qui n'ont pas moins compté, je dis :

Merci !



Image par Hugo Amatte
Instagram: @skero_

Table des matières

Remerciements	i
Table des matières	v
Listes des abréviations, figures et tableaux	xv
Liste des abréviations	xv
Liste des figures	xxi
Liste des tableaux.....	xxv
PARTIE I – INTRODUCTION	1
Chapitre 1 - La paramécie comme modèle d'étude	3
1.1 La paramécie parmi les eucaryotes	5
1.2 Le dimorphisme nucléaire des ciliés.....	5
1.2.1 Micronoyaux (MIC).....	7
1.2.2 Macronoyau (MAC).....	9
1.3 Cycle de vie chez la paramécie	11
1.3.1 Croissance végétative.....	11
1.3.2 Cycles sexuels	13
1.3.3 L'expression des gènes	17
Chapitre 2 - Les réarrangements programmés du génome (RPG)	21
2.1 Élimination d'ADN entre lignées germinale et somatique	21
2.1.1 Élimination de chromosomes	22

2.1.2	Élimination programmée d'ADN.....	22
2.1.3	Les réarrangements : mécanisme de défense et d'innovation	22
2.2	Les réarrangements du génome chez <i>Paramecium tetraurelia</i>	26
2.2.1	Les séquences éliminées de manière imprécise : les éléments transposables et les séquences répétées	27
2.2.2	Les séquences précisément éliminées : les IES.....	33
2.3	Différence entre les RPG de <i>Paramecium</i> et <i>Tetrahymena</i>	45
2.3.1	Fragmentation des chromosomes	46
2.3.2	Élimination des IES.....	46
Chapitre 3 - Machinerie d'élimination programmée d'ADN		49
3.1	Reconnaissance des séquences à éliminer.....	51
3.1.1	Le contrôle maternel de l'élimination d'ADN.....	51
3.1.2	Une boucle de rétrocontrôle ARN pour assurer une excision complète des IES	57
3.1.3	Modification de la chromatine et marques épigénétiques	60
3.1.4	Quelles sont les séquences éliminées dépendantes des facteurs épigénétiques ?.....	67
3.1.5	Une séquence consensus optimisée pour certaines IES	69
3.1.6	Modèle de reconnaissance des séquences à éliminer	71
3.2	Le complexe d'excision des IES.....	72
3.2.1	L'excision une affaire de transposase.....	73
3.2.2	Des PiggyMac (like) en cascade.....	77
3.2.3	Un couplage entre cassure et réparation	78
3.2.4	L'élimination des IES chez <i>Tetrahymena</i>	81
3.3	Des cassures prise en charge par le NHEJ	82
Projet : Comprendre comment le complexe d'excision reconnaît ses cibles		85
	Identifier de nouveaux acteurs	87
	Localiser le complexe d'excision sur la chromatine.....	88

PARTIE II – RÉSULTATS.....	89
Chapitre 1 - Mise en évidence du dialogue entre les générations de MAC91	
1.1 Bloquer l’excision des IES induit une surexpression des gènes impliqués dans les RPG.....	91
1.1.1 La dérégulation transcriptionnelle a lieu dans le MAC parental ..	91
1.1.2 Bloquer la réparation de l’ADN induit-elle une surexpression tardive de la machinerie d’excision ?	93
1.1.3 L’excision des IES tardives est bloquée par la rétention des plus précoces lors d’un ARNi contre <i>XRCC4</i>	93
1.2 Le programme transcriptionnel de l’ancien MAC est dépendant du développement des nouveaux noyaux.....	95
1.2.1 L’expression dans le MAC parental est liée au développement du noyau zygotique	97
1.3 Modèle de dialogue entre l’ancien et les nouveaux MAC.....	97
Chapitre 2 - Identification de gènes candidats potentiellement impliqués dans les RPG 99	
2.1 Un problème de génération de séquençage.....	99
2.1.1 Validation de l’outil ComBat-seq.....	101
2.1.2 Définition des groupes de répliquats	103
2.2 Un dialogue intergénérationnel entre les MAC et identification de nouveaux candidats aux RPG (article en préparation).....	106
Chapitre 3 - Caractérisation de l’effet d’un ARNi contre le complexe condensine 157	
3.1 Les condensines de la paramécie	159
3.2 Un complexe condensine développemental.....	163
3.2.1 Une dérégulation spécifique des condensines D	163
3.2.2 Effets d’un ARNi contre une condensine D	163
3.3 Un complexe condensine végétatif.....	167

3.3.1	Un complexe condensine V	167
3.3.2	Effets d'un ARNi contre une condensine V.....	167
Chapitre 4 - Mise au point du protocole de Cut&Tag chez <i>P. tetraurelia</i>.		171
4.1	Mise au point du Cut&Tag	173
4.1.1	Localisation des protéines cibles du Cut&Tag.....	175
4.1.2	Localisation de la pA-Tn5 lors du Cut&Tag	181
4.1.3	Preuve de principe du Cut&Tag : détection de la marque H3K4me3	181
4.2	Voies d'amélioration	185
4.2.1	Production d'une pA-Tn5 plus active	185
4.2.2	Stabiliser le complexe d'excision pour faciliter le Cut&Tag.....	187
PARTIE III – DISCUSSION.....		189
Chapitre 1 - Dialogue entre les générations de MAC et définition de gènes candidats		191
1.1	Un lien étroit entre cassure et réparation de l'ADN	191
1.2	Un dialogue entre l'ancien et les nouveaux MAC	193
1.3	Dérégulation des gènes lorsque l'excision des IES est incomplète.....	194
1.3.1	De nombreux gènes indispensables à l'excision des IES sont dérégulés	194
1.3.2	Les nouveaux MAC pourraient contribuer à la dérégulation observée au temps tardif.....	196
1.3.3	Une proximité des gènes dérégulés avec le complexe d'excision	198
1.3.4	L'importance du complexe PRC2.....	199
1.4	Régulation de la transcription dans l'ancien MAC par les nouveaux MAC201	
1.4.1	Les nouveaux MAC sont importants pour le démarrage du programme transcriptionnel dans l'ancien MAC.....	201

1.4.2	Les nouveaux MAC sont nécessaires à l'expression de plus de la moitié des gènes connus impliqués dans les RPG.....	202
1.4.3	Identification d'un régulateur commun.....	203
1.4.4	Un motif nécessaire à l'expression des gènes.....	205
1.4.5	Perspectives	207
Chapitre 2 - Les condensines dans les RPG.....		209
2.1	Une spécialisation des Smc4 chez la paramécie.....	209
2.2	Un complexe dédié au développement des nouveaux MAC : la condensine D	211
2.2.1	La condensine D nécessaire au recrutement de la machinerie d'excision	211
2.2.2	Un complexe proche des condensines de <i>Tetrahymena</i> en composition ou en fonction.....	215
2.2.3	Des complexes Smc impliqués dans l'introduction et la réparation des cassures double-brin.....	216
2.3	Un complexe végétatif : la condensine V	217
2.3.1	Des condensines nécessaires à la division amitotique du MAC.	217
2.3.2	Plusieurs complexes condensine V	218
2.3.3	Perspectives	220
PARTIE IV – MATÉRIEL ET MÉTHODES.....		221
Chapitre 1 - Protocoles de biologie moléculaire.....		223
1.1	Souches de bactéries utilisées	223
1.2	Préparation de bactéries électrocompétentes	224
1.3	Transformation par électroporation.....	225
1.4	Préparation d'ADN issue de plasmides	226
1.4.1	MiniPrep	226
1.4.2	MaxiPrep d'ADN	227

1.5	Polymerase Chain Reaction (PCR)	227
1.6	Purification de fragments de plasmide	228
1.7	Ligation	229
1.8	Préparation d'ADN pour injection	230
1.9	Construction des plasmides.....	231
1.9.1	p349_pUC_GFP-UTR_PGM.....	231
1.9.2	p364_pUC_GFP-UTR_PGM-Motif_Asel & p418_pUC_GFP-UTR_PGM-Motif_INVlg	231
1.9.3	p361_3xFlag-Pgm & p362_3xFlag-Pgm _{D401A}	232
	Chapitre 2 - Protocoles de paramécies.....	233
2.1	Souches de paramécies utilisées.....	233
2.2	Conditions de culture	234
2.2.1	Dilution des milieux de culture	235
2.2.2	Contrôler la croissance des paramécies	235
2.2.3	Déclencher et suivre l'autogamie	236
2.3	Test de survie.....	236
2.4	Induction d'ARNi par <i>feeding</i>	237
2.4.1	Préparation des milieux	238
2.4.2	Conditions de culture	240
2.4.3	Plasmides de <i>feeding</i>	240
2.5	Suivie de la croissance végétative	240
2.6	Extraction d'ADN génomique.....	241
2.7	Micro-injection de transgène	241
2.7.1	Micro-injection dans le MAC végétatif	241
2.7.2	Sélection des transformants	242
2.7.3	Plasmides utilisés pour injections	245
2.8	Western blot	245
2.8.1	Récoltes.....	245

2.8.2	Préparation d'extrait protéique	246
2.8.3	Migration des protéines et western blot.....	246
2.8.4	Immunomarquage de membrane de nitrocellulose	247
2.9	Immunofluorescence sur cellules entières.....	248
2.9.1	Récoltes.....	248
2.9.2	Perméabilisation et fixation	249
2.9.3	Immunomarquage.....	250
2.10	Lame DAPI	251
2.11	Immunofluorescence sur noyaux.....	252
2.11.1	Récoltes.....	252
2.11.2	Lyse des cellules	253
2.11.3	Centrifugation basse vitesse	253
2.11.4	Marquages des noyaux.....	253
2.12	Compositions des tampons.....	254
	Chapitre 3 - Cut&Tag	261
3.1	Production d'une pA-Tn5 active.....	261
3.1.1	Production de la pA-Tn5	261
3.1.2	Formation du complexe actif	263
3.1.3	Test d'activité	264
3.1.4	Composition des tampons.....	264
3.2	Marquage des protéines en Cut&Tag.....	267
3.2.1	Matériel utilisé pour la localisation des protéines cibles.....	267
3.2.2	Matériel utilisé pour la localisation de la pA-Tn5	267
3.2.3	Protocole	268
3.3	Cut& Tag sur H3K4me3.....	271
3.3.1	Matériel utilisé	271
3.3.2	Protocole	272
3.4	Solutions pour Cut&Tag.....	275

3.5	Test de la qualité des banque de Cut&Tag	277
Chapitre 4 - Analyse bio-informatique		281
4.1	Données ARN utilisées	281
4.1.1	Origines des ARN utilisés	281
4.1.2	Séquençage des ARN.....	281
4.2	Génomomes de référence.....	282
4.3	Analyse d'expression différentielle des gènes	282
4.3.1	Mise en forme des données de comptage.....	283
4.3.2	Définition des pseudo-réplicats.....	287
4.3.3	Identification des gènes dérégulés par DESeq2.....	289
4.4	Caractérisation des gènes dérégulés.....	291
4.4.1	Définition d'une liste de candidats sous R	291
4.4.2	Caractérisation des candidats sous R.....	291
4.5	Identification d'un motif conservé.....	293
4.5.1	Identification du motif par STREME	293
4.5.2	Définition des gènes avec motif par FIMO.....	294
4.5.3	Analyse R du motif dans les promoteurs des gènes	295
PARTIE V – ANNEXES.....		297
Chapitre 1 - Données supplémentaires utilisées dans l'introduction		299
1.1	Reconnaissance des séquences à éliminer par les ARNnc.....	300
1.2	Modification de la chromatine et marques épigénétiques	301
1.3	Le complexe d'excision des IES.....	302
1.4	Des cassures prise en charge par le NHEJ	303
Chapitre 2 - Données supplémentaires de la partie résultats		305
2.1	Identification de gènes candidats potentiellement impliqués dans les....	305
2.2	Caractérisation de l'effet d'un ARNi contre le complexe condensine.....	307
2.3	Mise au point du Cut&Tag	310

Chapitre 3 - Données supplémentaires de la partie discussion.....	313
Chapitre 4 - Données supplémentaires de la partie matériel et méthodes	
321	
4.1 Protocoles de paramécies	321
4.2 Cut&Tag	323
PARTIE VI – BIBLIOGRAPHIE.....	325

Table des matières

Listes des abréviations, figures et tableaux

LISTE DES ABRÉVIATIONS

-A-

A : nucléotide adénine

ACP : Analyse en Composantes Principales

ADN : Acide désoxyribonucléique

ARN : Acide ribonucléique

ARNdb : ARN double-brin

ARNnc : ARN non-codant

ADNg : ADN génomique

ARNi : interférence ARN

ARNies : ARN ciblant les IES (*iesRNA*)

ARNscn : ARN scan (*scnRNA*)

-B-

BET : Bromure d'éthidium

BSA : Sérum Albumine Bovin (*Bovin Serum Albumin*)

-C-

C : nucléotide cytosine

Caf1 : *Chromatin assembly factor 1*

Listes des abréviations, figures et tableaux

CBS : *Chromosome Breakage Site*

ChIP : *Chromatin Immuno-Precipitation*

CIP : *Calf Intestinal alkaline Phosphatase*

Cp : *Cycle point*

Cph1 ou CAP-H1 : *Chromosome Associated Protein H1*

CRD : *Domaine riche en cystéine (Cysteine Riche Domain)*

-D-

D : acide aspartique

DAPI : 4',6-diamidino-2-phénylindole

DNA-PKcs : *protéine-kinase ADN-dépendante (DNA-dependent protein kinase)*

DO : *Densité Optique*

-E-

E : acide glutamique

EDTA : *acide éthylène diamine tétra-acétique*

Eed/EED : *Embryonic ectoderm development*

EGTA : *éthylène glycol-bis(β-aminoéthyl éther)-N,N,N',N'- tétra-acétique*

Ez11/EZL1 : *Enhancer of Zest-Like*

-F-

FACT : *Facilitate Chromatin Transcription*

FANS: *Fluorescence Assisted Nuclear Sorting*

-G-

G : nucléotide guanine

GFP : *Green Florescent Protein*

-H-

HAWK : *HEAT proteins associated with kleisins*

HEAT : *Huntingtin/EF3/PP2A/Tor1 (Huntingtin / Elongation Factor 3 / Protein Phosphatase 2A/ Tor1*

Hi-C : *High-throughput chromosome conformation capture*

HiChIP : *Hi-C based Chromatin Immunoprecipitation*

HMM : *Hidden Markov Models*

-I-

IES : *Internal Eliminated Sequence*

IESmc : IES sous contrôle maternel (*maternally controlled IES, mclES*)

IESscn : IES dépendante des ARN scan

IPTG : Isopropyl β -D-1-thiogalactopyranoside

ISWI : *Imitation-Switch*

-K-

K : nucléotide G ou T

kb : kilobase

Kp : *Klebsiella pneumoniae*

-L-

LB (milieu) : *Lysogeny broth*

-M-

MAC : macronoyau (lignée somatique)

Mb : mégabase

MIC : micronoyaux (lignée germinale)

mtA : *mating type A*

mtF : *mating type F*

mtFL : *mating type F Like*

-N-

N : nucléotide A, T, C ou G

NHEJ : *Non-Homologous End-Joining*

Ni : Non-injectée

NMD : *Nonsense Mediated mRNA Decay*

-O-

ORF : *Open Reading Frame*

-P-

PAF : paraformaldéhyde

pA-Tn5 : protéine A couplée à la transposase de Tn5

pb : paire de bases

PCI : Phénol-Chloroforme-Isoamyl alcool

PCR : *Polymerase Chain Reaction*

PDE : *Programmed DNA Elimination*

Pdsg/PDGS : *Paramecium protein involved in Development of the Somatic Genome*

Pgm/PGM : PiggyMac

PgmL/PGML : PiggyMac-Like

PolX/POLX : ADN-polymérase X

PICS : *Protease Inhibitor Cocktail Set1*

PRC2 : *Polycomb Repressive Complex 2*

-R-

R : nucléotide A ou G (purine)

RBAP46/48 : *Retinoblastoma-associated proteins 46/48*

RPG : Réarrangements Programmés du Génome

Rpm : rotation par minute

RT-qPCR : *Real Time quantitative PCR*

-S-

S : nucléotide G ou C

SDS : Sodium-Dodécile Sulfate

Smc/SMC : *Structural Maintenance of Chromosome*

Spt16/SPT16 : *Suppressor of Ty 16*

Suz12/SUZ12 : *Suppressor of zest 12*

-T-

T : nucléotide thymine

TAD : *Topological Associated Domain*

TAR : site de télomérisation (*Telomere Addition Region*)

TIR : *Terminal Inverted Repeats*

TPB : *Tetrahymena PiggyBac-like*

TSS : *Transcription Start Site*

TTS : *Transcription Termination Site*

-V, W, X, Y, Z-

VST : *Variance Stabilizing Transformation*

WGD : duplication globale du génome (*Whole Genome Duplication*)

WGP : *Wheat Grass Powder*

WT : *wild-type*

XLF : *Xrcc4 Like Factor*

Y : nucléotide C ou T (pyrimidine)

LISTE DES FIGURES

Figure 1 - Images de paramécies	2
Figure 2 - Arbres phylogénétiques des eucaryotes.....	4
Figure 3 - Structures des chromosomes MIC et MAC.....	6
Figure 4 - Plusieurs WGD ont participé à la divergence des <i>aurelia</i>	8
Figure 5 - Division végétative de la paramécie	12
Figure 6 - Cycle sexuel de la paramécie : la conjugaison	14
Figure 7 - Cycle sexuel de la paramécie : l'autogamie	16
Figure 8 - Profils des différents groupes de régulation des gènes au cours de l'autogamie.....	18
Figure 9 - Formation des chromosomes MAC	28
Figure 10 - Allongement des télomères.....	28
Figure 11 - Histogramme de la répartition des tailles des IES de moins de 150bp	32
Figure 12 - Élimination d'ADN lors du développement des MAC	32
Figure 13 - La séquence consensus aux bornes des IES est similaire à celle des <i>Tc1/mariner</i>	34
Figure 14 - Acquisitions des IES par <i>P. tetraurelia</i> au cours de l'évolution.....	36
Figure 15 - Cinétique d'excision des IES pendant le développement du MAC	36
Figure 16 - Évolution des IES.....	38
Figure 17 - Détermination du type sexuel chez les <i>aurelia</i>	42
Figure 18 - Comparaison des RPG entre <i>P. tetraurelia</i> et <i>T. thermophila</i>	44
Figure 19 - Sauvetage de la délétion MAC de l' <i>antigène A</i>	50
Figure 20 - Induction de la rétention d'une IES sous contrôle maternel.....	52
Figure 21 - Modèle du scan de génome par les ARNscn chez <i>P. tetraurelia</i>	54
Figure 22 - Modèle des ARNies.....	58
Figure 23 - Recrutement du complexe PRC2 sur ses cibles par les ARNscn.....	62
Figure 24 - Modèles d'excision des IES sur la chromatine	64
Figure 25 - Dépendance des IES aux facteurs épigénétiques	68
Figure 26 - Séquences consensus aux bornes des IES	70
Figure 27 - Modèle de reconnaissance des séquences à éliminer sur la chromatine	70

Figure 28 - Comparaison de la géométrie des cassures aux bornes des IES avec celle des transposons <i>Tc1/mariner</i> et <i>PiggyBac</i>	73
Figure 29 - La transposase domestiquée <i>PiggyMac</i>	74
Figure 30 - Les <i>PiggyMac</i> -likes	76
Figure 31 - Hiérarchie d'assemblage des <i>PgmL</i>	78
Figure 32 - Mécanisme d'excision des IES de <i>P. tetraurelia</i>	80
Figure 33 - Bloquer la machinerie d'excision induit une surexpression de <i>PGM</i> et <i>KU80c</i>	86
Figure 34 - Le rapporteur GFP reproduit fidèlement le profil d'expression de <i>Pgm</i>	90
Figure 35 - Un ARNi contre <i>XRCC4</i> induit une surexpression de <i>Pgm</i> aux temps tardifs.....	92
Figure 36 - Rétention des IES lors d'un ARNi contre <i>XRCC4</i>	94
Figure 37 - Effet d'un ARNi dirigé contre <i>CtIP</i> sur le développement des noyaux au cours de l'autogamie	96
Figure 38 - Un ARNi contre <i>CtIP</i> conduit à une perte d'expression de <i>Pgm</i>	96
Figure 39 - Analyse en composantes principales (ACP) des données avant correction par ComBat-seq	98
Figure 40 - Validation de la correction par ComBat-seq	100
Figure 41 - ACP des données après correction par ComBat-seq.....	102
Figure 42 - <i>Clustering</i> hiérarchique	104
Figure 43 - Complexe condensine	156
Figure 44 - Profils d'expression des gènes du complexe condensine	158
Figure 45 - Profils d'expression des composants du complexe condensine D de <i>P. tetraurelia</i>	160
Figure 46 - <i>Smc4-2</i> est nécessaire au déroulement des RPG.....	162
Figure 47 - Déstabilisation du complexe d'excision lors d'un ARNi contre <i>SMC4-2</i>	164
Figure 48 - <i>Smc4-1</i> est nécessaire à la ségrégation des MAC lors de la croissance végétative	166
Figure 49 - Comparaison des protocoles de CHIP et de Cut&Tag.....	172
Figure 50 - Sélection de transformants exprimant un niveau endogène de <i>Pgm</i>	174
Figure 51 - Contrôle qualité des noyaux extraits	176

Figure 52 - Localisation des protéines au cours du Cut&Tag.....	178
Figure 53 - Contrôle des banques obtenues après un Cut&Tag contre H3K4me3	180
Figure 54 - Vue sur le browser Jbrowse de ParameciumDB.....	182
Figure 55 - Le Cut&Tag sur H3K4me3 révèle un marquage spécifique des premiers nucléosomes des gènes.....	182
Figure 56 - Test d'activité des pA-Tn5 : coupure d'un plasmide supère enroulé	184
Figure 57 - Mutant catalytique de Pgm.....	186
Figure 58 - Modèle de rétrocontrôle de l'expression génique entre l'ancien et le nouveau MAC.....	192
Figure 59 - Intersection des gènes identifiés en TurboID et par analyse transcriptomique	199
Figure 60 - Comparaison des analyses des gènes surexprimés au temps tardif lors d'un ARNi contre <i>EZL1</i>	200
Figure 61 - Gènes surexprimés aux temps tardifs lorsque l'excision des IES est incomplète	200
Figure 62 - Gènes dépendant de la formation des nouveaux MAC surexprimés aux temps tardifs lorsque l'excision des IES est incomplète.....	204
Figure 63 - Modèle possible de régulation de la transcription dans l'ancien MAC par les RPG	205
Figure 64 - Deux complexes condensine I de <i>P. tetraurelia</i>	208
Figure 65 - Comparaison des protéines des complexes condensines.....	208
Figure 66 - Conservation des Smc	210
Figure 67 - Modèle d'interaction entre le complexe d'excision et la condensine D	212
Figure 68 - Les complexes condensine I de <i>T. thermophila</i>	214
Figure 69 - Production d'ARNdb dans la souche <i>E. coli</i> HT115.....	224
Figure 70 - Carte du plasmide p349	231
Figure 71 - Carte du plasmide p361	232
Figure 72 - Test trich	233
Figure 73 - Paramécies végétative et autogame	236
Figure 74 - Microinjections d'ADN	242
Figure 75 - Montage de la cassette de transfert de western blot.....	247
Figure 76 - Définition des stades d'autogamie	252

Figure 77 - Contrôle qualité des noyaux utilisés en Cut&Tag pour vérifier la localisation de la pA-Tn5.....	266
Figure 78 - Contrôle qualité des noyaux utilisés en Cut&Tag contre H3K4me3 .	270
Figure 79 - Schéma de la formation des banques Cut&Tag.....	274
Figure 80 - Définition des pseudo-réplicats puis des gènes différentiellement exprimés.....	286
Figure 81 - Récapitulatif des stades publiés pour les différentes cinématiques utilisées.....	288
Figure 82 - Profils d'expression des gènes pouvant coder pour les Smc et les kleisines composant les complexes condensine de la paramécie.....	308
Figure 83 - Profils d'expression des gènes pouvant coder pour les protéines HAWK des complexes condensine de la paramécie.....	309
Figure 84 - Exemple de profil de noyaux abimés	310
Figure 85 - Distribution de la taille des inserts séquencés pour les banques de Cut&Tag.....	311
Figure 86 - Comparaison de la position des inserts séquencés pour les Cut&Tag avec 2h ou 3h de tagmentation.....	312

LISTE DES TABLEAUX

Tableau 1 - Définition des pseudo-réplicats	105
Tableau 2 - Sous-unités des condensines I des vertébrés et des ciliés	158
Tableau 3 - Nombre de divisions effectuées par les paramécies en fonction de la température de culture.....	234
Tableau 4 - Production de milieu HT115 sans plasmide.....	238
Tableau 5 - Production de milieu HT115 avec plasmide non-induit	239
Tableau 6 - Production de milieu HT115 avec plasmide induit.....	239
Tableau 7 - Plasmides de <i>feeding</i>	240
Tableau 8 - Amorces utilisées pour les RT-qPCR.....	244
Tableau 9 - Plasmides injectés	245
Tableau 10 - Anticorps primaires utilisés en western blot	247
Tableau 11 - Anticorps secondaires utilisés en western blot.....	248
Tableau 12 - Anticorps primaires utilisés en immunofluorescence.....	250
Tableau 13 - Anticorps secondaires utilisés en immunofluorescence	250
Tableau 14 - Anticorps primaires utilisés en immunofluorescence sur noyaux.....	254
Tableau 15 - Oligonucléotides chargés sur la pA-Tn5	263
Tableau 16 - Utilisation des amorces pour le Cut&Tag.....	274
Tableau 17 - Oligonucléotides utilisés pour la PCR semi-quantitative	278
Tableau 18 - Références des ARN prélevés au cours des cinétiques d'autogamie	280
Tableau 19 - Références des séquençages ARN utilisés	280
Tableau 20 - Extrait du Tableau 27 contenant les métadonnées utilisées pour l'analyse des gènes différentiellement exprimés	282
Tableau 21 - Métadonnées des échantillons utilisés pour valider la correction par ComBat-Seq	284
Tableau 22 - Liste des gènes impliqués dans la reconnaissance des IES par les ARNnc.....	300
Tableau 23 - Liste des gènes du complexe PRC2	301
Tableau 24 - Liste des gènes impliqués dans le recrutement du complexe d'excision à la chromatine	301
Tableau 25 - Liste des gènes composant le complexe d'excision des IES.....	302

Tableau 26 - Liste des gènes impliqués dans la réparation des cassures introduites lors des RPG	303
Tableau 27 - Métadonnées utilisées pour l'analyse des gènes différentiellement exprimés.....	306
Tableau 28 - Caractéristiques des gènes condensines identifiés chez la paramécie	307
Tableau 29 - Dérégulation des gènes impliqués dans les RPG	316
Tableau 30 - Liste d'autres gènes impliqués dans les PRG.....	317
Tableau 31 - Gènes connus porteurs du motif dans leur promoteur	319
Tableau 32 - Plasmides utilisé pour induire les ARNi	322
Tableau 33 - Séquences des amorces disponibles pour Cut&Tag	323

Partie I – Introduction

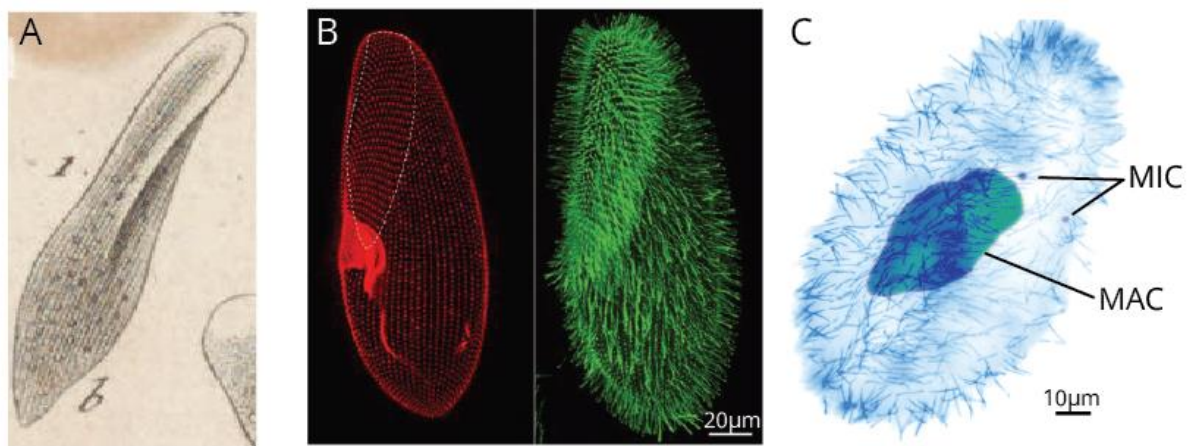


Figure 1 - Images de paramécies

A Une des premières représentations de paramécie du groupe des *aurelia* (image tirée de Bruguère, 1791). **B** Images d'immunofluorescence de *Paramecium tetraurelia* en utilisant un microscope confocal avec un marquage de la base des cils (gauche) et des cils (droite) (photo tirée de Bouhouche et al., 2021). **C** Image d'immunofluorescence de *Paramecium tetraurelia* réalisée par Janine Beisson montrant les cils couvrant la surface de la paramécie ainsi que les noyaux germinaux (MIC) et somatique (MAC).

Chapitre 1 - La paramécie comme modèle d'étude

Les paramécies font partie des premiers microorganismes identifiés avec l'apparition du microscope et les cultures de microorganismes dans des infusions de végétaux (foin, blé, avoine, thé,...). Elles furent décrites dès la fin du XVII^{ème} siècle (Van Leewenhoek, 1676) et la première illustration fut publiée en 1704 dans *Philosophical Transactions of the Royal Society* (Philosophical transactions, 1704). Il faut ensuite attendre 1752 pour que John Hill leur donne le nom de *Paramecium* d'après le grec *παραμήκης* (*paramēkēs*, "oblong") faisant référence à la forme de la cellule (Hill, 1752). En 1791, Jean-Guillaume Bruguère, sous la supervision de Jean-Baptiste Lamarck, définit les paramécies comme de très grosses cellules de forme ovale-oblongue bordée de cils, appartenant au groupe des infusoires.

Les paramécies sont des organismes eucaryotes unicellulaires qui mesurent entre 0,1 et 0,3 mm de long ce qui les rend facilement observables au microscope et isolables manuellement. Dans la nature, elles vivent en eau douce et se nourrissent d'autres microorganismes tels que des bactéries, des levures ou des algues. On les trouve principalement dans les lacs et les eaux stagnantes, elles peuvent être présentes dans des flaques d'eau dès lors qu'elles contiennent des végétaux en décomposition. La facilité d'observation et de culture des paramécies a permis leur étude dès le début du XVIII^{ème} siècle et en fait des organismes modèles encore très utilisés.

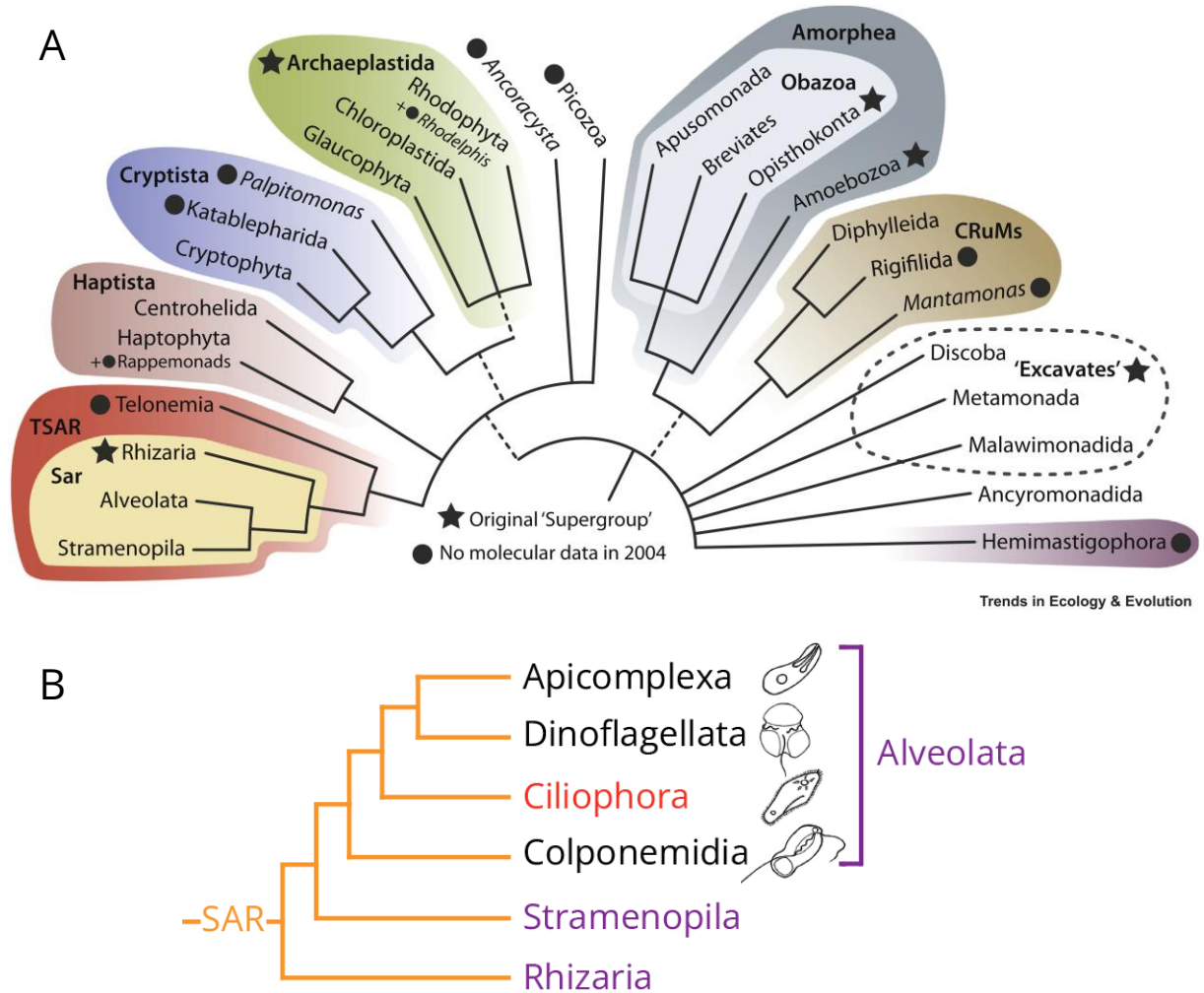


Figure 2 - Arbres phylogénétiques des eucaryotes

A Classification actuelle des eucaryotes sur laquelle les super groupes ont été colorés. Les lignes pointillées indiquent qu'il est incertain que les groupes en aval soient monophylétiques (figure tirée de Burki et al., 2020). **B** Zoom sur le supergroupe des SAR (*Stramenopila*, *Rhizaria* et *Alveolata*) et plus particulièrement sur le groupe des *Alveolata* dont font partie les ciliés (*Ciliophora*) (en rouge) (figure adaptée de Tikhonenkov et al., 2020).

1.1 LA PARAMÉCIE PARMIS LES EUCARYOTES

Aujourd'hui, les paramécies ne sont plus classées parmi les infusoires, mais parmi les alvéolés, au sein du grand groupe monophylétique des ciliés (Figure 2) (Adl et al., 2012; Burki et al., 2020; Grattepanche et al., 2018). Les ciliés regroupent plus de 8 000 espèces différentes d'eucaryotes unicellulaires qui, comme les paramécies, sont recouverts de cils. Les cils jouent un rôle très important dans la vie de la cellule car ils lui permettent de se déplacer, mais également de se nourrir et participent à la reproduction sexuée. En plus d'être pourvus de cils, les ciliés possèdent la particularité étonnante de présenter deux types de noyaux qui coexistent au sein de la cellule pendant tous les stades de son développement.

De nombreux groupes d'espèces ont rapidement été identifiés parmi les paramécies en s'appuyant sur leur morphologie. Ainsi, dès 1773, Otto Müller donne le nom d'*aurelia* au premier groupe d'espèces de paramécies (Müller, 1773). Il est décrit un peu plus tard parmi d'autres groupes identifiés selon des caractères morphologiques dans le Tableau encyclopédique et méthodique des trois règnes de la nature (Bruguière, 1791). Les nouvelles techniques permettent d'identifier toujours plus précisément les différentes espèces. Ainsi de nombreuses nouvelles espèces de ciliés ont été identifiées durant les trois siècles qui se sont écoulés depuis leur découverte. Le groupe des *aurelia* a ainsi été d'abord divisé en 12 espèces morphologiquement identiques par des approches génétiques (Sonneborn, 1974). Le séquençage des génomes a ensuite permis de confirmer ces espèces en ajoutant 3 espèces supplémentaires, portant le nombre total d'espèces d'*aurelia* à 15 (Catania et al., 2009; Gout et al., 2019; Sellis et al., 2021).

1.2 LE DIMORPHISME NUCLÉAIRE DES CILIÉS

Les ciliés ont la particularité de présenter deux types de noyaux dans leur cytoplasme. Ils sont formés à partir d'un même noyau zygotique et reflètent la séparation physique entre la lignée germinale et la lignée somatique. Le nombre de noyaux pour chacune des lignées peut varier d'une espèce de cilié à l'autre. *Paramecium tetraurelia* possède deux micronoyaux (MIC) qui forment la lignée germinale, et un macronoyau (MAC) qui correspond à la lignée somatique (Figure 3).

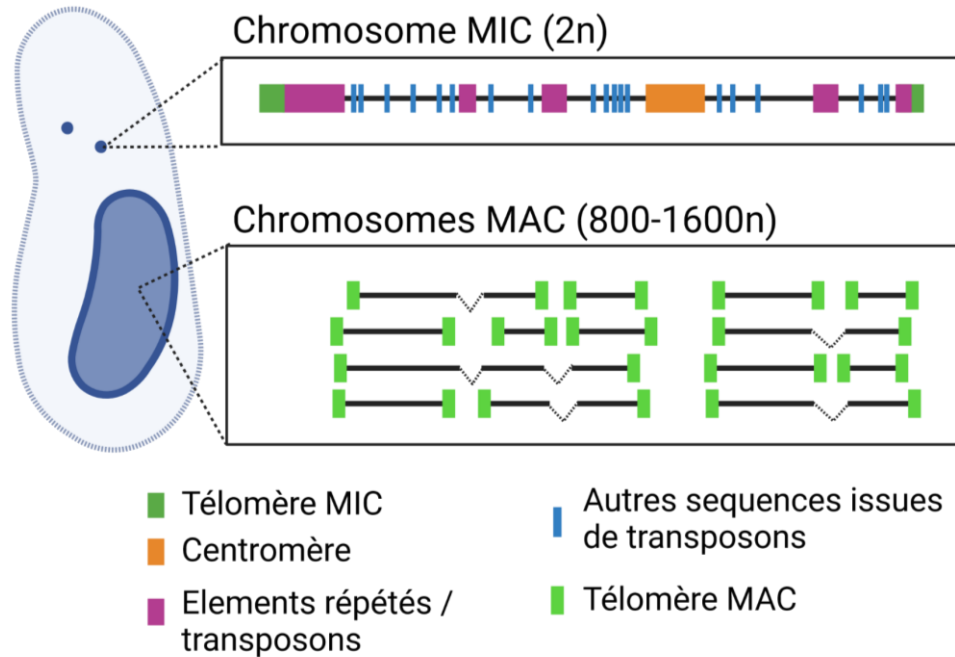


Figure 3 - Structures des chromosomes MIC et MAC

Les chromosomes MIC sont semblables à ceux des autres eucaryotes avec des centromères et des télomères. Ils possèdent également de nombreuses séquences issues de l'insertion de transposons et des séquences répétées (détails section 1.2.1). Les chromosomes MAC sont composés des mêmes séquences que les MIC, auxquelles il faut soustraire les séquences répétées et les séquences issues de transposons. Ils sont également dépourvus de centromère et de taille réduite par rapport aux chromosomes MIC (détails section 1.2.2). La nature et la localisation des centromères n'étant pas clairement établies, on suppose ici que le chromosome représenté ne possède qu'un centromère dont les séquences ADN associées sont éliminées des chromosomes MAC. (Created with BioRender.com)

1.2.1 Micronoyaux (MIC)

Les MIC sont de petits noyaux diploïdes (~3 µm de diamètre) qui correspondent à la lignée germinale de la paramécie. En effet, à chaque cycle sexuel seule l'information génétique contenue dans les MIC est transmise d'une génération à l'autre. L'étude de la structure des chromosomes et du génome MIC a été rendue possible par l'amélioration des techniques de purification des MIC et des techniques de séquençage. En effet, la combinaison d'étapes de centrifugation en gradients de Percoll et de cytométrie en flux a permis d'obtenir des MIC suffisamment purs pour en séquencer l'ADN (Duharcourt and Sperling, 2018; Guérin et al., 2017; Preer et al., 1992; Sellis et al., 2021). De plus, le séquençage Oxford Nanopore permet aujourd'hui d'obtenir de longues lectures permettant de résoudre des séquences répétées (Kono and Arakawa, 2019). L'assemblage du génome MIC est toujours en cours, mais il semblerait que les chromosomes MIC aient une structure assez similaire à celle des autres eucaryotes avec la présence de centromères (Lhuillier-Akakpo et al., 2016) sur des chromosomes linéaires bordés de télomères (même si la nature des centromères et télomères reste encore à déterminer).

Les séquençages de l'ADN des MIC ont permis d'estimer la complexité haploïde des noyaux germinaux entre 100 et 150 Mb chacun (Guérin et al., 2017; Sellis et al., 2021). L'analyse des séquences a permis de montrer la présence de nombreux transposons et séquences répétées témoignant de plusieurs vagues d'invasion du génome MIC par les transposons qui ont été conservés dans la lignée germinale. L'invasion par certains de ces transposons interrompt la phase codante de plus de la moitié des gènes de la paramécie. Le génome MIC ne peut donc pas produire une version viable de toutes les protéines. En pratique, le génome MIC n'est pas transcrit au cours de la vie végétative de la cellule, cette fonction étant laissée au noyau MAC. C'est cette spécificité qui a rendu possible l'intégration d'autant de transposons dans le génome germinal (Drews et al., 2022b).

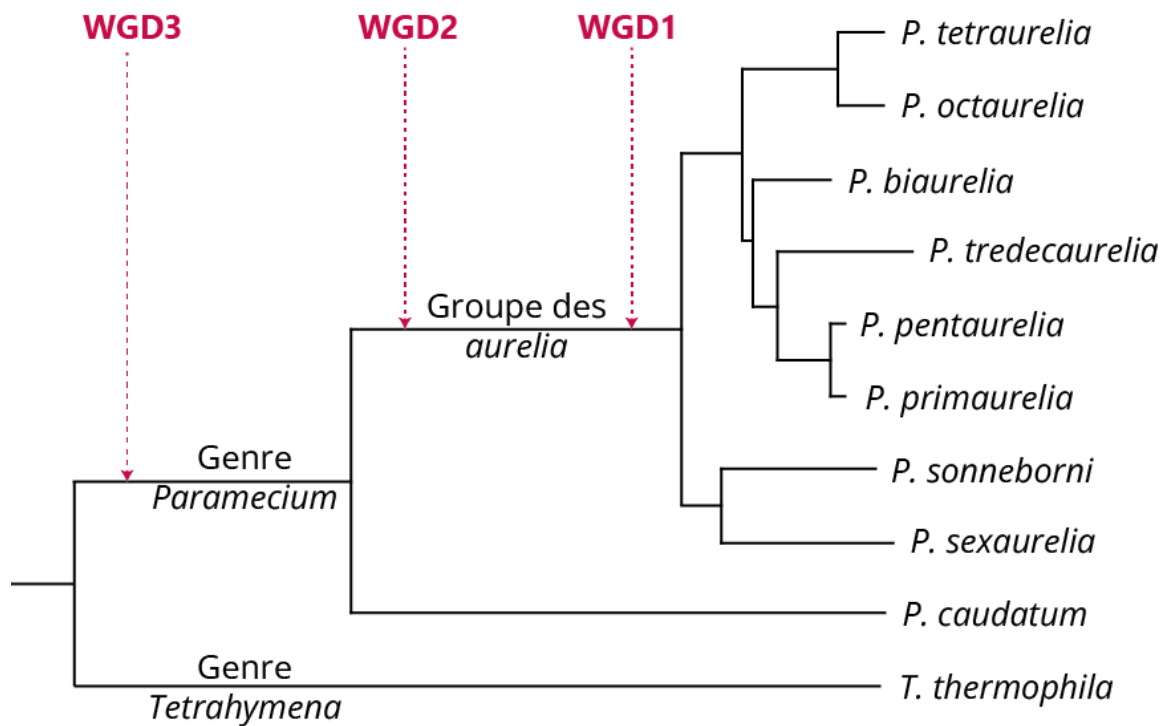


Figure 4 - Plusieurs WGD ont participé à la divergence des *aurelia*

Arbre phylogénétique des *oligohymenophora* (adapté de Aury et al., 2006; Sellis et al., 2021) sur lequel ont été replacées approximativement les duplications globales du génome (WGD).

1.2.2 Macronoyau (MAC)

Le noyau MAC correspond à la lignée somatique. Il sert à la production des transcrits nécessaires à la vie de la cellule pendant tout son cycle de vie, mais est détruit et un nouveau MAC est formé à chaque cycle sexuel (voir section 1.3.2). Sa taille est environ dix fois supérieure à celle des MIC (~30 μm , soit environ 1/5^{ème} de la taille de la paramécie).

Le MAC contient une version fragmentée des chromosomes MIC dépourvus de centromère et d'autres séquences répétées (Betermier and Duharcourt, 2014). Il est hautement polyploïde : sa ploïdie a d'abord été évaluée à ~800n (Guérin et al., 2017; Preer, 1976). Cependant, cette première estimation ne prenait pas en compte que près de 50% du génome MIC était éliminé pour former le génome MAC (Sellis et al., 2021). La ploïdie du MAC a donc été réévaluée à ~1600n.

1.2.2.1 Un génome riche en gènes grâce à des duplications globales

Le génome de la paramécie est naturellement riche en adénine et thymine (A et T). À la fois son génome MIC et son génome MAC sont composés à 70-75% d'A ou de T. Lors des premières identifications de gènes, une nouvelle caractéristique de la paramécie pour le moins étonnante a été mise à jour. En effet, la paramécie n'utilise pas le code génétique universel des codons. Seul le codon TGA est utilisé comme codon STOP, les codons TAA et TAG étant traduits en glutamine (Caron and Meyer, 1985; Preer et al., 1985). "Craquer" le code des codons de la paramécie a permis d'identifier 41 533 gènes (Arnaiz et al., 2017; Aury et al., 2006) faisant du génome MAC un génome particulièrement dense en gènes (~78% du génome codant pour quelques 40 460 protéines).

Ce grand nombre de gènes peut être en partie dû aux différentes duplications globales du génome (*WGD*, *Whole Genome Duplication*) qui ont eu lieu lors de l'évolution des paramécies (Aury et al., 2006). En effet, *P. tetraurelia* a subi au moins trois vagues de duplication (Figure 4). Les deux dernières auraient eu lieu avant la différenciation des espèces du groupe *aurelia* (McGrath et al., 2014a; McGrath et al., 2014b) mais après la divergence avec *P. caudatum*. La plus ancienne WGD a probablement eu lieu après la divergence entre *Paramecium* et *Tetrahymena* (Ruehle et al., 2016). De ces WGD résulte la duplication de nombreux gènes dont près de la moitié sont encore présents en deux copies issues de la

WGD la plus récente (Gout et al., 2010). Les familles de gènes résultant de duplications causées par les WGD sont appelées ohnologues (Ohno, 1970).

1.2.2.2 Un génome optimisé pour l'expression des gènes

Malgré sa forte densité en gènes et son volume bien supérieur au MIC, la taille du génome haploïde MAC est largement inférieure à celle du génome germinale avec seulement 72-75Mb (Guérin et al., 2017; Sellis et al., 2021). Les séquences présentes uniquement dans le génome MIC correspondent principalement à des anciennes insertions de transposons ou des séquences minisatellites répétées. L'élimination de ces séquences non-codantes permet l'optimisation du génome MAC pour la transcription et la restauration de l'intégrité des gènes dont la séquence a été interrompue dans le génome MIC.

De manière plus générale, il est facile de constater que le génome MAC a été optimisé pour la transcription tout au long de la vie de la cellule. Avec d'une part, une forte ploïdie permettant une production rapide des nombreux transcrits nécessaires à la production de protéines. Et d'autre part, une optimisation de la séquence avec l'élimination des séquences germinales délétères pour la vie de la cellule. Lors de sa vie végétative, la paramécie transcrit 65% de ses gènes à partir de son MAC (Arnaiz et al., 2017; Aury et al., 2006).

Même si des études restent à mener pour le confirmer, il semblerait que la chromatine dans le MAC de la paramécie soit particulière. En effet, elle semble très dense en nucléosomes et dépourvue d'histone H1 dans ses séquences *linker* très courtes (Drews et al., 2022a; Gnan et al., 2022). Cette observation est contraire au modèle selon lequel les régions peu couvertes par les nucléosomes sont plus transcrites. Cependant, la présence des marques de chromatine active H3K4me3 et H3K9ac et de la marque divalente H3K27me3 dans le MAC végétatif a conduit à un nouveau modèle selon lequel la présence des nucléosomes serait requise pour induire la transcription (Drews et al., 2022a; Gnan et al., 2022). Enfin, la chromatine ne porterait probablement pas de cytosine méthylée (5mC) typique de l'hétérochromatine (Singh et al., 2018). On trouve cependant des méthylations 6mAT au début des régions transcrites (Hardy et al., 2021).

1.3 CYCLE DE VIE CHEZ LA PARAMÉCIE

Le cycle de vie de la paramécie, comme pour tous les ciliés, peut être divisé en deux phases, avec d'une part une croissance végétative et d'autre part des cycles sexuels impliquant ou non un partenaire.

1.3.1 Croissance végétative

La croissance végétative (Figure 5) correspond à des divisions clonales. Au cours de cette phase, les noyaux MIC effectuent une mitose fermée (sans rupture de l'enveloppe nucléaire) afin de former de nouveaux noyaux qui vont se répartir dans les deux cellules filles (Drews et al., 2022b).

La division du MAC est, quant à elle, amitotique et n'est pas synchronisée avec celle des MIC. Le MAC commence à se répliquer bien avant que les MIC n'entrent en mitose (Berger and Schmidt, 1978). L'enveloppe nucléaire reste intacte pendant que le noyau s'allonge grâce à un réseau de microtubules intranucléaires (Tucker et al., 1980). Puis le MAC est séparé en deux au niveau du sillon de division pour se répartir dans les deux cellules filles. Ce processus ne semble pas impliquer de condensation des chromosomes. De plus, ces derniers semblent répartis aléatoirement entre les cellules puisqu'aucun fuseau mitotique ne permet d'assurer la ségrégation des chromosomes (Drews et al., 2022b). Le maintien de l'intégrité du génome lors des divisions végétatives semble surtout dépendre de la très grande ploïdie du MAC, qui assure une bonne probabilité qu'au moins une copie de chaque chromosome ségrège dans chacune des cellules filles. Une fois réparti entre les deux cellules filles, le MAC doit se répliquer afin de restaurer sa ploïdie globale (la ploïdie de chaque fragment de chromosome MIC peut être différente). Les premières études sur la paramécie semblent montrer qu'il existe un contrôle global de la quantité d'ADN dans les MAC (Berger and Schmidt, 1978). Cependant des études plus récentes ont montré que des transgènes injectés dans le MAC en quantités variables sont considérés comme des minichromosomes par la cellule. Ils sont répliqués et maintenus à une ploïdie constante au cours des divisions végétatives (Bourgain and Katinka, 1991; Garnier et al., 2004; Gotz et al., 2016). Ces résultats suggèrent qu'il pourrait exister un contrôle du maintien de la ploïdie au niveau des chromosomes.

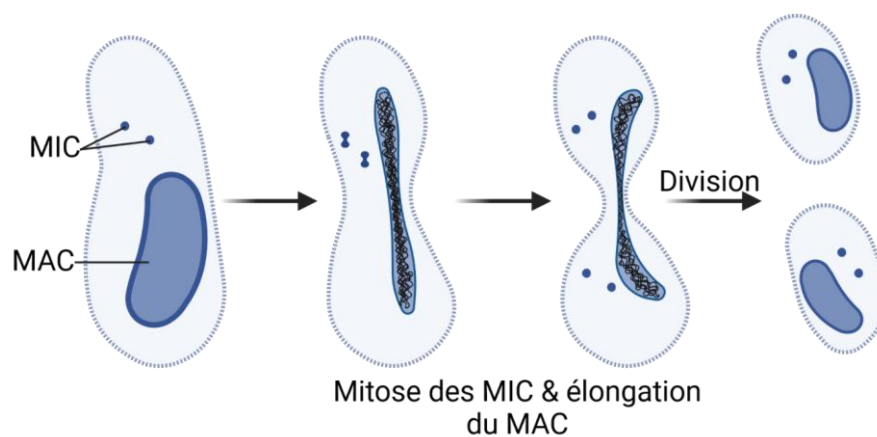


Figure 5 - Division végétative de la paramécie

Lors d'une division végétative, les MIC se divisent par mitose avant de se répartir entre les deux cellules filles alors que le MAC fait une division amitotique (détails section 1.3.1). Les cellules filles obtenues après une division végétative sont des clones de la cellule parentale. (Created with BioRender.com)

Enfin, la croissance végétative est une division clonale qui n'implique pas de méiose des MIC. Elle marque donc le vieillissement des cellules. En effet, l'âge des cellules est compté en nombre de divisions végétatives. Des travaux de transplantation de noyaux entre de jeunes et de vieilles cellules ont permis de déterminer que c'est le noyau MAC qui est le senseur de l'âge de la cellule (Aufderheide, 1986). Cependant, chez la paramécie, la taille des télomères de part et d'autre des chromosomes MAC reste constante durant toute la vie de la cellule (Gilley and Blackburn, 1994). En revanche, il semblerait qu'au cours des divisions les chromosomes MAC se fragmentent de plus en plus et des télomères sont ajoutés à certaines de ces nouvelles extrémités. La télomérase semble donc active tout au long de la vie de la cellule. Le senseur du vieillissement chez la paramécie ne résiderait donc pas dans la longueur de ses télomères, mais dans la fragmentation de ses chromosomes MAC. Si elles n'effectuent pas de cycle sexuel, les paramécies ont une espérance de vie moyenne de 150 divisions clonales après la division suivant la formation du noyau zygotique (appelée division caryonidale) (Figure 6).

1.3.2 Cycles sexuels

Les espèces du groupe *aurelia* possèdent deux types de cycle sexuel : l'un impliquant la présence d'un partenaire de type sexuel compatible (la conjugaison), l'autre n'impliquant pas de partenaire (l'autogamie) (Drews et al., 2022b).

1.3.2.1 Conjugaison

La reproduction par conjugaison engage deux partenaires de types sexuels différents : les types O (*odd*) et E (*even*) chez les *aurelia*. Afin que la conjugaison puisse avoir lieu, il faut que les partenaires soient jeunes (~10 divisions depuis le dernier cycle sexuel) et dans un état appelé réactivité (Singh et al., 2014; Sonneborn, 1937).

Lors de la conjugaison (Figure 6), les paramécies de types sexuels différents s'apparient par la bouche. La formation des couples déclenche la méiose des MIC tandis que le MAC reste intact. La méiose des MIC produit 8 noyaux haploïdes (équivalents des gamètes des métazoaires). Dans chacun des conjugants, un noyau haploïde migre vers le cône paroral où il se divise par mitose pour donner

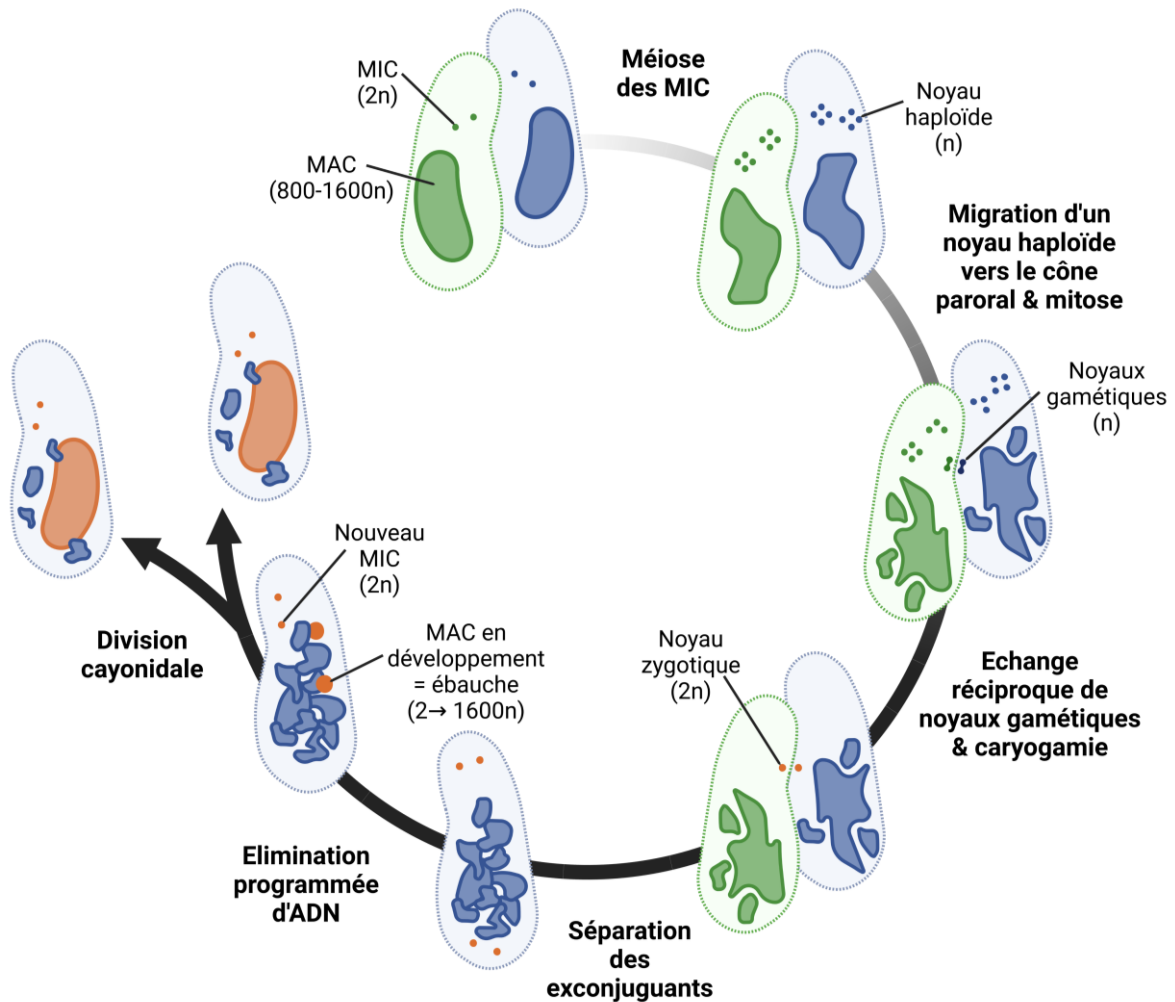


Figure 6 - Cycle sexuel de la paramécie : la conjugaison

La conjugaison (décrite dans la section 1.3.2.1) correspond à un cycle sexuel classique puisqu'elle implique que deux partenaires de type sexuels compatibles échangent des noyaux gamétiques qui fusionnent pour former le noyau zygotique. Les cellules obtenues à la suite d'un tel croisement possèdent donc 50% du génome « maternel » et 50% du génome « paternel ». (Created with BioRender.com)

deux noyaux gamétiques identiques. Les sept autres noyaux haploïdes sont perdus. Les deux partenaires effectuent un échange réciproque de leurs noyaux gamétiques. Le noyau gamétique résidant et celui issu de l'échange fusionnent ensuite par caryogamie pour former le noyau zygotique diploïde. Le noyau zygotique fait une première mitose puis, les deux partenaires se séparent. Ils seront restés appariés pendant ~5h (à 27°C).

Après la séparation des exconjugants, les noyaux zygotiques effectuent une seconde mitose pour former quatre noyaux diploïdes. Grâce à des gradients intracellulaires de potassium et de calcium, les deux noyaux du pôle antérieur de la cellule deviendront des MIC et ceux du pôle postérieur des MAC (Betermier and Duharcourt, 2014; Grandchamp and Beisson, 1981). Les noyaux zygotiques sont des noyaux diploïdes provenant des MIC. Comme eux, ils possèdent une complexité entre 100 et 150Mb (voir section 1.2.1). Cependant, les noyaux MAC matures sont hautement polyploïdes et possèdent une complexité de ~75Mb (voir section 1.2.2). Lors de leur développement en MAC, les noyaux zygotiques devront éliminer ~50% de leur génome par le biais de réarrangements programmés du génome (RPG). Dans le même temps, les MAC en développement vont suivre 6-7 cycles d'endoduplication (Berger, 1973) afin d'atteindre une ploïdie de ~128n ; en prenant en compte le fait que le MAC élimine 50% de son ADN, les 74c estimés dans l'article correspondent donc à ~150c. Ces premiers cycles d'endoduplication ont la particularité d'être synchrones créant des *pulses* de réplication (Guérin et al., 2017).

Les MIC effectuent ensuite une dernière mitose qui déclenche la division caryonidale de la paramécie. Chacune des cellules filles recevra deux MIC et un MAC en développement. Les nouveaux MAC terminent leur développement dans les cellules filles et effectuent 4-5 cycles d'endoduplication supplémentaires pour atteindre la ploïdie des MAC matures (Berger, 1973).

Dès l'apparition des 8 noyaux haploïdes, le MAC parental commence à se débobiner et continue sa fragmentation durant tout le développement des nouveaux MIC et MAC. Il faut attendre quelques divisions après la division caryonidale pour que les derniers fragments du MAC disparaissent. Malgré sa

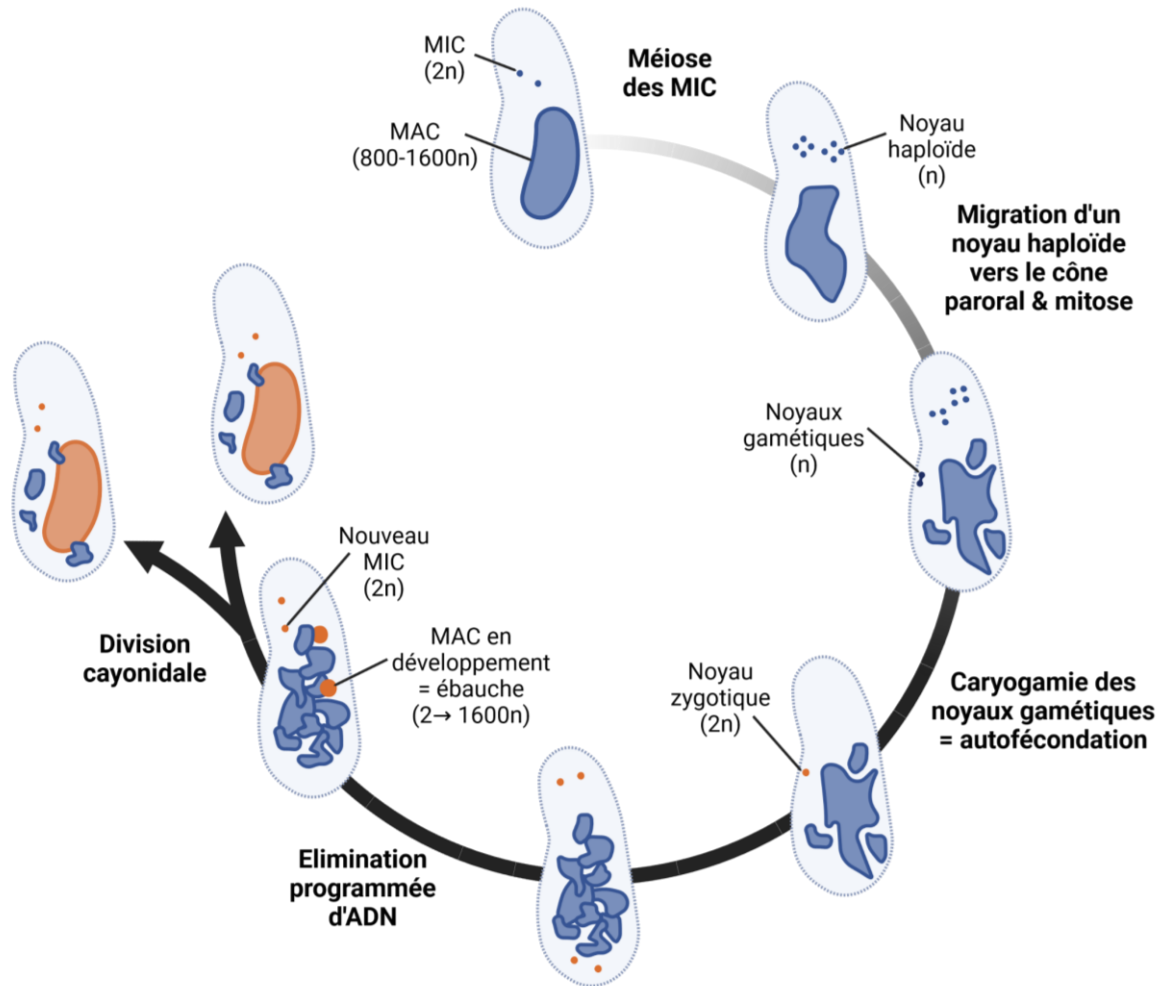


Figure 7 - Cycle sexuel de la paramécie : l'autogamie

Le cycle de l'autogamie (décrit dans la section 1.3.2.2) est très similaire à celui de la conjugaison. La différence réside en l'absence de partenaire. Après leur mitose dans le cône paroral, les noyaux gamétiques fusionnent par caryogamie. De ce fait, toutes les cellules résultant d'une autogamie sont complètement homozygotes. (Created with BioRender.com)

fragmentation, le MAC parental est toujours transcriptionnellement actif et produit les transcrits dont les nouveaux MAC ont besoin pour se développer (Betermier and Duharcourt, 2014).

1.3.2.2 Autogamie

L'autogamie est très similaire à la conjugaison, mais elle n'implique pas la présence de partenaire (Figure 7). De plus, afin de déclencher leur autogamie, les cellules ont besoin d'avoir effectué 20 à 30 divisions végétatives depuis leur dernier cycle sexuel. Enfin, le déclenchement de l'autogamie suit en général un stress comme la carence de nourriture. Le vieillissement des cellules peut également conduire au déclenchement de l'autogamie.

Comme la conjugaison, l'autogamie commence par la méiose des MIC dont l'un des produits migre dans le cône paroral avant de se diviser. Contrairement à la conjugaison, il n'y a pas d'échange de noyaux gamétiques. Les deux noyaux fusionnent simplement par caryogamie dans le cône paroral effectuant une étape d'autofécondation. La suite du processus est identique à celui de la conjugaison. Cependant, étant donné que l'autogamie implique la caryogamie de noyaux identiques, elle donne des cellules 100% homozygotes contrairement à la conjugaison (Betermier and Duharcourt, 2014).

Du fait de ses conditions de déclenchement, l'autogamie est moins synchrone que la conjugaison. Mais elle peut être déclenchée plus facilement dans un plus grand volume de culture. C'est pourquoi les études du développement des nouveaux MAC sont souvent réalisées lors de l'autogamie avec comme hypothèse que, à l'échelle moléculaire, les processus de RPG se déroulent de la même manière lors de la conjugaison et de l'autogamie.

1.3.3 L'expression des gènes

1.3.3.1 Profils d'expression des gènes au cours de l'autogamie

L'étude du transcriptome au cours de l'autogamie a montré que l'expression de nombreux gènes varie lors du cycle sexuel. Ces gènes peuvent être classés en six catégories déterminées par leur profil d'expression (Figure 8) (Arnaiz et al., 2017).

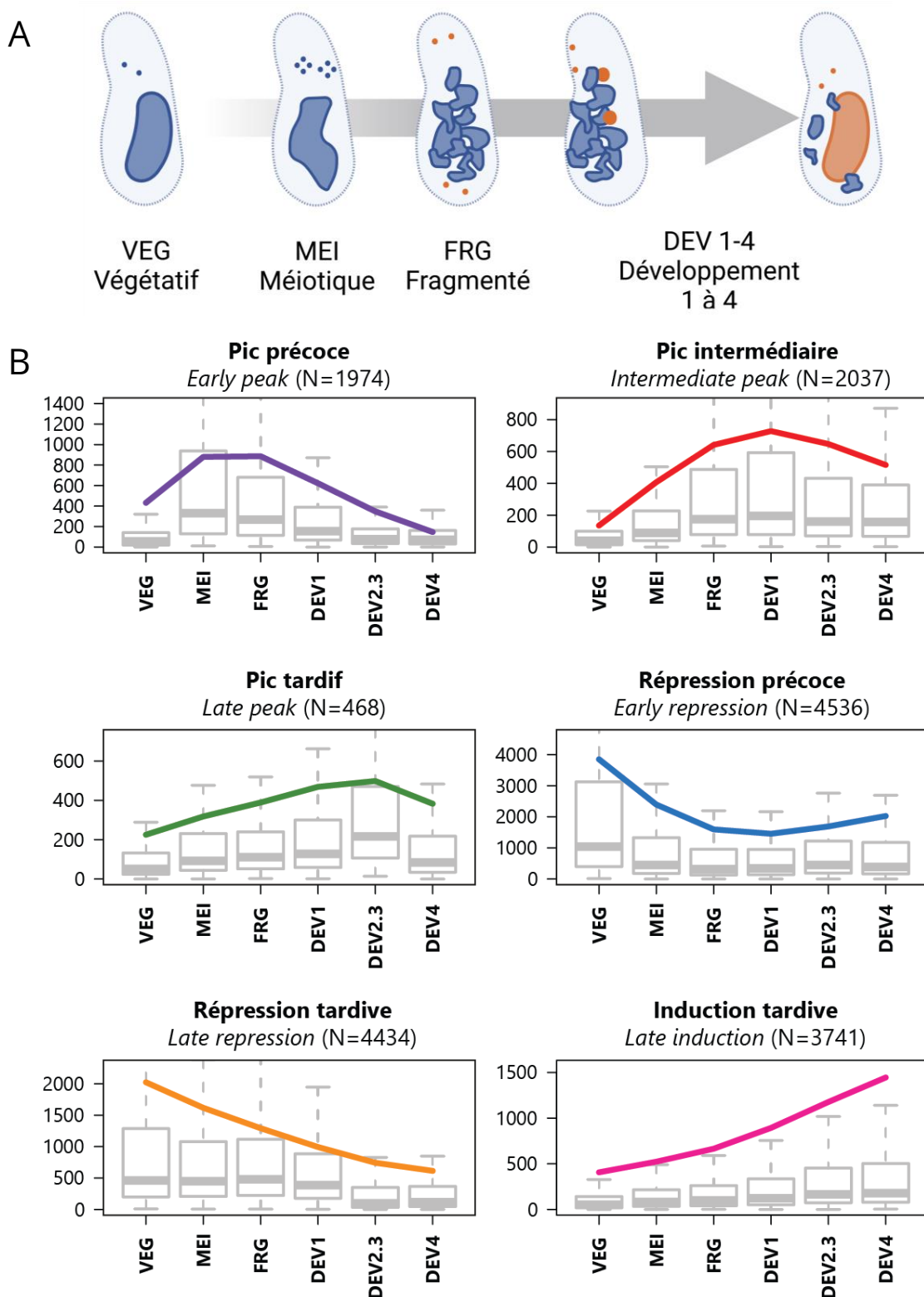


Figure 8 - Profils des différents groupes de régulation des gènes au cours de l'autogamie

A Description des stades de l'autogamie (*Created with BioRender.com*). B Profils obtenus lors de l'étude du transcriptome de *P. tetraurelia* au cours de l'autogamie. Les gènes dont l'expression varie lors de ces cycles ont été classés en six groupes décrits dans le texte (figure adaptée de Arnaiz et al., 2017).

- Pic précoce ou *early peak* sont les gènes dont l'expression est maximale au moment de la méiose et de la caryogamie,
- Pic intermédiaire ou *intermediate peak* sont les gènes dont le maximum d'expression est atteint durant le développement des nouveaux MAC avant de diminuer,
- Pic tardif ou *late peak* sont les gènes dont l'expression atteint son maximum à la fin du développement des nouveaux MAC et chute avant la fin de l'autogamie,
- Induction tardive ou *late induction* sont les gènes dont l'expression augmente de façon continue jusqu'à la fin du développement du MAC,
- Répression tardive ou *late repression* sont les gènes exprimés pendant la croissance végétative et dont l'expression diminue tout au long de l'autogamie pour atteindre son minimum à la fin du développement du MAC,
- Répression précoce ou *early repression* sont les gènes exprimés au cours de la vie végétative de la cellule et dont l'expression diminue au début de l'autogamie avant d'augmenter sur la fin du cycle sexuel.

A ces catégories, on peut ajouter une dernière catégorie qui comprend 59% des 41 533 gènes de la paramécie :

- Constitutif ou *none* sont des gènes dont l'expression ne varie pas au cours de l'autogamie.

1.3.3.2 Une régulation globale de l'expression

Bien que les profils d'expression des gènes au cours du développement de la paramécie aient pu être décrits, il ne faut cependant pas oublier que les noyaux MAC depuis lesquels ces gènes sont exprimés sont polyploïdes. Si avoir des gènes en un grand nombre de copies permet de produire efficacement des transcrits, il faut toutefois que l'expression de toutes ces copies soit régulée. Chez la paramécie, il est peu probable que toutes les copies d'un même gène soient coréglées, d'autant que le nombre de copies de chaque chromosome MAC pourrait varier d'un individu à l'autre. On peut alors imaginer une régulation plus globale qui se ferait au travers de dosage des gènes. Ainsi, chaque copie du gène peut avoir un niveau de transcription différent pour assurer un niveau global de transcrit stable (Drews et al., 2022a; Drews et al., 2022b). Mais le mécanisme derrière cette régulation reste encore méconnu.

Chapitre 2 - Les réarrangements programmés du génome (RPG)

Les noyaux somatiques de la paramécie (MAC) ont une complexité haploïde 30-50% inférieure à celle des noyaux germinaux (MIC) dont ils sont issus. Il y a donc une élimination d'ADN entre la lignée germinale et la lignée somatique. L'élimination d'ADN dans les lignées somatiques n'est pas particulière aux ciliés. En effet, de nombreux organismes multicellulaires éliminent une partie de leur ADN germinale dans leurs cellules somatiques (Wang and Davis, 2014). Cette élimination est généralement liée à la différenciation des cellules somatiques pendant le développement embryonnaire ou à la détermination du type sexuel de l'individu.

2.1 ÉLIMINATION D'ADN ENTRE LIGNÉES GERMINALE ET SOMATIQUE

Des différences de contenu génétique entre les lignées germinale et somatique ont pu être identifiées dans plus d'une centaine d'espèces largement réparties dans plusieurs grands groupes d'eucaryotes. Les mécanismes permettant la réduction de la complexité haploïde entre les lignées somatique et germinale peuvent être classés en deux grandes catégories, avec d'une part l'élimination programmée d'ADN et d'autre part l'élimination de chromosomes entiers.

2.1.1 Élimination de chromosomes

L'élimination de chromosomes entiers a été décrite chez de nombreux marsupiaux, mais également chez d'autres organismes. On trouve des exemples parmi les plantes, les oiseaux et les insectes. L'élimination de chromosomes peut avoir lieu dans les cellules précurseurs des cellules somatiques. Ainsi, à l'exception des cellules germinales, tout l'organisme aura le même nombre de chromosomes. Cette perte de chromosome peut notamment servir à la détermination du type sexuel de l'individu, comme c'est le cas chez certaines mouches et certains marsupiaux. Chez certaines plantes, l'élimination de chromosome peut également être effectuée de manière tissu-spécifique, principalement pour éliminer de l'ADN non-codant (Borodin et al., 2022; Kloc et al., 2022; Wang and Davis, 2014).

2.1.2 Élimination programmée d'ADN

Comme pour l'élimination de chromosomes, l'élimination programmée de chromatine (aussi appelée réarrangements programmés du génome (RPG) chez la paramécie) consiste en général en l'élimination de séquences d'ADN répétées, de transposons et de gènes spécifiques de la lignée germinale. L'élimination programmée d'ADN, comme celle de chromosomes, a été identifiée dans de nombreux groupes phylogénétiques différents. La quantité d'ADN éliminée varie de ~10% du génome germinale chez certains nématodes à plus de 90% chez les copépodes (Wang and Davis, 2014). On trouve également des exemples bien décrits d'élimination d'ADN parmi les vertébrés, avec les myxines et les lamproies qui éliminent une partie des séquences germinales de leurs tissus somatiques. Les mécanismes permettant ces réarrangements varient beaucoup d'un système à l'autre (Drotos et al., 2022; Kloc et al., 2022). Les réarrangements ne touchent pas forcément tous les tissus de la lignée somatique, comme le montre le système de l'immunité des vertébrés permettant de générer la forte diversité des anticorps (Christie et al., 2022).

2.1.3 Les réarrangements : mécanisme de défense et d'innovation

Le paragraphe précédent montre que nombreux organismes différents éliminent une partie de leur ADN dans leur lignée somatique. Mais il est justifié de se demander quel est l'intérêt pour les différents organismes de maintenir cet

ADN dans leur génome germinale et de le transmettre à leur descendance, si c'est pour l'éliminer à chaque génération.

Il faut d'abord rappeler que la majorité des séquences éliminées sont des séquences répétées ou des séquences issues d'invasion par des transposons qui pourraient être délétères pour le génome, si elles étaient exprimées et venaient à s'insérer dans des gènes (Borodin et al., 2022; Kloc et al., 2022; Wang and Davis, 2014). L'élimination de ces séquences du génome somatique de l'hôte est une façon radicale de s'assurer que les transposons ne seront pas transcrits. De plus, on peut imaginer qu'au cours de l'évolution les éléments transposables pourraient finir par être éliminés du génome germinale.

Cependant, les RPG permettent au génome germinale de fluctuer et d'accumuler des éléments plus rapidement puisque les séquences introduites sont par la suite éliminées du génome somatique (Drotos et al., 2022). Être capable de RPG pourrait donc permettre une évolution plus rapide des espèces et donner lieu à des événements de spéciation.

2.1.3.1 Élément transposable : de menace génomique à avantage évolutif

2.1.3.1.1 Rendre la menace silencieuse

Les transposons, par exemple, sont des séquences souvent considérées comme des menaces pour la stabilité du génome à cause de leur capacité à se déplacer dans celui-ci. Ils sont une des principales sources de mutation naturelle et peuvent donc causer des maladies telles que les cancers ou des maladies neurodégénératives. Même leur nature répétitive peut être source de mutation puisqu'elle augmente les risques de recombinaisons homologues entre chromosomes. Ces recombinaisons peuvent être à l'origine de délétions, de duplications, de translocations ou d'inversions de bras entier de chromosomes (Cosby et al., 2019; di Stefano, 2022).

Afin d'éviter le saut des transposons, les organismes vivants ont mis en place divers mécanismes permettant leur inactivation. Cette inactivation peut avoir lieu au niveau transcriptionnel avec la formation locale d'hétérochromatine ou au niveau post-transcriptionnel par le biais d'ARN interférence (ARNi) (Cosby et al., 2019; Délérès et al., 2021; di Stefano, 2022; Maupetit-Mehouas and Vaury, 2020;

Senti et al., 2015). De plus, au cours du temps, les séquences de transposons silencieux accumulent naturellement des mutations puisqu'elles ne subissent plus de pression de sélection. Les mutations peuvent alors s'accumuler jusqu'à ce que les éléments transposables dégénèrent et ne soient plus reconnaissables.

2.1.3.1.2 Les éléments transposables : sources de mutations et d'évolution

Mais ces sources de mutations sont également une importante source de variabilité et d'adaptation. C'est probablement l'une des raisons pour laquelle, dans de nombreuses espèces, les transposons sont réactivés lors d'une courte période du développement (Chuong et al., 2017; Maupetit-Mehouas and Vaury, 2020).

L'introduction de mutations en amont de certains gènes peut changer leur régulation et faire évoluer l'organisme. Par exemple, chez les plantes c'est l'insertion de transposons en amont de gènes qui a notamment permis la déclinaison d'orange en orange sanguine ou les différents cépages de raisins (Lisch, 2013). L'insertion de transposons en amont de différents gènes pourrait même être à l'origine de la corégulation de ces gènes. Les éléments transposables auraient donc la capacité de créer ou de modeler des réseaux de régulation chez leur hôte (Chuong et al., 2017; Goerner-Potvin and Bourque, 2018). On trouve de nombreux autres exemples chez les insectes pour qui les transposons sont à l'origine de nombreuses résistances aux pesticides (Gilbert et al., 2021).

L'insertion de transposon dans la séquence codante d'un gène peut parfois présenter un avantage évolutif comme pour la phalène du bouleau lors de la révolution industrielle. L'acquisition d'un transposon dans l'un des gènes responsables de la couleur blanche de ses ailes lui a permis d'obtenir des ailes foncées, moins facilement repérables sur les troncs des arbres noircis par le charbon (Hof et al., 2016).

Les transposons acquis dans le génome germinale peuvent également apporter un bénéfice à la cellule hôte à travers la cooptation ou la domestication de certaines de leurs protéines. La cooptation (*taming* en anglais) correspond à une protéine qui est détournée de sa fonction première (assurer la mobilité du transposon) pour accomplir une fonction cellulaire (Capy, 2021). Par exemple, chez les insectes des reverse-transcriptases codées par des rétrotransposons

(transposons à ARN) ont été cooptées pour participer à la réponse immunitaire contre les virus (Gilbert et al., 2021).

La domestication est un stade plus avancé de la mise des éléments transposables au service de l'organisme puisqu'elle résulte de la transformation d'une protéine qui perd ses fonctions initiales dans le transposon au seul bénéfice de l'hôte sans pour autant acquérir de nouvelles activités. Le gène devient alors un gène cellulaire et se spécialise au cours de son évolution (Capy, 2021). Ces transposases à ADN domestiquées sont aujourd'hui nécessaires à la formation du système immunitaire adaptatif chez les vertébrés (Jones and Gellert, 2004; Trancoso et al., 2020). De nombreux parallèles existent d'ailleurs entre ce mécanisme de réarrangement qui se produit lors de la différenciation des lymphocytes et les RPG qui ont lieu au cours du développement du noyau somatique de la paramécie (Bétermier et al., 2020).

Les transposons peuvent par ailleurs devenir nécessaires aux mécanismes des RPG qui permettent la maturation du génome de la lignée germinale vers un génome somatique (Alzohairy et al., 2013; Capy, 2021; Cosby et al., 2019). D'anciens éléments transposables ADN ont également été identifiés comme participant aujourd'hui aux voies de réparation de l'ADN (Alzohairy et al., 2013).

2.1.3.2 Les RPG un mécanisme de défense source d'évolution

Les RPG eux-mêmes peuvent avoir évolué vers un rôle bien précis pour être cooptés ou adaptés à d'autres rôles. Par exemple, la machinerie des RPG servant à l'origine à éliminer des éléments transposables peut être dérivée pour optimiser le génome en éliminant l'hétérochromatine. En général, plus le génome somatique est petit, moins il sera énergivore à amplifier lors du développement de l'individu. Supprimer des séquences non-transcrites au lieu de les compacter en hétérochromatine semble être une solution pour réduire la taille du génome sans impacter la transcription (Drotos et al., 2022).

Les RPG peuvent également avoir évolué pour reconstituer un génome somatique fonctionnel. Par exemple, chez certains nématodes, les chromosomes sexuels sont fusionnés dans la lignée germinale assurant qu'ils ségrègent toujours ensemble. Mais pour l'expression dans les lignées somatiques, ces chromosomes sont fragmentés pour reprendre leur forme ancestrale (Drotos et al., 2022).

Certains ciliés, en plus d'éliminer une partie de leur ADN, réordonnent les séquences de leur génome germinale pour former leur génome somatique (Drews et al., 2022b; Drotos et al., 2022; Yerlici and Landweber, 2014).

2.2 LES RÉARRANGEMENTS DU GÉNOME CHEZ *PARAMECIUM TETRAURELIA*

Les ciliés sont des organismes modèles pour l'étude des RPG. Ils sont les seuls unicellulaires réarrangeant entièrement leur génome entre la lignée germinale et la lignée somatique. Cela est dû à leur particularité d'avoir plusieurs noyaux au sein d'une même cellule (même si le nombre de noyaux germinaux peut varier en fonction des espèces). Lors de ma thèse, j'ai étudié une partie du mécanisme des RPG chez *Paramecium tetraurelia*, c'est donc sur cette espèce que je focaliserai cette introduction.

Lors des réarrangements de son génome, *P. tetraurelia* fragmente, au moins en partie, ses chromosomes germinaux. Ainsi, le MAC possède plus de chromosomes que les MIC, mais ceux-ci sont plus petits. En effet, environ 200 chromosomes ont été identifiés dans le MAC alors que le MIC n'en compterait qu'un peu plus de 50 (Aury et al., 2006). De plus, au cours des RPG, elle élimine 30 à 50% de son génome germinale haploïde, présent dans les MIC, pour former son génome somatique, présent dans le MAC.

La formation d'un MAC fonctionnel requiert l'élimination de nombreuses séquences (Figure 9). Au cours des réarrangements, sont éliminés de manière imprécise des régions « MIC-spécifiques » porteuses de nombreuses copies d'éléments transposables et d'autres éléments répétés correspondant à au moins 21-25Mb. Afin de reconstituer le génome, les RPG doivent aussi conduire à l'élimination précise d'environ 45 000 séquences non-codantes, pour la plupart issues de transposons, appelées IES (*Internal Eliminated Sequence*). Les IES représentent ~3,5Mb de séquences qui doivent être éliminées (Arnaiz et al., 2012; Betermier and Duhaucourt, 2014; Guérin et al., 2017). Cette élimination d'ADN a lieu en même temps que l'endoduplication du MAC, la chronologie d'excision peut donc être estimée en fonction de la ploïdie du MAC (Zangarelli et al., 2022).

2.2.1 Les séquences éliminées de manière imprécise : les éléments transposables et les séquences répétées

Les éliminations imprécises d'ADN au cours des RPG concernent principalement des éléments transposables et d'autres séquences répétées (minisatellites).

2.2.1.1 Fragmentation des chromosomes MIC

L'élimination de ces éléments conduit soit à la fragmentation du chromosome MIC soit à une délétion intrachromosomique (Guérin et al., 2017). Ces délétions débutent au cours du 6^{ème} cycle d'endoduplication des MAC en développement (entre ~64C et ~128C) (Zangarelli et al., 2022). L'élimination d'un même élément correspond donc à plusieurs événements d'élimination (un par copie). Puisque leur élimination est imprécise, chaque événement d'élimination donnera un chromosome MAC unique, créant une variabilité entre les chromosomes MAC issus d'un même chromosome MIC (Duret et al., 2008).

L'hétérogénéité des chromosomes MAC ne permet pas de définir de séquences précises pour chaque chromosome. On définit à la place des *scaffolds* qui correspondent à la plus grande séquence consensus qui ait pu être reconstruite (ils peuvent contenir des *gaps* entre les séquences assemblées). Les *scaffolds* représentent donc une population de molécules (Figure 9). Au total, 188 *scaffolds* dont la taille est supérieure à 45 kb ont pu être assemblés. Ils recouvrent 96% des 72Mb aujourd'hui assemblés. Parmi ces *scaffolds*, 115 sont bornés par des télomères suggérant qu'il s'agit de *scaffolds* MAC entiers (Arnaiz et al., 2012). Mais les *scaffolds* sont finalement loin de la réalité des chromosomes MAC individuels puisqu'ils ne sont que des consensus. Lors du séquençage du génome MAC, des lectures télomériques ont pu être repositionnées à l'intérieur des *scaffolds*. Cela suggère qu'un même *scaffold* peut recouvrir un ensemble de chromosomes MAC de tailles variables en raison de l'hétérogénéité de la fragmentation liée à l'élimination imprécise des régions riches en éléments transposables (Duharcourt and Sperling, 2018).

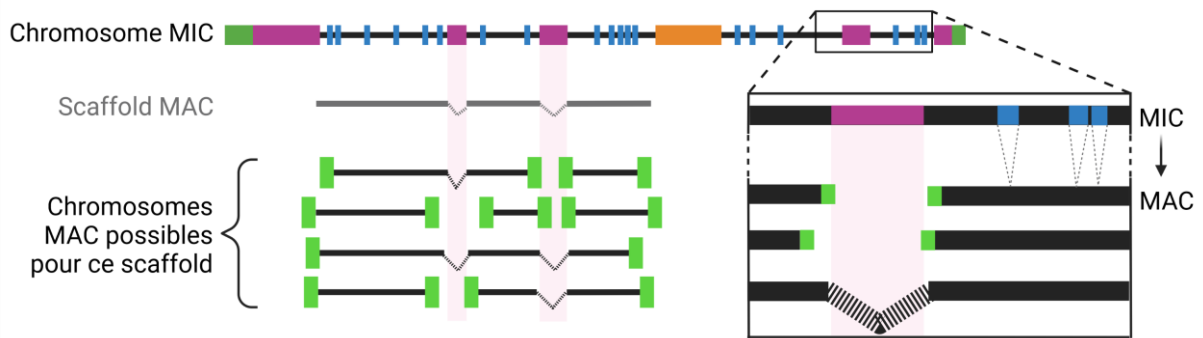


Figure 9 - Formation des chromosomes MAC

Afin de former les chromosomes MAC, les RPG doivent éliminer précisément les IES (*Internal Eliminated Sequences*; en bleu). Les séquences répétées (en rose) et les centromères (en orange) sont également éliminés, mais de manière imprécise, ce qui peut donner lieu à de la fragmentation. Enfin, des télomères (en vert) sont ajoutés aux nouvelles extrémités. La séquence MAC la plus longue ayant pu être reconstruite correspond à un *scaffold* MAC (en gris). (Created with BioRender.com)

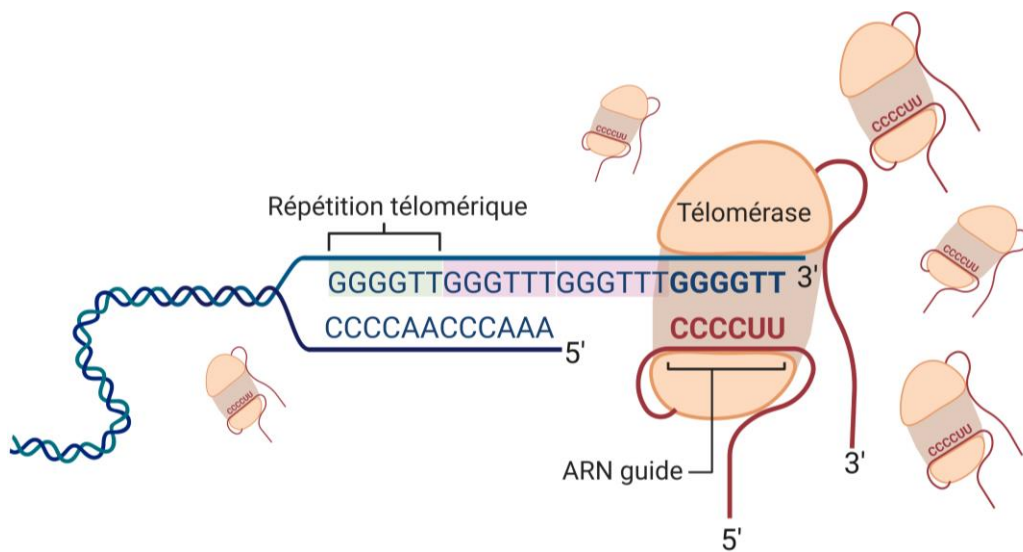


Figure 10 - Allongement des télomères

La télomérase de la paramécie a la particularité d'être active tout au long de la vie de la cellule, mais elle est prompte à introduire des erreurs. Ce sont uniquement ces erreurs qui sont à l'origine de l'hétérogénéité des répétitions télomériques de la paramécie, car celle-ci n'utilise qu'un même ARN guide porteur de la répétition 5'-C₄U₂-3' (McCormick-Graham and Romero, 1996). (Created with BioRender.com)

2.2.1.2 Néo-téломérisation des chromosomes MAC

La fragmentation des chromosomes implique l'addition de télomères aux nouvelles extrémités des chromosomes. Les télomères de la paramécie sont composés d'un mélange de répétitions 5'-C₄A₂-3' et 5'-C₃A₃-3' (Figure 10). Le nombre de répétitions est variable d'un télomère à l'autre. De ce fait, la longueur des télomères est également très variable, oscillant entre 200 et 300 pb (Baroin et al., 1987; McCormick-Graham, 1997). En plus d'un nombre variable de répétitions, les différents chromosomes MAC issus d'un même chromosome MIC peuvent porter des combinaisons différentes de répétitions s'ils sont issus d'événements de fragmentation distincts.

À cela s'ajoute le fait que la composition des télomères peut varier au cours des réplifications végétatives de la cellule, car la télomérase semble active tout au long de la vie de la cellule et qu'elle est prompte à introduire des erreurs (Gilley and Blackburn, 1994).

De plus, la position à laquelle les télomères s'ajoutent peut varier à deux niveaux. D'une part, les télomères peuvent être ajoutés au niveau de sites de télomérisation (TAR, *Telomere Addition Region*) éloignés de plusieurs kb les uns des autres. D'autre part, il existe également une variabilité au niveau de chaque site de télomérisation. En effet, les télomères ne sont pas ajoutés à une position précise, ils peuvent être ajoutés *de novo* dans une région variant de 200 bp à ~1 kb (Betermier and Duharcourt, 2014; Forney and Blackburn, 1988; le Mouël et al., 2003)

Trois niveaux d'hétérogénéité se superposent donc pour former les chromosomes MAC :

- Toutes les copies issues d'un même chromosome MIC ne sont pas fragmentées de la même façon,
- La séquence et le nombre de répétitions télomériques aux extrémités des chromosomes MAC peuvent varier entre les événements de télomérisation,
- Les répétitions télomériques peuvent être ajoutées dans des régions (les TAR) séparées de plusieurs milliers de bases.

2.2.1.3 Élimination imprécise des éléments transposables

2.2.1.3.1 Les séquences éliminées tardivement

La comparaison entre les séquences MIC et MAC a permis de mettre en évidence la présence de séquences ressemblant à des transposons ADN *Tc1/mariner* ou à des répétitions minisatellites éliminées de manière imprécise du MAC (le Mouël et al., 2003). De plus récents assemblages du génome MIC ont permis d'identifier plus précisément la présence d'éléments transposables similaires aux *Tc1/mariner Sardine, Thon* et *Tennessee* dont les TIR (*Terminal Inverted Repeats*) et les ORF (*Open Reading Frame*) sont partiels ou entiers dans le génome (Arnaiz et al., 2012).

Étant donné que l'élimination des éléments transposables est tardive, leur expression est d'abord réprimée grâce à des marques épigénétiques (H3K27me3 et H3K9m3) déposées par le complexe PRC2 (*Polycomb Repressive Complex 2*) de la paramécie (Frapporti et al., 2019). Le dépôt de ces marques serait guidé par des ARN-scan non-codants (ARNscn) (Lhuillier-Akakpo et al., 2014). Plus récemment, l'interaction du complexe PRC2 avec la protéine PIWI, Ptiwi09, capable d'interagir avec les ARNscn semble confirmer l'implication de ces petits ARN dans le dépôt des marques par le complexe PRC2 (Miró-Pina et al., 2022; Wang et al., 2022b). L'élimination des éléments transposables semble dépendre du dépôt de ces marques (Frapporti et al., 2019) (voir détail section 3.1).

Le mécanisme d'élimination imprécise des éléments transposables est actuellement peu décrit. L'endonucléase à l'origine des cassures aux bornes de ces séquences n'est pas encore identifiée avec certitude. La présence de l'endonucléase PiggyMac (Pgm) est nécessaire à leur élimination (Baudry et al., 2009), mais il n'est pas démontré si cette dépendance est liée directement à l'intervention de Pgm ou si elle est une conséquence indirecte du blocage des RPG à une étape en amont (l'excision des IES précède l'élimination des transposons (Zangarelli et al., 2022), voir section suivante). L'incertitude vient également du fait que l'élimination des éléments transposables est imprécise alors que celle des IES est extrêmement précise. Pgm serait-elle capable de couper l'ADN de manière précise ou non selon les contextes ? La réponse à cette question reste encore à déterminer. La différence pourrait également résider dans le mécanisme de

réparation de ces cassures qui pourrait être moins fidèle. On peut également imaginer un mécanisme n'impliquant pas d'endonucléase, où la fragmentation des chromosomes serait due à un arrêt de la réplication au niveau de ces séquences (ou au niveau de toutes séquences hétérochromatiniennes).

2.2.1.3.2 Les séquences éliminées précocement

Jusqu'à l'obtention récente d'un premier assemblage du génome MIC (Guérin et al., 2017), les éléments éliminés étaient identifiés en faisant la différence entre le génome MAC sauvage (WT, *wild-type*) et le génome MAC+IES. Ce dernier ayant été assemblé à partir du génome MAC dans des conditions où l'expression de *PGM*, le gène de la seule endonucléase connue impliquée dans les RPG, est bloquée par un ARNi (Arnaiz et al., 2012). C'est de cette manière que les séquences éliminées ont été identifiées.

Le premier assemblage du génome MIC permet d'estimer la taille du génome germinale entre 98 et 151Mb (Guérin et al., 2017; Sellis et al., 2021). La comparaison entre cet assemblage du génome MIC avec celui obtenu à partir des nouveaux MAC de cellules déplétées pour *PGM* a permis de confirmer qu'une grande partie des éléments transposables était éliminée de manière dépendante de Pgm. Cela a également mis en avant au moins 3Mb de séquences MIC dont l'élimination ne dépend pas de Pgm, mais dépendrait de la présence du complexe PRC2 (Guérin et al., 2017). Cependant, l'endonucléase responsable de l'élimination de ces séquences n'a pas encore été identifiée.

La très récente mise à jour de l'assemblage du génome MIC basé sur du séquençage Oxford Nanopore (Kono and Arakawa, 2019) a permis d'identifier la nature de ces séquences éliminées avant même que l'élimination des IES les plus précoces ne commence. Elles contiendraient principalement des éléments transposables et des minisatellites particuliers (Communication personnelle Éric Meyer, Olivier Arnaiz et Linda Sperling). On peut supposer que l'élimination de ces séquences implique un mécanisme différent de celui des séquences éliminées tardivement de manière imprécise. Il est également possible que les mécanismes d'élimination des éléments transposables précoces et tardifs soient les mêmes et qu'ils soient simplement mis en pause le temps d'éliminer précisément les IES.

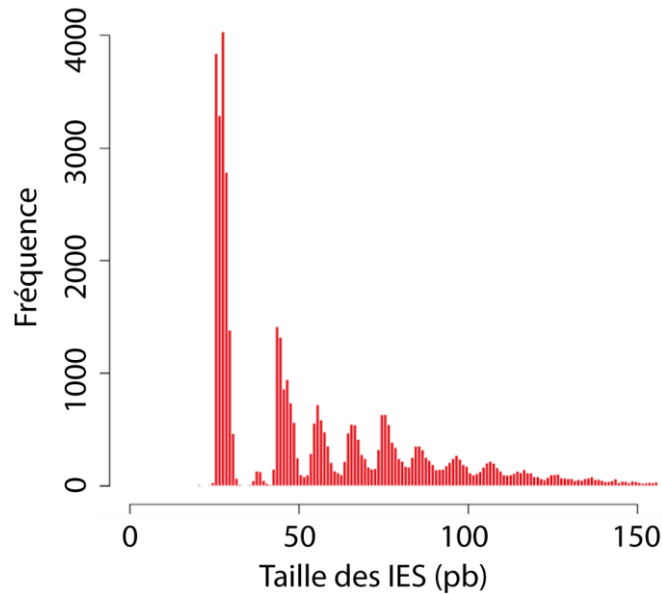


Figure 11 - Histogramme de la répartition des tailles des IES de moins de 150pb

La taille des IES n'est pas répartie de manière homogène, mais présente des pics toutes les 10 pb avec un biais vers le premier pic correspondant aux IES de 26 à 30 pb : les IES les plus petites. (Figure adaptée de Arnaiz et al., 2012)

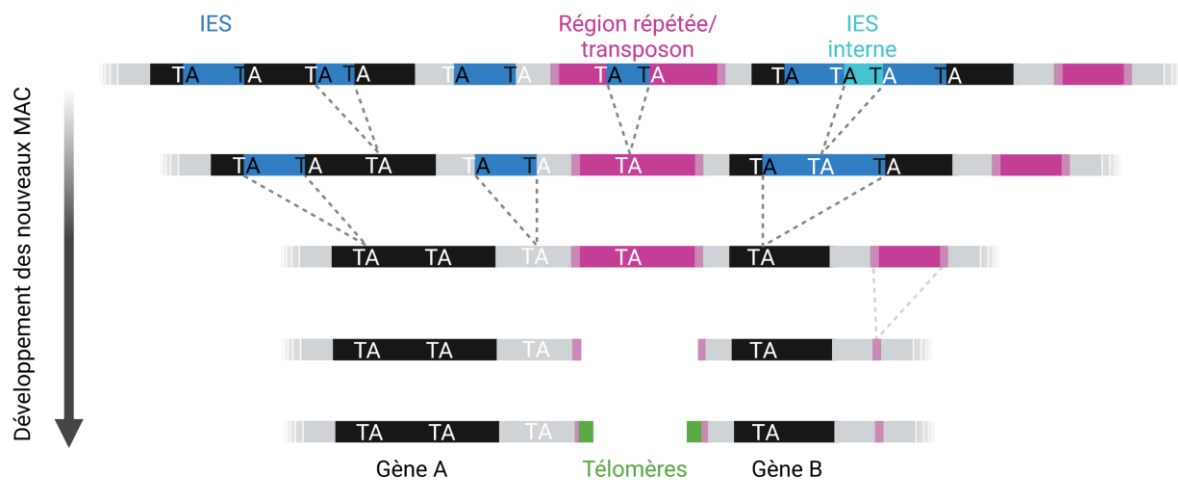


Figure 12 - Élimination d'ADN lors du développement des MAC

Les RPG ont lieu dans les nouveaux MAC au cours des stades DEV1-4 (Figure 8A). Lors de ces RPG, les IES (en bleu) sont progressivement éliminés de manière à reconstituer la séquence des gènes (en noir). Les séquences répétées et les transposons (en rose) sont éliminés de manière imprécise, ce qui donne lieu à la fragmentation du chromosome. Enfin, des télomères (en vert) sont ajoutés sur les extrémités nouvellement formées. Seuls les événements d'élimination dépendants de Pgm sont représentés sur ce schéma. (Created with BioRender.com)

2.2.2 Les séquences précisément éliminées : les IES

En plus des séquences éliminées de manière imprécise, 44 128 séquences « MIC-spécifiques » appelées IES, ont pu être identifiées en comparant la séquence du génome MIC à celle du MAC (Arnaiz et al., 2012; Guérin et al., 2017; Sellis et al., 2021). Chez la paramécie, ces séquences sont éliminées du génome avec une précision chirurgicale. Cette précision est absolument nécessaire, car les IES interrompent au moins une fois la phase codante de 47% des gènes de la paramécie. Elles sont toutefois plus rarement présentes dans les gènes fortement exprimés (Arnaiz et al., 2012).

2.2.2.1 Caractéristiques des IES

2.2.2.1.1 Taille des IES

Les IES correspondent à des séquences dont la taille varie de 26 pb pour les plus petites à quelques milliers de pb pour les plus grandes. Mais la grande majorité (93%) des IES ont une taille inférieure à 150 pb. Leur répartition de taille révèle une périodicité d'environ 10 pb correspondant à un tour d'hélice d'ADN double-brin (Figure 11). Cette périodicité pourrait être liée à la flexibilité du complexe d'excision qui imposerait des contraintes de conformation de l'ADN afin de pouvoir introduire les cassures aux deux bornes des IES (Arnaiz et al., 2012; Bischerour et al., 2018). La distribution des tailles des IES forme un pic entre 26 et 30 pb regroupant 35% des IES identifiées. Cependant, il semblerait que le pic suivant (36-40 bp) soit contre-sélectionné, les IES se répartissent ensuite dans les pics >50 bp avec des effectifs de plus en plus petits (Arnaiz et al., 2012).

2.2.2.1.2 Répartition des IES dans le génome

Les IES sont réparties tout au long du génome. Elles sont à la fois présentes dans les gènes et dans les régions intergéniques, voire dans d'autres IES (Figure 12). Observer la distribution des IES le long des *scaffold* MAC donne l'impression que la distribution des IES n'est pas homogène sur certains *scaffolds* (Arnaiz et al., 2012). En effet, l'alignement des *scaffolds* de plus de 30 kb a permis de montrer que la densité des IES est plus importante au niveau de l'une de leurs extrémités et qu'elle décroît tout au long du *scaffold* (Arnaiz et al., 2012). Il faut cependant

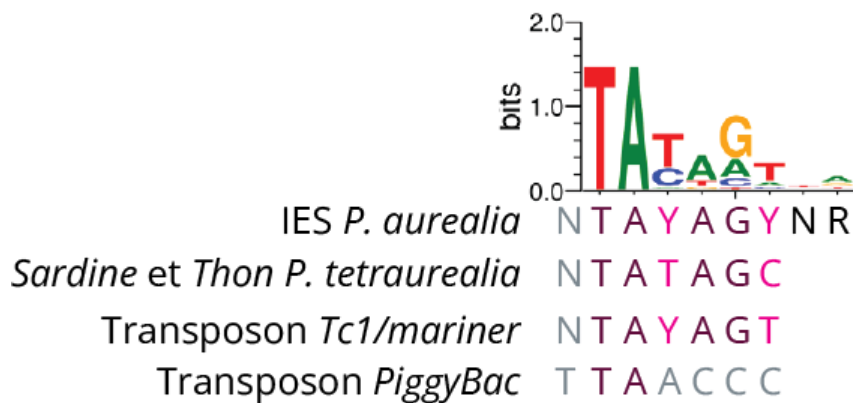


Figure 13 - La séquence consensus aux bornes des IES est similaire à celle des *Tc1/mariner*

Le logo obtenu en alignant les bornes des IES (Arnaiz et al., 2012) est représenté au-dessus de la séquence consensus des bornes des IES (Klobutcher and Herrick, 1995). Cette séquence est alignée sur des bornes de *Tc1/mariner* *Thon* et *Sardine* de *P. tetraurelia*, ainsi que les séquences consensus des transposons *Tc1/mariner* et *PiggyBac* (Dubois et al., 2012). (Y = C ou T ; R = A ou G ; N = A, T, C ou G)

prendre en compte que les chromosomes MAC sont des fragments des chromosomes MIC. La répartition des IES pourrait être liée à la structure des chromosomes MIC. Malheureusement, l'assemblage actuel du MIC ne permet pas encore de répondre à cette hypothèse, mais il a permis de mettre en évidence la présence d'IES dans les régions éliminées de manière imprécise et dans d'autres IES (Sellis et al., 2021; Zangarelli et al., 2022). Cela confirme que les IES semblent bien présentes dans tous types de séquences du génome MIC.

Si les IES sont réparties dans tout le génome, il semblerait que les gènes fortement exprimés comportent moins d'IES que les gènes les moins exprimés. De plus, si l'IES présente dans le gène ne possède pas forcément de codon stop (TGA) en phase avec l'ORF du gène, sa taille est rarement un multiple de trois comparée aux IES présentes dans les régions non-codantes (Arnaiz et al., 2012). Cela peut être un système de sécurité assurant que les gènes ne s'expriment pas tant que l'IES est présente à cause d'un changement de phase de lecture ou d'un codon stop précoce. Les ARN aberrants ainsi formés peuvent ensuite être pris en charge par le *Nonsense-Mediated mRNA Decay* (NMD) et dégradés (Kurosaki et al., 2019).

2.2.2.1.3 Composition nucléotidique des IES

La séquence des IES est significativement enrichie en AT (80%) comparée au reste du génome (72%). La seule constante est que les IES sont bornées de dinucléotides TA et ne laissent après leur excision qu'un TA. Cette particularité avait été mise en exergue avec l'étude des premières IES identifiées (Klobutcher and Herrick, 1995; Steele et al., 1994) et a été confirmée pour l'ensemble des IES (Arnaiz et al., 2012). En plus du dinucléotide TA, Klobutcher et Herrick (Klobutcher and Herrick, 1995) identifient un faible consensus, 5'-TAYAGYNR-3', aux bornes des IES. Ce consensus a par la suite été confirmé avec l'identification de l'ensemble des IES (Arnaiz et al., 2012). La séquence consensus établie par Klobutcher et Herrick rappelle la séquence consensus des extrémités des transposons *Tc1/mariner*, 5'-TACAGTKS-3' (Figure 13). Ceci leur a permis de poser l'hypothèse que les IES dériveraient de transposons intégrés dans le génome.

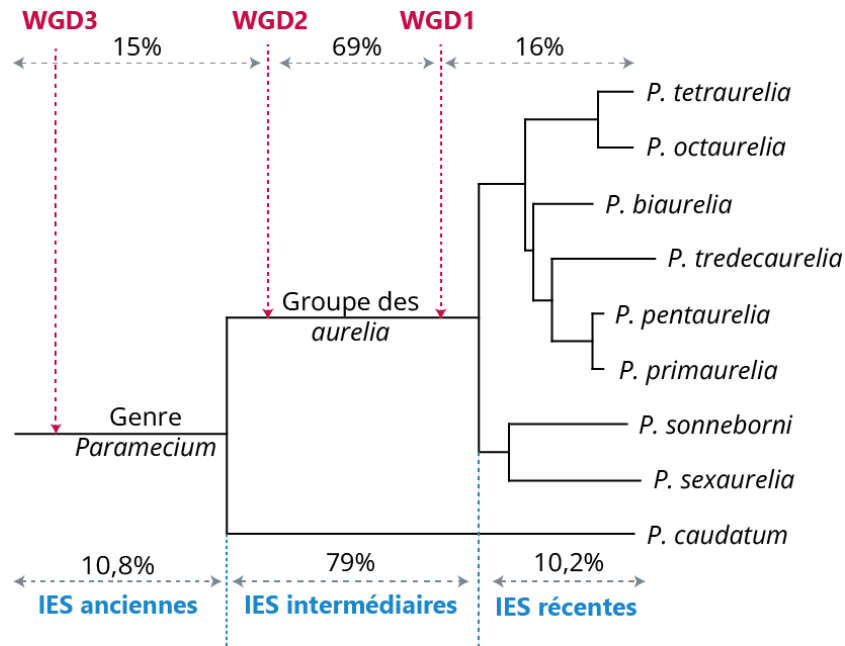


Figure 14 - Acquisitions des IES par *P. tetraurelia* au cours de l'évolution

La proportion d'IES acquises par *P. tetraurelia* entre les WGD est indiquée au-dessus de l'arbre (Arnaiz et al., 2012). De la même manière, sous l'arbre, est indiquée la proportion d'IES anciennes, intermédiaires et récentes définies par rapport à la divergence entre *P. caudatum* et l'ensemble des *aurelia* puis entre les *aurelia*. (Sellis et al., 2021).

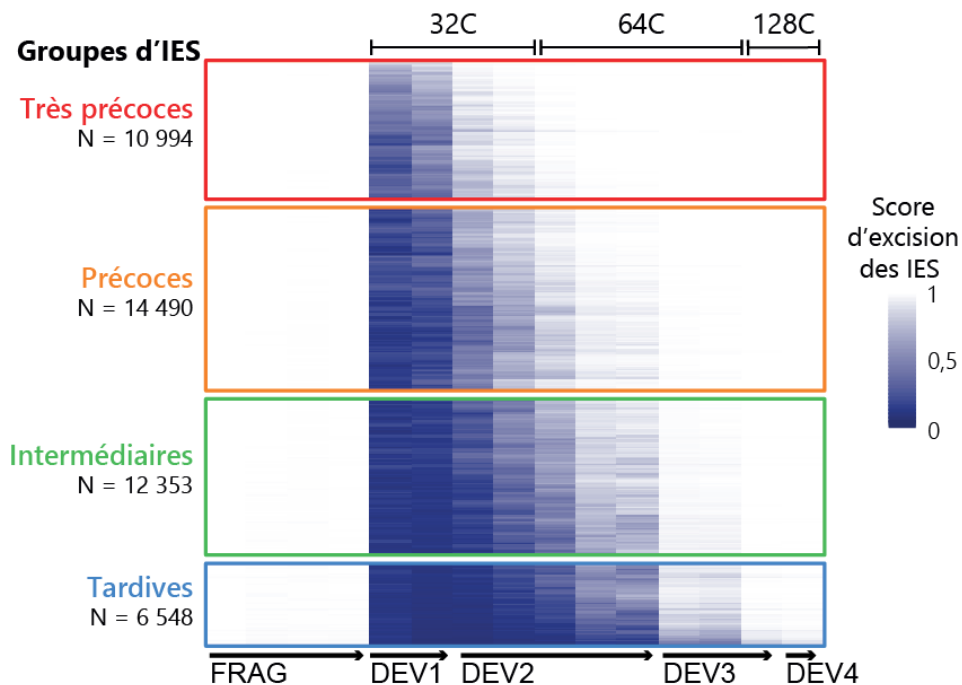


Figure 15 - Cinétique d'excision des IES pendant le développement du MAC

Heatmap représentant le score d'excision des IES (0 = pas excisées; 1 = complètement excisées). Chaque ligne représente une IES. Les IES sont ordonnées par *clustering* hiérarchique et séparées en quatre groupes par un *clustering* en *k-means* selon leur score d'excision. (figure adaptée de Zangarelli et al., 2022)

En-dehors de la faible séquence consensus à leurs bornes, les IES ne semblent pas partager de grandes similitudes de séquence entre elles. La présence d'IES dans les gènes et les différentes WGD qu'a subies le génome permettent d'étudier l'évolution de la séquence d'une IES intégrée dans des gènes ohnologues. Cette étude a montré que l'acquisition ou la perte d'IES dans le temps est assez rapide comparée à celle des introns. En effet, seuls 5,9% des gènes issus de la WGD la plus ancienne ont une IES conservée contre 84,5% pour les gènes issus de la WGD la plus récente. Cependant les IES présentes à un même locus dans des ohnologues ont des séquences très divergentes. La divergence observée est celle attendue dans le cas de séquences non-codantes évoluant sans pression de sélection (Arnaiz et al., 2012). La divergence des IES serait due à l'absence de pression de sélection en dehors de leurs extrémités (Zangarelli et al., 2022). Cela explique également qu'elles ne ressemblent que peu aux transposons qu'elles étaient probablement à l'origine.

2.2.2.2 Origine évolutive des IES

La présence d'un consensus aux bornes des IES proche de celui aux extrémités des éléments transposables *Tc1/mariner* et leur répartition sur tout le génome MIC de la paramécie sont deux arguments en faveur d'invasions du génome par des transposons qui sont ensuite réduits en IES.

2.2.2.2.1 Acquisition des IES

L'évolution des séquences d'IES étant très rapide, il est difficile de les comparer. Une première comparaison a pourtant pu être faite pour quelques IES multicopies présentes à différents *loci* non-homologues dans le génome de *P. tetraurelia*. Elle a mis en évidence des similitudes entre les séquences de certains groupes et celle de l'élément transposable *Tc1/mariner*, *Thon*. Elle a également permis d'identifier une nouvelle famille de transposons *Tc1/mariner* qui sont excisés précisément comme des IES, baptisés *Anchois*, (Arnaiz et al., 2012). Plus récemment, la comparaison des séquences des IES entre toutes les espèces d'*aurelia* a permis de grouper les quelques IES conservées à des *loci* non-homologues en 24 familles contenant au moins 10 copies d'IES. Six de ces groupes sont présents chez *P. tetraurelia*, dont seulement 3 contiennent plus de 10 copies

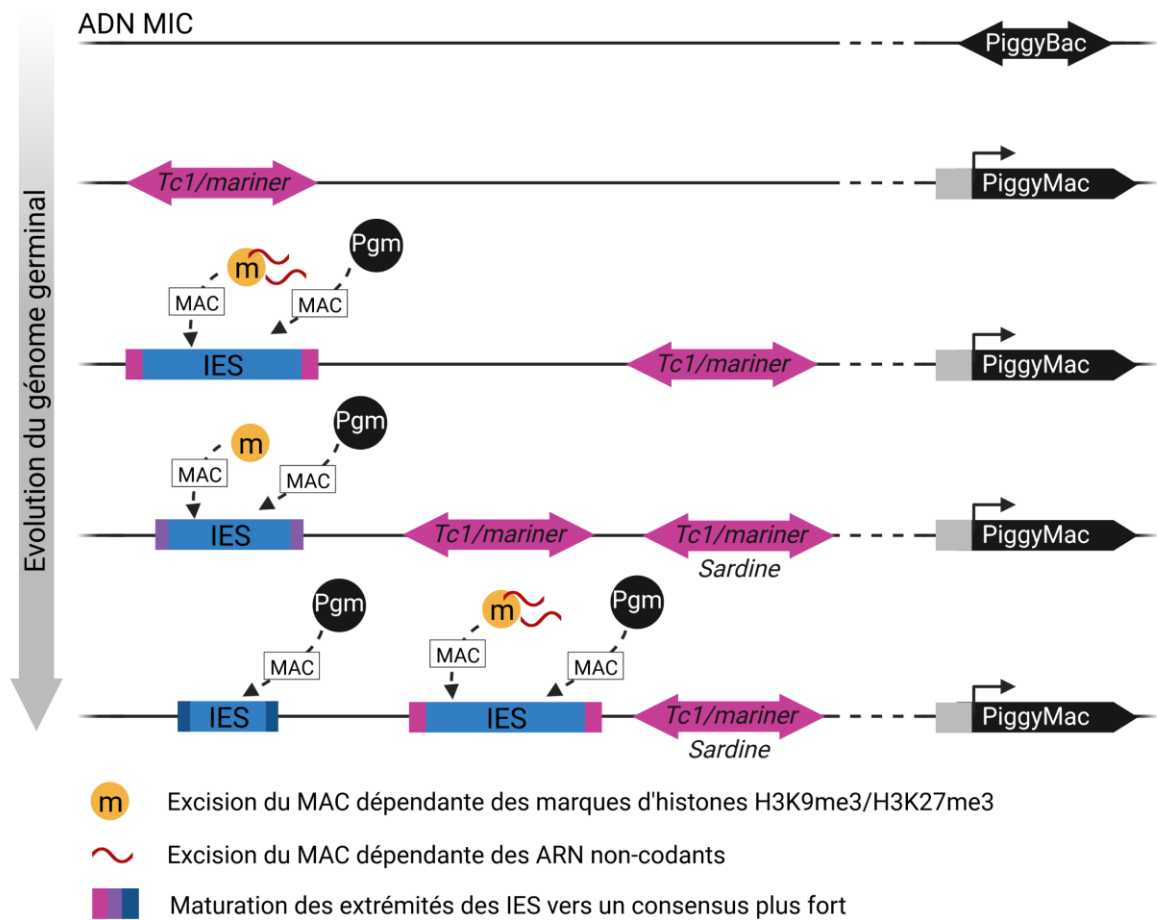


Figure 16 - Évolution des IES

Modèle de la formation des IES et de leur optimisation (Klobutcher and Herrick, 1997; Zangarelli et al., 2022). Les IES sont issues de transposons *Tc1/mariner* qui se sont intégrés dans le génome germlinal (MIC). À la suite de mutations, ils ont dégénéré pour donner des IES toujours capables d'être excisées par PiggyMac du génome somatique. La reconnaissance de ces IES dans le génome somatique (MAC) dépend d'abord de nombreux facteurs épigénétiques (ARN non-codants et marques d'histones). Les IES évoluent ensuite dans le MIC de manière à être de moins en moins dépendantes des facteurs épigénétiques jusqu'à ne plus dépendre que d'une forte séquence consensus à leurs bornes. Cette optimisation de la reconnaissance des IES permet également une diminution du nombre d'erreurs d'excision. (Created with BioRender.com)

d'IES. Parmi ces trois groupes, 2 présentent des similitudes avec des transposons *Tc1/mariner* Thon et Mérou (Sellis et al., 2021). Ces deux études semblent confirmer que les IES proviennent de transposons *Tc1/mariner* qui se seraient intégrés dans le génome germinale des paramécies.

L'étude des IES au sein des familles de gènes ohnologues (Arnaiz et al., 2012) et, plus récemment, une étude évolutive comparant le génome MIC de plusieurs espèces de paramécie (Sellis et al., 2021) ont permis de dater les IES. Dans les deux cas, l'étude suppose que deux IES présentes au même locus dans des gènes ohnologues/homologues résultent d'une insertion précédant la divergence des gènes à la suite des WGD ou de la spéciation.

La première étude se base sur les ohnologues encore présents en quatre exemplaires par suite des deux dernières WGD. Elle a permis de dater les IES en fonction des WGD. Ainsi, on estime que ~16% des IES auraient été acquises après la WGD la plus récente, ~69% entre la WGD la plus récente et l'intermédiaire, et ~15% auraient été acquises avant la première WGD (Arnaiz et al., 2012). Il semblerait donc qu'une majorité d'IES ait été acquise entre la première et la deuxième WGD. Ces résultats se confirment avec l'analyse faite cette fois-ci sur les gènes conservés entre les espèces de paramécie. Cette analyse date ~10% des IES après la divergence des *aurelia*, ~79% après la divergence de *Paramecium caudatum* et ~11% existaient avant cette divergence (Sellis et al., 2021). Combiner ces deux approches permet d'obtenir une idée plus précise de la vitesse d'acquisition des IES (Figure 14).

2.2.2.2 Optimisation des IES

La récente mise au point d'une méthode de tri des ébauches au cours des RPG par cytométrie en flux a permis d'établir une chronologie dans l'excision des IES (Zangarelli et al., 2022). Il a ainsi pu être établi que l'excision des IES a lieu pendant le 5^{ème} cycle d'endoduplication (entre 32C et 64C), soit un cycle d'endoduplication avant l'élimination des éléments transposables et des régions éliminées de manière imprécise. Les IES ont été classées en quatre groupes en fonction de leur chronologie d'excision (très précoce, précoce, intermédiaire et tardive). L'excision de ces séquences semble cependant être un processus continu (aucune rupture n'est observée entre les différents groupes) (Figure 15). Les

groupes sont donc artificiels mais permettent de comparer des IES selon leur temps d'excision.

Cette étude (Zangarelli et al., 2022), combinée à une approche évolutive (Sellis et al., 2021), a permis de généraliser l'idée que les IES acquises le plus récemment sont les plus longues et présentent plus de similitudes avec la séquence des transposons comparées aux IES les plus vieilles. Il semblerait donc qu'avec le temps la taille des IES diminue. Établir la chronologie d'excision des IES a également permis de montrer que les premières IES éliminées ont tendance à être plus courtes et plus anciennes que les IES éliminées tardivement. De plus, l'excision des séquences les plus précoces semble dépendre de moins de facteurs épigénétiques que les IES plus tardives. Enfin, les IES les plus précocement éliminées possèdent une séquence consensus à leurs bornes plus forte que les IES éliminées tardivement (même petites) (Figure 16). En effet, les IES éliminées tardivement ne possèdent à leurs bornes que le TA caractéristique des IES tandis que les IES éliminées plus précocement possèdent une séquence consensus qui dépend de la taille de l'IES (TATAG si elles font entre 25 et 33 pb et TACAG si elles font entre 42 et 140 pb) (Zangarelli et al., 2022).

Il semblerait donc que les IES subissent une optimisation au cours du temps : comparées aux IES récemment intégrées dans le génome, les IES plus anciennes tendent à être plus petites. Elles possèdent des séquences consensus plus fortes à leurs bornes, leur élimination dépend de moins de facteurs épigénétiques et a lieu plus tôt au cours des RPG.

2.2.2.3 Les séquences éliminées ont elles un rôle ?

Une grande partie des IES sont le résultat de l'invasion du génome MIC par des éléments transposables, suivie d'une optimisation de leur élimination afin qu'elle soit plus précoce et plus précise. Chez la paramécie, la présence du mécanisme des RPG a permis le maintien d'intégration d'éléments transposables à l'intérieur des gènes s'ils étaient capables d'être excisés sous forme d'IES. Les IES peuvent être perçues comme des reliques résultant de l'évolution des éléments transposables, toujours maintenues sous la pression de sélection imposée par les RPG. Cependant maintenir et éliminer ses IES présente un coût énergétique pour

la cellule, il semble donc justifié de se demander si les IES peuvent avoir une utilité pour la cellule.

Comme discuté dans la section 2.1.3, il faut garder à l'esprit que les RPG (en particulier les erreurs lors des RPG) sont une source de variabilité et donc d'évolution pour la cellule (Duret et al., 2008). L'autogamie pouvant être déclenchée par une carence de nourriture, on peut également se demander si ce n'est pas une stratégie pour la cellule que d'utiliser les nucléotides éliminés lors des RPG pour la synthèse d'ADN des cycles d'endoduplication suivants comme cela a été proposé chez les copépodes (Drotos et al., 2022). Les RPG peuvent aussi avoir été détournés pour éliminer des séquences qui ne sont pas issues d'invasion du génome. Dans ce cas, on parlera d'exaptation des RPG avec la cooptation d'IES vers une nouvelle fonction (Capy, 2021).

2.2.2.3.1 Un rôle de régulateur

Bien que la cinétique d'excision des IES ait été établie en fonction des cycles d'endoduplication des MAC en développement, les RPG ne sont pas régulés par la réplication elle-même. En effet bloquer la réplication au stade 32C n'empêche pas l'élimination ultérieure des séquences (Zangarelli et al., 2022).

Puisque les RPG ne sont pas régulés en fonction des cycles d'endoduplication, on peut imaginer qu'il existe une régulation interne aux RPG. D'une part, les marques épigénétiques nécessaires à la reconnaissance d'une partie des IES pourraient prendre du temps afin de se mettre en place. D'autre part, l'élimination des IES précoces pourrait activer l'expression de gènes qui seraient ensuite inactivés avec l'élimination des IES tardives. Cette régulation pourrait être liée à la présence de tout ou partie du gène dans une IES ou dans une séquence éliminée de manière imprécise. Ces gènes seraient alors exprimés à partir des MAC en développement jusqu'à ce qu'ils soient éliminés.

Quelques gènes pouvant correspondre à cette description ont déjà été identifiés. En effet, des gènes transcrits dont la séquence codante inclut une IES conservée ont été identifiés chez *Paramecium tetraurelia* (Sellis et al., 2021). Il reste cependant à déterminer s'ils sont bien exprimés à partir des MAC en développement et s'ils jouent un rôle dans les RPG ou leur régulation.

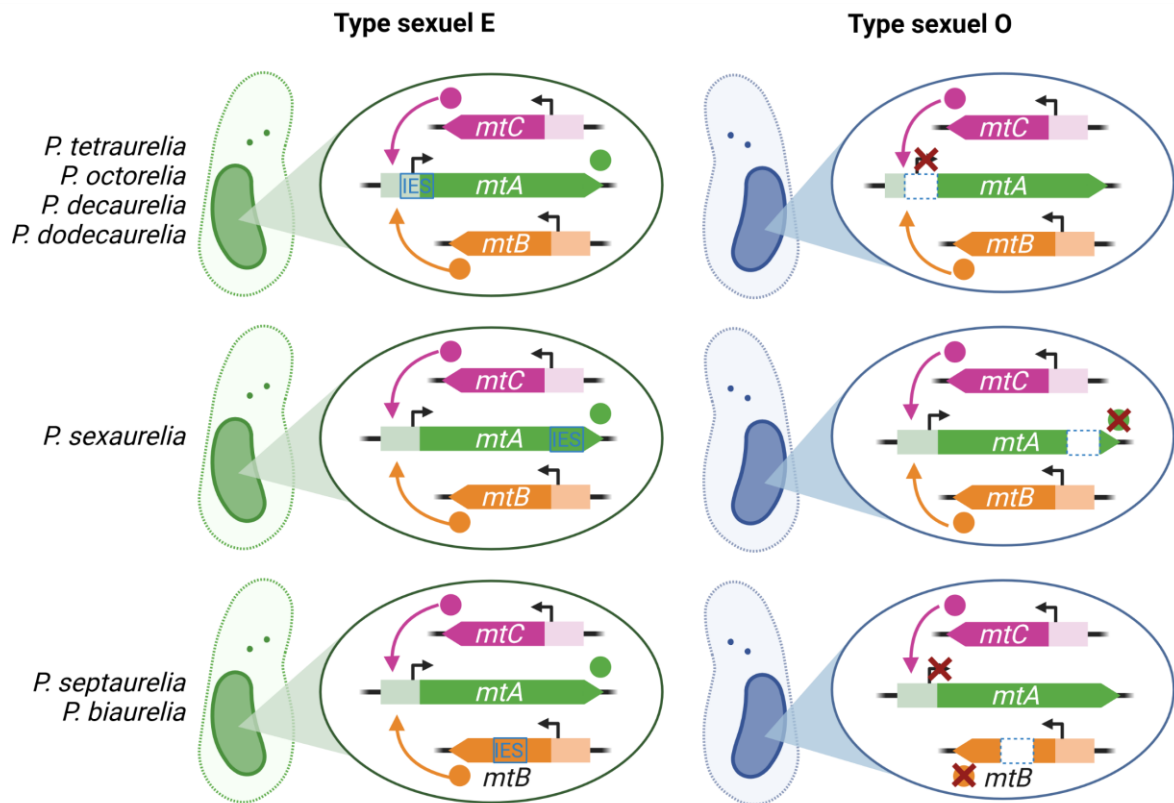


Figure 17 - Détermination du type sexuel chez les *aurelia*

Chez un certain nombre d'*aurelia*, le type sexuel est déterminé en fonction de l'excision ou non d'une séquence génomique cooptée en IES (encadrée en bleu). (Created with BioRender.com)

2.2.2.3.2 Un rôle fonctionnel

Quelques IES ont pu récemment être identifiées comme pouvant faire partie de la séquence codante de gènes (Sellis et al., 2021). Un cas de cooptation (*taming*) des RPG a pu être mis en évidence chez la paramécie (Singh et al., 2014).

Comme décrit dans la section 1.3.2.1, la conjugaison nécessite la présence de paramécies de types sexuels différents. Chez les *aurelia* on distingue deux types sexuels : le type O et le type E (Orias et al., 2017; Sonneborn, 1937). Le type sexuel est déterminé par le gène *mtA* (*mating type A*) exprimé uniquement dans les paramécies de type E.

Chez *Paramecium tetraurelia*, *octaurelia*, *decaurelia* et *dodecaurelia*, le gène *mtA* est interrompu par quatre IES qui sont éliminées au cours des RPG. En plus de ces IES dans sa séquence codante, une séquence de 195 pb contenant le TSS (*Transcription Start Site*) et le début de la séquence codante de *mtA* est éliminée du MAC dans les cellules de type O mais retenue dans celles de type E. Ainsi, le gène *mtA* ne s'exprime que dans les cellules de type E (Singh et al., 2014). La délétion de cette séquence dépend des mêmes facteurs que l'excision des IES. Cela suggère que ces 195 bp correspondent à une séquence MAC qui a évolué pour se comporter comme une IES. Les RPG ont donc été détournés de leur fonction d'origine (éliminer les séquences parasites envahissant le génome MIC) afin d'éliminer une séquence MAC et de différencier deux types sexuels (Figure 17).

De manière similaire, chez *Paramecium sexaurelia*, une partie, cette fois terminale, du gène *mtA* est excisée comme une IES. Chez *Paramecium septaurelia* et *biaurelia*, l'expression du gène *mtA* est contrôlée par l'excision d'une IES présente, non pas dans le promoteur du gène, mais dans la séquence de *mtB* (un des gènes dont dépend l'expression de *mtA*) (Figure 17) (Sawka-Gądek et al., 2021; Singh et al., 2014). Les RPG semblent avoir été cooptés plusieurs fois indépendamment au cours de l'évolution des différentes espèces du groupe des *aurelia* afin de déterminer les types sexuels.

2.2.2.3.3 Un rôle structural

Les centromères de la paramécie ne sont pas encore bien définis, que ce soit ceux des chromosomes MIC. La présence de centromères a toutefois été

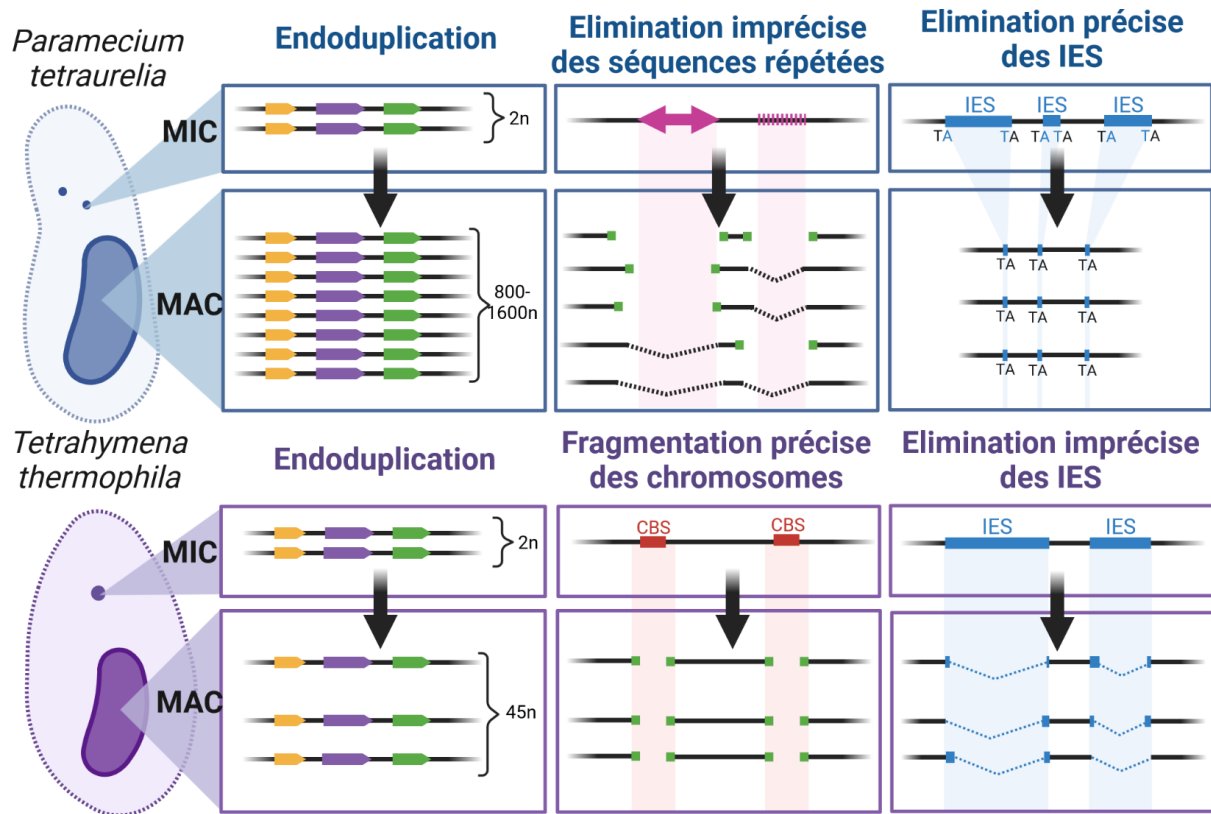


Figure 18 - Comparaison des RPG entre *P. tetraurelia* et *T. thermophila*

Comme tous les ciliés, les paramécies et *Tetrahymena* possèdent un dimorphisme nucléaire avec d'un côté la lignée germinale, représentée par le(s) MIC, et de l'autre la lignée somatique, le MAC. Après chacun de leurs cycles sexuels les deux ciliés détruisent leur MAC parental et en forment de nouveaux à partir de leur(s) MIC. La formation des nouveaux MAC nécessite d'une part des cycles d'endoduplication et d'autre part des RPG. Chez les paramécies, les RPG consistent en l'élimination précise des IES (en bleu) et une élimination imprécise des transposons et des séquences répétées (en rose) qui peut conduire à la fragmentation des chromosomes MIC. Les RPG, chez *Tetrahymena*, conduisent également à l'élimination des IES, mais cette fois de manière imprécise et à la fragmentation des chromosomes MIC. Cette dernière étant déterminée par la présence de CBS (*Chromosome Breaking Site* ; en rouge). (Created with BioRender.com)

confirmée dans les MIC végétatifs par le marquage du variant d'histone H3 centromérique, CenH3 (par un anticorps ou par une fusion GFP) (Lhuillier-Akakpo et al., 2016). Mais aucun centromère n'a pour l'instant été identifié dans les MAC.

Au cours de l'autogamie, CenH3 est visible dans les MIC méiotiques puis dans les noyaux zygotiques. Il disparaît ensuite progressivement au cours du développement des MAC et disparaît complètement 35h après le début de l'autogamie. La disparition du signal CenH3 coïncide donc avec l'élimination imprécise des éléments transposables et autres séquences répétées. De plus, empêcher l'élimination de ces séquences, par un ARNi dirigé contre *PGM* ou *EZL1*, conduit au maintien tardif de CenH3 dans les nouveaux MAC. La perte de CenH3 correspond probablement à l'élimination des séquences associées aux centromères (Lhuillier-Akakpo et al., 2016). Il est donc possible que les régions riches en séquences répétées ou en IES soient utilisées comme centromères dans les chromosomes MIC. Les chromosomes MAC seraient quant à eux dépourvus de centromères.

2.3 DIFFÉRENCE ENTRE LES RPG DE *PARAMECIUM* ET *TETRAHYMENA*

Tetrahymena thermophila, comme la paramécie, est un cilié appartenant au groupe des *oligohymenophorea*, c'est donc un proche parent des paramécies (Figure 4). De nombreuses similitudes existent entre ces deux espèces ce qui rend leur comparaison intéressante (Wang et al., 2017).

Comme tous les ciliés, *Tetrahymena* présente un dimorphisme nucléaire : elle possède un unique MIC et un MAC. Comme pour la paramécie, son MIC est nécessaire pour le maintien du génome au travers des générations sexuelles, mais transcriptionnellement silencieux au cours de la vie végétative de la cellule. C'est le MAC qui est responsable de la transcription, mais il est détruit à chaque cycle sexuel. Afin de former des nouveaux MAC, *Tetrahymena* ne peut effectuer que la conjugaison. Contrairement à la paramécie, elle ne possède pas la capacité de faire une autogamie.

Au cours du développement de son MAC, *Tetrahymena* élimine 30% des séquences germinales. Comme chez la paramécie, cette élimination s'accompagne également d'une fragmentation des chromosomes et de cycles d'endoduplication

(Figure 18). Cependant, *Tetrahymena* effectue moins de cycles d'endoduplication puisque son MAC atteint une ploïdie de « seulement » 45-50n. Au cours des RPG, le MAC parental dégénère par un processus d'autophagie sans se fragmenter (Noto and Mochizuki, 2017; Yao et al., 2014).

2.3.1 Fragmentation des chromosomes

Alors que la fragmentation des chromosomes chez la paramécie est hétérogène, et a lieu en général au niveau des séquences répétées, chez *Tetrahymena* les cassures des chromosomes ont lieu au niveau de ~200 CBS (*Chromosome Breakage Site*) présentant un motif conservé. La fragmentation est suivie de l'élimination d'une vingtaine de paires de bases avant l'addition de télomères permettant de fragmenter les 5 chromosomes MIC en près de 230 chromosomes MAC. Une cinquantaine des produits de la fragmentation est directement dégradée avant même l'addition de télomères ou au cours des premières divisions végétatives de la cellule (Noto and Mochizuki, 2017; Yao et al., 2014).

Les CBS correspondent à une séquence de 15 pb conservée chez les différentes espèces de *Tetrahymena*, mais aucune CBS n'a pu être identifiée chez la paramécie. De plus des techniques de Hi-C (Rao et al., 2014) et de HiChIP (Mumbach et al., 2016) ont permis de mettre au jour une partie de l'organisation 3D des génomes MIC et MAC (Luo et al., 2020). Cette étude montre que les TAD (*Topological Associated Domain*) identifiés dans le MIC correspondent globalement aux chromosomes MAC, les CBS se trouvant à la frontière entre les TAD. Ceci suggère que l'organisation 3D du génome pourrait jouer un rôle lors du développement des noyaux somatiques. Il n'est cependant pas exclu que cette observation soit due à une contamination lors de l'extraction des noyaux MIC par les noyaux MAC.

2.3.2 Élimination des IES

De la même manière que la paramécie, *Tetrahymena* élimine des IES de son génome. Ces IES sont également des séquences dérivées de transposons *Tc1/mariner*, mais aussi des hélitrons et des transposons de la famille *Maverick/Tlr* (Hamilton et al., 2016). Cependant dans le cas de *Tetrahymena*, ces quelques

12 000 IES sont éliminées pour la grande majorité de manière imprécise. Les IES de *Tetrahymena* sont plus grandes que celles de la paramécie avec une taille variant de 136 pb à 43,4 kb avec une moyenne autour de 2,8 kb. Enfin, la grande partie de ces IES se trouve dans des séquences non-codantes (Noto and Mochizuki, 2017; Yao et al., 2014). Il existe cependant quelques exceptions, en effet, 12 IES sont précisément excisées chez *Tetrahymena*. Ces IES ont la particularité de se trouver dans la séquence codante de gènes (Wang et al., 2017).

Ces différences pourraient expliquer la divergence dans la précision d'excision des IES entre les deux genres. Avec d'une part des IES pouvant interrompre les séquences codantes qui doivent être éliminés très précisément pour maintenir l'intégrité du génome et dont l'excision est optimisée avec le temps (voir section 2.2.2.2) (Sellis et al., 2021; Zangarelli et al., 2022). D'autre part, des transposons et des IES de grande taille éliminées imprécisément mais se trouvant en dehors des séquences codantes (Noto and Mochizuki, 2017; Yao et al., 2014).

Il paraît également important de mettre en avant la différence de caractéristiques des IES entre ces deux ciliés. En effet, la définition des IES chez *Tetrahymena* recouvre toutes formes de séquences internes éliminées (précisément ou non) tandis que, chez la paramécie, cette définition est restreinte aux séquences internes éliminées de façon précise. Chez la paramécie, les séquences éliminées de façon imprécise et donnant lieu à des délétions internes ne sont pas comptées parmi les IES, notamment car leur élimination peut aussi donner lieu à de la fragmentation des chromosomes.

Chapitre 3 - Machinerie d'élimination programmée d'ADN

L'étude de l'élimination programmée d'ADN chez la paramécie a permis de mettre au jour une partie de son mécanisme, autant pour l'identification des séquences à éliminer que pour la description de la machinerie-cœur réalisant l'élimination.

De nombreux acteurs impliqués directement ou non dans les RPG ont déjà été identifiés. Ces gènes font partie des gènes différentiellement exprimés lors de l'autogamie : leur expression varie en fonction de l'avancement de l'autogamie (voir détails section 1.3.3 et Tableau 23 en annexe). Les gènes impliqués dans la reconnaissance des séquences à éliminer sont généralement des gènes du pic précoce (*early peak*, spécifiquement induits lors de la méiose) tandis que les gènes de la machinerie-cœur font plutôt partie des gènes du pic intermédiaire (*intermediate peak*, spécifiquement exprimés lors des RPG) (Tableau 25 en annexe).

De manière plus surprenante, un certain nombre des acteurs impliqués dans la réparation finale des cassures de l'ADN générées lors des RPG ont leur pic d'expression lors de la méiose (expression *early peak*) (Tableau 26 en annexe). Les profils ayant été établis sur des données de séquençage d'ARN, cela ne traduit pas directement le profil d'expression de la protéine (Arnaiz et al., 2017). Cela a été confirmé pour les protéines Ligase IV (Kapusta et al., 2011) mais également pour les Cernunos, homologues de XLF (*Xrcc4 Like Factor*) : Cera et Cerb (données non-publiées de Émeline Ostyn, Nicolas Méthais et Laura Le Brun). On peut y voir un système de sécurité pour la cellule qui exprime les éléments nécessaires à la réparation des cassures avant d'exprimer les acteurs permettant de les introduire.

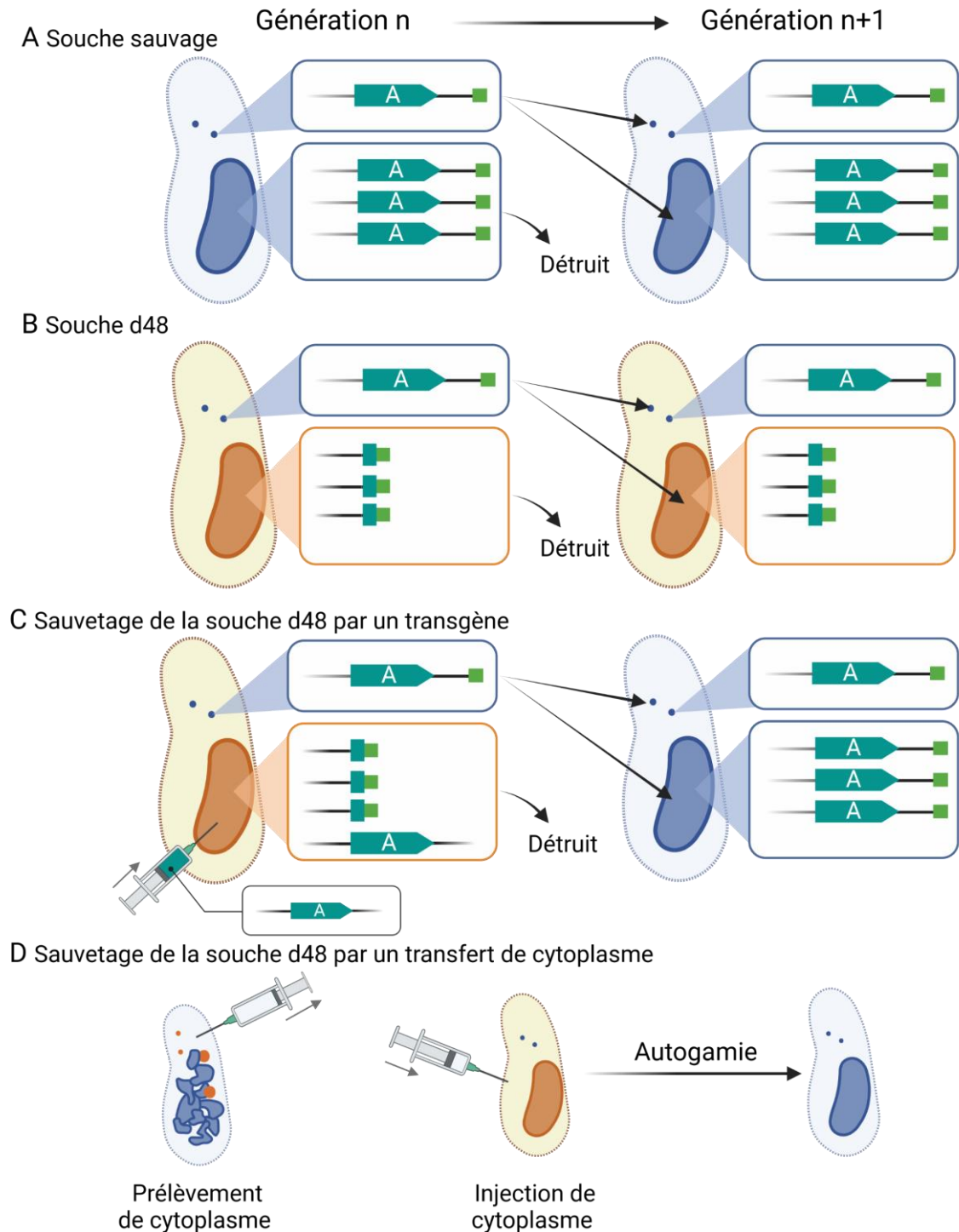


Figure 19 - Sauvetage de la délétion MAC de l'antigène A

A Héritéité de l'antigène A dans une souche sauvage et B dans la souche d48. C La transformation d'une d48 avec un transgène porteur de l'antigène A permet de restaurer l'amplification de l'antigène A dans sa descendance. D Le transfert de cytoplasme d'une souche sauvage autogame dans une d48 permet de restaurer l'expression de l'antigène A dans sa descendance. (Les télomères sont représentés par des carrés vert clair). (Created with BioRender.com)

3.1 RECONNAISSANCE DES SÉQUENCES À ÉLIMINER

Les RPG de la paramécie se traduisent principalement par l'élimination précise des IES et l'élimination imprécise des séquences répétées (voir section 2.2). Puisqu'aucune séquence consensus forte ne permet de discriminer toutes les séquences à éliminer, la reconnaissance de ces éléments passe principalement par des marques épigénétiques (Drews et al., 2022b). Les mécanismes impliqués dans la reconnaissance des IES sont conservés entre la paramécie et *Tetrahymena* (Noto and Mochizuki, 2017; Wang et al., 2017; Zhao et al., 2019).

3.1.1 Le contrôle maternel de l'élimination d'ADN

L'existence d'un contrôle maternel de l'élimination programmée d'ADN a été mise en évidence en partant de l'observation de l'hérédité non-mendélienne de certains caractères.

3.1.1.1 Exemples d'élimination d'ADN sous contrôle maternel

3.1.1.1.1 La délétion MAC du gène de l'antigène de surface A

Le premier cas à avoir été décrit au niveau moléculaire est celui de l'hérédité de l'*antigène A* grâce au variant d48 (Forney et al., 1983) (Figure 19B). Cette souche ne peut exprimer l'antigène A à cause d'une délétion d'une partie du chromosome MAC alors même que le chromosome MIC est intact (Forney and Blackburn, 1988), suggérant une hérédité maternelle plutôt que mendélienne.

Cette hypothèse a ensuite été vérifiée, d'abord en injectant, dans le MAC du variant d48, un transgène porteur de la version sauvage du gène. Le transgène est maintenu dans le MAC comme n'importe quel chromosome et suffit à restaurer de manière stable l'expression de l'antigène A dans les générations suivantes (Godiska et al., 1987). Cette première observation suggère une influence du MAC parental sur le développement des nouveaux MAC (Figure 19D).

D'autre part, l'injection de cytoplasme d'une cellule sauvage autogame dans le cytoplasme du variant d48 permet également de restaurer stablement

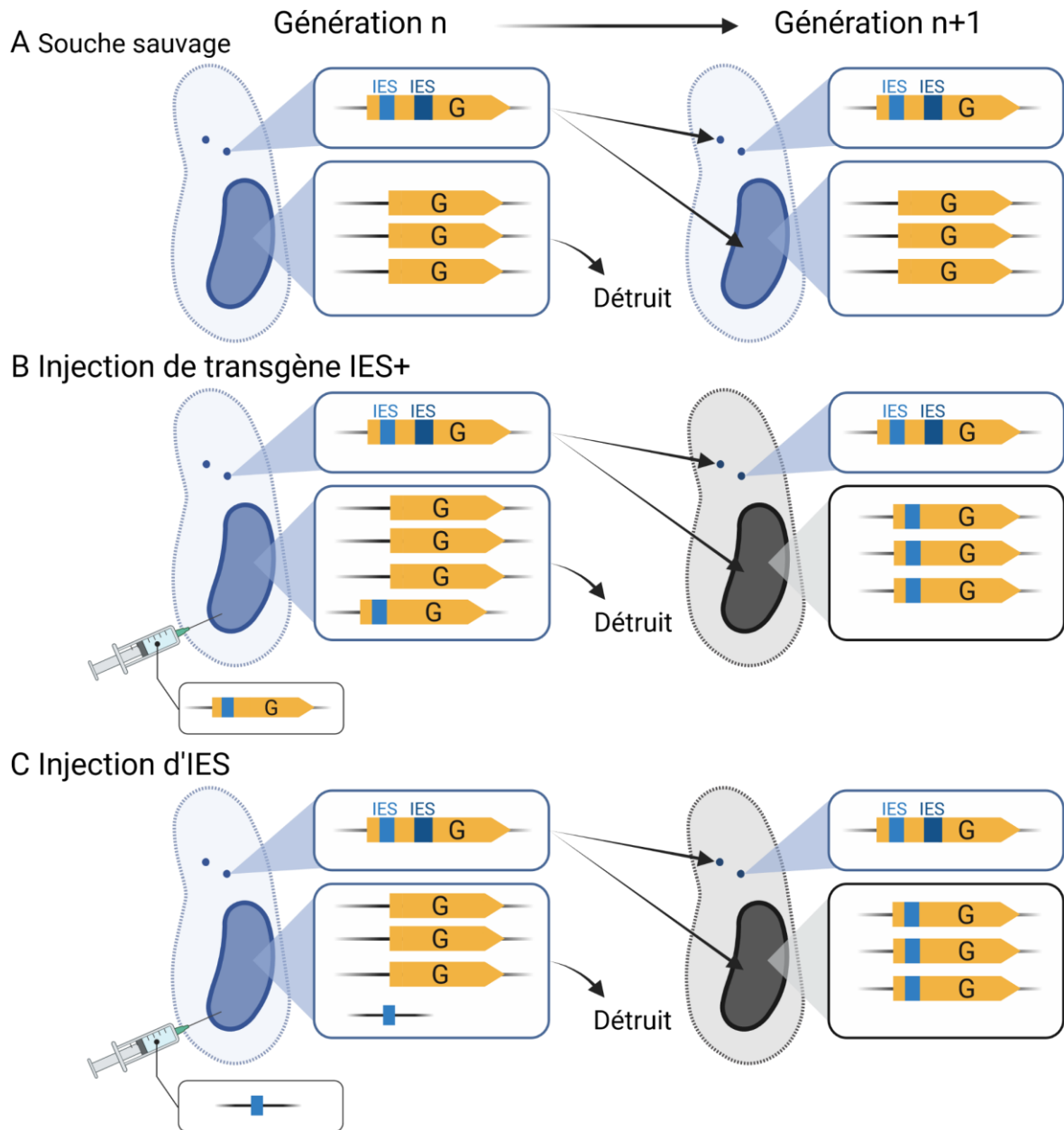


Figure 20 - Induction de la rétention d'une IES sous contrôle maternel

A Au cours du développement du MAC, les IES (en bleu) présentes dans l'antigène G sont excisées de manière efficace dans des conditions sauvages. **B** L'injection, dans le MAC, d'un transgène porteur de la séquence de l'antigène G avec une IES peut inhiber l'excision de cette IES dans le MAC de la descendance. **C** De la même manière, l'injection d'un transgène simplement porteur de l'IES permet d'induire sa rétention dans le MAC de la descendance. (Created with BioRender.com)

l'expression de l'antigène A dans la descendance (Koizumi and Kobayashi, 1989) (Figure 19D). Cette seconde observation appuie l'hypothèse de l'implication d'un facteur cytoplasmique dans l'hérédité maternelle de la délétion MAC.

3.1.1.1.2 L'excision des IES sous contrôle maternel

L'injection dans le MAC de la paramécie ou de *Tetrahymena* d'un transgène porteur de la séquence de certaines IES a permis de mettre en évidence une hérédité maternelle pour la rétention ou l'excision de certaines IES appelées IES sous contrôle maternel (IES_{mc}, *maternally controlled IES*) (Chalker and Yao, 1996; Duharcourt et al., 1995; Duharcourt et al., 1998). L'hérédité de la rétention de ces IES semble dépendre du contenu du MAC parental, toujours présent lors de la formation des nouveaux MAC (Figure 20).

La notion d'IES_{mc} a permis notamment d'expliquer l'hérédité non pas mendélienne mais maternelle du type sexuel (Nanney, 1957; Sonneborn, 1937). En effet, comme présenté dans la section 2.2.2.3.2, le type sexuel de la paramécie est déterminé par l'expression ou non du gène *mtA*. Son expression n'est possible que si l'IES portant une partie de son promoteur et de sa séquence codante est retenue (Singh et al., 2014). Il s'avère que cette IES est sous contrôle maternel, elle n'est donc retenue que si elle est présente dans le MAC parental. C'est donc le MAC du parent qui détermine le type sexuel de la descendance.

Par la suite, des travaux ont permis de montrer que l'excision de ces IES_{mc} ainsi que l'élimination de nombreux transposons étaient dépendantes de la présence de deux protéines ohnologues de liaison à l'ARN : Nowa1 et Nowa2 (Nowacki et al., 2005). Bien que cela reste à démontrer à l'échelle du génome, par extension toutes les IES dépendantes des protéines Nowa1/2 ont été considérées comme étant sous contrôle maternel (Betermier and Duharcourt, 2014; Nowacki et al., 2005).

En plus d'avoir un effet sur l'excision des IES, des fusions Nowa1-GFP ont permis de montrer que, durant l'autogamie, Nowa1 se localise d'abord dans les fragments de l'ancien MAC avant de passer progressivement dans les MAC en développement (Nowacki et al., 2005). Cette observation, additionnée aux effets de la composition de l'ancien MAC sur l'élimination d'ADN dans les nouveaux MAC,

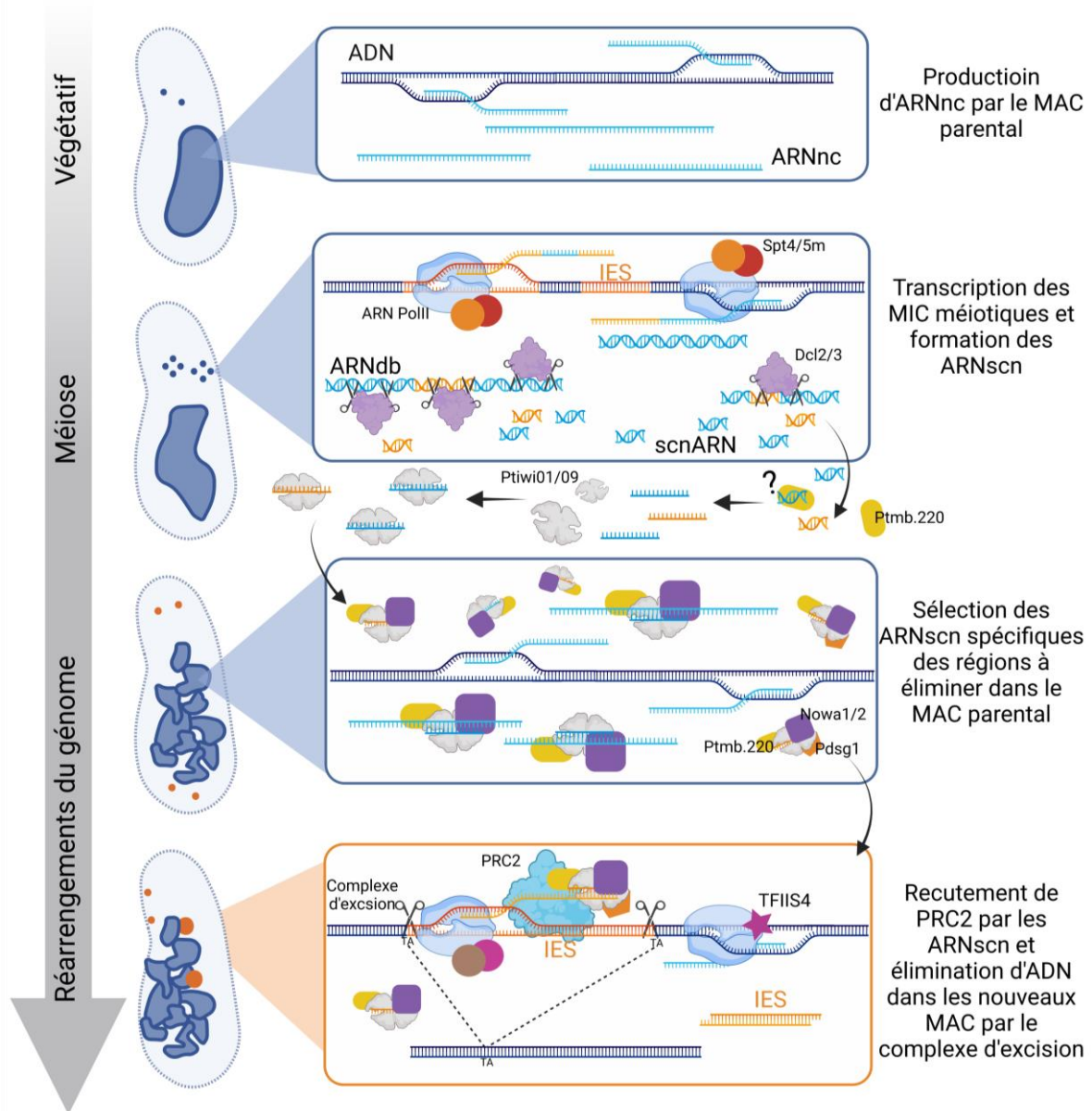


Figure 21 - Modèle du scan de génome par les ARNscn chez *P. tetraurelia*

Des ARN non-codants (ARNnc) sont formés de manière constitutive dans les MAC. Lorsque les MIC entrent en méiose, le génome germinal est transcrit dans les deux sens permettant la formation d'ARN double-brin (ARNdb) pris en charge par des Dicer-like (Dcl2/3) pour former les ARN scan (ARNscn). Ceux-ci sont ensuite transportés sous forme simple brin vers le MAC parental par des Piwi (Ptiwi01/09). Les Ptiwi s'associent alors à Nowa1/2 (et Pdsg1) et les ARNscn qui s'apparient sur les ARNnc MAC sont dégradés. L'hélicase Ptmb.220 pourrait séparer les deux brins des ARNscn pour qu'ils puissent s'associer aux Ptiwi et/ou former un complexe avec les Nowa et résoudre les éventuelles structures secondaires des ARNnc facilitant ainsi l'appariement des ARNscn. Les complexes restants, spécifiques des séquences à éliminer, sont ensuite transportés vers les nouveaux MAC où ils s'apparient sur les ARN naissants correspondant aux séquences à éliminer et y recrutent PRC2 (*Polycom Repressive Complex 2*). Par la suite, le complexe d'excision se fixe sur ces cibles pour les exciser. (Created with BioRender.com)

a permis de poser l'hypothèse qu'il existe une comparaison globale entre le génome germinale et le génome somatique parental identifiant des séquences à éliminer. Puisque les protéines Nowa sont des protéines de liaison à l'ARN, la comparaison entre les deux génomes pourrait faire intervenir des ARN.

3.1.1.2 Le modèle du scan du génome par ARN non-codants

Les processus de comparaison entre les génomes ont été largement étudiés au cours des vingt dernières années et de nombreux acteurs ont été identifiés ce qui a permis d'établir un modèle assez complet (Figure 21). Ce modèle s'est révélé être vrai à la fois pour la paramécie et pour *Tetrahymena* (Coyne et al., 2012; Drews et al., 2022b). Une des avancées majeures ayant permis de mettre en place ce modèle est la découverte des ARN scans (ARNscn) d'abord chez *Tetrahymena* (Mochizuki et al., 2002) puis chez la paramécie (Lepère et al., 2008; Nowacki et al., 2005).

Au cours de la vie de la cellule, le MIC est transcriptionnellement inactif. La seule exception se situe au moment de la méiose. Au cours de la méiose, les deux brins des chromosomes du MIC sont entièrement transcrits formant des ARN double-brin (ARNdb) couvrant l'intégralité du génome MIC. Cette transcription est effectuée par l'ARN-polymérase II associée aux facteurs d'élongation méiotiques Spt4/Spt5m (Gruchota et al., 2017; Owsian et al., 2022). Ces longs ARNdb sont ensuite coupés par les protéines Dicer-like, Dcl2 et Dcl3, en petit ARNdb de ~25 pb pour former les ARNscn (Lepère et al., 2009; Sandoval et al., 2014).

Les ARNscn sont ensuite séparés en ARN simple brin et pris en charge par les ohnologues Ptiwi01 et Ptiwi09, des protéines spécifiques de la méiose appartenant à la famille des protéines Piwi (Bouhouche et al., 2011). Les ARNscn sont ainsi transportés jusqu'au MAC parental. Les protéines de liaison à l'ARN Nowa1 et 2 essentielles à l'excision des IES dépendantes des ARNscn (IESscn) et des transposons étant localisées dans le MAC parental (Nowacki et al., 2005; Swart et al., 2017), on suppose aussi qu'elles participent à l'étape de comparaison entre les transcrits germinaux et somatiques, mais cela reste à démontrer. A ce complexe s'ajoute également l'hélicase putative Ptmb.220 (Nowak et al., 2011). Ptmb.220 pourrait faciliter l'interaction entre les Ptiwi et les ARNscn en séparant les deux brins d'ARN. Elle pourrait également interagir avec les transcrits naissants

dans le nouveau ou l'ancien MAC afin de permettre l'association des ARNscn (Figure 21). Les ARNscn sont comparés à des transcrits produits à partir du MAC parental (Lepère et al., 2008). Cette étape fait également intervenir *Pdsg1* (*Paramecium protein involved in Development of the Somatic Genome 1*) qui possède une localisation similaire à celle de *Nowa1*, mais qui, de manière surprenante, n'est nécessaire que pour l'excision d'une partie des IES dépendantes des scans mais pas des transposons. Sur les quelques IES testées, on remarque que seules les IESmc sont touchées. De plus, plus l'IESmc est grande plus elle est affectée par un ARNi contre *PDSG1* (Arambasic et al., 2014). Cet effet pourrait être dû à un gradient de dépendance des IESmc aux ARNscn, à leur taille ou leur chronologie d'excision.

La production des longs ARN non-codants (ARNnc) à partir du MAC parental dépendrait des facteurs d'élongation *Spt4vA* et *Spt5v* également présents dans les fragments lors du développement (Gruchota et al., 2017; Owsian et al., 2022). Les ARNscn trouvant une complémentarité sur les ARNnc du MAC parental sont dégradés. Le MAC parental étant une version réarrangée du MIC, seuls les ARNscn spécifiques des séquences à éliminer sont conservés.

Les protéines de fusion GFP avec *Nowa1*, *Ptiwi01/09*, *Ptmb.220* et *Spt4/5*, transcrites à partir de l'ancien MAC, se localisent dans les nouveaux MAC (Gruchota et al., 2017; Nowacki et al., 2005; Nowak et al., 2011; Owsian et al., 2022). On suppose que les complexes protéiques transportent les ARNscn restés intacts vers les nouveaux MAC. Au cours de leur développement, les nouveaux MAC produisent des ARNnc dépendant d'un régulateur de l'ARN polymérase II : *TFIS4* (Maliszewska-Olejniczak et al., 2015). Les ARNscn s'apparient sur ces transcrits naissants au niveau des séquences à éliminer ou directement sur les séquences afin de recruter localement la machinerie d'excision par l'intermédiaire de *PRC2* (Miró-Pina et al., 2022; Swart et al., 2017; Wang et al., 2022b).

Le modèle des ARNscn concorde avec l'hypothèse qu'un composant cytoplasmique soit impliqué dans l'hérédité observée de la délétion dans l'antigène A (voir section 3.1.1.1). Le composant dans le cytoplasme des cellules autogames sauvages ayant permis de restaurer l'intégrité du gène de l'antigène A dans le MAC du variant d48 (Koizumi and Kobayashi, 1989) devait très probablement contenir des ARNnc issus des fragments de l'ancien MAC sauvage.

Ces ARNnc n'étant pas porteurs de la délétion sur l'*antigène A*, les ARNscn provenant des MIC du variant d48 se sont appariés sur toute la longueur du gène. De cette façon les ARNscn correspondant à la portion délétée de l'*antigène A* dans d48 sont détruits et l'expression de l'*antigène A* est restaurée dans la descendance.

3.1.1.3 Les IES dépendantes des ARNscn (IESscn)

Le séquençage de noyaux après un ARNi contre *DCL2/3* a permis d'identifier précisément les IES dépendantes des ARNscn. Le taux de rétention des IES dans cet ARNi est globalement assez faible, mais 4 972 IES montrent au moins 5% de rétention (Lhuillier-Akakpo et al., 2014; Sandoval et al., 2014). Ces IES dépendantes des ARNscn ont été confirmées par le séquençage de noyaux après un ARNi ciblant *PTIWI01/09* (Furrer et al., 2017).

Sur la base de ces résultats, le modèle des ARNscn permet d'expliquer l'excision d'uniquement ~11% des IES. Cependant, un ARNi contre *TFIS4* induit la rétention de ~48% des IES incluant 97% des IES sensibles à l'ARNi contre *DCL2/3* (Maliszewska-Olejniczak et al., 2015). Cela suggère que d'autres ARN non-codants pourraient permettre la reconnaissance des IES n'étant pas sous contrôle maternel. Il est également possible que l'efficacité des ARNi contre *DCL2/3* ou des *PTIWI01/09* n'ait été que partielle, laissant la possibilité à certaines IES d'être reconnues et excisées dans ces conditions.

3.1.2 Une boucle de rétrocontrôle ARN pour assurer une excision complète des IES

Rappelons que les IES interrompent près de la moitié des gènes de la paramécie, leur rétention a donc un effet négatif sur la survie des paramécies (Arnaiz et al., 2012). De plus, l'excision des IES se fait alors que l'endoduplication des nouveaux MAC a déjà commencé (voir détail section 2.2.2). Ce n'est donc pas 2 mais 32 à 64 copies des quelques 45 000 IES qui doivent être précisément éliminées (Zangarelli et al., 2022). Afin d'assurer l'élimination complète des IES la paramécie possède une boucle de rétrocontrôle positif (Figure 22) (Allen and Nowacki, 2017).

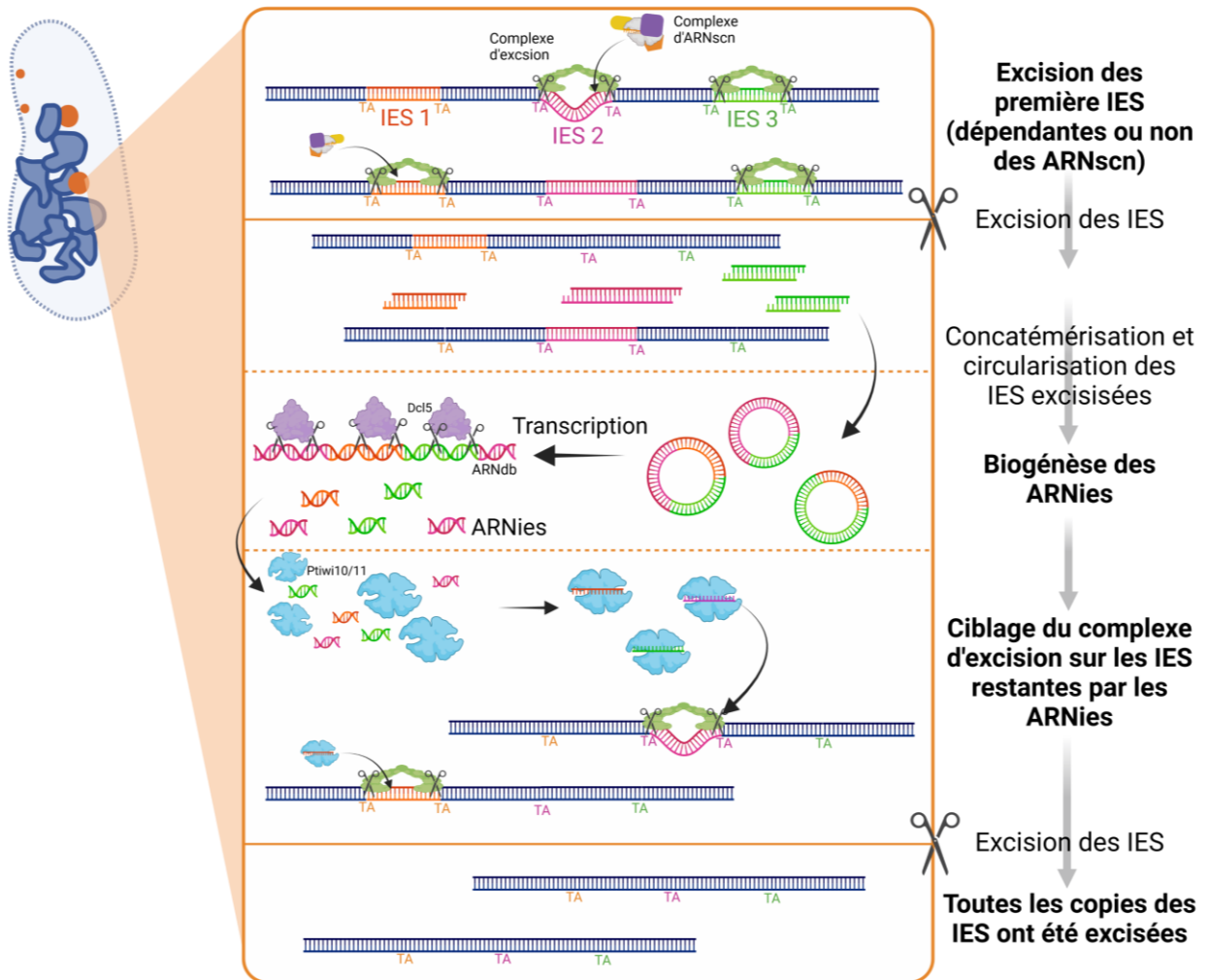


Figure 22 - Modèle des ARNies

Après un premier cycle de reconnaissance, les IES sont excisées. Les IES libres sont ensuite circularisées et éventuellement concatémérisées avant d'être transcrites pour former des ARNdb. Puis, ces ARNdb sont coupés par Dcl5 afin de former les ARNies qui sont ensuite pris en charge par Piwi10/11 et un des brins est éliminé. Ces ARN simple-brin vont participer à la reconnaissance des IES restant à exciser. (Created with BioRender.com)

3.1.2.1 Le modèle des ARNies

Une fois excisées les IES sont relarguées comme des molécules linéaires puis éventuellement concatémérisées avant d'être circularisées (Allen et al., 2017; Bétermier et al., 2000; Kapusta et al., 2011). Ces cercles d'IES ne sont pas répliqués mais sont transcrits permettant la formation d'un autre type de petits ARNnc de 25-30 pb, appelés ARNies. Contrairement aux ARNscn, les ARNies s'accumulent aux stades tardifs du développement des nouveaux MAC (un fois que l'excision des IES a commencé). La formation des ARNies à partir des ARNnc double-brin produits depuis les cercles d'IES fait intervenir non pas Dcl2/3 mais Dcl5 (Sandoval et al., 2014). Elle fait également intervenir Pdsg2, nécessaire à la stabilité de Dcl5 dans les MAC en développement (Arambasic et al., 2014). Les ARNies sont ensuite pris en charge par les protéines ohnologues Ptiwi10 et Ptiwi11 (Furrer et al., 2017) et viennent aider les ARNscn dans leur marquage des régions du génome à éliminer. Ils marquent plus spécifiquement les extrémités des IES (Sandoval et al., 2014). L'élimination imprécise semble, quant à elle, indépendante de ces ARNies.

L'expression tardive des *PTIWI10* et *11* et l'échec d'expression de *PTIWI10* à partir d'un transgène (sans IES) injecté dans le MAC parental, suggèrent que ces gènes sont exprimés à partir des nouveaux MAC après l'excision des IES présentes dans leur séquence codante (Furrer et al., 2017). Cette hypothèse est soutenue par le fait que les IES présentes dans les gènes *PTIWI10* et *11* font partie des premières IES excisées (Zangarelli et al., 2022).

3.1.2.2 L'impact des ARNies sur l'excision des IES

Comme pour les IESscn, les IES dont l'excision dépend des ARNies sont peu nombreuses. En effet, le séquençage d'ADN issu de préparations de noyaux enrichies en nouveaux MAC obtenues après un ARNi contre *DCL5*, a permis d'identifier 6 630 IES ayant un score de rétention supérieur à 5%. Cet ensemble d'IES est différent de celui obtenu après un ARNi contre *DCL2/3* mais est toujours inclus, à 97%, parmi les IES sensibles à un ARNi contre *TFIIS4* (Maliszewska-Olejniczak et al., 2015; Sandoval et al., 2014). Ce faible taux de rétention peut être dû au fait que les ARNies sont un système accessoire facilitant l'excision de toutes les IES mais n'étant strictement nécessaire qu'à un petit nombre d'IES. Les IES dépendant des ARNies sont des IES éliminées tardivement (Zangarelli et al., 2022)

laissant la possibilité au cycle d'endoduplication suivant d'être effectué. Il y aurait donc deux fois plus de copies de ces IES à exciser, ce qui serait facilité par les ARNies.

Il est également possible que les mécanismes faisant intervenir les ARNscn et ARNies se complètent l'un l'autre. En effet, éteindre les deux voies, par un ARNi contre les *DCL2/3* et *5*, conduit à une rétention similaire à celle observée lors d'un ARNi dirigé contre *TFIS4* (Swart et al., 2017). On peut donc penser que lors d'un ARNi contre *DCL2/3* il y a une compensation par les ARNies et réciproquement lors d'un ARNi contre *DCL5* par les ARNscn. Ce qui expliquerait le petit nombre d'IES impacté par les ARNi conduisant à l'interruption d'une seule des deux voies.

3.1.3 Modification de la chromatine et marques épigénétiques

3.1.3.1 Le dépôt des marques d'histone

Chez les eucaryotes, la transcription des transposons est maintenue silencieuse grâce à la formation locale d'hétérochromatine (voir section 2.1.3.1). Chez les ciliés le système de répression de l'expression des éléments transposables a été poussé à l'extrême avec l'élimination de ces séquences du génome MAC (voir section 2.2.1). Cette élimination repose également sur la formation d'hétérochromatine avec le dépôt de marques H3K9me3 et H3K27me3 (Lhuillier-Akakpo et al., 2014).

Ces deux marques d'histone sont considérées depuis longtemps comme ayant des rôles distincts chez la plupart des eucaryotes multicellulaires. Cependant, les récentes mises en évidence d'une connexion entre les deux voies tendent à nuancer ce propos (Délérís et al., 2021). Chez la paramécie, les marques H3K9me3 et H3K27me3 sont déposées par PRC2 au niveau des séquences éliminées (Frapporti et al., 2019; Lhuillier-Akakpo et al., 2014).

3.1.3.1.1 Le complexe PRC2 de la paramécie

L'unité catalytique du complexe a été la première identifiée. Il s'agit d'Ezl1 (*Enhancer-of-zest-like 1*) une méthyltransférase capable de tri-métyler à la fois la 9^{ème} et la 27^{ème} lysine de la queue N-terminale de l'histone 3 (Frapporti et al., 2019; Lhuillier-Akakpo et al., 2014). Ezl1 seule n'est pas capable de métyler ses cibles,

elle agit au travers du complexe PRC2. Le complexe PRC2 est très conservé chez les eucaryotes, il se compose des sous-unités Ezl1, Suz12 (*Suppressor of zest 12*) et Eed (*Embryonic ectoderm development*) (Vijayanathan et al., 2022) dont les homologues ont été récemment identifiés chez la paramécie (Figure 23) (Miró-Pina et al., 2022; Wang et al., 2022b).

Au complexe s'ajoute Caf1 (*Chromatin assembly factor 1*), orthologue chez la paramécie de RBAP46/48 (*Retinoblastoma-associated proteins 46/48*), et Rf2 (*Ring finger 2*) deux sous-unités du complexe PRC2 qui stabilisent le complexe dans les MAC en développement (Figure 23). Caf1 par son interaction avec les Ptiwi01/09 participe à la stabilisation de PRC2 sur ses cibles marquées par les ARNscn (voir section 3.1.3.1.2.2). Rf2 semble quant à elle nécessaire à la stabilité de PRC2 sur la chromatine des nouveaux MAC (Ignarski et al., 2014; Miró-Pina et al., 2022; Wang et al., 2022b).

En plus des co-facteurs essentiels précités, la paramécie possède les facteurs Eap1 (*Ezl1-associated protein 1*) et Rf4 (*Ring finger 4*) (Figure 23). Ces facteurs stabilisent le complexe PRC2 sur la chromatine et permettent ainsi l'établissement ou le maintien des marques H3K9me3 et H3K27me2 au niveau d'une partie des séquences à éliminer sans pour autant être totalement nécessaires à leur dépôt. Rf4 participe notamment à la connexion entre le complexe PRC2 et Ptiwi09 permettant à PRC2 de reconnaître ses cibles par le biais des ARNscn (voir section 3.1.3.1.2.2) (Miró-Pina et al., 2022).

3.1.3.1.2 Interaction entre PRC2 et les ARNscn

3.1.3.1.2.1 PRC2 impliqué dans la maturation des ARNscn dans le MAC maternel

Puisque PRC2 est responsable du marquage des régions à éliminer dans les nouveaux MAC, il est normal de trouver les éléments composant le complexe et ses co-facteurs dans les MAC en développement. Cependant les protéines de fusion GFP avec les différents composants du complexe et ses co-facteurs montrent également un marquage du MAC parental avant sa fragmentation (c'est-à-dire lors de la méiose des MIC et le tout début de la formation des nouveaux MAC) (Ignarski et al., 2014; Lhuillier-Akakpo et al., 2014; Miró-Pina et al., 2022; Wang et al., 2022b).

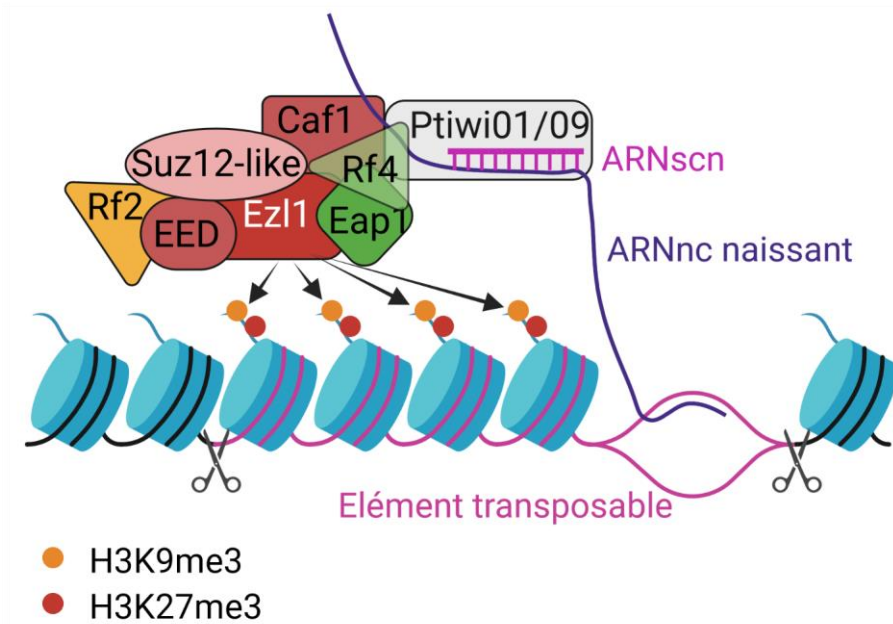


Figure 23 - Recrutement du complexe PRC2 sur ses cibles par les ARNscn

Le complexe PRC2 de la paramécie est formé de Ezh1, EED, Suz12-like, Caf1 et Rf2. Au cours des RPG, les Ptiwi01/09 sur lesquels sont fixés des ARNscn s'apparient au niveau des ARNnc naissants correspondant à des éléments à éliminer (comme des éléments transposables). Le complexe PRC2 est alors recruté au niveau des Ptiwi01/09 par l'intermédiaire de Rf4 et Eap1. La sous-unité catalytique de PRC2, Ezh1, peut alors déposer les marques H3K9me3 et H3K27me3 au niveau des éléments transposables et d'autres séquences à éliminer. Le lien entre ces marques d'histone et le recrutement de la machinerie d'excision reste encore à établir. (Created with BioRender.com)

La fonction du complexe dans le MAC parental reste encore mal comprise, le complexe pourrait aussi bien participer à la régulation transcriptionnelle des gènes dans le MAC parental que participer à la biogénèse des ARNnc produits dans le MAC parental dans le modèle des ARNscn (voir section 0). En effet, Caf1 a d'abord été décrite pour son effet sur la sélection des ARNscn MIC-spécifiques (Ignarski et al., 2014; Swart et al., 2017) avant d'être identifiée comme faisant partie du complexe PRC2 (Miró-Pina et al., 2022; Wang et al., 2022b). L'idée d'une implication directe ou indirecte dans la dégradation des ARNscn s'hybridant sur le MAC parental s'est alors étendue à tout le complexe PRC2. Cette hypothèse s'appuie sur le fait que des ARNi dirigés contre *EZL1* affectent également la sélection des ARNscn (Lhuillier-Akakpo et al., 2014; Swart et al., 2017).

3.1.3.1.2.2 Recrutement de PRC2 dans les MAC en développement

Dans les MAC en développement, le recrutement de PRC2 sur ses cibles se fait au travers des ARNscn portés par les Ptiwi01/09 (voir section 3.1.1.2). Ces ARNscn permettent de marquer les séquences à éliminer en se fixant au niveau des ARN naissants produits à partir des MAC en développement (Figure 21). Étant donné que la quasi-totalité des IES retenues dans un ARNi *TFIIS4* est comprise dans les IES retenues lors d'un ARNi contre *EZL1*, on peut supposer que la production de bon nombre de ces ARNnc dépend du facteur d'élongation TFIIS4 (Maliszewska-Olejniczak et al., 2015).

Le lien entre le complexe PRC2 et les ARNscn a récemment été identifié grâce à des immunoprécipitations suivies de spectrométrie de masse. C'est la protéine Rf4 qui servirait d'intermédiaire entre le complexe PRC2 et les Ptiwi01/09 (Figure 23). Mais cet intermédiaire ne semble pas nécessaire à l'activité enzymatique du complexe PRC2 puisque des modifications H3K9me3 et H3K27me3 sont observées dans le MAC en développement même dans des conditions d'ARNi contre *RF4*. Cependant, les modifications ne semblent pas être correctement localisées sur la chromatine puisque dans le même temps les éléments ne transposables ne sont plus enrichis en marques H3K9me3 et H3K27me3 comparé au reste du génome. Il semblerait que Rf4 serve de connexion entre les complexes PRC2 et ceux des ARNscn. Il permet ainsi le marquage des séquences à éliminer (Miró-Pina et al., 2022; Swart et al., 2014). PRC2 pourrait

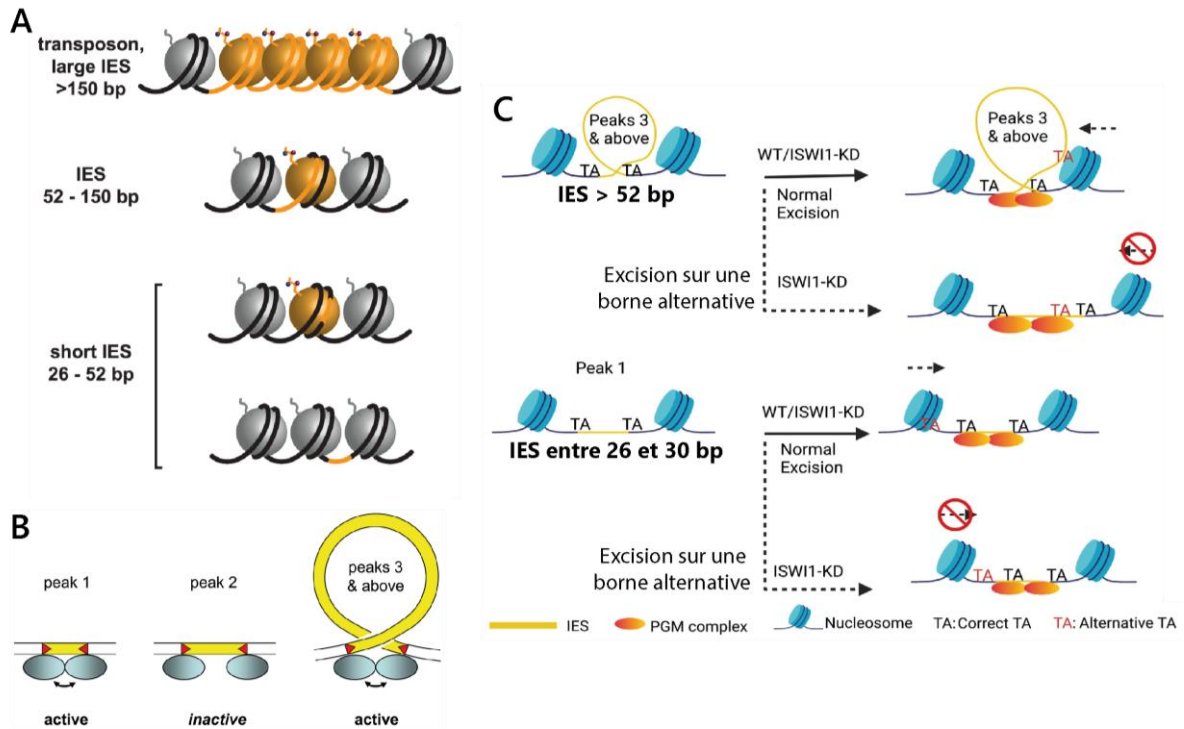


Figure 24 - Modèles d'excision des IES sur la chromatine

A Modèle expliquant le biais de taille des IES (en orange) dépendantes des marques d'histones H3K4me3 et H3K27me3 (rond rouge ou noir) (Lhuillier-Akakpo et al., 2014). **B** Modèle d'assemblage d'un complexe actif sur l'ADN nu en fonction de la taille de l'IES (Arnaiz et al., 2012). **C** Modèle du remodelage de la chromatine permettant une excision précise des IES (Figure adaptée de Singh et al., 2022).

également avoir un effet indirect en agissant sur la transcription des ARNnc maternels sur lesquels les ARNscn vont s'hybrider. L'absence de *scanning* conduirait à la rétention de tous les ARNscn. Une fois transférés dans les nouveaux MAC, ces derniers couvriraient alors toutes les séquences, ne permettant plus de discriminer les séquences destinées au MAC et celles spécifique des MIC.

3.1.3.1.3 Des marques épigénétiques nécessaires à l'élimination de nombreuses séquences

Les modifications d'histone H3K9me3 et H3K27me3, déposées par PRC2, sont nécessaires à l'élimination imprécise des régions porteuses d'éléments transposables, mais également à l'élimination de ~26 000 IES (soit ~70% des IES) (Frapporti et al., 2019; Lhuillier-Akakpo et al., 2014). Ces modifications devraient être positionnées de manière très précise sur les IES car la quasi-totalité des IES est plus petite que l'ADN enroulé autour d'un nucléosome (< 150 pb). Un unique nucléosome devrait donc être marqué pour identifier précisément la localisation de l'IES. Ce qui a conduit au modèle selon lequel, d'une part, le complexe d'excision couperait l'ADN aux extrémités des régions couvertes d'hétérochromatine. D'autre part, le complexe pourrait être recruté localement pour exciser les IES plus petites (Figure 24A). Même si le mécanisme permettant le recrutement du complexe d'excision au niveau des IES reste encore à éclaircir.

Si les séquences éliminées de manière imprécise sont bien enrichies en marques H3K9me3 et H3K27me3 déposées par PRC2, aucune donnée n'indique directement que des IES sont couvertes par des nucléosomes marqués. De plus, ces marques se trouvent également dans les fragments du MAC parental (Ignarski et al., 2014). Il est donc possible que le rôle de PRC2 dans l'excision des IES soit indirect.

Enfin, ce modèle n'explique pas comment une IES dont une borne se trouve enroulée autour d'un nucléosome peut être précisément excisée. Deux mécanismes sont alors possibles : premièrement, on peut supposer que la borne puisse être coupée sur le nucléosome mais avec plus de difficulté. Il est également possible que la machinerie ne puisse couper précisément une borne d'IES que si elle est en dehors d'un nucléosome. Auquel cas les IES de plus de 150 pb pourraient avoir été contre-sélectionnées car la nécessité de déplacer (ou

supprimer) plus d'un nucléosome ou de devoir couper la borne enroulée autour d'un nucléosome pourrait rendre leur excision plus difficile.

3.1.3.2 Le remodelage local de la chromatine

Une récente étude a permis de définir la position des nucléosomes sur la chromatine du MAC végétatif, révélant une périodicité d'environ 150 pb avec seulement quelques paires de bases entre chaque nucléosome (Gnan et al., 2022). Même si la position des nucléosomes au cours des RPG n'est pas connue, on peut tout de même voir la similitude entre la taille des IES (inférieure à 150 bp) et celle des nucléosomes. L'excision des IES est peut-être plus compliquée si l'IES s'enroule sur plus d'un nucléosome. Il est possible que la chromatine ne soit pas suffisamment accessible dès lors que l'IES est couverte par plus d'un nucléosome, ou que les nucléosomes rendent la chromatine trop rigide pour permettre une interaction entre les deux bornes de l'IES. Plusieurs découvertes récentes pourraient faire pencher la balance en la faveur de cette seconde hypothèse. En effet, deux protéines impliquées dans le remodelage de la chromatine ont été identifiées comme nécessaires au bon déroulement des RPG.

La première est un homologue de Spt16 (*Suppressor of Ty 16*) exprimé spécifiquement lors des RPG et se localisant dans les MAC en développement (Spt16-1) (de Vanssay et al., 2020). Spt16 est connu comme faisant partie du complexe FACT (*Facilitate Chromatin Transcription*) notamment impliqué dans l'ouverture de la chromatine afin de faciliter la réplication de l'ADN, mais également la transcription (Hammond et al., 2017). Chez la paramécie, un ARNi contre *SPT16-1* empêche Pgm de se localiser stablement dans les MAC en développement sans pour autant modifier son expression, ce qui a pour conséquence la rétention de tous les éléments normalement éliminés par Pgm (voire section 2.2.1.3). La perturbation de la localisation de Pgm en l'absence de Spt16-1 semble agir indépendamment ou en aval des ARNscn et des marques H3K9me3 et H3K27me3 qui ne sont pas impactés par l'ARNi contre *SPT16-1*. Spt16-1 est donc nécessaire à la localisation de Pgm, soit parce qu'elle permet au complexe d'excision d'interagir avec la chromatine, soit en remodelant la chromatine de manière que le complexe puisse avoir accès à l'ADN (de Vanssay et al., 2020).

La seconde est un homologue du remodeleur de la chromatine Iswi exprimé spécifiquement lors des RPG et localisé dans les MAC en développement : Iswi1a (*Imitation-Switch*) (Singh et al., 2022). Un ARNi contre *ISWI1a* induit la rétention de 78% des IES avec un biais vers les grandes IES. Ces IES représentent une partie des IES dépendantes du complexe PRC2 et de TFIIIS4. En effet, les IES retenues dans un ARNi contre *ISWI1a* sont pratiquement toutes incluses dans les IES dépendantes de PRC2 et de TFIIIS4 (Analyse complémentaire Olivier Arnaiz). Étant donné qu'une immunoprécipitation d'Iswi1a suivie de spectrométrie de masse a permis d'identifier des acteurs impliqués dans la biogénèse des ARNscn, Nowa1 et Ptiwi01 (voir section 3.1.1.2). Il est proposé que la localisation d'Iswi au niveau des IES puisse être dirigée directement ou indirectement par les ARNscn (Singh et al., 2022). Davantage de résultats (notamment sur la localisation de Iswi1a en l'absence d'ARNscn) pourraient permettre de confirmer ou non cette hypothèse. De plus, on observe une forte augmentation du nombre d'excisions sur un TA alternatif (Singh et al., 2022). Ces observations ont remis en avant le modèle initial selon lequel l'excision de IES se ferait sur de l'ADN nu (Figure 24B) (Arnaiz et al., 2012) en y ajoutant l'importance d'un remodelage local de la chromatine afin de masquer les bornes alternatives (Figure 24C).

3.1.4 Quelles sont les séquences éliminées dépendantes des facteurs épigénétiques ?

Les séquences éliminées par la paramécie peuvent être séparées en deux groupes distincts avec d'une part, les éléments éliminés imprécisément (principalement des transposons et des séquences répétées) et d'autre part, les IES éliminées précisément (voir section 2.2.2).

L'identification des facteurs permettant la reconnaissance des éléments à éliminer a permis de déterminer les marques nécessaires à la reconnaissance des transposons avant leur élimination. En effet, l'élimination de toutes ces séquences est dépendante à la fois du complexe PRC2 et des ARNscn (Arambasic et al., 2014; Frapporti et al., 2019; Miró-Pina et al., 2022; Zangarelli et al., 2022). L'image est cependant moins claire pour les IES. En effet les IES regroupent un très grand nombre de séquences de taille et d'ancienneté différentes (voir section 2.2.2.1).

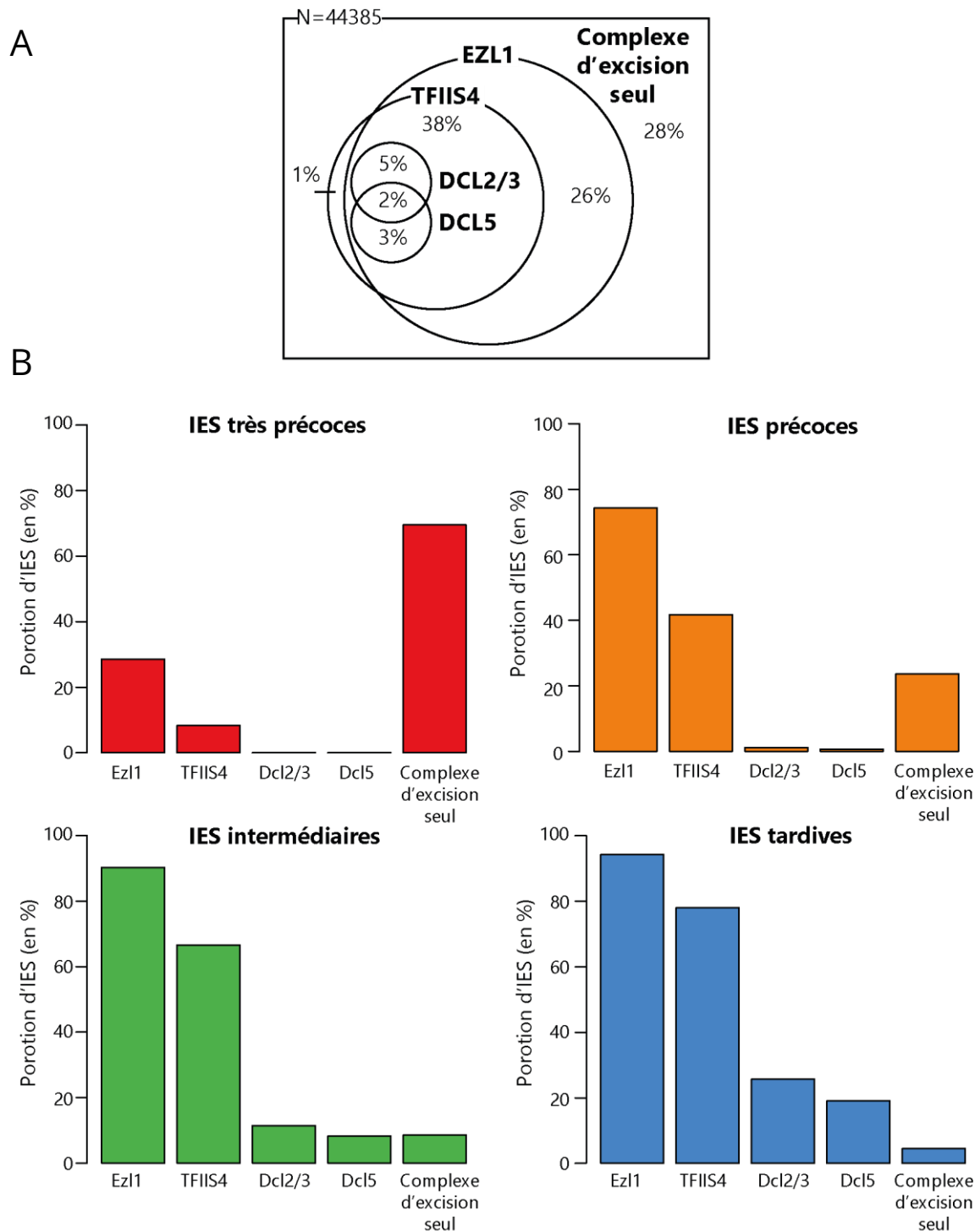


Figure 25 - Dépendance des IES aux facteurs épigénétiques

A Diagramme de Venn montrant la dépendance des IES aux facteurs épigénétiques, obtenu après des ARNi contre *EZL1*, *TFIIIS4*, *DCL2/3* ou *DCL5*. Les IES ne dépendant pas de ces facteurs sont classées dans le groupe « complexe d'excision seul ». **B** Les histogrammes représentent la proportion d'IES dépendant des facteurs épigénétiques en fonction de leur chronologie d'excision. (Figures adaptées de Zangarelli et al., 2022)

De plus, leur excision peut dépendre de plusieurs facteurs épigénétiques précédemment décrits (des ARNnc : ARNscn et/ou ARNies ; des marques d'histone : H3K9me3 et H3K27me3) (Figure 25A).

La chronologie d'excision et la conservation des séquences ont permis de montrer que les IES intégrées depuis longtemps dans le génome ont tendance à être plus courtes et à faire partie des IES les plus efficacement excisées (précoces et précises), suggérant une optimisation des séquences (Sellis et al., 2021; Zangarelli et al., 2022). Le modèle d'une optimisation est par ailleurs soutenu par le fait que les IES nécessitant la machinerie des ARNscn et ARNies sont des IES plutôt grandes (Sandoval et al., 2014; Swart et al., 2014). Le même biais étant observé lors des ARNi contre *EZL1* (Lhuillier-Akakpo et al., 2014), il semblerait que les IES excisées le plus précocement ne nécessitent pas les marques épigénétiques connues (Figure 25B) (Zangarelli et al., 2022). On peut alors supposer qu'une IES n'est marquée par des modifications d'histones que si elle est incluse dans un nucléosome (Figure 24A) et qu'elle n'a pas encore été optimisée.

Malgré les nombreux mécanismes épigénétiques identifiés, il reste encore près de 30% des IES, insensibles à l'état de la chromatine ou aux ARN non-codants, pour lesquelles on ignore le mécanisme de reconnaissance.

3.1.5 Une séquence consensus optimisée pour certaines IES

De manière générale, les IES ont une séquence consensus faible à leurs bornes qui est peu conservée d'une IES à l'autre (Arnaiz et al., 2012). La mise en évidence de différentes catégories d'IES a cependant permis d'identifier des motifs un peu plus conservés pour les IES excisées le plus précocement (Figure 26) (Zangarelli et al., 2022). Ces IES sont enrichies en IES de petite taille (26-33 pb) intégrées depuis longtemps dans le génome. De plus, leur excision dépend de peu de facteurs épigénétiques, la majorité ne dépendant que de la machinerie d'excision. Enfin, moins d'erreurs d'excision sont observées pour cette catégorie d'IES. Il semblerait donc que ces IES aient été optimisées avec le temps afin de faciliter leur excision. Une des hypothèses est que les séquences consensus plus fortes aux bornes de ces IES soient un des résultats de cette optimisation (Sellis et al., 2021; Zangarelli et al., 2022).

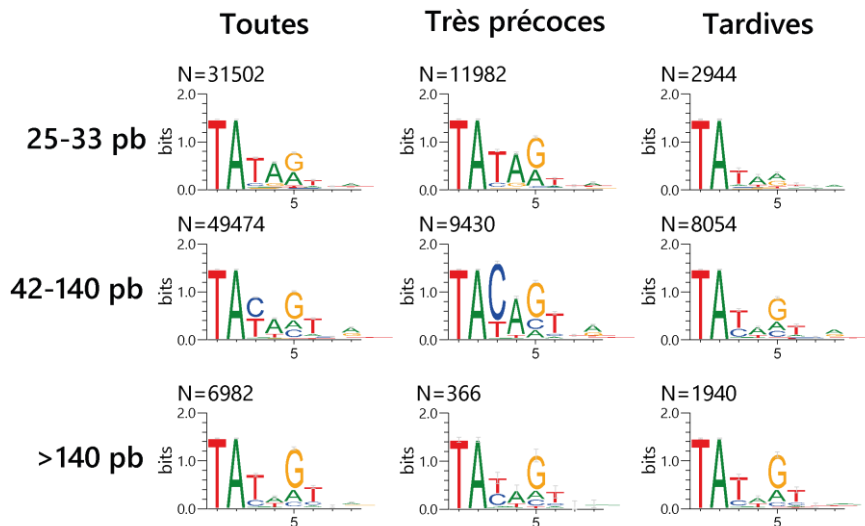


Figure 26 - Séquences consensus aux bornes des IES

Logos des séquences aux bornes de toutes les IES, des IES excisées très précocement (Très précoces) ou des IES excisées tardivement (Tardives) en fonction de leur groupe de taille. (Figure tirée de Zangarelli et al., 2022)

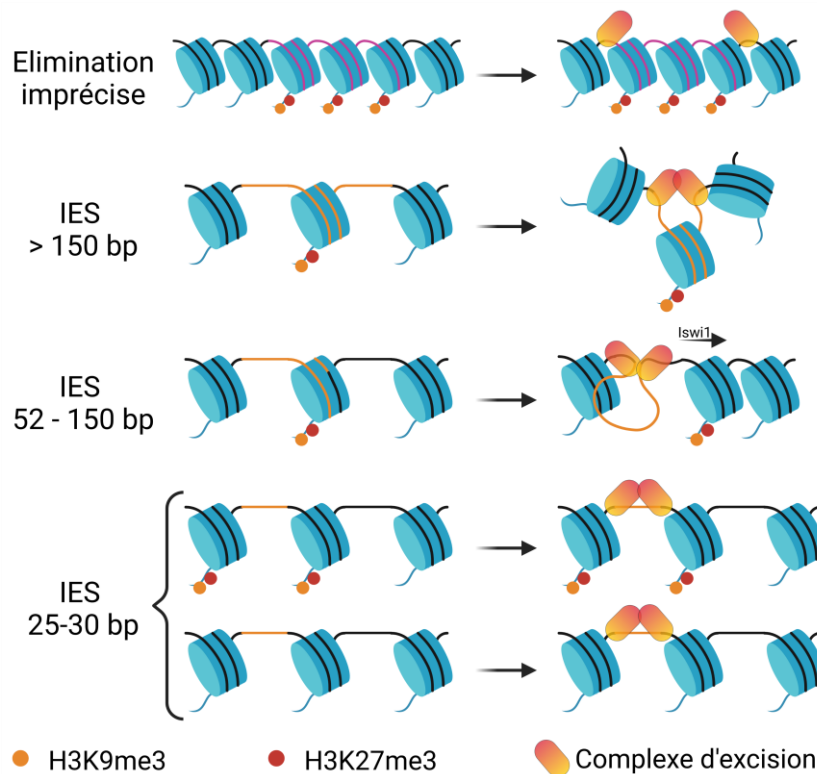


Figure 27 - Modèle de reconnaissance des séquences à éliminer sur la chromatine

Les séquences éliminées de façon imprécise (en rose) sont délimitées uniquement par l'hétérochromatine aux bornes de laquelle l'ADN est coupé. Les mécanismes de reconnaissance des IES doivent cependant être plus précis et ne peuvent reposer uniquement sur le marquage des nucléosomes, plus grand que la plupart des IES. Le mécanisme pourrait varier en fonction de la taille des IES et faire intervenir des remodeleurs de la chromatine (voir détails section 3.1.3). (Created with BioRender.com)

3.1.6 Modèle de reconnaissance des séquences à éliminer

La présence des marques d'histone déposées par PRC2 sur les séquences éliminées de façon imprécise n'est plus à établir (Frapporti et al., 2019). Il est donc possible que pour ces éléments PRC2 joue un rôle direct. Auquel cas on peut imaginer que leur élimination est imprécise, car les coupures sont simplement introduites de part et d'autre des régions sous forme d'hétérochromatine (Figure 27).

L'implication directe de PRC2 dans l'excision des IES n'a cependant pas encore été démontrée. Il est possible que la rétention des IES observée lors d'ARNi dirigés contre *EZL1* résulte du blocage de la voie des ARNscn et des ARNies. En effet, les IES dépendantes de *Ezl1* incluent les IES retenues lors d'ARNi dirigé contre *TFIIS4* (Maliszewska-Olejniczak et al., 2015). De plus, les IES retenues dans ces ARNi ont tendance à être plus grandes et à être excisées plus tardivement (Sellis et al., 2021; Zangarelli et al., 2022). Étant donné que ces IES ont tendance à être plus grandes, il est facilement imaginable qu'elles soient au moins partiellement incluses dans un nucléosome. Mais très peu d'IES font plus de 150 pb, elles ne s'enrouleraient donc pas sur plus d'un nucléosome (Arnaiz et al., 2012). Si PRC2 marque l'unique nucléosome porteur d'une portion de l'IES, il est improbable que la machinerie d'excision coupe aux bornes des marques d'hétérochromatine pour les IES. De plus, la périodicité de la taille des IES (~10 pb) correspondant à un tour d'hélice d'ADN, il est fort probable que l'excision des IES ait lieu sur de l'ADN nu (Arnaiz et al., 2012). Étant donné que les IES dépendantes d'*Iswi* sont également dépendantes d'*Ezl1* et de *TFIIS4*, on peut imaginer que PRC2 marque le nucléosome qu'il faut déplacer afin de libérer l'IES et permettre son excision. En l'absence d'*Iswi*, le nucléosome resterait en place et le complexe d'excision couperait sur un TA libre proche.

Mais, si plus de 90% des IES dépendantes de *TFIIS4* (donc des ARNscn et/ou des ARNies) sont également dépendantes de *Ezl1* (donc de PRC2), 33% des IES dépendant de PRC2 sont indépendantes de *TFIIS4*. Comment PRC2 est dirigé sur ces 9 103 IES reste encore assez flou. Étant donné que ces IES sont également indépendantes d'*Iswi*, il est possible que le nucléosome marqué ne soit pas porteur directement de l'IES, mais que l'IES soit contenue dans la séquence linker

entre deux nucléosomes marqués. Auquel cas Iswi n'aurait pas besoin de déplacer les nucléosomes.

Toutes ces marques chromatiniennes pourraient prendre du temps à se mettre en place. Ce qui expliquerait pourquoi les IES dépendantes de Ezl1 et TFII54 sont plus longues à être excisées (Zangarelli et al., 2022). Les premières IES à être excisées étant plus courtes et ne dépendant que rarement de facteurs épigénétiques, on peut supposer qu'elles se situent dans des régions *linker* entre les nucléosomes (ce qu'il faudrait vérifier expérimentalement). La séquence consensus à leurs bornes pourrait être soit une séquence directement reconnue par le complexe d'excision, soit une séquence qui doit être exclue des nucléosomes.

Enfin, il me paraît également important de remarquer une certaine interdépendance entre l'élimination d'ADN et la réplication. D'une part, si l'élimination d'ADN se poursuit même lorsque la réplication est bloquée (Zangarelli et al., 2022), elle ne démarre qu'après 5 cycles d'endoduplication. D'autre part, l'endoduplication ne semble pas pouvoir se poursuivre dans certaines conditions où les RPG ne s'effectuent pas correctement (Kapusta et al., 2011). De plus, les facteurs Caf1, Iswi1 et Spt16, précédemment décrits pour leur implication dans l'excision de nombreuses IES sont également connus pour leur rôle dans la structuration de la chromatine (Bellush and Whitehouse, 2017). Spt16, dans le complexe FACT, agit en amont de la fourche de réplication et participe à l'ouverture la chromatine afin de faciliter la réplication de l'ADN. Caf1 et Iswi1 agissent en aval de la fourche de réplication en positionnant les nucléosomes sur l'ADN néosynthétisé. Il est donc possible que les nucléosomes de l'ADN néosynthétisé soient positionnés de manière à exclure les IES optimisées par leur taille et leurs bornes.

3.2 LE COMPLEXE D'EXCISION DES IES

Des expériences d'ARNi contre *PiggyMac* (PGM) ont permis de proposer que Pgm soit l'endonucléase responsable de l'introduction précise des cassures double-brin aux bornes des IES (Arnaiz et al., 2012; Baudry et al., 2009). L'élimination imprécise tardive semble également dépendre de la présence de

Pgm mais il reste à déterminer si cette dépendance est directe ou indirecte. Étant donné que la caractérisation de la machinerie d'excision imprécise n'en est qu'à ses débuts, je me focaliserai ici sur le mécanisme d'excision des IES.

3.2.1 L'excision une affaire de transposase

La majorité des IES présentes dans le génome sont issues de l'insertion de transposons *Tc1/mariner* qui auraient dégénéré, ne laissant qu'un faible consensus 5'-TAYAGYNR-3' aux bornes des IES où seul le TA est strictement conservé (voir section 2.2.2.2). Cependant, Pgm est une transposase domestiquée de la famille des transposases PiggyBac exprimée au cours des RPG. Contrairement aux *Tc1/mariner*, les transposons *PiggyBac* sont éliminés précisément, restaurant leur site d'insertion 5'-TTAA-3' (Figure 28) (Baudry et al., 2009; Bouallègue et al., 2017; Dubois et al., 2012; Mitra et al., 2008).

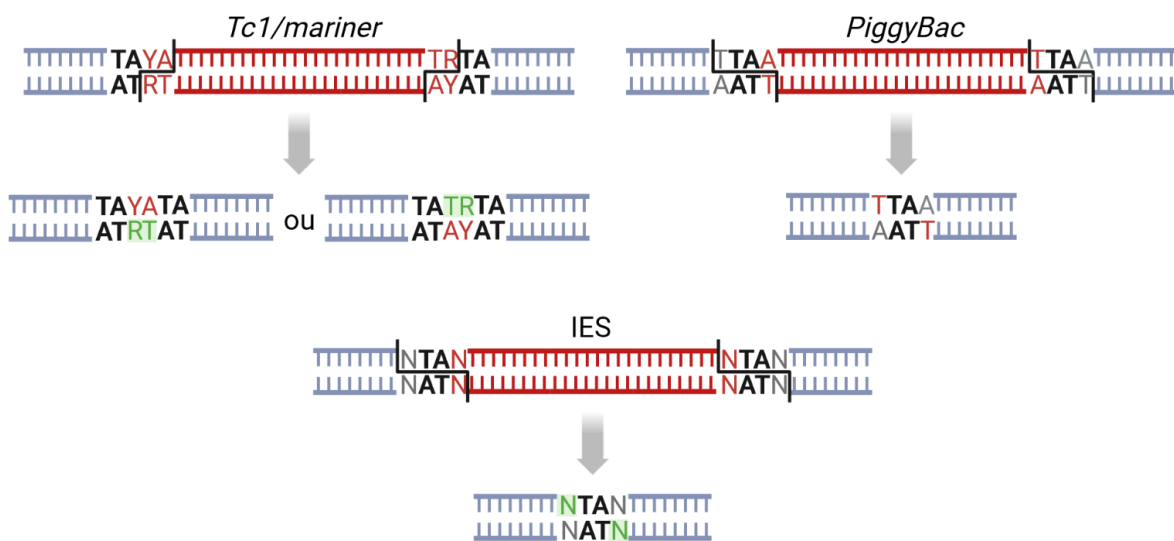


Figure 28 - Comparaison de la géométrie des cassures aux bornes des IES avec celle des transposons *Tc1/mariner* et *PiggyBac*

Représentation schématique de la géométrie des cassures double-brin introduites par les transposases *Tc1/mariner* (en haut à gauche) et *PiggyBac* (en haut à droite) ou lors de l'excision des IES (en bas). Les TA conservés entre les différents sites de coupure sont en lettres grasses noires. Le transposon ou l'IES insérés dans le génome sont représentés en rouge. Les lettres en vert représentent les bases qui sont ajoutées *de novo* lors de la réparation du site d'excision. (Created with BioRender.com)

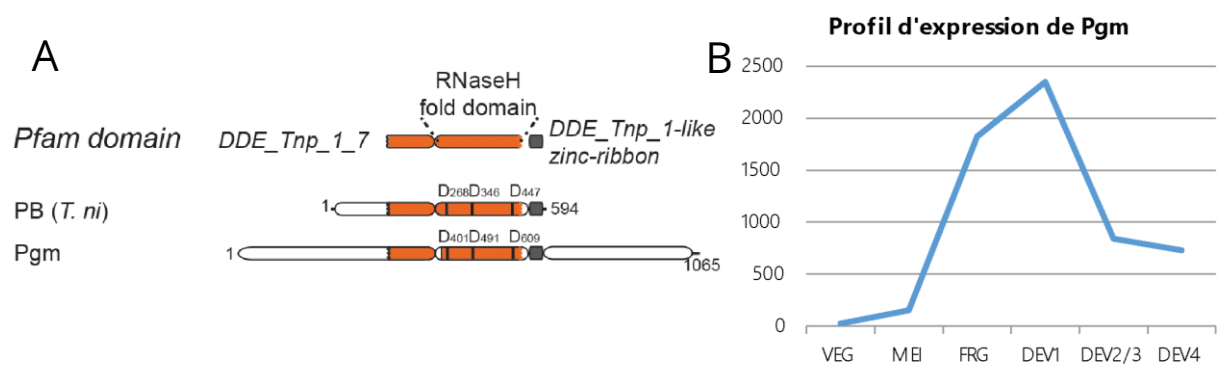


Figure 29 - La transposase domestiquée PiggyMac

A PiggyMac (Pgm) possède les domaines caractéristiques de la transposase PiggyBac (PB) de *Trichoplusia ni*. Le domaine Pfam DDE_Tnp_1_7 (PF13843) correspond aux domaines en orange, le domaine de droite correspondant aux domaines RNaseH porteur des trois acides aspartiques catalytiques (barre grise). Auquel s'ajoute le domaine Pfam DDE_Tnp_1_7-like zinc ribbon (PF13842) (carré gris) correspondant au domaine riche en cystéine (CRD) de PB. Pgm porte un variant de ce domaine (Figure tirée de Bischerour et al., 2018). B Profil d'expression de Pgm au cours de l'autogamie déterminé à partir de séquençages ARN (Arnaiz et al., 2017).

La transposase domestiquée Pgm possède la triade catalytique DD(D/E) dans son domaine RNaseH caractéristique des transposases PiggyBac. Elle possède également un domaine riche en cystéine (CRD, *Cysteine-Rich Domain*) comme la plupart des transposases PiggyBac, suivi d'un long domaine *coiled-coil* C-terminal spécifique des ciliés (Figure 29A). La triade catalytique et le domaine CRD de Pgm sont strictement nécessaires à son activité sans pour autant altérer sa localisation. (Dubois et al., 2017)

Une étude plus approfondie de la structure de Pgm a permis de mettre en évidence la particularité de son CRD. En effet, le CRD de Pgm diffère de celui de nombreuses transposases PiggyBac (domestiquées ou non). Contrairement à d'autres transposases PiggyBac, le CRD de Pgm lui permettrait non pas de se lier avec l'ADN mais plutôt d'interagir avec la queue N-terminale de l'histone H3 quel que soit son état de méthylation (Guérineau et al., 2021). Ni Pgm, ni les PiggyMac-Likes (PgmL, détaillées dans la section suivante) ne semblent donc être en mesure de reconnaître spécifiquement les modifications H3K9me3 et H3K27me3 censées marquer une partie des séquences à éliminer (voir section 3.1.3). Leur CRD pourrait simplement servir à stabiliser le complexe d'excision sur la chromatine, tandis que les marques seraient lues par un autre acteur qui n'a pas été encore identifié.

De plus, les constructions Pgm inactives (délétées pour la triade catalytique ou le CRD) ont un effet toxique même en présence de Pgm endogène, suggérant une compétition avec le Pgm endogène. La délétion du domaine C-terminal ne semble pas modifier la localisation de Pgm, ni abolir complètement l'activité du complexe d'excision (Dubois et al., 2017). L'activité partielle du complexe d'excision dans ces conditions est soumise à la présence d'un niveau résiduel de Pgm endogène dans la cellule. En effet, Pgm délétée de son domaine C-terminal seul ne peut pas compléter un ARNi PGM. Cette différente observation suggère que Pgm multimérise dans les complexes d'excision et qu'il faut *a minima* une sous-unité porteuse du domaine C-terminal pour permettre l'activité du complexe (Bischerour et al., 2018; Bouallègue et al., 2017; Dubois et al., 2017). Le rôle précis de ce domaine spécifique des ciliés n'est cependant pas connu à l'heure actuelle. Il pourrait être lié à la domestication de Pgm et/ou nécessaire à l'interaction de Pgm avec d'autres membres du complexe.

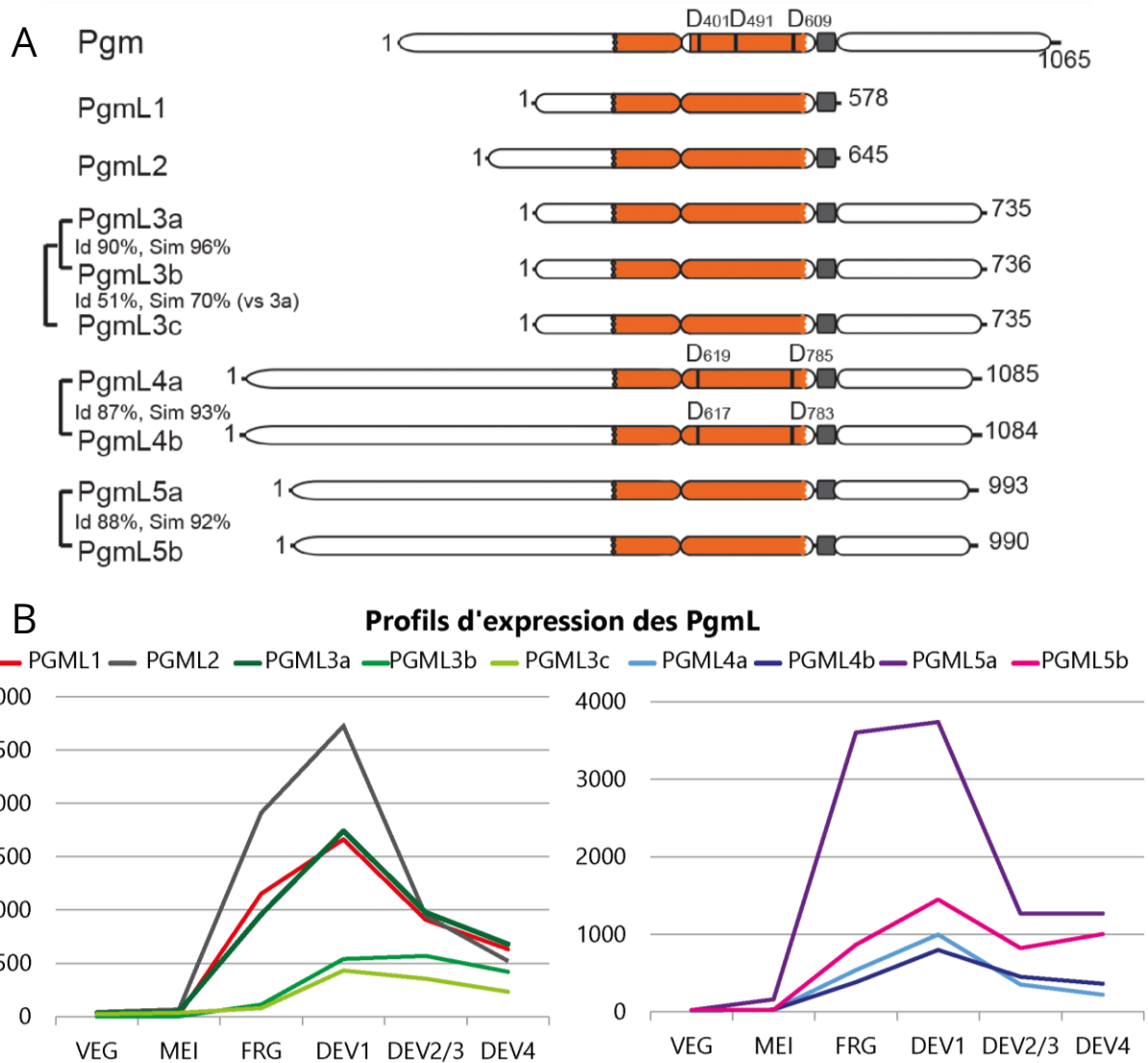


Figure 30 - Les PiggyMac-likes

A Comme Pgm, les PiggyMac-likes (PgmL) possèdent le domaine Pfam DDE_Tnp_1_7 (en orange), mais aucune n'a conservé l'ensemble des acides aminés catalytiques (barres grises). Id = pourcentage d'identité des séquences en acides aminés ; Sim = pourcentage de similarité. (Figure tirée de Bischerour et al., 2018). **B** Profil d'expression des PgmL au cours de l'autogamie déterminé à partir de séquençages ARN (Arnaiz et al., 2017).

Les récentes études ont montré qu'en plus de multimériser, Pgm a besoin de partenaires pour introduire les cassures aux bornes des IES. En leur absence, l'excision des IES n'a pas lieu, rendant impossible la production d'une descendance sexuelle viable après l'autogamie. Ce phénotype a été largement utilisé afin d'identifier les composants de la machinerie d'excision.

3.2.2 Des PiggyMac (like) en cascade

En plus de Pgm, d'autres transposases PiggyBac domestiquées ont pu être identifiées chez *P. tetraurelia* et d'autres espèces de paramécies. Elles sont toutes exprimées spécifiquement au cours des RPG. De plus, elles sont porteuses des domaines RNaseH et CRD spécifiques retrouvés chez PiggyMac. Cependant elles ne possèdent pas (ou seulement partiellement) la triade catalytique DD(D/E) (Figure 30) laissant supposer qu'elles sont catalytiquement inactives d'où leur nom de PiggyMac-like (PgmL) (Bischerour et al., 2018; Guérineau et al., 2021).

Bien que catalytiquement inactives, toutes les PgmL sont capables d'interagir directement avec Pgm et permettent de stabiliser cette dernière dans les MAC en développement. Elles participent ainsi à la formation d'une descendance viable. Toutes les PgmL ne semblent cependant pas avoir le même rôle au sein du complexe. En effet, un ARNi contre *PGML1* a un effet plus faible sur la localisation de Pgm dans les ébauches que les ARNi contre les autres *PGML*. Ceci laisse supposer que PgmL1 est moins important pour la localisation de Pgm que les autres PgmL, ou qu'une autre PgmL peut la remplacer.

Le séquençage des MAC en développement dans des conditions d'ARNi contre *PGML1* ou *PGML3ab* (les ohnologues *PGML3* les plus exprimés) révèle une plus faible rétention des IES que lors de l'ARNi des autres *PGML* (notamment pour les petites IES) mais avec 4 à 5 fois plus d'erreurs d'excision que dans des conditions sauvages, l'excision se faisant sur un TA alternatif interne à l'IES de référence. PgmL1 et 3 augmentent donc la stabilité du complexe d'excision dans les MAC en développement et diminuent le risque d'erreur (Bischerour et al., 2018). Une hypothèse serait que PgmL1 et 3 apportent de la flexibilité au complexe d'excision, lui permettant d'ajuster la distance entre ses sous-unités catalytiques Pgm. Ainsi, le complexe pourrait mieux s'adapter à la distance entre les bornes et

éviter l'introduction de cassure sur des TA ne correspondant pas aux bornes de l'IES.

De récents résultats non-publiés de l'équipe (produits par Julien Bischerour, Vinciane Régner et Coralie Zangarelli) ont permis de proposer un ordre d'assemblage du complexe dans les MAC en développement (Figure 31). PgmL1 et 3 seraient les premières à se fixer stablement sur la chromatine des MAC en développement permettant la stabilisation de PgmL2 et 5. Enfin PgmL4 et Pgm se fixeraient sur la chromatine afin d'exciser les IES. Deux modèles sont alors possibles. Le premier supposerait que tous les Pgm et PgmL s'assemblent pour former un gros complexe d'excision. Il est également possible que le complexe soit dynamique et que les PgmL1 et 3 soient nécessaires à localisation stable de PgmL5 et 2 dans les nouveaux MAC, mais qu'ensuite les PgmL1 et 3 se dissocient du complexe et se fixent sur leur cible suivante, ou soient dégradées. De la même manière, PgmL5 et 2 pourraient stabiliser PgmL4 et Pgm sur la chromatine avant de se dissocier du complexe et permettre l'excision des IES.

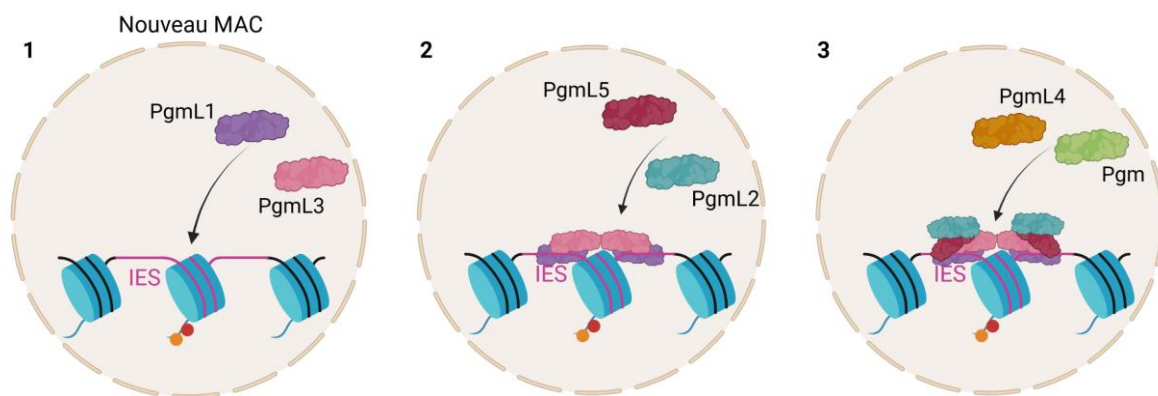


Figure 31 - Hiérarchie d'assemblage des PgmL

1 PgmL1 et PgmL3 sont les premières à s'associer à la chromatine. 2 PgmL2 et PgmL5 s'associent ensuite sur la chromatine. 3 Enfin, PgmL4 et Pgm s'ajouteraient au complexe d'excision afin que Pgm puisse introduire les cassures aux bornes des IES. La stœchiométrie du complexe reste cependant à définir. (Created with BioRender.com)

3.2.3 Un couplage entre cassure et réparation

Le complexe d'excision n'est pas composé uniquement de transposases PiggyBac domestiquées. En effet, de manière surprenante, des expériences d'ARNi

ont montré que l'hétérodimère Ku70/Ku80 semble également nécessaire à l'activité du complexe (Dubois et al., 2012; Marmignon et al., 2014). Ces protéines sont plus connues pour leur rôle dans la voie de réparation des cassures double-brin d'ADN appelée NHEJ pour *Non-Homologous End-Joining* (Zahid et al., 2021).

La paramécie ayant subi plusieurs duplications globales du génome (voir section 1.2.2.1), les gènes *KU* s'en sont trouvés dédoublés. Elle possède ainsi deux ohnologues de *KU70* issus de la dernière WGD (*KU70a* et *KU70b*) codant pour des protéines aux séquences primaires identiques à 98%. Bien que *Ku70a* soit plus fortement exprimée que *Ku70b* durant l'autogamie, il semblerait que les deux protéines aient des fonctions redondantes. Dans le génome de la paramécie on trouve également trois copies de *KU80* : *KU80a*, *KU80b* et *KU80c*. *Ku80a* et *Ku80b* sont issues de la dernière WGD, elles présentent près de 90% d'identité dans leur séquence primaire (Marmignon et al., 2014). Leur expression est constitutive dans les cellules végétatives et reste stable au cours de l'autogamie. Leur déplétion par ARNi ne semble affecter ni la localisation de Pgm, ni la viabilité de la descendance. De plus, contrairement à *Ku80c*, *Ku80a* ne se localise pas stablement dans les MAC en développement et n'est pas capable de compléter un ARNi contre *KU80c*. *Ku80a* et *b* n'interviennent donc pas spécifiquement lors des RPG. Enfin, *Ku80c* est exprimée spécifiquement lors des RPG et est indispensable à la localisation de Pgm et l'élimination d'ADN (précise ou non). Il existe donc une spécialisation de *Ku80c* pour les RPG (Abello et al., 2020; Marmignon et al., 2014). De récentes données d'immunofluorescence laissent penser que l'hétérodimère *Ku70/Ku80c* ancrerait PgmL4 et Pgm aux complexes formés par les PgmL1, 2, 3 et 5 (voir Figure 31, données non-publiées de l'équipe produites par Valerio Vitali, Julien Bischerour et Coralie Zangarelli).

Par ailleurs, il a été montré que, dans un système hétérologue, *Ku70* et *Ku80c* forment un complexe avec Pgm, laissant supposer que *Ku80c* n'est pas simplement recruté au niveau des cassures double-brin générées par la transposase domestiquée, mais fait partie du complexe d'excision (Marmignon et al., 2014). Des études de chimères entre *Ku80a* et *Ku80c* ont permis de déterminer que la spécialisation de *Ku80c* dans les RPG est due à l'évolution de son domaine α/β présent en N-terminal (Abello et al., 2020).

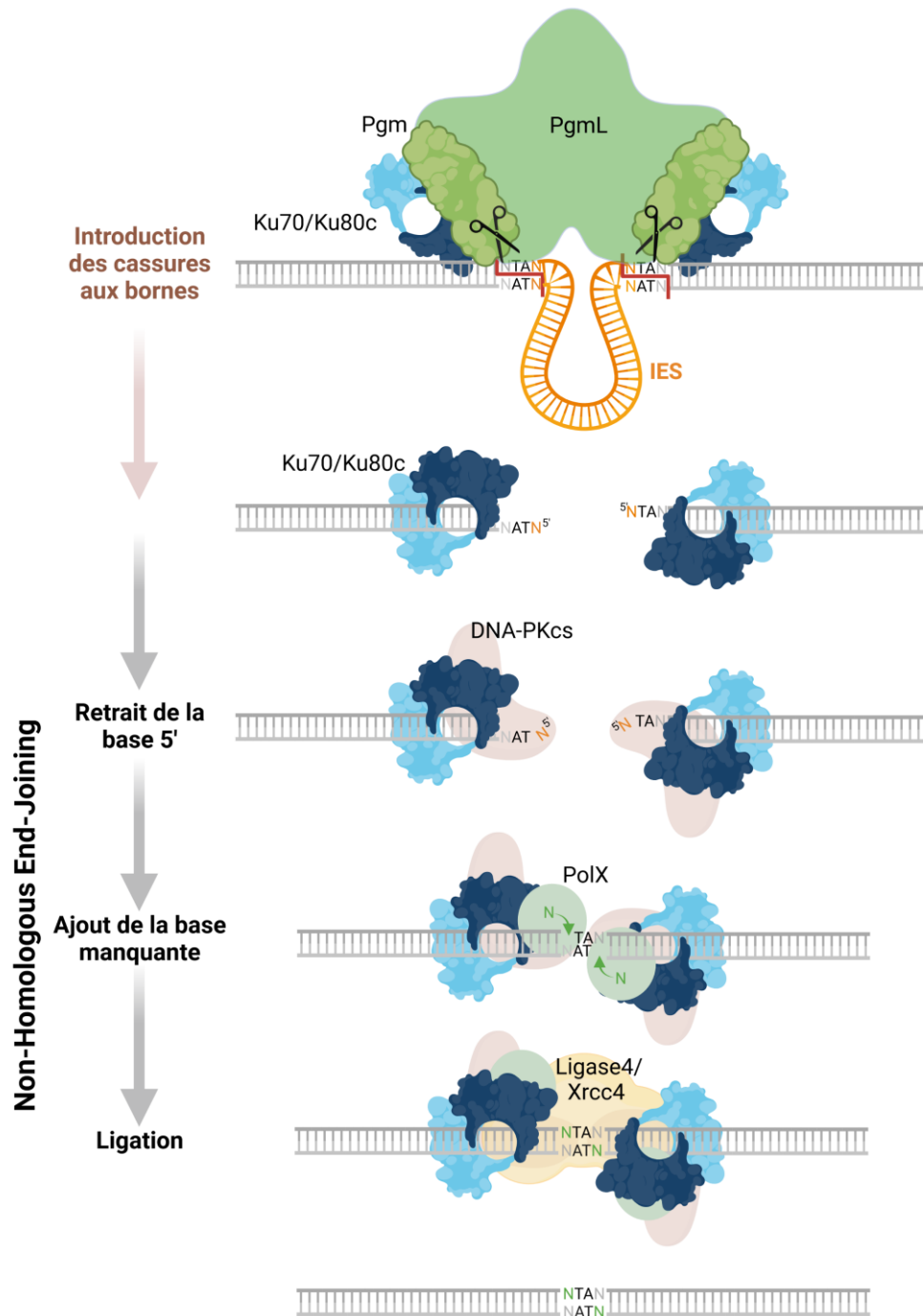


Figure 32 - Mécanisme d'excision des IES de *P. tetraurelia*

Après l'introduction des cassures aux bornes des IES par Pgm, l'hétérodimère Ku70/Ku80c recrute les acteurs du NHEJ (*Non-Homologous-End-Joining*). Dans un premier temps, la DNA-PKcs est recrutée au niveau de la cassure double-brin. La base en 5' est retirée par une nucléase qui reste à identifier. Les deux brins peuvent ensuite être réunis et s'apparier au niveau des TA. Une DNA-polymérase X (PolX) ajoute ensuite la base manquante. Le rapprochement des brins et/ou l'addition de la base manquante sont dépendants du complexe Ligase4/Xrcc4. Ce complexe effectue ensuite la ligation afin de réparer la cassure. (Created with BioRender.com)

La présence de Ku80c dans le complexe d'excision suggère qu'il existe un couplage entre l'introduction des cassures par les transposases domestiquées et la réparation de celles-ci par le NHEJ. Pgm ne peut introduire les cassures sans que la machinerie de réparation (ou au moins l'hétérodimère Ku70/K80) ne soit présente. Étant donné le nombre de cassures introduites dans le génome par Pgm pour éliminer les ~45 000 IES, on peut y voir un système de sécurité pour la cellule qui éviterait une fragmentation totale de son génome et des réarrangements aberrants (translocations).

3.2.4 L'élimination des IES chez *Tetrahymena*

3.2.4.1 Élimination imprécise des IES par Tpb2

Les IES de *T. thermophila* sont également excisées grâce à des transposases PiggyBac domestiquées. Tpb2 (*Tetrahymena PiggyBac like 2*) est l'orthologue de Pgm, elle catalyse l'excision des ~10 000 IES excisées de manière imprécise (Hamilton et al., 2016; Vogt and Mochizuki, 2013). L'étude de Tpb2 révèle une organisation très proche de celle de Pgm. Toutes deux possèdent le domaine RNaseH spécifique des transposases PiggyBac ainsi qu'un CRD, suivi d'un *coiled-coil* C-terminal. De plus les deux transposases domestiquées sont prédites comme étant capables de multimériser (Cheng et al., 2010; Dubois et al., 2017; Guérineau et al., 2021; Vogt and Mochizuki, 2013).

D'autre part, Lia5, une nouvelle transposase de la famille PiggyBac, pourrait être orthologue aux PgmL, a été identifiée chez *T. thermophila*. Comme les PgmL, Lia5 n'a pas conservé la triade catalytique mais reste nécessaire à l'activité de Tpb2 (Bischerour et al., 2018; Shieh and Chalker, 2013). La formation de complexes composés de plusieurs transposases domestiquées, l'une catalytiquement active et l'autre inactive semble donc être conservée également chez *T. thermophila*.

Les cassures introduites par Tpb2 sont prises en charge par les protéines Ku et ensuite réparées par le NHEJ. Cependant, contrairement à Pgm, l'hétérodimère Ku70/Ku80 ne semble pas indispensable à l'activité de coupure de Tpb2 (Lin et al., 2012).

3.2.4.2 Élimination précise des IES grâce au couplage avec le NHEJ

Ce sont deux autres transposases domestiquées orthologues à Pgm, Tpb1 et Tpb6, qui ont la charge de l'élimination précise des 12 IES situées dans les séquences codantes (Cheng et al., 2016; Feng et al., 2017).

Les cassures introduites par Tpb1 et Tpb6 sont réparées par le NHEJ, comme celles introduites par Tpb2. La particularité de ces deux transposases domestiquées réside dans la présence d'un domaine homologue à Ku dans leur séquence codante. Ces deux protéines seraient donc le résultat d'une fusion entre une PiggyBac domestiquée et une protéine Ku, poussant à l'extrême le couplage entre cassure et réparation de l'ADN (Cheng et al., 2016; Feng et al., 2017). Étant donné que ces deux protéines permettent l'élimination précise d'ADN chez *Tetrahymena*, on peut supposer que le couplage entre Pgm et l'hétérodimère Ku70/Ku80c participe également à la précision dans l'excision observée chez la paramécie.

3.3 DES CASSURES PRISE EN CHARGE PAR LE NHEJ

Étant donné le couplage existant entre Pgm et l'hétérodimère Ku70/Ku80c, il n'est aujourd'hui pas surprenant que les cassures double-brin de l'ADN générées par Pgm lors des RPG soient prises en charge et réparées par le NHEJ (Dubois et al., 2012; Marmignon et al., 2014). Cela a pourtant été longtemps remis en question étant donné que la voie du NHEJ est connue pour être un mécanisme introduisant des erreurs lors de la réparation des cassures double-brin (Ceccaldi et al., 2016). Cela contraste avec la précision avec laquelle les IES sont excisées. La précision de l'élimination d'ADN (et donc la stabilité du génome malgré les cassures double-brin) ne serait pas inhérente au NHEJ de la paramécie. Elle serait directement liée au couplage entre cassure et réparation (Figure 32), comme cela a pu être rapporté dans différents systèmes (Bétermier et al., 2020). De plus, le NHEJ ne semble pas mutagène par lui-même ; il introduirait des mutations lors de la ligation des extrémités incompatibles (Bétermier et al., 2014). La présence de TA aux bornes de toutes les IES, qui assure une complémentarité partielle des extrémités devant être liguées par le NHEJ, participe probablement aussi à la fidélité de la réparation.

Aujourd'hui, il est admis que les protéines Ku nécessaires à l'introduction des cassures recrutent par la suite la machinerie du NHEJ afin de réparer les cassures double-brin. Pgm introduit des cassures décalées centrées sur les TA laissant 4 pb sortantes en 5' (Gratias and Bétermier, 2003). L'hétérodimère Ku70/Ku80c recrute ensuite la protéine-kinase ADN-dépendante (DNA-PKcs) permettant ainsi la stabilisation des deux brins à ligaturer (Betermier and Duharcourt, 2014; Xu et al., 2020). La base en 5' est ensuite éliminée par une endonucléase qui n'a pas encore été identifiée. Les deux brins peuvent alors s'apparier et la base manquante en 3' peut être ajoutée par les ADN-polymérases de la famille X (PolX) (données non-publiées de l'équipe produites par Baptiste Verron et Julien Bischerour). La ligation s'effectue ensuite grâce au complexe Ligase IV/Xrcc4 (Kapusta et al., 2011).

Des ARNi dirigés contre *XRCC4* et *LIG4ab* ont montré que le complexe Ligase IV/Xrcc4 est nécessaire à l'étape de fermeture de la cassure double-brin introduite lors des RPG. Il semblerait cependant que l'étape d'addition de la base manquante avant la ligation nécessite aussi la présence des deux protéines (Kapusta et al., 2011). Une autre observation remarquable lors de l'ARNi contre *XRCC4* ou *LIG4ab* est une forte diminution des marquages au DAPI ou à l'iodure de propidium dans les MAC en développement, suggérant un défaut d'amplification de l'ADN (Abello et al., 2020; Kapusta et al., 2011). Il semblerait donc qu'il soit difficile pour la paramécie de continuer à endodupliquer son génome lorsque celui-ci est morcelé par l'activité du complexe d'excision. Cette observation met donc en avant l'importance d'autant plus grande du couplage entre cassure et réparation lors des RPG.

Projet : Comprendre comment le complexe d'excision reconnaît ses cibles

De nombreux acteurs nécessaires à la reconnaissance et à l'excision des IES sont aujourd'hui connus. Mais il reste encore de nombreuses questions en suspens. Par exemple, si on sait aujourd'hui que l'excision de ~72% des IES dépend de facteurs épigénétiques (marques d'histone et/ou ARNnc) (voir section 3.1), on ne sait toujours pas comment sont reconnues les quelques 12 500 IES restantes. Parmi ces IES, certaines portent une séquence consensus plus forte à leurs bornes (voir section 3.1.5). Mais on ne sait pas si cette séquence est suffisante à la reconnaissance et l'excision précise de ces IES, ni quels seraient les acteurs impliqués dans cette reconnaissance. Il reste également une zone d'ombre sur ~30% des IES dont l'excision dépend de PRC2, car le mécanisme guidant PRC2 sur ces ~10 000 IES n'est pas connu (voir section 3.1.4). De plus, le mécanisme de recrutement du complexe d'excision sur les marques déposées par PRC2 reste encore à établir.

Le mécanisme permettant la reconnaissance des séquences à éliminer de manière imprécise semble quant à lui plus clair que celui des IES puisqu'il est établi que ces séquences dépendent des marques déposées par PRC2 qui seraient dirigées sur ces cibles par le biais d'ARNscn (voir section 3.1.3.1). Cependant, puisqu'une partie de ces éléments est éliminée de façon indépendante de Pgm (voir section 2.2.1.3.2), il n'est toujours pas certain que ce soit Pgm qui introduise les cassures à leurs extrémités.

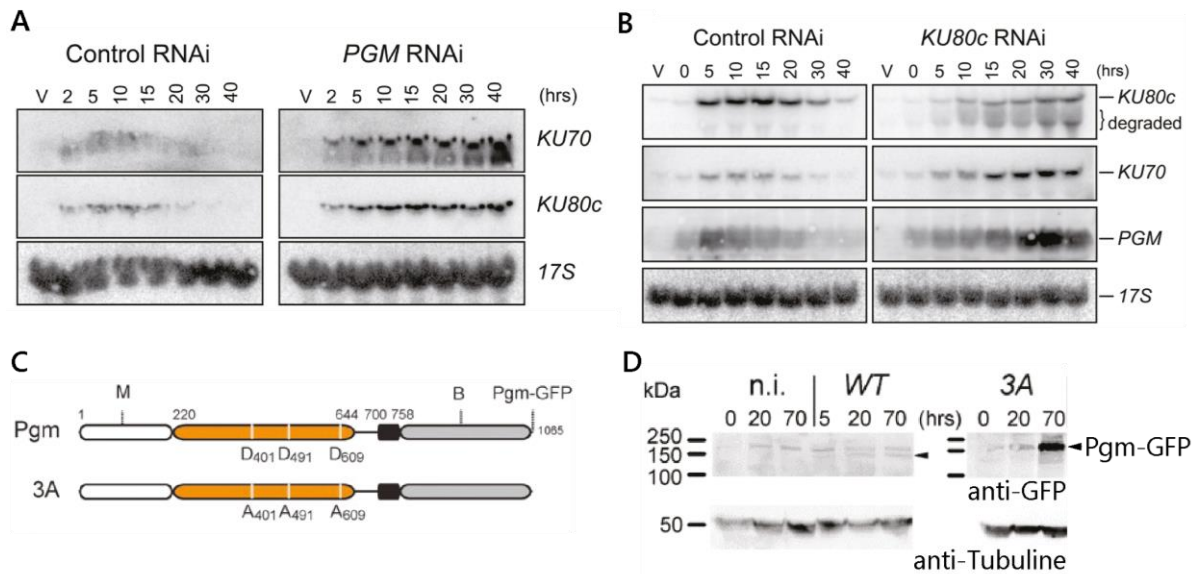


Figure 33 - Bloquer la machinerie d'excision induit une surexpression de *PGM* et *KU80c*

A Southern blot montrant l'accumulation des transcrits *KU70* et *KU80c* lors d'un ARNi contre *PGM* et **(B)** des transcrits *PGM* et *KU70* lors d'un ARNi contre *KU80c* (Marmignon et al., 2014). **C** Représentation des constructions utilisées lors de l'étude fonctionnelle des domaines de Pgm (figure adaptée de Dubois et al., 2017). **D** Western blot montrant l'expression sauvage (WT) de la fusion Pgm*-GFP ainsi que l'accumulation aberrante de Pgm-GFP lors de la complémentation d'un ARNi *PGM* par l'injection du mutant catalytique de Pgm fusionné à la GFP, PGM*3A-GFP (3A) (figure adaptée de Dubois et al., 2017).

Il y a donc d'un côté les IES pour lesquelles le mécanisme d'excision semble assez clair, mais pas la manière dont le complexe est guidé vers les séquences à reconnaître. D'un autre côté, se trouvent les éléments éliminés de façon imprécise pour lesquels, à l'inverse des IES, le mécanisme de reconnaissance semble clair, mais pas les acteurs de leur élimination.

IDENTIFIER DE NOUVEAUX ACTEURS

Une manière de mieux comprendre un mécanisme est d'en trouver les différents acteurs. J'ai donc cherché si des gènes avaient le même comportement que les acteurs connus dans des conditions sauvages et lorsque les RPG se déroulent mal. En effet, la plupart des gènes impliqués dans les RPG font partie des gènes des pics précoce ou intermédiaire (*early* ou *intermediate peak*, voir Tableau 24 et Tableau 25 en annexe). De plus, ils semblent dérégulés lorsque certains acteurs manquent, ce qui compromet le déroulement des RPG.

Les premières observations de dérégulation de la machinerie d'expression ont été faites par *northern blot* dans des conditions d'ARNi contre *PGM* et contre *KU80c* : réduire l'expression de *PGM* par ARNi conduit à une accumulation des transcrits *KU70* et *KU80c* (Figure 33A). Réciproquement, un ARNi dirigé contre *KU80c* induit une accumulation des transcrits *PGM* (et *KU70*) (Figure 33B) (Marmignon et al., 2014).

Par la suite, les tentatives de complémentation d'un ARNi dirigé contre *PGM* par un mutant catalytique de Pgm fusionné à la GFP (Figure 33C et D) ont permis de montrer que, si l'expression de la protéine mutante est induite correctement au moment des RPG, elle s'accumule en fort excès aux stades plus tardifs (Dubois et al., 2017).

De plus, le séquençage d'ADN extrait d'ébauches enrichies a permis de montrer que dans des conditions d'ARNi contre *PGM* (Guérin et al., 2017) et *KU80c* (Abello et al., 2020) l'excision de toutes les IES est bloquée. Il semblerait donc que retirer ou muter un composant de la machinerie cœur (le complexe nécessaire à l'excision de toutes les IES) induise la surexpression des autres facteurs nécessaires à l'introduction des cassures. Lors de ma thèse, j'ai recherché tous les

gènes dérégulés lorsque la machinerie d'excision est bloquée afin d'identifier de nouveaux acteurs potentiellement impliqués dans les RPG.

LOCALISER LE COMPLEXE D'EXCISION SUR LA CHROMATINE

Identifier de nouveaux acteurs ne permet que de donner des pistes pour trouver des nouvelles pièces du puzzle. Afin de savoir comment s'assemblent les différents éléments, il est également intéressant de visualiser la position du complexe d'excision sur la chromatine. De cette manière, on pourrait, d'une part, identifier la dynamique temporelle et les conditions de fixation de Pgm (et de ses partenaires) (voir section 3.2) sur ses cibles et, d'autre part, les comparer à la cinétique d'élimination des séquences d'ADN et de dépôt des marques d'histone. Cela permettrait de savoir quels facteurs suivent la chronologie d'élimination d'ADN et quels facteurs se fixent sur toutes les séquences sans forcément déclencher de coupure de l'ADN.

D'autre part, visualiser la position du complexe d'excision sur la chromatine permettrait de déterminer la nature de ses cibles. De cette manière, il serait possible de trancher sur l'implication de Pgm dans l'excision des éléments éliminés tardivement de manière imprécise (voir section 3.2).

Enfin, observer les variations de localisation du complexe sur la chromatine en fonction des acteurs présents permettrait de déterminer les conditions de liaison du complexe sur ses cibles. On pourrait également identifier le mode d'action des facteurs nécessaires à l'excision des différentes classes d'IES : permettent-ils de guider le complexe d'excision sur ses cibles ? ou activent-ils localement le complexe ?

Dans une deuxième partie de ma thèse, j'ai donc cherché à mettre au point une technique permettant de localiser Pgm (ou ses partenaires) sur la chromatine : le *Cut&Tag* (Henikoff et al., 2020; Kaya-Okur et al., 2019). Cette technique n'ayant jamais été réalisée sur un cilié, j'ai adapté la technique à la chromatine de paramécie, très dense en nucléosomes. Une fois mise en place, cette technique permettra de suivre la localisation de Pgm et de nombreuses autres protéines du complexe des RPG.

Partie II – Résultats

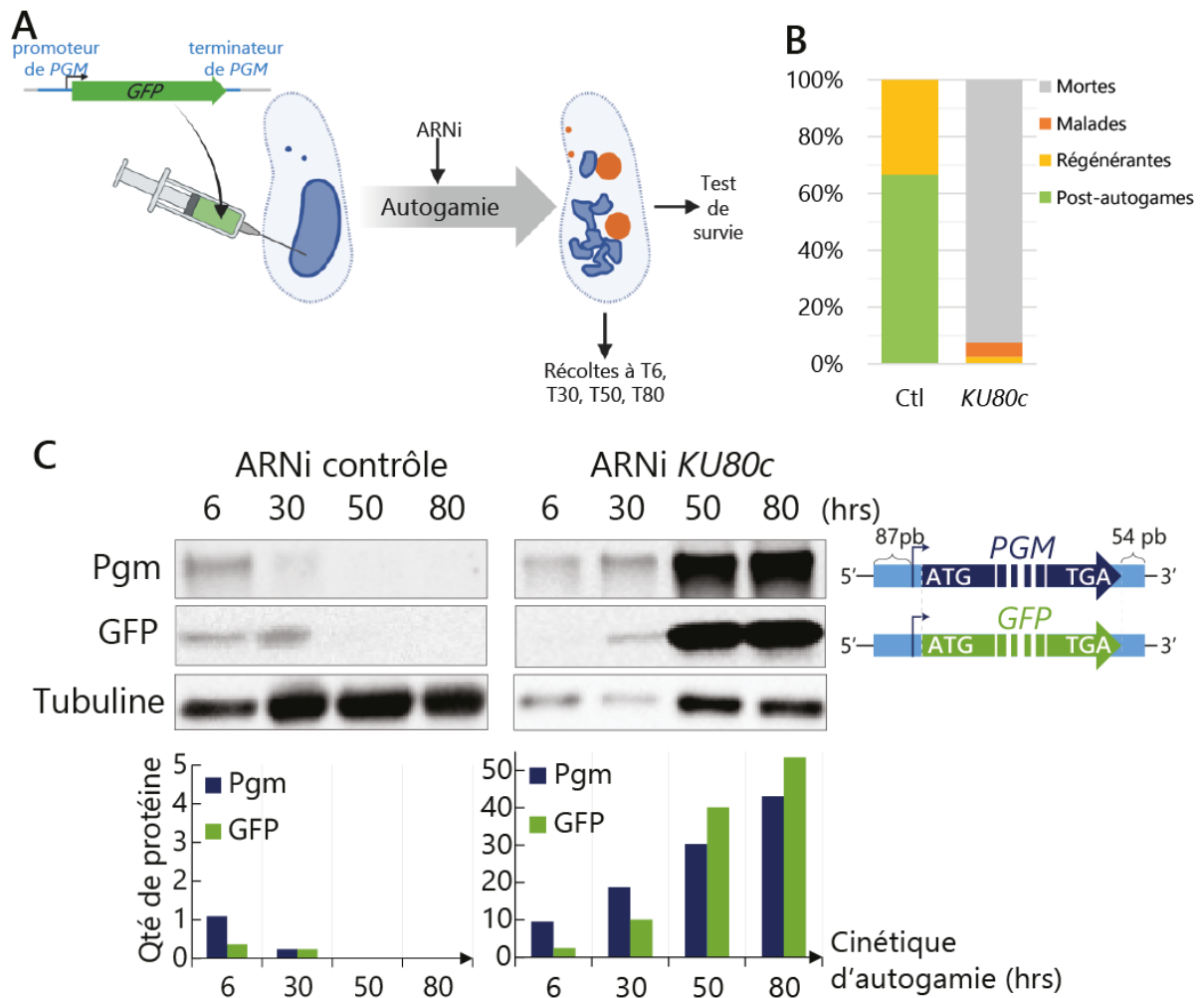


Figure 34 - Le rapporteur GFP reproduit fidèlement le profil d'expression de Pgm

A Schéma d'expérience d'injection de transgène *GFP* et récolte après passage de l'autogamie. (Created with Biorender.com) **B** Test de survie effectué après l'injection du transgène *GFP* (1976 cphg) et passage de l'autogamie dans des conditions d'ARNi contrôle (ctl) et *KU80c*. **C** Quantification par western blot de la quantité de Pgm et de GFP au cours de l'autogamie dans des conditions d'ARNi contrôle et lors d'un ARNi dirigé contre *KU80c*. Le temps est donné en heures suivant le T0.

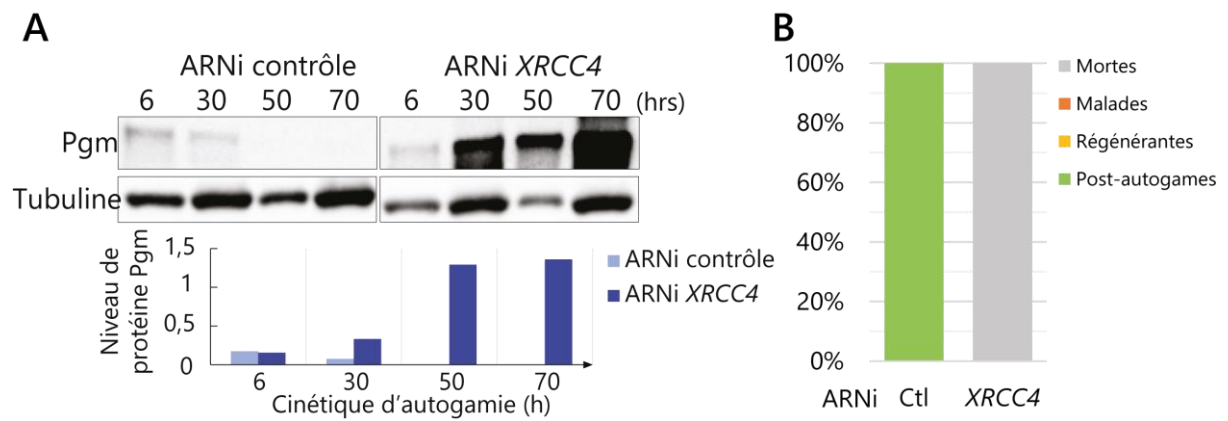
Chapitre 1 - Mise en évidence du dialogue entre les générations de MAC

1.1 BLOQUER L'EXCISION DES IES INDUIT UNE SUREXPRESSION DES GÈNES IMPLIQUÉS DANS LES RPG

Pgm et l'hétérodimère Ku70/Ku80c sont des composants essentiels à l'introduction des cassures double-brin aux bornes des IES (Baudry et al., 2009; Marmignon et al., 2014). En leur absence, l'ensemble des IES sont retenues dans le génome (Abello et al., 2020; Arnaiz et al., 2012). Les études précédentes ont permis de mettre en évidence que bloquer l'activité catalytique de Pgm ou empêcher le complexe d'excision de se former induit une surexpression tardive de plusieurs composants de la machinerie d'excision, avec une accumulation dans les MAC en développement (Dubois et al., 2017; Marmignon et al., 2014).

1.1.1 La dérégulation transcriptionnelle a lieu dans le MAC parental

Un transgène porteur du rapporteur *GFP* sous le contrôle des promoteur et terminateur de *PGM* est injecté en grand nombre de copies (1 976 copies per haploid genome, cphg) dans le MAC de cellules végétatives. On observe ensuite l'expression de ce rapporteur de l'expression de Pgm au cours de l'autogamie. L'expression du rapporteur *GFP* à partir des fragments de l'ancien MAC reproduit le profil d'expression de Pgm, que ce soit dans des conditions d'ARNi contrôle ou lorsque les RPG sont bloqués par un ARNi contre *KU80c* (Figure 34). La dérégulation des gènes lorsque les RPG sont bloqués semble donc bien avoir lieu dans l'ancien MAC.



1.1.2 Bloquer la réparation de l'ADN induit-elle une surexpression tardive de la machinerie d'excision ?

Bloquer l'introduction des cassures aux bornes des IES induit une accumulation de Pgm et de l'hétérodimère Ku70/Ku80c dans les nouveaux MAC. Afin d'examiner si interférer avec la réparation des cassures double-brin introduites par l'élimination des IES a le même effet sur l'expression de Pgm, j'ai effectué un ARNi contre *XRCC4* qui intervient à la dernière étape de la réparation des cassures. Un western blot effectué sur des extraits protéiques totaux prélevés à différents temps de l'autogamie dans ces conditions montre une surexpression tardive de la protéine Pgm (Figure 35). De plus, des observations préliminaires de microscopie ont montré que bloquer l'excision des IES à l'étape de réparation par un ARNi contre *XRCC4* (voir section 3.3) induit une accumulation de Ku80c dans les MAC en développement (Abello et al., 2020). Il semblerait donc qu'un ARNi contre *XRCC4* induise également une surexpression des composants de la machinerie cœur.

Cependant, le séquençage d'ADN extrait d'ébauches obtenues après un ARNi contre *XRCC4* et purifiées par cytométrie en flux (FANS : *Fluorescence Assisted Nuclear Sorting*) révèle que 65% des IES sont retenues dans le génome (Figure 36A). Cette expérience ne permet donc pas de savoir si bloquer uniquement l'étape de réparation après l'excision des IES est suffisante pour induire une surexpression des composants de la machinerie cœur.

1.1.3 L'excision des IES tardives est bloquée par la rétention des plus précoces lors d'un ARNi contre *XRCC4*

Si on regarde la chronologie d'excision des 35% des IES non-retenues en ARNi contre *XRCC4*, on remarque que plus de la moitié d'entre elles sont des IES qui s'excisent normalement très précocement en conditions contrôle (Figure 36B). Cela montre que si les premières IES sont coupées à leurs bornes sans qu'il y ait réparation, l'excision des IES plus tardives est inhibée. Cela pourrait être une conséquence du couplage étroit existant entre cassure et réparation. En effet, l'équipe a montré que la présence de l'hétérodimère Ku70/Ku80c est nécessaire à l'introduction des cassures aux bornes des IES par Pgm (Abello et al., 2020; Marmignon et al., 2014). Puisque l'hétérodimère Ku70/K80c fait également partie

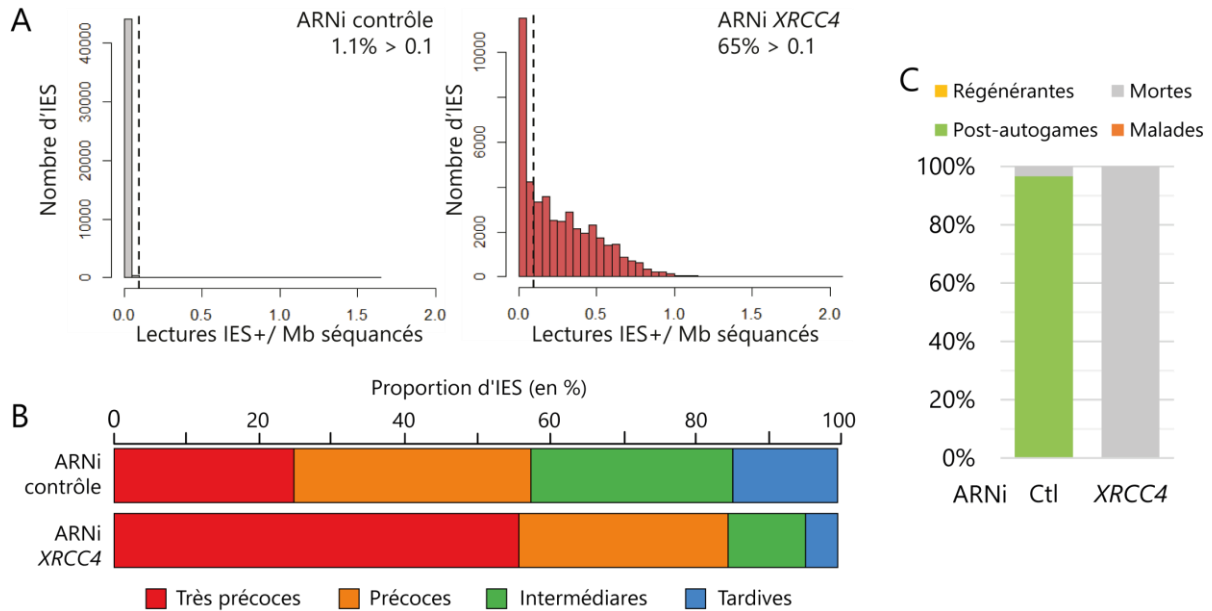


Figure 36 - Rétention des IES lors d'un ARNi contre XRCC4

A Des IES sont retenues lors d'un ARNi contre XRCC4. Les histogrammes représentent le nombre d'IES en fonction du nombre de lectures IES+ normalisées (intervalle de 0,5 pb) dans des conditions contrôle (gauche, en gris) et après un ARNi contre XRCC4 (droite, en rouge). Pour chaque condition, la proportion d'IES pour lesquelles on observe plus de 0,1 lecture IES+ par Mb séquencées est indiquée. Noter que les ordonnées sont différentes entre les deux histogrammes. **B** Les IES excisées (avec moins de 0,1 lecture IES+ par Mb séquencées) dans un ARNi contre XRCC4 sont enrichies en IES excisées très précocement en conditions normales (groupes définis dans Zangarelli et al., 2022). **C** Résultat du test de survie pour les cultures utilisées pour purifier les ébauches dont ont été extraits les ADN séquencés.

de la voie NHEJ de réparation de l'ADN, on peut supposer qu'il n'est relargué de l'ADN qu'après la réparation de la cassure au site d'excision alors que Pgm pourrait être libérée dès l'ADN coupé. La présence de Ku sur les cassures aux bornes des IES précoces pourrait empêcher de nouveau complexe d'excision de se fixer sur les bornes des IES plus tardives.

En résumé, lorsque l'excision des IES est bloquée par un ARNi contre *PGM*, *KU80c* ou *XRCC4*, un système de compensation se met en place, induisant la surexpression de composants connus de la machinerie d'excision des IES (Pgm et Ku80c). Ces observations ont conduit à l'hypothèse que bloquer l'induction ou la progression des réarrangements induirait la surexpression des gènes impliqués dans le processus. Il semblerait ainsi qu'il existe un dialogue entre les MAC en développement et les fragments de l'ancien MAC : En cas de défaut des RPG, les MAC en développement induiraient le maintien de la transcription dans les fragments de l'ancien MAC de gènes nécessaires aux réarrangements. Déterminer quels sont les gènes surexprimés dans ces conditions pourrait donc permettre l'identification de nouveaux gènes impliqués dans l'excision des IES.

1.2 LE PROGRAMME TRANSCRIPTIONNEL DE L'ANCIEN MAC EST DÉPENDANT DU DÉVELOPPEMENT DES NOUVEAUX NOYAUX

La plupart des gènes connus pour être impliqués dans les réarrangements programmés du génome, et plus particulièrement les gènes faisant partie de la machinerie cœur d'excision, font partie des gènes du pic intermédiaire (Tableau 25 en annexe) (Arnaiz et al., 2017). J'ai mis en évidence ci-dessus que la transcription dans les fragments de l'ancien MAC est influencée par le déroulement des réarrangements programmés qui ont lieu dans les MAC en développement. Cela expliquerait que la transcription des gènes du pic intermédiaire s'arrête une fois les réarrangements achevés et que si ceux-ci ne vont pas à leur terme, la transcription est maintenue. Cependant, il reste à déterminer ce qui induit l'expression des gènes.

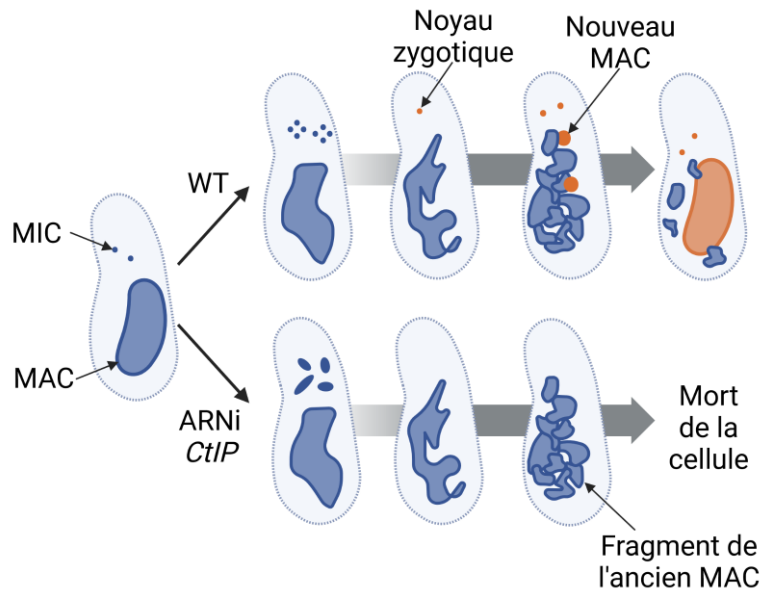


Figure 37 - Effet d'un ARNi dirigé contre *CtIP* sur le développement des noyaux au cours de l'autogamie

Représentation schématique du déroulement de l'autogamie dans des conditions sauvages (WT) et lors d'un ARNi contre *CtIP* (Godau et al., 2019). (Created with BioRender.com)

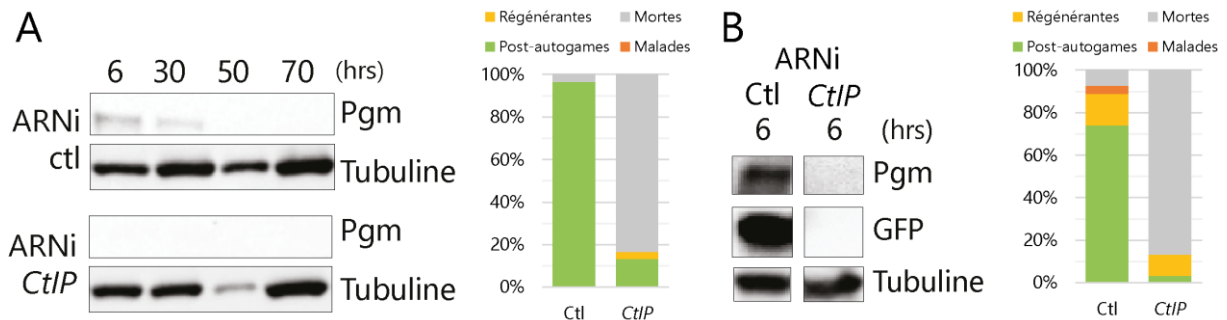


Figure 38 - Un ARNi contre *CtIP* conduit à une perte d'expression de Pgm

A (à gauche) Quantification par western blot de la quantité de Pgm au cours de l'autogamie dans des conditions d'ARNi contrôle (ctl) et lors d'un ARNi dirigé contre *CtIP*. Le temps est donné en heures suivant le T0. (à droite) Test de survie correspondant à la culture ayant permis de faire le western blot. **B** (à gauche) Quantification par western blot de la quantité de Pgm et de la GFP (416 cphg) présents 6h après le T0 de l'autogamie dans des conditions d'ARNi contrôle (ctl) et lors d'un ARNi dirigé contre *CtIP*. (à droite) Test de survie correspondant à la culture ayant permis de faire le western blot.

1.2.1 L'expression dans le MAC parental est liée au développement du noyau zygotique

Puisque ces gènes sont nécessaires aux RPG dans les nouveaux MAC en développement, le début des réarrangements ne peut déclencher leur activation. J'ai donc cherché si la transcription de ces gènes était dépendante du développement des nouveaux MAC. Pour cela, j'ai ciblé *CtIP* par ARNi. *CtIP* est un gène essentiel au maintien de l'intégrité du génome dans le MIC. En effet, il permet la réparation des cassures double-brin introduites par Spo11 au cours de la méiose (Godau et al., 2019). En l'absence de *CtIP*, la méiose des MIC est lourdement impactée empêchant la formation du noyau zygotique d'où sont originaires les nouveaux MIC et MAC. Il n'y a donc pas de formation de nouveaux MAC lors d'un ARNi contre *CtIP* (Figure 37).

Des western blots sur des extraits protéiques prélevés lors d'une cinétique d'autogamie en condition d'ARNi contre *CtIP* (Figure 38A) montre une absence de la protéine Pgm. De plus, l'utilisation d'un rapporteur *GFP* injecté dans le MAC (Figure 38B) permet de montrer que l'absence de développement du noyau zygotique empêche l'induction de la transcription de *PGM* dans les fragments de l'ancien MAC. Les observations du comportement de *PGM* dans ces conditions permettent donc d'ajouter une étape au modèle précédemment décrit.

1.3 MODÈLE DE DIALOGUE ENTRE L'ANCIEN ET LES NOUVEAUX MAC

La transcription de *PGM* dans les fragments de l'ancien MAC serait dépendante du développement du noyau zygotique (ou des nouveaux MIC ou MAC). Une fois le gène *PGM* transcrit dans les fragments, la protéine Pgm est produite dans le cytoplasme, puis se localise dans les MAC en développement pour y réarranger le génome. *PGM* reste exprimé dans les fragments tant que les réarrangements dans les MAC en développement ne sont pas achevés. Mais ce comportement peut-il être généralisé aux autres composants de la machinerie cœur ? Si oui, cela peut être une caractéristique supplémentaire permettant d'identifier de nouveaux facteurs impliqués dans les réarrangements programmés du génome.

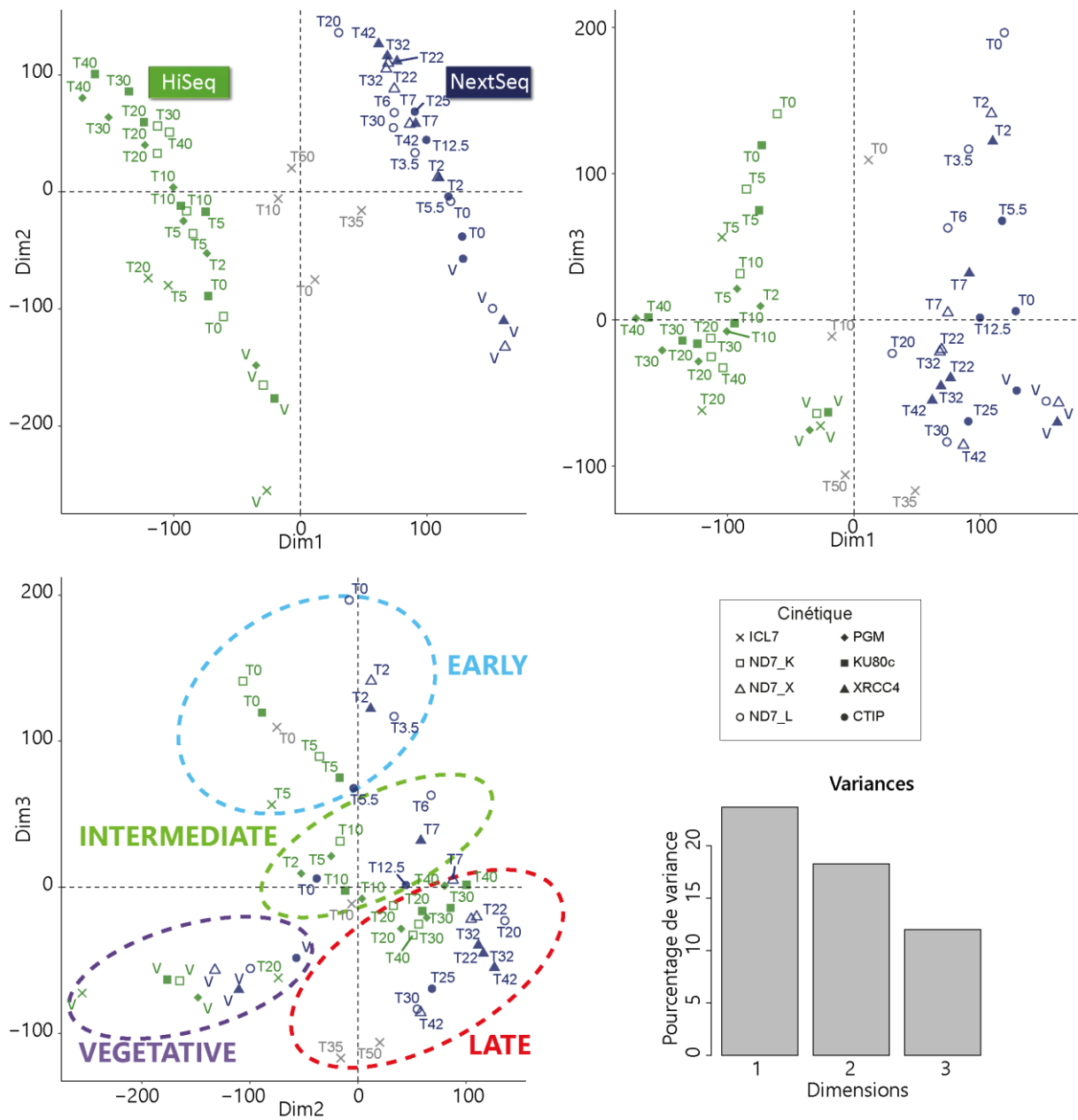


Figure 39 - Analyse en composantes principales (ACP) des données avant correction par ComBat-seq

Représentation des intersections des trois premières dimensions des ACP des points des cinétiques normalisés avant la correction par ComBat-seq. Les points des cinétiques colorés en vert correspondent aux échantillons séquencés par la méthode HiSeq. Ceux représentés en bleu ont été séquencés par la méthode NextSeq. Enfin, les points en gris correspondent aux échantillons qui ont été séquencés par les deux méthodes et collapés en tant que réplcats techniques. Sur la dernière ACP, sont également encadrés en pointillés, des groupes de points correspondant à des mêmes stades d'autogamie : Végétatif (*VEGETATIVE*), en violet ; Précoce (*EARLY*), en bleu clair ; Intermédiaire (*INTERMEDIATE*), en vert clair ; Tardif (*LATE*), en rouge.

Chapitre 2 - Identification de gènes candidats potentiellement impliqués dans les RPG

Précédemment j'ai mis en évidence que des gènes connus impliqués dans les RPG restent induits tant que ces derniers ne sont pas achevés. De plus, il semblerait que la transcription de ces gènes dépende de la formation du noyau zygotique. Dans cette partie, j'ai donc cherché à généraliser cette approche afin d'identifier de nouveaux candidats intéressants pour les RPG. Afin de généraliser l'observation des dérégulations, j'ai étudié des données de séquençage d'ARN prélevés à différents temps au cours de l'autogamie pour identifier des gènes différentiellement exprimés (ou dérégulés) dans trois ARNi différents perturbant les RPG (contre *PGM*, *KU80c* et *XRCC4*) par rapport aux cinétiques contrôles.

2.1 UN PROBLÈME DE GÉNÉRATION DE SÉQUENÇAGE

Les données à partir desquelles j'ai travaillé proviennent de plusieurs laboratoires et les échantillons n'ont pas été séquencés dans les mêmes conditions. En effet, les cinétiques ND7_K, PGM et KU80c (Marmignon et al., 2014) et les cinétiques ICL7 et EZL1 (Frapporti et al., 2019) ont été séquencés avec la technologie HiSeq d'Illumina. Alors que les cinétiques ND7_C, ND7_X, XRCC4 (Kapusta et al., 2011) et CTIP (Godau et al., 2019) et quelques points des cinétiques ICL7 et EZL1 (Miró-Pina et al., 2022) ont été séquencés plus récemment en utilisant cette fois la technologie NextSeq d'Illumina.

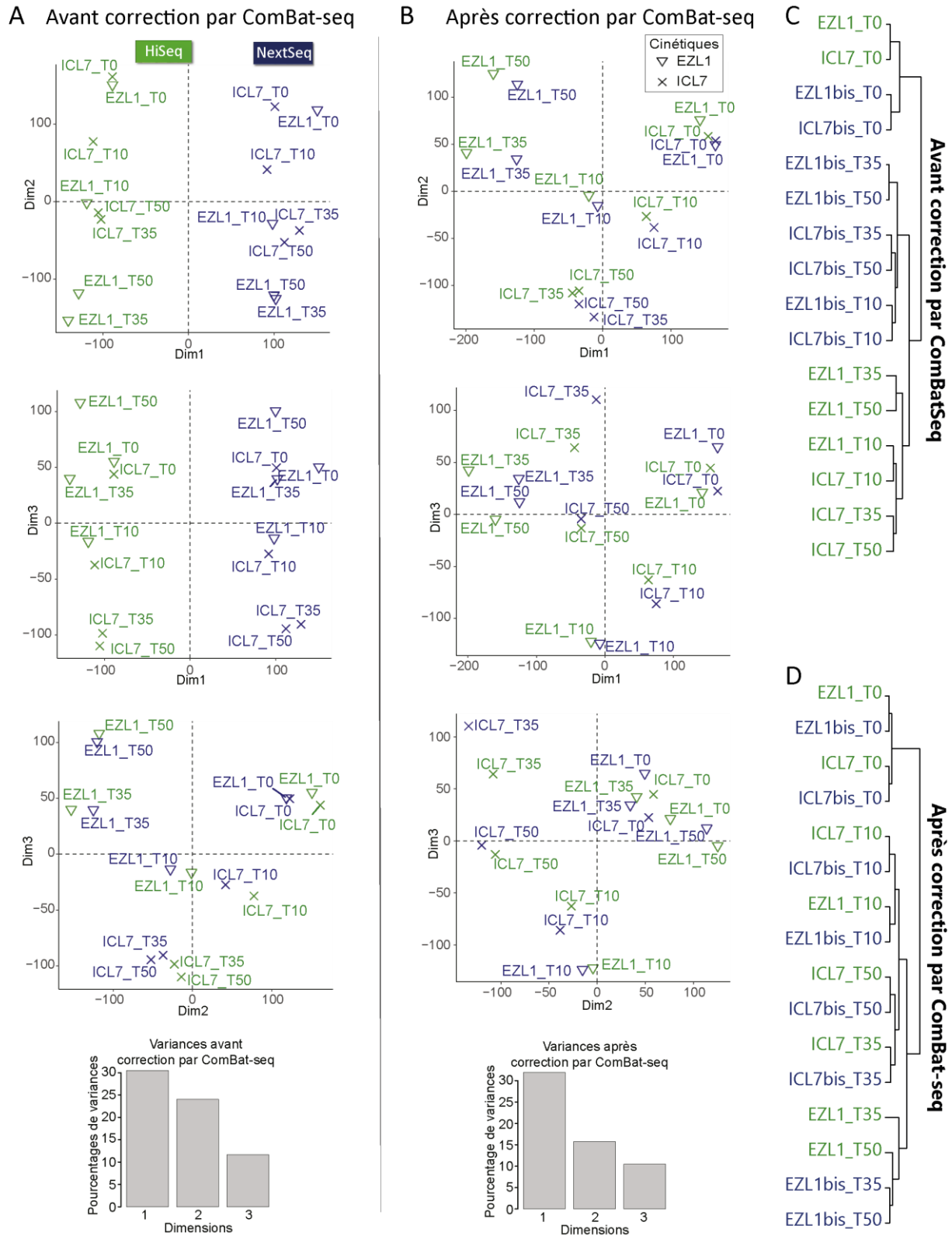


Figure 40 - Validation de la correction par ComBat-seq

Résultat des ACP des *clustering* hiérarchiques pour les échantillons séquencés par HiSeq (Lhuillier-Akakpo et al., 2014) (en vert) et en NextSeq (Miró-Pina et al., 2022) (en bleu) avant la correction par ComBat-seq (A et C) et après cette correction (B et D).

Il s'avère que lorsque l'on visualise ces données par une analyse en composantes principales (ACP), la première dimension discrimine les données uniquement selon la technologie de séquençage utilisée (Figure 39). Il est donc impossible d'analyser directement toutes les cinétiques en même temps, pour cela, il faut d'abord corriger l'effet *batch*. J'ai donc corrigé les données à ma disposition en utilisant le package R : ComBat-seq (Zhang et al., 2020).

2.1.1 Validation de l'outil ComBat-seq

Afin de vérifier la correction des données, j'ai utilisé les échantillons des cinétiques ICL7 et EZL1 (Frapporti et al., 2019) qui ont été séquencés à la fois en HiSeq (Lhuillier-Akakpo et al., 2014) et en NextSeq (Miró-Pina et al., 2022). Les ACP avant la correction par ComBat-seq (Figure 40A et B) mettent en évidence que, comme pour mes échantillons, la première dimension sépare les échantillons en fonction de la méthode de séquençage. Cependant, la seconde dimension semble permettre de séparer les échantillons selon le temps de la cinétique. Enfin, la troisième dimension tend à séparer l'ARNi contrôle ICL7 de l'ARNi contre EZL1, elle transcrit donc l'effet de l'ARNi sur l'expression des gènes lors de ces cinétiques. De ce fait, lorsque l'on représente les 2^{ème} et 3^{ème} dimensions de l'ACP, les répliquats techniques et les échantillons prélevés au même moment de la cinétique se regroupent (hormis pour les points après T35 qui se séparent dû à l'effet de l'ARNi). Il semblerait donc que la correction de l'effet *batch* soit possible.

2.1.1.1 Comparaison des échantillons par ACP

La correction par ComBat-seq permet de rapprocher les répliquats techniques sur les ACP. Il semblerait donc que cette correction soit efficace. De plus, on peut remarquer une forte ressemblance entre le graphique des dimensions 2 et 3 avant la correction (Figure 39) et celui des dimensions 1 et 2 après la correction (Figure 41). La correction de l'effet batch ne semble donc pas avoir altéré les données, mais simplement décalé les dimensions identifiées en ACP en supprimant la première dimension qui était principalement due à la méthode de séquençage. Les points proches des cinétiques d'autogamie peuvent donc être rapprochés pour former des pseudo-répliquats en utilisant les 1^{ère} et 2^{ème} dimensions de l'ACP.

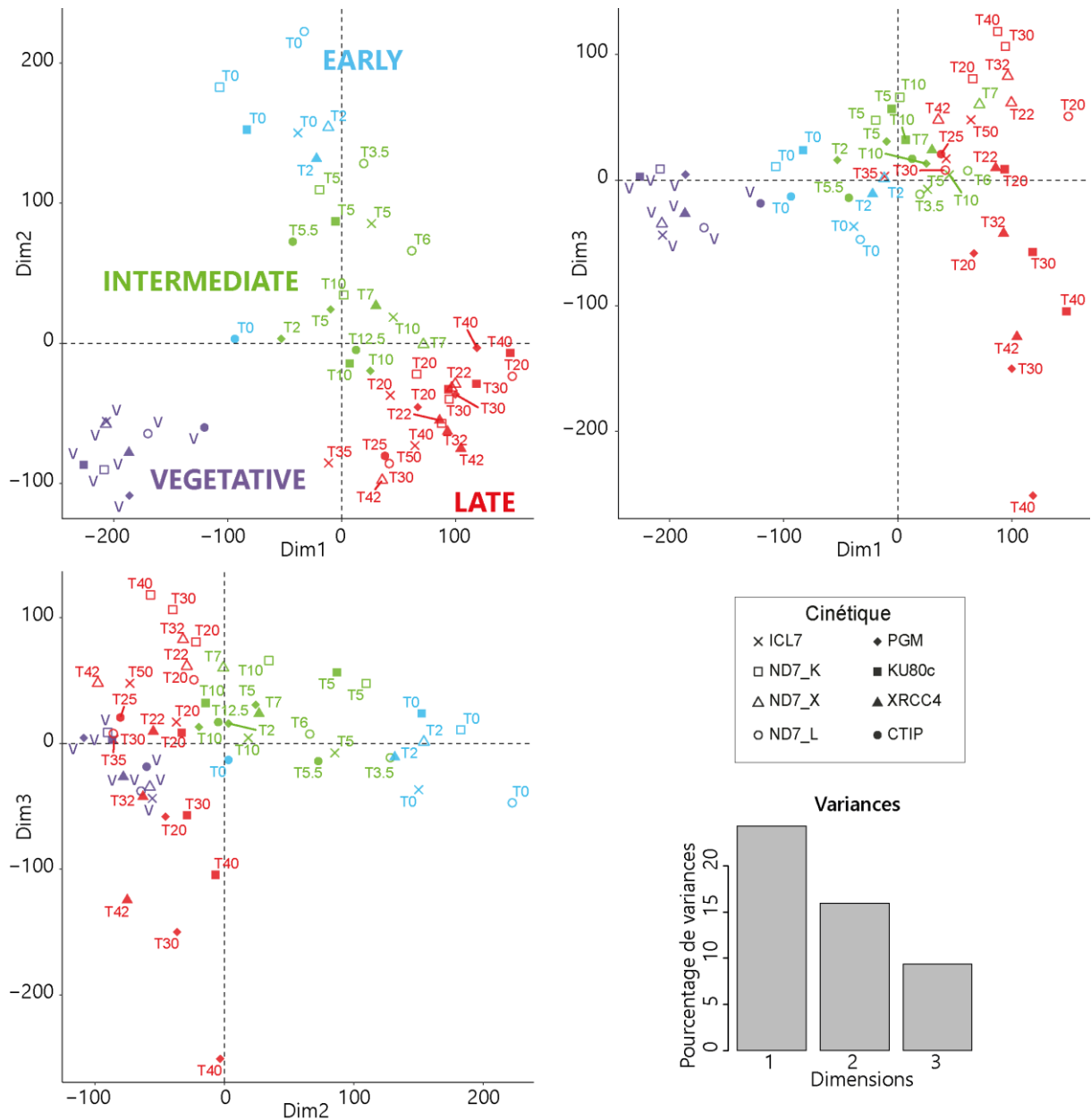


Figure 41 - ACP des données après correction par ComBat-seq

Représentation des intersections des trois premières dimensions des ACP des points des cinétiques après la correction par ComBat-seq et normalisation. Les points des cinétiques sont colorés en fonction du stade de l'autogamie auxquels ils appartiennent : végétatif (*VEGETATIVE*), en violet ; précoce (*EARLY*), en bleu clair ; intermédiaire (*INTERMEDIATE*), en vert clair ; tardif (*LATE*), en rouge.

2.1.1.2 Validation de la correction par *clustering* hiérarchique

Afin de valider la correction, j'ai également comparé les données avant et après correction par ComBat-seq via un calcul des distances de Pearson représentées sur un *clustering* hiérarchique.

Le *clustering* hiérarchique avant la correction (Figure 40C) montre que les échantillons se regroupent majoritairement selon la méthode de séquençage. Cependant, une fois la correction appliquée, les échantillons groupent avec leur réplicats techniques (Figure 40D). La correction semble donc être efficace. On peut cependant remarquer que la correction est moins efficace pour les échantillons tardifs de la cinétique EZL1 comme cela avait été constaté sur les ACP. Malgré cela, les T0 et T10 des cinétiques EZL1 et ICL7 groupent ensemble. Tous les temps tardifs de chacune des cinétiques sont également regroupés. On s'attend en effet à ce que la cinétique contrôle, ICL7, diverge de la cinétique pour les temps tardifs car l'effet de l'ARNi contre *EZL1* sur l'expression des gènes est plus fort.

2.1.2 Définition des groupes de réplicats

Après avoir corrigé les données pour le biais de séquençage, je les ai analysées ensemble. Cependant, les cinétiques en condition d'ARNi à ma disposition n'ayant pas de réplicats biologiques, j'ai défini des groupes de points proches les uns des autres dans chaque cinétique qui seront considérés comme des réplicats lors de l'analyse des gènes différentiellement exprimés. C'est ce que je considère ici comme étant des pseudo-réplicats.

Afin de définir les pseudo-réplicats, j'ai décidé de m'appuyer d'abord sur les ACP (Figure 41). Les ACP permettent de grouper les points par stade de la cinétique d'autogamie, hormis pour le T0 de la cinétique CTIP. Les points végétatifs se détachent bien des autres points de la cinétique. On observe ensuite une progression des temps précoces vers les temps tardifs. Cela reflète le processus continu qu'est l'autogamie mais complique la définition des stades. De plus, sur les ACP, on peut remarquer que la 3^{ème} dimension reflète la dérégulation au temps tardif puisque les cinétiques contrôles s'éloignent des cinétiques dans les différentes conditions d'ARNi perturbant les RPG ou le développement des nouveaux MAC.

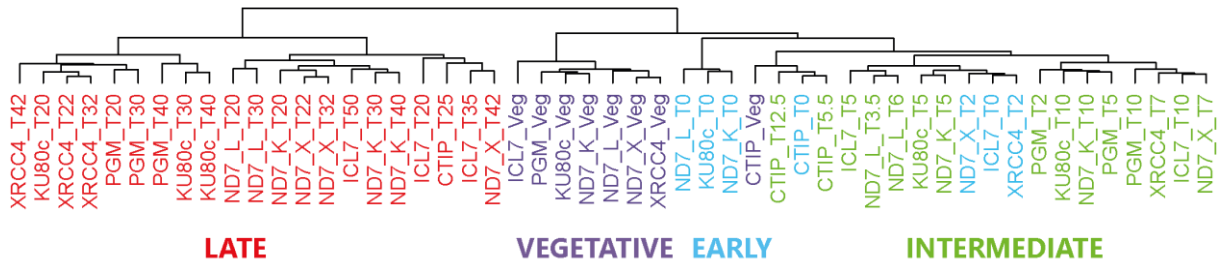


Figure 42 - Clustering hiérarchique

Clustering hiérarchique des données corrigées par ComBat-seq et normalisées, mettant en évidence les stades tardif (*LATE*) et végétatif (*VEGETATIVE*), la différence entre les stades précoce (*EARLY*) et intermédiaire (*INTERMEDIATE*) est moins claire sur cette représentation que sur les ACP. Les points de la cinétique CTIP sont également tous groupés ensemble (à l'exception du point tardif).

J'ai également réalisé un calcul des distances de Pearson représenté sous la forme d'un *clustering* hiérarchique (Figure 42) afin de vérifier si les points proches de la cinétique peuvent bien être considérés comme des pseudo-réplicats. La représentation hiérarchique permet de définir une frontière assez nette entre les stades intermédiaire et tardif. La frontière reste cependant assez floue entre les stades précoce et intermédiaire.

Afin de définir la frontière entre les stades précoce et intermédiaire, je me suis donc appuyée à la fois sur les ACP et le *clustering* hiérarchique mais également sur les données de cytologie en comparant les diagrammes des stades (voir Figure 81 de la partie Matériel et méthodes) pour valider que les groupes (Tableau 1) correspondent à la réalité biologique de l'état d'avancement de l'autogamie dans les échantillons.

De plus, le *clustering* hiérarchique montre que les points de la cinétique CTIP groupent entre eux mais pas avec les autres stades (contrairement à ce que l'on voit pour la plupart des points sur les ACP). Pour définir les stades de cette cinétique, je me suis donc appuyée sur l'ACP et les données de cytologie en comparant les diagrammes des stades des différents points des cinétiques (voir Figure 81 de la partie Matériel et méthodes). La définition finale des stades d'autogamie regroupant les pseudo-réplicats est résumée dans le Tableau 1.

Cinétiques Stade de l'autogamie	Contrôles				CTIP	XRCC4	PGM	KU80c
	ICL7	ND7_K	ND7_X	ND7_L				
VEGETATIVE	Veg	Veg	Veg	Veg	Veg	Veg	Veg	Veg
EARLY	T0	T0	T2	T0	T0	T2	-	T0
INTERMEDIATE	T5 - T10	T5 - T10	T7	T3.5 - T6	T5.5 - T12.5	T7	T2 - T5 - T10	T5 - T10
LATE	T20 - T35 - T50	T20 - T30 - T40	T22 - T32 - T42	T20 - T30	T25	T22 - T32 - T42	T20 - T30 - T40	T20 - T30 - T40

Tableau 1 - Définition des pseudo-réplicats

Pour toutes les conditions, les points dans une même case sont considérés comme des réplicats pour un même stade d'autogamie. Toutes les cinétiques contrôles sont ensuite utilisées comme réplicats biologiques les unes des autres. Enfin les quatre points (en gras) de la cinétique ICL7 qui ont été séquencés plusieurs fois sont considérés comme des réplicats techniques.

2.2 UN DIALOGUE INTERGÉNÉRATIONNEL ENTRE LES MAC ET IDENTIFICATION DE NOUVEAUX CANDIDATS AUX RPG (ARTICLE EN PRÉPARATION)

Je présente ici une première version du manuscrit sur mes résultats de thèse. Une discussion de ces résultats est également proposée dans le Chapitre 1 -de la partie Discussion de ce manuscrit.

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

Mélanie Bazin-Gélis 1, Evangelia Eleftheriou 1,2, Coralie Zangarelli 1, Linda Sperling 1, Olivier Arnaiz 1, Mireille Bétermier 1*

1. Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC),
91198, Gif-sur-Yvette, France
2. Present address: Institut Pasteur....

*To whom correspondence should be addressed: mireille.betermier@i2bc.paris-saclay.fr

ABSTRACT

Multinucleated cells are found in various eukaryotes but the mechanisms by which nuclei interact with each other and regulate their functions remain poorly understood. The ciliate *Paramecium tetraurelia* is a unicellular organism harboring two germline micronuclei (MIC) and a somatic macronucleus (MAC). At each sexual cycle, the parental MAC is progressively destroyed while new MACs differentiate from the MIC. During new MAC development, the parental MAC ensures most of gene expression. Concomitantly, programmed DNA elimination (PDE) streamlines the new MAC genome for gene expression by removing 45000 Internal Eliminated Sequences (IESs) that interrupt half of the genes.

Using RNA sequencing, we show that 628 genes (out of 41533) are up-deregulated when essential IES excision genes are knocked down, suggesting that impairing PDE in the new MAC triggers a compensatory gene up-deregulation in the parental MAC. Moreover, we found that the transcription of 2365 genes, including 42% of the up-deregulated genes, depends on the presence of new MACs. We conclude that a regulatory crosstalk is established between the two generations of somatic nuclei that coexist in the cell.

Moreover, many known PDE genes are found among the up-deregulated genes, suggesting that this set of genes includes novel actors of PDE.

INTRODUCTION

Eukaryotic cells generally harbor a single nucleus that serves two major functions: gene expression and the transmission of genetic information to post-mitotic daughter cells or post-meiotic gametic cells. However, many examples of multinucleate cells have been documented (reviewed in (1)), for instance the muscular fiber of animals or the syncytium of filamentous fungi. In these giant cells, in situ hybridization experiments or, more recently, single-nucleus RNA sequencing data have pointed to a heterogeneous transcriptional behavior among individual nuclei or the nuclear subtypes that occupy different regions of the cytoplasm (2–4). Whether and how a crosstalk is established between the different nuclei of a multinucleate cell to coordinate their transcriptional activities still appears to be poorly known.

Ciliates are binucleate unicellular eukaryotes, in which nuclear germline and somatic functions are distributed between distinct types of nuclei (5). All gene transcription takes place in the somatic macronucleus (MAC), a highly polyploid nucleus that is essential for cell survival. The germline micronucleus (MIC, $2n$) is transcriptionally silent, but plays a crucial role in sexual processes: it undergoes meiosis and, following fertilization and karyogamy, transmits the germline genome to the zygotic nucleus of the next generation. At each sexual cycle, the old (parental) MAC is destroyed and, in the same cell, new MACs develop anew from mitotic copies of the zygotic nucleus. MAC development has been studied in several model ciliates, including *Paramecium* (6). In *Paramecium tetraurelia*, two sexual processes have been described: conjugation between reactive partners with compatible mating types, and autogamy, a self-fertilization that takes place in a single cell. At the start of each sexual cycle, MIC meiosis is accompanied by the fragmentation of the old MAC into ~30 fragments

that persist in the cytoplasm and continue to be transcribed throughout new MAC development (7). Gene transcription, therefore, progressively switches from the old to the new MACs to ensure progeny survival when vegetative growth resumes and old MAC fragments are eventually lost. A previous analysis of differential gene expression established that the expression pattern of 17190 genes (out of 41533) follows a developmental program during autogamy (8), with six clusters of differentially expressed genes. An early peak of expression takes place during MIC meiosis and old MAC fragmentation. During MAC development, intermediate and late peaks of gene expression precede a late induction cluster. The gene expression program also includes early and late repression clusters.

In *P. tetraurelia*, MAC development involves successive rounds of genome endoduplication to reach the final ploidy ($\sim 1000n$) of the mature new MAC (7). Most of the Programmed DNA Elimination (PDE) starts when the new MAC genome has reached a $32n$ ploidy and extends over several consecutive endoduplication rounds (9). As a whole, PDE removes 30 to 50% of germline DNA (reviewed in (6)), reducing the haploid genome complexity from 100 to 150 Mbp in the MIC (10) to 72 to 76 Mbp in the MAC (11, 12). Eliminated sequences include multiple DNA repeats (transposable elements (TE), minisatellites) that are removed in an imprecise manner, leading to the formation of variable MAC chromosome ends or internal deletions (13). In addition, around 45000 Internal Eliminated Sequences (IES), mostly short, non-coding and originating from ancient TEs, are excised from new MAC chromosomes (10, 14) according to a defined temporal program (9). IES ends define a poorly conserved and degenerate consensus sequence, which includes the conserved TA dinucleotides that are present at their boundaries with flanking DNA (14, 15). Since 47% of genes are interrupted by at least one IES in the MIC, precise IES excision is necessary to ensure that functional genes are assembled in the new MAC. Thus, PDE not only shapes the structure of MAC

chromosomes but also rearranges the somatic genome to make it suitable for gene expression.

Numerous studies have focused on the identification of essential actors involved in PDE. A first important line of research has unraveled the role played by non-coding RNAs in specifying the elimination of germline-restricted sequences from the developing MAC. MIC-specific small RNAs, called scnRNAs, produced during the meiosis of the parental MICs, are selected through genomic subtraction against transcripts from the rearranged genome of the old MAC (16, 17). ScnRNAs have been proposed to drive the Ezl1 histone-methyltransferase to their homologous sequences through pairing to nascent transcripts in the new developing MAC (18), which triggers the deposition of H3K9me3 and H3K27me3 histone modifications on associated nucleosomes (19, 20). Ezl1-dependent heterochromatin formation is necessary for the elimination of repeated DNA and ~75% of IESs (21). Other small RNAs, the iesRNAs, are produced from the first excised IES molecules and stimulate excision of the remaining copies that have not yet been excised (22). Of note, 25% of IESs, which exhibit stronger nucleotide sequence signals at their ends, are excised with high efficiency and precision, independently of non-coding RNAs and histone marks (9). Additional factors, including the chromatin remodeling complexes FACT (23) and ISWI (24), may control the accessibility of eliminated DNA to the enzymatic machinery that carries out PDE.

A second line of research has aimed at identifying key components of the core IES excision machinery, which are also required for most of imprecise DNA elimination (12). A complex of six domesticated PiggyBac transposases, the catalytically active PiggyMac (Pgm) subunit (25, 26) and its five Pgm-like (PgmL) partners (27) has been proposed to form the endonuclease

that introduces DNA double-strand breaks (DSB) at IES ends (28). Once IESs are removed, broken excision sites are repaired through the non-homologous end joining (NHEJ) pathway and, following limited DNA end-processing, the NHEJ-specific LigaseIV-Xrcc4 complex stitches together the two flanking chromosome ends (29). In addition to playing its expected role in DSB repair, the specialized Ku70/Ku80c heterodimer appears to be part of the DNA cleavage complex: it interacts with Pgm, stabilizes its localization in the developing new MAC and is indispensable for DSB introduction at IES ends (30, 31).

Most of the proteins that were shown to participate in the epigenetic control or catalysis of PDE are encoded by genes from the early, intermediate or late peaks, which are switched off at the end of autogamy in control conditions (8, 32). Previous observations suggested that depleting essential actors of PDE may perturb this gene expression program (30). Indeed, northern blot experiments revealed that *KU80c* and *KU70* mRNAs, which are expressed in the intermediate peak, accumulate continuously until late stages of MAC development in a *PGM* knockdown (KD). Reciprocally, *KU70* and *PGM* mRNAs are up-deregulated in a *KU80c* KD. To generalize these observations, we studied the *P. tetraurelia* transcriptome during autogamy of cells, in which three different genes involved in PDE were knocked down independently. We report that impairing the normal course of PDE in the new MAC results in up-deregulation of 628 genes, among which we found many known PDE genes, suggesting that the list may contain other important PDE genes. We also show that subsets of genes from the early, intermediate and late peaks are no longer expressed when new MACs fail to develop. Based on these findings, we propose that a crosstalk between the two generations of MACs that coexist in autogamous cells links the transcription program in the old MAC to the progression of PDE in the new MAC.

MATERIALS AND METHODS

Paramecium strains and culture conditions

Paramecium tetraurelia 51 new (28) or its mutant derivative 51 *nd7-1* (26) were grown at 27°C in a standard wheat grass infusion medium inoculated with *Klebsiella pneumoniae* and supplemented with 0.8 µg/mL β-sitosterol and 100 µg/mL ampicillin (33). Autogamy was triggered by starvation (34) and the progression of old MAC fragmentation was monitored using DAPI staining of fixed cells (31). In all autogamy time-courses, the T0 time-point is defined as the time when 50% of cells have fragmented their MAC.

Microinjection of the *GFP*-reporter transgene

All *GFP*-reporter transgenes were constructed from the *PGM-GFP* and *GFP-PGM* constructs published in (26). In plasmid p0349, the *GFP* gene was cloned under the control of *PGM* transcription signals (96 bp upstream of the start codon and 54 bp downstream of the stop codon of the *PGM* gene) (Figure 1A). The wildtype promoter of p0349 was then modified using synthetic DNA fragments (from Integrated DNA Technologies) to give plasmids p0364 and p0418, carrying a mutant or an inverted motif in the *PGM* promoter, respectively. Plasmid maps and sequences are available upon request. All plasmids were linearized by *Bsa*AI and co-injected into the MAC of vegetative *P. tetraurelia* 51 *nd7-1* with a 10-fold lower amount (1:10 mass ratio) of a linearized ND7-complementing plasmid, as described in (26). Transformants were screened for trichocyst discharge (trich+) after 2 to 3 cell divisions. Transformed cells were grown for 8-10 vegetative divisions in standard conditions and their genomic DNA was extracted using the NucleoSpin Tissue kit (Macherey-Nagel). Transgene injection levels (copy number per haploid genome: cphg) were determined using real-time

qPCR. Oligonucleotide primers used to amplify *PGM* and *GFP*, with *KU80c* as a genomic reference are listed in Supplementary Table S1. Since there is no endogenous *GFP* gene in *Paramecium*, we used genomic DNA extracted from cells transformed by the *PGM-GFP* fusion as a reference, with a $cphg=6$ based on Southern blot hybridization (26) that was confirmed by qPCR using *PGM* primers. This DNA was used as a *GFP* standard in all further qPCR reactions. For all *GFP* transformants, we calculated the apparent $cphg$ from the C_p value provided by real-time qPCR:

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

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

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

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

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

RNAi-mediated gene knockdowns (KD) during autogamy

Double-strand RNA (dsRNA)-induced RNAi was carried out using the feeding procedure (35).

Transformed or non-injected *Paramecium* cells were grown in standard *K. pneumoniae* medium for 4 vegetative divisions, then transferred to wheat grass infusion medium containing plasmid-free *Escherichia coli* HT115 for 4-5 divisions (36). RNAi was induced as described in (26). Each culture was split and fed for 4-5 divisions with non-induced HT115 containing RNAi plasmids (against target or control genes), then with dsRNA-producing

HT115 for 10-15 more divisions before cells starve and start autogamy. Final volumes were 200 mL for medium-scale or 1 L for large-scale experiments. Three days after autogamy has started, the ability of sexual progeny to develop a functional new MAC was monitored by performing a survival assay as described (26) (Supplementary Figure S1).

Control experiments were performed using plasmid pOND7c, which targets RNAi against the non-essential *ND7* gene (37). The plasmids used for *PGM*, *KU80c*, *XRCC4* and *CtIPa* + *CtIPb* (= *CtIP*) RNAi were respectively pPBL49HN (25), pL4440-KU80C-2 (30), pXRCC4-R (29), pL4440-PtCtIPa and pL4440-PtCtIPb (38).

Cell collection and protein extraction for western blot analysis

50-mL aliquots (~1 to 2 x 10⁵ cells) were collected from medium-scale cultures at different autogamy time-points. Cells were washed for 20 min in Dryl's buffer (2 mM sodium citrate, 1 mM NaH₂PO₄, 1 mM Na₂HPO₄, 1 mM CaCl₂) before concentration in a final volume of ~40 µL. 20-µL aliquots were flash-frozen in liquid nitrogen and stored at -80°C. For western blots, 20 µL of frozen cells were lysed by a 3 min incubation at 100°C in 5% boiling SDS with 2x final concentration of Protease Inhibitor Cocktail Set 1 (Merck Chemicals). After centrifugation (5 min 15000g), Laemmli buffer (Bio-Rad) complemented with β-mercaptoethanol (354 mM final concentration) was added to the supernatant before a 3-min incubation at 100°C. Electrophoresis was performed in Criterion TGX Stain-free 4-15% precast SDS-Page gels (Bio-Rad) in migration buffer (25 mM Tris-base, 200mM Glycin, 0.1% SDS). Following migration, proteins were electro-transferred to 0.45 µm NC Protran nitrocellulose blotting membranes (Amersham) in transfer buffer (20 mM NaH₂PO₄, 20 mM Na₂HPO₄, pH 6.7).

Pgm was immuno-detected using α -Pgm 2659-GP guinea pig polyclonal antibodies (1:5000 or 1:1000) (26), GFP using the rabbit polyclonal antibody padg1 (1:2000, ChromoTek), tubulin using the monoclonal α -alpha Tubulin TEU435 (1:10000 or 1:5000) (39). Antibodies were diluted in TBST (50 mM Tris-HCl pH 8, 150 mM NaCl, 0.5% Tween) containing 4% low fat milk powder. Primary antibodies were revealed using species-appropriate HRP-conjugated secondary antibodies (anti-rabbit or anti-mouse IgG (1:5000, Promega), or anti-guinea pig IgG (1:5000, Thermo Scientific)). The signal was acquired using the ChemiDoc Touch Imaging System (Bio-Rad) and quantified with the Image Lab software (Bio-Rad).

RNA sequencing

All RNA preparations used for sequencing were extracted previously from autogamy time-course experiments, in which the following genes were knocked down using RNAi: *EZL1* and *ICL7* control (21), *PGM*, *KU80c* & *ND7* control (= *ND7_K*) (30), *CtIP* (38), *XRCC4* and two independent *ND7* controls (= *ND7_X* and *ND7_L*) (29).

RNAseq data from *EZL1* and *ICL7* KDs were extracted from previously published work (Supplementary Table S2) (19, 40). Poly A RNAs extracted from the *KU80c*, *PGM* and *ND7_K* time-courses were sequenced from oriented Illumina TruSeq libraries using a HiSeq sequencer (2 x 100 bp paired end), as described (14). For the *XRCC4*, *CtIP*, *ND7_X* and *ND7_L* time-courses, RNA sequencing was performed on a NextSeq 500 sequencer (2 x 75 bp paired end), as described (19) (Supplementary Table S2).

Reference genomes

The sequencing data were mapped on *P. tetraurelia* strain 51 MAC v1 (ptetraurelia_mac_51.fa) and MAC+IES v1 (ptetraurelia_mac_51_with_ies.fa) reference genomes using Hisat2 (v2.1.0, --rna-strandness FR --min-intronlen 20) (8). Gene annotation v2.0 (ptetraurelia_mac_51_annotation_v2.0.gff3) and IES annotation v1 (internal_eliminated_sequence_PGM_ParTIES.pt_51.gff3) were used in this study (8).

Estimation of IES retention from DNaseq reads

For DNA sequencing, developing new MACs were collected at T30 (30 hours after T0) from 500 mL of large-scale cultures of *P. tetraurelia* 51 new grown in *XRCC4* or *ND7* RNAi-inducing media and enriched by centrifugation through a 2.1 M sucrose layer (9, 14). Total DNA from flow cytometry-sorted new MACs (Supplementary Figure S2) was sequenced on a NextSeq 500 sequencer as described (9) (Supplementary Table S2). IES retention was evaluated by counting the reads mapping on each IES left boundary normalized by total read coverage per Mbp of the MAC+IES genome.

Differential expression (DE) analysis

For each RNAseq dataset, raw fragment counts were determined for each gene using htseq-count (v0.11.2, --stranded=yes --mode=intersection-nonempty) on filtered BAM files (samtools v1.3.1) as described (8). For the correction of batch effects due to differences in RNA extraction conditions and sequencing protocols (HiSeq or NextSeq), we first validated the ComBat-seq procedure (with no group condition ((41), sva v3.38.0, R package) on

technical replicates from the EZL1 and ICL7 time courses that were sequenced using both HiSeq (40) and NextSeq (19) (Supplementary Figure S3, Supplementary Table S3). We then corrected all RNAseq datasets (except for the EZL1 time course) as the first step of our analysis workflow (Supplementary Figure S4), before completing a first round of DESeq2 analysis (42), DESeq2 v1.30.1, R package). Since biological replicates were available only for control KDs (ND7 and ICL7 time-courses), we used hierarchical clustering (Pearson correlation matrix, stats v4.0.3, R package) and principal component analysis (PCA, FactoMiner v2.4, R package) to group the samples in 4 autogamy stages for each RNAi condition (40) (Part 1 of Supplementary Figure S4). The resulting groups were manually curated based on cytological observations of autogamous cells (Supplementary Figure S5). For the ICL7 time course, the time points sequenced both by HiSeq and NextSeq were considered as technical replicates and collapsed. The samples of all control time-courses grouping in the same autogamy stage were treated together as biological replicates. For each KD, the time-points that were attributed to the same autogamy stage were considered as replicates for all DE analyses. Following a second DESeq2 round, we validated the final set of 19 replicate groups by hierarchical clustering and PCA analysis (Supplementary Figure S6). The results of DE analysis were extracted using both RNAi conditions and autogamy stages as contrast (Part 2 of Supplementary Figure S4). Genes were considered deregulated if their fold-change reached 2 (or more) compared to controls, with a p-value < 0.05 (gene lists in Supplementary Tables S4 & S5).

Identification of a conserved motif

We used STREME 5.4.1 (43) to search for 5 to 20 bp-long sequence motifs (with a p-value <0.05) that are enriched in the promoters (i.e. the 150-bp window upstream of the transcription start site (TSS, see (8)) of the 180 genes that are co-regulated with *PGM* (Co-*PGM* genes: intermediate peak genes, up-deregulated in *XRCC4*, *PGM* and *KU80c* KDs and down-regulated in a *CtIP* KD). The search was repeated 5 times against different control sets of the same number of randomly-picked gene promoters. The .meme files corresponding to the first hits of each STREME round were aligned and trimmed to the shortest motif size (18 bp) (`trim_motifs`, `universalmotif v1.12.3`, R package). A consensus motif was defined by merging the five motifs (`merge_motif`, `universalmotif v1.12.3`, R package). For each round, STREME defined an enrichment factor as the fraction of genes with at least one match to the consensus in their promoter. The significance threshold for further screening was based on the mean of the 5 enrichment factors.

Analysis of the motif in gene promoters

To identify all genes carrying the motif in their promoter, we used the FIMO 5.4.1 software (44) with default settings. We first extracted the data obtained for the set of 180 Co-*PGM* genes and defined a cut-off p-value that matches the significance threshold calculated from STREME enrichment factors (defined above). We then applied the resulting cut-off to the full list of motifs identified by FIMO.

To compare the position of the motifs in their respective promoter, we considered the position of the first 5' nucleotide of the motif relative to the TSS. We counted the number of motifs found in successive 5-bp windows within gene promoters. Each window count was

compared to all others and an experimental p-value was calculated using the hypergeometric law (phyper function, stat v4.1.3, R package). To limit the false discovery rate due to multiple tests, we compared the resulting data to simulated p-values obtained using 1000 sets of an equivalent number of randomly selected genes carrying a motif in their promoter (45). For each window, the adjusted p-value corresponds to the fraction of simulated p-values that are inferior to the experimental p-value. The motif was considered significantly enriched in a 5-bp window if the adjusted p-value is < 0.05 .

RESULTS

Inhibition of Programmed DNA Elimination (PDE) triggers overexpression of known core machinery genes

Following up on previous experiments (30), we first confirmed on western blots that the observed up-deregulation of *KU70* and *PGM* mRNAs in a *KU80c* KD results in accumulation of Pgm at late developmental stages (Figure 1A, Supplementary Figure S1A for survival test). To investigate whether the deregulation takes place at the transcriptional level, a *GFP*-reporter transgene expressed under the control of *PGM* transcription start and termination signals was micro-injected into the MAC of vegetative *P. tetraurelia* cells. GFP levels were monitored on western blots after the onset of autogamy (Figure 1A, Supplementary Figure S1A for survival test). We observed that accumulation of GFP parallels that of endogenous Pgm in a *KU80c* KD, indicating that gene deregulation mostly takes place in the old MAC. We can, however, not ascertain whether the accumulation of GFP results from transcriptional or post-transcriptional deregulation, since the mRNA produced from the transgene carries both the 5' and 3' UTRs of *PGM*.

Pgm and Ku70/Ku80c are essential for DNA cleavage upon initiation of IES excision (25, 30). As a consequence, RNAi-mediated silencing of *PGM* or *KU80c* each cause genome-wide IES retention (14, 31). To examine whether interfering with IES excision downstream of DSB introduction would also up-deregulate *PGM*, we knocked down *XRCC4*, which encodes an essential partner of the NHEJ-specific Ligase 4. Previous molecular analyses established that DSBs introduced at IES ends in an *XRCC4* KD are not repaired, resulting in the persistence of broken ends and a failure to assemble IES- junctions (29). As shown on western blots, we observed a strong increase in Pgm levels in an *XRCC4* KD relative to control conditions (Figure 1B, Supplementary Figure S1B for survival test). To monitor the excision status of each IES, we could not use the standard IES retention score (46) because calculating this score relies on the counting of de novo IES- junctions, the formation of which is impaired after DSB introduction in *Xrcc4*-depleted cells. We therefore used high-throughput DNA sequencing data from purified new MACs to evaluate IES boundary coverage, by counting the IES+ reads corresponding to uncut molecules and normalizing the resulting value by total sequencing coverage (Figure 1C, Supplementary Figure S1C for survival test). We considered that an IES is at least partially retained when its boundary coverage is >0.1 . Surprisingly, we found that 65% of IESs were above this threshold at T30 in an *XRCC4* KD, in contrast to a control KD. Very early and early excised IESs were overrepresented among the 8854 IESs that exhibited little or no retention (Figure 1D). This suggests that inhibiting DSB repair at early IES excision sites impinges on the normal progression of programmed DNA elimination. Taken together, our data indicate that knocking down *XRCC4* also deregulates *PGM* expression during MAC development, even if IES excision is not affected through the same mechanism as in a *KU80c* KD.

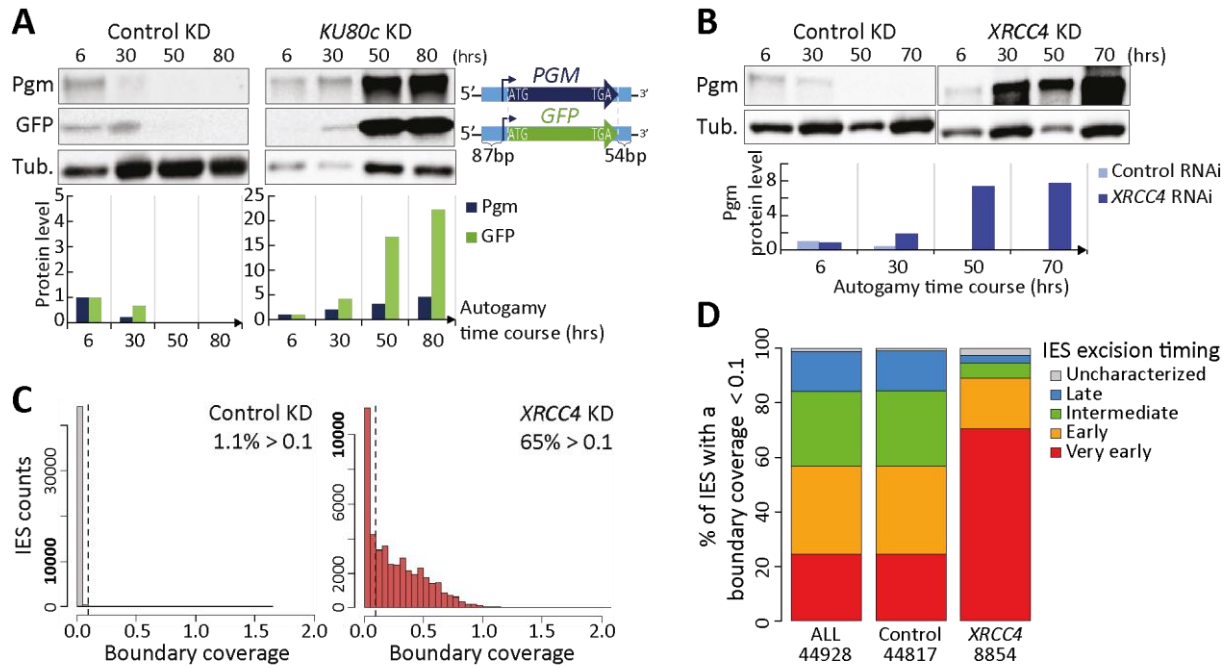


Figure 1. Pgm overexpression in *KU80c* and *XRCC4* KDs

In panels A and B, time-points are in hours (hrs) following T0 (defined in Materials and Methods). Normalized values relative to the α -tubulin signal (Tub.) are plotted below each western blot. (A) Western blot quantification of the levels of endogenous Pgm and GFP during an autogamy time-course of cells subjected to control or *KU80c* RNAi. The GFP is expressed from a *GFP*-reporter gene under the control of *PGM* transcription signals (1976 cphg). For each primary antibody used the T6 values is set at 1 as reference level. (B) Western blot quantification of endogenous Pgm levels during an autogamy time-course of cells subjected to control or *XRCC4* RNAi. The level of Pgm at T6 in control RNAi is set to 1 as reference level. (C) Evidence for IES retention in an *XRCC4* KD. The histograms represent the number of IESs in each 0.05-interval of normalized IES boundary coverage, following control (gray) of *XRCC4* (red) RNAi. The % in each panel indicates the fraction of IESs with a normalized IES boundary coverage > 0.1. Note that y-axes are not to the same scale in both panels. (D) Repartition among the excision timing groups (9) of the IESs with a boundary coverage < 0.1 in control and *XRCC4* KD compared to all IESs (ALL). The number of IESs in each category is given under the plot. (all IES groups are significantly different from ALL and control KD in *XRCC4* KD, p-value < 1.10⁻²⁰).

Genome-wide identification of up-deregulated autogamy genes in *PGM*, *KU80c* and *XRCC4*

KDs

PGM belongs to a cluster of 2037 differentially expressed genes that are transiently induced in an intermediate expression peak during autogamy, at the stage when PDE takes place in the developing MAC (8). To gain broader insight into how many genes are up-deregulated in the three KDs that we used in the present study to alter the normal course of PDE, we collected RNA sequencing data from independent autogamy time-course experiments performed on cells subjected to *KU80c*, *PGM* (30) or *XRCC4* KDs (29) and compared them with four independent control time-courses. As described in Materials and Methods, all sequencing data were first corrected using the ComBat-seq procedure (41) to compensate for batch effects due to the variable conditions (cell culture and sequencing method) in which the different datasets were obtained (Supplementary Figure S3 & Supplementary Figure S4). We then applied principal component analysis and hierarchical clustering to the normalized datasets to group the samples in four developmental stages (VEGETATIVE, EARLY, INTERMEDIATE and LATE) (Supplementary Figure S6). To define the sets of deregulated genes in *PGM*, *KU80c* and *XRCC4* KDs, we focused on the LATE stage, which yielded the largest number of deregulated genes in each KD relative to control conditions (Supplementary Table S4). We found that many genes are significantly up-deregulated at the LATE stage in a *PGM*, *KU80c* or *XRCC4* KD, with an intersection of 628 genes, referred to as UP PKX genes, that are up-deregulated in all three KDs (Figure 2A). Almost half of these 628 UP PKX genes belong to the early (15%) or intermediate (37%) expression peaks ((8); Figure 2B), an enrichment that can be detected in each individual KD (Supplementary Figure S7A-D). We also noted a bias in the germline structure of the LATE up-deregulated genes, 86% being devoid of any IES compared with 53% for all genes (Figure 2C, Supplementary Figure

S7E-H). The 235 LATE up-deregulated genes from the intermediate peak, which represent only ~12% of the expression cluster, include almost all of the genes encoding known components of the core IES excision machinery (Supplementary Table S4): *PGML1*, *PGML2*, *PGML3a*, *PGML4a&b*, *PGML5a&b* (27) and *KU70a* (30). Only *PGM* (25) and *KU80c* (30) are missing since they are not overexpressed in their own respective KD. Our results suggest that mRNA from genes involved in PDE accumulate (or persist) at LATE stages in *PGM*, *KU80c* or *XRCC4* KDs, as a response to defective IES excision in the new MAC.

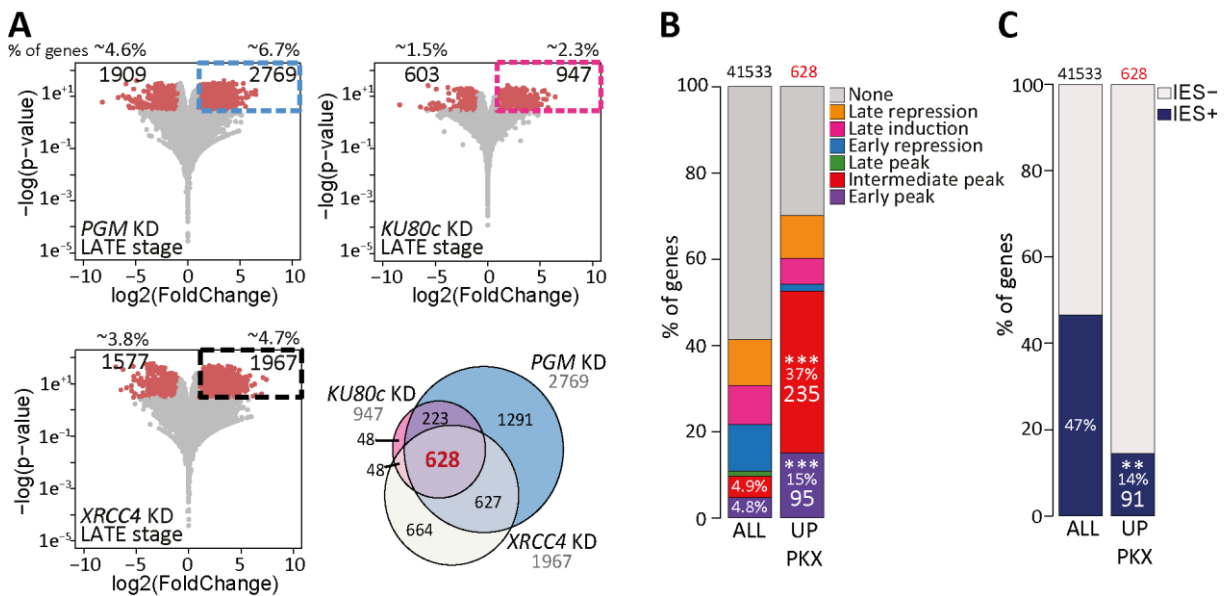


Figure 2. Identification of a set of 330 genes co-regulated with known core IES excision machinery genes

(A) Volcano plots showing differentially expressed genes at the LATE stage in *PGM* (top left), *KU80c* (top right) or *XRCC4* (bottom left) KDs relative to the controls. Significantly deregulated genes are indicated (in red), with their number and percentage relative to all 41533 genes. Up-deregulated genes are boxed on each plot. The Venn diagram shows the intersection of the sets of genes that are up-deregulated at LATE stages in all three KDs. (B) Proportions of gene from each previously identified expression cluster (8) among the different sets of deregulated genes. ALL: all annotated genes; UP PKX: up-deregulated genes at LATE stage in *PGM*, *KU80c* and *XRCC4* KDs, ***: p-value < 1.10⁻³. (C) Percentage of genes without IES (IES-) or carrying at least one IES (IES+) among the two categories described in B; **: p-value < 1.10⁻²⁰.

The developing new MAC controls the transcriptional program during MAC development

As shown above, the shut-off of a subset of 330 genes from the early and intermediate expression peaks at the LATE stage of autogamy seems to critically depend upon the correct completion of PDE in the new developing MAC (Figure 2B), a behavior similar to that of endogenous *PGM* and its derived *GFP* transgene (Figure 1A and B). This suggests that gene transcription in the old MAC is controlled by some signal(s) coming from the new developing MAC. To investigate this possibility, we examined the production of Pgm during autogamy in cells depleted in *CtIP*, an essential protein involved in MIC meiosis (38). In a *CtIP* KD, the formation of the zygotic nucleus is inhibited and no new MAC is formed (Figure 3A, Supplementary Figure S5). In contrast to control conditions, we detected no expression of endogenous Pgm on western blots in a *CtIP* KD, whichever autogamy stage was examined (Figure 3B; Supplementary Figure S1D for the survival test). Moreover, following micro-injection into the old MAC of a reporter *GFP* transgene expressed under the control of *PGM* transcription signals, we observed only a background GFP signal in a *CtIP* KD (Figure 3C and Supplementary Figure S8; Supplementary Figure S1E for the survival test). Our data suggest that *PGM* expression from the old MAC is abolished when no new MAC develops after the onset of autogamy. We then extended our study to the whole transcriptome by sequencing poly A-transcripts that were extracted during an autogamy time-course of *CtIP*-silenced cells (Supplementary Table S2) and performing DE analyses. To identify the genes that, similar to *PGM*, are down-deregulated in a *CtIP* KD, we concentrated on the EARLY and INTERMEDIATE autogamy stages (Figure 3D). We found 2365 significantly down-deregulated genes at either of these stages, with an enrichment in genes from the early, intermediate and late induction peaks (Figure 3E and Supplementary Figure S7I). Of note, 263 of the EARLY or INTERMEDIATE down-deregulated genes in a *CtIP* KD are also up-deregulated at LATE autogamy stages in

PGM, *KU80c* and *XRCC4* KDs (Figure 3D). Among these, 180 belong to the intermediate induction peak and, therefore, share the same expression pattern as *PGM* under all induction peak and, therefore, share the same expression pattern as *PGM* under all conditions that were tested in this study. We conclude that, at least for this subset of 180 intermediate peak genes, hereafter designated as ‘Co-*PGM*’ genes, expression from the old MAC is switched on by an activating signal from the new developing MAC and remains on until PDE is completed.

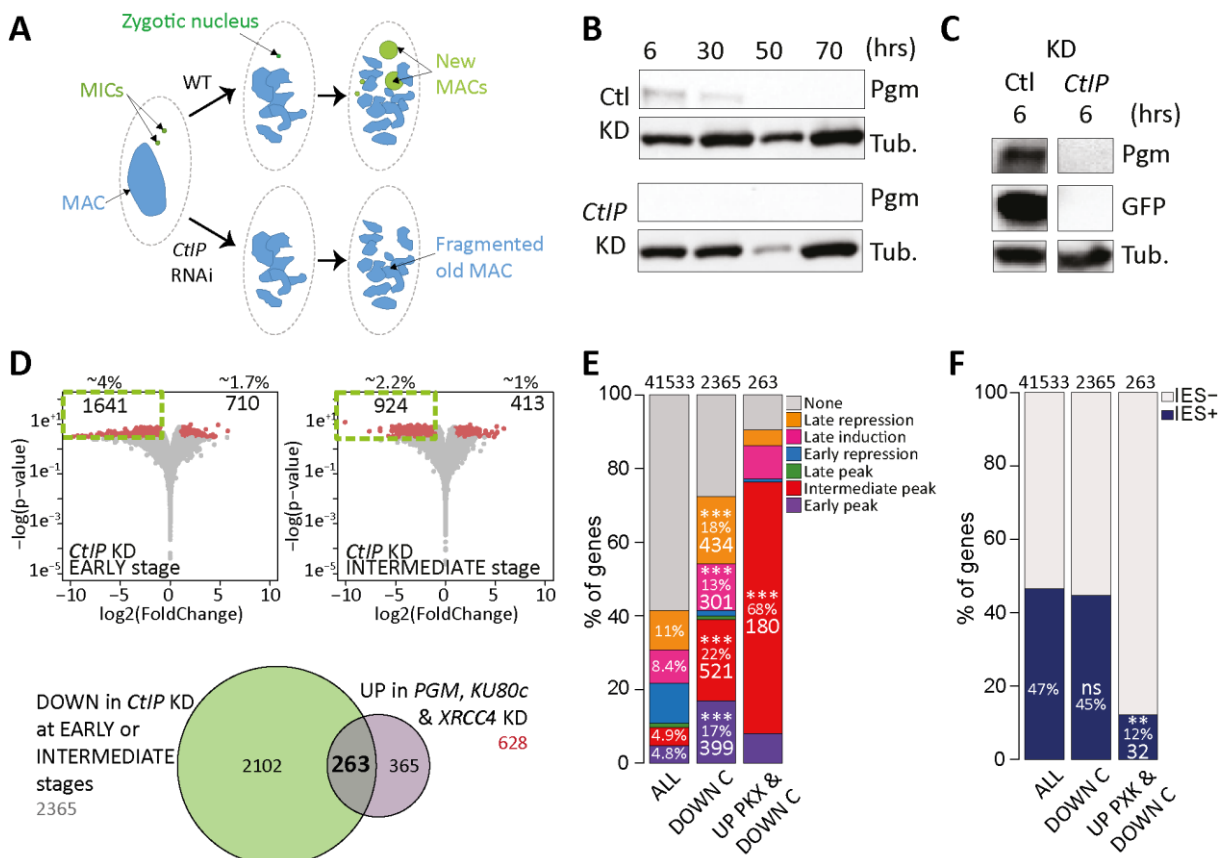


Figure 3. *Pgm* is underexpressed in a *CtIP* KD

(A) *CtIP* RNAi inhibits the formation of the zygotic nucleus and blocks the development of new MACs, while old MAC fragmentation still takes place (38). (B) Western blots showing *Pgm* levels during an autogamy time-course of cells subjected to control (Ctl) or *CtIP* RNAi. Time-points are in hours (hrs) following T0. The α -tubulin (Tub.) was used as a loading control (C) Comparison of the endogenous *Pgm* and GFP signals obtained on western blots at T6 in cells harboring the *GFP*-reporter gene (416 cphg), following control or *CtIP* RNAi. The whole membrane is shown in Supplementary Figure S8A. (D) Volcano plots showing

differentially expressed genes at the EARLY (left) or INTERMEDIATE (right) stages in a *CtIP* KD relative to the controls. Significantly deregulated genes are indicated (in red), with their number and percentage relative to all 41533 genes. Down-deregulated genes are boxed. The Venn diagram shows the intersection of the sets of genes that are down-deregulated at EARLY or INTERMEDIATE stages in *CtIP* KD and up-deregulated at LATE stages in all three conditions displayed in Figure 2. (E) Proportions of genes from each previously identified expression cluster (8) among the different sets of deregulated genes. ALL: all annotated genes; DOWN C: down-deregulated genes at EARLY or INTERMEDIATE stages in *CtIP* KD; UP PKX & DOWN C: intersection the previously described DOWN C genes and the up-deregulated genes at LATE stage in *PGM*, *KU80c* and *XRCC4* KDs. ***: p-value < 1.10⁻³ (F) Percentage of genes without IES (IES-) or carrying at least one IES (IES+) among the three categories described in E; ns: not significant, **: p-value < 1.10⁻²⁰.

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

To get further insight into the possible mechanisms underlying the transcriptional control of Co-*PGM* genes, we used STREME to search for conserved nucleotide sequence motifs in their promoter (Materials and Methods), which we defined as the 150-bp window upstream of their transcription start site (TSS). We found a significantly enriched 18-bp non-palindromic consensus motif (5'-AAAATCMTTdWAATAWTT-3', where M: A/C; d: A/G/T; W: A/T), present in either orientation in 98 out of 180 promoters (54%; Figure 4A). Of note, a total of 3704 genes carried one (94% genes) or more (6% genes) copies of the motif (Supplementary Table S6). Apart from these 98 motif-carrying Co-*PGM* genes, 3606 other genes were found to also harbor the motif in their promoter (Figure 4B), among which 261 intermediate peak (InterP*) genes that were not deregulated in all KDs tested in this study (Figure 4C). No particular orientation bias was noted for the motif (Figure 4D). However, the position of the

motif relative to the TSS was markedly different for Co-*PGM* genes, with a significantly higher density in the -60/-50-bp region (Figure 4E).

To address the biological significance of the conserved motif, we focused on the *PGM* gene itself. Indeed, a copy of the motif (5'-AAAATCATTAAAATATTT-3') is present 55 bp upstream of the *PGM* TSS (Figure 5A; Supplementary Figure S9). We mutagenized this sequence upstream of our *GFP*-reporter transgene and, following micro-injection into the vegetative MAC, we followed on western blots the production of GFP during autogamy of transformed cells harboring either the wild-type or the mutant construct (Figure 5B). No GFP production was detected at any stage from the transgene carrying the mutant motif, in the control as well as the *KU80c* KD, suggesting that the motif is essential for gene transcription. We then tested the effect of changing the orientation of the motif by replacing it with its reverse complement upstream of the *GFP*-reporter gene (Figure 1A). In the control KD, the pattern of GFP production was unchanged during autogamy, with a maximum at T7 and a decrease at late stages, even though protein levels were overall lower than with the wild-type construct (Figure 5C, left panels). In the *KU80c* KD, *GFP* expression from the promoter carrying the inverted motif was up-deregulated at late autogamy stages, similar to what we observed for the wild-type promoter (Figure 5C, right panel). We conclude from these experiments that the conserved motif is required for *PGM* gene expression in the intermediate induction peak, independent of its orientation.

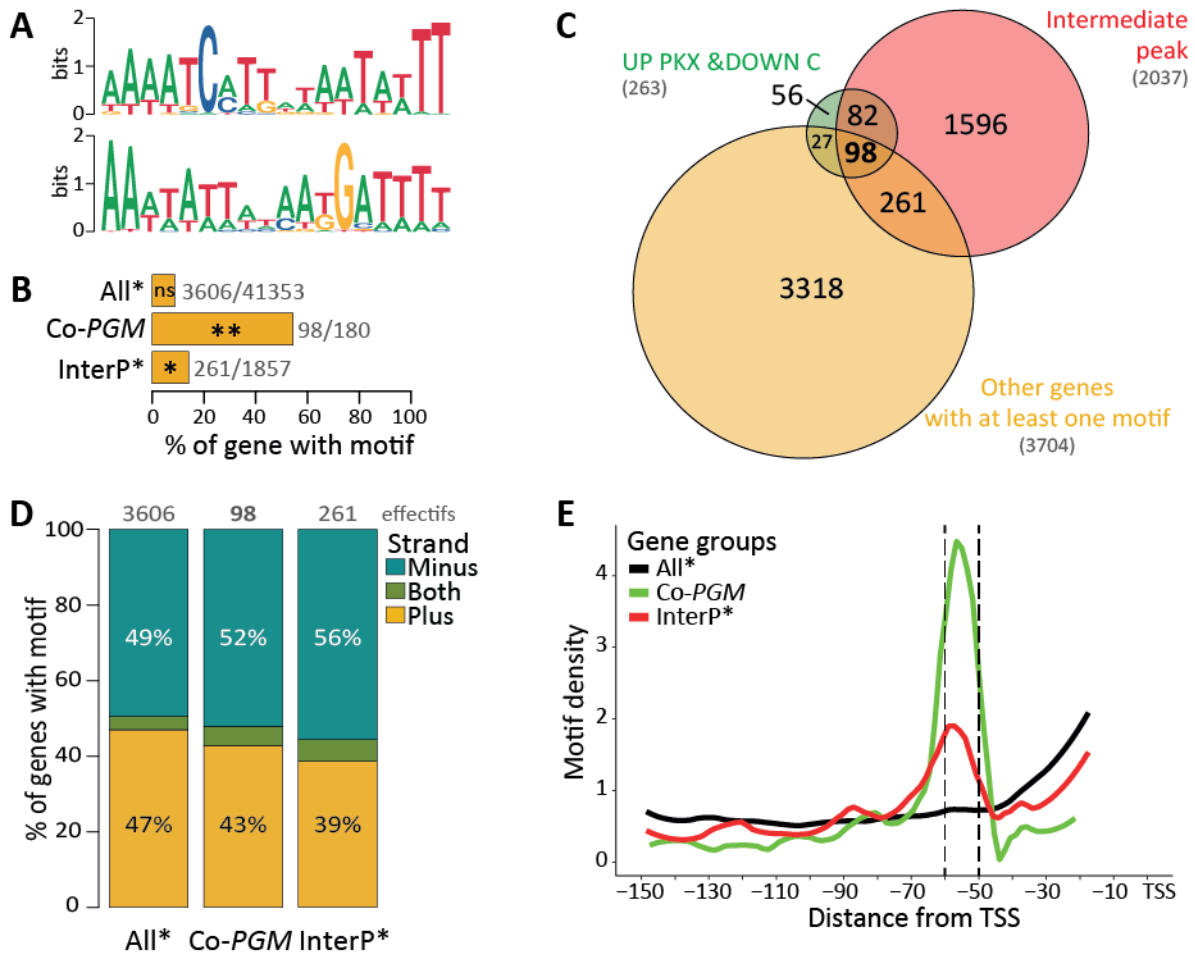


Figure 4. Identification and characterization of an enriched motif in the promoter of genes coregulated with PGM (Co-PGM genes)

(A) Logo of the 18bp non-palindromic merged motif (top) and its reverse complement (bottom) identified by STREME in the 150bp upstream of gene TSSs (Transcription Start Sites). (B) The enrichment percentage of genes with at least one motif in their promoter is shown relative to all 41533 genes (All*: all genes except Co-PGM genes, InterP*: all intermediate peak genes excluding Co-PGM genes). ns: not significant, *: p-value <1.10-10, **: p-value <1.10-20. (C) Venn diagram showing the intersection of the set of genes up-deregulated in PGM, KU80c and XRCC4 KD at LATE stage and down regulated in CtIP KD at EARLY or INTERMEDIATE stage (green) with the intermediate peak expression cluster (red) and all genes carrying at least one motif in their promoter (yellow). (D) Orientation of the motif in gene promoters for the three gene categories described in A. (Minus: all motifs are on the minus strand; Plus: all motifs are on the plus strand; Both: the promoter carries at least two motifs in different orientations). (E) Motif density as a function of the distance of the 5' nucleotide from the TSS, for the three gene categories described in B.

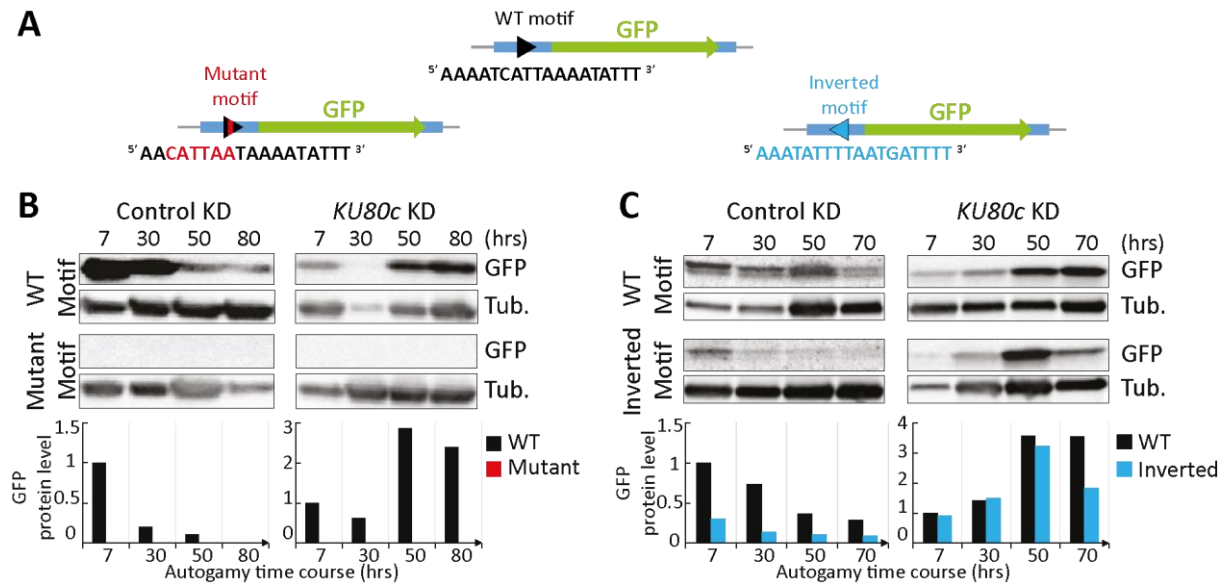


Figure 5. Functional analysis of the motif using the *GFP*-reporter transgene

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

DISCUSSION

Paramecium provides an intriguing example of a multinucleate cell, in which structurally distinct nuclei fulfill different functions. In this study, we have focused on a particular stage of the sexual cycle, during which two consecutive generations of somatic nuclei coexist in the same cell. At this stage, the old (parental) MAC ensures most of gene expression, while the new (zygotic) MAC differentiates during a process that involves massive PDE in its genome, before taking over transcription in the progeny. Sequential waves of gene induction and repression were described during autogamy (8). However, what controls the progression of the transcription program during the sexual cycle is still poorly understood. We show here that abolishing the formation of new MACs (*CtIP* KD) or perturbing the normal course of PDE in the new MAC (*PGM*, *KU80c* or *XRCC4* KDs) induce significant variations of the P. tetraurelia transcriptome during autogamy. Our results provide evidence that the new MAC controls, at two different levels, the progression of the developmental transcription program in the old MAC.

Developing new MACs control gene expression during autogamy

The first level of trans-nuclear control takes place at the EARLY and INTERMEDIATE stages of autogamy. Several published transgene microinjection experiments confirmed that, consistent with a former study performed during conjugation (7), the old MAC contributes most of the mRNAs of known PDE genes during autogamy, which is sufficient to ensure that a functional new MAC develops in the offspring (e.g. (19, 20, 23, 26, 31, 40)). In a *CtIP* KD, the failure to form a zygotic nucleus completely abolishes new MAC development without impinging on the fragmentation of the old MAC (38), from which all mRNAs are produced

during autogamy. Our transcriptome analysis indicates that the absence of new developing MACs reduces the steady-state mRNA levels of 2365 genes at the EARLY and INTERMEDIATE stages (Figure 3), half of which are specifically induced during autogamy under control conditions (8), including a fraction of genes from the early and intermediate peaks (20% and 26%, respectively) and the late induction cluster (8%).

The second level of control exerted by the new MAC depends on the outcome of PDE at LATE autogamy stages. Our work extends previous observations that impeding PDE in the new MAC triggers gene up-deregulation (30). We report that 628 so-called UP PKX genes are significantly up-deregulated at the LATE autogamy stage in *PGM*, *KU80c* and *XRCC4* KDs, instead of being switched off as in the control (Supplementary Table S4). Although PDE is also inhibited in an *EZL1* KD (40, 47), *EZL1* RNA-seq data were not included in our DE analysis because Ezl1-mediated heterochromatin formation is known to repress gene expression at LATE autogamy stages (40). We nevertheless compared the set of UP PKX genes with the list of LATE up-deregulated genes in an *EZL1* KD (Supplementary Table S3) and found a large overlap between the two sets of genes, with UP PKX genes representing ~20% of up-deregulated genes in an *EZL1* KD (Supplementary Figure S11). Thus, two effects of *EZL1* KD on gene expression may be distinguished in our analyses: LATE overexpression of UP PKX genes seems to be a consequence of PDE inhibition, whereas up-deregulation of the other genes is probably due to loss of Ezl1-mediated chromatin modification independently of PDE.

Under control conditions, 15% of UP PKX genes are expressed in the early peak and 37% in the intermediate peak, which corresponds to the expression pattern of most known PDE genes (8). In addition to *PGM* and *KU80c*, 8 UP PKX genes (*PGML1*, *PGML2*, *PGML3a*,

PGML4a&b, *PGML5a&b* and *KU70a*) encode essential components of the core elimination machinery that cleaves DNA at IES ends (27, 30). Ten more UP PKX genes are involved in RNA-mediated targeting of eliminated DNA (*NOWA1* & *NOWA2* (48), *TFIS4* (18), *PDSG2* (49)), chromatin remodeling (*SPT16.1* (23), *ISWI1* (24)) or take part in PDE through still unclear mechanisms (*DIE5a* (50), *mtGa&b* (51), *SPO11* (52)). Based on these observations, we propose that the set of 628 UP PKX genes may include new candidates, whose function in PDE should be explored further.

Since a *GFP*-reporter transgene expressed from the old MAC under the control of *PGM* transcription signals shows a similar expression pattern to endogenous *PGM*, we conclude that LATE up-deregulation applies to old MAC transcripts. However, the observation that 86% of UP PKX genes are IES-free (Figure 2C) and 74% of down-deregulated genes carry at least one IES (Supplementary Figure S7E-H) suggests that mRNA production from the new MAC is also de-regulated at the LATE stage in *PGM*, *KU80c* or *XRCC4* KDs. The biases in IES content may be explained by the selective degradation of IES+ transcripts through Nonsense-Mediated Decay (NMD), because IES retention frequently introduces premature termination codons in mRNAs (14). As a consequence of NMD, IES-containing genes expressed from the new MAC may have been overlooked in our search for UP PKX genes. Consistent with a previous report (53), we indeed found that *PTIWI10*, a gene involved in iesRNA-mediated control of IES excision, is no longer expressed at LATE stages in *PGM*, *KU80c* and *XRCC4* KDs (Supplementary Table S4). In addition to carrying an IES upstream of its start codon, *PTIWI10* is interrupted by an NMD-visible IES in the non-rearranged genome and was shown to be specifically expressed from the new MAC (53).

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

Several lines of evidence already established that the old MAC controls new MAC development through at least two pathways. First, as discussed above, coding transcription in the old MAC ensures the synthesis of proteins that carry out PDE in the new MAC. Second, maternal non-coding transcripts produced by the old MAC serve as pairing templates for the selection of MIC-restricted scnRNAs that are transferred to the new MAC (16), where they drive the recognition of the sequences to be eliminated through Ezi1-mediated formation of heterochromatin (19, 20).

Reciprocally, we provide evidence here that the new developing MAC controls the onset and progression of the gene expression program in the old MAC throughout the different stages of autogamy (Figure 6). In particular, we have identified a group of 180 genes that are co-regulated with *PGM* under all tested conditions (Figure 3E): these “co-*PGM*” genes are specifically expressed in the intermediate induction peak, their induction depends on the presence of new developing MACs in the cell and their mRNA levels drop down after the successful completion of PDE. Future studies will unravel the mechanism(s) that underlie the control of gene expression by the new MAC. An attractive hypothesis would be that the production of activating factors is triggered by the developing new MAC to turn on gene transcription in the old MAC at EARLY and INTERMEDIATE stages. Our finding that, like *PGM* itself, 98 out of 180 co-*PGM* genes carry a conserved motif in their promoter suggests that genes from this subset are activated through the sequence-specific binding of a transcription factor to the motif. The observed LATE up-deregulation of co-*PGM* genes when PDE do not proceed normally may indicate that the activator is expressed from the new MAC before its

genome is rearranged and/or that a repressor produced after PDE is completed is necessary to switch off gene expression.

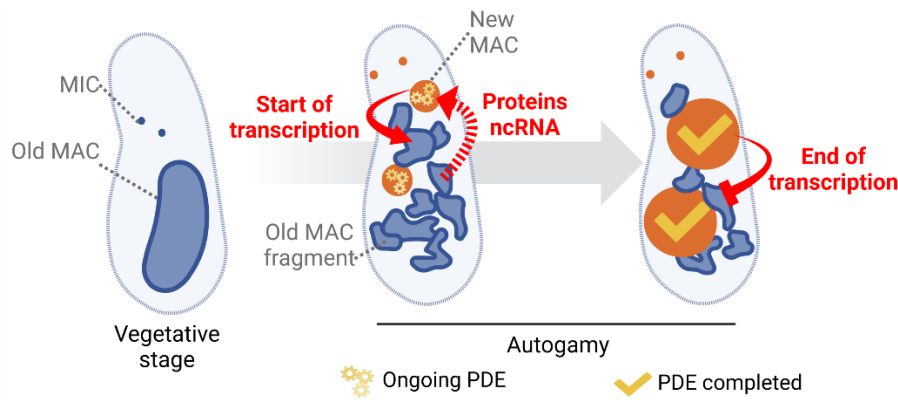


Figure 6. Bi-directional crosstalk between the old and new somatic MACs in *Paramecium*

Upon starvation, vegetative cells (left) become committed to starting their sexual cycle. During autogamy, the self-fertilization process that we have focused on, the MICs undergo meiosis and eventually give rise to the new MICs and MACs (orange), while the parental MAC (blue) is fragmented into ~30 fragments. Most of mRNAs are produced by the old MAC fragments, while programmed DNA elimination (PDE) takes place in the new MAC. Maternal ncRNAs that are constitutively produced from the old MAC participate in the targeting of eliminated sequences in the new MAC (16). We propose that, at EARLY and INTERMEDIATE stages (middle), the new MAC transmits an inducing signal to old MAC fragments (full arrow) to start the transcription of development-specific genes. This drives the production of proteins that carry out PDE in the new MAC (dotted arrow). We also propose that, at the LATE stage (right), a signal from the new MAC arrests the expression of a subset of genes when PDE is completed. (Created with BioRender.com).

The existence of a bi-directional cross-communication between two successive generations of somatic nuclei throughout the *Paramecium* sexual cycle is reminiscent of the mother-embryo crosstalk that has been reported during mammalian development (for a review, see for example (54)). In both systems, the parental and zygotic somatic lines appear to mutually control each other through the production of coding and non-coding RNAs, proteins and

other regulatory factors. A notable difference is that inter-generational exchanges in *Paramecium* are intra-cellular: how a continuous dialogue is established at the molecular level between the old and the new developing MACs will require further investigation. We propose that inter-nuclear crosstalk allows gene transcription in the old MAC and genome rearrangements in the new MAC to proceed in a concerted manner, in order to give rise to viable sexual progeny.

ACKNOWLEDGEMENTS

We would like to thank Sandra Duharcourt, Eric Meyer and Gaëlle Lelandais for fruitful discussions, sharing data and parts of code ahead of publication and providing helpful advice and support as members of MBG.'s thesis advisory committee. We thank all members of the Bétermier lab for stimulating discussions, Pascaline Tirand and Cindy Mathon for technical assistance and Tim Kamara and Camille Poitrenaud for help in plasmid construction during their short-term internships.

This work was supported by intramural funding from the Centre National de la Recherche Scientifique (CNRS) and grants from the Agence Nationale de la Recherche (“PIGGYPACK” ANR-14-CE10-0005 to MB, LS & OA, “LaMarque” ANR-18-CE12-0005-02 and “CURE” ANR-21-CE12-0019-01 to MB, “POLYCHROME” ANR-19-CE12-0015 to OA) and the Fondation pour la Recherche Médicale (FRM EQU202103012766 to MB). MBG was supported by a PhD fellowship from Paris-Saclay University (Ecole Doctorale n°577 SDSV) and a training grant from the Department of Genome Biology of I2BC to attend the “DU-Omique” course in bioinformatics (Université Paris-Cité). We acknowledge the sequencing and bioinformatics expertise of the I2BC High-throughput sequencing facility, supported by France Génomique

(funded by the French National Program “Investissement d’Avenir” ANR-10-INBS-09).

Experiments involving fluorescence-assisted nuclear sorting benefited from the expertise of the Imagerie - Gif core facility, supported by the Agence Nationale de la Recherche (ANR-11-EQPX-0029/Morphoscope, ANR-10-INBS-04/FranceBioImaging, ANR - 11 - IDEX - 0003 - 02/ Saclay Plant Sciences).

REFERENCES

1. Padilla,J.R., Ferreira,L.M. and Folker,E.S. (2022) Nuclear movement in multinucleated cells. *Development*, 149.
2. Dundon,S.E.R., Chang,S.-S., Kumar,A., Occhipinti,P., Shroff,H., Roper,M. and Gladfelter,A.S. (2016) Clustered nuclei maintain autonomy and nucleocytoplasmic ratio control in a syncytium. *Mol Biol Cell*, 27, 2000–2007.
3. Kim,M., Franke,V., Brandt,B., Lowenstein,E.D., Schöwel,V., Spuler,S., Akalin,A. and Birchmeier,C. (2020) Single-nucleus transcriptomics reveals functional compartmentalization in syncytial skeletal muscle cells. *Nat Commun*, 11, 6375.
4. Petrany,M.J., Swoboda,C.O., Sun,C., Chetal,K., Chen,X., Weirauch,M.T., Salomonis,N. and Millay,D.P. (2020) Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. *Nat Commun*, 11, 6374.
5. Prescott,D.M. (1994) The DNA of ciliated protozoa. *Microbiol Rev*, 58, 233–267.

6. Bétermier, M. and Duhaucourt, S. (2014) Programmed Rearrangement in Ciliates: *Paramecium*. *Microbiol Spectr*, 2, MDNA-0035-2014.
7. Berger, J.D. (1973) Nuclear Differentiation and Nucleic Acid Synthesis in Well-fed Exconjugants of *Paramecium aurelia*.
8. Arnaiz, O., van Dijk, E., Bétermier, M., Lhuillier-Akakpo, M., de Vanssay, A., Duhaucourt, S., Sallet, E., Gouzy, J. and Sperling, L. (2017) Improved methods and resources for *paramecium* genomics: transcription units, gene annotation and gene expression. *BMC Genomics*, 18, 483.
9. Zangarelli, C., Arnaiz, O., Bourge, M., Gorrichon, K., Jaszczyszyn, Y., Mathy, N., Escoriza, L., Bétermier, M. and Régnier, V. (2022) Developmental timing of programmed DNA elimination in *Paramecium tetraurelia* recapitulates germline transposon evolutionary dynamics. *Genome Res*, 32, 2028–2042.
10. Sellis, D., Guérin, F., Arnaiz, O., Pett, W., Lerat, E., Boggetto, N., Krenek, S., Berendonk, T., Couloux, A., Aury, J.-M., et al. (2021) Massive colonization of protein-coding exons by selfish genetic elements in *Paramecium* germline genomes. *PLoS Biol*, 19, e3001309.
11. Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., et al. (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, 444, 171–178.
12. Guérin, F., Arnaiz, O., Boggetto, N., Denby Wilkes, C., Meyer, E., Sperling, L. and Duhaucourt, S. (2017) Flow cytometry sorting of nuclei enables the first global characterization of *Paramecium* germline DNA and transposable elements. *BMC Genomics*, 18, 327.

13. le Mouël,A., Butler,A., Caron,F. and Meyer,E. (2003) Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia. *Eukaryot Cell*, 2, 1076–90.
14. Arnaiz,O., Mathy,N., Baudry,C., Malinsky,S., Aury,J.-M.M., Denby Wilkes,C., Garnier,O., Labadie,K., Lauderdale,B.E., le Mouël,A., et al. (2012) The Paramecium germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet*, 8, e1002984.
15. Klobutcher,L.A. and Herrick,G. (1995) Consensus inverted terminal repeat sequence of Paramecium IESs: resemblance to termini of Tc1-related and Euplotes Tec transposons. *Nucleic Acids Res*, 23, 2006–13.
16. Lepère,G., Bétermier,M., Meyer,E. and Duharcourt,S. (2008) Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in Paramecium tetraurelia. *Genes Dev*, 22, 1501–1512.
17. Lepère,G., Nowacki,M., Serrano,V., Gout,J.-F., Guglielmi,G., Duharcourt,S. and Meyer,E. (2009) Silencing-associated and meiosis-specific small RNA pathways in Paramecium tetraurelia. *Nucleic Acids Res*, 37, 903–915.
18. Maliszewska-Olejniczak,K., Gruchota,J., Gromadka,R., Denby Wilkes,C., Arnaiz,O., Mathy,N., Duharcourt,S., Bétermier,M. and Nowak,J.K. (2015) TFIS-Dependent Non-coding Transcription Regulates Developmental Genome Rearrangements. *PLoS Genet*, 11, e1005383.
19. Miró-Pina,C., Charmant,O., Kawaguchi,T., Holoch,D., Michaud,A., Cohen,I., Humbert,A., Jaszczyszyn,Y., Chevreux,G., del Maestro,L., et al. (2022) Paramecium Polycomb repressive

complex 2 physically interacts with the small RNA-binding PIWI protein to repress transposable elements. *Dev Cell*, 57, 1037-1052.e8.

20. Wang,C., Solberg,T., Maurer-Alcalá,X.X., Swart,E.C., Gao,F. and Nowacki,M. (2022) A small RNA-guided PRC2 complex eliminates DNA as an extreme form of transposon silencing. *Cell Rep*, 40, 111263.

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

22. Sandoval,P.Y., Swart,E.C., Arambasic,M. and Nowacki,M. (2014) Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting. *Dev Cell*, 28, 174–88.

23. de Vanssay,A., Touzeau,A., Arnaiz,O., Frapporti,A., Phipps,J. and Duharcourt,S. (2020) The Paramecium histone chaperone Spt16-1 is required for Pgm endonuclease function in programmed genome rearrangements. *PLoS Genet*, 16, e1008949.

24. Singh,A., Maurer-Alcalá,X.X., Solberg,T., Häußermann,L., Gisler,S., Ignarski,M., Swart,E.C. and Nowacki,M. (2022) Chromatin remodeling is required for sRNA-guided DNA elimination in Paramecium. *EMBO J*, 41, e111839.

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

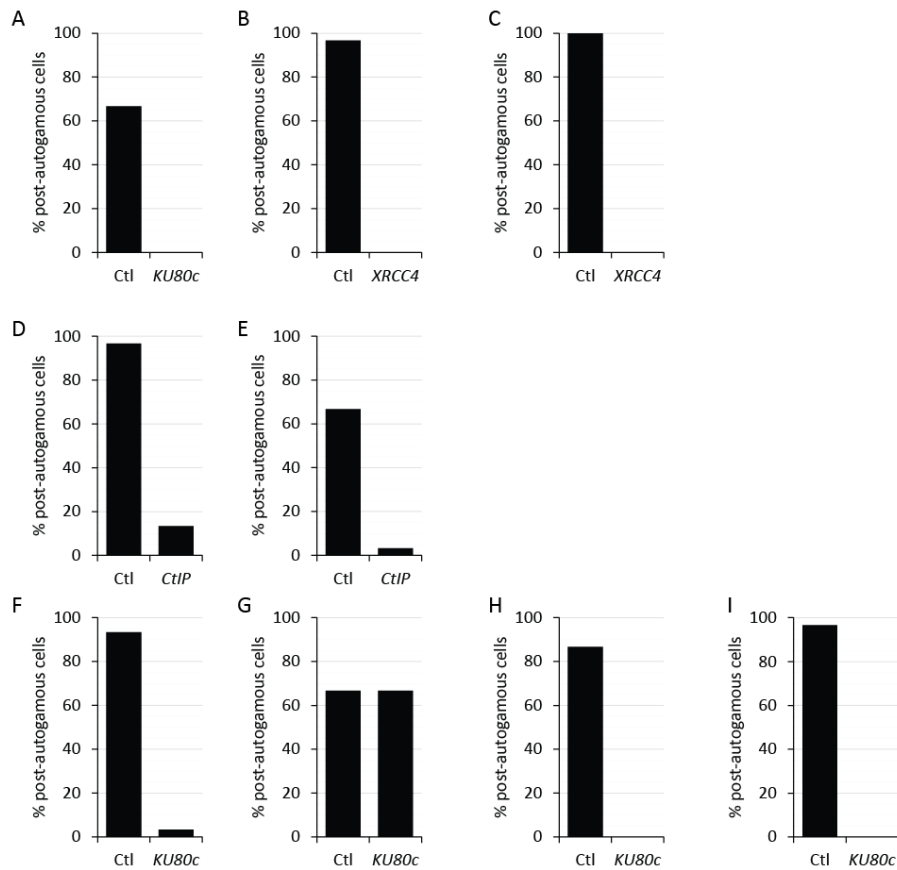
26. Dubois,E., Mathy,N., Régnier,V., Bischerour,J., Baudry,C., Trouslard,R. and Bétermier,M. (2017) Multimerization properties of PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements. *Nucleic Acids Res*, 45, gkw1359.
27. Bischerour,J., Bhullar,S., Denby Wilkes,C., Régnier,V., Mathy,N., Dubois,E., Singh,A., Swart,E., Arnaiz,O., Sperling,L., et al. (2018) Six domesticated PiggyBac transposases together carry out programmed DNA elimination in *Paramecium*. *Elife*, 7, 1–24.
28. Gratiás,A. and Bétermier,M. (2003) Processing of Double-Strand Breaks Is Involved in the Precise Excision of *Paramecium* Internal Eliminated Sequences. *Mol Cell Biol*, 23, 7152–7162.
29. Kapusta,A., Matsuda,A., Marmignon,A., Ku,M., Silve,A., Meyer,E., Forney,J.D., Malinsky,S. and Bétermier,M. (2011) Highly Precise and Developmentally Programmed Genome Assembly in *Paramecium* Requires Ligase IV–Dependent End Joining. *PLoS Genet*, 7, e1002049.
30. Marmignon,A., Bischerour,J., Silve,A., Fojcik,C., Dubois,E., Arnaiz,O., Kapusta,A., Malinsky,S. and Bétermier,M. (2014) Ku-Mediated Coupling of DNA Cleavage and Repair during Programmed Genome Rearrangements in the Ciliate *Paramecium tetraurelia*. *PLoS Genet*, 10, e1004552.
31. Abello,A., Régnier,V., Arnaiz,O., le Bars,R., Bétermier,M. and Bischerour,J. (2020) Functional diversification of *Paramecium* Ku80 paralogs safeguards genome integrity during precise programmed DNA elimination. *PLoS Genet*, 16, e1008723.
32. Arnaiz,O., Meyer,E. and Sperling,L. (2019) *ParameciumDB* 2019: integrating genomic data across the genus for functional and evolutionary biology. *Nucleic Acids Res*, 48, 599–605.

33. Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., et al. (2010) Silencing Specific *Paramecium tetraurelia* Genes by Feeding Double-Stranded RNA. *Cold Spring Harb Protoc*, 2010, pdb.prot5363.
34. Bétermier, M., Duharcourt, S., Seitz, H. and Meyer, E. (2000) Timing of Developmentally Programmed Excision and Circularization of *Paramecium* Internal Eliminated Sequences. *Mol Cell Biol*, 20, 1553–1561.
35. Galvani, A. and Sperling, L. (2002) RNA interference by feeding in *Paramecium*. *Trends in Genetics*, 18, 11–12.
36. Timmons, L., Court, D.L. and Fire, A. (2001) Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene*, 263, 103–112.
37. Garnier, O., Serrano, V., Duharcourt, S. and Meyer, E. (2004) RNA-Mediated Programming of Developmental Genome Rearrangements in *Paramecium tetraurelia*. *Mol Cell Biol*, 24, 7370–7379.
38. Godau, J., Ferretti, L.P., Trenner, A., Dubois, E., von Aesch, C., Marmignon, A., Simon, L., Kapusta, A., Guérois, R., Bétermier, M., et al. (2019) Identification of a miniature *Sae2/Ctp1/CtIP* ortholog from *Paramecium tetraurelia* required for sexual reproduction and DNA double-strand break repair. *DNA Repair (Amst)*, 77, 96–108.
39. Callen, A.M., Adoutte, A., Andrew, J.M., Baroin-Tourancheau, A., Bré, M.H., Ruiz, P.C., Clérot, J.C., Delgado, P., Fleury, A. and Jeanmaire-Wolf, R. (1994) Isolation and characterization of libraries of monoclonal antibodies directed against various forms of tubulin in *Paramecium*. *Biol Cell*, 81, 95–119.

40. Frapporti,A., Miró Pina,C., Arnaiz,O., Holoch,D., Kawaguchi,T., Humbert,A., Eleftheriou,E., Lombard,B., Loew,D., Sperling,L., et al. (2019) The Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in *Paramecium*. *Nat Commun*, 10, 2710.
41. Zhang,Y., Parmigiani,G. and Johnson,W.E. (2020) ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform*, 2, 1–10.
42. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15, 550.
43. Bailey,T.L. (2021) STREME: Accurate and versatile sequence motif discovery. *Bioinformatics*, 37, 2834–2840.
44. Grant,C.E., Bailey,T.L. and Noble,W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, 27, 1017.
45. Boyle,E.I., Weng,S., Gollub,J., Jin,H., Botstein,D., Cherry,J.M. and Sherlock,G. (2004) GO::TermFinder - Open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, 20, 3710–3715.
46. Denby Wilkes,C., Arnaiz,O. and Sperling,L. (2016) ParTIES: a toolbox for *Paramecium* interspersed DNA elimination studies. *Bioinformatics*, 32, 599–601.
47. Lhuillier-Akakpo,M., Guérin,F., Frapporti,A. and Duharcourt,S. (2016) DNA deletion as a mechanism for developmentally programmed centromere loss. *Nucleic Acids Res*, 44, 1553–65.

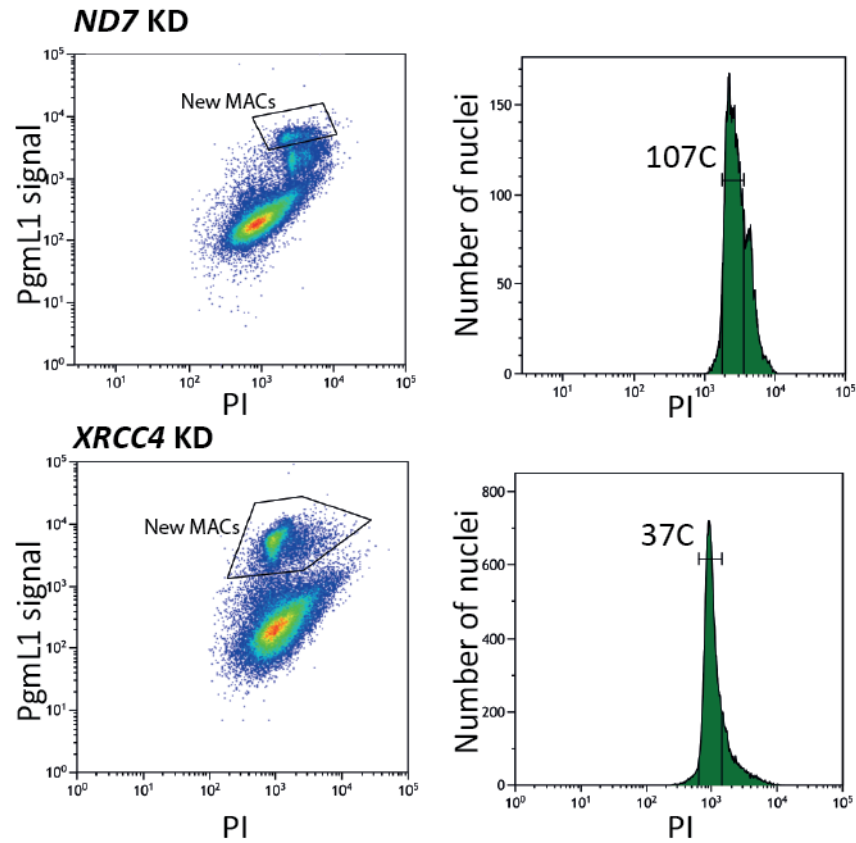
48. Nowacki, M., Zagorski-Ostoja, W. and Meyer, E. (2005) Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in *Paramecium tetraurelia*. *Curr Biol*, 15, 1616–28.
49. Arambasic, M., Sandoval, P.Y., Hoehener, C., Singh, A., Swart, E.C. and Nowacki, M. (2014) Pdsg1 and Pdsg2, novel proteins involved in developmental genome remodelling in *Paramecium*. *PLoS One*, 9, e112899.
50. Matsuda, A., Shieh, A.W.-Y., Chalker, D.L. and Forney, J.D. (2010) The Conjugation-Specific Die5 Protein Is Required for Development of the Somatic Nucleus in both *Paramecium* and *Tetrahymena*. *Eukaryot Cell*, 9, 1087–1099.
51. Bhullar, S., Denby Wilkes, C., Arnaiz, O., Nowacki, M., Sperling, L. and Meyer, E. (2018) A mating-type mutagenesis screen identifies a zinc-finger protein required for specific DNA excision events in *Paramecium*. *Nucleic Acids Res*, 46, 9550–9562.
52. Rzeszutek, I., Swart, E.C., Pabian-Jewuła, S., Russo, A. and Nowacki, M. (2022) Early developmental, meiosis-specific proteins — Spo11, Msh4-1, and Msh5 — Affect subsequent genome reorganization in *Paramecium tetraurelia*. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1869, 119239.
53. Furrer, D.I., Swart, E.C., Kraft, M.F., Sandoval, P.Y. and Nowacki, M. (2017) Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA. *Cell Rep*, 20, 505–520.
54. Idelevich, A. and Vilella, F. (2020) Mother and Embryo Cross-Communication. *Genes (Basel)*, 11, 376.

55. Gout,J.-F., Johri,P., Arnaiz,O., Doak,T.G., Bhullar,S., Couloux,A., Guérin,F., Malinsky,S., Sperling,L., Labadie,K., et al. (2019) Universal trends of post-duplication evolution revealed by the genomes of 13 Paramecium species sharing an ancestral whole-genome duplication. *bioRxiv*, 10.1101/573576.
56. McGrath,C.L., Gout,J.-F., Johri,P., Doak,T.G. and Lynch,M. (2014) Differential retention and divergent resolution of duplicate genes following whole-genome duplication. *Genome Res*, 24, 1665–1675.
57. McGrath,C.L., Gout,J.-F., Doak,T.G., Yanagi,A. and Lynch,M. (2014) Insights into Three Whole-Genome Duplications Gleaned from the Paramecium caudatum Genome Sequence. *Genetics*, 197, 1417–1428.

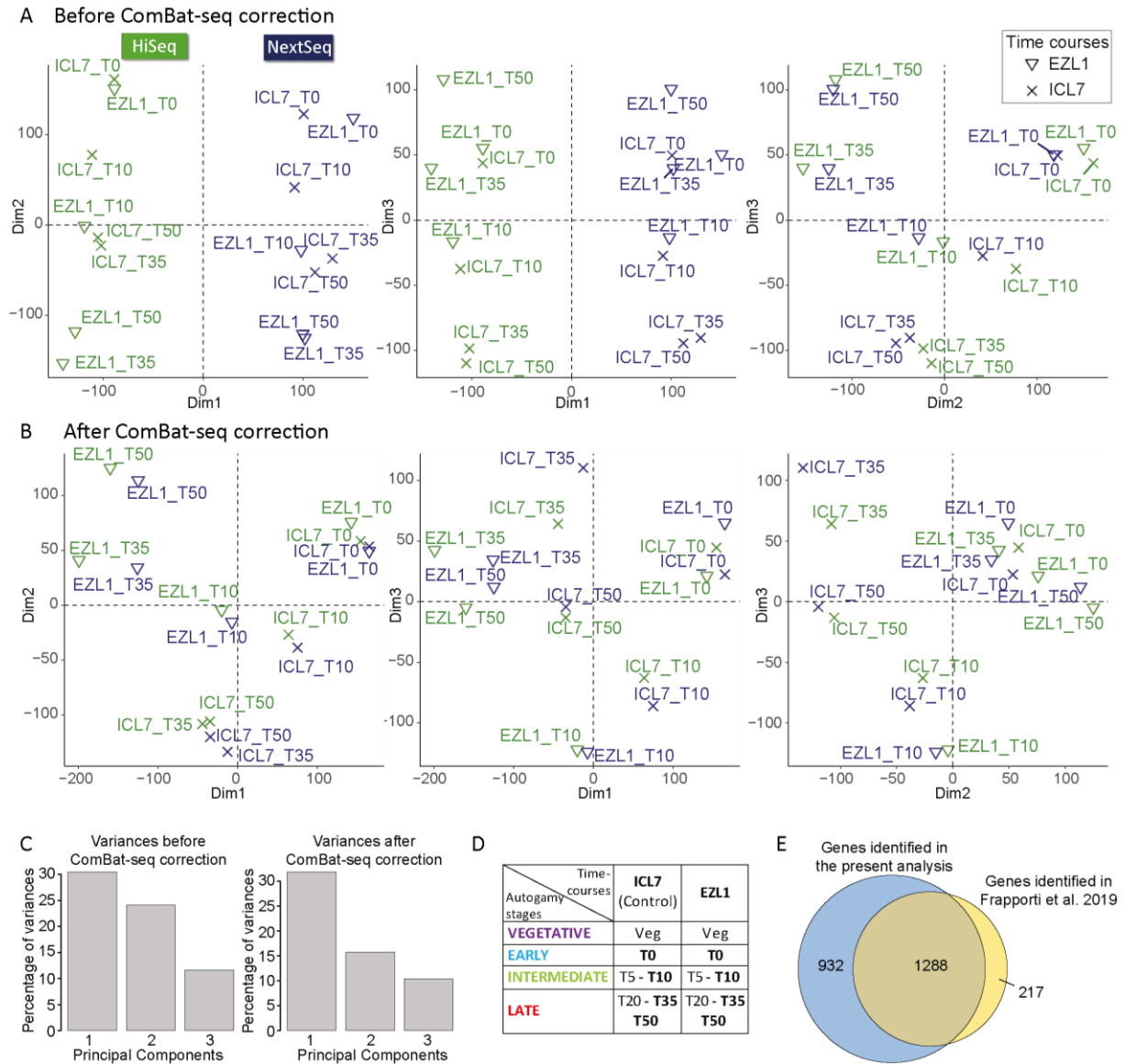


Supplementary Figure S1. Survival of post-autogamous sexual progeny in the different RNAi experiments performed in this study

Each panel shows the fraction of post-autogamous survivors with a functional new MAC, confirming the efficiency of all KDs (Ctl: control KD). (A) *KU80c* RNAi on cells injected with the *GFP*-reporter gene (1976 cphg), used for western blot analysis (Figure 1A). (B) *XRCC4* RNAi experiment used for western blot analysis of endogenous Pgm amounts (Figure 1B). (C) *XRCC4* RNAi experiment used for anlagen purification and DNA sequencing (Figure 1C). (D) *CtIP* RNAi experiment used for western blot analysis of endogenous Pgm amounts (Figure 3B). (E) *CtIP* RNAi experiment on cells injected with the *GFP*-reporter gene (416 cphg), used for western blot analysis (Figure 3C & Supplementary Figure S8). (F) *KU80c* RNAi on cells injected with the *GFP*-reporter gene carrying the WT motif in its promoter (695 cphg), used for western blot analysis (Figure 5B). (G) *KU80c* RNAi on cells injected with the *GFP*-reporter gene carrying the mutant motif in its promoter (2524 cphg), used for western blot analysis (Figure 5B). (H) *KU80c* RNAi on cells injected with the *GFP*-reporter gene carrying the WT motif in its promoter (204 cphg), used for western blot analysis (Figure 5C). (I) *KU80c* RNAi on cells injected with the *GFP*-reporter gene carrying the inverted motif in its promoter (188 cphg), used for western blot analysis (Figure 5C).

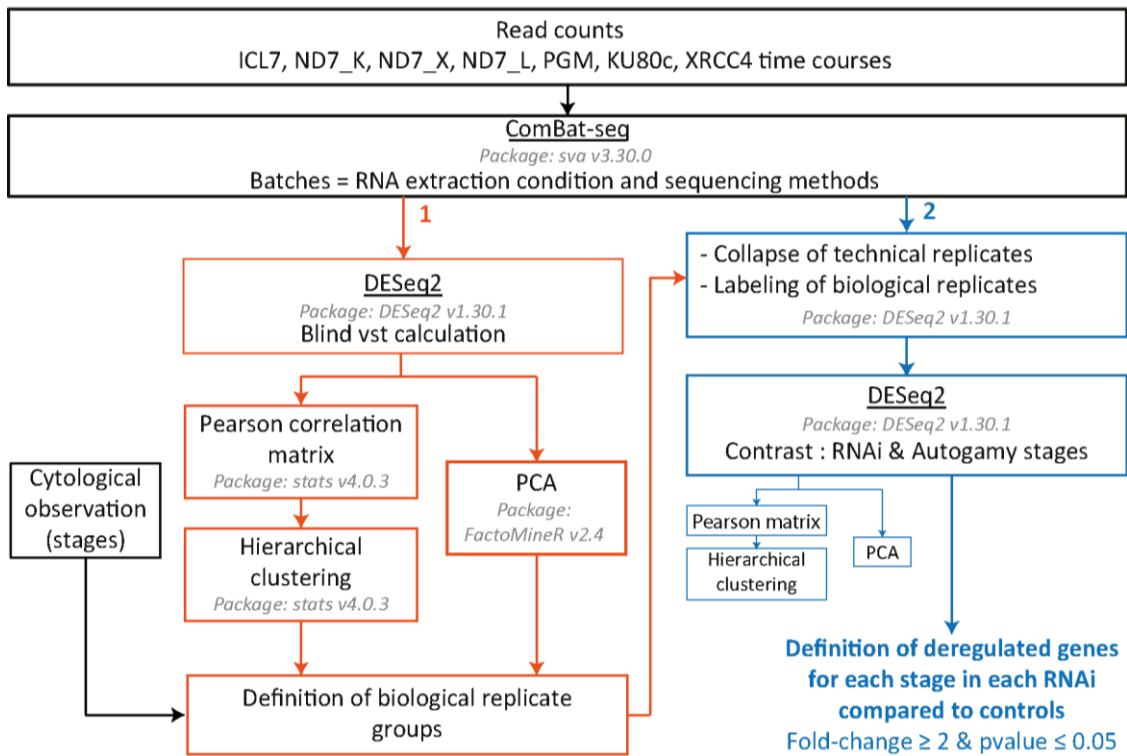


Supplementary Figure S2. Flow cytometry on the nuclei sorted for DNA-sequencing
(Left) Gates used to sort the new MACs from nucleus extraction at T30 of cell culture in control KD (*ND7* KD) or *XRCC4* KD. (Right) Evaluation of the ploidy of the new MACs.



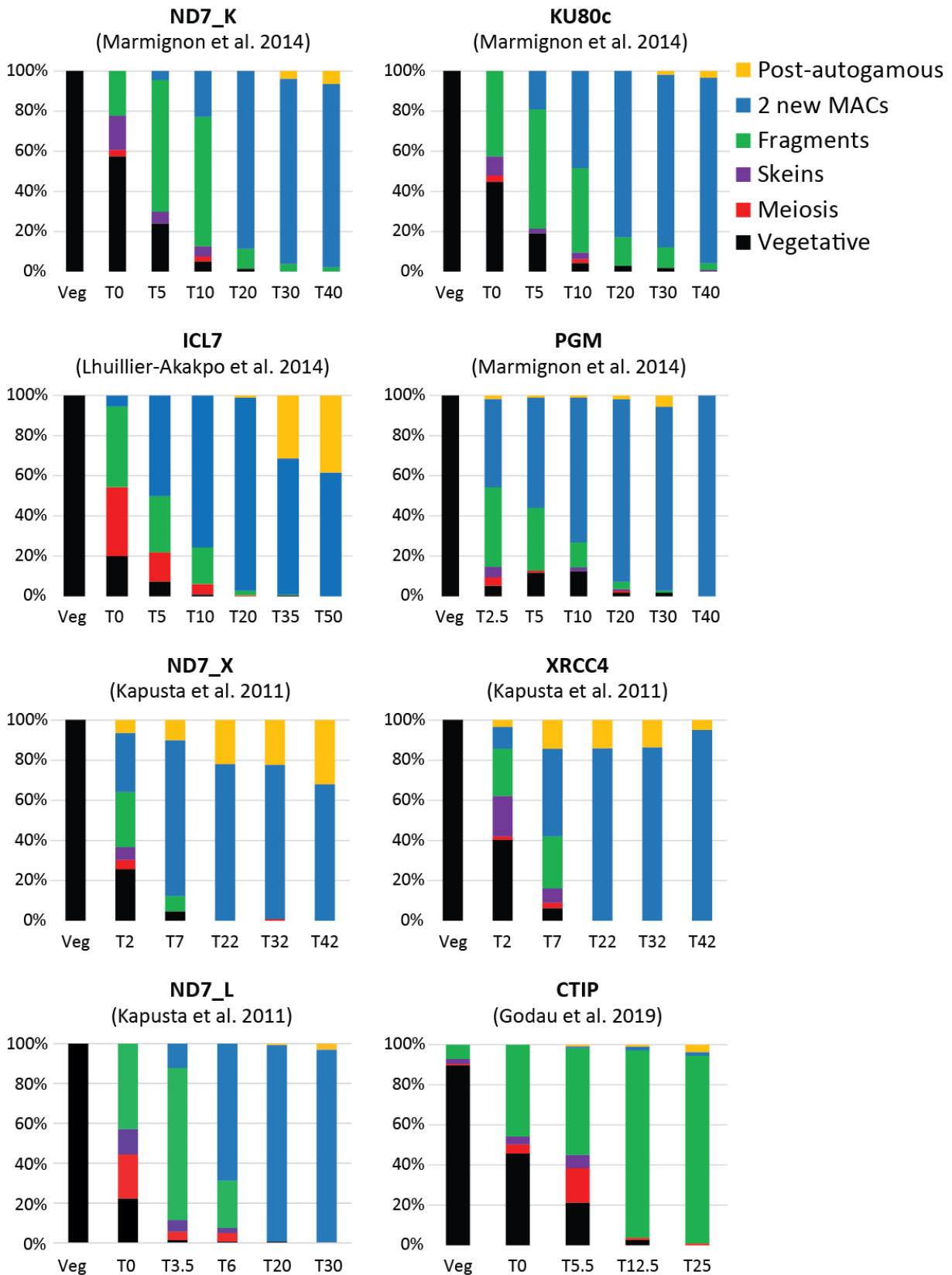
Supplementary Figure S3. Validation of ComBat-seq using *ICL7* (control) and *EZL1* RNAi samples

(A) & (B) HiSeq (green) and NextSeq (dark blue) RNA sequencing data were obtained at T0, T10, T35 and T50 during *ICL7* and *EZL1* autogamy time-courses (19, 40). We extracted the vst (variance stabilizing transformation)-normalized counts from DESeq2 analyses performed before (A) or following (B) ComBat-seq correction ((41), sva 3.38.0 R package). Three dimensions of the PCA are shown. (C) Percentages of variances for the PCAs shown in A & B. (D) Table showing the 8 groups of replicates used for DE analysis, after manual curation and collapse of technical replicates (in bold). (E) Venn diagram showing the comparison of the set of previously identified overexpressed genes at the LATE stage (40) with those found in the present analysis, with a fold-change significance threshold >2.



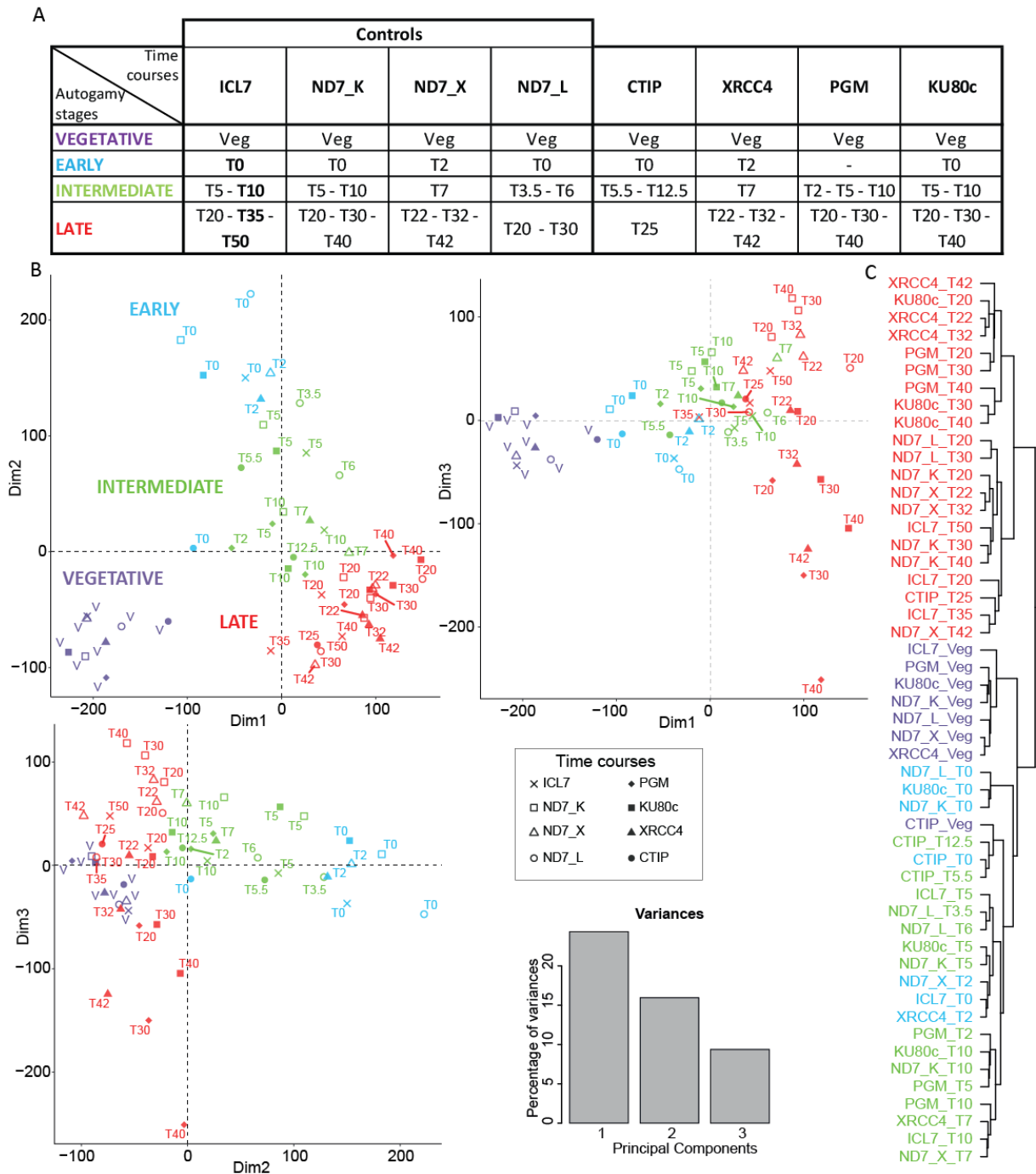
Supplementary Figure S4. Transcriptome analysis workflow for the identification of deregulated genes

All read counts from the indicated autogamy time-courses were first corrected for batch effects using ComBat-seq. (1) We performed a first DESeq2 analysis using the corrected data and extracted the vst-normalized values. Groups of replicates were defined using both Pearson correlation followed by hierarchical clustering, and PCA. We manually curated the groups using cytological observations of autogamy stages (Supplementary Figure S5). (2) After collapsing the technical replicates, we labelled the replicates and ran a second DESeq2 round. The groups of replicates were confirmed using PCA and hierarchical clustering of vst-normalized counts. The results were extracted using RNAi conditions and autogamy stages (VEGETATIVE, EARLY, INTERMEDIATE, LATE) as contrast to define deregulated genes in each RNAi relative to control conditions, at each autogamy stage.



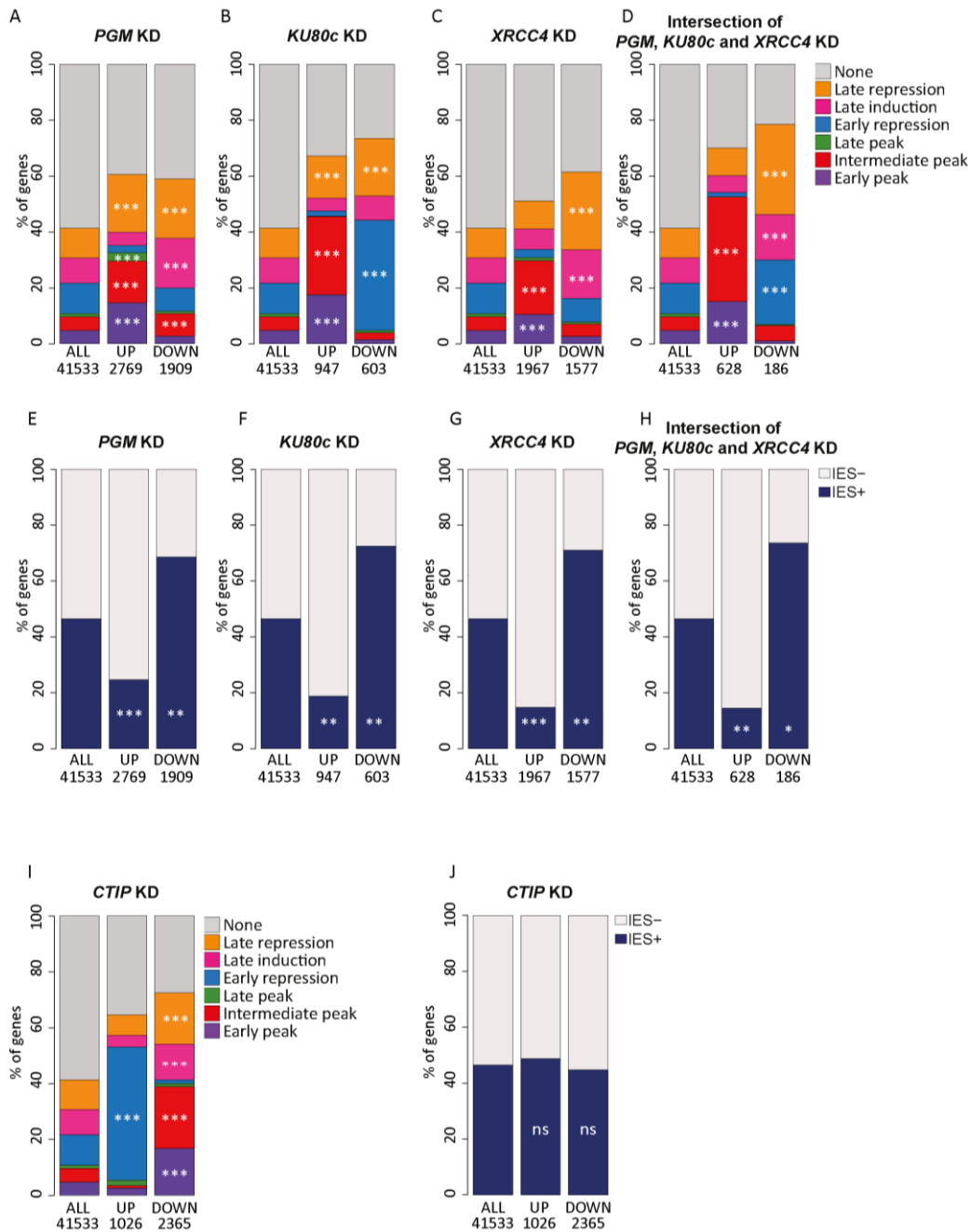
Supplementary Figure S5. Progression of autogamy stages in the different time-courses

All data were extracted from previously published work (references are indicated above each diagram).



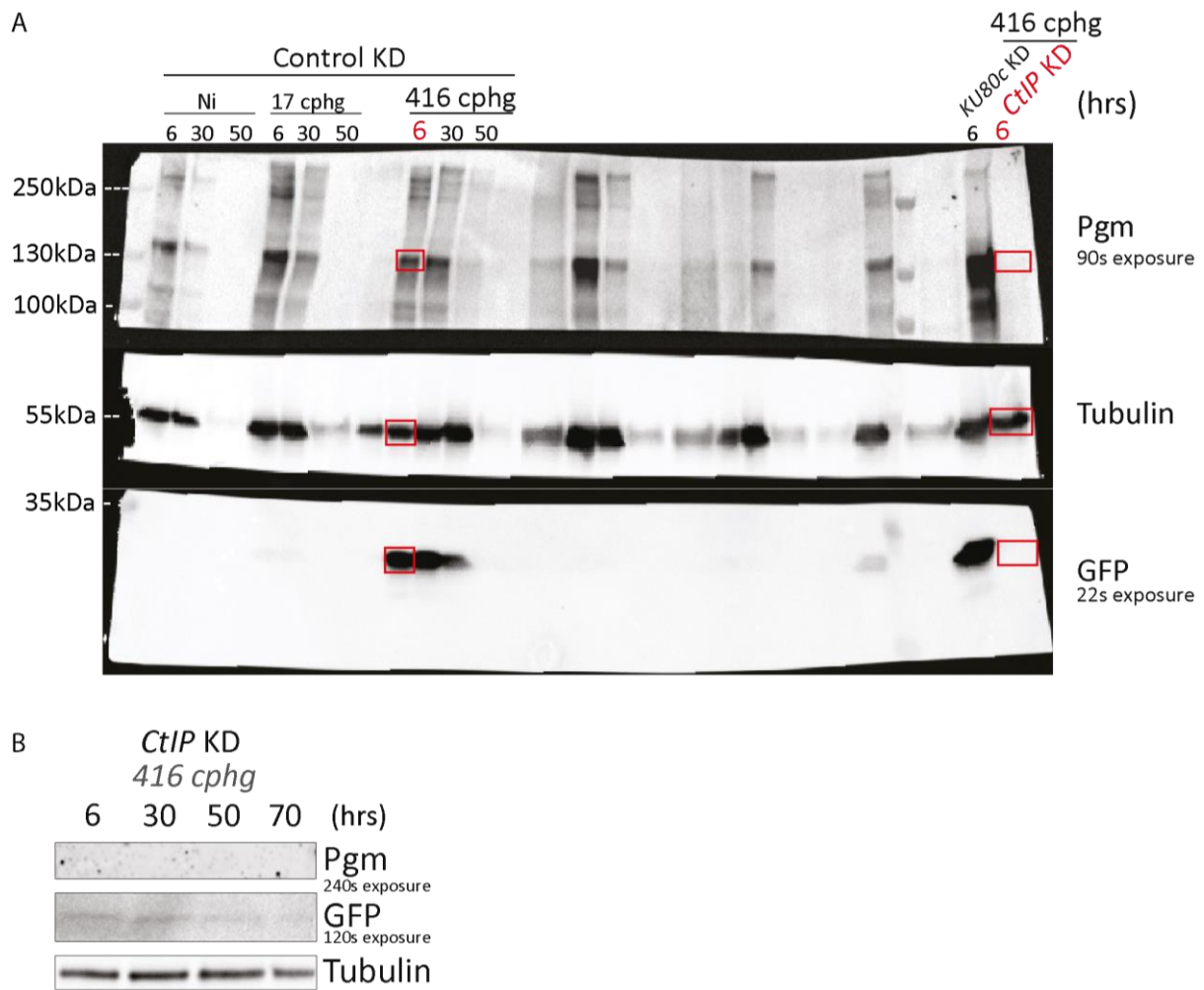
Supplementary Figure S6. Definition of the groups of biological replicates

(A) Table showing the 19 groups of replicates used for DE analysis, after manual curation and collapse of technical replicates for the ICL7 time-course (in bold). For each autogamy stage, all controls are grouped as biological replicates. (B) PCA of the vst normalized data extracted from the second round of DESeq2 (see part 2 of Supplementary Figure S4). (C) Hierarchical clustering (Pearson correlation matrix) of all datasets used in B (same color code). In B and C, the different time-points are colored according to the autogamy stages defined in A (VEGETATIVE, EARLY, INTERMEDIATE & LATE).



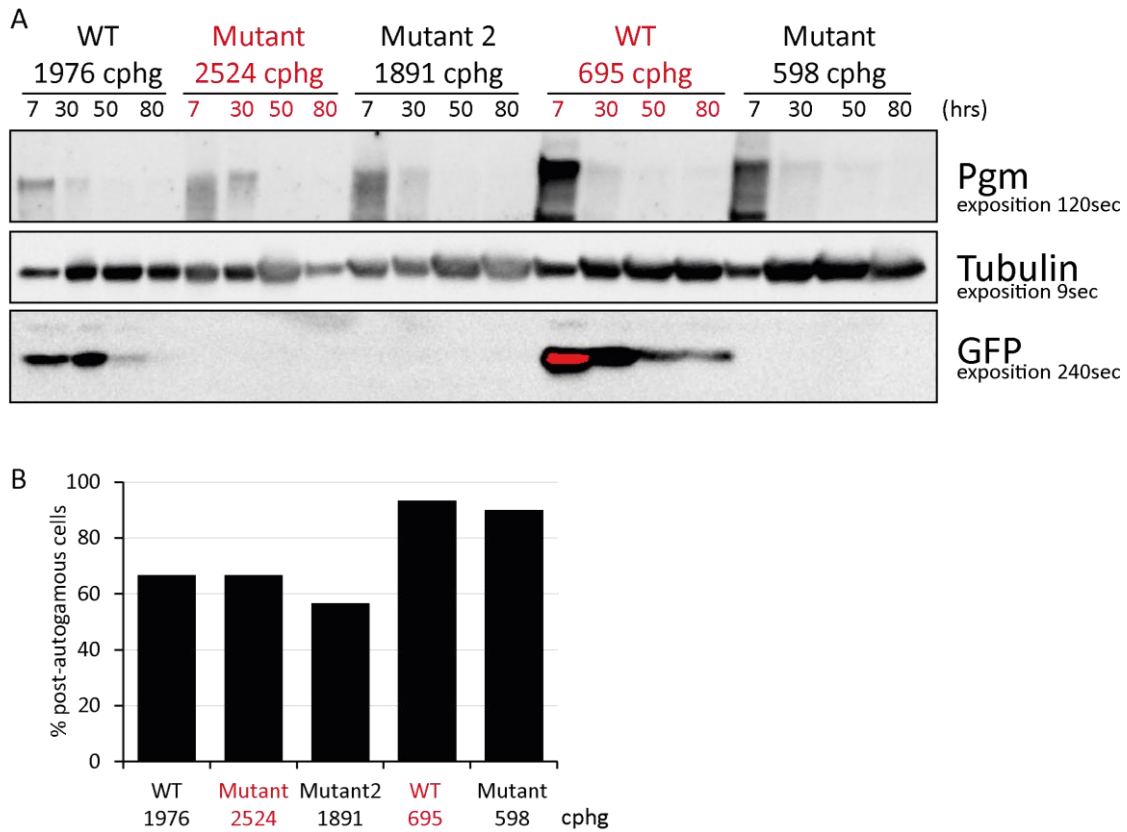
Supplementary Figure S7. Proportions of gene expression clusters and genes with IESs in their MIC version among the combinations of deregulated genes in the different KDs

(A-H) Proportion of each gene expression cluster (A-D) and genes with IES (E-H) among the deregulated genes at LATE stage in the different KD and the combination of the three KD. (I-J) Proportion of each gene expression cluster (I) and genes with IESs (J) among the deregulated genes at EARLY or INTERMEDIATE stage in *CtIP* KD. In A-D & I, *** stands for a p-value < 1.10⁻³. In E-H & J, ns stands for “not significant”, * for a p-value < 1.10⁻¹⁰, ** for a p-value < 1.10⁻²⁰ and *** for a p-value < 1.10⁻¹⁰⁰.



Supplementary Figure S8. Western blot analysis of Pgm downregulation in a *CtIP* KD

(A) Full-size image of the membrane used in Figure 3C. The progression of autogamy was followed for different injected clones carrying the *GFP*-reporter construct. The samples corresponding to the clone shown in Figure 3C (416 cphg) are boxed in red for the control and *CtIP* KD. The membrane was cut in three to simultaneously monitor the production of Pgm (top), Tubulin (middle) and GFP (bottom) using appropriate antibodies. (B) Western blot analysis of Pgm and GFP production during autogamy in a *CtIP* RNAi for the injected clone shown in A.



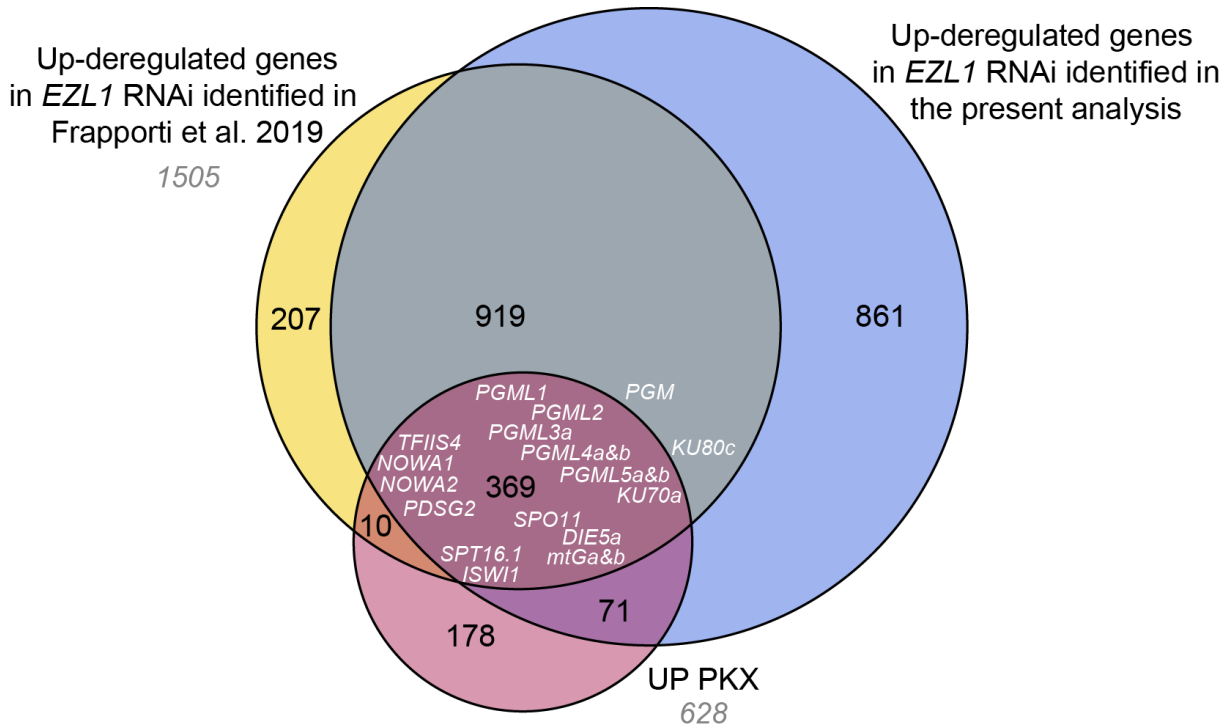
Supplementary Figure S9. Effect of the mutated motif on the *GFP*-reporter transgene expression

(A) Western blot quantification of the levels of endogenous Pgm and GFP expressed from the reporter constructs injected in different clones at various cphg, during an autogamy time-course of cells subjected to control RNAi. The *GFP*-reporter transgenes carry in their promoters the wild-type motif (WT: AAAATCATTAAATATTT) or mutated versions (mutant: AACATTAATAAAATATTT, mutant 2: AAATGATTTTAAATATTT, the mutated portion of the motif are underlined). Time-points are in hours (hrs) following T0. The time courses in red correspond to those presented in the Figure 5B. In the GFP panel (bottom), the red label indicates that the intensity of the signal was saturating. (B) Survival tests of each clone presented in panel A after a control RNAi.



Supplementary Figure S10. Alignment of the nucleotide sequences of the *PGM* gene start and its upstream region from 11 *Paramecium* species.

Sequences were extracted from ParameciumDB ((32), based on (55–57)). The *P. tetraurelia* sequence is annotated below the alignment and the nucleotide substitutions found in other species are highlighted in pink. The conserved motif is boxed in red, the *PGM* start codon and the stop codon of the upstream gene are boxed in black.



Supplementary Figure S11. Genes overexpressed in *PGM*, *KU80c*, *XRCC4* and *EZL1* RNAi.

Venn diagram showing the intersection of the genes identified as overexpressed in *PGM*, *KU80c* and *XRCC4* RNAi (UP PKX) with the genes overexpressed in *EZL1* RNAi in (40) and in this study. The 18 known UP PKX genes and the *PGM* and *KU80c* genes are indicated in white.

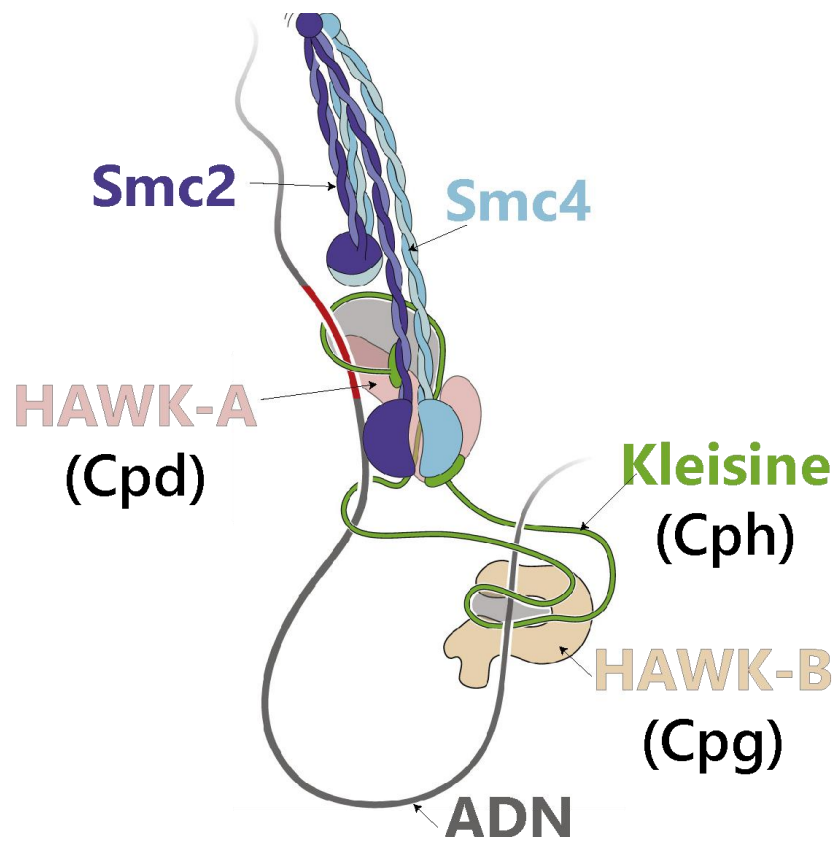


Figure 43 - Complexe condensine

Complexe condensine (Oldenkamp and Rowland, 2022). Les noms des protéines de paramécie correspondants sont indiqués entre parenthèses.

Chapitre 3 - Caractérisation de l'effet d'un ARNi contre le complexe condensine

Les condensines sont des complexes protéiques extrêmement conservés puisqu'on les trouve dans les trois domaines du vivant : les eucaryotes, les procaryotes et les archées (Hirano, 2012). Elles sont principalement connues pour leur rôle dans la compaction de l'ADN sous forme de chromosomes lors de la mitose et la méiose des cellules. Il existe deux complexes condensine chez les eucaryotes : condensine I et condensine II. La condensine II se localise dans les noyaux, elle permet une première condensation des chromosomes en formant de larges boucles d'ADN au cours de la prophase ce qui permet le raccourcissement des chromosomes. La condensine I est quant à elle présente uniquement dans le cytoplasme et n'interagit avec l'ADN qu'après la disparition de l'enveloppe nucléaire lors de la prométaphase. Elle permet de former des boucles sur les boucles formées par la condensine II augmentant ainsi le niveau de compaction de l'ADN (Davidson and Peters, 2021; Ganji et al., 2018; Hirano, 2012; Oldenkamp and Rowland, 2022). La condensine I est présente chez tous les eucaryotes mais pas la condensine II. En effet, celle-ci est notamment absente chez les champignons, les excavâtes et le cilié *T. thermophila* (Hirano, 2012).

Les condensines font partie des complexes Smc (*Structural Maintenance of Chromosome*). Chez les eucaryotes les condensines forment des pentamères (Figure 43) composés des sous-unités Smc2 et Smc4, d'une kleisine et de deux

	Vertébrés	<i>T. thermophila</i>	<i>P. tetraurelia</i>
Smc2	Smc2	Smc2	Smc2
Smc4	Smc4	Smc4	Smc4-1, Smc4-2
Kleisine	CAP-H	Cph1, Cph2, Cph3, Cph4	Cph3a, Cph3b, Cph4a, Cph4b
HAWK-A	CAP-D2	Cpd1, Cpd2	Cpd1-1a, Cpd1-1b, Cpd1-1c, Cpd1-2a, Cpd1-2b, Cpd2
HAWK-B	CAP-G	Cpg1	Cpg1

Tableau 2 - Sous-unités des condensines I des vertébrés et des ciliés

Noms des sous-unités du complexe identifiées chez les vertébrés et *Tetrahymena* (Hirano, 2012) ainsi que les homologues identifiés chez la paramécie (Communication personnelle Valerio Vitali).

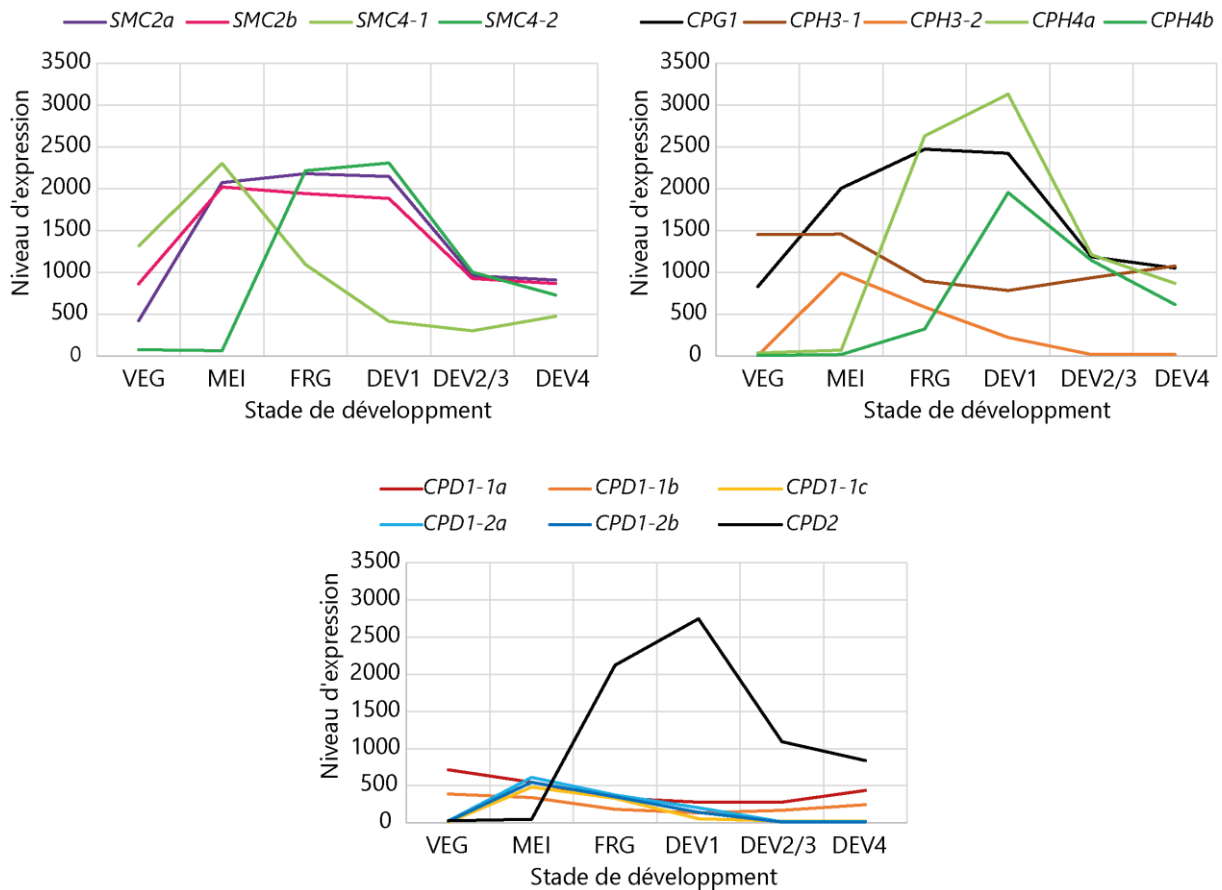


Figure 44 - Profils d'expression des gènes du complexe condensine

Profils d'expression moyens (Arnaiz et al., 2017) des gènes prédits comme participant aux complexes condensine de la paramécie. VEG = végétatif, MEI = méiotique, FRG = fragment, DEV = développement des nouveaux MAC (DEV1 : T5-T10, DEV2/3 : T11-T19, DEV4 : T20-T50) (Le niveau d'expression correspond aux comptages normalisés de DESeq2 (Love et al., 2014)).

protéines porteuses de répétitions HEAT (*Huntingtin/EF3/PP2A/Tor1*) : les HAWK (*HEAT proteins associated with kleisins*). Les complexes des condensines I et II diffèrent par leur kleisine et leurs HAWK. La condensine I comporte la kleisine Cph1 (ou CAP-H1, *Chromosome Associated Protein H1*) et les HAWK Cpd2 (ou CAP-D2) et Cpg (ou CAP-G) tandis que la condensine II inclut la kleisine Cph2 et les HAWK Cpd2 et Cpg2 (Davidson and Peters, 2021; Hirano, 2012; Yatskevich et al., 2019).

Chez le cilié *T. thermophila*, un homologue a pu être identifié pour les Smc2 et 4 ainsi que pour Cpg (Hirano, 2012). De plus, des homologues de Cpd2 et Cph1 ont pu être identifiés suggérant que le complexe correspond bien à une condensine I (Cervantes et al., 2006; Howard-Till and Loidl, 2018; Howard-Till et al., 2019). Il y aurait plusieurs complexes condensine I selon les homologues de Cpd2 et Cph1 présents dans le complexe. La particularité étant que tous ces complexes condensine I ont une localisation nucléaire chez le cilié.

3.1 LES CONDENSINES DE LA PARAMÉCIE

Comme *Tetrahymena*, la paramécie ne possède pas de complexe condensine II mais uniquement des homologues au complexe condensine I (Tableau 2). Une première analyse du génome a permis l'identification d'homologues aux Smc2 et 4 (ParameciumDB). Plus récemment, une analyse par alignement de séquences et recherche de régions conservées par HMM (*Hidden Markov Models*) a également permis l'identification des homologues chez la paramécie pour les sous-unités spécifiques des condensines I (Analyse réalisée par Valerio Vitali) (Tableau 28 en annexe).

La paramécie possède deux homologues pour *SMC2* (*SMC2a* et *SMC2b*) exprimés au stade végétatif et dont l'expression est maximale durant la méiose et le développement des nouveaux MAC (Figure 44). Elle possède deux paralogues pour *SMC4* (*SMC4-1* et *SMC4-2*). *SMC4-1* est exprimé au stage végétatif et lors de la méiose des MIC tandis que *SMC4-2* est exprimé spécifiquement lors du développement des nouveaux MAC (Figure 44). En plus des Smc, quatre homologues de la kleisine Cph1 ont été identifiés : les *CPH3-1*, *CPH3-2*, *CPH4a* et

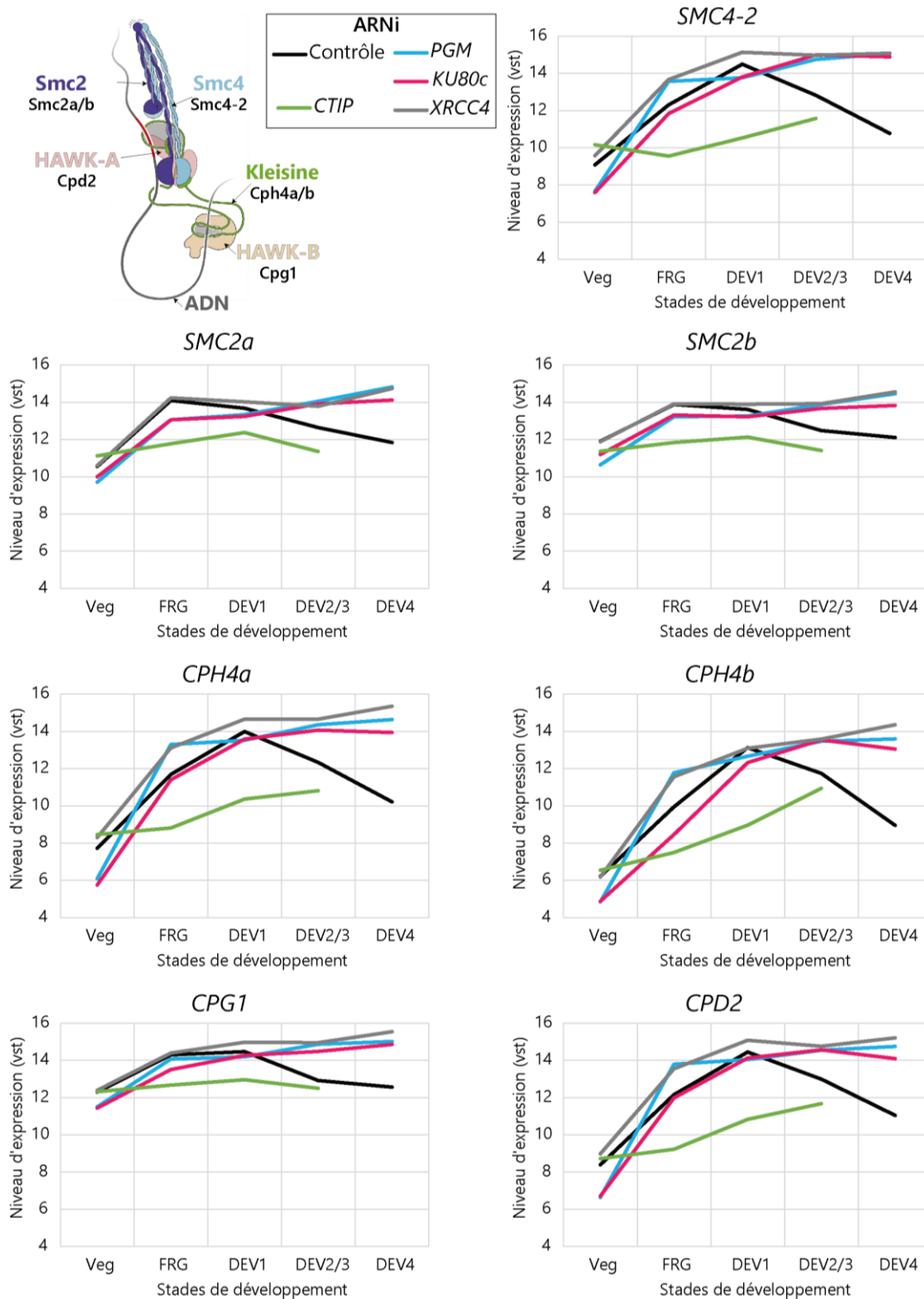


Figure 45 - Profils d'expression des composants du complexe condensine D de *P. tetraurelia*

Complexe condensine (Oldenkamp and Rowland, 2022) avec les noms des protéines de paramécies (en noir). Les profils d'expression des gènes exprimés spécifiquement au cours des RPG après des ARNi contrôles (moyenne des ARNi contre *ND7* ou *ICL7*) ou des ARNi dirigés contre *PGM*, *KU80c* ou *XRCC4*. Veg = Végétatif, FRG = Fragment (T0-T4), DEV : Développement des nouveaux MAC (DEV1 : T5-T10, DEV2/3 : T11-T19, DEV4 : T20-T50). Le niveau d'expression correspond à la moyenne des normalisations vst (*variance stabilizing transformation*, DESeq2 (Love et al., 2014)) des points de chaque stade.

CPH4b. *CPH3-1* est exprimé pendant toute la vie de la cellule mais son expression décroît lors du développement des nouveaux MAC. Son paralogue, *CPH3-2*, est exprimé spécifiquement pendant la méiose des MIC mais à plus faible niveau. Tandis que *CPH4a* et son ohnologue *CPH4b* sont exprimés spécifiquement pendant le développement des nouveaux MAC (Figure 44). La paramécie possède six homologues de Cpd2 répartis en trois groupes paralogues (Cpd1-1, Cpd1-2 et Cpd2). *CPD1-1a* et son ohnologue *CPD1-1b* sont principalement exprimés lors de la vie végétative de la cellule. Le troisième ohnologue, *CPD1-1c* et les deux ohnologues *CPD1-2a* et *CPD1-2b* ne sont exprimés que lors de la méiose des MIC. Enfin, un unique homologue a été identifié pour Cpg : Cpg1 dont le profil d'expression du gène ressemble à celui des *SMC2*.

Bien que la paramécie ne possède pas de condensine II, il est très probable qu'elle possède différents complexes condensine I ayant chacun leur spécialisation comme cela a été proposé chez *T. thermophila* (Howard-Till and Loidl, 2018; Howard-Till et al., 2019). Basé sur les profils d'expression, on peut d'ores et déjà séparer les gènes en deux groupes avec d'un côté : *SMC4-1*, les *CPH3*, les *CPD1* exprimés au cours de la vie végétative et lors de la méiose des MIC, et de l'autre *SMC4-2*, les *CPH4* et *CPD2* exprimés spécifiquement lors du développement des nouveaux MAC. A ces deux groupes s'ajoutent les gènes *SMC2* et *CPG1* qui sont exprimés lors de la vie végétative de la cellule et dont l'expression est maximale lors de la méiose des MIC et le développement de nouveaux MAC. Par la suite, ces deux groupes seront respectivement nommés condensine V (végétative) et condensine D (développementale) (les *SMC2* et *CPG1* sont considérés comme faisant partie des deux groupes). Il est intéressant de remarquer que chacun recouvre l'ensemble des composants de la condensine I. On peut donc supposer que des protéines forment des complexes condensine spécifiques du développement des nouveaux MAC (Schéma Figure 45) et de la phase de croissance végétative/méiose.

Une des principales différences entre les protéines condensine I de la paramécie par rapport aux complexes des vertébrés ou de *Tetrahymena* est la présence de deux Smc4 (Tableau 2) : Smc4-1 qui fait partie des condensines V et Smc4-2 qui fait partie des condensines D. J'ai donc utilisé des ARNi contre ces deux protéines afin d'étudier les deux complexes théoriques.

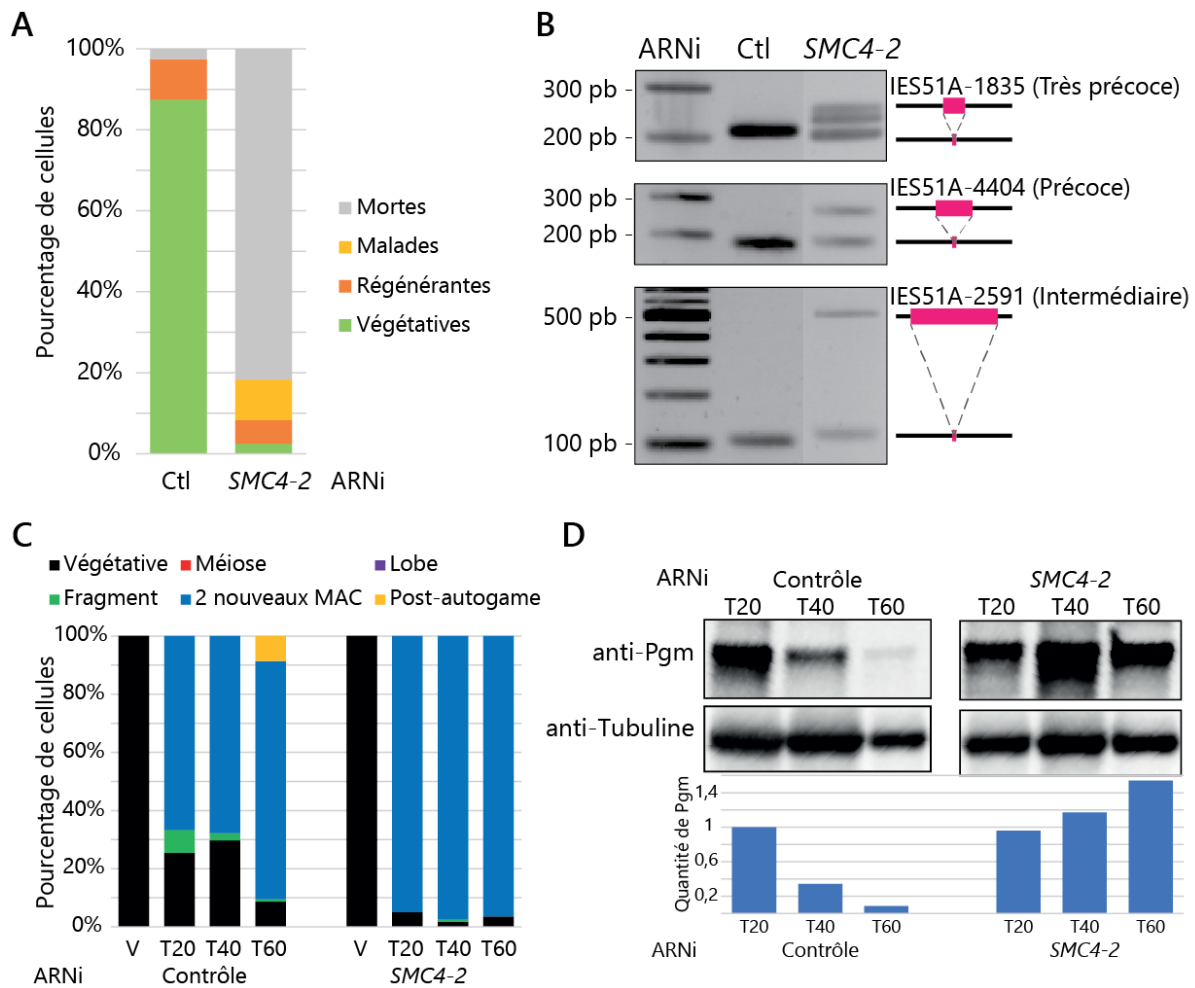


Figure 46 - Smc4-2 est nécessaire au déroulement des RPG

A Test de survie après un ARNi contrôle (contre *ND7* ou en utilisant le plasmide L4440 vide). **B** Test de rétention des IES dans des conditions d'ARNi contrôle (L4440). Entre parenthèses sont indiqués les groupes d'excision auxquels appartiennent les IES (voir section 2.2.2.2 de l'Introduction). **C** Suivi des stades de développement après passage de l'autogamie en condition contrôle (ARNi *ND7*) et lors d'un ARNi contre *SMC4-2*. Le temps est indiqué en heures après le T0. **D** (haut) Western blot avec un anti-Pgm et anti-Tubuline sur la cinétique présentée en C. Le temps est indiqué en heures après le T0. (bas) Quantification du signal Pgm normalisée sur la tubuline.

Les gels présentés dans le panneau B ont été réalisés par Valerio Vitali et les résultats des panneaux C et D ont été produits par Aurélie Camprodon sous ma supervision.

3.2 UN COMPLEXE CONDENSINE DÉVELOPPEMENTAL

3.2.1 Une dérégulation spécifique des condensines D

Dans le chapitre précédent, j'ai identifié une liste de gènes candidats, pouvant être impliqués dans les RPG. Pour cela, les gènes doivent être exprimés spécifiquement lors des réarrangements (*early* ou *intermediate peaks*) et être surexprimés au temps tardif lors d'un ARNi contre *PGM*, *KU80c* ou *XRCC4*. Une partie des condensines précédemment identifiées se trouve parmi ces gènes (Figure 45). Plus particulièrement, on retrouve les gènes *CPD2*, *CPG1*, *CPH4a*, *CPH4b*, *SMC2a*, *SMC2b* et *SMC4-2*, c'est-à-dire l'ensemble des gènes condensines spécifiquement exprimés lors des RPG. Seules les condensines D sont affectées par des défauts dans le déroulement des RPG, ce qui suggère que ces condensines pourraient avoir une fonction qui diffère des condensines V.

De plus, l'expression de toutes les condensines D est diminuée lors d'ARNi dirigé contre *CtIP* (Figure 45). Leur expression, comme celle de *PGM*, dépend donc de la présence des nouveaux MAC en développement (ou du noyau zygotique).

Enfin, *CPD2*, *CPH4a*, *CPH4b* et *SMC4-2* (toutes protéines spécifiques des condensines D) portent dans leur promoteur une version du motif consensus précédemment identifié à la position enrichie pour les gènes ayant le même profil d'expression que *PGM* (Tableau 28 en annexe). Il est donc possible que ces gènes soient corégulés avec *PGM* par l'intermédiaire d'un même facteur de transcription.

3.2.2 Effets d'un ARNi contre une condensine D

Afin d'évaluer l'effet de l'absence d'une condensine D sur le déroulement de l'autogamie, les cultures sont montées en volume. Après avoir passé trois jours dans du milieu induisant un ARNi contre *SMC4-2*, elles sont placées en carence de nourriture afin d'induire le passage de l'autogamie.

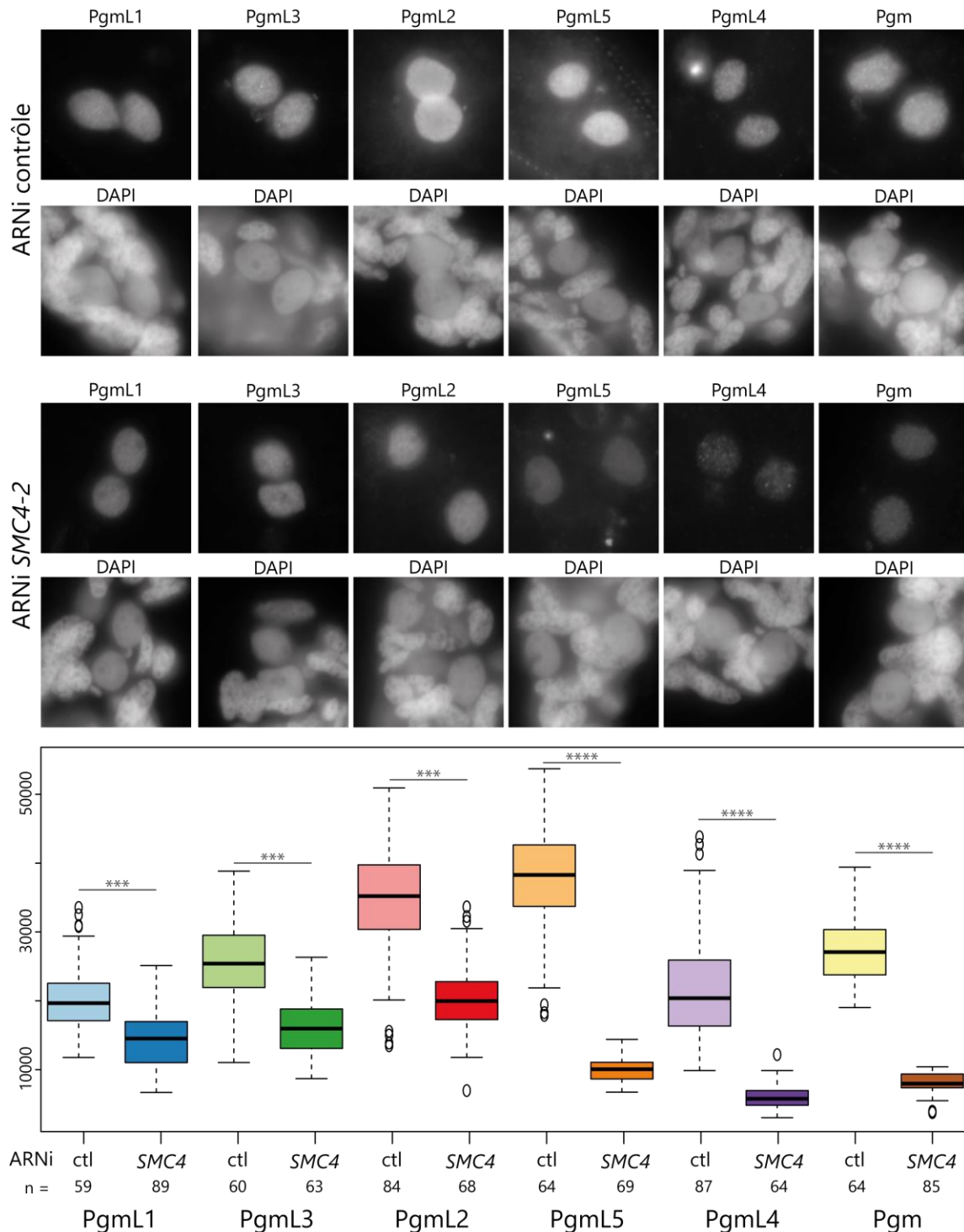


Figure 47 - Déstabilisation du complexe d'excision lors d'un ARNi contre *SMC4-2*

Immunofluorescence avec des pré-extractions de noyaux prélevés à T7 dans des conditions d'ARNi contrôle (L4440) ou contre *SMC4-2* marqué par des anticorps dirigés contre Pgm ou les PgmL (**Haut**) Image des différentes conditions réalisées par Valerio Vitali. (**Bas**) Quantification du signal des anticorps dans les nouveaux MAC (ctl = condition contrôle, n= indique le nombre de noyaux quantifiés pour chaque conditions). L'ordre de Pgm et des PgmL correspond à la hiérarchie d'assemblage du complexe d'excision (voir section 3.2.2 de l'Introduction). La significativité de la différence entre les conditions contrôle et ARNi *SMC4-2* est mesurée par un test de Mann-Whitney (***) = p-value < 10⁻¹⁰, **** = p-value < 10⁻²⁰).

3.2.2.1 Déficit du développement des nouveaux MAC

La qualité du passage de l'autogamie est mesurée par un test de survie. Ce test révèle qu'en l'absence de *Smc4-2* la descendance issue de l'autogamie n'est pas viable (Figure 46A). En effet, un test d'excision des IES sur de l'ADN génomique (ADNg) prélevé quatre jours après le début de l'autogamie révèle la présence de formes non-excisées des IES 51A1835, 51A4404 et 51A2591 (Figure 46B). La présence de la forme excisée peut être due à une excision partielle de l'IES, mais elle est plus probablement due à la contamination par les fragments de l'ancien MAC. Dans tous les cas, la présence d'une forme retenue est le signe que *Smc4-2* est nécessaire à la bonne réalisation de l'excision des IES. De plus, parmi les IES testées, 51A1835 fait partie des IES très précoces, elle est donc normalement très efficacement excisée. Le fait que de la rétention soit observée pour cette IES suggère un rôle important de *Smc4-2* dans les RPG.

3.2.2.2 Dérégulation de la machinerie d'excision

Comme exposé dans le Chapitre 1 -de la partie Résultat, un défaut dans les RPG semble conduire à une surexpression de la machinerie d'excision même si toutes les IES ne sont pas complètement retenues. Lors d'un ARNi contre *SMC4-2*, les deux nouveaux MAC se forment correctement (Figure 46C). Cependant, le défaut dans l'excision des IES est suffisamment important pour induire une surexpression de *Pgm* aux temps tardifs, visible par western blot (Figure 46D).

3.2.2.3 Déstabilisation des *Pgm* et *PgmL* dans les nouveaux MAC

Les RPG s'accomplissent correctement lorsque le complexe d'excision se localise stablement dans les nouveaux MAC. Un défaut dans l'excision des IES peut être dû à un défaut de la localisation d'éléments du complexe d'excision. Une immunofluorescence avec pré-extraction au Triton ciblant *Pgm* et les différentes *PgmL* lors d'un ARNi contre *SMC4-2* montre une baisse significative de la localisation de ces protéines dans les nouveaux MAC à 7 heures après le début de l'autogamie (T7) par rapport aux conditions contrôle.

Mais, bien que l'intensité du signal soit réduite, le signal pour toutes les *PgmL* et pour *Pgm* reste qualitativement le même : le signal est homogène et légèrement granuleux pour les *PgmL*, à l'exception de *PgmL4* qui, comme *Pgm*,

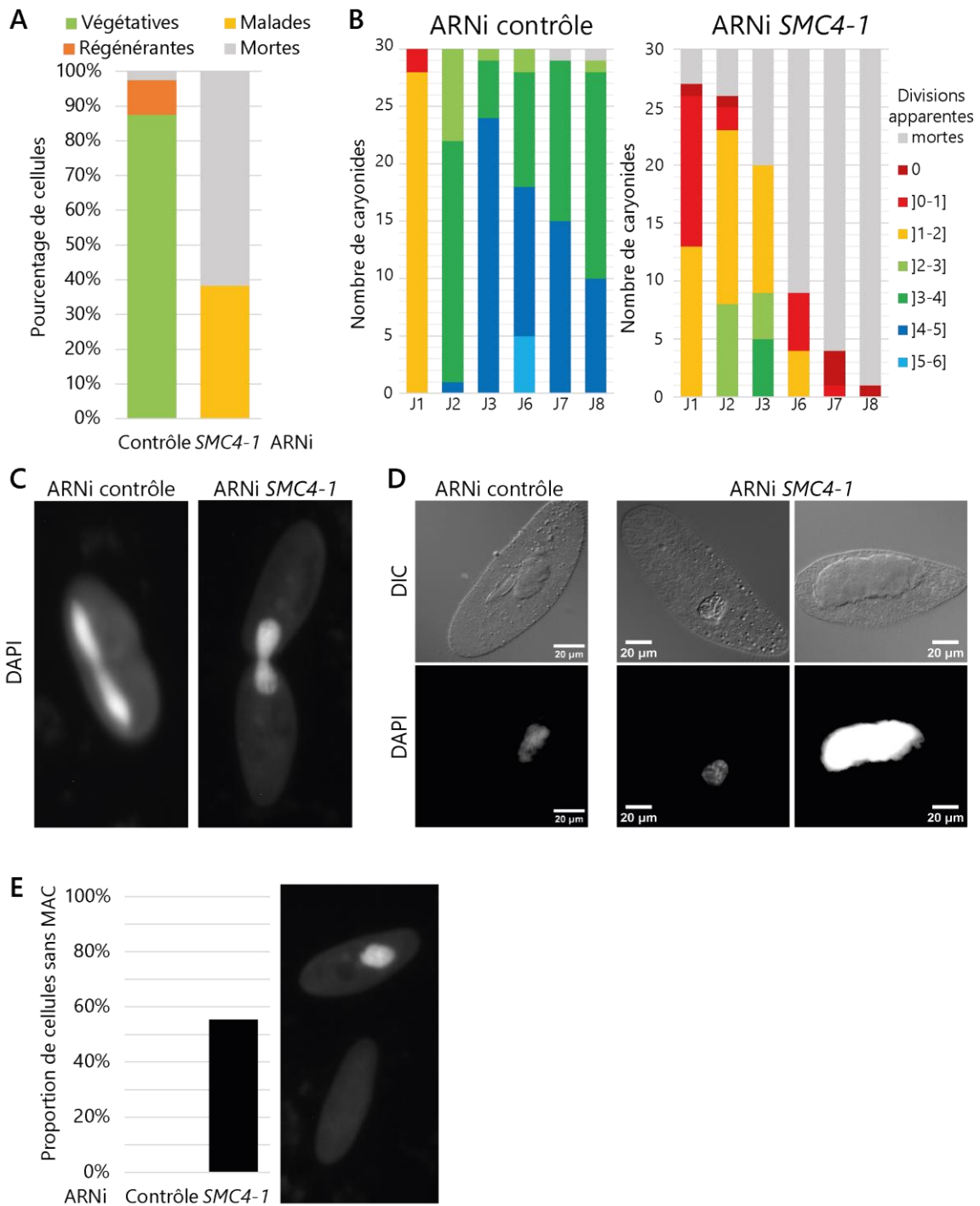


Figure 48 - *Smc4-1* est nécessaire à la ségrégation des MAC lors de la croissance végétative

A Test de survie de cellules après 3 jours d'ARNi contrôle (contre *ND7*) ou contre *SMC4-1*. **B** Suivi de la croissance végétative en condition d'ARNi contrôle (contre *ND7*) ou contre *SMC4-1*. **C** Image d'un diviseur en condition contrôle et lors d'un ARNi *SMC4-1* observé à la loupe à épifluorescence après un marquage DAPI. **D** Exemple de cellules marquées au DAPI après un ARNi contrôle ou contre *SMC4-1*. **E** (droite) Proportion de cellules sans noyau. (gauche) Cellules sans noyau observées à la loupe à épifluorescence lors d'un ARNi contre *SMC4-1* après marquage au DAPI. (Données produites par Aurélie Campronon sous ma supervision)

présente des foci. Bien que plus faiblement associés aux noyaux, les différents composants du complexe semblent toujours se localiser correctement. Étant donné que l'efficacité de l'ARNi contre *SMC4-2* n'a pas été mesurée, il est possible qu'une faible quantité de Smc4-2 soit toujours produite. Les Smc4-2 résiduelles pourraient être à l'origine du faible signal Pgm et PgmL observé dans les noyaux. Il est également possible que le complexe soit stabilisé par d'autres protéines.

3.3 UN COMPLEXE CONDENSINE VÉGÉTATIF

Smc4-2 semble donc être nécessaire au développement des nouveaux MAC. Son paralogue, Smc4-1 est quant à lui exprimé durant la vie végétative de la cellule et au cours de la méiose des MIC. Il pourrait donc jouer un rôle distinct.

3.3.1 Un complexe condensine V

Le groupe des condensines V se compose de : *SMC4-1*, les *CPH3*, les *CPD1* ainsi que *CPG1* et *SMC2*. Ces gènes codent pour l'ensemble des protéines nécessaires à la formation d'un complexe condensine I complet (Tableau 2) étant donné que les *CPH3* codent pour des homologues de la kleisine, les gènes *CPD1* et le gène *CPG1* sont respectivement des homologues des protéines HAWK A et B. Il est donc possible que, comme pour les gènes de condensine D, les gènes du groupe des condensines V codent pour les différents partenaires d'un même complexe.

3.3.2 Effets d'un ARNi contre une condensine V

En plus d'avoir des profils d'expression différents de celui des condensines D, l'expression des condensines V n'est pas affectée par des défauts dans l'accomplissement des RPG (Tableau 28 en annexe). Le rôle de ces condensines n'est donc probablement pas directement lié aux RPG.

3.3.2.1 Mortalité lors de la croissance végétative

Afin d'évaluer l'effet de l'absence d'une condensine V sur la survie des cellules, j'ai effectué un test de survie des cellules après trois jours d'ARNi contre *SMC4-1* (Figure 48A). Ces trois jours d'ARNi suffisent pour induire une mort d'une portion importante des cellules. Les cellules survivantes montrant des problèmes

pour se diviser sont considérées comme malades. En effet, de nombreuses cellules ne se divisent pas à la vitesse habituelle. De plus, les diviseurs restent souvent accrochés les uns à la suite des autres pouvant former des chaînes de plusieurs cellules.

La croissance végétative de 30 paramécies dans des conditions d'ARNi contre *SMC4-1* a ensuite été suivie sur plusieurs jours (Figure 48B). Le phénotype le plus important reste la mort des cellules. En effet après seulement huit jours dans le milieu d'ARNi contre *SMC4-1*, une cellule unique a survécu sur les 30 paramécies initialement suivies. En plus de l'effet sur la survie des cellules, on remarque également que les cellules semblent se diviser deux fois moins vite dans les conditions d'ARNi contre *SMC4-1* comparé aux conditions contrôles. Étant donné la mortalité observée, il est possible que les cellules se divisent à la même vitesse mais que la moitié d'entre-elles meurent. De plus, comme lors du test de survie, les diviseurs semblent rester longtemps attachés ensemble dans les conditions d'ARNi contre *SMC4-1*. Regarder les cellules ayant passé deux jours dans le milieu induisant un ARNi contre *SMC4-1*, montre une augmentation du nombre de cellules en chaîne et des chaînes plus longues (plus de 5 cellules peuvent se trouver les unes à la suite des autres).

3.3.2.2 Problème de ségrégation du MAC parental

Une observation des cellules à la loupe (Figure 48C) ou au microscope (Figure 48D) après un marquage des noyaux au DAPI révèle un problème de ségrégation des MAC au cours des divisions végétatives. En effet, au cours de la division végétative dans des conditions contrôle, le MAC s'étire sur tout le long du diviseur (Figure 48C). Cependant, dans des conditions d'ARNi contre *SMC4-1*, le MAC ne s'étire pas entre les diviseurs et semble être bloqué au niveau du sillon de division. Cela pourrait expliquer les chaînes de cellules observées lors du test de survie et lors de la croissance végétative. La cytokinèse est bloquée par la présence du MAC dans le sillon de division.

Le problème de ségrégation du MAC lors d'ARNi *SMC4-1* donne des cellules avec un MAC proportionnellement plus petit ou plus gros comparé au contrôle (Figure 48D). Cela pourrait être le résultat d'une séparation arbitraire du MAC

entre les deux cellules filles ou à un défaut d'amplification majeur de la portion du MAC reçu.

De plus, les problèmes de ségrégation du MAC entraînent également l'apparition de cellules sans MAC dans les conditions d'ARNi contre *SMC4-1* (Figure 48E). Plus de la moitié des cellules ne possèdent pas de MAC. Cela pourrait expliquer les problèmes de division apparente. En effet, il est peu probable que les cellules sans MAC survivent. Dans les conditions d'ARNi contre *SMC4-1*, les cellules semblent donc se diviser deux fois moins vite car la moitié des cellules meurent à chaque division. *Smc4-1* est donc nécessaire à la bonne ségrégation du MAC lors de la croissance végétative de la cellule. Elle est donc strictement nécessaire pour la survie de la population à long terme.

Chapitre 4 - Mise au point du protocole de Cut&Tag chez *P. tetraurelia*

L'identification de nouveaux acteurs permet d'avoir une image plus précise des mécanismes impliqués dans les RPG. Mais les RPG sont un processus global impliquant l'introduction de très nombreuses cassures double-brin qui doivent ensuite être réparées. Les RPG sont finement régulés pour la reconnaissance des séquences à éliminer, mais aussi pour l'ordre d'élimination des IES (voir Chapitre 3 -de la partie Introduction).

Le modèle actuel de la hiérarchie d'assemblage du complexe d'excision se base uniquement sur des données d'immunofluorescence. La comparaison d'immunofluorescence avec et sans pré-extraction a permis de déterminer les conditions d'association stable des composants du complexe d'excision dans les MAC en développement. Si la protéine est stablement localisée dans le noyau, dans des conditions d'immunofluorescence avec pré-extraction, on suppose par extension qu'elle est stablement associée à la chromatine. Mais les techniques d'immunofluorescence ne permettent pas de définir précisément où les protéines se localisent sur la chromatine. Pour cela, il faut en général passer par des techniques d'immunoprécipitation de la chromatine (ChIP, *Chromatin Immunoprecipitation*) nécessitant une quantité importante de noyaux.

Au cours de l'autogamie, trois types de noyaux co-existent dans la cellule : les nouveaux MIC, les nouveaux MAC et les fragments de l'ancien MAC. Les MIC étant beaucoup plus petits que les nouveaux MAC, il est plus facile de les éliminer par centrifugation. De plus, les MIC étant diploïdes, la contamination par les MIC

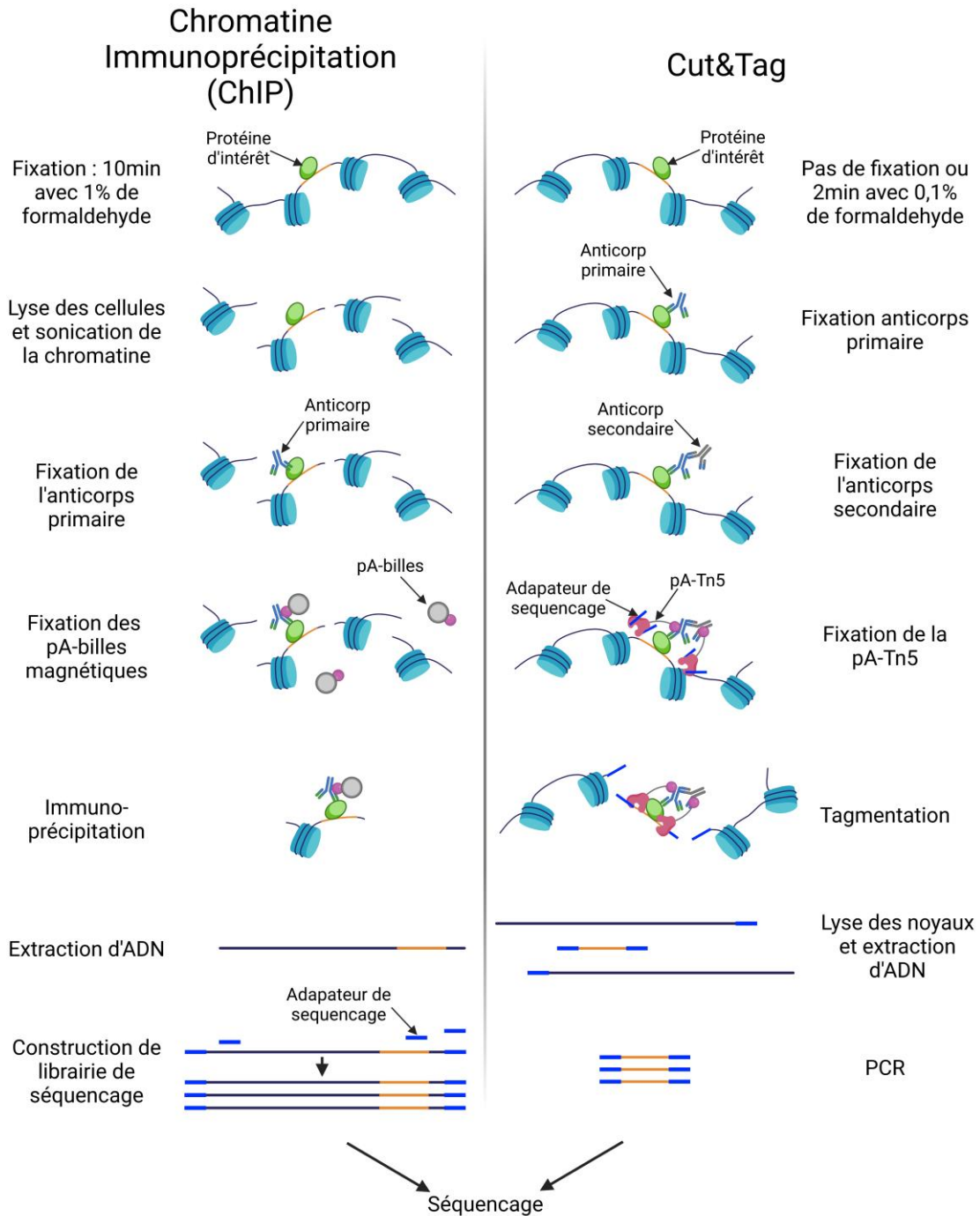


Figure 49 - Comparaison des protocoles de ChIP et de Cut&Tag

A **gauche**, le protocole d'immunoprécipitation de chromatine (ChIP) ciblant une protéine d'intérêt reconnue par un anticorps. L'anticorps est ensuite fixé à des billes magnétiques couplées à la protéine A (pA-billes). L'ADN est ensuite purifié et des adaptateurs de séquençage sont ajoutés lors de la préparation des bibliothèques. A **droite**, le Cut&Tag avec des étapes successives de fixation de l'anticorps primaire puis secondaire suivies de la fixation de la transposase Tn5 fusionnée à la protéine A (pA-Tn5). Au cours de l'étape de tagmentation la pA-Tn5 ajoute directement les adaptateurs de séquençage et la bibliothèque est amplifiée par PCR. (Created with BioRender.com)

est faible face aux nouveaux MAC (30 à 120n au moment des RPG). La contamination la plus préoccupante est celle provenant des fragments de l'ancien MAC qui, en plus de faire une taille proche de celle des nouveaux MAC, contiennent une quantité d'ADN équivalente (~50n par fragment). Les techniques actuelles de purification des MAC en développement demandent de passer par de grands volumes de culture de paramécie suivie de tri par FANS (Zangarelli et al., 2022). Obtenir de grandes quantités de nouveaux MAC purifiés reste donc un défi.

Puisque la ChIP n'a été utilisée sur la paramécie que pour visualiser les marques d'histone et qu'elle nécessite une grande quantité de noyaux (Frapporti et al., 2019), j'ai décidé de me tourner vers la mise au point d'une technique nécessitant moins de matériel : le Cut&Tag (Henikoff et al., 2020; Kaya-Okur et al., 2020), afin de tenter de localiser le complexe d'excision sur la chromatine.

4.1 MISE AU POINT DU CUT&TAG

Comme la ChIP, le Cut&Tag se base sur la reconnaissance de la protéine d'intérêt par des anticorps. Mais contrairement à la ChIP, le Cut&Tag ne commence pas par une étape de fragmentation de la chromatine et ne nécessite donc que peu ou pas de fixation, limitant ainsi le bruit de fond.

La première étape du Cut&Tag est la fixation de l'anticorps primaire reconnaissant la protéine d'intérêt. Le signal est ensuite amplifié par l'utilisation d'un anticorps secondaire. Les anticorps sont ensuite reconnus par la protéine A fusionnée à la transposase Tn5. La Tn5 est donc recrutée au niveau de la protéine d'intérêt. La Tn5 est chargée avec des extrémités du transposon *Tn5* auxquelles sont fusionnés des adaptateurs de séquençage. Les adaptateurs sont donc directement liés à l'ADN au moment de l'intégration locale du transposon pA-Tn5 (tagmentation). De ce fait, l'ADN couvert par la protéine d'intérêt est directement marqué de part et d'autre par les adaptateurs de séquençage lors de la coupure introduite par l'intégration de la transposase. Une PCR permet ensuite d'amplifier directement uniquement l'ADN fragmenté lié aux adaptateurs en une librairie qui peut être séquencée (Figure 49).

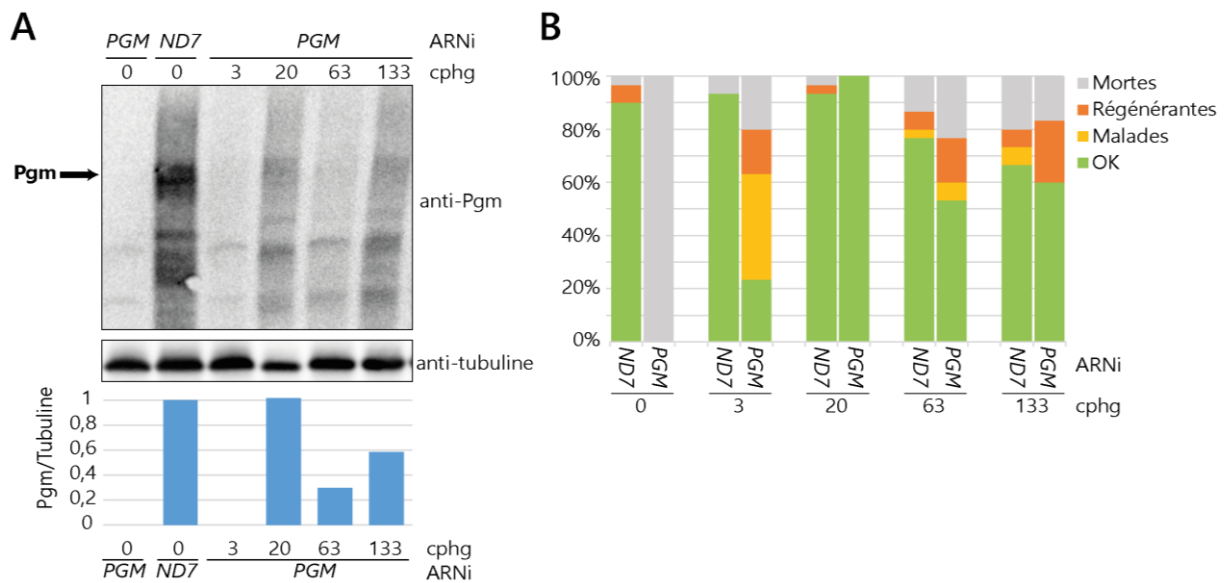


Figure 50 - Sélection de transformants exprimant un niveau endogène de Pgm

Les paramécies sont transformées avec un transgène 3xFlag-Pgm à différents nombres de copies (cphg, copie par génome haploïde). Elles sont ensuite cultivées dans un volume final de 50mL et récoltées à T6 pendant l'autogamie. **A** Sélection par western blot des transformants exprimant un niveau endogène de Pgm. **B** Test de survie. Tous les transformants (excepté celui à 3 cphg) sont issus du même caryonide.

Le Cut&Tag ne demandant pas d'étape de fixation, il est important de vérifier que la protéine d'intérêt est accessible par les anticorps dans les solutions du Cut&Tag. Le Cut&Tag se déroule principalement dans deux tampons HEPES qui ne diffèrent que par leur concentration en sel : 150 ou 300 mM NaCl. La concentration en sel est maintenue à 150mM NaCl jusqu'à l'ajout de la pA-Tn5 (protéine A fusionnée à la transposase 5). Une fois la pA-Tn5 ajoutée, la concentration en sel est augmentée à 300mM NaCl afin d'éviter que la Tn5 ne se fixe sur l'ADN de manière non-spécifique.

Le Cut&Tag est une technique dérivée du protocole de Cut&Run (Meers et al., 2019) qui utilise la MNase (*Micrococcal Nuclease*) à la place de la Tn5. L'utilisation de la MNase permet de s'affranchir de la solution à 300 mM NaCl, elle est donc souvent préférée au Cut&Tag pour visualiser des facteurs de transcription sur la chromatine. Cependant, lors du Cut&Run, les petits fragments d'ADN qui diffusent à l'extérieur des noyaux sont utilisés pour faire les librairies qui sont ensuite séquencées. Or, dans le cas des RPG de la paramécie, les IES excisées feraient également partie des fragments d'ADN capable de diffuser ce qui ajouterait un bruit de fond important. C'est la raison pour laquelle je me suis tournée vers le Cut&Tag plutôt que le Cut&Run.

4.1.1 Localisation des protéines cibles du Cut&Tag

Le Cut&Tag nécessite que les protéines que l'on souhaite visualiser restent associées à la chromatine dans des solutions à forte teneur en sel. Des expériences de biochimie menées par Julien Bischerour et Joël Acker tendent à montrer que Pgm et les PgmL forment un complexe qui reste stable même à de très fortes concentrations de sel (500 mM NaCl) (données non-publiées). Cependant, *in vitro* Pgm se détache de l'ADN passé les 250mM de NaCl (Bischerour et al., 2018). On peut donc se poser des questions sur la stabilité de Pgm sur la chromatine dans ces conditions, bien qu'*in vivo* d'autres facteurs (absents lors des tests *in vitro*) puissent augmenter la stabilité de Pgm sur la chromatine.

Dans un premier temps, j'ai visualisé la localisation des protéines et le bon fonctionnement des anticorps dans les conditions du Cut&Tag. Afin de vérifier que les anticorps restent présents dans les noyaux tout au long de la procédure, j'ai

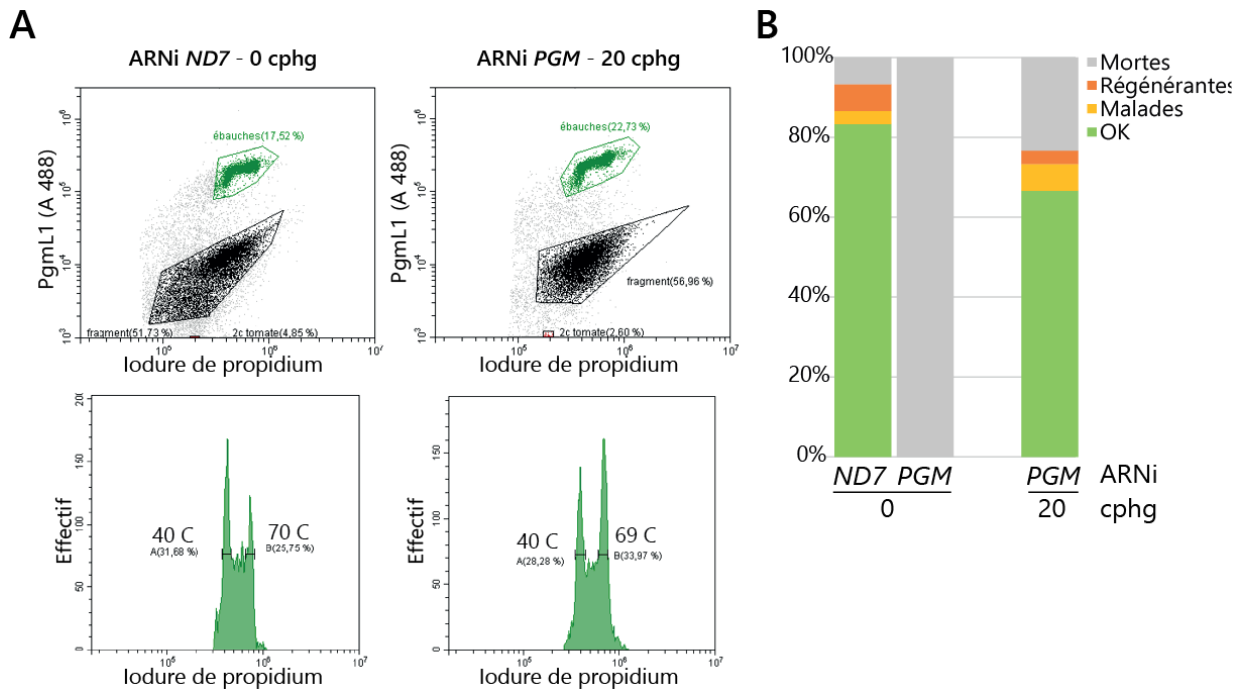


Figure 51 - Contrôle qualité des noyaux extraits

Les cultures sélectionnées Figure 50 sont ensuite cultivées dans un volume final de 500mL afin d'en extraire les noyaux. **A** La qualité et la ploïdie des noyaux extraits sont contrôlés par cytométrie en flux après marquage des noyaux par un anticorps anti-PgmL1 et un anticorps secondaire couplé à un fluorophore AF488 (Zangarelli et al., 2022). Les noyaux correspondant aux nouveaux MAC (ébauches) sont identifiés en vert (haut). Le contenu en ADN de ces noyaux est ensuite représenté pour évaluer la ploïdie (bas) **B** La qualité des cultures et du passage de l'autogamie est contrôlé par un test de survie.

utilisé des anticorps anti-H3K4me3. Pour marquer Pgm dans les noyaux deux marquages étaient possibles : utiliser des anticorps dirigés contre Pgm ou utiliser un transgène exprimant Pgm couplé à une étiquette.

4.1.1.1 Utilisation d'un transgène 3xFlag-Pgm

Afin de tester à la fois le marquage par l'anti-Pgm et l'anti-Flag, j'ai utilisé des cellules transformées avec un transgène porteur d'une version de *PGM* résistante à l'ARNi et précédée d'un 3xFlag (NSR-DYKDHD-G-DYKDHD-I-DYKDDD-DK) : 3xFlag-Pgm. J'ai effectué une première montée en volume afin d'extraire les protéines à T6 et de sélectionner les paramécies exprimant 3xFlag-Pgm à un niveau équivalent à l'expression endogène de Pgm. La quantité de Pgm dans les différentes cultures est ensuite évaluée par western blot (Figure 50A). Le clone injecté à 20 cphg (copies par génome haploïde) dans des conditions de complémentation (ARNi *PGM*) exprime une quantité similaire de Pgm comparé à un clone non injecté (0 cphg) dans des conditions d'ARNi contrôle (ARNi *ND7*). Le clone à 20 cphg est donc considéré comme exprimant Pgm à un niveau endogène. De plus, le test de survie (Figure 50B) montre qu'à 20 cphg la construction complémente bien sans avoir d'effet toxique. On voit également que le nombre de copies de transgène injecté ne reflète pas forcément la quantité de protéine exprimée et donc pas le niveau de complémentation.

Le clone à 20 cphg est ensuite monté une 2^{ème} fois en volume dans des conditions de complémentation (ARNi *PGM*) cette fois jusqu'à un volume de 500mL et les noyaux sont extraits à T6. Une partie des noyaux est ensuite utilisée pour faire un contrôle de la qualité de la préparation. Les noyaux sont marqués à l'iodure de propidium et par des anticorps anti-PgmL1 révélés par un anticorps secondaire couplé à un fluorophore AF488. Le marquage des noyaux est ensuite observé en cytométrie en flux comme décrit dans (Zangarelli et al., 2022) (Figure 51A). On évalue ainsi le taux de dégradation des nouveaux MAC grâce au marquage PgmL1. Des noyaux dégradés présenteront une perte du marquage PgmL1 visible lors de la cytométrie (voir Figure 84 en annexe). Ici, les noyaux ne sont pas dégradés. L'analyse par cytométrie en flux des noyaux identifiés comme étant des nouveaux MAC permet également d'évaluer la quantité d'ADN présent dans les nouveaux MAC et donc de suivre les cycles d'endoduplication. Ici les

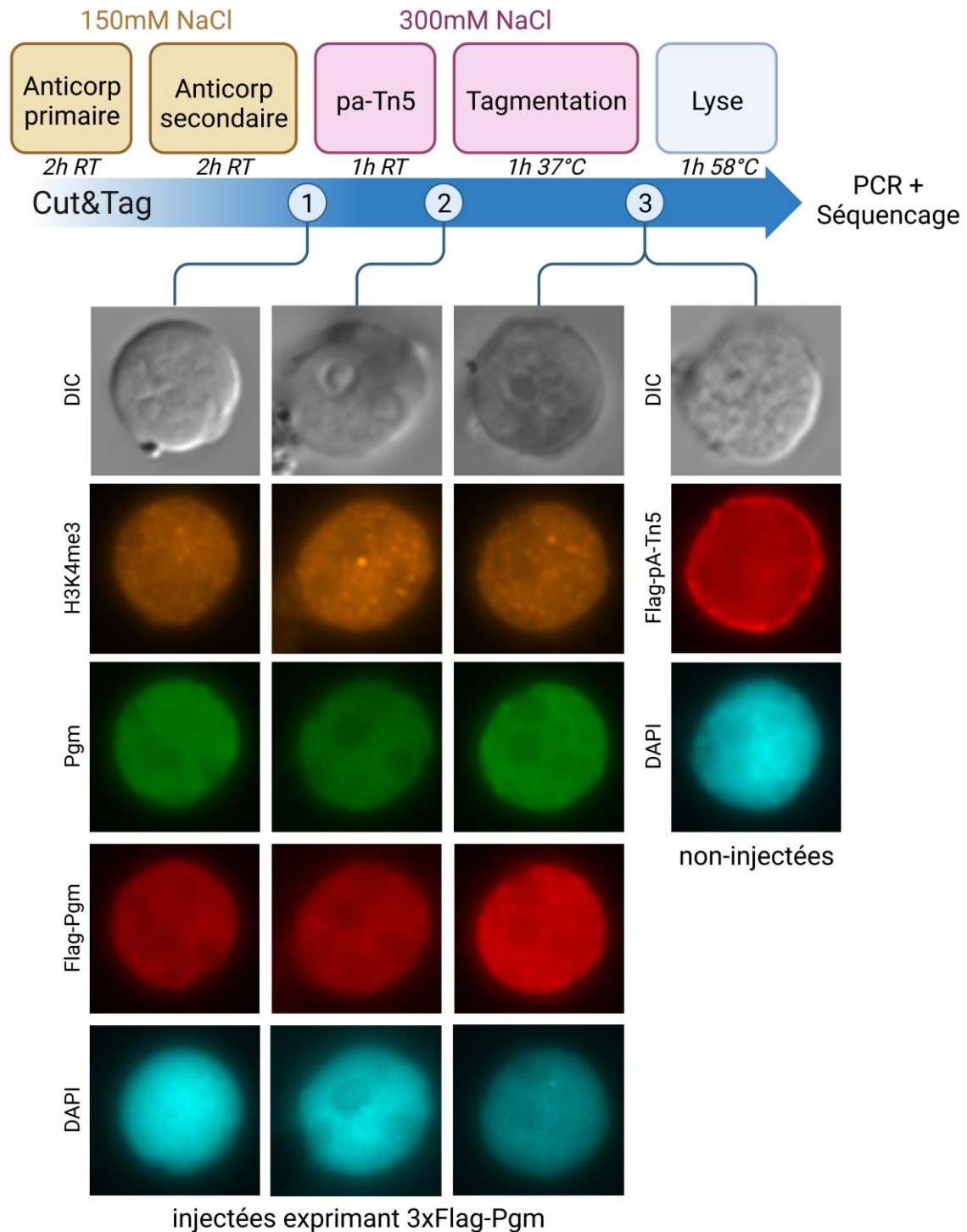


Figure 52 - Localisation des protéines au cours du Cut&Tag

(Haut) Représentation schématique des différentes étapes du Cut&Tag, avec les étapes dans les solutions à 150mM NaCl en jaune, celle à 300mM NaCl en rose et celle sans NaCl en bleu. Les noyaux ont été prélevés et marqués après les étapes de fixation des anticorps (1), après la fixation de la pA-Tn5 (2) et après l'étape de tagmentation (3). À **gauche**, les protéines cibles du Cut&Tag ont été marquées : H3K4me3 par un anti-H3K4me3 et Pgm par un anti-Pgm ou un anti-Flag. À **droite**, la pA-Tn5 a été marquée par un anti-Flag. (Created with BioRender.com)

noyaux atteignent ~40°C et ~70°C ce qui est attendu pour des cultures à T6 (Zangarelli et al., 2022). La qualité du passage de l'autogamie des grandes cultures est enfin évaluée par un test de survie (Figure 51B).

Ces noyaux sont ensuite utilisés pour faire les tests de localisation des protéines au cours du Cut&Tag.

4.1.1.2 Marquage des protéines lors du Cut&Tag

Après fixation des noyaux sur les billes de concanavaline et marquage par les anticorps primaires (anti-Pgm, anti-Flag et anti-H3K4me3), j'ai utilisé des anticorps secondaires couplés à un fluorophore (les anticorps secondaires normalement utilisés pour le Cut&Tag n'étant pas couplés). Enfin, je n'ai pas ajouté la pA-Tn5 dans les solutions car cette dernière est également fusionnée à un Flag et fausserait donc la visualisation du transgène 3xFlag-Pgm dans les noyaux. Les noyaux sont ensuite prélevés après les marquages dans la solution à 150 mM de sel, après passage dans la solution à 300mM de sel (sans pA-Tn5) et après une incubation d'une heure à 37°C (voir protocole détaillé section 3.2 du Matériel et méthodes).

J'ai ainsi visualisé les protéines Pgm et H3K4me3 dans les MAC en développement après différentes étapes du Cut&Tag (Figure 52, gauche). Pour les trois prélèvements, le marquage de H3K4me3 est bien présent dans les MAC en développement où il forme des foci. La localisation de H3K4me3 dans les nouveaux MAC ne semble donc pas être impactée par les différentes étapes du Cut&Tag. De plus, visualiser la localisation de Pgm avec l'anti-Pgm ou l'anti-Flag, montre que Pgm reste dans les MAC en développement même après des incubations dans un tampon à 300 mM NaCl. Cela suggère que la solution à 300 mM de sel ne sera pas un problème pour visualiser la localisation de Pgm sur la chromatine des noyaux en développement. De plus, l'incubation d'une heure à 37°C ne change pas le signal observé. Les noyaux ne sont donc pas endommagés par cette incubation.

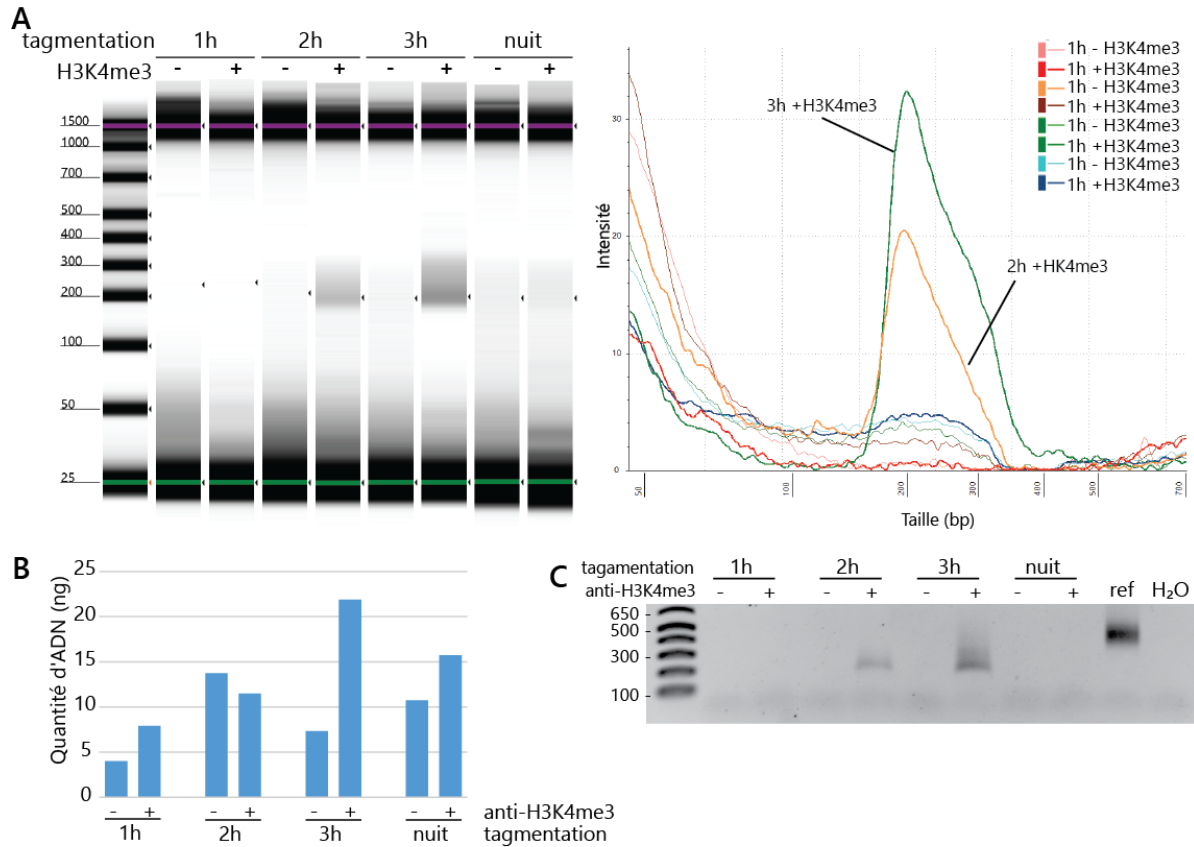


Figure 53 - Contrôle des banques obtenues après un Cut&Tag contre H3K4me3

Tests sur les banques obtenues après un Cut&Tag ciblant H3K4me3. Les conditions sans anti-H3K4me3 (-) correspondent à des contrôles négatifs, témoins de la spécificité de la réaction. Différents temps de tagmentation ont été testés pour faire ces banques, variant d'une heure à toute la nuit. **A** Les résultats d'électrophorèse en capillaire des banques sont représentés sous forme de gel (à gauche) ou de graphique (à droite). **B** La quantité d'ADN obtenue dans chacune des banques est représentée sur un histogramme. **C** Test d'amplification des banques par PCR-semi quantitative et comparaison avec une banque dont l'amplification est bonne (piste « ref ») et un contrôle négatif de la PCR (H₂O).

4.1.2 Localisation de la pA-Tn5 lors du Cut&Tag

J'ai cherché à confirmer l'entrée de la pA-Tn5 dans les MAC en développement. Pour cela, j'ai utilisé des noyaux issus de cellules sauvages non-injectées marqués avec l'anticorps primaire dirigé contre H3K4me3 et un anticorps secondaire sans fluorophore. Après la tagmentation, j'ai réalisé un marquage de la pA-Tn5 en utilisant un anti-Flag (voir protocole détaillé section 3.2 du Matériel et méthodes).

Malgré un marquage en périphérie autour des noyaux, la pA-Tn5 semble bien présente dans les nouveaux MAC (Figure 52, droite). De plus, il semblerait que le marquage forme des foci de la même manière qu'H3K4me3. Ceci suggère que la pA-Tn5 se localise bien au niveau de cette marque d'histone.

4.1.3 Preuve de principe du Cut&Tag : détection de la marque H3K4me3

Le contrôle positif utilisé lors du Cut&Tag est en général un marquage par un anticorps anti-H3K4me3. J'ai donc utilisé cet anticorps afin de vérifier que le Cut&Tag pouvait être réalisé sur la chromatine de paramécie.

Pour cela, j'ai utilisé des noyaux issus de cellules dans lesquelles aucun transgène n'a été injecté et dont j'ai vérifié la qualité par cytométrie en flux (voir Figure 78 en annexe). J'ai ensuite appliqué le protocole développé par l'équipe d'Henikoff (Henikoff et al., 2020; Kaya-Okur et al., 2019) en faisant varier le temps de tagmentation (voir détail du protocole section 3.3 du Matériel et Méthodes).

Une fois la banque de Cut&Tag obtenue, j'ai vérifié la qualité des banques. L'électrophorèse en capillaire montre qu'un signal n'est obtenu qu'après une tagmentation de 2 ou 3h lorsque l'anti-H3K4me3 a été ajouté (Figure 53A). La présence de l'anti-H3K4me3 est donc nécessaire à l'obtention d'un signal ce qui suggère que la réaction est spécifique. Contrairement au protocole initial de Cut&Tag, une heure de tagmentation n'est pas suffisante pour obtenir une banque. De plus, une tagmentation trop longue (sur la nuit) conduit également à une perte du signal. Cette perte pourrait être due à une dégradation de l'ADN par la Tn5. Si on regarde la quantité d'ADN obtenue dans les banques (Figure 53B), on

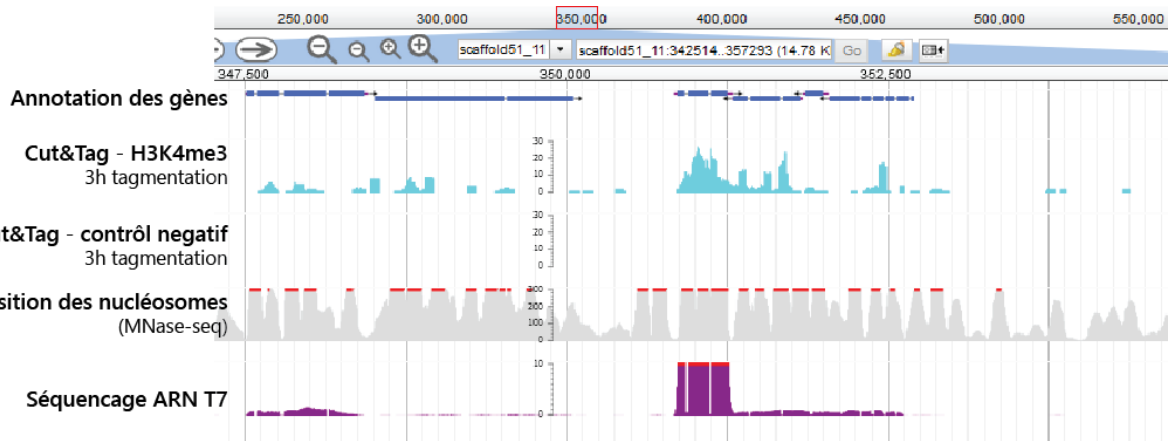


Figure 54 - Vue sur le browser Jbrowse de ParameciumDB

Capture d'écran du Jbrowse de ParameciumDB avec la position des gènes annotés et la position des lectures séquencées pour des banques de Cut&Tag contre H3K4me3 (bleu) et son contrôle négatif (rose) comparé au MNase-seq (gris) (Gnan et al., 2022) et au séquençage d'ARN prélevé 7h après le début de l'autogamie (violet).

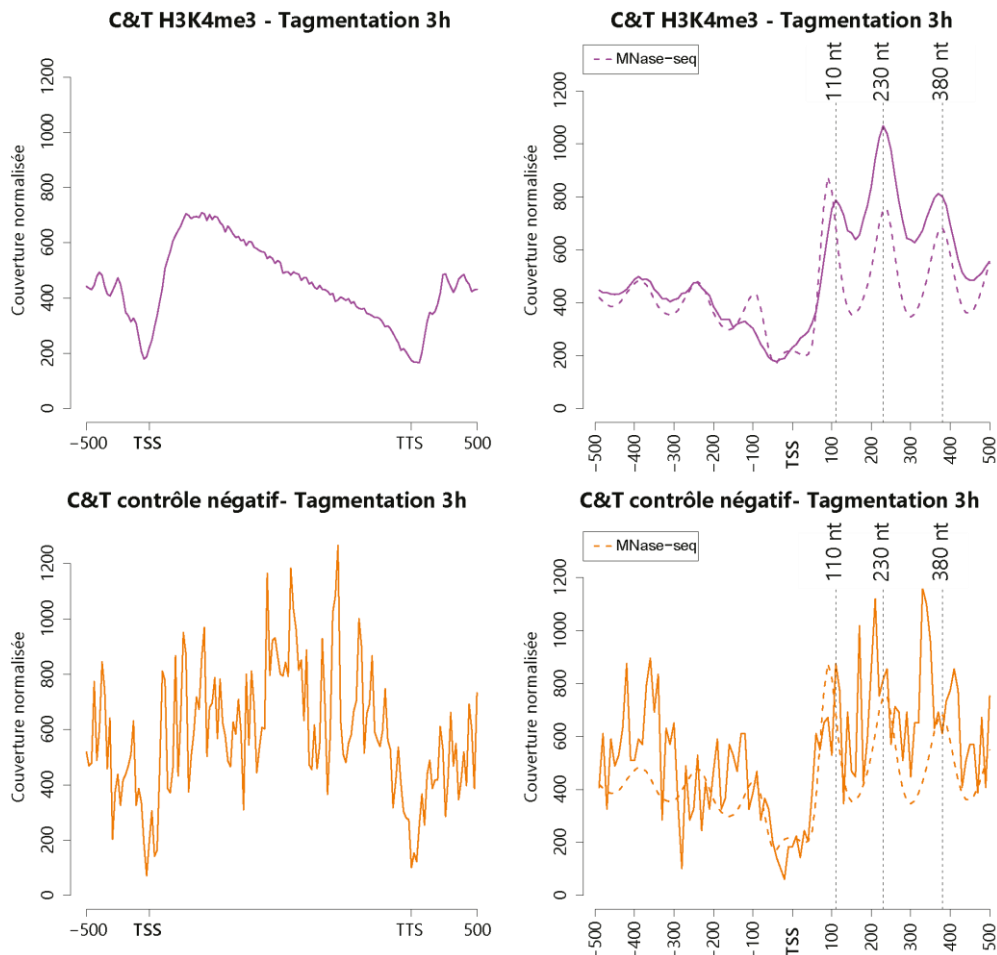


Figure 55 - Le Cut&Tag sur H3K4me3 révèle un marquage spécifique des premiers nucléosomes des gènes

Les gènes ont été alignés entre leur TSS et TTS. La couverture au niveau de tous les gènes a été additionnée pour représenter la couverture globale au niveau des gènes (**gauche**). Zooms centrés sur le TSS (**droite**) comparé à la position des nucléosomes (MNase-seq) (Gnan et al., 2022). (Seule la position des inserts de 50 à 200 pb ont été représentés).

constate que la quantité est généralement inférieure dans les contrôles négatifs suggérant qu'en l'absence d'anticorps primaire, la pA-Tn5 tagmente moins l'ADN. Enfin, l'amplification des banques est testée par PCR semi-quantitative afin de vérifier si elles peuvent être amplifiées correctement lors du séquençage (Figure 53C). Une fois encore, aucun contrôle négatif ne permet d'obtenir de signal et seules les conditions avec 2 ou 3h de tagmentation permettent d'obtenir des banques correctes. On remarque cependant que le niveau d'amplification de ces deux banques est moins bon que celui de la banque de référence. Il est possible que les banques contiennent de l'ADN qui a été fragmenté, mais pas tagmenté, et qui se trouve être moins bien amplifiable par PCR. Une autre possibilité est qu'il resterait des dimères d'oligonucléotides qui n'auraient pas été éliminés lors de la purification de la banque après la PCR.

En conclusion, il faut une tagmentation de 2 à 3h pour obtenir une banque à la suite d'un Cut&Tag dirigé contre la marque H3K4me3 sur la chromatine de paramécie. Les banques pour ces deux temps de tagmentation ont donc été séquencées.

Les séquençages des banques ont montré que moins de lectures sont obtenues pour les contrôles négatifs comparé aux conditions avec l'anticorps primaire (Figure 85 en annexe). Il y a donc peu de bruit de fond. De plus, la taille des inserts varie entre 50 et 200 pb avec un pic autour de 70 pb. Seuls les inserts de 50 à 200 pb ont donc été analysés.

Les lectures correspondant aux inserts ont été cartographiées sur le génome de la paramécie. La couverture obtenue à partir des banques de Cut&Tag ne permet pas de positionner précisément les histones marquées (Figure 54). On remarque cependant qu'il y a peu de signal dans les régions sans gènes. De plus, la position des marques semble coïncider avec la position des nucléosomes précédemment établie (Gnan et al., 2022). Les noyaux utilisés ont été prélevés 6h après le début de leur autogamie. Comparer la présence d'H3K4me3 sur les gènes avec le niveau de transcription au même temps, met en évidence que le signal obtenu est plus fort sur les gènes plus transcrits (Figure 54). Ce qui est cohérent avec le fait qu'H3K4me3 est connue pour être une marque de la transcription active.

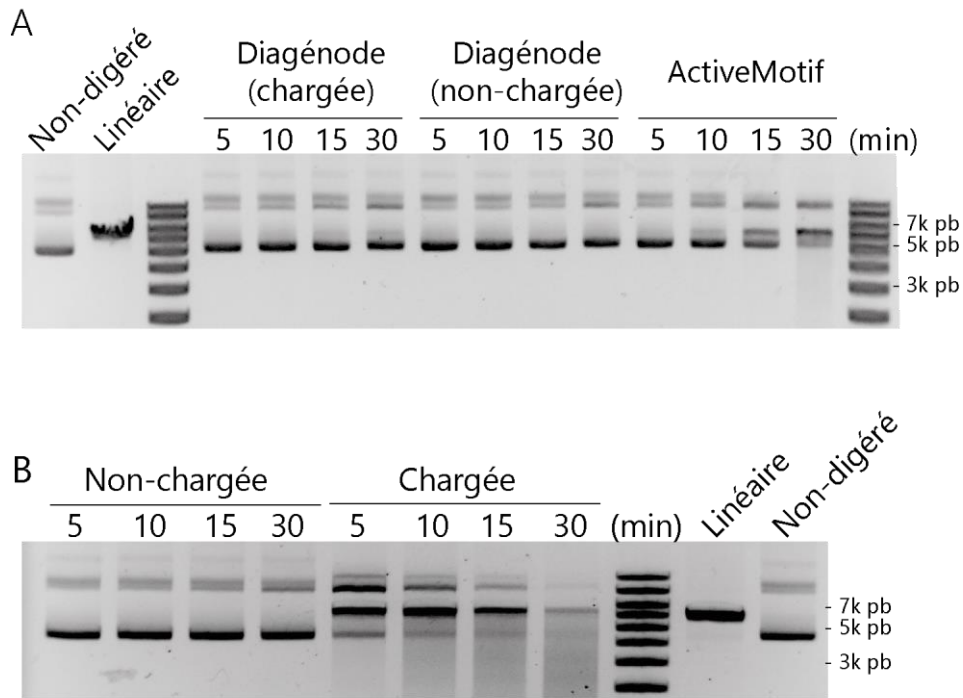


Figure 56 - Test d'activité des pA-Tn5 : coupure d'un plasmide supère enroulé

A Comparaison de l'activité de pA-Tn5 commerciale (Diagénode ou ActiveMotif) sur un plasmide. **B** Test de l'activité de la pA-Tn5 produite dans le laboratoire. Les conditions "non-chargée" correspondent aux pA-Tn5 qui n'ont pas été chargées en ADN et ne sont normalement pas capables de s'intégrer. Les pistes "non-digéré" correspondent au plasmide natif. Les pistes "linéaire" contiennent le plasmide linéarisé par une enzyme de restriction.

Afin d'avoir une vue plus globale de la position de la marque H3K4me3 sur les gènes, la couverture normalisée sur tous les gènes a été compilée. Les gènes ont ensuite été étirés pour que leurs TSS et TTS (*Transcription Termination Site*) se superposent (Figure 55 gauche). La périodicité des nucléosomes n'est pas visible car les gènes ont été étirés. On remarque cependant un enrichissement en marque H3K4me3 sur la région 5' des gènes. Enfin, si on regarde la couverture cumulée autour des TSS (Figure 55 droite), on peut voir la position des trois premiers nucléosomes. La position des 2^e et 3^e nucléosomes correspond à la position des nucléosomes obtenue par MNase-seq (Gnan et al., 2022). Le 1^{er} nucléosome identifié avec la marque H3K4me3 semble cependant légèrement décalé vers la droite, comparé aux données MNase-seq. Cela pourrait être lié au fait qu'H3K4me3 est une marque spécifique de la transcription. Elle pourrait causer un déplacement du 1^{er} nucléosome des gènes afin de faciliter l'accessibilité du TSS à la machinerie de transcription.

Des résultats similaires ont été obtenus par le séquençage des banques pour lesquels la tagmentation a duré 2 heures, mais avec une profondeur de séquençage au niveau des gènes un peu inférieure à celle obtenue en cumulant les lectures obtenues pour la banque du Cut&Tag avec 3h de tagmentation (Figure 86 en annexe). Le Cut&Tag semble donc fonctionner pour la chromatine de paramécie sur une marque d'histone si on allonge un peu le temps de tagmentation. Il reste à déterminer s'il permettra de localiser Pgm sur la chromatine.

4.2 VOIES D'AMÉLIORATION

4.2.1 Production d'une pA-Tn5 plus active

Le Cut&Tag fonctionne pour cartographier les marques H3K4me3 sur la chromatine de la paramécie. Mais, pour obtenir un résultat, il a fallu grandement augmenter le temps de tagmentation. Il est donc possible que l'activité de la pA-Tn5 commerciale ne soit pas optimale pour tagmenter la chromatine de la paramécie. En effet, le Cut&Tag repose sur l'activité de la pA-Tn5 qui permet une intégration locale des adaptateurs de séquençage. La qualité de la pA-Tn5 est donc un point critique.

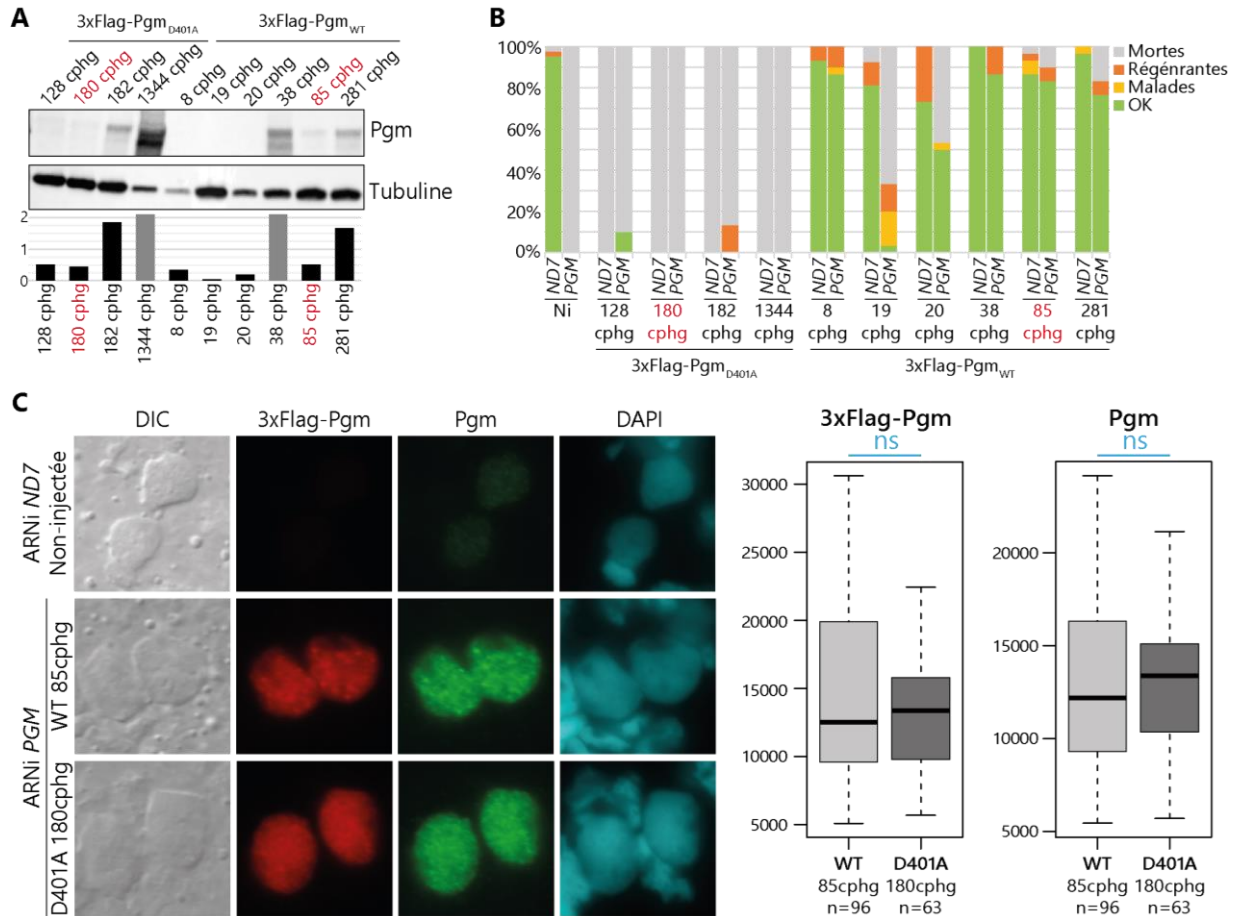


Figure 57 - Mutant catalytique de Pgm

A Estimation par western blot du niveau de protéines produites dans les différentes cellules transformées par le transgène *PGM* sauvage (3xFlag-PGM_{WT}) et le mutant catalytique (3xFlag-PGM_{D401A}) dans des conditions de complémentation (ARNi *PGM*). La quantification du signal normalisé est montrée en dessous. Les barres grises vont au-delà de 2. **B** Test de survie des cellules transformées avec différents niveaux d'injection des transgènes 3xFlag-PGM_{WT} ou 3xFlag-PGM_{D401A} dans des conditions contrôle (ARNi ND7) ou des conditions de complémentation (ARNi *PGM*). En rouge, les transformants utilisés pour l'immunofluorescence. **C** Immunofluorescence sur cellules entières avec marquage par un anti-flag (en rouge) et un anti-Pgm (en vert). La quantification des signaux dans les nouveaux MAC transformés par 3xFlag-PGM_{WT} (WT) ou 3xFlag-PGM_{D401A} (D401A) est présentée par des boxplots. Un test de Mann-Whitney ne révèle aucune différence significative (p -value > 0,1).

Au cours de ma thèse, j'ai utilisé différentes pA-Tn5. J'ai d'abord utilisé une pA-Tn5 produite par l'équipe de Carl Mann (I2BC) pour les tests de localisation et les premiers tests de protocole Cut&Tag. Par la suite, j'ai utilisé des pA-Tn5 commerciales afin de mettre au point le Cut&Tag. Malheureusement, les tests d'activité ont révélé que d'un lot à l'autre l'activité de la pA-Tn5 était très variable obligeant à recalibrer entièrement l'étape de tagmentation (augmenter la quantité de Tn5 et/ou allonger le temps de tagmentation). J'ai donc produit ma propre pA-Tn5 (Voir protocole section 3.1 du Matériel et méthodes) ce qui permet d'obtenir bien plus de pA-Tn5 dans un même lot. De plus, comparée à la pA-Tn5 commerciale, la pA-Tn5 produite dans l'équipe semble beaucoup plus active (Figure 56). Utiliser la pA-Tn5 produite dans le laboratoire permettrait donc de pouvoir retourner à un temps de tagmentation plus court (passer de 3h de tagmentation à 1h comme indiqué dans le protocole initial) et d'améliorer l'efficacité de la tagmentation.

4.2.2 Stabiliser le complexe d'excision pour faciliter le Cut&Tag

Aucune donnée n'est aujourd'hui disponible que ce soit sur la stœchiométrie du complexe ou sur la vitesse de réaction. Étant donné le nombre de cassures double-brin introduites lors des RPG par Pgm, il est possible que la réaction soit assez rapide. Dans le but de faire un Cut&Tag dirigé contre Pgm, il pourrait donc être intéressant d'essayer de bloquer Pgm sur ses cibles. Pour cela, deux solutions sont envisageables. La première consiste à ajouter une étape de fixation légère des noyaux (0,1% de formaldéhyde pendant 2 min). La seconde serait l'utilisation d'un mutant catalytique de Pgm. La mutation empêcherait l'introduction des cassures et pourrait ainsi bloquer Pgm plus longtemps sur ses cibles ce qui faciliterait la visualisation de Pgm sur la chromatine.

J'ai utilisé un transgène porteur d'un simple mutant D401A de Pgm (3xFlag-Pgm_{D401A}) que j'ai injecté dans des cellules végétatives. J'ai estimé par western blot la quantité de protéine produite à partir des transgènes 3xFlag-Pgm_{WT} et 3xFlag-Pgm_{D401A} (Figure 57A). Le niveau d'injection calculé par qPCR (cphg) ne traduit pas fidèlement la quantité de protéine produite au cours de l'autogamie à partir des transgènes. La quantité de protéine estimée par western blot est cohérente avec les résultats obtenus sur le test de survie (Figure 56B). Peu importe la quantité de

transgène injecté, la construction ne montre pas de toxicité en condition d'ARNi contrôle (ARNi contre *ND7*). De plus, il faut un niveau suffisant de protéine 3xFlag-Pgm_{WT} afin d'avoir un bon niveau de complémentation en condition d'ARNi contre *PGM*. Le test de survie montre également que pour un niveau de protéine équivalent 3xFlag-Pgm_{D401A} (marqué en rouge) le mutant D401A ne complémente pas. Ce qui suggère qu'il est bien catalytiquement inactif. De plus, le plus faible niveau de 3xFlag-Pgm_{D401A} suffit à induire une mort des cellules lors du test de survie après un ARNi contrôle (*ND7*). Il est donc fortement probable que le mutant catalytique interagisse avec les complexes d'excision endogènes, empoisonnant la réaction. Enfin, j'ai comparé la localisation des transformants 3xFlag-Pgm_{WT} à 85 cphg et 3xFlag-Pgm_{D401A} à 180 cphg (produisant une quantité similaire de protéine) (Figure 56C). Les deux constructions se localisent bien dans les nouveaux MAC. La quantification du signal dans les noyaux montre qu'il n'y a pas de défaut de localisation du mutant catalytique comparé à la construction sauvage.

La construction 3x-Flag-Pgm_{D401A} pourrait donc être une bonne alternative afin de stabiliser Pgm plus longtemps sur ses cibles et, ainsi, faciliter la visualisation de la localisation de Pgm sur la chromatine par Cut&Tag.

Partie III – Discussion

Chapitre 1 - Dialogue entre les générations de MAC et définition de gènes candidats

1.1 UN LIEN ÉTROIT ENTRE CASSURE ET RÉPARATION DE L'ADN

Le développement des nouveaux MAC nécessite l'élimination précise de ~45 000 d'IES ce qui implique l'introduction de très nombreuses cassures double-brin de l'ADN dans un temps limité. Afin de pouvoir réparer toutes ces cassures, les paramécies expriment les protéines nécessaires à leur réparation (LigaseIV, Xrcc4 et Cernunos) avant celles nécessaires pour leur introduction (Pgm, et les PgmL) (Baudry et al., 2009; Bischerour et al., 2018; Kapusta et al., 2011; Marmignon et al., 2014; données non-publiées Nicolas Méthais, Émeline Ostyn et Laura Lebrun). Le couplage entre cassure et réparation est d'autant plus fort qu'en l'absence de l'hétérodimère Ku70/Ku80c, le complexe d'excision est inactif (Ku80c étant exprimé uniquement pendant le développement des nouveaux MAC) (Marmignon et al., 2014).

Cependant, l'abolition de la dernière étape de réparation de l'ADN par le NHEJ, la ligation, par un ARNi contre *XRCC4*, n'empêche pas totalement l'activité du complexe d'excision. En effet, ~35% des IES parviennent à s'exciser dans ces conditions même si les cassures introduites à leurs sites d'excision sont mal réparées. Ces IES sont principalement des IES précoces ce qui suggère qu'il existe un rétrocontrôle négatif empêchant l'excision des IES plus tardives lorsque l'ADN, au niveau des premières IES excisées, n'est pas correctement réparé.

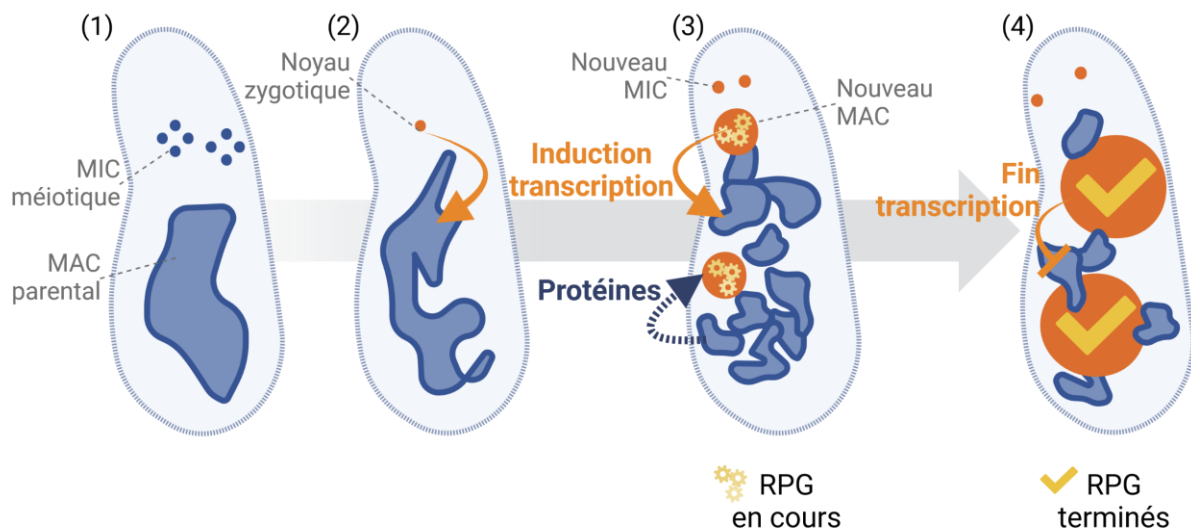


Figure 58 - Modèle de rétrocontrôle de l'expression génique entre l'ancien et le nouveau MAC

Au cours de l'autogamie, il existe un dialogue entre les deux générations de MAC qui co-existent dans la cellule. **(1)** Les MIC parentaux font une méiose pour former des noyaux haploïdes (MIC méiotique). **(2)** Deux de ces noyaux fusionnent pour former un noyau zygotique. Le noyau zygotique ou les jeunes nouveaux MAC envoient un signal induisant le démarrage du programme transcriptionnel dans le MAC parental. **(3)** Le MAC parental se fragmente et produit les transcrits codant pour les protéines indispensables au bon déroulement des RPG dans les nouveaux MAC. **(4)** Une fois les RPG terminés, les nouveaux MAC transmettent un signal vers les anciens MAC pour mettre fin à la transcription des gènes codant pour les protéines nécessaires aux RPG. (Created with Biorender.com)

La fragmentation des chromosomes MAC a normalement lieu après l'excision de la majeure partie des IES. Ce rétrocontrôle pourrait donc simplement consister en un senseur de la fragmentation des chromosomes. Si les cassures aux bornes des IES sont introduites mais pas réparées du fait de l'absence du complexe Xrcc4/Ligase IV, on peut supposer que l'hétérodimère Ku70/Ku80c reste bloqué sur les extrémités de l'ADN ce qui titre la quantité de Ku80c disponible dans la cellule. Puisque Ku80c est nécessaire à l'activité du complexe d'excision (Marmignon et al., 2014), cela conduirait à une diminution du nombre de complexes d'excision actifs et donc à la rétention d'une partie des IES. Il est également possible que l'ARNi contre *XRCC4* ne soit pas complet et qu'il ne reste suffisamment de protéine que pour la réparation des premières cassures introduites aux bornes des IES. Dans tous les cas, le blocage partiel de l'élimination des IES par un ARNi contre *XRCC4* est suffisant pour déclencher la surproduction des protéines nécessaires aux RPG.

1.2 UN DIALOGUE ENTRE L'ANCIEN ET LES NOUVEAUX MAC

Les ciliés ont la particularité de faire coexister dans une même cellule au moins un noyau germinal (MIC) et un noyau somatique (MAC). Au cours de leur développement, trois types de noyaux coexistent puisque le MAC parental est toujours présent lors du développement des nouveaux MAC. L'utilisation de transgènes rapporteurs exprimés uniquement à partir de l'ancien MAC montre que le démarrage du programme transcriptionnel dans le MAC parental dépend de la formation du noyau zygotique. Il pourrait donc y avoir un signal produit par le noyau zygotique, ou les jeunes nouveaux MAC, qui serait transmis à l'ancien MAC afin que le programme transcriptionnel puisse commencer dans celui-ci (Arnaiz et al., 2017) (Figure 58 (2)).

L'injection de transgène dans l'ancien MAC dans des conditions de complémentarité montre que le MAC parental est capable de fournir les protéines nécessaires aux RPG qui ont lieu dans les nouveaux MAC, au moins aux stades précoces (Baudry et al., 2009; Bischerour et al., 2018; Marmignon et al., 2014) (Figure 58 (3)). J'ai montré que lorsque l'excision des IES dans les nouveaux MAC n'est pas complète, il y a un système de compensation qui se met en place induisant une surexpression des protéines nécessaires pour les RPG. L'utilisation

d'un transgène rapporteur, exprimé uniquement à partir de l'ancien MAC, montre que l'ancien MAC participe à l'accumulation de transcrits observée lorsque les RPG ne se déroulent pas correctement. On peut alors supposer que sans signal provenant des nouveaux MAC pour signifier l'achèvement des RPG, les anciens MAC continuent de transcrire les gènes codant pour les protéines nécessaires aux RPG. Cela implique qu'une fois les RPG terminés, les nouveaux MAC produisent un signal pour mettre fin au programme transcriptionnel spécifique des RPG dans les nouveaux MAC (Figure 58 (4)).

L'ensemble de ces données suggère un dialogue entre l'ancien et les nouveaux MAC au cours de l'autogamie. On peut donc proposer un modèle selon lequel les nouveaux MAC réguleraient la transcription dans l'ancien MAC des gènes nécessaires à leurs RPG. Ce dernier fournit ensuite les transcrits nécessaires à la production des protéines requises pour éliminer les IES présentes dans les nouveaux MAC (Figure 58).

1.3 DÉRÉGULATION DES GÈNES LORSQUE L'EXCISION DES IES EST INCOMPLÈTE

Le modèle (Figure 58) suggère que les gènes impliqués dans les réarrangements sont surexprimés lorsque ces derniers sont bloqués (même partiellement). Croiser trois conditions d'ARNi distincts, compromettant le déroulement des RPG, permet d'attribuer la dérégulation des gènes non pas à un ARNi en particulier mais bien à un défaut des RPG. L'identification des gènes surexprimés dans ces conditions m'a permis de définir une liste de 628 gènes. Il faut ajouter *PGM* et *KU80c* à cette liste puisqu'ils ne peuvent pas être surexprimés dans leur ARNi respectif mais qu'ils respectent les autres critères. C'est donc une liste de 630 gènes dont la régulation transcriptionnelle est influencée par le déroulement des RPG que j'obtiens via l'analyse transcriptomique.

1.3.1 De nombreux gènes indispensables à l'excision des IES sont dérégulés

Cette liste de 630 gènes surexprimés inclut la quasi-totalité des gènes codant pour les composants connus du complexe d'excision : *Pgm*, les *PgmL* et

l'hétérodimère Ku70/Ku80c (Tableau 29 en annexe) (Baudry et al., 2009; Bischerour et al., 2018; Marmignon et al., 2014). Les seuls gènes manquants correspondent à certains ohnologues moins exprimés qui codent probablement pour des protéines redondantes (*PGML3b* et *c*) ou des pseudogènes (*KU70b*). Par contre, seule une partie des gènes dédiés à la réparation sont surexprimés : ceux exprimés spécifiquement au moment de l'excision des IES (*intermediate peak*). Cela laisse penser que les gènes de la réparation exprimés plus précocement sont régulés indépendamment des RPG. Ils peuvent toutefois exercer une forte influence sur eux. Par exemple, l'ARNi contre *XRCC4* induit la rétention d'une partie des IES et est suffisant pour induire le phénomène de compensation, alors qu'*XRCC4* ne fait pas partie des gènes surexprimés lors d'un ARNi contre *KU80c*. La liste des 630 gènes contient donc de nouveaux candidats intéressants mais elle n'est pas exclusive.

Parmi les gènes absents de cette liste, figurent notamment des gènes essentiels qui ne sont pas directement impliqués dans la réaction d'excision des IES elles-mêmes. En effet, seule une petite partie des gènes impliqués dans la reconnaissance des IES par des marques épigénétiques (marque d'histone ou ARNnc) est dérégulée (Tableau 29 en annexe) (Betermier and Duhaucourt, 2014; Miró-Pina et al., 2022; Wang et al., 2022b). En revanche, les gènes codant pour des protéines Iswi et Spt16 qui pourraient être impliquées dans le remodelage local de la chromatine pour permettre au complexe d'excision d'accéder aux bornes des IES (de Vanssay et al., 2020; Singh et al., 2022) font bien partie de la liste des 630 gènes ; tout comme l'ensemble des gènes codant pour les condensines D spécifiques du développement des nouveaux MAC. Les gènes dérégulés pourraient donc principalement correspondre à des gènes nécessaires à l'introduction des cassures. Il semblerait logique que le système surexprime principalement les gènes nécessaires à l'introduction des cassures dans le cas où celles-ci ne sont pas correctement introduites.

En plus des gènes, dont la fonction est connue, l'équipe d'Éric Meyer a récemment identifié une catégorie de gènes pouvant être impliquée dans l'excision non pas de toutes les IES mais de certains groupes d'IES (données non-publiées) : les mtFL (*mating type F like*, Tableau 30 en annexe). Au total 25 mtFL réparties dans 18 familles d'ohnologues ont été identifiées. Sur ces 18 familles, 10

possèdent au moins un ohnologue dans la liste de 630 gènes candidats ce qui correspond à la majeure partie des mtFL exprimées spécifiquement au moment de l'excision des IES (*intermediate peak*).

Parmi les 630 gènes, 41 sont déjà connus pour leur implication dans les RPG, ce qui laisse encore 589 gènes à explorer. De manière générale, cette liste est biaisée vers les gènes des pics précoce ou intermédiaire (*early* ou *intermediate peak*), c'est-à-dire en gènes exprimés spécifiquement au cours de l'autogamie. De manière surprenante, les gènes dont l'expression est réduite dans les mêmes conditions, sont principalement enrichis en gènes déjà réprimés à ce stade (*early* et *late repression*). Parmi les gènes réprimés au stade tardif, on trouve également un enrichissement en gènes induits tardivement (*late induction*). Cela suggère que le programme transcriptionnel est bloqué lorsque les RPG n'ont pas lieu. Cependant, seule une petite fraction (< 30%) de chaque groupe de profils d'expression est dérégulée. Le programme transcriptionnel ne serait donc pas globalement régulé par les RPG. Il est possible que seuls les gènes importants pour les RPG ou dépendants de ces derniers soient dérégulés.

1.3.2 Les nouveaux MAC pourraient contribuer à la dérégulation observée au temps tardif

Cette liste de 630 gènes est enrichie en gènes dépourvus d'IES dans leur séquence codante. Cela suggère une préférence pour des gènes qui n'ont pas besoin des RPG pour être exprimés et donc une potentielle implication des nouveaux MAC dans la surproduction des transcrits au temps tardif.

La grande majorité des IES est éliminée au cours des 24h suivant l'autogamie (Zangarelli et al., 2022). Passé ce délai le génome dans les nouveaux MAC est théoriquement fonctionnel et peut alors fournir des transcrits codants pour tous les gènes. Il est donc possible qu'au temps tardif (plus de 40h après le début de l'autogamie) une partie des transcrits provienne des nouveaux MAC. Des contraintes techniques empêchent aujourd'hui de déterminer précisément à quel moment la transcription codante démarre dans les nouveaux MAC. Cependant, une expression depuis les nouveaux MAC est supposée pour certains gènes. Par exemple, compléter un ARNi contre *PTIWI10* en injectant un transgène porteur de la version réarrangée de la séquence codante de *PTIWI10* et de ses

séquences régulatrices s'est révélé inefficace (Furrer et al., 2017). Alors même qu'une construction similaire avec le promoteur d'un autre gène permet la complémentation. Cela laisse supposer que l'expression de *PTIWI10* dépend d'un régulateur présent uniquement dans les nouveaux MAC et ne peut donc avoir lieu qu'à partir de ces derniers. Mais cela reste encore à démontrer. *PTIWI10* fait également partie des trois seuls gènes connus des RPG étant exprimés tardivement (*late induction*) et possédant une IES (avec *PTIWI11* et *mtFL14A*). Ces trois gènes ont également la particularité d'être les seuls parmi les gènes connus des dont l'expression est réduite lorsque toutes les IES ne sont pas excisées (Tableau 29 en annexe). On peut alors supposer que ces trois gènes sont exprimés uniquement depuis les nouveaux MAC une fois que ces derniers ont été réarrangés. Il pourrait en être de même pour les 61 autres gènes porteurs d'une IES dans leur séquence codante, exprimés tardivement (*late induction*) et dont l'expression est réduite dans au moins deux des trois ARNi induisant de la rétention d'IES. De nombreux gènes pourraient donc être exprimés uniquement à partir des nouveaux MAC réarrangés (ou en cours de réarrangement) alors que l'ancien MAC est encore présent et transcrit.

De façon complémentaire, les gènes dont l'expression est réduite lorsque l'excision des IES est incomplète ont plus tendance à être porteurs d'IES dans leur séquence codante. Les IES dans la séquence codante des gènes génèrent souvent un codon stop précoce ce qui conduit à la destruction du transcrit par le NMD (Arnaiz et al., 2012; Yi et al., 2021). Dans ces conditions, il est compliqué de voir une accumulation de transcrit pour ces gènes. Parmi les gènes connus, les seuls qui soient porteurs d'une IES et pour lesquels une accumulation de transcrit est observée lorsque des IES sont retenues sont *PGML5a* et *b*. Cependant, leurs IES étant dans leur 3'UTR, elles sont donc potentiellement invisibles au NMD. Il est possible que les autres gènes surexprimés, possédant une IES dans leur séquence codante, soient invisibles au NMD. La baisse de quantité de transcrits observée peut donc être due à un mécanisme de régulation de la transcription ou à la présence des IES qui induisent une dégradation des ARNm. Cette deuxième hypothèse concorderait avec une contribution des nouveaux MAC dans la dérégulation observée puisque les anciens MAC sont dépourvus d'IES.

Les gènes de la paramécie ont été annotés sur la version réarrangée du génome. De ce fait, tous les gènes identifiés sont des gènes codants uniquement en l'absence de leur IES. Afin de pouvoir identifier des gènes codant dans leur version non-réarrangée, il faudrait réannoter le génome. La recherche de transcrits codants non-réarrangés a permis de mettre en évidence quelques protéines pouvant être exprimées depuis les MAC en développement uniquement avant l'excision de leur IES (données non-publiées Olivier Arnaiz) et dont le rôle n'est pas encore connu. De tels transcrits ont été mis en évidence chez un autre cilié, *Oxytricha* (Miller et al., 2021). Cependant le rôle de ces transcrits reste encore à déterminer.

1.3.3 Une proximité des gènes dérégulés avec le complexe d'excision

La technique de TurboID permet d'identifier l'interactome d'une protéine cible fusionnée à une biotine-ligase (May et al., 2020). Les résultats de TurboID ciblant Pgm ou PgmL4 permettent de donner une idée du contexte dans lequel ces deux protéines évoluent (données non-publiées produites dans l'équipe par Marc Guérineau). Croiser ma liste de 630 gènes surexprimés dans les ARNi contre *PGM*, *KU80c* et *XRCC4*, avec les gènes identifiés par des approches de TurboID dirigé contre Pgm ou PgmL4 ne montre qu'une faible intersection (Figure 59).

Il faut cependant souligner que les analyses de TurboID n'ont pas permis d'identifier toutes les protéines du complexe d'excision (Tableau 29 en annexe). Cela peut suggérer que le complexe est dynamique, auquel cas la TurboID ne permet d'avoir qu'une image à un temps donné de l'état du complexe. Il se pourrait donc que les protéines codées par les nombreux gènes candidats qui ne font pas partie de ceux identifiés par la TurboID ne soient simplement pas localisées à proximité du complexe d'excision au moment de l'introduction des cassures ; mais elles pourraient participer à sa régulation ou à l'établissement d'un environnement favorable à son activité (remodelage de la chromatine, par exemple). De plus, l'intersection entre les données des deux TurboID est également assez faible (Figure 59), il est donc possible qu'il y ait de nombreux faux positifs ou faux négatifs. Parmi les 12 gènes à l'intersection, seulement la moitié est (ou a été) étudiée. De nouveaux candidats intéressants pourraient donc se trouver dans cette nouvelle liste restreinte.

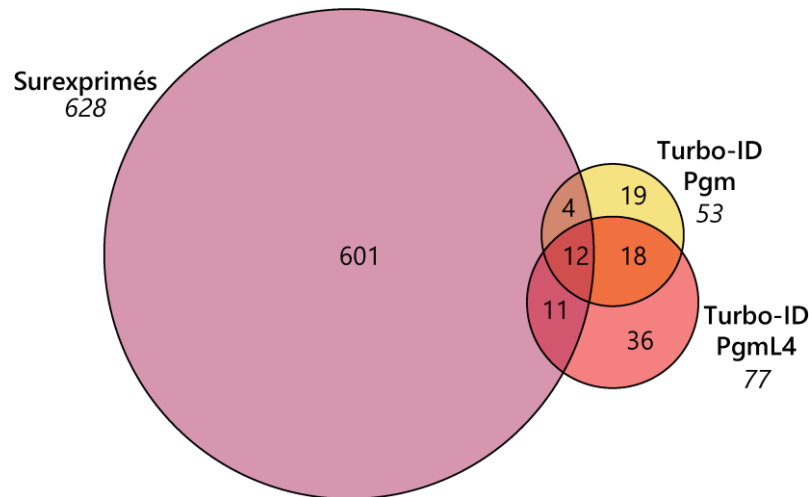


Figure 59 - Intersection des gènes identifiés en TurboID et par analyse transcriptomique

Diagramme de Venn montrant l'intersection entre les gènes identifiés comme surexprimés dans les ARNi contre *PGM*, *KU80c* et *XRCC4* (Surexprimés) et les gènes codant pour les protéines identifiées lors des TurboID ciblant Pgm ou PgmL4 (Communication personnelle, Marc Guérineau).

1.3.4 L'importance du complexe PRC2

Afin de valider la méthode utilisée pour définir les gènes dérégulés, j'ai utilisé la cinétique *EZL1* qui a déjà été publiée (Lhuillier-Akakpo et al., 2014) et j'ai comparé mes résultats avec ceux publiés (Frapporti et al., 2019). Pour ce faire, j'ai ajouté les réplicats techniques reséquencés (Miró-Pina et al., 2022). L'ajout des réplicats techniques diminue légèrement le *fold change* médian, mais augmente le *fold change* maximum observé (Figure 60A). La p-values médiane est également plus basse (Figure 60B). La sensibilité de l'analyse est donc légèrement améliorée. Étant donné que l'ajout des réplicats techniques diminue la p-value, j'obtiens un plus grand nombre de gènes dans mon analyse (Figure 60C). Plus de 85% des gènes identifiés dans les premières analyses publiées sont inclus dans mon analyse. Les ~15% de gènes qui ne sont pas dans mon analyse correspondent probablement à des faux positifs.

Comparer les gènes connus identifiés dans les deux analyses (Tableau 29 en annexe) montre que la plupart des gènes connus pour leur implication dans les RPG sont présents dans les deux analyses (Frapporti et al., 2019). De plus, ~80% des gènes surexprimés lorsque l'excision des IES est incomplète sont également surexprimés lors d'un ARNi contre *EZL1* et se trouvent à l'intersection des deux analyses (Figure 61).

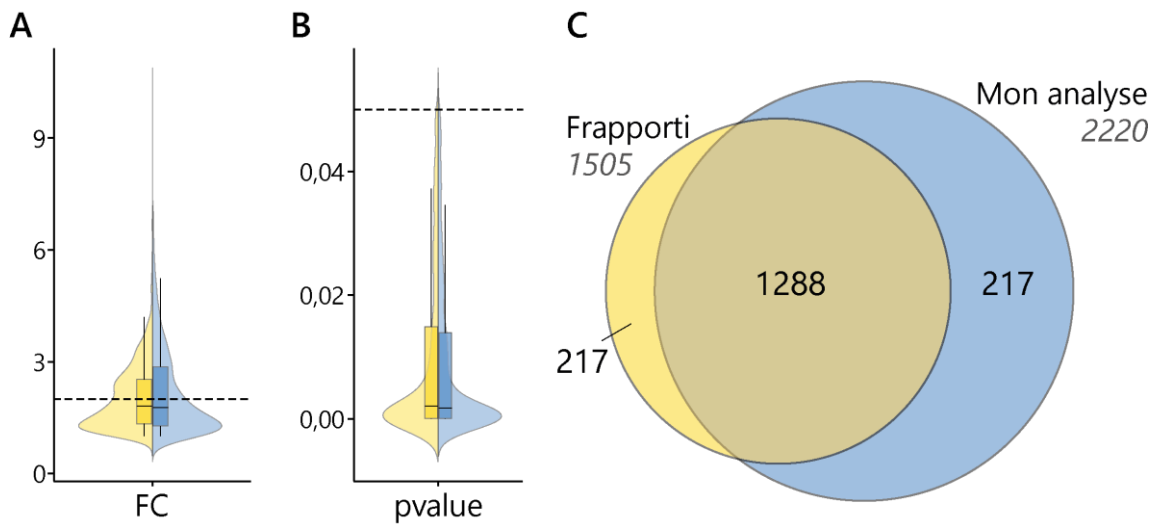


Figure 60 - Comparaison des analyses des gènes surexprimés au temps tardif lors d'un ARNi contre *EZL1*

Pour les trois panneaux, les données issues de l'analyse publiée (Frapporti et al., 2019) sont représentées en jaune et les données de mon analyse en bleu. **A** Distribution des *fold change* (FC). La ligne pointillée représente le seuil de FC = 2. **B** Distribution des p-value. La ligne pointillée représente le seuil de p-value = 0,05. **C** Diagramme de Venn représentant l'intersection des gènes identifiés comme surexprimés au temps tardif lors d'un ARNi contre *EZL1* selon l'analyse publiée (Frapporti) ou mon analyse.

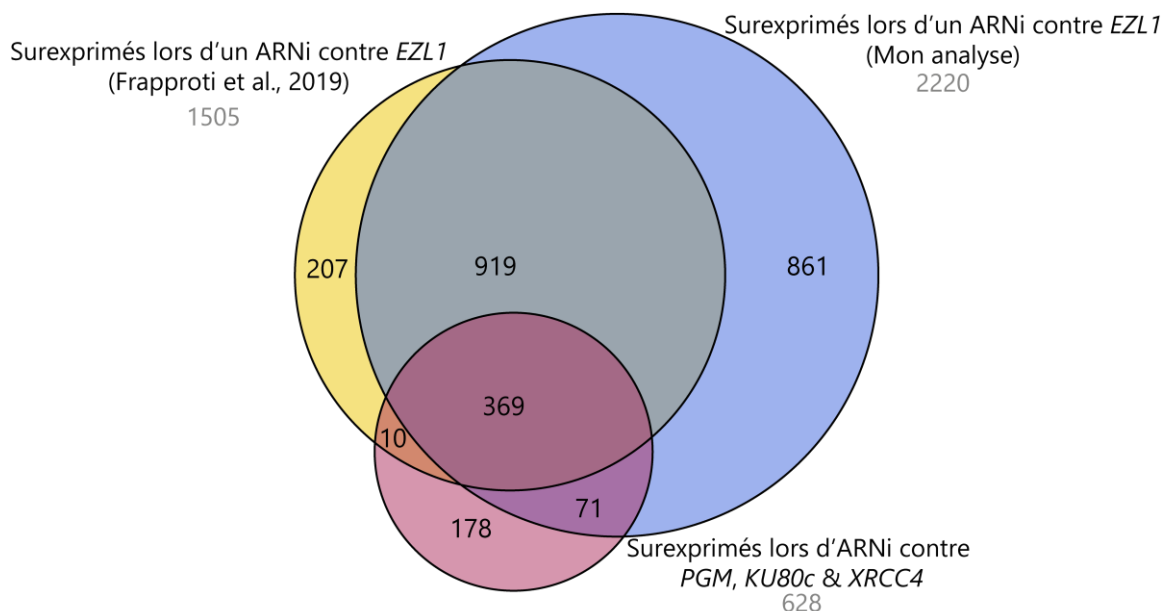


Figure 61 - Gènes surexprimés aux temps tardifs lorsque l'excision des IES est incomplète

Diagramme de Venn montrant les gènes définis comme surexprimés lors d'un ARNi contre *EZL1* dans mon analyse (bleu) et celle publiée (jaune) et lorsque l'excision des IES est incomplète dû à un ARNi contre *PGM, KU80c* et *XRCC4* (prune).

De manière plus globale, ~70% des gènes dérégulés lors d'un ARNi contre *PGM*, *KU80c* et *XRCC4* le sont également lors d'un ARNi contre *EZL1* (Figure 61). Comme lors des ARNi contre *PGM*, *KU80c* et *XRCC4*, l'excision des IES est incomplète lors d'un ARNi contre *EZL1* (Lhuillier-Akakpo et al., 2014). La rétention des IES dans ces conditions induit donc également le système de compensation conduisant à la surexpression de nombreux gènes nécessaires à l'excision des IES. *Ezl1* étant l'unité catalytique du complexe répresseur PRC2 de la paramécie (Frapporti et al., 2019; Lhuillier-Akakpo et al., 2014; Miró-Pina et al., 2022), les surexpressions observées ne sont probablement pas uniquement dues aux problèmes d'excision des IES. Elles reflètent également la transcription de nombreux autres gènes normalement réprimés au temps tardif par le dépôt des marques H3K9me3 et/ou H3K27me3 par le complexe PRC2.

1.4 RÉGULATION DE LA TRANSCRIPTION DANS L'ANCIEN MAC PAR LES NOUVEAUX MAC

Le modèle Figure 58 propose un dialogue entre l'ancien et le nouveau MAC au cours des RPG. Il s'appuie sur la dérégulation de la transcription dans l'ancien MAC lorsque les réarrangements de l'ADN dans les nouveaux MAC ne se déroulent pas correctement. Il s'appuie également sur la nécessité de la présence des nouveaux MAC afin que la transcription des ~2 400 gènes puisse commencer. Il y aurait donc un signal positif provenant des nouveaux MAC (ou du noyau zygotique) induisant l'expression des gènes dans l'ancien MAC.

1.4.1 Les nouveaux MAC sont importants pour le démarrage du programme transcriptionnel dans l'ancien MAC

Contrairement à l'accumulation de transcrits visibles au temps tardif lorsque les IES ne sont pas correctement excisées, les ARNm mesurés dans un ARNi contre *CtIP* ne peuvent pas provenir des nouveaux MAC puisque ces derniers ne se développent pas dans ces conditions. De plus, l'utilisation d'un transgène rapporteur exprimé uniquement à partir du MAC confirme que les nouveaux MAC sont nécessaires à l'induction de la transcription dans les fragments de l'ancien MAC. Enfin, les gènes dérégulés (positivement ou négativement) dans ces

conditions d'ARNi ne présentent pas d'enrichissement particulier en gènes porteurs d'IES dans leur séquence codante ou non. Cela suggère que la présence des IES dans la séquence codante est transparente pour le système. Cela est un argument supplémentaire en faveur de l'idée que la dérégulation transcriptionnelle observée est due à une dérégulation du MAC parental. Ce dernier ne possédant pas d'IES, la présence d'IES ou non dans la séquence codante germinale des gènes n'influence pas sa transcription ou la stabilité des transcrits qu'il produit.

Les gènes les plus impactés par l'absence de nouveaux MAC sont des gènes exprimés lors du développement en conditions contrôles. Les seules classes de gènes du développement qui ne sont pas enrichies parmi les gènes dont l'expression est réduite au stade précoce ou intermédiaire lors d'un ARNi contre *CTIP* sont les groupes de gènes *late peak* et *early repression*. Les premiers, ne sont pas exprimés à ces stades, ce qui explique pourquoi ils ne peuvent pas être identifiés comme étant réprimés. Les seconds sont normalement réprimés lorsque l'autogamie commence, mais ils sont largement enrichis parmi les gènes surexprimés lorsque les nouveaux MAC ne se développent pas. Il semblerait donc que la régulation du programme transcriptionnel dans les anciens MAC dépende de la formation du noyau zygotique ou des ébauches. Il faut toutefois souligner que la dérégulation observée représente une petite portion des gènes de chaque groupe d'expression (< 30%). Tous les gènes d'un même groupe ne sont donc pas soumis aux mêmes régulateurs. On peut supposer que les gènes dont l'induction du programme de transcription dépend de la formation des nouveaux MAC sont impliqués dans leur développement.

1.4.2 Les nouveaux MAC sont nécessaires à l'expression de plus de la moitié des gènes connus impliqués dans les RPG

En l'absence de nouveaux MAC, entre 5 et 6% des gènes présentent une diminution de la quantité de leur ARNm aux stades précoce ou intermédiaire de l'autogamie. Cela peut être dû à une dégradation des transcrits ou une diminution de la transcription. Dans un cas comme dans l'autre, cela résulte en une absence (ou une diminution) de la production des protéines correspondantes. L'expression de 58 des 83 gènes connus dépend de la présence des nouveaux MAC ou du noyau

zygotique (Tableau 29 en annexe). Parmi ces gènes, on retrouve les gènes codants pour le complexe d'excision, les protéines de réparation exprimées spécifiquement pendant l'excision des IES, les condensines D, les gènes impliqués dans le remodelage de la chromatine ou encore ceux impliqués dans la reconnaissance des IES via les ARNies et enfin la quasi-totalité des mtFL identifiées par Éric Meyer. Les seuls gènes connus dont l'expression ne dépend pas du développement des nouveaux MAC sont les gènes nécessaires au mécanisme de *scanning* (Drews et al., 2022b) et les gènes de la réparation dont l'expression est précoce. Ce qui concorde avec l'hypothèse que les gènes de réparation précoces (*early peak*) et intermédiaires (*intermediate peak*) sont régulés de manières différentes. Les premiers ne dépendant ni de la formation des nouveaux MAC, ni du déroulement des RPG contrairement aux seconds.

1.4.3 Identification d'un régulateur commun

De manière plus générale, l'expression de ~40% des 630 gènes, surexprimés aux temps tardifs lorsque l'excision des IES est incomplète, dépend de la formation des nouveaux MAC. Ces 263 gènes sont très largement enrichis en gènes exprimés spécifiquement au moment de l'excision des IES (*intermediate peak*). Ils contiennent probablement une liste restreinte de gènes nécessaires au bon déroulement des RPG. La plupart de ces gènes sont également dérégulés lors d'un ARNi contre *EZL1* (Figure 62). On peut notamment souligner la présence des gènes du complexe d'excision, des condensines D et ceux codants pour les protéines de réparation spécifique du développement des ébauches (Tableau 29 en annexe).

Puisque l'expression de ces gènes est dérégulée de la même façon lorsque les nouveaux MAC ne se forment pas ou lorsque les IES ne sont pas correctement excisées, on peut imaginer que les gènes d'un même groupe d'expression sont contrôlés par un régulateur commun. Le régulateur suivra logiquement les mêmes profils d'expression et les mêmes dérégulations. Pour les gènes exprimés spécifiquement lors de l'excision des IES (*intermediate peak*) ce régulateur pourra se trouver parmi la liste de 180 gènes ayant le même comportement. À cette liste il faut rajouter les gènes *PGM* et *KU80c*. Le régulateur pourrait également être exprimé un peu avant le pic d'expression de ces gènes. On peut donc ajouter à

cette liste les gènes du pic précoce (*early peak*), ce qui porterait le nombre de candidats à 203 (dont 190 sont également dérégulés lors d'un ARNi contre *EZL1*). Parmi ces gènes, 36 sont (ou ont déjà été) étudiés et ne semblent pas être le régulateur.

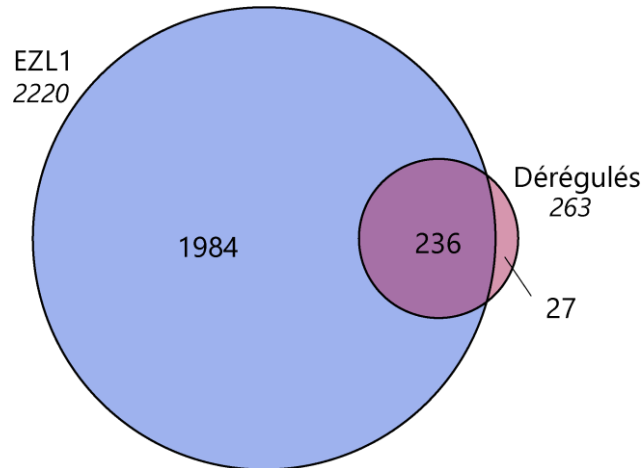


Figure 62 - Gènes dépendant de la formation des nouveaux MAC surexprimés aux temps tardifs lorsque l'excision des IES est incomplète

Diagramme de Venn montrant les gènes définis comme dérégulés dans mon analyse de la cinétique d'ARNi contre *EZL1* (EZL1) et les gènes dont l'expression est réduite lors d'un ARNi contre *CTIP* et qui sont surexprimés lorsque l'excision des IES est incomplète dû à un ARNi contre *PGM*, *KU80c* et *XRCC4* (Dérégulés).

On ne s'attend pas à ce que le régulateur se trouve à proximité du complexe d'excision, il ne semble donc pas avisé de croiser ces données avec les données de TurboID afin de réduire la liste. Une analyse de l'ontologie des gènes semble plus adaptée afin de trouver des gènes qui pourraient coder pour des facteurs de transcription et donc au moins être prédits pour se lier à l'ADN. Cependant, ces annotations sont peu nombreuses chez la paramécie. Beaucoup de gènes ne possèdent aucune prédiction probablement en raison des spécificités des protéines des ciliés qui ne sont pas assez proches des modèles les plus étudiés. Par exemple, sur les 236 gènes surexprimés lors d'ARNi contre *PGM*, *KU80c*, *XRCC4* et *EZL1* et dont l'expression dépend de la formation des nouveaux MAC, seuls 87 possèdent une annotation pour une fonction moléculaire et 33 sont associés à un processus biologique. S'appuyer sur l'ontologie des gènes conduit probablement à manquer de nombreux gènes intéressants. De plus, ces problèmes d'annotation rendent les statistiques d'enrichissement peu fiables. De manière qualitative, on

remarque deux gènes identifiés comme pouvant participer à la régulation de la transcription (PTET.51.1.G0040036 et PTET.51.1.G0120377) auxquels s'ajoutent une dizaine de gènes qui pourraient interagir avec l'ADN.

En plus de ces gènes, il est également possible que le régulateur n'ait pas encore été annoté, auquel cas il ne peut pas se trouver dans ces listes de gènes. En effet, on peut imaginer que le régulateur soit exprimé à partir des jeunes nouveaux MAC. De plus, afin de pouvoir agir comme senseur du déroulement des RPG, il pourrait posséder une IES (probablement tardive) dans sa séquence codante. Le régulateur ne serait alors exprimé qu'une fois les nouveaux MAC formés et seulement tant que l'IES est présente. Il induirait ainsi transitoirement l'expression des gènes nécessaires aux RPG (Figure 63). Si les nouveaux MAC ne se forment pas alors le régulateur n'est pas exprimé et les gènes nécessaires aux RPG ne sont pas induits. Par contre, si les MAC se développent mais que l'excision des IES est incomplète, alors le régulateur reste induit, conduisant à la surexpression des gènes impliqués dans les RPG.

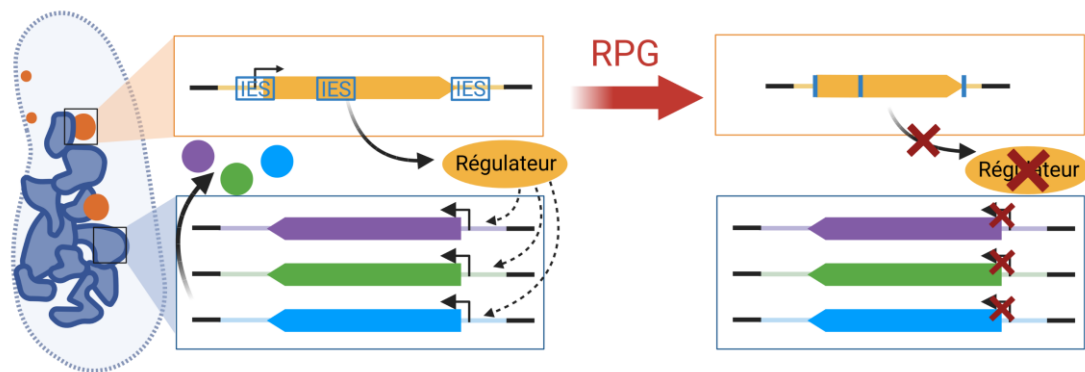


Figure 63 – Modèle possible de régulation de la transcription dans l'ancien MAC par les RPG

Les nouveaux MAC pourraient produire un régulateur (en jaune) qui ne serait exprimé qu'à partir d'une forme non-réarrangée du génome (IES représentées par des carrés bleus). Tant que les IES sont présentes, le régulateur est exprimé et induit l'expression des gènes dans les fragments du MAC parental. L'excision des IES indiquerait la fin des RPG, le régulateur ne pourrait alors plus être produit et la transcription des gènes qu'il contrôle dans l'ancien MAC s'arrêterait. (Created with Biorender.com)

1.4.4 Un motif nécessaire à l'expression des gènes

Si on suppose que l'expression des gènes nécessaires aux RPG est contrôlée par un même régulateur, on peut imaginer que ce dernier reconnaisse une séquence spécifique présente dans leur promoteur. En effet, j'ai pu identifier un

motif enrichi parmi les 180 gènes exprimés spécifiquement pendant l'excision des IES (*intermediate peak*), dont l'expression nécessite la présence des nouveaux MAC et qui sont surexprimés lorsque l'excision des IES est incomplète. Ce qui est en faveur d'un régulateur commun, au moins pour les 98 gènes porteurs du motif (auxquels on ajoute *PGM* qui ne peut pas être surexprimé dans son propre ARNi).

Le motif identifié est non-palindromique mais ne possède pas d'orientation préférentielle. Il est présent dans le promoteur de nombreux autres gènes qui ne suivent pas les mêmes profils de régulation, que ce soit dans les conditions sauvages ou dans les différents ARNi. Si l'orientation du motif ne semble pas être importante sa position semble l'être. En effet, le motif des 99 gènes dérégulés commence entre 50 et 60 pb en amont de leur TSS, contrairement aux autres gènes. Une étude intégrative des motifs identifiés dans les promoteurs d'*Arabidopsis thaliana* a montré que, de manière générale, ce n'est pas l'orientation des motifs (qu'ils soient palindromiques ou non) qui est importante pour la régulation des gènes mais leur position dans le promoteur (Lis and Walther, 2016).

Un enrichissement pour cette position est également trouvé pour les gènes du pic intermédiaire. Ces gènes sont aussi plus enrichis en gènes porteurs du motif dans leur promoteur que les autres groupes d'expression. Le biais de position est donc probablement dû à cet enrichissement. La recherche de motif a été faite au sein d'une population de gènes assez restreinte qui correspond à des gènes pouvant être impliqués dans les RPG, ce qui a pu limiter la découverte de motifs. Cependant le déroulement des RPG fait partie du processus développemental de la cellule au cours de l'autogamie. Il est donc possible que des gènes non impliqués dans les RPG soient corégulés avec ces derniers de manière à maintenir un développement de la cellule cohérent entre les noyaux et la cellule.

Parmi la liste des gènes connus, 42 portent le motif dans leur promoteur et près de la moitié font également partie des 180 gènes dont l'expression dépend de la formation des nouveaux MAC et qui sont surexprimés lorsque les IES ne sont pas correctement excisées (Tableau 31 en annexe). Parmi les 99 gènes corégulés porteurs du motif, 22 sont connus pour leur implication dans les RPG. Cependant près de la moitié des 180 gènes corégulés ne portent pas le motif dont de nombreux gènes connus pour leur implication dans les RPG. Il est possible que

ces gènes ne portent pas le motif car leur expression ne dépend pas du même régulateur. Il est également possible que parmi les gènes porteurs du motif se trouve un gène qui régulerait les gènes qui ne sont pas porteurs du motif. Il y aurait alors une cascade de régulateurs qui pourraient aussi bien augmenter la transcription de ces gènes que la stabilité des transcrits lorsque les RPG ne se déroulent pas correctement. En l'absence du développement des nouveaux MAC, ces régulateurs porteurs du motif ne seront pas exprimés et ne pourraient pas induire la transcription (ou stabiliser les transcrits) des gènes ne possédant pas le motif. Cela induirait une diminution de ces derniers en l'absence de la formation des nouveaux MAC.

1.4.5 Perspectives

Identifier le (ou les) régulateur(s) pourrait être la première étape pour vérifier le modèle proposé dans la section précédente. Dans un premier temps, des ARNi contre les différents régulateurs candidats pourraient être réalisés afin d'observer le déroulement de l'autogamie et l'expression des gènes du complexe d'excision dans ces conditions. Une autre manière d'aborder la question serait de mettre en place des expériences *in vitro* afin d'identifier les facteurs présents dans des extraits nucléaires de paramécie et capables de se lier à des oligonucléotides portant le motif.

D'un autre côté, la liste des gènes corégulés avec Pgm, pourrait être explorée afin d'identifier de nouveaux acteurs des RPG, en faisant des ARNi contre les candidats. La liste pourrait être limitée dans un premier temps aux gènes possédant un nom ou une description intéressante, ou étant également présents dans les autres analyses (ARNi contre *EZL1*, donnée de TurboID).

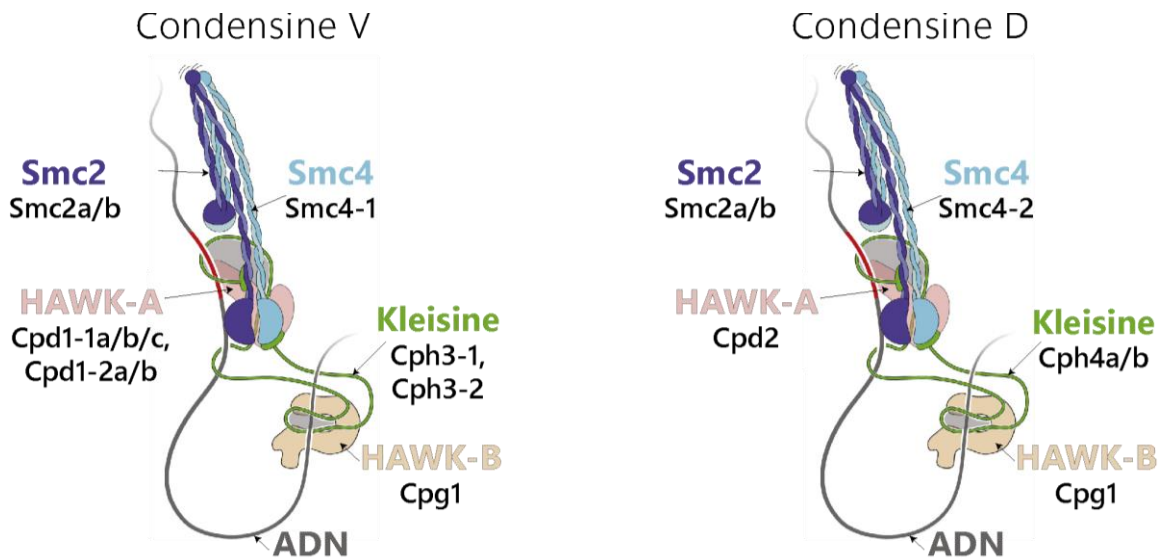


Figure 64 - Deux complexes condensine I de *P. tetraurelia*

La paramécie pourrait posséder au moins deux complexes condensine I : un complexe nécessaire à la vie végétative de la cellule et la méiose des MIC (condensine V) et un complexe requis uniquement pour le développement des nouveaux MAC (condensine D). (Figure adaptée de Oldenkamp and Rowland, 2022)

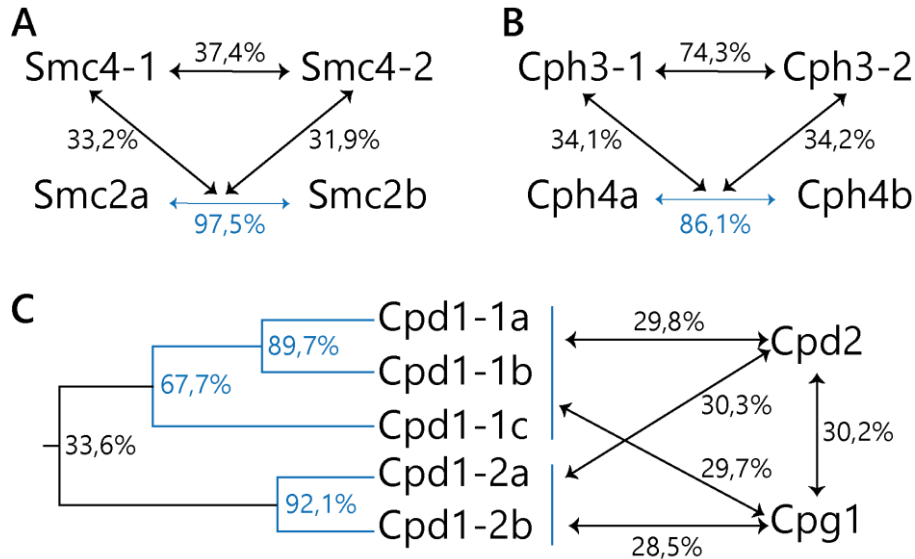


Figure 65 - Comparaison des protéines des complexes condensine

Le pourcentage indique l'identité entre les séquences protéiques des Smc (A), des kleisines (B) et des HAWK (C). En bleu sont représentés les liens entre les ohnologues et en noir ceux entre les paralogues. Lorsque les flèches indiquent un groupe de gènes, le pourcentage moyen d'identité avec le groupe d'ohnologue est indiqué.

Chapitre 2 - Les condensines dans les RPG

Des gènes homologues aux complexes condensine I de *Tetrahymena thermophila* (Howard-Till and Loidl, 2018; Howard-Till et al., 2019) ont été identifiés chez la paramécie (Figure 64). En fonction de leur profil d'expression on peut classer ces gènes en deux groupes : les condensines V exprimées pendant la vie végétative de la cellule et lors de la méiose, et les condensines D exprimées au cours du développement des nouveaux MAC. De plus, les sous-unités spécifiques des condensines D ou V ne présentent qu'entre 30 et 40% de similarité (Figure 65, Smc4-1 est comparé à Smc4-2, les Cph3 aux Cph4 et les Cpd1 aux Cpd2). En plus d'avoir des profils d'expression différents, les gènes des deux groupes codent pour des protéines bien différentes. Chacun des deux groupes possède toutes les sous-unités nécessaires pour former un complexe condensine I (Figure 64). Il est donc possible que la paramécie possède au moins deux complexes condensine partageant les mêmes sous-unités Smc2 et HAWK-B (Cpg1). Étant donné le niveau d'identité des ohnologues de Smc2 (Figure 65A) et leurs profils d'expression il est très probable que ces protéines soient redondantes (Cpg1 ne possède pas d'ohnologue).

2.1 UNE SPÉCIALISATION DES SMC4 CHEZ LA PARAMÉCIE

La plupart des espèces ne possèdent qu'une paire de protéines Smc2/4. Les complexes condensine varient ensuite par les sous-unités kleisine et HAWK qui leur sont associées. Cependant, la paramécie possède deux Smc4 paralogues :

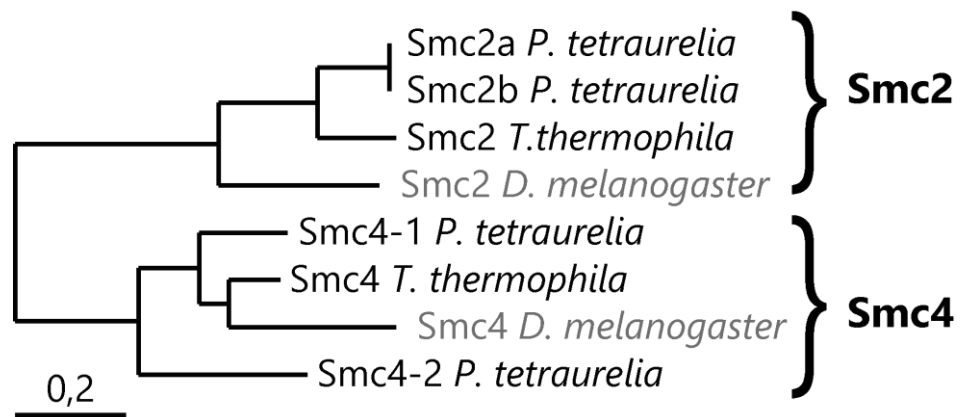


Figure 66 - Conservation des Smc

Arbre phylogénétique des séquences primaires des protéines Smc de *Tetrahymena* (*T. thermophila*), de la paramécie (*P. tetraurelia*) et de la drosophile (*D. melanogaster*). (Généré sur phylogeny.fr)

Smc4-1 et Smc4-2. Smc4-1 est nécessaire à la croissance végétative de la cellule tandis que Smc4-2 n'est requis qu'au cours du développement du noyau somatique. Smc4-1 étant la plus proche de la Smc4 de *Tetrahymena* et de la drosophile (Figure 66), il est possible qu'elle corresponde à la Smc4 ancestrale. Smc4-2 aurait ensuite divergé pour se spécialiser dans le développement des nouveaux MAC et ne participe plus à la vie végétative de la cellule.

2.2 UN COMPLEXE DÉDIÉ AU DÉVELOPPEMENT DES NOUVEAUX MAC : LA CONDENSINE D

Dans le complexe condensine D de la paramécie, seules Smc2 et la kleisine Cph4 sont codées par deux ohnologues ayant des séquences primaires très proches (98% et 86% d'identité respectivement (Figure 65B)) et suivant le même profil d'expression (Figure 44). Il est donc possible qu'ils aient des fonctions redondantes. Il n'y aurait donc qu'un seul complexe condensine D nécessaire au développement des nouveaux MAC.

2.2.1 La condensine D nécessaire au recrutement de la machinerie d'excision

La déstabilisation de Pgm et des différentes PgmL en l'absence de Smc4-2 n'a pas la même intensité selon la sous-unité du complexe d'excision touchée. Les sous-unités PgmL1 et 3 sont les moins touchées (respectivement 1,3 et 1,6 fois moins de signal dans l'ARNi contre *SMC4-2* comparé aux contrôle) ; PgmL2 est un peu plus impactée (1,8 fois moins de signal) ; Pgm et les PgmL4 et 5 sont les plus déstabilisées en l'absence de Smc4-2 (respectivement 3,4, 3,6 et 3,8 fois moins de signal). Cela fait écho à la hiérarchie d'assemblage du complexe d'excision (voir section 3.2.2 de l'introduction). La condensine D semble plus importante pour la localisation stable des dernières Pgm/PgmL assemblées. Il est possible que les PgmL1 et 3 soient naturellement plus stablement localisées dans le noyau que le reste du complexe d'excision. Il est possible que PgmL1 et 3 s'associent d'elles-mêmes à la chromatine. Elles seraient ensuite stabilisées par la condensine D avant de recruter les autres sous-unités du complexe d'excision.

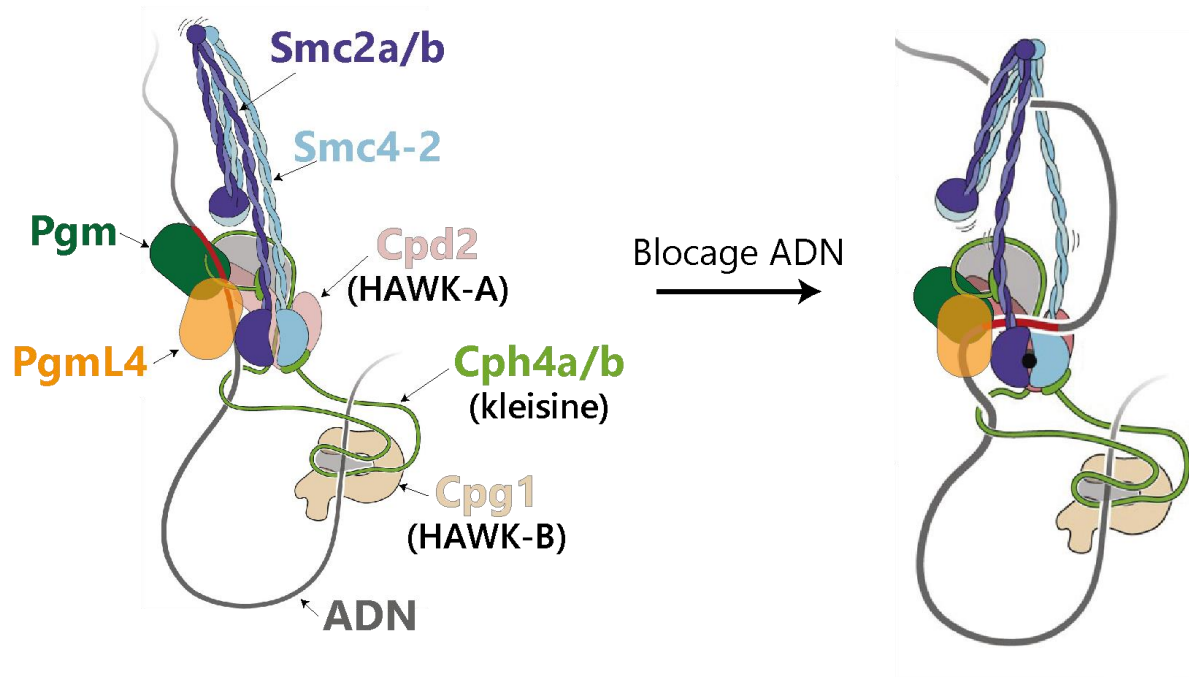


Figure 67 - Modèle d'interaction entre le complexe d'excision et la condensine D

Les données de bio-ID (communication personnelle Marc Guérineau) suggèrent que Pgm interagit avec les Cpd2 et Cph4a tandis que PgmL4 interagirait en plus avec les Smc2. Si on considère également le modèle du complexe condensine proposé (Oldenkamp and Rowland, 2022), il faudrait que Pgm et PgmL4 interagissent avec le complexe condensine D au niveau de la portion mobile de l'ADN (en rouge) qui est transloqué par la condensine. La courbure donnée localement au brin d'ADN pourrait alors faciliter la coupure par Pgm aux bornes des IES.

Par ailleurs, des données d'immunofluorescences montre qu'un ARNi contre *CPD2* donne un phénotype similaire à celui observé lors d'ARNi ciblant *SMC4-2* (données non-publiées Valerio Vitali). Ces protéines pourraient donc bien fonctionner ensemble. De plus, des expériences de Bio-ID révèlent une proximité entre le complexe d'excision et les protéines Cpd2 et Cph4a (Données non-publiées de Marc Guérineau, Tableau 28 en annexe) et entre PgmL4 et les deux ohnologues de Smc2. D'une part, cela renforce l'idée que les condensines D forment bien un complexe, même si cela reste à confirmer par des approches biochimiques. D'autre part, cela suggère une interaction directe entre le complexe d'excision et la condensine D spécifiquement (Figure 67). Chez les procaryotes, une interaction entre une condensine et une nucléase a déjà été observée (Deep et al., 2022; Liu et al., 2022b). La condensine sert alors de senseur de la topologie de l'ADN, permettant à la nucléase de faire la différence entre l'ADN chromosomique bactérien et des plasmides. Étant donné qu'aucune séquence consensus n'est suffisante pour définir les bornes des IES, il est possible que ce soit une structure particulière de l'ADN qui soit reconnue par le complexe d'excision. Auquel cas, la condensine D pourrait servir de senseur en plus de stabiliser le complexe.

Au vu des résultats de TurboID (Tableau 28 en annexe), il semblerait que l'interaction entre la condensine D et la machinerie d'excision se fasse par le biais des sous unités Cpd2 et Cph4 spécifiques à ce complexe (Figure 67). Le complexe condensine D permettrait ainsi la stabilisation du complexe d'excision sur ses cibles. De plus, des études structurales de *PiggyBac* révèlent que l'introduction de la cassure de l'ADN nécessite une importante courbure du site de coupure (Chen et al., 2020). La condensine pourrait donc participer à la courbure de l'ADN, ce qui faciliterait (ou permettrait) l'introduction des cassures double-brin par Pgm (Figure 67). Enfin, la condensine serait capable de former des dimères (Kong et al., 2020). Le modèle actuel propose qu'un complexe d'excision se trouve à chaque borne des IES pour y introduire une cassure double-brin (Arnaiz et al., 2012; Wang et al., 2022a). On peut imaginer que chacun des deux complexes d'excision soit relié à un complexe condensine garantissant une certaine flexibilité à l'ensemble.

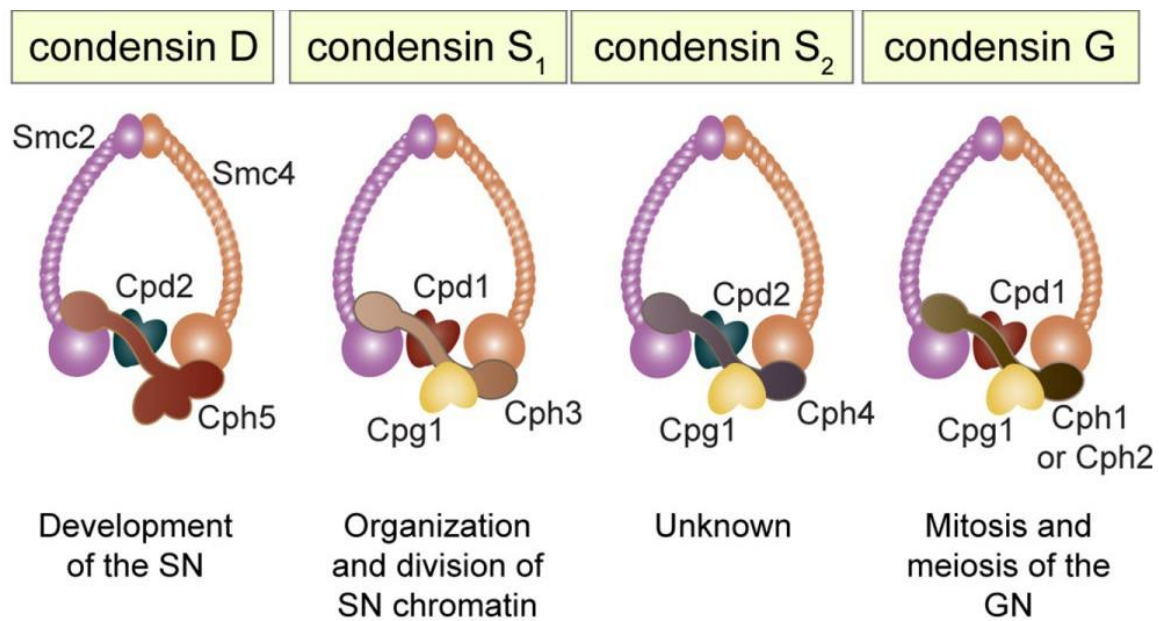


Figure 68 - Les complexes condensine I de *T. thermophila*

Quatre complexes de condensine ont été identifiés chez *T. thermophila*. Tous possèdent les Smc2 et 4 (Structural Maintenance of Chromosomes), une Cpd 1 ou 2 (HAWK-A), une Cph1-5 (kleisine) et éventuellement une Cpg1 (HAWK-B). SN : noyau somatique, GN : noyau germinal. (Figure tirée de Howard-Till et al., 2019)

2.2.2 Un complexe proche des condensines de *Tetrahymena* en composition ou en fonction

Il est intéressant de remarquer que les composants du complexe condensine D de la paramécie correspondent aux homologues du complexe S_2 de *Tetrahymena* (Figure 68) dont la fonction reste inconnue. De plus, aucun homologue des sous-unités spécifiques du complexe D de *Tetrahymena* n'a pu être identifié chez la paramécie. Il semblerait que chez la paramécie, ce soit un complexe homologue au complexe S_2 qui participe au développement du noyau somatique.

La spécialisation des Smc4 chez la paramécie permet de distinguer plus facilement les rôles végétatif ou développemental de ses complexes condensine. Il est possible que chez *Tetrahymena*, le complexe S_2 participe au développement du noyau somatique mais que son effet soit masqué par le complexe D. On peut imaginer que le complexe S_2 soit à l'origine spécifique au développement du noyau somatique. Chez *Tetrahymena* la kleisine Cph5 aurait évolué pour remplir à la fois son rôle de kleisine et celui d'une protéine HAWK-B et participer à l'élimination imprécise d'ADN lors des RPG. Chez la paramécie, c'est la protéine Smc4 du complexe S_2 qui aurait évolué pour se spécialiser dans l'élimination précise des IES au cours des RPG.

Une des principales différences entre les RPG de *Tetrahymena* et de la paramécie réside dans la précision de l'élimination des IES. Il est possible que chez *Tetrahymena*, la condensine S_2 ne participe qu'à l'excision des 12 IES précisément éliminées et que son complexe D soit nécessaire à l'élimination des ~12 000 IES éliminées de façon imprécise. La paramécie ne posséderait que le complexe S_2 puisqu'elle élimine précisément toutes ses IES. Il n'est cependant pas exclu que les condensines D de la paramécie soient également nécessaires à l'élimination imprécise de l'ADN. Le séquençage de l'ADN des nouveaux MAC dans des conditions où l'une des sous-unités du complexe condensine D est manquante permettrait de répondre partiellement à la question. En effet, il n'est toujours pas déterminé si la rétention des IES empêche la poursuite du programme d'élimination de l'ADN et donc les éliminations imprécises d'ADN, ni si l'élimination imprécise dépend des mêmes acteurs que l'excision des IES. Le séquençage des

nouveaux MAC permettra donc simplement de généraliser l'impact des condensines D aux RPG de la paramécie.

2.2.3 Des complexes Smc impliqués dans l'introduction et la réparation des cassures double-brin.

Les complexes Smc ont tous la particularité de posséder un hétérodimère de protéines Smc. Les différents complexes sont définis en fonction de la composition de cet hétérodimère (Oldenkamp and Rowland, 2022; Yatskevich et al., 2019). Les cohésines ont une structure similaire aux condensines et se composent d'un hétérodimère de Smc, d'une kleisine et de deux protéines HAWK. Contrairement aux complexes condensine, l'hétérodimère Smc se compose d'une Smc1 et d'une Smc3 également présent chez la paramécie. Jusqu'à récemment, on pensait que le rôle de la cohésine se limitait à relier les chromatides sœurs lors des divisions nucléaires (par opposition aux condensines qui lient des portions d'un même chromosome pour le condenser). Cependant, des études récentes ont montré que les cohésines sont capables de former des boucles intrachromosomiques. Elles participent ainsi à l'organisation de la chromatine pendant toute la vie de la cellule (Matityahu and Onn, 2021; Oldenkamp and Rowland, 2022). En plus de participer à la formation des TAD, les boucles formées par les cohésines sont essentielles à la recombinaison V(D)J permettant la formation des immunoglobulines (Oldenkamp and Rowland, 2022; Rogers et al., 2021). De nombreux parallèles existent entre le V(D)J et l'excision des IES (Bétermier et al., 2020). Dans le cas du V(D)J, la formation de boucle par la cohésine est nécessaire pour organiser les sites de recombinaison pour qu'ils puissent être reconnus par la transposase domestiquée Rag1 (Liu et al., 2022a; Zhang et al., 2019). Cela appuie l'idée que les complexes Smc puissent organiser l'ADN afin de permettre l'introduction des cassures double-brin par une transposase domestiquée. Chez la paramécie, on pourrait imaginer que ce soit le complexe condensine D qui mette en forme les bornes des IES afin qu'elles puissent être reconnues par le complexe d'excision puis clivées par la transposase domestiquée Pgm.

La découverte du complexe Smc5/6 a mis en évidence un rôle non-structural d'un complexe Smc. En effet, ce complexe a d'abord été identifié pour

son rôle dans la réparation des cassures double-brin (Yatskevich et al., 2019). Il s'avère qu'il peut servir de plateforme de recrutement et initier des processus divers tels que la réparation de l'ADN, mais également, la réplication ou structuration de l'ADN (Matityahu and Onn, 2021; Solé-Soler and Torres-Rosell, 2020). Des études assez récentes ont également montré une implication de ces complexes Smc dans la réparation des cassures double-brin. En effet, ils participent à la formation des foci d'ADN endommagés (Arnould et al., 2021). Étant donné que Smc4-2 diverge des Smc4 classiques (Figure 66), on peut imaginer que certaines des propriétés des cohésines ou du complexe Smc5/6 aient pu être acquises par Smc4-2. Cependant les images d'immunofluorescence montrent que les *foci* formés par Pgm ou certaines PgmL sont toujours présents dans un ARNi Smc4-2. Il est donc peu probable que ce soit le complexe condensine D qui joue un rôle dans la formation des *foci*. Ceci laisse la possibilité d'une intervention des cohésines dans ce processus. Les *foci* pourraient également ne pas refléter une accumulation d'ADN endommagé à un même locus mais une accumulation de protéines inactives. En revanche, étant donné l'importance de Smc4-2 dans la stabilité de Pgm et des PgmL, il n'est pas exclu que la condensine D puisse servir de plateforme de recrutement pour le complexe d'excision, de la même manière que le complexe Smc5/6.

2.3 UN COMPLEXE VÉGÉTATIF : LA CONDENSINE V

2.3.1 Des condensines nécessaires à la division amitotique du MAC

Le complexe condensine V possède une Smc4 spécifique : Smc4-1. Il se distingue également du complexe condensine D par ses kleisines, homologues à Cph3 de *Tetrahymena*, et ses HAWK-A, homologues à Cpd1. Chez *Tetrahymena*, ces sous-unités sont spécifiques du complexe condensine S₁ nécessaire à la division amitotique du MAC lors de sa croissance végétative (Figure 68). Muter les gènes *SMC4*, *SMC2* ou *CPH3* (Cervantes et al., 2006; Howard-Till and Loidl, 2018) de *Tetrahymena* donne un phénotype similaire à un ARNi contre *SMC4-1* chez la paramécie.

Chez *Tetrahymena*, le complexe condensine S₁ est notamment nécessaire à la mise en place du réseau de microtubules intranucléaires lors de la division

amitotique du noyau somatique, même si le lien entre condensine et microtubules reste à déterminer (Cervantes et al., 2006). Lors de la mitose et de la méiose, les condensines I participent à la condensation des centromères et l'attachement des kinétochores pour permettre une ségrégation correcte des chromatides sœurs (Hirano, 2012). Une étude chez le xénope a montré qu'en l'absence des protéines du complexe condensine I, la formation du réseau de microtubules formant le fuseau de division est perturbée, voire inhibée. Cette perturbation n'est pas due au problème de chargement des kinétochores mais semble bien être inhérente à la condensine (Wignall et al., 2003). Bien que cette observation ait été faite lors de la mitose, elle met en avant un rôle de la condensine dans la formation, ou la stabilisation, d'un réseau de microtubules nécessaire à la division nucléaire.

Des études de l'organisation des chromosomes polytènes de la drosophile ont mis en évidence un rôle contre-intuitif de la condensine. En effet il semblerait que la condensine II participe à l'individualisation des chromosomes polytènes au cours de la méiose. De plus, elle permettrait de séparer les chromatides sœurs tout au long de la vie de la cellule afin d'éviter les transvections (Hartl et al., 2008a; Hartl et al., 2008b; Hirano, 2012). Une récente étude chez *Tetrahymena* suggère un rôle spécifique de la condensine I dans la répartition des ~50 copies des chromosomes dans le noyau, permettant une ségrégation relativement équitable lors de la division du noyau somatique (Howard-Till and Loidl, 2018). Les ciliés ne possédant pas de condensine II, on peut supposer que le rôle de la condensine II dans l'organisation de la chromatine est assuré par les condensines I. Ce rôle pourrait être d'autant plus important chez les ciliés qu'ils possèdent un haut niveau de ploïdie dans leur MAC (~50n pour *Tetrahymena* et ~1600n pour la paramécie).

2.3.2 Plusieurs complexes condensine V

La condensine G joue le rôle classique des condensines chez *Tetrahymena* : elle est impliquée dans la compaction de l'ADN lors de la mitose du noyau germinal. Aucun homologue aux sous-unités Cph1 et Cph2, spécifiques de cette condensine, n'a pu être identifié chez la paramécie (Figure 68). Le rôle de Smc4-1 dans la mitose des MIC de la paramécie reste encore à démontrer. Cependant, plusieurs homologues de la kleisine Cph3 et de la protéine HAWK-A, Cpd1, de

Tetrahymena ont pu être identifiés chez la paramécie. Les deux paralogues Cph3-1 et Cph3-2 restent relativement proches avec ~74% d'identité (Figure 65B). En revanche, il y a une importante divergence entre les Cpd1.

Deux groupes de paralogues contenant respectivement deux et trois ohnologues ont été identifiés pour la protéine HAWK A Cpd1 (Cpd1-1 et Cpd1-2). Les deux groupes d'ohnologues présentent à peine plus d'identité entre eux (34%) qu'avec Cpd2 (28-30%), la protéine HAWK-A spécifique des condensines D (Figure 65C). De plus, les profils d'expression sauvage des gènes faisant partie des condensines V permettent de définir deux sous-groupes. D'une part, Cph3-1 et les Cpd1-1a/b exprimés aux stades végétatif et méiotique (sous-unité v/m). D'autre part, Cph3-2, Cpd1-1c et les Cpd1-2a/b exprimés uniquement au stade méiotique. Il est donc possible que la paramécie possède deux complexes condensine V distincts. Les sous-unités v/m participeraient à la mitose des MIC et la division amitotique du MAC lors de la croissance végétative. Les sous-unités méiotiques peuvent venir suppléer les condensines v/m lors de la méiose des MIC. Auquel cas, les deux groupes de sous-unités n'auraient pas de fonction distincte. Il est également possible que les sous-unités méiotiques aient une fonction spécifique de la méiose. Lors de l'autogamie, la méiose des MIC, la formation du noyau zygotique et ses deux divisions mitotiques se déroulent en peu de temps (ils correspondent aux stades MEI et FRG). Il est donc possible que les sous-unités méiotiques servent à la méiose des MIC et que les sous-unités v/m soient ensuite mobilisées pour les deux divisions mitotiques du noyau zygotique. Auquel cas il y aurait des sous-unités spécifiques de la mitose et d'autres de la méiose.

Au vu des profils d'expression, il est peu probable que la paramécie possède des complexes condensine distincts pour les noyaux germinaux et somatiques contrairement à ce qui a été observé pour *Tetrahymena*. Cependant le rôle de Smc4-1 dans les divisions des MIC (par mitose ou méiose) n'est pas encore déterminé. Localiser ces protéines dans la cellule au cours de la croissance végétative et lors de l'autogamie permettrait de s'en assurer.

2.3.3 Perspectives

Afin de s'assurer que tout le complexe condensine D est bien impliqué dans l'autogamie, il faudrait réaliser des ARNi contre chacun des gènes du complexe. Il sera également important de vérifier la localisation des différents composants du complexe au cours de l'autogamie et leur interaction avec le complexe d'excision formé par Ku70/Ku80c, Pgm et les PgmL. Enfin, comparer les complexes des condensines V et D pourrait permettre de mieux identifier les spécificités du complexe importantes pour le bon déroulement des RPG.

Partie IV – Matériel et méthodes

Chapitre 1 - Protocoles de biologie moléculaire

1.1 SOUCHES DE BACTÉRIES UTILISÉES

Différentes souches bactériennes sont utilisées dans le laboratoire :

- ***Klebsiella pneumoniae* (Kp)** : utilisée uniquement pour renourrir les paramécies
- ***Escherichia coli* DH5α** : Cette souche est utilisée pour les constructions de plasmides. Elle est notamment porteuse de la mutation *recA1* empêchant la recombinaison homologue et de la mutation *endA1* empêchant la dégradation des plasmides insérés.
- ***Escherichia coli* SURE** : Cette souche est utilisée pour les constructions de plasmides porteurs de longues séquences de paramécies (riche en AT). En plus de la mutation *endA1*, elle possède des mutations dans les gènes *recB* et *recJ* nécessaires aux réarrangements et délétion d'ADN formant des structures secondaires et tertiaires inhabituelles pour les bactéries mais pas chez les eucaryotes. Ces mutations sont sous pression de sélection grâce l'introduction du gène de résistance à la kanamycine.
- ***Escherichia coli* HT115** (Timmons et al., 2001) : Cette souche est utilisée pour induire l'ARNi chez les paramécies. Elle est optimisée pour la production d'ARNdb (Figure 69). Sous contrôle du promoteur inductible *Plac*, elle possède le gène codant pour ARN-polymérase 7 (ARN-Pol7) permettant la formation d'ARN double-brin à partir de séquence ADN encadrée par des promoteurs T7 en présence d'IPTG (Isopropyl β-D-1-thiogalactopyranoside). De plus, son gène codant pour la RNase III est

interrompu par un gène de résistance à la tétracycline (Tet^R) limitant la dégradation des ARNdb.

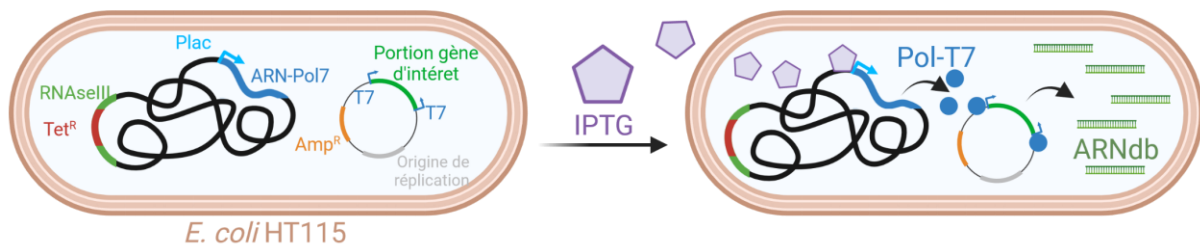


Figure 69 - Production d'ARNdb dans la souche *E. coli* HT115

Les HT115 ont sur leur chromosome le gène codant pour l'ARN polymérase 7 (ARN-Pol7) sous le contrôle du promoteur inducible *Plac* et un gène de résistance à la tétracycline (Tet^R) interrompant la séquence du gène *RNase III*. La production d'ARNdb est induit par l'IPTG dans des HT115 transformées par un plasmide porteur d'une portion du gène à éteindre entre deux promoteurs T7 et du gène de résistance à l'ampicilline (*AmpR*). (Created with Biorender.com)

Dans une culture de LB, les *Kp* et les DH5 α sont donc cultivées sans antibiotique particulier. Les HT115 sont maintenues en présence de 20 $\mu\text{g}/\text{mL}$ de tétracycline et les SURE en présence de 25 $\mu\text{g}/\text{mL}$ de kanamycine. C'est deux antibiotiques ne sont pas ajoutés lors de l'inoculation de WGP par ces souches car si les paramécies tolèrent bien l'ampicilline, la tétracycline et la kanamycine sont toxiques pour elles.

1.2 PRÉPARATION DE BACTÉRIES ÉLECTROCOMPÉTENTES

Préparation : Préparer en amont une dilution de glycérol à 10% dans de l'eau MilliQ qui sera autoclavée. Pour ce protocole il faudra également faire très attention à maintenir les bactéries au froid le plus possible car elles deviennent très fragiles (mettre les différents récipients et tubes au froid en avance et travailler en chambre froide).

Culture : La veille, inoculer 15 mL de LB (+ antibiotiques nécessaires) avec une colonie. Laisser pousser sur la nuit à 37°C avec agitation. Le jour de la préparation, inoculer 1 L de LB (+ antibiotiques nécessaires) avec 10 mL de la culture sur la nuit et placer à 37°C avec agitation jusqu'à ce que la $DO_{600\text{nm}}$ atteigne $\sim 0,5$.

Lavages : Une fois la DO_{600nm} autour de 0,5, mettre la culture 20 min sur glace avec agitation pour que la culture refroidisse vite puis la centrifuger 15 min à 4 000 rpm et 4°C. Jeter le surnageant et resuspendre le culot dans 1L d'eau stérile glacée avant de faire une seconde centrifugation dans les mêmes conditions. Après la centrifugation, jeter le surnageant pour ne garder que 15 à 20 mL d'eau. Resuspendre le culot dans l'eau restante et séparer le volume dans deux Falcons 50 mL. Centrifuger les bactéries à 4 000 rpm pendant 10 min à 4°C et jeter le surnageant. Resuspendre ensuite les culots dans 10 mL de glycérol 10% stérile avant une seconde centrifugation dans les mêmes conditions que précédemment. Jeter à nouveau le surnageant et resuspendre les culots dans 1 à 5 mL de glycérol 10%. Aliquoter par 50 μ L et congeler dans l'azote avant de les transférer dans une boîte sur carboglace stockée à -80°C.

1.3 TRANSFORMATION PAR ÉLECTROPORATION

Préparation : Disposer sur glace autant d'aliquotes de bactéries électrocompétentes et de cuves d'électroporation que de transformations à effectuer, de manière à décongeler sur glace les bactéries et à refroidir les cuves d'électroporation. Diluer 10 ng d'ADN dans 50 μ L d'eau MilliQ ou diluer le mix de ligation pour atteindre les 50 μ L finale. Laisser l'ADN refroidir sur glace.

Electroporation : A la flamme, transférer rapidement les 50 μ L d'ADN dans les 50 μ L de bactéries compétentes. Homogénéiser rapidement en faisant quelques allers-retours au cône avant de mettre le mélange dans la cuve. Il faut procéder rapidement à cette étape afin que les bactéries ne chauffent pas. Appliquer un choc électrique avec l'électroporateur. Resuspendre les cellules dans 450 μ L de LB et transférer le volume dans un Falcon propre. Laver la cuve avec 450 μ L supplémentaire de LB qui seront ajoutés dans le Falcon. Incuber les cellules 1h à 37°C avant de les étaler sur boîtes de LB (+ antibiotiques appropriés). Différentes dilutions peuvent être élevées sur les boîtes.

De manière générale, les plasmides utilisés dans le laboratoire portent le gène de résistance à l'ampicilline. Les transformants sont donc sélectionnés sur des boîtes contenant *a minima* de l'ampicilline.

1.4 PRÉPARATION D'ADN ISSUE DE PLASMIDES

La préparation d'ADN plasmidique à partir de culture bactérienne s'effectue sur des colonnes de silice. Selon la taille de la culture initiale, des MiniPrep ou des MaxiPrep sont respectivement effectuées à l'aide des kits Macherey-Nagel™ Nucleospin™ *plasmid* ou NucleoBon Xtra midi.

1.4.1 MiniPrep

Culture : La veille, effectuer une préculture dans 5 mL de LB (+ antibiotiques appropriés). Laisser les bactéries poussées à 37°C avec agitation sur la nuit.

Récolte : Dans un tube Eppendorf 2 mL centrifuger 2 mL de préculture 2 min à 11 000 g. Jeter le surnageant et centrifuger à nouveau 2 mL de culture si besoin. 1,5 mL de culture peut être préservée pour être conservée à -80°C après l'addition de 300 µL de glycérol 87%.

Extraction et purification : Lyser les cellules en ajoutant 250 µL de tampon A1 sur le culot bactérien. Ajouter ensuite 250 µL de tampon A2 et laisser agir 5 min pour extraire l'ADN puis 300 µL de tampon A3 pour faire précipiter l'ADN génomique (ADNg). Centrifuger les tubes 5 min à 11 000 g pour faire sédimenter l'ADNg et transférer le surnageant sur une colonne de silice. Centrifuger la colonne 1 min à 11 000 g pour fixer l'Adn sur les colonnes. Les colonnes sont ensuite lavées une première fois avec 500 µL d'AW chauffé à 53°C puis 600 µL de tampon A4. La colonne est ensuite séchée par centrifugation en effectuant deux centrifugations consécutives en retournant les tubes entre deux. Enfin la colonne est laissée ouverte 5 min à température ambiante pour laisser s'évaporer l'éthanol qui pourrait rester.

Élution : L'ADN est ensuite élué en ajoutant 50 à 100 µL de tampon AE à 70°C et incubé 5 min à 70°C. Enfin, l'ADN est récupéré dans une Eppendorf propre après une centrifugation d'1 min à 11 000 g. La qualité de la purification de l'ADN et la quantité obtenue seront évaluées au Nanodrop.

1.4.2 MaxiPrep d'ADN

Culture : La veille au matin, effectuer une préculture dans 5 mL de LB (+ antibiotiques appropriés). Laisser les bactéries poussées à 37°C avec agitation sur la journée. Le soir, diluer au centième cette préculture dans le volume désiré de LB (+ antibiotiques appropriés). En général 300 mL de LB sont inoculés avec 3 mL de préculture et laissés sur la nuit à 37°C avec agitation.

Récolte : Répartir la culture dans une ou plusieurs pots de centrifugation et centrifuger 15 min à 4 000 rpm et à 4°C. Puis tout le surnageant est jeté.

Extraction et purification : Lyser les cellules en resuspendant le culot bactérien dans 12 mL de tampon RES. Ajouter ensuite 12 mL de tampon LYS et laisser agir 5 min pour extraire l'ADN puis 12 mL de tampon NEU pour faire précipiter l'ADN génomique (ADNg). Verser ensuite doucement dans le filtre préalablement humidifié avec 25 mL de tampon EQU et laisser élué par gravité. Jeter l'éluât et laver doucement le filtre avec 15 mL de tampon EQU. Une fois le tampon élué, jeter le filtre et laver la colonne de silice avec 25 mL de tampon WASH. Enfin, dans un Falcon50 propre, élué l'ADN par gravité avec 10 mL de tampon ELU chauffé à 50°C.

Re-concentration : Ajouter 10,5 mL d'isopropanol au 50 mL d'éluât et vortexer fortement. Centrifuger au moins 30 min à 4°C et 9 000 g pour faire précipiter l'ADN des plasmides. Rincer ensuite le culot avec 4 mL d'éthanol 70% et centrifuger à nouveau 20 à 30 min. Jeter le surnageant et faire sécher le culot en retournant le Falcon sur du papier absorbant puis en le laissant ouvert 5 min à température ambiante ou à 50°C. Enfin, l'ADN est resuspendu dans 200 µL de TE (Tris 10 mM, 0,5 mM d'EDTA pH8) et transféré dans un tube Eppendorf. La qualité de la purification de l'ADN et la quantité obtenue seront évaluées au Nanodrop.

1.5 POLYMERASE CHAIN REACTION (PCR)

Les PCR sont réalisées avec 0,22 U/µl d'ADN polymérase Phusion™ (ThermoFisher™) dans un mix contenant 10 ng d'ADN matrice, 200 µM de dNTP, 0,5 µM de chacune des amorces (commandée chez Eurofins) et le tampon spécifique de la polymérase. La dénaturation initiale s'effectue à 98°C pendant

30 sec. Elle est suivie d'une seconde dénaturation de 10 sec à 98°C, d'une hybridation de 30 sec à la température d'hybridation des amorces calculées sur le site www.thermofisher.com et d'1 min d'extension à 72°C. Le cycle est répété 25 fois. Les élongations sont terminées à 72°C pendant 5 min.

1.6 PURIFICATION DE FRAGMENTS DE PLASMIDE

Sélection de fragment sur gel : Lors de la construction de nouveaux plasmides, les fragments issus de digestion par des enzymes de restriction sont purifiés sur gel de manière à ne conserver que le fragment de la taille qui nous intéresse. Pour cela le produit de digestion est déposé sur un gel de mois de 2% agarose dans un tampon TAE et migré dans une cuve d'électrophorèse. Le gel est ensuite passé dans un bain de bromure d'éthidium (BET) frais pendant 5 min afin de marquer l'ADN. La bande correspondant au fragment de plasmide d'intérêt est ensuite découpée sous UV et placée dans un tube Eppendorf. L'ADN est ensuite extrait en utilisant le kit Nucleospin™ *Gel and PCR clean-up* de Macherey-Nagel™. Afin de pouvoir en extraire l'ADN, l'agarose est fondue à 50°C avec 200 µL de tampon NT1 pour 100 mg d'agarose.

Purification de produit de PCR : La construction de plasmide implique souvent l'amplification par PCR de fragment d'ADN génomique ou d'autres portions de plasmide. Les produits de PCR sont d'abord purifiés une première fois afin d'éliminer les amorces et de changer le tampon dans lequel se trouve l'ADN. L'ADN est digéré de manière à pouvoir être intégré dans un vecteur et à nouveau purifié après la digestion afin de reconcentrer l'insert et d'éliminer les enzymes de restriction qui auront été préalablement inactivées à la chaleur (si cela est possible). Dans les deux cas, le kit Nucleospin™ *Gel and PCR clean-up* de Macherey-Nagel™ est utilisé et le mix de PCR ou de digestion est dilué dans deux volumes de NT1.

Purification de l'ADN : Le mix dilué ou l'agarose fondu dans le NT1 est ensuite chargé sur une colonne de silice par une centrifugation d'1 min à 11 000 g et l'éluât est jeté. L'opération est répétée jusqu'à ce que tout le volume soit chargé sur la colonne. La colonne est ensuite lavée deux fois avec 700 µL de NT3 puis séchée en effectuant deux centrifugations consécutives d'1 min à 11 000 g en

retournant simplement le tube dans la centrifugeuse entre les deux. La colonne est finie de séchée 5 min à 70°C afin d'être sûr que tout l'éthanol contenu dans les tampons se soit évaporé.

Élution : L'ADN est élué dans un volume choisi de tampon NE chauffé à 70°C. Pour une élution classique, on choisit en général un volume de 20 µL. Mais lorsque l'ADN doit être concentré afin de faire une ligation, deux éluions de la colonne sont effectuées avec 10 puis 15 µL pour les inserts et 20 puis 30 µL pour les vecteurs. L'ADN purifié est élué par une centrifugation d'1 min à 11 000 g. Afin d'améliorer le rendement, lorsque l'ADN est purifié après une extraction sur gel ou lorsque de grands fragments d'ADN sont purifiés, une centrifugation d'1 min à 50 g peut être ajoutée avant la centrifugation à 11 000 g. La qualité de la purification de l'ADN et la quantité obtenue seront évaluées au Nanodrop.

1.7 LIGATION

Après digestion par les enzymes de restriction voulu, les fragments de plasmides peuvent être réassemblés pour former les nouvelles constructions. Pour cela, 150 ng de vecteur digéré sont mélangés avec 3 fois plus de molécules d'insert calculées de la manière suivante :

$$\text{ng d'insert} = 150 \text{ ng} \times \frac{\text{taille de l'insert en bp}}{\text{taille du vecteur en bp}} \times 3$$

Afin d'éviter que le vecteur ne se religue sur lui-même, son extrémité 5' est déphosphorylée en ajoutant de la Quick-CIP (NEB #M0525) au mix de digestion pour la dernière demi-heure à 37°C. Le vecteur est ensuite purifié sur gel (protocole section 1.5).

Pour la ligation les quantités adéquates d'insert et de vecteurs sont mélangées dans un volume final de 8 µL maximum. Au mix est ajouté 1 µL de tampon 10X T4 DNA-ligase (NEB #B0202) et 1 µL de T4 DNA-ligase (NEB #M0202). Le volume est ajusté à 10 µL final. La réaction est ensuite laissée 2 h à température ambiante ou sur la nuit à 18°C avant d'être arrêtée par une incubation de 30 min à 65°C. Enfin, le produit de ligation est électroporé dans des bactéries compétentes (protocole section 1.3).

1.8 PRÉPARATION D'ADN POUR INJECTION

Linéarisation : Afin que l'ADN injecté soit conservé par la paramécie comme un minichromosome, les transgènes sont d'abord linéarisés (voir section 1.3.1 de l'introduction). 75 µg de plasmide portant la construction d'intérêt sont digérés dans 400 µL. Dans le même temps, 7,5 µg de plasmide porteur du gène *ND7* sauvage (pGemND7 (Baudry et al., 2009)) sont également linéarisés par digestion. Tous les plasmides utilisés ici ont été linéarisés par *BsaI*-HF®v2 (NEB #R3733) mais d'autres enzymes de restriction peuvent être utilisés. Après la digestion complète, la réaction est arrêtée en ajoutant 10 mM final d'EDTA et les deux ADN digérés sont mélangés.

Extraction phénol : 100 mM final de NaCl sont ajoutés au mix d'ADN digéré. Afin d'éliminer les enzymes de digestion, 1 volume de PCI (Phénol-Chlor-Isoamyl alcool) est ajouté au mix. Les deux phases sont fortement mélangées au vortex puis centrifugées 2 min à 13 000 g.

Filtration : Afin de pouvoir être injecté, l'ADN produit doit posséder le moins de poussière possible. Pour cela passer la phase aqueuse contenant l'ADN sur un filtre 220 µm (Millipore Ultrafree, Durapore, PVDF 0,22mm) préalablement lavé avec deux fois 500 µL d'eau MilliQ. Le filtre est alors centrifugé 2 min à 12 000 g pour faire passer la solution.

Après cette étape, afin de limiter le plus possible la présence de poussière, il est préférable de travailler avec une boîte de cône et de tubes Eppendorf neufs et de rincer les cônes à l'eau avant leur utilisation. Les solutions trop vieilles ayant pu être ouvertes de nombreuses fois sont également à éviter.

Concentration de l'ADN : L'ADN est ensuite concentré par précipitation éthanol dans un petit volume. Deux volumes éthanol 100% sont mélangés à l'ADN puis centrifugés à 25 000 g à 4°C pendant au moins 30 min. Le surnageant est ensuite jeté et remplacé par 500 µL d'éthanol 70%. L'ADN est ensuite à nouveau centrifugé à 25 000 g et 4°C pendant 10 min. Le surnageant est ensuite éliminé et le culot est rapidement séché sous hotte stérile pour éliminer l'éthanol restant. Enfin, l'ADN est resuspendu dans 10 à 15 mL d'eau MilliQ.

1.9 CONSTRUCTION DES PLASMIDES

1.9.1 p349_pUC_GFP-UTR_PGM

Le plasmide p349 contenant le rapporteur GFP sous contrôle du promoteur (les 96 pb en amont de *PGM*) et du terminateur de *PGM* (les 54 pb en aval de *PGM*) a été construit à partir de deux plasmides existants : pPBL49::5'UTR-GFP-PGM-3'UTR (p216) et pPBL49::5'UTR-PGM-GFP-3'UTR (p219) (construits par Céline Baudry dans l'équipe). Les deux plasmides ont été digérés par les enzymes de restriction *SacI* et *Accl*. Le fragment de 560 pb p216 et le fragment de 3264 pb de p219 sont purifiés sur gel (protocole section 1.5) et ligués ensemble (protocole section 1.7) avant d'être électroporés dans une DH5 α (protocole section 1.3). Le plasmide a ensuite été extrait (protocoles section 1.4) et séquencé d'abord partiellement par Eurofins puis entièrement par Plasmidsaurus (carte Figure 70).

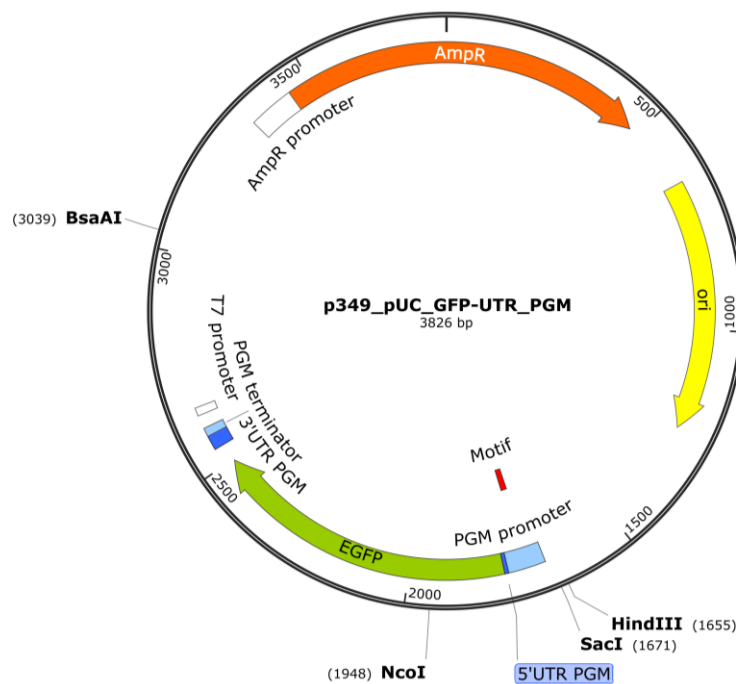


Figure 70 - Carte du plasmide p349

1.9.2 p364_pUC_GFP-UTR_PGM-Motif_AseI & p418_pUC_GFP-UTR_PGM-Motif_INVlg

Le promoteur sauvage en amont de la *GFP* sur le plasmide p349 a été modifié en utilisant des fragments d'ADN synthétique (commandés chez IDT) porteurs de la mutation *AseI* du motif ou motif inversé dans le promoteur de *PGM*.

p349 et les fragments synthétiques ont été digérés par HindIII et NcoI purifiés sur gel (protocole section 1.5) et ligués ensemble (protocole section 1.7) avant d'être électroporés dans une DH5 α (protocole section 1.3). Les plasmides ont ensuite été extraits (protocoles section 1.4). Le plasmide p364 a été d'abord vérifié par digestion puisque la mutation du motif a permis l'insertion d'un site de restriction AseI sans changer la composition du motif. Le promoteur mutant de *PGM* a ensuite été séquencé par Eurofins pour confirmer l'intégrité de la construction. Les plasmides porteurs de la mutation inversé ont, quant à eux, été directement sélectionnés par séquençage. Enfin les deux plasmides ont été entièrement séquencés par Plasmidsaurus.

1.9.3 p361_3xFlag-Pgm & p362_3xFlag-Pgm_{D401A}

Les plasmides p361 et p362 dérivent tous deux du plasmide p113_Pgm(Abs). Ce plasmide porte la séquence de *PGM* dans laquelle ont été introduites des mutations silencieuses afin de le rendre résistant au ARNi (Dubois et al., 2017). Un tag 3xFlag a été ajouté en N-ter de Pgm (atgaattctagagactacaaa gacatgacggtgattataaagatcatgacatcgactacaaggatgacgatgataag) pour former p361_3xFlag-Pgm. Le 3xFlag a également été ajouté en N-terminal d'une version mutante de la séquence de *PGM* pour former p362_3xFlag-Pgm_{D401A}, qui code pour une version de Pgm où l'acide aspartique 401 (1^{er} des 3 acides aminés catalytiques) est remplacé par une alanine.

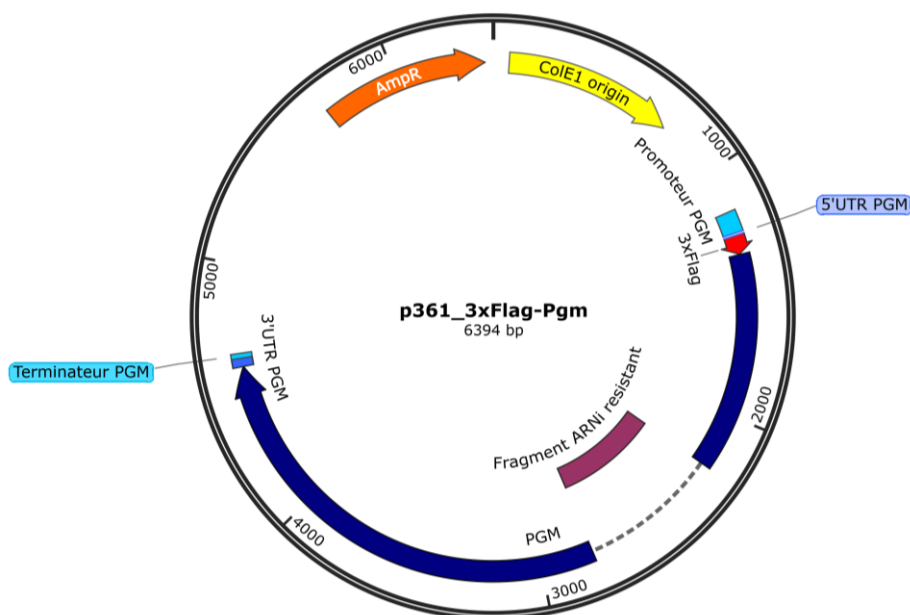


Figure 71 - Carte du plasmide p361

Chapitre 2 - Protocoles de paramécies

Les protocoles de cultures de paramécies ont été établis et affinés au fil des ans. De nombreux détails sur les protocoles et les milieux utilisés lors de culture de paramécies en condition normal (section 2.2) ou lors de l'induction d'ARNi (section 2.4) sont disponibles sur le lien suivant : <https://paramecium.i2bc.paris-saclay.fr/docs/index.php/Protocols>. Les tampons utilisés dans les différents protocoles sont décrits dans la section 2.12.

2.1 SOUCHES DE PARAMÉCIES UTILISÉES

Lors de ma thèse j'ai utilisé deux souches différentes de paramécies : *Paramecium tetraurelia* 51 new (Gratias and Bétermier, 2003) et, un mutant pour le gène *ND7* dérivé de cette souche, 51 *nd7-1* (Dubois et al., 2017). *ND7* est un gène nécessaire pour l'exocytose des trichocystes (Skouri and Cohen, 1997) (Figure 72) mais qui est non-essentiel à la vie de la paramécie et n'intervient pas lors de l'autogamie.

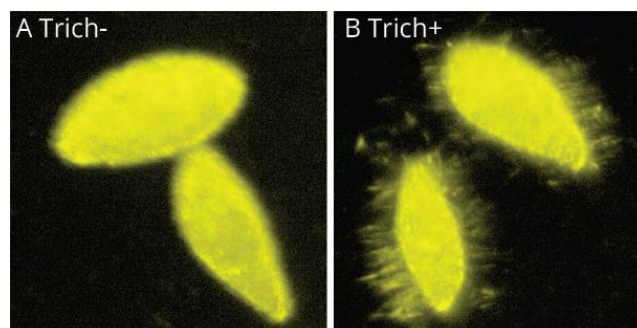


Figure 72 - Test trich

En réaction à l'acide picrique les paramécies relâchent des trichocystes. A La souche 51 *nd7-1* étant mutée pour le gène *ND7*, elle ne relâche pas de trichocystes contrairement à la souche sauvage, 51new (B).

2.2 CONDITIONS DE CULTURE

Dans des conditions optimales, les paramécies sont cultivées à 27°C dans un milieu WGP (*Wheat Grass Powder*) bactérisé avec *Klebsiella pneumoniae* (*Kp*) et complétement par 0,8 µg/mL de β-sitostérol et 100 µg/mL d'ampicilline (Beisson et al., 2010). Sauf précision contraire, de l'ampicilline et du β-sitostérol sont ajoutés dans tous les milieux WGP.

La température de culture des paramécies peut être changée selon la vitesse de croissance souhaitée (Tableau 3). Deux concentrations de WGP sont également utilisées : 0,2X pour les cultures de moins d'1 mL et 1X pour les cultures en plus grand volume.

Température de culture	Nombre de divisions en 24h
27°C	4
20°C	3
18°C	2
14°C	1

Tableau 3 - Nombre de divisions effectuées par les paramécies en fonction de la température de culture.

Jusqu'à 1 mL les paramécies sont cultivées dans des dépressions sur une lame de verre. Au-delà de ce volume, les paramécies sont cultivées dans des erlenmeyers pouvant contenir 10 fois le volume actuel de culture. Par exemple, on utilise un erlenmeyer de 50 mL pour 5 mL de culture de paramécie. De cette manière le volume de culture dans l'erlenmeyer ne dépasse pas les quelques centimètres d'épaisseur permettant une bonne croissance des cellules.

Afin de contrôler la croissance des cellules, la densité en bactérie des milieux est mesurée et ajustée à une DO à 600 nm (DO_{600nm}) 0,1 et les paramécies sont comptées quotidiennement.

2.2.1 Dilution des milieux de culture

La DO_{600nm} du milieu est mesurée afin d'être ajustée à 0,1. Pour ajuster la DO_{600nm} , une partie du milieu est centrifugée 15 min à 4 000 rpm à température ambiante. La portion de milieu à centrifuger est calculée de la manière suivante :

$$V \text{ à centrifuger} = V \text{ de culture} - V \text{ de culture} \times \frac{DO_{600nm} \text{ désirée}}{DO_{600nm} \text{ de la culture}}$$

Si la DO_{600nm} était trop faible, on jette le surnageant et on resuspend le culot bactérien dans la culture restante.

Si la DO_{600nm} était trop élevée, on transfère le surnageant dans le volume de culture restant. Les milieux ne peuvent pas être directement dilués en ajoutant du WPG frais car le WPG dans lequel aucune bactérie n'a poussé est légèrement toxique pour les paramécies. On suppose que les bactéries dégradent ce qui rend le WPG toxique pour les paramécies.

2.2.2 Contrôler la croissance des paramécies

Afin de contrôler la croissance des cellules, toutes les cellules sont comptées lorsque la culture fait moins d'1 mL, au-delà on compte les cellules présentes dans trois aliquotes de 100 μ L de culture et on fait la moyenne des trois pour estimer la densité en paramécie de la culture. De cette manière, on contrôle à la fois le nombre de division effectuée par les cellules et la densité de cellules. Contrôler ces deux paramètres est importants lors de la culture des paramécies. Les paramécies doivent avoir au moins 20 divisions afin de pouvoir entrer en autogamie et au-delà de 35-40 divisions elles peuvent déclencher leur autogamie de manière aléatoire. Il faut donc rester dans cette fenêtre pour déclencher l'autogamie.

La densité en paramécies et en bactéries est également importante afin d'éviter d'induire une carence trop précocement. Lorsque les paramécies sont renourries avec un milieu à une $DO_{600nm} = 0,1$, elles sont en carence alimentaire passer une densité 1 000 cellules/mL. Il faut donc maintenir la densité cellulaire en dessous de ce seuil lors de la montée en volume des cultures de paramécies.

2.2.3 Déclencher et suivre l'autogamie

Lorsque le volume de culture désiré est atteint et que les cellules ont au moins 20 divisions, on peut induire l'autogamie de la culture par une carence de nourriture (Bétermier et al., 2000) (voir section 1.3.2.2 de l'Introduction pour les détails du déroulement de l'autogamie). Afin d'induire la carence, la veille, les cellules sont diluées à 250-300 cellules/mL dans environ 50% de milieu frais. De cette manière les paramécies déclenchent leur autogamie environ 24 h après la dilution (Dubois et al., 2017; Zangarelli et al., 2022). La progression de l'autogamie au sein de la culture est suivie en évaluant la proportion de cellules ayant fragmenté leur MAC (Figure 73). Pour cela les paramécies de 20-30 μ L de culture sont fixées et marquées au DAPI en y ajoutant 4-6 μ L de DAPI-PAF. Après 10-15 min d'incubation, les MAC des cellules sont observés à la loupe à épifluorescence (Abello et al., 2020).

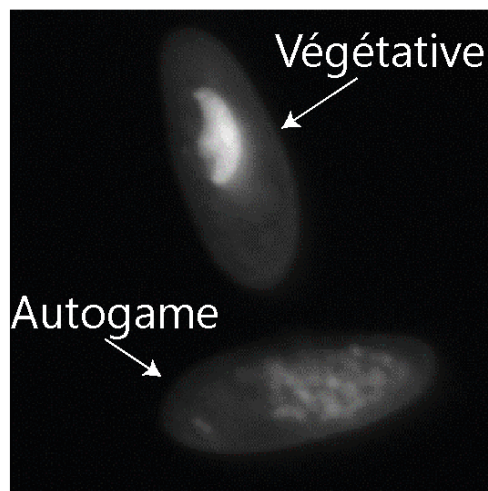


Figure 73 - Paramécies végétative et autogame

Image de paramécies marquées au DAPI et visualisées à la loupe à épifluorescence avec un grossissement x120. Sur l'image on peut observer un exemple de l'état des noyaux observé dans une culture de paramécies en train de passer autogame. La cellule en haut n'a pas son MAC fragmenté, elle est considérée comme végétative. Celle en dessous est considérée comme autogame car son MAC est clairement fragmenté.

2.3 TEST DE SURVIE

Afin de contrôler que l'autogamie s'est correctement déroulée, les paramécies sont gardées quatre jours à 27°C. Le quatrième jour, 30 cellules sont repiquées individuellement dans des dépressions contenant 200-300 μ L de

WGP 0,2X complétement en β -sitostérol et ampicilline et placé à 27°C. La croissance et l'aspect des cellules est ensuite suivie sur quatre jours supplémentaires. Dans ces conditions, les cellules ayant une croissance normale se retrouvent à nouveau en carence alimentaire au bout de 2-3 jours.

Le quatrième jour, les cellules sont fixées et marquées au DAPI afin de regarder l'état de leur MAC. Pour cela, 20 μ L de culture de chaque dépression est prélevé et déposé sur une plaque auquel on ajoute 4 μ L de DAPI-PAF. Après 10-15min d'incubation, l'état des MAC des cellules est observé à la loupe à épifluorescence.

Au bout de quatre jours dans ces conditions, les cellules ont au maximum ~16 divisions, elles sont donc trop jeunes pour relancer une nouvelle autogamie. Si le marquage DAPI révèle que leur MAC est fragmenté, c'est que l'autogamie ne s'est pas correctement déroulée et qu'au lieu de former de nouveaux MAC, la paramécie a régénéré l'ancien MAC.

Ces différentes observations permettent au quatrième jour de définir 4 catégories :

- Mortes : les cellules n'ayant pas survécu jusqu'au terme des quatre jours,
- Malades : des cellules ayant un retard de croissance les empêchant d'être à nouveau en carence de nourrir au bout de quatre jours ou une morphologie anormale,
- Régénérantes : des cellules ayant eu une croissance correcte mais dont le MAC est fragmenté au quatrième jour du test de survie,
- Post-autogames : des cellules ayant une croissance correcte, une morphologie normale et ayant gardé un MAC intègre le quatrième jour.

2.4 INDUCTION D'ARNi PAR FEEDING

Chez la paramécie, il est possible d'induire un ARNi en adaptant le type de bactérie qu'elle mange c'est ce qu'on appelle communément le *feeding* (Galvani and Sperling, 2002).

Afin d'induire l'ARNi, les paramécies sont progressivement changées de milieu. Élevées en Kp, les paramécies sont d'abord passées dans un milieu avec

des HT115 sans plasmide, plus des HT115 avec plasmides non-induits. Enfin les paramécies sont cultivées 2-3 jour en milieu HT115 avec plasmides induits. Chaque ARNi est fait à partir d'un plasmide différent (Résumé des plasmides utilisés dans le Tableau 32 en annexe).

2.4.1 Préparation des milieux

2.4.1.1 Milieu HT115 sans plasmide

Les HT115 sont d'abord étalées sur du LB-agar contenant de la tétracycline (20 µg/mL) et laissées une nuit à 37°C afin d'avoir des colonies pour faire les précultures. La boîte est ensuite gardée quelques semaines à 4°C. Le milieu est ensuite préparé pour le jour J comme indiqué dans le Tableau 4.

J-2	Préparer une préculture de HT115 en LB + tétracycline (20 µg/mL). Incuber à 37°C avec agitation (~150 rpm) sur la nuit.
J-1	Diluer au 1/100 ^{ème} la préculture de LB dans du WPG 1X (sans ampicilline). Incuber à 30°C avec agitation (~150 rpm) sur la nuit.
	Ajuster la DO _{600nm} à 0,1
J	Ajouter le β-sitostérol à 0,8 µg/mL final Diluer les paramécies dans le milieu

Tableau 4 - Production de milieu HT115 sans plasmide

2.4.1.2 Milieu HT115 avec plasmide non-induit

Des HT115 électrocompétentes sont transformées avec les plasmides de *feeding* appropriés puis étalées sur du LB-agar contenant de la tétracycline (20 µg/mL) et de l'ampicilline (100 µg/mL) et laissées une nuit à 37°C afin d'avoir des colonies pour faire les précultures. La boîte est ensuite gardée quelques semaines à 4°C. Le milieu est ensuite préparé pour le jour J comme indiqué dans le Tableau 5. Si l'on souhaite éteindre plus d'un gène plusieurs plasmides peuvent être utilisés, des cultures indépendantes sont préparées à J-2 et J-1. Le jour J, les DO des cultures sont ajustés, puis les cultures sont mélangées à part.

	Préparer une préculture de HT115 transformées en LB + tétracycline
J-2	(20 µg/mL) + ampicilline (100 µg/mL). Incuber à 37°C avec agitation (~150 rpm) sur la nuit.
	Diluer au 1/100 ^{ème} la préculture de LB dans du WPG 1X + ampicilline
J-1	(100 µg/mL). Incuber à 30°C avec agitation (~150 rpm) sur la nuit.
	Ajuster la DO _{600nm} à 0,1
J	Ajouter le β-sitostérol à 0,8 µg/mL final Diluer les paramécies dans le milieu

Tableau 5 - Production de milieu HT115 avec plasmide non-induit

2.4.1.3 Milieu HT115 avec plasmide induit

Les mêmes colonies d'HT115 transformées avec le plasmide de *feeding*, que dans la section précédente, sont utilisées ici pour inoculer la préculture. Le milieu est ensuite préparé pour le jour J comme indiqué dans le Tableau 6. Si l'on souhaite éteindre plus d'un gène plusieurs plasmides peuvent être utilisés, des cultures indépendantes sont préparées à J-2 et J-1. Le jour J, les cultures sont mélangées à part égale après que leurs DO ont été ajustées.

	Préparer une préculture de HT115 transformées en LB + tétracycline
J-2	(20 µg/mL) + ampicilline (100 µg/mL). Incuber à 37°C avec agitation (~150 rpm) sur la nuit.
	Le matin : Culture sur la journée
	Diluer au 1/10 ^{ème} la préculture de LB dans du WPG 1X + ampicilline (100 µg/mL). Incuber à 37°C avec agitation (~150 rpm) sur la journée.
J-1	Le soir : Induction du milieu
	Diluer la culture de la journée dans du WPG 1X + ampicilline (100 µg/mL) pour atteindre une DO _{600nm} de 0,04. Ajouter de l'IPTG à 0,4 mM final et à 30°C avec agitation (~150 rpm) sur la nuit.
	Ajuster la DO _{600nm} à 0,1
J	Ajouter le β-sitostérol à 0,8 µg/mL final Diluer les paramécies dans le milieu

Tableau 6 - Production de milieu HT115 avec plasmide induit

2.4.2 Conditions de culture

Les paramécies sont mises en culture dans un milieu WPG + *Kp* le temps de 4 divisions végétatives. Elles sont ensuite transférées dans du milieu WPG + HT115 sans plasmide pour 4-5 divisions végétatives. La culture est ensuite séparée pour être renourrie avec des milieux WPG + HT115 avec plasmides. Une partie est renourrie avec des HT115 contenant un plasmide de *feeding* contrôle et l'autre partie avec des HT115 contenant des plasmides de *feeding* correspondant aux gènes à éteindre. Après 4-5 divisions végétatives, les paramécies sont enfin renourries avec du milieu WPG + HT115 avec plasmides induits. Elles sont maintenues dans le milieu induit pendant 10 à 15 divisions supplémentaires avant d'induire le déclenchement de l'autogamie par carence alimentaire (voir section 2.2.3).

2.4.3 Plasmides de *feeding*

Les différents plasmides utilisés pour induire des ARNi sont résumés dans le Tableau 7. La séquence des inserts est présentée dans le Tableau 32 en annexe.

n°	Nom	Gène ciblé
p078	p0ND7c (Garnier et al., 2004)	<i>ND7</i>
p079	pPBL49HN monomérique (Baudry et al., 2009)	<i>PGM</i>
p135	pL4440-KU80C-2 (Marmignon et al., 2014)	<i>KU80c</i>
p250	pXRCC4-R (Kapusta et al., 2011)	<i>XRCC4</i>
p076	pL4440-PtCtIPa (Godau et al., 2019)	<i>CtIPa</i>
p077	pL4440-PtCtIPb (Godau et al., 2019)	<i>CtIPb</i>
p424	L4440_SMC4-1	<i>SMC4-1</i>
p421	L4440_SMC4-2	<i>SMC4-2</i>

Tableau 7 - Plasmides de *feeding*

2.5 SUIVIE DE LA CROISSANCE VÉGÉTATIVE

Afin de suivre la croissance végétative des cellules, 30 cellules prélevées dans les tubes stock du cycle sont lavées 4 fois 20 min dans du milieu induit avant d'être placées à 27°C. Les cellules sont ensuite comptées tous les jours et s'il y a des survivants, une cellule est repiquée au hasard dans 200 µL de milieu induit

frais et replacée à 27°C. Pour passer le week-end, elles sont placées à 18°C le vendredi et repiquées le lundi suivant.

Afin de préparer les milieux induits, une préculture de HT115 transformées avec les plasmides d'intérêt est placée à 37°C. Le lendemain le WGP1X + ampicilline est inoculée au 1/200^{ème} en début d'après-midi et placé à 30°C avec agitation. Le soir, lorsque la DO_{600 nm} atteint ~0,04 l'IPTG est ajouté à 0,4 mM final et à 30°C avec agitation (~150 rpm) sur la nuit. Le matin, la DO_{600 nm} est mesurée pour s'assurer que les différentes cultures ont poussé de la même manière. Mais la DO_{600 nm} des milieux n'est pas réajustée à 0,1 avant leur utilisation.

2.6 EXTRACTION D'ADN GÉNOMIQUE

Extraction d'ADN : L'extraction d'ADNg est effectuée en utilisant le kit Nucleospin™ Tissue de Macherey-Nagel™. Jusqu'à 10⁵ de paramécies sont reconcentrées dans 200 µL de milieu, auquel est ajouté 24 µL de protéinase K (20 mg/mL) dilué dans le tampon PB. Les paramécies sont vortexées une première fois, puis 200 µL de tampon B3 est ajouté et les cellules sont incubées 10 min à 70°C pour être lysées.

Purification : 210 µL d'éthanol 100% sont ensuite ajoutés et au mix pour mettre l'accrochage de l'ADN sur la colonne de silice. Le mix est alors déposé sur la colonne par une centrifugation d'1 min à 11 000 g. La colonne est ensuite lavée d'abord avec 500 µL de tampon BW puis avec 600 µL de tampon B5. Elle est ensuite séchée par deux centrifugations consécutives d'1 min à 11 000 g.

Élution : Afin d'éluer l'ADN, 100 µL de tampon d'élution BE chauffé à 70°C sont ajoutés sur la colonne et incubés 1 min à température ambiante. L'ADN est ensuite élué par une dernière centrifugation d'1 min à 11 000 g.

2.7 MICRO-INJECTION DE TRANSGÈNE

2.7.1 Micro-injection dans le MAC végétatif

Préparation de l'aiguille : L'ADN est préparé comme décrit dans la section 1.5 et centrifugé au moins 10 min à 18°C et 25 000 g pour faire précipiter les

poussières. À l'aide d'un cône long, 1-2 μL d'ADN est prélevé en surface et transféré dans une aiguille d'injection Femtotips™ d'Eppendorf. L'aiguille est ensuite fixée sur un micro-injecteur FemtoJet® 4 d'Eppendorf et l'ADN est poussé au bout de l'aiguille.

Préparation des paramécies : Les paramécies sont préalablement transférées à 18°C. Elles sont ensuite lavées pendant 5 à 10 min dans de l'eau Volvic avec 0,5% de BSA toujours à 18°C. Des gouttes, contenant une unique paramécie avec un minimum de liquide, sont ensuite disposées sur une lamelle puis recouvertes de paraffine.

Injection : Sous un microscope, le milieu est retiré à l'aide de l'aiguille d'aspiration jusqu'à ce que la paramécie soit immobilisée sous la paraffine (Figure 74). L'ADN est ensuite injecté dans le MAC à l'aide de l'aiguille d'injection. Le milieu est ensuite relargué avec l'aiguille d'aspiration pour permettre à la paramécie de nager à nouveau. Une fois toutes les cellules d'une lame injectées, les cellules survivantes sont repiquées dans 200 μL de WGP 0,2X + *Kp* et placées à 27°C.

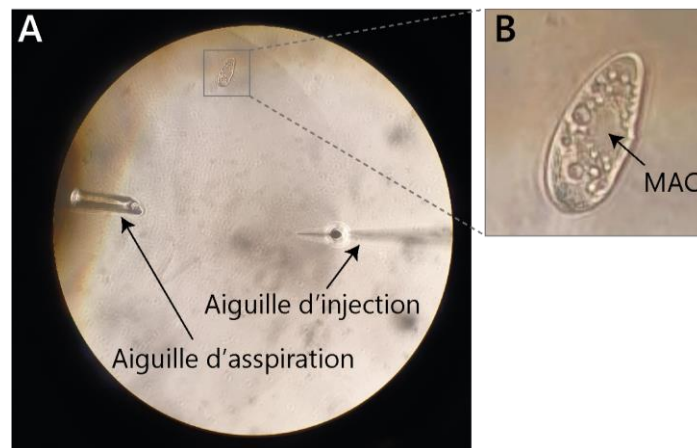


Figure 74 - Microinjections d'ADN

A Aperçu du champ visible au microscope avec à gauche l'aiguille d'aspiration et à droite l'aiguille Femtotips™ servant à l'injection d'ADN. Encadrée en haut, une paramécie bloquée sous la paraffine. **B** Zoom sur la paramécie où l'on peut visualiser la MAC où l'ADN doit être injectée ainsi que des vacuoles.

2.7.2 Sélection des transformants

2.7.2.1 Test trich

Afin d'avoir un contrôle rapide de la transformation des cellules. Les injections sont effectuées dans la souche 51 *nd7-1* mutée pour le gène *ND7* et un

transgène porteur du gène *ND7* sauvage est co-injecté avec la construction. Ainsi seules les paramécies injectées seront capables de relarguer leur trichocyste en réponse à l'addition d'acide picrique (Figure 72).

Le lendemain de l'injection, quelques cellules sont prélevées et placées sur une plaque. On ajoute ensuite une goutte d'acide picrique saturée et on observe le relargage des trichocystes. Seules les cellules Trich+ sont conservées. Les autres sont écartées, elles n'ont probablement pas été correctement injectées.

2.7.2.2 Calcul du nombre de copie par génome haploïde (cphg)

Extraction d'ADN : Pour chaque transformants Trich+, ainsi que des contrôles non-injectés (Ni), une cellule est repiquée dans une dépression contenant 200 μ L de WPG 1X bactérisée avec du *Kp* et placée à 27°C. 24 h plus tard, les cellules ont fait ~4 divisions, 800 μ L de WPG 1X bactérisée avec du *Kp* frais est alors ajouté pour renourrir les cellules. Lorsque la dépression est pleine (32 à 24 h après), l'ADNg est extrait à partir du millilitre de cellules (protocole section 2.5).

Mesure de la quantité de transgène : Afin de mesurer la quantité de transgènes injectée, on effectue une PCR quantitative en temps réelle (RT-qPCR : *Real Time quantitative PCR*). Pour évaluer la quantité de transgènes injectés, il faut mesurer la quantité de transgènes amplifiée et la comparer à l'amplification d'un locus endogène. Pour chaque RT-qPCR deux mix sont donc effectués : l'un avec les amorces permettant d'amplifier un locus endogène (ici *KU80c*) et l'autre avec des amorces permettant d'amplifier le transgène (Tableau 8).

Chaque réaction contiendra 0,5 μ L d'amorce à 10 μ M, 2,75 μ L d'eau MilliQ, 5 μ L de mix 2X Luna® pour qPCR (NEB #M3003) et 1,25 d'ADNg. Chacune des cultures Trich+ et des contrôles Ni sont testés en triplicat. Pour chacune des deux paires d'amorces auxquelles s'ajoutent des triplicats sans ADN comme contrôle négatif. Un mix sans l'ADN est donc effectué pour chacune des paires d'amorces utilisées en tenant compte du nombre de réaction à effectuer. Par exemple pour tester 6 cultures il faut prévoir suffisamment de mix pour 2x21 réactions. Les mix avec l'ADN sont déposés sur une plaque 384 puits. Une fois la plaque complétée elle est sellée à l'aide d'un film plastique et centrifugée 1 min à 1 000 rpm. Enfin, la plaque est placée dans un Light Cycler 480 (Roche) qui permettra de calculer le

Cp (*Cycle point* : le nombre de cycles nécessaires pour que le signal sorte du bruit de fond).

N°	Gène ciblé	Séquence
OMB433	<i>PGM</i>	GCAGTGGGAACTATTAGACATAACAGAG
OMB434		TGGAAATACATGAGCATAAGTTCATTTG
OMB1176	<i>KU80c</i>	AGCTCCTGATGGCAAGACAT
OMB1177		AGTCTGTCCACAACCAGCTG
OMB1255	<i>GFP</i>	GCCAACACTTGTCACTACTTTAA
OMB1256		ACTTCAGCTCTTGTCTTGTAGT

Tableau 8 - Amorces utilisées pour les RT-qPCR

Calcul du cphg lorsque qu'un locus endogène est présent (exemple transgène porteur du gène *PGM*) :

$$\text{cphg} = 2^{(\text{Cp}_{\text{genomic ref}} - \text{Cp}_{\text{PGM}})} - \text{nb de locus endogène}$$

Calcul du cphg en l'absence de locus endogène (exemple transgène porteur du gène *GFP*) :

$$\text{cphg}_{\text{GFP}} = 2^{(\text{Cp}_{\text{genomic ref}} - \text{Cp}_{\text{GFP}})}$$

Le cphg doit ensuite être corrigé par un facteur (f) calculé à partir d'un standard. Comme standard j'ai utilisé l'ADNg i15 issu de paramécies injectées avec une fusion *PGM-GFP* pour lesquelles un cphg=6 a été déterminé par Southern blot (Dubois et al., 2017) puis confirmé par RT-qPCR. À chaque RT-qPCR, le standard est ajouté et le cphg_{standard} est calculé à partir du Cp obtenu.

$$\text{cphg}_{\text{standard}} \times f = 6 \leftrightarrow f = \frac{6}{\text{cphg}_{\text{standard}}}$$

Enfin le cphg_{GFP} obtenu est corrigé par le facteur f. C'est ce cphg qui sera directement indiqué dans les résultats.

$$\text{cphg}_{\text{GFP corrigé}} = \text{cphg}_{\text{GFP}} \times f$$

2.7.3 Plasmides utilisés pour injections

Les différents plasmides injectés sont résumés dans le Tableau 9.

n°	Nom
p112	pGemND7 / pND7 (Baudry et al., 2009)
p349	pUC_GFP-UTR_PGM
p364	pUC_GFP-UTR_PGM-Motif_Asel / Motif mutant
p418	pUC_GFP-UTR_PGM-Motif_INVlg / Motif inversé

Tableau 9 - Plasmides injectés

2.8 WESTERN BLOT

2.8.1 Récoltes

Filtration : 50 à 100 mL de culture de cellules autogames (~3 000 cellules/mL) sont filtrés sur des compresses de gaz préalablement mouillées au DRYL 1X (composition section 2.12.5). Rincer rapidement la compresse avec un peu de DRYL 1X et transférer les cellules dans un tube à centrifugé en forme de poire.

Lavage : Centrifuger les cellules à 800 rpm pendant 1 min pour les faire tomber au fond. À l'aide d'une baguette de verre, bloquer les 2 mL au fond de la poire et retirer le reste du milieu. Resuspendre ensuite les cellules dans 50 à 100 mL de DRYL 1X pour un lavage de 20 min à température ambiante.

Récolte : Centrifuger ensuite les cellules à 800 rpm pendant 1 min et utiliser la baguette de verre pour retirer un maximum de surnageant. Centrifuger une seconde fois à 800 rpm, mais pendant 2 min, pour reconcentrer les cellules et récupérer rapidement les cellules concentrées au fond avec une pipette pasteur. Répartir les cellules dans deux tubes Eppendorf 1,5 mL avant de les centrifuger 1 in à 14 C et 6 000 rpm dans une centrifugeuse *swinging*. Enfin retirer le surnageant pour ne garder que 20 µL et jeter les tubes dans de l'azote liquide pour les congeler. L'heure de récolte considérée et celle à laquelle les cellules sont placées dans l'azote. Les culots de cellules sont ensuite conservés à -80°C. Ils pourront être utilisés ultérieurement.

2.8.2 Préparation d'extrait protéique

Les culots de 20 μ L de paramécies sont sortis sur carboglace. On y ajoute ensuite rapidement 1 μ L d'inhibiteur de protéase PICS 100X (*Protease Inhibitor Cocktail Set1*, Merck Chemicals) et 20 μ L de SDS 10% bouillant. Le mix est ensuite rapidement vortexé et placé dans un bain d'eau bouillante pendant 3 min pour lyser les cellules. Le lysat est ensuite centrifugé 5 min à 25 000 g. Le surnageant est aliquoté par 11 μ L et 4 μ L de tampon Leampli 4X (Bio Rad) complémenté à 354 mM de β -mercaptoéthanol est ajouté. Le lysat est alors bouilli pendant 3 min. Il peut ensuite être directement chargé.

Les lysats peuvent également être conservés à -80°C avant ou après avoir été bouilli dans le Leampli. Cependant le β -mercaptoéthanol peu activer les métalloprotéases. Si les lysats doivent être gardés longtemps, il est donc préférable de garder les lysats avant l'ajout du Leampli.

2.8.3 Migration des protéines et western blot

Migration : Les protéines sont ensuite déposées sur des gels SDS-Page Criterion™ TGX *Stain-Free*™ 4-15% pré-coulés (Bio Rad) et migrés dans un tampon de migration (composition section 2.12) dans une cuve d'électrophorèse. Les gels utilisés étant *Stain-Free*, on peut vérifier l'état globale des préparations de protéines après migration au ChemiDocBR *Touch Imaging System* (Bio Rad) après une étape d'activation de 5 min.

Transfert : Le gel est ensuite transféré sur une membrane de nitrocellulose (*NC Protean nitrocellulose blotting membranes*, Amersham). La membrane est d'abord réhydratée quelques minutes dans le tampon de transfert (composition section 2.12.7) puis on monte la membrane et le gel entre deux papiers buvard et deux éponges dans une cassette (Figure 75). La cassette est ensuite placée dans une cuve de transfert remplie de tampons de transfert. Le transfert s'effectue en 1 h à 4°C et 500 mA.

Coloration de la membrane : Afin de vérifier l'efficacité du transfert, la membrane est plongée dans un bain de rouge ponceau (ThermoFisher™) pendant quelques minutes. L'excédent de ponceau est ensuite retiré rapidement à l'eau.

Une fois l'image du gel capturée, le reste de ponceau est rapidement éliminé par le tampon de lavage (composition section 2.12.8).

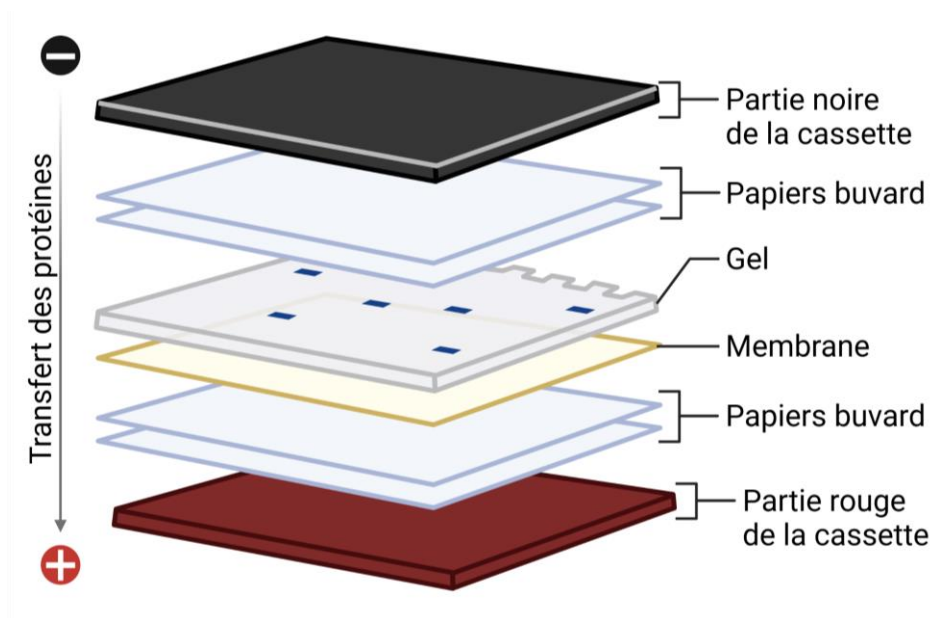


Figure 75 - Montage de la cassette de transfert de western blot
(Created with Biorender.com)

2.8.4 Immunomarquage de membrane de nitrocellulose

La membrane est d'abord bloquée par une incubation d'1h à température ambiante sous-agitation dans du tampon de lavage TBS-T (composition section 2.12.8) auquel est ajouté 4 g de Régilait écrémé pour 100 mL (appelé TBST-lait).

Marquage : L'anticorps primaire est ensuite ajouté, dilué à la concentration indiquée (Tableau 10), dans du TBST-lait et incubé sous agitation 1h à température ambiante ou toute la nuit à 4°C.

Anticorps	Dilution	Référence	Anticorps secondaire à utiliser
Anti-Tubuline	1/10 000	TEU435 (Callen et al., 1994)	Souris
Anti-Pgm	1/5 000	GP3 (Dubois et al., 2017)	Cochon d'inde
Anti-GFP	1/2 000	Padg1 (Chromo Tek)	Lapin
Anti-PgmL4	1/5 000	GP2	Cochon d'inde

Tableau 10 - Anticorps primaires utilisés en western blot

La membrane est ensuite lavée deux fois rapidement avec du tampon de lavage. Puis lavée deux fois 10 min avec agitation à température ambiante dans ce même tampon.

La membrane est incubée sous agitation 1h à température ambiante avec une dilution appropriée de l'anticorps secondaire dans du TBST-lait (Tableau 11). Le même processus de lavage, que décrit précédemment, est répété et la membrane est placée dans du TBS 1X (sans Tween) avant d'être révélée.

Anticorps	Dilution	Fournisseur
IgG Anti-Souris HRP	1/5 000	Promega
IgG Anti-Cochon d'inde HRP	1/5 000	ThermoScientific
IgG Anti-Lapin HRP	1/5 000	Promega #W4011

Tableau 11 - Anticorps secondaires utilisés en western blot

Révélation : Les anticorps secondaires utilisés étant couplés à la HRP (Peroxydase de raifort ou *Horseradish peroxidase*), ils sont révélés en utilisant le kit WesternBright ECL (Advansta). Les deux réactifs sont mélangés avec un ratio 1:1 et déposés afin de recouvrir la membrane. Après 2min, l'excédent de solution est éliminé et la membrane est révélée au ChemiDocBR.

2.9 IMMUNOFLUORESCENCE SUR CELLULES ENTIÈRES

Au cours des différentes étapes de ce protocole il est possible de perdre un grand nombre de cellules. Afin de limiter la perte de cellule il faut être très précautionneux lorsque les surnageants sont retirés, cette étape peut être effectuée sous loupe binoculaire pour voir le culot de cellule. Dans tous les cas, il faut penser à regarder régulièrement à la loupe si le culot de cellule est toujours présent et si elles n'ont pas été perdues lors des lavages.

2.9.1 Récoltes

Filtration : 20 à 50 mL de culture de cellules autogames (~3 000 cellules/mL) sont filtrés sur des compresses de gaz préalablement

mouillées au DRYL 1X (composition section 2.12.5). Rincer rapidement la compresse avec un peu de DRYL 1X et transférer les cellules dans une poire.

Lavage : Centrifuger les cellules à 800 rpm pendant 1 min pour les faire tomber au fond. À l'aide d'une baguette de verre bloquer les 2 mL au fond de la poire et retirer le reste du milieu. Resuspendre ensuite les cellules dans 50 à 100 mL de DRYL 1X pour un lavage de 20 min à température ambiante.

Récolte : Centrifuger ensuite les cellules à 800 rpm pendant 1 min et utiliser la baguette de verre pour retirer un maximum de surnageant. Centrifuger une seconde fois à 800 rpm pendant 1 min pour reconcentrer les cellules. Retirer rapidement le surnageant avec une pipette pasteur pour ne garder que 2 500 µL de cellules concentrées. Transférer les cellules dans un tube Eppendorf propre.

2.9.2 Perméabilisation et fixation

Afin d'observer la localisation stable des protéines présentes dans les ébauches, les cellules sont perméabilisées avant d'être fixées. Les protéines qui ne sont donc pas stablement associées sont donc lavées. Le temps de récolte considéré sera l'heure à laquelle les cellules ont été perméabilisées.

Perméabilisation : Prélever 40 µL de cellules concentrées et y ajouter 400 µL de PHEM-T froid (composition section 2.12.9). Mélanger en tapotant sur le tube et incuber les cellules 4 min à 18-20°C. Centrifuger 1 min à 200 g dans une centrifugeuse *swinging* pour faire tomber les cellules avant de retirer un maximum de surnageant.

Fixation : Ajouter 400 µL de PHEM-PAF froide (composition section 2.12.3) et resuspendre doucement les cellules en tapotant sur le tube. Incuber 15 min à 18-20°C. Centrifuger 1 min à 200 g dans une centrifugeuse *swinging* pour faire tomber les cellules avant de retirer un maximum de surnageant.

Lavages : Les lavages se font dans du TBST-BSA (composition section 2.12.11). 500 µL de TBST-BSA est ajouté sur le culot de cellules. Les cellules sont resuspendues doucement par tapotement du tube et incubées 10 min à 18-20°C. Centrifuger 1 min à 200 g dans une centrifugeuse *swinging* pour faire tomber les cellules avant de retirer un maximum de surnageant. Le lavage est effectué une seconde fois. À la fin du deuxième lavage, les cellules sont, soit remise en

suspension dans 500 μ L de TBST-BSA et garder la nuit à 4°C pour être marquées le lendemain ; soit un troisième lavage est effectué pour marquer les cellules directement.

2.9.3 Immunomarquage

Anticorps primaires : Les 200 μ L de cellules sont centrifugées 1 min à 1 000 rpm dans une centrifugeuse *swinging* à 14°C pour faire tomber les cellules et un maximum de surnageants est retiré. 200 μ L de TBST-BSA (composition section 2.12.11) contenant les anticorps primaires à la dilution souhaitée (Tableau 12) sont ajoutés et incubés 2h à 18-20°C.

Anticorps	Dilution	Référence	Anticorps secondaire à utiliser
Anti-Pgm	1/500	GP3 (Dubois et al., 2017)	Cochon d'inde
Anti-Flag	1/500	M2 (Sigma #F1804)	Souris

Tableau 12 - Anticorps primaires utilisés en immunofluorescence

Lavages : Les cellules sont centrifugées 1 min à 1 000 rpm dans une centrifugeuse *swinging* à 14°C et le surnageant est retiré. Elles sont ensuite resuspendues et incubées 10 min à 18-20°C dans 400 μ L de TBST-BSA avant d'être à nouveau centrifugées. Un second lavage est effectué de la même manière.

Anticorps secondaires : 200 μ L de TBST-BSA contenant les anticorps secondaires à la dilution souhaitée (Tableau 13) sont ajoutés sur le culot de cellule et incubé 45 min à 18-20°C dans le noir.

Anticorps	Dilution	Fournisseur
IgG Anti-Souris AF-568	1/500	ThermoFisher Scientific
IgG Anti-Cochon d'inde AF-488	1/500	ThermoFisher Scientific

Tableau 13 - Anticorps secondaires utilisés en immunofluorescence

Marquage DAPI : Les cellules sont lavées deux fois comme décrit précédemment mais dans le noir. Elles sont ensuite resuspendues dans du TBST-BSA contenant 0,5 μ g/mL de DAPI et incubées 5 min à 18-20°C dans le noir.

Montage sur lame : Les cellules sont centrifugées 1 min à 1 000 rpm dans une centrifugeuse *swinging* à 14°C. 4 μ L de cellules sont prélevés au fond du tube

et déposé sur une goutte de 8 μL de Citifulore AF2 sur une lame de microscopie. Les cellules sont ensuite recouvertes d'une lamelle (24x32 mm). La lamelle est fixée en vernissant les quatre coins. La lame est par la suite conservée à 4°C dans le noir. Le lendemain, la lamelle est scellée par l'ajout de vernis.

2.10 LAME DAPI

Lorsque l'on veut simplement regarder l'état des noyaux des paramécies pour mesurer l'avancement de l'autogamie, on peut utiliser un protocole simplifié avec un marquage uniquement au DAPI. Pour cela il faut 40 μL de cellules concentrées (comme pour le protocole d'immunofluorescence, voir section 2.9.1).

Perméabilisation et fixation : 200 μL de PHEM-T froid sont ajoutés directement sur les 40 μL de cellules concentrées. Les cellules sont incubées 4 min à température ambiante. Sans étape de centrifugation intermédiaire, on ajoute directement 400 μL de PHEM-PAF et on incube les cellules 10 min supplémentaire à température ambiante.

Lavages : Après une centrifugation d'1 min à 1 000 g dans une centrifugeuse *swinging*, un maximum de surnageants est retiré. On ajoute ensuite 500 μL de TBST-BSA sur le culot de cellules. Les cellules sont resuspendues doucement par tapotement du tube et incubées 5 min à température ambiante. Un second lavage est effectué de la même manière.

Marquage DAPI : Les cellules sont ensuite resuspendues dans du TBST-BSA contenant 0,5 $\mu\text{g}/\text{mL}$ de DAPI et incubées 5 min à température ambiante dans le noir.

Montage sur lame : Le montage s'effectue exactement comme décrit dans le protocole d'immunofluorescence (voir section 2.9.1).

Observation des lames : Les lames sont ensuite observées aux microscopes à épifluorescence. Une centaine de cellules choisies aléatoirement est sélectionnée et classée selon leur stade. On évalue ainsi la proportion de chaque stade (Figure 76) dans la culture et donc la progression de l'autogamie.

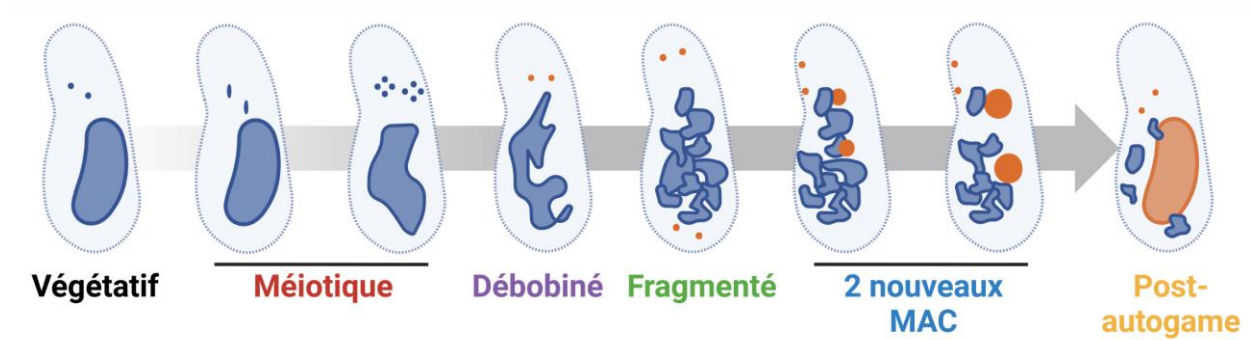


Figure 76 - Définition des stades d'autogamie
(Created with Biorender.com)

2.11 IMMUNOFLUORESCENCE SUR NOYAUX

2.11.1 Récoltes

Filtration : 200 mL à plusieurs litres de culture de cellules autogames (~3 000 cellules/mL) sont filtrés sur des compresses de gaz préalablement mouillées au Tris 10mM pH7,4. Rincer rapidement la compresse avec un peu de Tris 10mM pH7,4 et transférer les cellules dans une poire.

Lavage : Centrifuger les cellules à 800 rpm pendant 1 min pour les faire tomber au fond. À l'aide d'une baguette de verre bloquer les 2 mL au fond de la poire et retirer le reste du milieu. Transférer les 2 mL dans une poire propre et recommencer l'opération jusqu'à ce que tout le volume de culture soit dans la poire. Centrifuger la poire avec les cellules concentrées à 800 rpm pendant 1 min et garder uniquement les 2 mL contenant les cellules. Resuspendre ensuite les cellules dans 100 mL de Tris 10 mM pH7,4 froids pour un lavage rapide.

Récolte : Centrifuger ensuite les cellules à 800 rpm pendant 1 min et utiliser la baguette de verre pour retirer un maximum de surnageants. Centrifuger une seconde fois à 800 rpm pendant 2 min pour reconcentrer les cellules. Retirer un maximum de surnageants avec une pipette et évaluer le volume de cellules concentrées obtenu.

2.11.2 Lyse des cellules

Les cellules concentrées sont transférées dans un potter (préalablement refroidis sur glace) et 10 volumes de tampon de lyse sont ajoutés doucement (composition section 2.12.12). La poire est rincée avec une partie du volume afin de récupérer les cellules restantes. Après 10 à 15min d'incubation dans la glace, les cellules sont lysées au potter. La lyse est vérifiée à la loupe à épifluorescence en ajoutant 2 μL de DAPI 0,2 mg/mL dans 4 μL de lysat. Lorsque plus aucune cellule n'est intacte, on considère que la lyse est complète. Il ne faut cependant pas mettre trop de coup de potter afin d'éviter d'abimer les noyaux.

2.11.3 Centrifugation basse vitesse

Le lysat est lavé et enrichi en nouveaux MAC grâce à une succession de centrifugation à basse vitesse. Le lysat est transféré dans un ou deux tubes Falcon 15 mL et centrifugé à 1000 g pendant 2 min dans une centrifugeuse *swinging* à 4°C. Le surnageant est retiré et le culot est resuspendu en tapotant doucement sur le tube. Puis on ajoute 10 mL de tampon de lavage pour noyaux (composition section 2.12.13) avant de centrifuger dans les mêmes conditions que précédemment. Les lavages sont répétés jusqu'à ce que le surnageant soit claire (en général 4 à 5 lavages suffisent).

Les culots sont ensuite resuspendus dans le tampon de conservation (composition section 2.12.14) avant d'être aliquotés et congelés à l'aide d'un Mr. Frosty™ (ThermoFisher 5100-0001). Les noyaux sont conservés à -80°C.

2.11.4 Marquages des noyaux

Décongeler un aliquot de 20 μL de noyaux sur glace. Ajouter 400 μL de TBS-BSA (composition section 2.12.16) et centrifuger à 500 g pendant 1 min dans une centrifugeuse *swinging* à 4°C. Suivant la taille du culot, répartir les noyaux dans plusieurs tubes.

Anticorps primaires : Centrifuger à nouveau dans les mêmes conditions pour faire tomber les noyaux et retirer un maximum de surnageants. 400 μL de TBS-BSA, contenant les anticorps primaires à la dilution souhaitée (Tableau 14), sont ajoutés et incubés 1 h à 4°C.

Anticorps	Dilution	Référence	Anticorps secondaire à utiliser
Anti-PgmL1	1/1000	GP1 (Zangarelli et al., 2022)	Cochon d'inde

Tableau 14 - Anticorps primaires utilisés en immunofluorescence sur noyaux

Lavages : Les noyaux sont centrifugés dans les mêmes conditions que précédemment et le surnageant est retiré. Ils sont ensuite resuspendus en tapotant sur le tube et incubés 10 min à 4°C dans 400 µL de TBS-BSA avant d'être à nouveau centrifugés. Un second lavage est effectué de la même manière.

Anticorps secondaires : 400 µL de TBS-BSA contenant les anticorps secondaires à la dilution souhaitée (Tableau 13) sont ajoutés sur le culot de cellules et incubés 45 min à 4°C dans le noir.

Lavages : Les noyaux sont centrifugés dans les mêmes conditions que précédemment et le surnageant est retiré. Ils sont ensuite resuspendus en tapotant sur le tube et incubés dans le noir 10 min à 4°C dans 400 µL de TBS-BSA.

Marquage Iodure de propidium : ~15 min avant d'aller au FACS, centrifuger les noyaux toujours avec les mêmes paramètres. Un maximum de surnageants est retiré. Les noyaux sont resuspendus en tapotant sur le tube avant d'ajouter 400 µL de tampon PI/RNase (BD Pharmingen, réf. 550825). Les tubes sont conservés sur glace dans le noir jusqu'à passage au FACS analytique ou au trieur.

2.12 COMPOSITIONS DES TAMPONS

Dans cette partie sont regroupées les compositions des différentes solutions citées dans les protocoles ainsi que des exemples de préparation de ces solutions.

2.12.1 WGP 20X

Faire infuser 100g de poudre WPG dans 800 mL d'eau MilliQ bouillante pendant 20 min. L'infusion est ensuite filtrée d'abord sur une épaisse couche de gaze en coton puis dans un filtre en cellulose (VELIN PRAT DUMAS France)

recouvert d'un préfiltre en fibre de verre (Millipore AP2512450) et d'une épaisseur de gaze.

Infusion WGP filtrée	800 mL
Tris base	15 g
Na ₂ HPO ₄	15 g
NaH ₂ PO ₄	4 g
Eau MilliQ	Qsp 1 L

Le pH est ensuite équilibré à 7 en utilisant du HCl. Enfin la solution est stérilisée à 120°C pendant 20min.

2.12.2 PHEM 5X

60 mM Pipes, 25 mM Hepes, 10 mM EGTA, 2 mM MgCl₂ pH6,9 (NaOH)

Pipes (302,37 g/mol)	907,1mg
Hepes (238,3 g/mol)	397,9 mg
EGTA (380,35 g/mol)	190,2 mg
MgCl ₂ 1 M	100 µL
Eau MilliQ	Qsp 50 mL

La solution reste laiteuse tant que le pH n'est pas ajusté à 6,9. La solution est donc assez longue à s'homogénéiser. Une fois complètement homogène, la solution est filtrée sur 0,2 µm et conservée à -20°C ou 4°C.

2.12.3 PHEM-PAF

PHEM 1X, 0,2% paraformaldéhyde

PHEM 5X	2 mL
Paraformaldéhyde 36,8-38%	540 µL
Eau MilliQ	Qsp 10 mL

Filtré à 0,2 µm et conservé à -20°C ou 4°C pour faire la solution de DAPI-PAF. Toutefois, lorsque l'on utilise le PHEM-PAF pour faire de l'immunofluorescence, on utilisera du PHEM-PAF frais du jour ou de la veille. De plus, il faut veiller à ce que le paraformaldéhyde n'ait pas été ouverte depuis plus d'un mois.

2.12.4 DAPI-PAF

20 µg/mL DAPI, 0,2% Paraformaldéhyde

DAPI 1 mg/mL	20 µL
PHEM-PAF	100 µL
Eau MilliQ	80 µL

Le DAPI-PAF peut être conservé 1-2 semaine à 4°C dans le noir.

2.12.5 DRYL

Afin de faire 1 L de DRYL 1X, 10 mL de DRYL stock 100X sont dilués dans 900 mL d'eau. On ajoute ensuite 1,5 mL de CaCl₂ 1M avant d'ajuster le volume final à 1 L avec de l'eau. Le DRYL stock 100X et le CaCl₂ ne doivent surtout pas être mélangé avant d'être dilué dans l'eau car cela induit la précipitation des sels.

DRYL stock 100X

100 mM Na₂H-PO₄, 100 mM NaH₂-PO₄, 200 mM Na₃-citrate

Na ₂ H-PO ₄ (177,99 g/mol)	8,9 g
NaH ₂ -PO ₄ (138,99 g/mol)	7 g
Na ₃ -citrate (294 g/mol)	29,4 g
Eau MilliQ	Qsp 500 mL

Le Na₂H-PO₄ doit être ajouté doucement pour ne pas précipiter. La solution est ensuite autoclavée avant d'être utilisée.

CaCl₂

1 M CaCl₂

CaCl ₂ (110,98 g/mol)	5,5 g
Eau MilliQ	Qsp 50 mL

La solution est ensuite autoclavée avant d'être utilisée.

2.12.6 Tampon de migration

Tampon de migration 1X : 25 mM Tris, 200 mM Glycine, 0,1% SDS

De manière général, un volume assez important de tampon est utilisé pour chaque migration de gel de protéine (~1 L). Un stock 10X est donc préparé de la manière suivante :

Tris-base (121,14 g/mol)	30 g
Glycine (75,07 g/mol)	144 g
SDS 20%	10 mL
Eau MilliQ	Qsp 1 L

Le tampon de migration est ensuite dilué à 1X dans de l'eau avant son utilisation.

2.12.7 Tampon de transfert

Tampon de transfert 1X : 20mM phosphate pH6,7

De manière général, un volume assez important de tampon est utilisé pour chaque transfert de gel de protéine (~1,5 L). Un stock 50X est donc préparé de la manière suivante :

Na ₂ H-PO ₄ (177,99 g/mol)	77,4 g
NaH ₂ -PO ₄ (138,99 g/mol)	78,5 g
Eau MilliQ	Qsp 1 L

Le Na₂H-PO₄ doit être ajouté doucement pour ne pas précipiter. Le tampon de migration est ensuite dilué à 1X dans de l'eau avant son utilisation.

2.12.8 Tampon de lavage TBS-T

Tampon de lavage 1X : 50 mM Tris pH8, 150 mM NaCl, 0,5% Tween

Le tampon de lavage va permettre de laver les membranes de western blot. Il est donc utilisé en grande quantité. Une solution de TBS stock à 10X de la manière suivante :

NaCl 5 M	30 mL
Tris-base 1M pH8	500 mL
Eau MilliQ	Qsp 1 L

Pour préparer le tampon de lavage on dilue 100 mL de TBS 10X avec 900 mL d'eau pour avoir du TBS 1X auquel on ajoute 5 mL de Tween 20.

2.12.9 PHEM-T

PHEM 1X, 1% Triton X100

Ce tampon est préparé la veille des immunofluorescences car le triton est assez long à resuspendre. De petits volumes sont nécessaires, on fera par exemple :

PHEM 5X	2 mL
Triton X100	100 µL
Eau MilliQ	Qsp 10 mL

La solution est ensuite filtrée sur un filtre 0,2 µm avant d'être utilisée. Elle peut aussi être préparée en plus grand volume et être conservée à -20°C avant d'être utilisée.

2.12.10 TBST

10 mM Tris pH7,4, 150 mM NaCl, 0,1% Tween

Attention, même s'il s'agit d'un tampon TBS-T, comme le tampon de lavage de western blot, les concentrations en tris et tween divergent ainsi que les pH utilisés. Les deux TBS-T ne sont donc pas interchangeables. Pour le TBS-T pour les immunofluorescences on procédera de la manière suivante :

NaCl 5 M	7,5 mL
Tris-base 1M pH7,4	2,5 mL
Eau MilliQ	Qsp 250 mL
Tween 20	250 µL

Le tween 20 est ajouté en dernier et la solution est ensuite filtrée sur un filtre 0,2 µm avant d'être utilisé.

2.12.11 TBST-BSA

TBST 1X, 3% BSA

Pour cette solution, 3 g de BSA (*Bovin Serum Albumin*) sont ajoutés pour 100 mL de TBST 1X. La BSA étant assez longue à se solubiliser, cette solution est généralement préparée en grand volume (500 mL) et filtrée sur un filtre 0,2 µm avant d'être aliquotée par 50 mL et conservée à -2°C pour les différentes utilisations. Elle peut aussi être préparée la veille des récoltes de cellules pour immunofluorescence.

2.12.12 Tampon de lyse

10 mM Tris pH6,8, 250 mM Sucrose, 10 mM MgCl₂, 0,2 % NP-40, 2X Inhibiteur de protéase Calbiochem

Sucrose 2,5 M	1 mL
Tris-base 1M pH6,8	2,5 mL
MgCl ₂ 1 M	100 µL
NP-40	20 µL
Eau MilliQ	Qsp 10 mL
Inhibiteur Calbiochem 100X	200 µL

Le tampon de lyse peut être préparé la veille de l'utilisation mais les inhibiteurs sont ajoutés le jour de l'utilisation et la solution est ensuite filtrée sur un filtre 0,2 µm. Il est gardé à 4°C avant d'être utilisé.

2.12.13 Tampon de lavage pour noyaux

10 mM Tris pH6,8, 250 mM Sucrose, 10 mM MgCl₂

Sucrose 2,5 M	25 mL
Tris-base 1M pH6,8	2,5 mL
MgCl ₂ 1 M	2,5 µL
NP-40	20 µL
Eau MilliQ	Qsp 250 mL

Le tampon de lavage est fait la veille de la récolte et gardé à 4°C.

2.12.14 Tampon de conservation

10 mM Tris pH6,8, 250 mM Sucrose, 10 mM MgCl₂, 0,2 % NP-40, 2X Inhibiteur de protéase Calbiochem, 15% glycérol

Pour cette solution XX µL de glycérol 100% et 200 µL d'inhibiteur de protéase 100X sont ajoutés dans 10mL de tampon de lavage le jour de la récolte. Il est gardé à 4°C avant d'être utilisé.

2.12.15 TBS

10 mM Tris pH7,4, 150 mM NaCl

Attention, même si le tampon est très proche de celui utilisé pour les cellules entières, utiliser un tampon TBS-T sur les noyaux risque de les abimer. Pour le TBS pour les immunofluorescences sur noyaux on procédera de la manière suivante :

NaCl 5 M	7,5 mL
Tris-base 1M pH7,4	2,5 mL
Eau MilliQ	Qsp 250 mL

La solution est ensuite filtrée sur un filtre 0,2 µm avant d'être utilisée ou aliquotée par 50 mL et congelée à -20°C.

2.12.16 TBS-BSA

TBS 1X, 3% BSA

Pour cette solution, 3 g de BSA sont ajoutés pour 100 mL de TBS 1X. La BSA étant assez longue à se solubiliser, cette solution est généralement préparée en grand volume (500 mL) et filtrée sur un filtre 0,2 µm avant d'être aliquotée par 50 mL et conservée à -2°C pour les différentes utilisations. Elle peut aussi être préparée la veille des récoltes de cellules pour immunofluorescence.

Chapitre 3 - Cut&Tag

Une importante partie de ma thèse a été consacrée à la mise au point du protocole de Cut&Tag sur la paramécie. Dans cette partie sont décrits plus en détail les versions protocoles ayant permis d'obtenir des résultats.

3.1 PRODUCTION D'UNE PA-Tn5 ACTIVE

3.1.1 Production de la pA-Tn5

3.1.1.1 Production bactérienne

Culture : Des *E. coli* dérivée de BI21 (C3013 NEB) sont transformées par le plasmide p382_pA-Tn5 (Addgene 124601) et étalées sur boîte LB avec 100 µg/mL d'ampicilline (LBA). Une colonie est ensuite repiquée dans 30 mL de LBA et incubée 3h à 37°C avec agitation. La culture est ensuite diluée dans 2 L final de LBA préchauffé à 37°C et incubée à 30°C avec agitation jusqu'à atteindre une DO_{600nm} de 0,5. La culture est ensuite placée 1h à 4°C avant d'ajouter 0,25 mM final d'IPTG. On laisse ensuite la culture pousser sur la nuit à 18°C avec agitation.

Récolte : La culture est centrifugée à 4 000 rpm pendant 30 min à 4°C répartie dans quatre pots de 500 mL. Un maximum de surnageants est retiré et chaque culot est resuspendu dans 10 mL d'HEGX + inhibiteur froid (composition section 3.1.4.1). Les culots resuspendus sont ensuite répartis dans 2 Falcon 50 mL et centrifugés 4 000 rpm pendant 15 min à 4°C. Le surnageant est retiré et le culot sec est congelé dans un bain d'éthanol et carboglace avant d'être conservé à -80°C.

3.1.1.2 Extraction pA-Tn5

La pA-Tn5 produite est fusionnée en C-terminal à un domaine intéine suivi d'un domaine de liaison à la chitine facilitant sa purification. Toutes les étapes suivantes se font à 4°C ou sur la glace et toutes les solutions utilisées sont préalablement placées à 4°C afin d'être utilisées froides.

Lyse : Un culot (correspondant à ~1 L de culture) est décongelé sur glace et resuspendu dans 10 mL final avec de HEGX + inhibiteur. Une première lyse à la presse de Carver est effectuée. Le lysat est récupéré dans un Falcon 50 mL et décongelé à température ambiante. Le lysat étant encore très visqueux, 10 mL d'HEGX + inhibiteur sont ensuite ajoutés avant de finaliser la lyse et fragmenté l'ADN bactérien par sonication. Le lysat obtenu est alors fluide et légèrement rosé. Le lysat est centrifugé à 200 000 g pendant 90 min à 4°C dans un rotor à angle fixe (50.2 Ti). On obtient alors un surnageant jaune et un culot rose foncé.

Précipitation ADN bactérien : Du polyethyleneimine (PEI) à 0,25% w/v final est ajouté dans le surnageant avant une centrifugation à 200 000 g pendant 60 min à 4°C dans un rotor à angle fixe (50.2 Ti). L'ADN précipite et forme un culot jaune translucide. Le surnageant est conservé et transféré dans une Falcon 50mL.

Fixation sur la chitine : Avant leur utilisation, 6 mL de résine de chitine (S6651S NEB) sont lavés trois fois avec 10 mL d'HEGX + inhibiteur. Entre chaque lavage la résine est précipitée par une centrifugation de 30 sec à 400 g et 4°C. Le culot de chitine tassé correspond à environ 4 mL. Après le dernier lavage, le lysat est transféré sur le culot de bille de résine et incubé une nuit à 4°C sur roue.

Lavages : La chitine est centrifugée 30 sec à 400 g et 4°C et le surnageant est éliminé. La chitine est ensuite resuspendue dans 5 mL d'HEGX sans inhibiteur et répartie dans deux colonnes d'élution (Poly-prep Bio-Rad). La chitine est lavée deux fois avec 20 mL d'HEGX par colonne. Une fois l'HEGX passé, 750 µL d'HEGX avec 130 mM de DTT est ajouté sur chacune des colonnes.

Élution : La colonne est fermée et 1,5 mL d'HEGX avec 100 mM de DTT est ajouté sur chaque colonne. Le DTT vas induire le clivage de l'intéine et permettre la libération de la pA-Tn5. Les colonnes sont incubées sans agitation à 4°C pendant 48 h ou sur le week-end. L'éluât des deux colonnes est ensuite récupéré par gravité dans un même Falcon 15 mL. Puis 750 µL d'HEGX est ajouté sur chacune

des colonnes pour chasser le volume mort et récupérer un maximum de protéines.

Dialyse : Afin de reconcentrer la pA-Tn5 et de la passer dans un tampon riche en glycérol permettant sa conservation, l'éluât est placé dans un boudin de dialyse et incubé une nuit avec agitation à 4°C dans le tampon de dialyse (composition section 3.1.4.2). La pA-Tn5 est ensuite transférée dans des tubes Eppendorf et stockée à -20°C.

3.1.2 Formation du complexe actif

Afin d'être active la Tn5 a besoin d'être chargée avec des oligonucléotides double-brin. Pour son utilisation dans le protocole de Cut&Tag la pA-Tn5 est chargée avec des oligonucléotides porteurs des adaptateurs de séquençage illumina A ou B afin de permettre la formation des banques directement après la tagmentation.

Les oligonucléotides simple-brin (Tableau 15) ont été commandé chez Eurofins avec une purification HPLC. Ils sont reçus lyophilisés et resuspendus à 100 µM final dans de l'eau MilliQ.

Identifiant	Nom	Séquence (5' → 3')
OMB1366	ME-A	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
OMB1367	ME-B	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
OMB1368	ME-rev	[PHO]-CTGTCTCTTATACACATCT (5' phosphate)

Tableau 15 - Oligonucléotides chargés sur la pA-Tn5

Hybridation : Dans un mix on mélange dans des tubes PCR : 22,5µL d'ME-A, 22,5 µL d'ME-rev et 5 µL de tampon d'hybridation (composition section 3.1.4.3). Le même mix est fait avec l'oligonucléotides ME-B. Les mix sont ensuite placés dans un bloque PCR et suivent le programme ci-après : 95°C 5 min, 65°C 5 min, 4°C. Les oligonucléotides double-brin sont ensuite conservés à 4°C.

Assemblage du complexe : 50 µL de pA-Tn5 produite dans le laboratoire sont mélangés avec 25 µL d'oligonucléotides double-brin ME-A/rev et 25 µL de ME-B/rev. Le mix est incubé 50 min dans un bloque PCR à 23°C puis conservé à 4°C.

3.1.3 Test d'activité

Afin de vérifier que le complexe qui vient d'être assemblé est bien actif, on effectue un test d'activité. De plus, le complexe pA-Tn5 assemblé est peu stable, il est donc nécessaire de vérifier régulièrement son activité.

L'activité de la pA-Tn5 assemblée est testée par digestion d'un plasmide. Pour cela on réalise un mix de 40 μL contenant 600 ng de plasmides, la quantité de pA-Tn5 souhaitée (2 μL en général) et du tampon de tagmentation 1X final (composition section 3.1.4.4). Le mix est aliquoté par 10 μL dans des tubes PCR.

Les quatre aliquots sont placés à 55°C pendant 5, 10, 15 ou 30 min, afin d'obtenir une cinétique de tagmentation. La réaction est ensuite arrêtée en ajoutant 1 μL de SDS 1% et 0,5 μL de protéinase K 20 $\mu\text{g}/\mu\text{L}$. Le mix est vortexé et rapidement centrifugé avant d'être placé à 55°C pendant 15 min. Une fois les incubations terminées, les différentes conditions sont gardées sur glace avant d'être déposés sur gel. Comme contrôle négatif, le test d'activité sera également effectué sur la pA-Tn5 non chargée.

Après l'ajout d'un tampon de charge, les différentes conditions sont déposées sur un gel d'agarose 1% (TAE 1X) et migrés 45 min à 135 V.

3.1.4 Composition des tampons

Dans cette partie sont regroupées les compositions des différentes solutions citées dans les protocoles ainsi que des exemples de préparation de ces solutions.

3.1.4.1 HEGX

20 M HEPES-KOH pH 7,2, 1 mM NaCl, 1 mM EDTA, 10% glycérol, 0,2% triton X-100

HEPES-KOH 1 M pH7,2	10 mL
NaCl 5 M	100 mL
EDTA 0,5 M	1 mL
Glycérol 100%	50 mL
Triton X-100	1 mL
Eau MilliQ	Qsp 500 mL

L'HEGX est un tampon HEPES, il est donc conservé à 4°C et fait à partir d'une solution stock d'HEPES également conservée à 4°C. Avant son utilisation, avec 15 µL d'inhibiteur de protéase Calbiochem, 100X par mL d'HEGX peut être ajouté pour faire du HEXG + inhibiteur.

3.1.4.2 Tampon de dialyse

40 mM HEPES-KOH pH 7,2, 100 mM NaCl, 0,1 mM EDTA, 50% glycérol, 0,05% triton X-100

Les dialyses s'effectuent contre 2 L de tampon :

HEPES-KOH 1 M pH7,2	80 mL
NaCl 5 M	40 mL
EDTA 0,5 M	0,4 mL
Glycérol 100%	1 L
Triton X-100	1 mL
Eau MilliQ	Qsp 2 L

Le tampon de dialyse est un tampon HEPES, il est donc conservé à 4°C et fait à partir d'une solution stock d'HEPES également conservée à 4°C. Resuspendre le glycérol étant assez long, le tampon de dialyse doit être préparé en avance.

3.1.4.3 Tampon d'hybridation

100 mM Tris pH8, 500 mM NaCl, 10 mM EDTA

Une solution stock 10X est constituée de la manière suivante :

Tris 1 M pH8	500 µL
NaCl 5 M	500 µL
EDTA 0,5 M pH8	125 µL
Eau MilliQ	Qsp 5 mL

Le tampon est ensuite utilisé à 1X final dans le mix d'hybridation des oligonucléotides.

3.1.4.4 Tampon de tagmentation

50 mM Tris pH8, 25 mM MgCl₂, 40% (w/v) PEG-4000

Tampon issue du protocole de : Picelli et al., 2014.

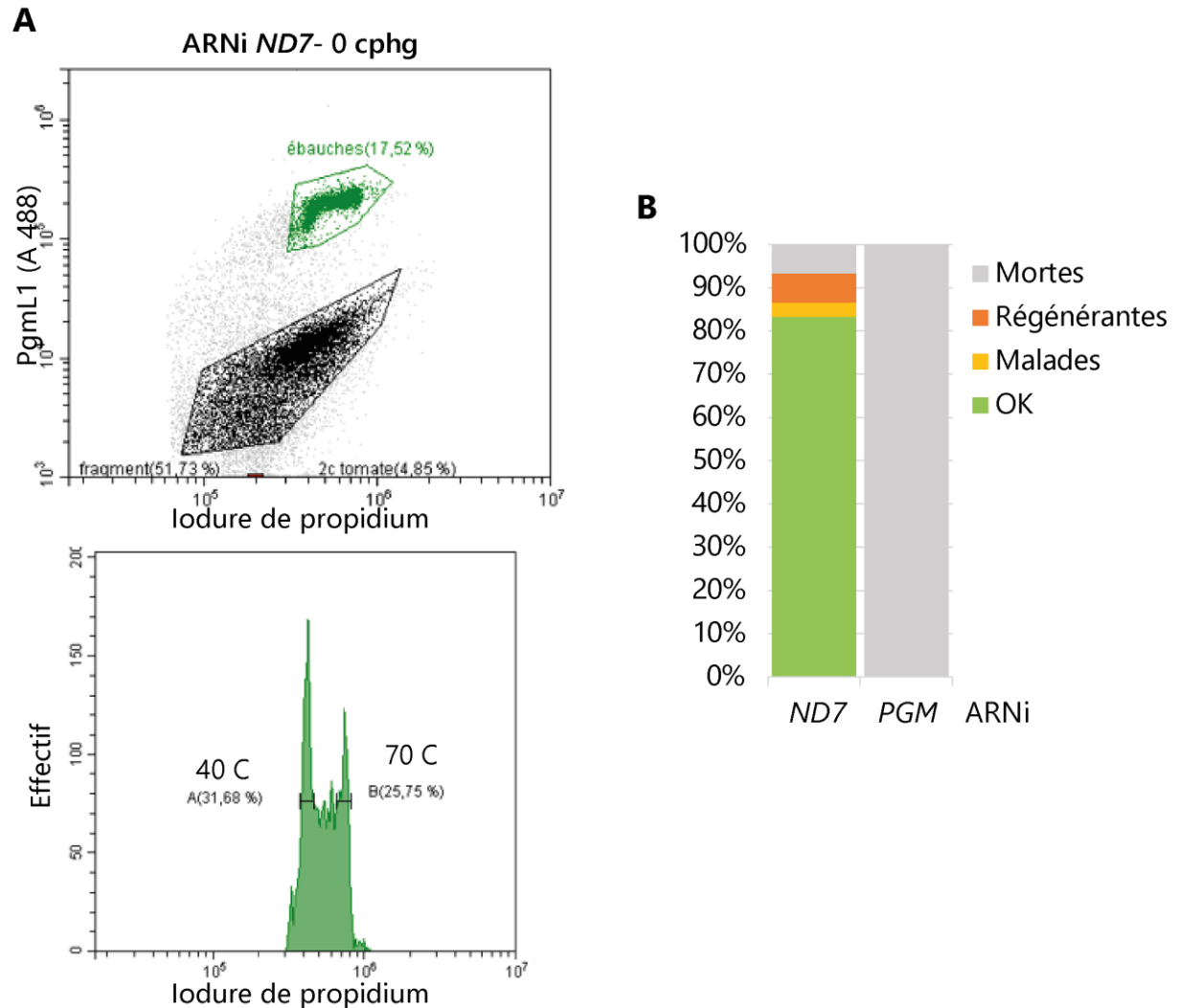


Figure 77 - Contrôle qualité des noyaux utilisés en Cut&Tag pour vérifier la localisation de la pA-Tn5

A La qualité et la ploïdie des noyaux extraits à T6 sont contrôlées par flow cytométrie après marquage des noyaux par un anti-PgmL1 et un anticorps secondaire couplé à un fluorophore AF488. Les noyaux correspondants aux nouveaux MAC (ébauches) sont identifiés en vert (haut). Le contenu en ADN de ces noyaux est ensuite représenté pour évaluer la ploïdie (bas) B La qualité des cultures et du passage de l'autogamie est contrôlée par un test de survie.

3.2 MARQUAGE DES PROTÉINES EN CUT&TAG

Afin de vérifier la localisation des protéines au cours du Cut&Tag, j'ai réalisé un protocole à partir du protocole de Cut&Tag pour marquer les protéines. Deux versions du protocole ont été effectués, un pour marquer les protéine cible du Cut&Tag et un pour marquer la pA-Tn5.

3.2.1 Matériel utilisé pour la localisation des protéines cibles

Noyaux : Les noyaux sont issus de cellules transformées avec le plasmide linéarisé p361_3xFlag-Pgm et cultivés en condition d'ARNi contre *PGM*. La qualité des noyaux a été vérifiée par cytométrie en flux (Figure 51, partie Résultats). La densité en noyaux a été estimée à ~50 000 noyaux pour 20 μ L. Un culot de 20 μ L de noyaux est utilisé pour faire 4 conditions de Cut&Tag. Les noyaux sont décongelés à température ambiante et sont resuspendus dans 180 μ L de tampon de lavage (composition section 3.4.1) est ajouté sur chaque culot. Tous les noyaux sont mélangés avant d'être divisée en aliquots de 50 μ L dans des tubes de 500 μ L.

Transposase : Aucun pA-Tn5 n'est ajouté.

Anticorps primaire : anti-H3K4me3 (ab184677, Abcam, lot : GR3346934-4), anti-Flag (F1804 clone M2, Sigma, lot : SLB7654) et anti-Pgm (GP3, Promega, (Dubois et al., 2017)). Les anticorps primaires sont dilués au 1 :100^{ème} dans le tampon pour anticorps (composition section 3.4.3).

Anticorps secondaire respectifs : IgG Anti-Lapin AF-647 ThermoFisher Scientific), IgG Anti-Souris AF-568 (ThermoFisher Scientific), IgG Anti-Cochon d'inde AF-488 (ThermoFisher Scientific). Les anticorps secondaires couplés à un fluorophore sont dilués au 1 :2000^{ème} dans le tampon de lavage (composition section 3.4.1).

3.2.2 Matériel utilisé pour la localisation de la pA-Tn5

Noyaux : Les noyaux utilisés sont issus de cellules non-transformées et cultivées en condition d'ARNi contre *ND7*. La qualité des noyaux a été vérifiée par cytométrie en flux (Figure 77). La densité en noyaux a été estimée à ~50 000 noyaux pour 20 μ L. Un culot de 20 μ L de noyaux est utilisé pour faire 4 conditions

de Cut&Tag. Les noyaux sont décongelés à température ambiante et sont resuspendus dans 180 μ L de tampon de lavage (composition section 3.4.1) est ajouter sur chaque culot. Tous les noyaux sont mélangés avant d'être divisés en aliquots de 50 μ L dans des tubes de 500 μ L.

Transposase : Une pA-Tn5 flaguée produite dans l'équipe de Carl Mann diluée au 1 :50^{ème} dans du tampon de lavage-300 (composition section 3.4.2).

Anticorps primaire : anti-H3K4me3 (ab184677, Abcam, lot : GR3346934-4) et anti-Flag (F1804 clone M2, Sigma, lot : SLB7654). Les anticorps primaires sont dilués au 1 :100^{ème} dans le tampon pour anticorps (composition section 3.4.3).

Anticorps secondaire respectifs : anti-lapin (ABIN101961, Antibodies-online, lot : NE-200-082001), IgG Anti-Souris AF-568 (ThermoFisher Scientific). L'anticorps secondaire non-coupé est dilué au 1 :100^{ème} dans le tampon de lavage (composition section 3.4.1). L'anticorps secondaire couplé à un fluorophore est dilué au 1 :2000^{ème} dans le même tampon.

3.2.3 Protocole

Sauf indication contraire, tout le protocole de Cut&Tag se passe à température ambiante et toutes les solutions utilisées sont conservées sur glace.

Préparation de billes : 5 μ L de billes magnétiques de concanavaline A (C7642, Sigma-Aldrich) est utilisé par réaction. La quantité nécessaire pour toutes les conditions est resuspendu par pipetage dans 1,5 mL de tampon de fixation (composition section 3.4.4). Le tube est placé sur un portoir aimanté pendant 2 min pour que toutes les billes se fixent. Le surnageant est retiré puis les billes sont resuspendues par pipetage dans 1 mL de tampon de fixation. Le tube est à nouveau placé sur le portoir aimanté le temps que toutes les billes se fixent sur la paroi et le surnageant est retiré. Les billes sont resuspendus dans un volume de tampon de fixation correspondant au volume de billes prélevé. Si des billes sont préparées pour 20 réactions de Cut&Tag, 105 μ L de billes sont lavées et finalement resuspendues dans 105 μ L de tampon de fixation.

Fixation des noyaux sur billes : 5 μ L de billes sont ajoutés dans chacun des tubes contenant 50 μ L de noyaux. Les noyaux sont incubés avec les billes sur roue pendant 10 min. Les tubes sont ensuite centrifugés rapidement avant d'être

placés sur le portoir aimanté. Une fois toutes les billes contre la paroi, le surnageant est retiré.

Anticorps primaire : Les noyaux liés aux billes sont resuspendus dans 50 μL de tampon pour anticorps contenant l'anticorps primaire. Pour chaque condition un contrôle négatif sans anticorps primaire est également réalisé. Les noyaux sont incubés sur roue toute la nuit à 4°C.

Lavage : Après une rapide centrifugation, les tubes sont placés sur le portoir aimanté. Une fois les billes fixées, le surnageant est retiré et 500 μL de tampon de lavage (composition section 3.4.1) est ajouté sans perturber les billes. Le tampon de lavage est retrié puis, les noyaux sont resuspendus dans 50 μL de tampon de lavage propre et incubés 10 min sur roue. Les tubes sont ensuite centrifugés rapidement et placés sur le portoir aimanté. Une fois toutes les billes contre la paroi, le surnageant est retiré.

Anticorps secondaire : Les noyaux sont resuspendus dans 50 μL de tampon de lavage contenant l'anticorps secondaire adéquat. Des anticorps secondaires sont également ajoutés dans les contrôles négatifs sans anticorps primaire. Les noyaux sont incubés 1h sur roue avec les anticorps secondaires.

Lavage : Après une rapide centrifugation, les tubes sont placés sur le portoir aimanté jusqu'à fixation des billes. Le surnageant est ensuite retiré et remplacé par 500 μL de tampon de lavage sans perturber les billes. Le tampon de lavage est ensuite retiré.

Pour chaque condition d'anticorps un tube est utilisé pour être maqué dans du tampon de lavage avec du DAPI 0,5 mg/mL final. Après 5 min d'incubation dans le noir et une centrifugation d'1 min à 500 g, 2 μL prélevé au fond du tube sont montés entre lame et lamelle dans 4 μL de Citifluore.

Fixation de la pA-Tn5 : Les noyaux sont resuspendus dans 50 μL de tampon de lavage-300 (composition section 3.4.2) avec ou sans pA-Tn5 et incubés 1h sur roue. Les billes sont ensuite lavées une fois comme décrit précédemment mais dans du tampon de lavage-300.

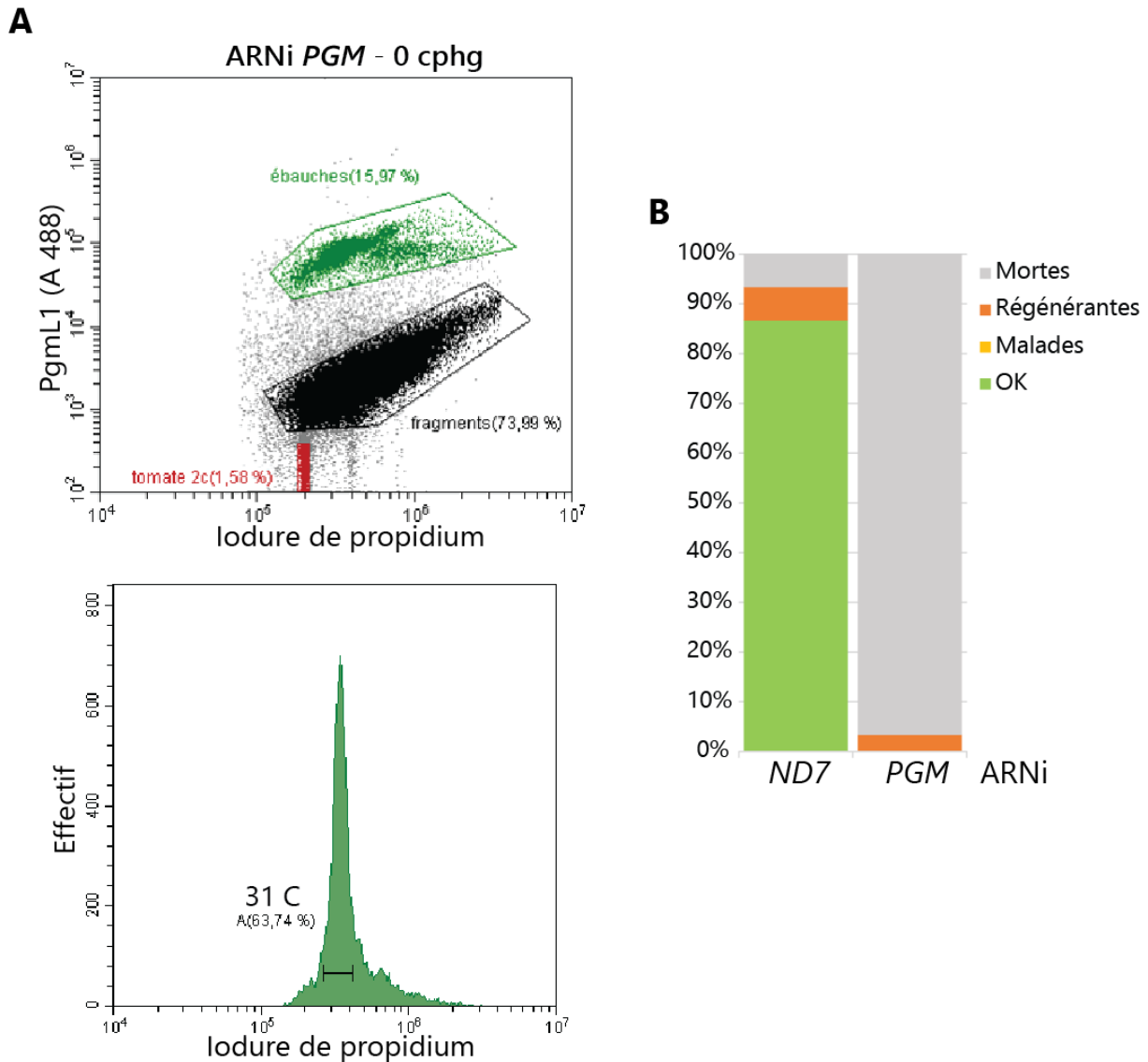


Figure 78 - Contrôle qualité des noyaux utilisés en Cut&Tag contre H3K4me3

A La qualité et la ploïdie des noyaux extraits à T6 sont contrôlées par flow cytométrie après marquage des noyaux par un anti-PgmL1 et un anticorps secondaire couplé à un fluorophore AF488. Les noyaux correspondants aux nouveaux MAC (ébauches) sont identifiés en vert (haut). Le contenu en ADN de ces noyaux est ensuite représenté pour évaluer la ploïdie (bas) **B** La qualité des cultures et du passage de l'autogamie est contrôlé par un test de survie.

Pour chaque condition d'anticorps un tube est utilisé pour être marqué au DAPI et monté entre lame et lamelle comme décrit précédemment. Pour les conditions avec pA-Tn5, un tube est marqué avec un anti-Flag puis un anti-souris AF-568 comme précédemment décrit.

Tagmentation : Les noyaux sont resuspendus dans 50 µL de tampon de tagmentation (composition section 3.4.5) et incubés 1 h à 37°C dans un bloque PCR. Les tubes sont ensuite rapidement centrifugés et placés sur le portoir aimanté. Une fois toutes les billes fixées, le surnageant est retiré et les noyaux sont resuspendus dans 50 µL de tampon TAPS-EDTA (composition section 3.4.6) pour arrêter la réaction. Après une courte centrifugation, les tubes sont à nouveau placés sur le portoir aimanté et le surnageant est retiré.

Pour chaque condition d'anticorps un tube est utilisé pour être marqué au DAPI et monté entre lame et lamelle comme décrit précédemment. Pour les conditions avec pA-Tn5, un tube est marqué avec un anti-Flag puis un anti-souris AF-568 comme précédemment décrit.

Lyse des noyaux : Les noyaux sont resuspendus dans 5 µL de tampon SDS (composition section 3.4.7) en humidifiant bien la paroi du tube de manière à en décoller les billes qui auraient pu rester accrochées. Après une rapide centrifugation, les noyaux sont incubés 1 h à 58°C dans un bloque PCR.

Pour chaque condition d'anticorps un tube est utilisé pour être marqué au DAPI et monté entre lame et lamelle comme décrit précédemment

3.3 CUT&TAG SUR H3K4ME3

Le protocole de Cut&Tag a été adapté à partir du protocole publié par l'équipe d'Henikoff et disponible sur le site protocol.io : [dx.doi.org/10.17504/protocols.io.bqwvmxe6](https://doi.org/10.17504/protocols.io.bqwvmxe6).

3.3.1 Matériel utilisé

Noyaux : Les noyaux utilisés sont issus de cellules non-transformées et cultivées en condition d'ARNi contre PGM. La qualité des noyaux a été vérifiée par cytométrie en flux (Figure 78). La densité en noyaux a été estimée à ~190 000

noyaux pour 20 μL . Un culot de 20 μL de noyaux est utilisé pour faire 4 conditions de Cut&Tag. Les noyaux sont décongelés à température ambiante et sont resuspendus dans 180 μL de tampon de lavage (composition section 3.4.1) et ajoutés sur chaque culot. Tous les noyaux sont mélangés avant d'être divisés en aliquots de 50 μL dans des tubes de 500 μL .

Transposase : La transposase commerciale de Diagénodé (C01070002, lot : 1/c) chargée avec les oligonucléotides (Tableau 15) en utilisant le protocole disponible sur le site du fournisseur (Picelli et al., 2014). La transposase est diluée au 1 :250^{ème} dans du tampon de lavage-300 (composition section 3.4.2).

Anticorps primaire : anti-H3K4me3 (39060, ActiveMotif, lot : 31420006) dilué au 1 :50^{ème} dans le tampon pour anticorps (composition section 3.4.3).

Anticorps secondaire : anti-lapin (ABIN101961, Antibodies-online, lot : NE-200-082001) dilué au 1 :100^{ème} dans le tampon de lavage (composition section 3.4.1).

3.3.2 Protocole

Sauf indication contraire, tout le protocole de Cut&Tag se passe à température ambiante et toutes les solutions utilisées sont conservées sur glace.

Préparation de billes : Les billes sont préparées de manière identique à ce qui a été décrit dans la section 3.2.3.

Fixation des noyaux sur billes : 4 μL de billes sont ajoutés dans chacun des tubes contenant 50 μL de noyaux. Les noyaux sont incubés avec les billes sur roue pendant 10 min. Les tubes sont ensuite centrifugés rapidement avant d'être placés sur le portoir aimanté. Une fois toutes les billes contre la paroi, le surnageant est retiré.

Anticorps primaire : Les noyaux liés aux billes sont resuspendus dans 50 μL de tampon pour anticorps contenant l'anticorps primaire. Pour chaque condition un contrôle négatif sans anticorps primaire est également réalisé. Les noyaux sont incubés sur roue pendant 2h puis sur roue toute la nuit à 4°C. Les tubes sont ensuite centrifugés rapidement et placés sur le portoir aimanté. Une fois toutes les billes contre la paroi, le surnageant est retiré.

Anticorps secondaire : L'incubation dans l'anticorps secondaire est identique à ce qui a été décrit dans la section 3.2.3.

Lavage : Après une rapide centrifugation, les tubes sont placés sur le portoir aimanté jusqu'à fixation des billes. Le surnageant est ensuite retiré et remplacé par 500 μ L de tampon de lavage sans perturber les billes. Le tampon de lavage est ensuite retiré.

Fixation de la pA-Tn5 : Les noyaux sont resuspendus dans 50 μ L de tampon de lavage-300 (composition section 3.4.2) avec ou sans pA-Tn5 et incubés 2h sur roue. Les billes sont ensuite lavées une fois comme décrit précédemment mais dans du tampon de lavage-300.

Tagmentation : Les noyaux sont resuspendus dans 50 μ L de tampon de tagmentation (composition section 3.4.5) et incubés 1h, 2h, 3h ou 24h à 37°C dans un bloque PCR. La suite de cette étape est identique à ce qui a été décrit dans la section 3.2.3.

Lyse des noyaux : La lyse des noyaux s'effectue comme ce qui a été décrit dans la section 3.2.3.

Amplification par PCR : Les 5 μ L de noyaux lysés sont resuspendus dans 15 μ L de solution de neutralisation (composition section 3.4.8) auxquels sont ajoutés 2 μ L de deux amorces de séquençage portant un code-barre (Tableau 16). On ajoute ensuite 25 μ L de mix PCR 2X (NEB-next non-hot star, M0541). La PCR a ensuite lieu dans un bloque PCR avec le programme suivant : 58°C 5 min, 72°C 5 min, 98°C 5 min suivie de 14 cycles de 98°C 10 sec, 60°C 10 sec pour terminer avec 1 min à 72°C. La PCR permet de former des banques qui pourront être séquencées directement après purification puisque les adaptateurs de séquençage sont intégrés lors de la PCR (Figure 79).

Purification de la banque : Après que les produits de PCR aient refroidi, 65 μ L de billes AMPure XP (A63880, Beckmann Coulter) y sont mélangées par pipetage. Après une rapide centrifugation les billes sont incubées 10 min. Les tubes sont ensuite placés sur le portoir aimanté jusqu'à ce que toutes les billes soient fixées. Le surnageant est éliminé. Puis les billes sont lavées deux fois avec 200 μ L d'éthanol 80% sans perturber les billes. Après le deuxième lavage les tubes sont laissés ouverts jusqu'à ce que l'éthanol restant s'évapore.

Élution : Les billes sont incubées pendant 1h dans 22 µL de Tris-HCl 10mM pH8. Les tubes sont ensuite placés sur le portoir aimanté jusqu'à ce que toutes les billes soient fixées. Le surnageant est transféré dans un tube Eppendorf propre et conservé à -20°C.

Conditions	Amorces i5	Amorces i7
Anti-Rb seul – 1h	i5.1 -> OMB1369	i7.1 -> OMB1379
Anti-H3K4me3 – 1h	i5.5 -> OMB1373	i7.5 -> OMB1383
Anti-Rb seul – 2h	i5.3 -> OMB1371	i7.3 -> OMB1381
Anti-H3K4me3 – 2h	i5.4 -> OMB1372	i7.4 -> OMB1382
Anti-Rb seul – 3h	i5.7 -> OMB1375	i7.7 -> OMB1385
Anti-H3K4me3 – 3h	i5.8 -> OMB1376	i7.8 -> OMB1386
Anti-Rb seul – ON	i5.9 -> OMB1377	i7.9 -> OMB1387
Anti-H3K4me3 – ON	i5.10 -> OMB1378	i7.10 -> OMB1388

Tableau 16 - Utilisation des amorces pour le Cut&Tag

Les conditions anti-Rb seules correspondent aux contrôles négatifs sans anticorps primaire. À la suite des anticorps utilisés sont indiqués les temps de tagmentation (ON = sur la nuit). La séquences des amorces utilisées est disponible dans le Tableau 33 en annexe.

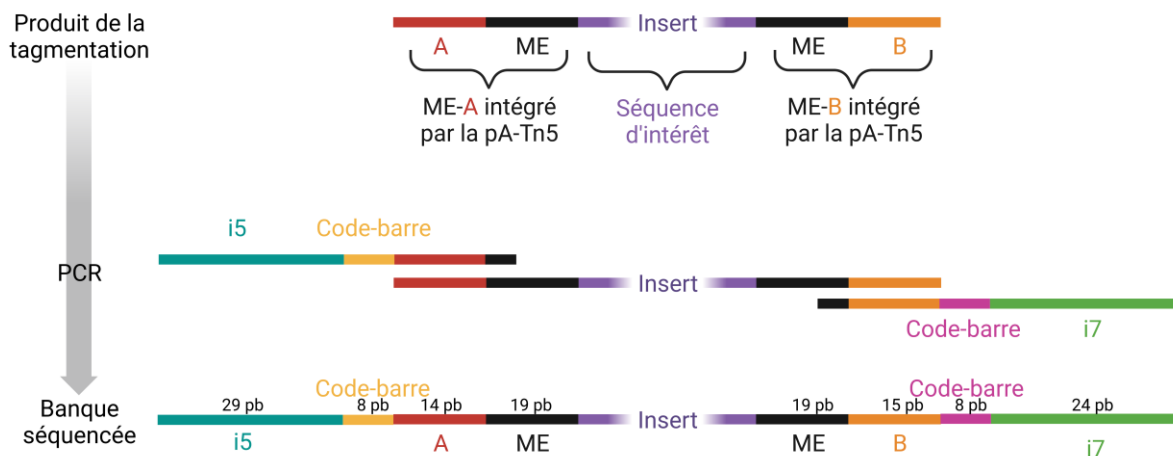


Figure 79 - Schéma de la formation des banques Cut&Tag

Après tagmentation par la pA-Tn5 au niveau de la protéine d'intérêt, les fragments d'ADN au niveau de cette protéine sont bordés par ME-A et ME-B précédemment chargés sur la pA-Tn5 (ME : Mosaic End). L'étape de PCR amplifie spécifiquement ces fragments grâce à des amorces qui contiennent les adaptateurs A ou B. La PCR permet également l'ajout de code-barre unique pour permettre de multiplexer les banques à séquencer et les adaptateurs de séquençage illumina i5 ou i7. La banque obtenue contient donc des fragments d'ADN qui peuvent être directement séquencés. L'ajout de ces séquences augmente la taille de l'insert de 136 pb. (Created with Biorender.com)

3.4 SOLUTIONS POUR CUT&TAG

3.4.1 Tampon de lavage

HEPES-KOH 20 mM pH7,5, NaCl 150 mM, Spermidine 0,5 mM

HEPES-KOH 1 M pH7,5	500 μ L
NaCl 5 M	750 μ L
Spermidine 2 M	6,25 μ L
Eau MilliQ	Qsp 25 mL
Roche Complete Protease inhibitor EDTA-free	1 pastille

La solution peut être conservée à 4°C sur 2 jours. La spermidine permet de maintenir l'intégrité des noyaux dans la solution.

3.4.2 Tampon de lavage-300

HEPES-KOH 20 mM pH7,5, NaCl 300 mM, Spermidine 0,5 mM

HEPES-KOH 1 M pH7,5	500 μ L
NaCl 5 M	1,5 mL
Spermidine 2 M	6,25 μ L
Eau MilliQ	Qsp 25 mL
Roche Complete Protease inhibitor EDTA-free	1 pastille

La solution peut être conservée à 4°C sur 2 jours. La spermidine permet de maintenir l'intégrité des noyaux dans la solution.

3.4.3 Tampon pour anticorps

HEPES-KOH 20 mM pH7,5, NaCl 150 mM, Spermidine 0,5 mM, BSA 0,1%

Le tampon pour anticorps est fait le jour de son utilisation en ajoutant 10 μ L de BSA 100X (NEB) dans 1 mL de tampon de lavage. Il est ensuite conservé sur glace.

3.4.4 Tampon de fixation

HEPES-KOH 20 mM pH7,9, KCl 100 mM, CaCl₂ 1 mM, MnCl₂ 1 mM

HEPES-KOH 1 M pH7,9	200 µL
KCl 1 M	100 µL
CaCl ₂ 1 M	10 µL
MnCl ₂ 1 M	10 µL
Eau MilliQ	Qsp 25 mL

La solution peut être conservée à 4°C sur quelque mois.

3.4.5 Tampon de tagmentation

HEPES-KOH 20 mM pH7,5, NaCl 300 mM, Spermidine 0,5 mM ; MgCl₂ 10mM

Le tampon de tagmentation est fait le jour de son utilisation en ajoutant 10 µL de MgCl₂ dans 1 mL de tampon de lavage-300. Il est ensuite conservé sur glace.

3.4.6 Tampon TAPS-EDTA

TAPS 10 mM pH8,5, EDTA 0,2 mM

TAPS 1 M pH8,5	10 µL
EDTA 500 mM pH8	0,4 µL
Eau MilliQ	1 mL

Le tampon de tagmentation est fait le jour de son utilisation.

3.4.7 Tampon SDS

TAPS 10 mM pH8,5, SDS 0,1%

TAPS 1 M pH8,5	10 µL
SDS 20%	5 µL
Eau MilliQ	1 mL

Le tampon de tagmentation est fait le jour de son utilisation.

3.4.8 Solution de neutralisation

Triton-X100 0,67%

Triton-X100	67 µL
Eau MilliQ	933 µL

Le tampon de tagmentation est fait le jour de son utilisation.

3.5 TEST DE LA QUALITÉ DES BANQUE DE CUT&TAG

Quantification des banques : La quantité d'ADN obtenue à la suite du Cut&Tag est mesurée au Qbit HS (ThermoFisher) à partir de 2 μ L de banque.

Qualité de banque : La qualité des banques est vérifiée par électrophorèse en capillaire à l'aide de cassette ADN D1000 (Agilent) et des réactifs BR correspondant puis migrée dans un appareil de Tape Station 4150 (Agilent).

Amplification des banques : Pour simuler l'amplification de la banque sur le séquenceur, on peut faire une PCR semi-quantitative, sur les banques à tester, et sur une banque de référence qui sert de témoin positif. Le principe est de réaliser un test d'amplification avec une petite quantité connue de chaque banque. On utilise une Taq-polymérase (ThermoFisher) de base pour cette amplification.

Pour chaque banque, et pour la référence on prépare un mix contenant :

Banque ADN	0,25 g
Tampon 5x Taq-polymérase	5 μ L
Oligo C 10 μ M	0,5 μ L
Oligo D 10 μ M	0,5 μ L
dNTP 10 mM	0,75 μ L
Taq-polymérase	0,25 μ L
Eau MilliQ	Qsp 25 μ L

La séquence des oligonucléotides utilisée est présentée dans le Tableau 17.

Les mix PCR sont incubés dans un bloque PCR avec le programme suivant : 98°C 30 sec suivie de 10 cycles : 98°C 40 sec, 60°C 30 sec, 72°C 30 sec puis 5 min à 72°C pour terminer l'élongation.

Après ajout du tampon de charge, 10 μ L de produit de PCR sont déposés sur un gel d'agarose 1,5% pour vérifier l'homogénéité des tailles et de l'amplification des banques testées. On vérifiera également que la banque

s'amplifie à un niveau similaire à la banque de référence en mesurant la quantité d'ADN obtenue après PCR au Qbit HS.

Identifiant	Nom	Séquence (5' → 3')
OMB1366	Oligo C	ATGATACGGCGACCACCGAGATCTACAC
OMB1367	Oligo D	CAAGCAGAAGACGGCATACGAGAT

Tableau 17 - Oligonucléotides utilisés pour la PCR semi-quantitative

Nom de la cinétique	ARNi contre...	Temps de récoltes	Référence
ICL7	<i>ICL7</i>	Veg, T0, T5, T10, T20, T35, T50	(Lhuillier-Akakpo et al., 2014)
EZL1	<i>EZL1</i>	Veg, T0, T5, T10, T20, T35, T50	
ND7_K	<i>ND7</i>	Veg, T0, T5, T10, T20, T30, T40	(Marmignon et al., 2014)
PGM	<i>PGM</i>	Veg, T0, T5, T10, T20, T30, T40	
KU80c	<i>KU80c</i>	Veg, T0, T5, T10, T20, T30, T40	
ND7_X	<i>ND7</i>	Veg, T2, T7, T22, T32, T42	(Kapusta et al., 2011)
XRCC4	<i>XRCC4</i>	Veg, T2, T7, T22, T32, T42	
ND7_L	<i>ND7</i>	Veg, T0, T3.5, T6, T20, T60	
CTIP	<i>CtIP</i>	Veg, T0, T5.5, T12.5, T25	(Godau et al., 2019)

Tableau 18 - Références des ARN prélevés au cours des cinétiques d'autogamie
Veg correspond à des cellules prélevées avant le début de l'autogamie. Les temps de récolte (T) sont indiqués en heure après le T0 (50% de cellules autogames).

Nom de la cinétique	Séquenceur utilisé	Temps de récoltes séquencés	Référence N° d'accension
ICL7	HiSeq	Veg, T0, T5, T10, T20, T35, T50	(Frapporti et al., 2019) PRJEB21555 (ENA)
ICL7	NextSeq	T0, T10, T35, T50	(Miró-Pina et al., 2022) PRJEB4660 (ENA)
EZL1	HiSeq	Veg, T0, T5, T10, T20, T35, T50	(Frapporti et al., 2019) PRJEB21555 (ENA)
EZL1	NextSeq	T0, T10, T35, T50	(Miró-Pina et al., 2022) PRJEB4660 (ENA)
ND7_K	HiSeq	Veg, T0, T5, T10, T20, T30, T40	Non-publié
KU80c	HiSeq	Veg, T0, T5, T10, T20, T30, T40	Non-publié
PGM	HiSeq	Veg, T2, T5, T10, T20, T30, T40	Non-publié
ND7_X	NextSeq	Veg, T2, T7, T22, T32, T42	Non-publié
XRCC4	NextSeq	Veg, T2, T7, T22, T32, T42	Non-publié
ND7_L	NextSeq	Veg, T0, T3.5, T6, T20, T60	Non-publié
CTIP	NextSeq	Veg, T0, T5.5, T12.5, T25	Non-publié

Tableau 19 - Références des séquençages ARN utilisés

Chapitre 4 - Analyse bio-informatique

Lors de ma thèse j'ai effectué les analyses à partir des tableaux de comptage des gènes dans chacune des conditions. Sauf indication contraire, toutes les analyses ont été effectuées sur R.4.1.2.

4.1 DONNÉES ARN UTILISÉES

4.1.1 Origines des ARN utilisés

Les analyses transcriptomiques ont été réalisées sur le séquençage d'ARN prélevé lors de cinétiques d'autogamie déjà publiées (Tableau 18).

4.1.2 Séquençage des ARN

Les séquençages ARN utilisés lors de ma thèse proviennent de sources variées (Tableau 19). Les cinétiques ICL7 et EZL1 utilisées lors de ma thèse ont été extraites de données publiées (Frapporti et al., 2019; Miró-Pina et al., 2022). Un séquençage 2x100 bp *paired-end* à partir de banques Illumina TruSeq sur un séquenceur HiSeq comme décrit dans (Arnaiz et al., 2012) avaient déjà été effectué avant le début de ma thèse pour les cinétiques d'ARN polyA extrait des cinétiques PGM, KU80c et ND7_X. Enfin, de nouveaux séquençages ont été réalisés pour ma thèse, avec le séquençage 2x75 bp *paired-end* des cinétiques ND7_X, XRCC4, ND7_L et CTIP sur un séquenceur NextSeq comme décrit dans (Miró-Pina et al., 2022).

4.2 GÉNOMES DE RÉFÉRENCE

Les séquençages ARN ont été mappés sur le génome de *P. tetraurelia* 51 MAC v1 (ptetraurelia_mac_51.fa) et MAC+IES v1 (ptetraurelia_mac_51_with_ies.fa) (Arnaiz et al., 2012). L'annotation des gènes utilisé lors de ma thèse est la version 2.0 (ptetraurelia_mac_51_annotation_v2.0.gff3) et l'annotation des IES est la version 1 (internal_eliminated_sequence_PGM_ParTIES.pt_51.gff3) (Arnaiz et al., 2012). Toutes ces annotations sont accessibles sur : <https://paramecium.i2bc.paris-saclay.fr/cgi/browse/jbrowse>.

4.3 ANALYSE D'EXPRESSION DIFFÉRENTIELLE DES GÈNES

Pour chacun des échantillons, le nombre de fragments correspondant à chacun des gènes a été déterminé par htseq-count sur les fichiers BAM filtrés par samtools, comme décrit dans (Arnaiz et al., 2017).

De plus, afin de pouvoir faire l'analyse des gènes différentiellement exprimés dans les différentes conditions, j'ai fait le tableau des métadonnées pour les échantillons de l'étude (Tableau 27 en annexe, un extrait contenant uniquement les données pour les cinétiques ICL7 est présenté dans le Tableau 20) chaque ligne du tableau correspond à un échantillon séquencé unique.

Names	Samples	Knock Down	Timing	Autog Stage	Condition	Seq_method	Labo
ICL7_Veg	ICL7_Veg	ctrl	Veg	VEG	VEG_ctrl	HiSeq	Duharcourt
ICL7_T0	ICL7_T0	ctrl	0	EARLY	EARLY_ctrl	HiSeq	Duharcourt
ICL7_T5	ICL7_T5	ctrl	5	INTER	INTER_ctrl	HiSeq	Duharcourt
ICL7_T10	ICL7_T10	ctrl	10	INTER	INTER_ctrl	HiSeq	Duharcourt
ICL7_T20	ICL7_T20	ctrl	20	LATE	LATE_ctrl	HiSeq	Duharcourt
ICL7_T35	ICL7_T35	ctrl	35	LATE	LATE_ctrl	HiSeq	Duharcourt
ICL7_T50	ICL7_T50	ctrl	50	LATE	LATE_ctrl	HiSeq	Duharcourt
ICL7bis_T0	ICL7_T0	ctrl	0	EARLY	EARLY_ctrl	NextSeq	Duharcourt
ICL7bis_T10	ICL7_T10	ctrl	10	INTER	INTER_ctrl	NextSeq	Duharcourt
ICL7bis_T35	ICL7_T35	ctrl	35	LATE	LATE_ctrl	NextSeq	Duharcourt
ICL7bis_T50	ICL7_T50	ctrl	50	LATE	LATE_ctrl	NextSeq	Duharcourt

Tableau 20 - Extrait du Tableau 27 contenant les métadonnées utilisées pour l'analyse des gènes différentiellement exprimés

4.3.1 Mise en forme des données de comptage

À cause des différences dans les méthodes de séquençages utilisées pour les différentes cinétiques (HiSeq ou NextSeq), les données de comptages sont corrigées pour un effet *batch* avant de pouvoir faire l'analyse des gènes différentiellement exprimés.

4.3.1.1 Correction des données par ComBat-seq

Les données ont été corrigées en utilisant la fonction ComBat-seq (bibliothèque R : sva v3.38.0) (Zhang et al., 2020). La fonction a été utilisée de la manière suivante :

```
ComBat_seq(countdata, batch = batch)
```

La variable *countdata* correspond au tableau regroupant les comptages pour chacun des gènes pour tous les points des cinétiques étudiées ici. La variable *batch* correspond aux séquenceurs utilisés et au laboratoire d'origine. Il est la fusion des deux dernières colonnes du tableau des métadonnées pour chacune des cinétiques (exemple : 'HiSeq_Duharcourt' ou 'NextSeq_Duharcourt'). Aucune information sur les groupes n'a été fournie pour les corrections.

4.3.1.2 Normalisation des données

Toutes les données utilisées pour représenter graphiquement les données de comptage sont effectuées sur les comptages après normalisation VST (*Variance Stabilizing Transformation*). Cette normalisation est extraite de l'objet *DESeqTransform* générée par la fonction DESeq2 (bibliothèque R : DESeq2 v1.30.1) (Love et al., 2014). Les données ont été normalisées de la manière suivante :

```
deseq = DESeqDataSetFromMatrix(countData = countdata,
                                colData = infodata, design = ~ Samples)
deseq = DESeq(deseq)
data_tab = assay(vst(deseq, blind = TRUE))
```

Les deux premières lignes servent à créer de l'objet *DESeqTransform* à partir des tableaux de comptage *countdata* corrigés ou non par ComBat-seq et des métadonnées contenues dans *infodata* (Tableau 21). De cette manière l'analyse

Names	Samples	Knockdown	Timing	Autog Stage	Condition	Seq_ method	Labo
EZL1_Veg	EZL1_Veg	EZL1	Veg	VEG	VEG_EZL1	HiSeq	Duharcourt
EZL1_T0	EZL1_T0	EZL1	0	EARLY	EARLY_EZL1	HiSeq	Duharcourt
EZL1_T5	EZL1_T5	EZL1	5	INTER	INTER_EZL1	HiSeq	Duharcourt
EZL1_T10	EZL1_T10	EZL1	10	INTER	INTER_EZL1	HiSeq	Duharcourt
EZL1_T20	EZL1_T20	EZL1	20	LATE	LATE_EZL1	HiSeq	Duharcourt
EZL1_T35	EZL1_T35	EZL1	35	LATE	LATE_EZL1	HiSeq	Duharcourt
EZL1_T50	EZL1_T50	EZL1	50	LATE	LATE_EZL1	HiSeq	Duharcourt
EZL1bis_T0	EZL1_T0	EZL1	0	EARLY	EARLY_EZL1	NextSeq	Duharcourt
EZL1bis_T10	EZL1_T10	EZL1	10	INTER	INTER_EZL1	NextSeq	Duharcourt
EZL1bis_T35	EZL1_T35	EZL1	35	LATE	LATE_EZL1	NextSeq	Duharcourt
EZL1bis_T50	EZL1_T50	EZL1	50	LATE	LATE_EZL1	NextSeq	Duharcourt
ICL7_Veg	ICL7_Veg	ctrl	Veg	VEG	VEG_ctrl	HiSeq	Duharcourt
ICL7_T0	ICL7_T0	ctrl	0	EARLY	EARLY_ctrl	HiSeq	Duharcourt
ICL7_T5	ICL7_T5	ctrl	5	INTER	INTER_ctrl	HiSeq	Duharcourt
ICL7_T10	ICL7_T10	ctrl	10	INTER	INTER_ctrl	HiSeq	Duharcourt
ICL7_T20	ICL7_T20	ctrl	20	LATE	LATE_ctrl	HiSeq	Duharcourt
ICL7_T35	ICL7_T35	ctrl	35	LATE	LATE_ctrl	HiSeq	Duharcourt
ICL7_T50	ICL7_T50	ctrl	50	LATE	LATE_ctrl	HiSeq	Duharcourt
ICL7bis_T0	ICL7_T0	ctrl	0	EARLY	EARLY_ctrl	NextSeq	Duharcourt
ICL7bis_T10	ICL7_T10	ctrl	10	INTER	INTER_ctrl	NextSeq	Duharcourt
ICL7bis_T35	ICL7_T35	ctrl	35	LATE	LATE_ctrl	NextSeq	Duharcourt
ICL7bis_T50	ICL7_T50	ctrl	50	LATE	LATE_ctrl	NextSeq	Duharcourt

Tableau 21 - Métadonnées des échantillons utilisés pour valider la correction par ComBat-Seq

DESeq2 n'est biaisée par aucune particularité des échantillons car ils sont tous uniques pour la fonction lorsque *Samples* est utilisé comme 'design'. Enfin, la fonction 'vst()' (librairie R : *DESeq2* v1.30.1) permet d'extraire les données normalisée VST. Préciser 'blind = *TRUE*' permet de s'assurer que les données normalisées ne tiendront pas compte des informations dans la variable 'design'. La fonction 'assay()' (librairie R : *DESeq2* v1.30.1) permet d'extraire les données sous forme de tableau.

4.3.1.3 Comparaison et visualisation des données corrigées et non corrigées

La correction par ComBat-seq a d'abord été appliquée uniquement sur les données des cinétique EZL1 et ICL7 qui ont été séquencées par les deux méthodes de séquençage (métadonnées correspondantes : Tableau 21). L'efficacité de la correction de l'effet *batch* a été validée en comparant les données corrigées et non-corrigées normalisée par VST.

4.3.1.3.1 Analyse en Composantes Principales (ACP)

Les données sont d'abord comparées par ACP. Les ACP sont générés à partir des données corrigées ou non sont comparées à l'aide de la fonction PCA (librairie R : *FactoMiner* v2.4). Les pourcentages de variance capturée par les trois premières dimensions sont représentés par un histogramme en barre en utilisant la fonction 'barplot()' (librairie R : *graphics* v4.1.2). Les différentes combinatoires des trois premières dimensions sont ensuite représentées en utilisant la fonction 'ggplot()' (librairie R : *ggplot2* v3.3.6).

4.3.1.3.2 Matrice de distance de Pearson

La matrice de corrélation de Pearson est générée en utilisant la fonction en utilisant la fonction 'cor()' (librairies R : *stats* v4.1.2). Elle est ensuite transformée en matrice de distance par la fonction 'as.dist()' de la même librairie. Les fonctions sont utilisées de manière suivante :

```
matDist = as.dist(1-cor(log2(data_tab+1), method="pearson"))
```

La variable *data_tab*, correspond aux tableaux de données corrigées ou non et normalisées par VST (comme indiqué dans la section 4.3.1.2).

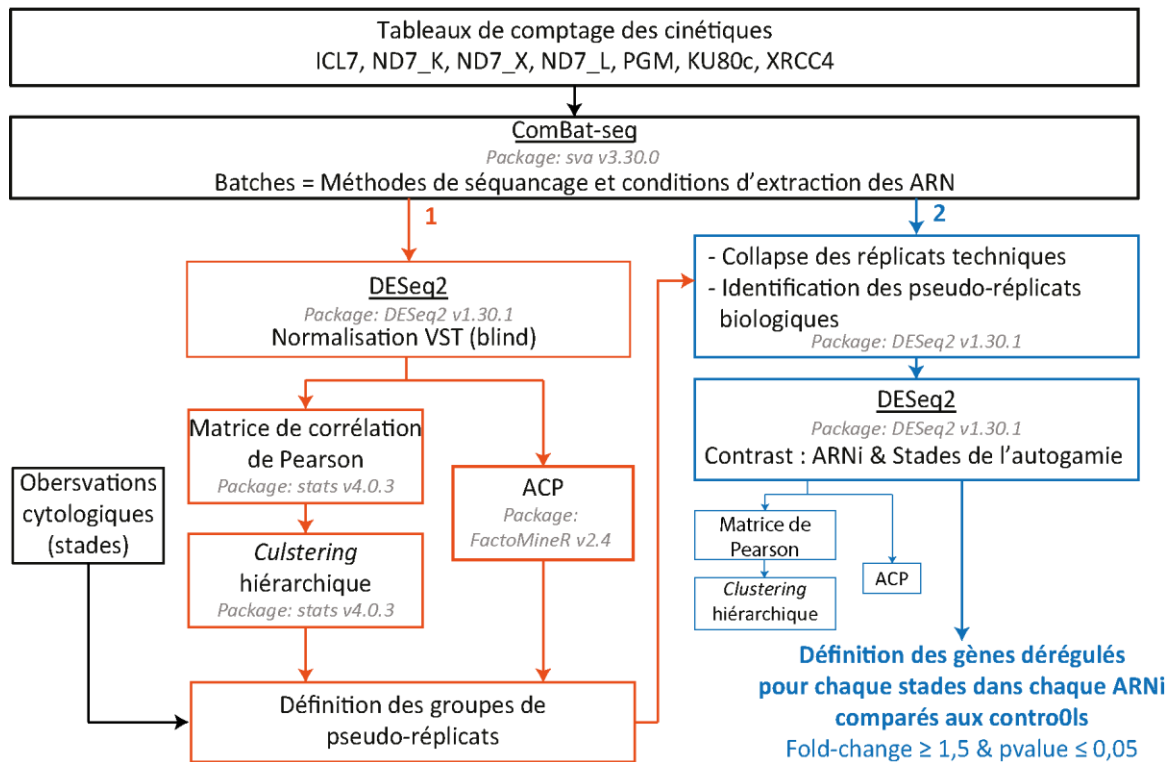


Figure 80 - Définition des pseudo-réplicats puis des gènes différentiellement exprimés

Les encadrés en noir correspondent aux données qui ont été utilisées pour l'analyse. La voie 1, en orange, représente les étapes qui ont permis la définition des groupes de pseudo-réplicats. Ces pseudo-réplicats ont ensuite été utilisés lors de la définition des gènes différentiellement exprimée (voie 2, en bleu).

4.3.1.3.3 *Clustering* hiérarchique

La matrice de distance (*matDist*) est ensuite traitée par la fonction 'hclust()' (bibliothèque R: *stats* v4.1.2). Pour faire le clustering hiérarchique qui est ensuite représenté sous forme d'arbre par la fonction 'as.dendrogram()' de la même bibliothèque.

```
res = hclust(matDist)
res = as.dendrogram(res)
```

4.3.2 Définition des pseudo-réplicats

Les ACP et les *clustering* hiérarchiques générés à partir des données de comptages des cinétiques d'intérêt (Tableau 27), corrigées par ComBat-seq pour le biais de méthode de séquençage, sont ensuite utilisés afin de définir des pseudo-réplicats (Figure 80). Ces pseudo-réplicats correspondent à des points proches dans les différentes cinétiques qui seront considérés comme des réplicats biologiques. De cette manière, j'ai défini des temps : VEG = végétatif, ERLAY = précoce, INTER = intermédiaire et LATE = tardifs à partir des cinétiques en heure.

Les groupes de pseudo-réplicats ainsi obtenus ont ensuite été affinés manuellement en s'appuyant sur les données biologiques (Figure 81). En effet, pour chaque cinétique d'autogamie utilisée, la progression au cours de l'autogamie a été suivie permettant d'évaluer la proportion de chaque stade aux différents temps. Cette évaluation est faite manuellement en observant au microscope une centaine de cellules marquée au DAPI. Les cellules sont ensuite classées dans les différents stades en fonction de l'état de leurs noyaux (protocole section 2.10).

Les groupes temps des cinétiques sont indiqués dans la colonne *AutogStage* du tableau des métadonnées utilisées dans l'analyse (Tableau 27, extrait Tableau 20).

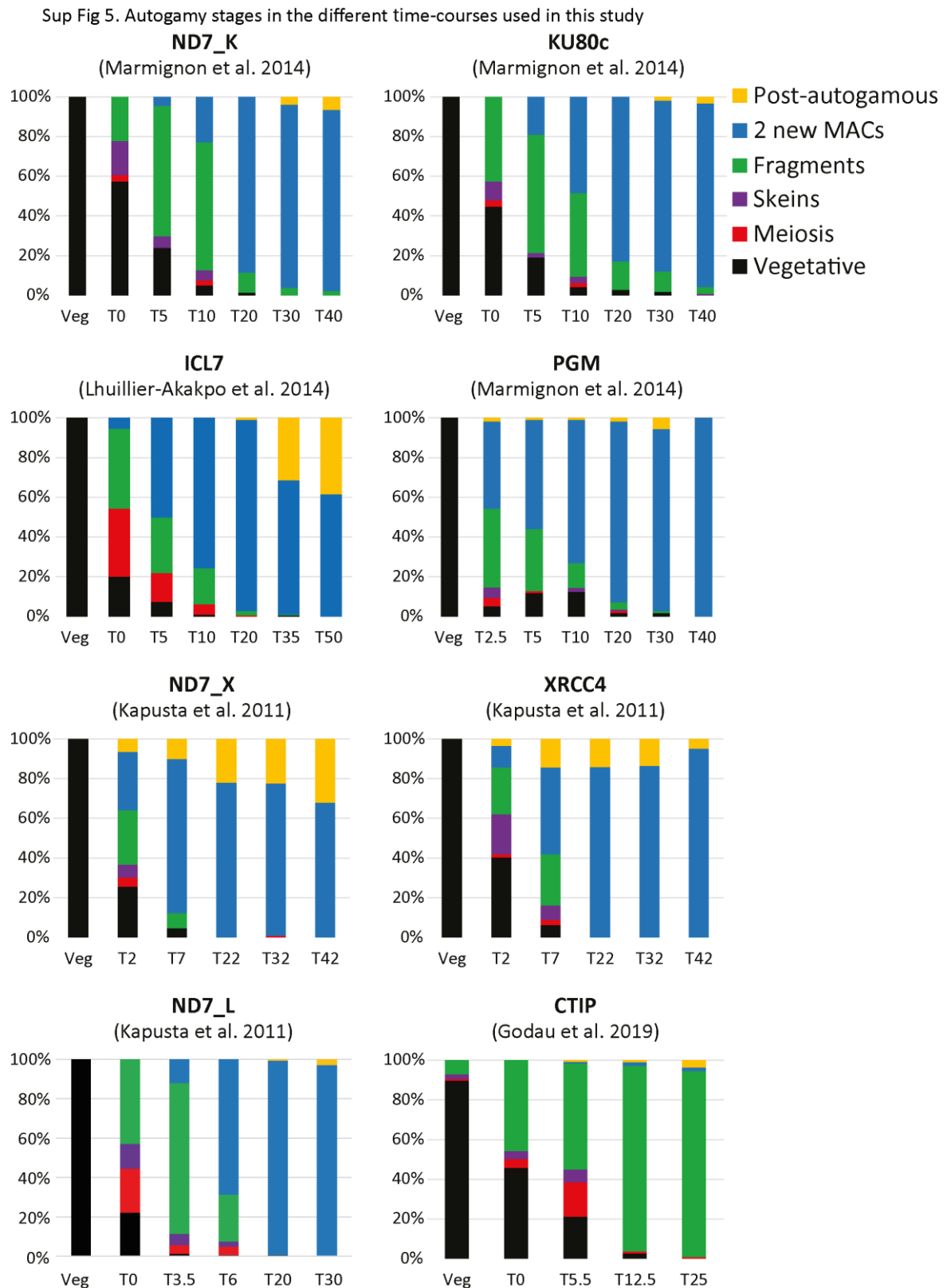


Figure 81 - Récapitulatif des stades publiés pour les différentes cinématiques utilisées

Le temps de récolte est indiqué en heure suivant le T0.

4.3.3 Identification des gènes dérégulés par DESeq2

L'identification des gènes différentiellement exprimé passe par une étape de mise en forme des données puis d'analyse à laquelle j'ai ajouté des étapes permettant de représenter graphiquement les données (Figure 80).

4.3.3.1 Mise en forme des données

Dans un premier temps, afin de définir les points qui seront considérés comme des répliquats biologiques lors de l'analyse, j'ai utilisé la colonne *Condition* du tableau des métadonnées. L'analyse (Tableau 27, extrait Tableau 20) qui fusionne les informations des groupes de pseudo-répliquats indiqués dans la colonne *AutogStage* avec celle des vrais répliquats biologiques indiqués dans la colonne *KnockDown*. Ces informations sont transformées en objet *DESeq* par la fonction 'DESeqDataSetFromMatrix()' (bibliothèque R : *DESeq2* v1.30.1) de la manière suivante :

```
deseq = DESeqDataSetFromMatrix(countData = countdata,
                                colData = infodata, design = ~ Condition)
```

La fonction utilise le tableau de comptage *countdata* corrigée par ComBat-seq et des métadonnées contenues dans *infodata* l'analyse (Tableau 27, extrait Tableau 20). On attribue à la variable 'design' les données présentes dans la colonne *Condition* du tableau des métadonnées. Elle sert à définir les points de comparaison des cinétiques et les répliquats.

Avant l'identification des gènes dérégulés, les temps de la cinétique ICL7 qui ont été séquencés par les deux méthodes sont identifiées comme des répliquats techniques par la fonction 'collapseReplicates()' (bibliothèque R : *DESeq2* v1.30.1) de la manière suivante :

```
deseq = collapseReplicates(deseq, groupby = deseq$Samples,
                            run = deseq$Names)
```

La fonction corrige le tableau des métadonnées contenu dans l'objet *deseq*, créée précédemment, en y ajoutant une colonne *runsCollapsed* qui contient la liste des échantillons qui doivent être considérés comme des répliquats techniques.

Dans le tableau des métadonnées, les colonnes *Samples* et *Names* ne diffèrent que par la présence de la mention « bis » dans les noms des échantillons séquencés une deuxième fois. Dans la colonne *Samples* les échantillons séquencés plusieurs fois ont donc le même nom. On indique pour *groupby* le nom des échantillons d'où proviennent les séquences et dans *run* le nom du séquençage.

4.3.3.2 Analyse

Une fois les données mises en forme, elles sont analysées par la fonction 'DESeq()' (bibliothèque R : *DESeq2* v1.30.1) simplement de la manière suivante :

```
deseq = DESeq(deseq)
```

4.3.3.3 Extraction des résultats

Les résultats pertinents de l'analyse sont ensuite extraits de manière à comparer chaque ARNi aux cinétiques contrôle pour les quatre périodes définies précédemment (VEG, EARLY, INTER et LATE). Pour cela on utilise la fonction 'result()' (bibliothèque R : *DESeq2* v1.30.1) de la manière suivante :

```
resContrast=results(deseq, contrast=c("Condition", kd, ctl))
```

La fonction est utilisée autant de fois qu'il y a de conditions à comparer pour extraire les données de l'analyse stockée dans l'objet *deseq*. Les variables *ctl* et *kd* utilisées varient d'une comparaison à l'autre. Elles prennent tour à tour les noms indiqués dans la colonne *Condition* du tableau des métadonnées (Tableau 27, extrait Tableau 20). Les variables *ctl* et *kd* permettent de comparer les différents éléments. Par exemple, définir les gènes différentiellement exprimés au temps tardif dans un ARNi *PGM*, on *ctl* sera égale à *LATE_ctl* et *kd* à *LATE_PGM*.

Une fois le tableau extrait, on y applique des seuils de différence (*Fold change, FC*) et de p-value afin d'identifier les gènes dont l'expression diffère significativement du contrôle. Dans cette analyse, j'ai utilisé par défaut un $FC \geq 2$ et $p\text{-value} < 0,05$. Les seuils sont définis de la manière suivante :

```
FC = 2
p-value = 0,05
```

```
res = resContrast[resContrast$padj < p-value ,]
res = res[res$log2FoldChange >= log2(FC) |
          res$log2FoldChange <= log2(1/FC),]
```

Les variables *FC* et *p-value* permettent de définir les seuils utilisés ensuite pour sélectionner uniquement les gènes dont la p-value corrigée (*padj*) est inférieure au seuil. Puis les gènes pour lesquels le *fold change* (positive ou négative) dépasse le seuil fixé. Si le *log2FoldChange* est ≥ 1 alors le gène sera considéré comme surexprimé. S'il est ≤ -1 alors l'expression du gène est considérée comme étant réduite par rapport au contrôle.

Le *fold change* et la p-value corrigée des gènes sont représentés par un volcanoplot en utilisant la fonction 'plot()' (bibliothèque R : *graphics* v4.1.2) avec les *log2FoldChange* en abscisse et $-\log(\text{padj})$ en ordonnée.

4.4 CARACTÉRISATION DES GÈNES DÉRÉGULÉS

4.4.1 Définition d'une liste de candidats sous R

Une fois la liste des gènes significativement dérégulés dans les différentes conditions établies, il reste à établir la liste des gènes candidats. Pour cela, j'ai utilisé R afin de déterminer quels sont les gènes à l'intersection des listes d'intérêt : surexprimé au temps tardif (LATE) dans les ARNi *PGM* ; *KU80c* et *XRRC4* et ayant une expression réduite lors les temps précoce (EARLY) et intermédiaire (INTER) lors d'un ARNi *CTIP*. J'ai représenté l'intersection de ces conditions par un digramme de Venn proportionnel en utilisant les fonctions 'draw.pairwise.venn()' et 'draw.triple.venn()' (bibliothèque R : *VennDiagram* v1.7.3).

4.4.2 Caractérisation des candidats sous R

4.4.2.1 Profils d'expressions

Différentes caractéristiques de la liste de gènes candidats ont été testées. En premier lieu, j'ai regardé par des diagrammes empilés la distribution des différents groupes de profils d'expression des gènes. Pour cela j'ai utilisé les

groupes prédéfinis (Arnaiz et al., 2017) et j'ai regardé la représentation des groupes parmi tous les gènes et parmi mes gènes d'intérêt afin de les comparer.

J'ai ensuite réalisé un test d'enrichissement (Boyle et al., 2004) pour définir si la différence de proportion par rapport à ce qui est observé pour l'ensemble des gènes est significative. Pour cela j'ai d'abord calculé la p-value expérimentale en utilisant la loi hypergéométrique ; en utilisant la fonction 'phyper()' (bibliothèque R : *stats* v4.1.2). Afin de limiter les faux positifs dû aux tests multiples, la p-value est ensuite corrigée par le résultat de 1 000 simulations : J'ai comparé les p-values obtenue pour 1 000 listes de gènes sélectionnés aléatoirement comprenant autant de gènes que ma liste d'intérêt. La p-value ajustée pour chaque profil d'expression correspond à la fraction de p-value simulée inférieure à la p-value expérimentale.

4.4.2.2 Présence d'IES dans la séquence codante des gènes

De la même manière, j'ai utilisé l'annotation des IES dans la séquence codante des gènes (Arnaiz et al., 2012) afin de comparer la proportion de gènes possédant une IES dans mes listes de gènes d'intérêt à celle dans tout le génome. J'ai effectué un test Chi2 afin d'évaluer si la différence observée était significative ou non en utilisant la fonction 'chisq.test()' (bibliothèque R : *stats* v4.1.2).

4.4.2.3 Ontologie des gènes

Enfin, j'ai regardé si certaines fonctions étaient enrichies dans ma liste de gènes candidats. Pour cela j'ai commencé par faire une représentation en histogramme des différents GOtermes pour les fonctions moléculaires en utilisant ggplot (bibliothèque R : *ggplot2* v3.3.6). J'ai ensuite établi pour quel fonction l'enrichissement était significatif utilisant la bibliothèque R *topGO* v2.46.0 et les GOtherms disponibles sur <https://paramecium.i2bc.paris-saclay.fr/>. Je me suis intéressée plus particulièrement aux *Molecular function (MF)*.

```
GOdata = new("topGOdata", ontology = "MF", allGenes = geneList,
            annot = annFUN.gene2GO, gene2GO = geneID2GO)

resultFisher = runTest(GOdata, algorithm = "classic", statistic =
            "fisher")
```

```
results.table = GenTable(GOdata, classicFisher = resultFisher ,
  topNodes = length(resultFisher@score))
```

La fonction 'new()' de la librairie permet de créer un objet *topGOdata* contenant la liste des gènes d'intérêt (*geneList*) et les termes GO pour tous les gènes de la paramécie (*geneID2GO*). L'objet est ensuite pris en charge par la fonction 'runTest()' qui va effectuer le test statistique de Fisher et déterminer quels sont les fonctions significativement sur-représentées dans ma liste de gènes d'intérêt. Enfin, les données sont extraites sous forme de tableau par la fonction 'GeneTable()'.

4.5 IDENTIFICATION D'UN MOTIF CONSERVÉ

J'ai cherché si un motif était enrichi dans les promoteurs de mes 112 gènes d'intérêts : surexprimés tardivement lors d'ARNi *PGM*, *KU80c* et *XRCC4*, dont l'expression diminue en condition d'ARNi contre *CtIP* et faisant parti des gènes *intermediate peak*. Les promoteurs de la paramécie n'étant pas encore clairement définis, j'ai considéré les 150 pb en amont du TSS des gènes comme étant des promoteurs (Arnaiz et al., 2017).

4.5.1 Identification du motif par STREME

J'ai utilisé le logiciel en ligne STREME 5.4.1 (Bailey, 2021) de la suite MEME afin de chercher un motif de 5 à 20 pb enrichi dans les promoteurs de mes 112 gènes d'intérêt par rapport à cinq séries de 112 promoteurs de gènes choisis aléatoirement (p-value > 0,05). Pour chacune des cinq recherches STREME, le fichier '.meme' correspondant au premier motif est ensuite extrait.

Les cinq motifs sont ensuite alignés et rognés à la taille du motif le plus petit en utilisant la fonction 'trim_motifs()' (librairie R : universalmotif v1.12.3) sur la variable *motifs* correspondant à la liste des cinq fichiers '.meme' des motifs identifiés.

```
motifs_t = trim_motifs(motifs, min.ic = 1, trim.from = "both")
```

La fonction 'merge_motif()' (bibliothèque R : universalmotif v1.12.3) permet ensuite de fusionner les cinq motifs (`motif_t`) afin d'extraire une nouvelle séquence consensus (`MergeMotif`). Le complémentaire de cette séquence (`Merge_rev`) est calculé par la fonction 'motif_rc()' de la même bibliothèque. Enfin, le logo correspondant à la séquence consensus, et son complémentaire, sont générés en utilisant la fonction 'view_motifs()' de la même bibliothèque :

```
MergeMotif = merge_motifs(motifs_t)
view_motifs(MergeMotif, show.positions = F)
Merge_rev = motif_rc(MergeMotif)
view_motifs(Merge_rev, show.positions = F)
```

Le motif consensus (`MergeMotif`) est ensuite sauvegardé sous la forme d'un fichier '.meme' en utilisant la fonction 'write_meme()' (bibliothèque R : universalmotif v1.12.3) :

```
write_meme(MergeMotif, paste0(save_path2, "0-
MergeMotif.meme"), overwrite = T)
```

4.5.2 Définition des gènes avec motif par FIMO

Le logiciel en ligne FIMO 5.4.1 (Grant et al., 2011) de la suite MEME m'a permis d'identifier tous les gènes porteurs d'une version du motif consensus dans leur promoteur. Pour cela j'ai chargé le fichier "0-MergeMotif.meme" précédemment créé et l'ensemble des promoteurs de tous les gènes identifiés de la paramécie sur FIMO et lancé la recherche avec les paramètres par défaut.

Afin de comparer les motifs identifiés par STREME et FIMO et de définir une p-value seuil, j'ai calculé la moyenne du nombre de gènes porteurs du motif identifié par STREME parmi mes 112 gènes d'intérêt. Ce qui m'a permis d'obtenir l'enrichissement STREME moyen. J'ai ensuite extrait les statistiques des motifs identifiés par FIMO pour mes 112 gènes d'intérêt et sectionné la p-value qui me permettait d'obtenir un enrichissement au plus proche de l'enrichissement STREME moyen calculé précédemment. La p-value seuil est ensuite appliquée à la liste des motifs identifié par FIMO pour définir la liste des gènes porteur du motif.

L'intersection des gènes dérégulés avec les gènes du pic intermédiaire et les gènes porteurs d'au moins un motif dans leur promoteur a été représentée par un diagramme de Venn proportionnel en utilisant la fonction 'draw.triple.venn()' (bibliothèque R : VennDiagram v1.7.3).

4.5.3 Analyse R du motif dans les promoteurs des gènes

4.5.3.1 Enrichissement du motif

STREME permet l'identification d'un motif dans un groupe de séquences données comparées à des séquences contrôles. Afin de vérifier que la liste de gènes identifiée par FIMO reflétait toujours cet enrichissement, j'ai calculé l'enrichissement en motif pour mes 112 gènes d'intérêt (Candidats), pour l'ensemble des autres gènes (All*) et pour les gènes du pic intermédiaire ne faisant pas partie de mes candidats (InterP*). J'ai représenté les enrichissements par un diagramme en barre généré par la fonction 'barplot()' (bibliothèque R : *graphics* v4.1.2) et testé la significativité de cet enrichissement par un test du Chi2 en utilisant la fonction 'chisq.test()' (bibliothèque R : *stats* v4.1.2).

4.5.3.2 Orientation du motif

Étant donné que STREME permet l'identification de motif sans tenir compte de son orientation, j'ai cherché si le motif avait une orientation préférentielle dans les promoteurs des trois groupes de gènes définis précédemment (Candidats, All* et InterP*). Pour cela j'ai défini trois catégories : *plus* (les gènes contenant au moins un motif sur le brin +), *minus* (les gènes contenant au moins un motif sur le brin -), et *both* (les gènes contenant au moins un motif sur le brin + et un sur le brin -). J'ai ensuite représenté par un diagramme empilé la proportion de chacune de ces trois catégories pour les gènes des groupes Candidats, All* et InterP* en utilisant la fonction 'barplot()' (bibliothèque R : *graphics* v4.1.2).

4.5.3.3 Position du motif

Enfin, j'ai cherché si le motif avait une position préférentielle dans le promoteur de mes gènes d'intérêt comparé aux autres gènes. Pour cela j'ai considéré la position du nucléotide en 5' le plus proche du TSS et représenté la densité en motif le long des promoteurs des trois groupes de gènes définis

précédemment (Candidats, All* et InterP*). La densité a été extraite en utilisant la fonction 'hist()' (bibliothèque R : *graphics* v4.1.2) et représenté graphiquement en utilisant la fonction 'ggplot()' (bibliothèque R : *ggplot2* v3.3.6).

J'ai ensuite découpé les 150 pb des promoteurs en fraction de 5 pb afin de définir si le motif était significativement enrichi dans certaines de ces fractions. Pour cela j'ai utilisé le même test d'enrichissement que décrit dans section 4.4.2 (Boyle et al., 2004).

Partie V – Annexes

Chapitre 1 - Données supplémentaires utilisées dans l'introduction

Dans ce chapitre sont regroupés sous forme de tableaux les liste des gènes connus pour leur implication dans les différents processus des RPG. Les tableaux reprennent les articles décrivant les gènes en question ainsi que leur localisation si elle est connue. Les gènes sont regroupés par fonction :

1. Reconnaissance des séquences à éliminer par les ARNnc

Cette catégorie regroupe les gènes impliqués dans la biogénèse des ARNscn et ARNies ainsi que le mécanisme par lequel ils permettent d'identifier les séquences à éliminer.

2. Modification de la chromatine et marques épigénétiques

Cette catégorie regroupe d'une part les gènes identifiés comme faisant partie du complexe PRC2 et d'autre part, les autres gènes connus pouvant affecter la structure de la chromatine.

3. Le complexe d'excision des IES

Cette catégorie regroupe l'ensemble des gènes identifier comme étant nécessaire à l'introduction des cassures aux bornes des IES.

4. Des cassures prises en charge par le NHEJ

Cette dernière catégorie regroupe les gènes impliqués dans la réparation des cassures introduite par le complexe d'excision. L'hétérodimère Ku70/Ku80c n'est pas présent dans cette partie car il est déjà décrit dans la partie précédente. Mais il est fortement supposé qu'il est nécessaire à la fois pour introduire les cassures double brins aux bornes des IES et pour la réparation de ces cassures.

1.1 RECONNAISSANCE DES SÉQUENCES À ÉLIMINER PAR LES ARNnc

ID <i>Nom du gène</i>	Profil d'expression	Localisation et référence
PTET.51.1.G0770102 <i>SPT5m</i>	Early peak	MIC méiotique → début du développement des nouveaux MIC et MAC (Gruchota et al., 2017)
PTET.51.1.G0010243 <i>SPT4mA</i>	Early peak	- (Owsian et al., 2022)
PTET.51.1.G0080193 <i>SPT4mB</i>	Early peak	MIC méiotique → fragments du MAC parental & début du développement des nouveaux MAC (Owsian et al., 2022)
PTET.51.1.G0320167 <i>SPT4vA</i>	Late repression	MAC parental végétatif → MIC méiotique → fragments du MAC parental & nouveaux MAC (Owsian et al., 2022)
PTET.51.1.G0210241 <i>DCL2</i>	Early peak	MIC méiotique et cytoplasme (Lepère et al., 2009)
PTET.51.1.G0990073 <i>DCL3</i>	Early peak	- (Lepère et al., 2009)
PTET.51.1.G0710112 <i>PTIWI01</i>	Early peak	Cytoplasme → MAC parental & nouveaux MAC (Bouhouche et al., 2011; Furrer et al., 2017)
PTET.51.1.G0660118 <i>PTIWI09</i>	Early peak	Cytoplasme → MAC parental (avant fragmentation) → nouveaux MAC (Bouhouche et al., 2011)
PTET.51.1.G0170077 <i>NOWA1</i>	Early peak	Fragments du MAC parental → nouveau MAC (Nowacki et al., 2005)
PTET.51.1.G0490206 <i>NOWA2</i>	Early peak	- (Nowacki et al., 2005)
PTET.51.1.G0010313 <i>PTMB.220</i>	Early peak	MAC parental (avant fragmentation) → nouveaux MAC (Nowak et al., 2011)
PTET.51.1.G0300085 <i>PDSG1</i>	Early peak	MAC parental (avant fragmentation) → nouveaux MAC (Arambasic et al., 2014)
PTET.51.1.G0380048 <i>SPT5v</i>	Early peak	MAC parental végétatif → fragments du MAC parental & nouveaux MAC (Gruchota et al., 2017)
PTET.51.1.G0900102 <i>TFIIS4</i>	Early / Intermediate peak	Nouveaux MAC (Maliszewska-Olejniczak et al., 2015)
PTET.51.1.G0070121 <i>DCL5</i>	Intermediate peak	Nouveaux MAC (foci) (Sandoval et al., 2014)
PTET.51.1.G0110289 <i>PDSG2</i>	Intermediate peak	Nouveaux MAC (Arambasic et al., 2014)
PTET.51.1.G0240239 <i>PTIWI10</i>	Late induction	Nouveaux MAC (Furrer et al., 2017)
PTET.51.1.G0620215 <i>PTIWI11</i>	Late induction	- (Bouhouche et al., 2011; Furrer et al., 2017)

Tableau 22 - Liste des gènes impliqués dans la reconnaissance des IES par les ARNnc

Ce tableau regroupe le numéro d'accèsion (ID), le groupe de profil d'expression (Arnaiz et al., 2017) et les données de localisation des protéines correspondant aux gènes impliqués dans la reconnaissance par les ARNscn (décrit section 0) et les ARNies (décrit section 3.1.2) des séquences à éliminer lors des RPG.

1.2 MODIFICATION DE LA CHROMATINE ET MARQUES ÉPIGÉNÉTIQUES

ID <i>Nom du gène</i>	Profil d'expression	Localisation et référence
PTET.51.1.G1740049 <i>EZL1</i>	Early peak	MAC parental (avant fragmentation) → nouveaux MAC (foci) (Lhuillier-Akakpo et al., 2014)
PTET.51.1.G0190277 <i>SUZ12</i>	Early peak	MAC parental (avant fragmentation) → nouveaux MAC (Wang et al., 2022b)
PTET.51.1.G0240079 <i>EED</i>	Early peak	
PTET.51.1.G0780031 <i>CAF1</i>	Early peak	MAC parental (avant fragmentation) → nouveaux MAC (Ignarski et al., 2014)
PTET.51.1.G1190062 <i>RF2</i>	Early peak	MAC parental (avant fragmentation) → nouveaux MAC (Wang et al., 2022b)
PTET.51.1.G1310069 <i>EAP1</i>	Early peak	MAC parental (avant fragmentation) → nouveaux MAC (Miró-Pina et al., 2022)
PTET.51.1.G0570234 <i>RF4</i>	Early peak	Fragments du MAC parental → nouveaux MAC (tardif) (Miró-Pina et al., 2022)

Tableau 23 - Liste des gènes du complexe PRC2

Ce tableau regroupe le numéro d'accèsion (ID), le groupe de profil d'expression (Arnaiz et al., 2017) et les données de localisation des protéines correspondant aux gènes identifiés comme faisant partie du complexe PRC2 de *P. tetraurelia* (décrit section 3.1.3).

ID <i>Nom du gène</i>	Profil d'expression	Localisation et référence
PTET.51.1.G0710091 <i>SPT16-1</i>	Intermediate peak	Nouveau MAC (De Vanssay et al., 2020)
PTET.51.1.G0140243 <i>ISWI1a</i>	Intermediate peak	Nouveaux MAC (Singh et al., 2022)
PTET.51.1.G0420126 <i>ISWI1b</i>	Early repression	- (Singh et al., 2022)
PTET.51.1.G0290211 <i>ISWI1c</i>	Early repression	- (Singh et al., 2022)
PTET.51.1.G0410183 <i>ISWI1d</i>	none	- (Singh et al., 2022)
PTET.51.1.G1470016 <i>ISWI2</i>	Early peak	- (Singh et al., 2022)

Tableau 24 - Liste des gènes impliqués dans le recrutement du complexe d'excision à la chromatine

Ce tableau regroupe le numéro d'accèsion (ID), le groupe de profil d'expression (Arnaiz et al., 2017) et les données de localisation des protéines correspondant aux gènes participant au recrutement du complexe d'excision sur la chromatine (décrit section 3.1.3) et de leurs orthologues.

1.3 LE COMPLEXE D'EXCISION DES IES

ID <i>Nom du gène</i>	Profil d'expression	Localisation et référence
PTET.51.1.G0490162 <i>PGM</i>	Intermediate peak	Nouveaux MAC (foci) (Baudry et al., 2009; Dubois et al., 2017)
PTET.51.1.G0110267 <i>PGML1</i>	Intermediate peak	Nouveaux MAC (Bischerour et al., 2018)
PTET.51.1.G0380073 <i>PGML2</i>	Intermediate peak	Nouveaux MAC (Bischerour et al., 2018)
PTET.51.1.G0010374 <i>PGML3a</i>	Intermediate peak	Nouveaux MAC (Bischerour et al., 2018)
PTET.51.1.G0080308 <i>PGML3b</i>	Intermediate peak	- (Bischerour et al., 2018)
PTET.51.1.G0020217 <i>PGML3c</i>	Late induction	- (Bischerour et al., 2018)
PTET.51.1.G0340197 <i>PGML4a</i>	Intermediate peak	Nouveaux MAC (Bischerour et al., 2018)
PTET.51.1.G0480099 <i>PGML4b</i>	Intermediate peak	- (Bischerour et al., 2018)
PTET.51.1.G0570051 <i>PGML5a</i>	Intermediate peak	Nouveaux MAC (Bischerour et al., 2018)
PTET.51.1.G0510172 <i>PGML5b</i>	Intermediate peak	Nouveaux MAC (Bischerour et al., 2018)
PTET.51.1.G1140146 <i>KU80c</i>	Intermediate peak	Nouveaux MAC (foci) (Marmignon et al., 2014)
PTET.51.1.G0150242 <i>KU70a</i>	Intermediate peak	MAC végétatif → Nouveaux MAC (foci) (Marmignon et al., 2014)
PTET.51.1.G0250220 <i>KU70b</i>	None	- (Marmignon et al., 2014)

Tableau 25 - Liste des gènes composant le complexe d'excision des IES

Ce tableau regroupe le numéro d'accèsion (ID), le groupe de profil d'expression (Arnaiz et al., 2017) et les données de localisation des protéines correspondant aux gènes identifiés étant nécessaire à l'introduction des cassures aux bornes des IES (décrit section 3.2).

1.4 DES CASSURES PRISE EN CHARGE PAR LE NHEJ

ID <i>Nom du gène</i>	Profil d'expression	Localisation et référence
PTET.51.1.G0190231 <i>DNAPKcs</i>	Early peak	-
PTET.51.1.P0210235 <i>POLXa</i>	Intermediate peak	Nouveaux MAC (Données non publiées de l'équipe produites par Baptiste Verron et Julien Bischerour)
PTET.51.1.P0360066 <i>POLXb</i>	Intermediate peak	
PTET.51.1.P0460033 <i>POLXc</i>	none	
PTET.51.1.P1010039 <i>POLXd</i>	Late repression	
PTET.51.1.G0020380 <i>CERa</i>	Early peak	- (Données non publiée de l'équipe)
PTET.51.1.G0220178 <i>CERb</i>	Early peak	- (Données non publiée de l'équipe)
PTET.51.1.G0540024 <i>LIGASE4a</i>	Early peak	Cytoplasmique & Nouveaux MAC (Kapusta et al., 2011)
PTET.51.1.G1110086 <i>XRCC4</i>	Early peak	- (Kapusta et al., 2011)
PTET.51.1.G0860013 <i>DIE5a</i>	Intermediate peak	- (Matsuda et al., 2010)
PTET.51.1.G0950081 <i>DIE5b</i>	Intermediate peak	- (Matsuda et al., 2010)

Tableau 26 - Liste des gènes impliqués dans la réparation des cassures introduites lors des RPG

Ce tableau regroupe le numéro d'accèsion (ID), le groupe de profil d'expression (Arnaiz et al., 2017) et les données de localisation des protéines correspondant aux gènes identifiés étant nécessaire à la réparation des cassures doubles brins introduite au cours des RPG (décrit section 3.3) et de leurs orthologues.

Chapitre 2 - Données supplémentaires de la partie résultats

2.1 IDENTIFICATION DE GÈNES CANDIDATS POTENTIELLEMENT IMPLIQUÉS DANS LES

Names	Stages	Knock Down	Timing	Autog Stage	Condition	Seq_method	Labo
CTIP_Veg	CTIP_Veg	CTIP	Veg	VEG	VEG_CTIP	NextSeq	Betermier
CTIP_T0	CTIP_T0	CTIP	0	EARLY	EARLY_CTIP	NextSeq	Betermier
CTIP_T5.5	CTIP_T5.5	CTIP	5.5	INTER	INTER_CTIP	NextSeq	Betermier
CTIP_T12.5	CTIP_T12.5	CTIP	12.5	INTER	INTER_CTIP	NextSeq	Betermier
CTIP_T25	CTIP_T25	CTIP	25	LATE	LATE_CTIP	NextSeq	Betermier
ICL7_Veg	ICL7_Veg	ctrl	Veg	VEG	VEG_ctrl	HiSeq	Duharcourt
ICL7_T0	ICL7_T0	ctrl	0	EARLY	EARLY_ctrl	HiSeq	Duharcourt
ICL7_T5	ICL7_T5	ctrl	5	INTER	INTER_ctrl	HiSeq	Duharcourt
ICL7_T10	ICL7_T10	ctrl	10	INTER	INTER_ctrl	HiSeq	Duharcourt
ICL7_T20	ICL7_T20	ctrl	20	LATE	LATE_ctrl	HiSeq	Duharcourt
ICL7_T35	ICL7_T35	ctrl	35	LATE	LATE_ctrl	HiSeq	Duharcourt
ICL7_T50	ICL7_T50	ctrl	50	LATE	LATE_ctrl	HiSeq	Duharcourt
ICL7bis_T0	ICL7_T0	ctrl	0	EARLY	EARLY_ctrl	NextSeq	Duharcourt
ICL7bis_T10	ICL7_T10	ctrl	10	INTER	INTER_ctrl	NextSeq	Duharcourt
ICL7bis_T35	ICL7_T35	ctrl	35	LATE	LATE_ctrl	NextSeq	Duharcourt
ICL7bis_T50	ICL7_T50	ctrl	50	LATE	LATE_ctrl	NextSeq	Duharcourt
KU80c_Veg	KU80c_Veg	KU80c	Veg	VEG	VEG_KU80c	HiSeq	Betermier
KU80c_T0	KU80c_T0	KU80c	0	EARLY	EARLY_KU80c	HiSeq	Betermier
KU80c_T5	KU80c_T5	KU80c	5	INTER	INTER_KU80c	HiSeq	Betermier
KU80c_T10	KU80c_T10	KU80c	10	INTER	INTER_KU80c	HiSeq	Betermier
KU80c_T20	KU80c_T20	KU80c	20	LATE	LATE_KU80c	HiSeq	Betermier
KU80c_T30	KU80c_T30	KU80c	30	LATE	LATE_KU80c	HiSeq	Betermier
KU80c_T40	KU80c_T40	KU80c	40	LATE	LATE_KU80c	HiSeq	Betermier

Names	Stages	Knock Down	Timing	Autog Stage	Condition	Seq_ method	Labo
ND7_K_Veg	ND7_K_Veg	ctrl	Veg	VEG	VEG_ctrl	HiSeq	Betermier
ND7_K_T0	ND7_K_T0	ctrl	0	EARLY	EARLY_ctrl	HiSeq	Betermier
ND7_K_T5	ND7_K_T5	ctrl	5	INTER	INTER_ctrl	HiSeq	Betermier
ND7_K_T10	ND7_K_T10	ctrl	10	INTER	INTER_ctrl	HiSeq	Betermier
ND7_K_T20	ND7_K_T20	ctrl	20	LATE	LATE_ctrl	HiSeq	Betermier
ND7_K_T30	ND7_K_T30	ctrl	30	LATE	LATE_ctrl	HiSeq	Betermier
ND7_K_T40	ND7_K_T40	ctrl	40	LATE	LATE_ctrl	HiSeq	Betermier
ND7_L_Veg	ND7_L_Veg	ctrl	Veg	VEG	VEG_ctrl	NextSeq	Betermier
ND7_L_T0	ND7_L_T0	ctrl	0	EARLY	EARLY_ctrl	NextSeq	Betermier
ND7_L_T3.5	ND7_L_T3.5	ctrl	3.5	INTER	INTER_ctrl	NextSeq	Betermier
ND7_L_T6	ND7_L_T6	ctrl	6	INTER	INTER_ctrl	NextSeq	Betermier
ND7_L_T20	ND7_L_T20	ctrl	20	LATE	LATE_ctrl	NextSeq	Betermier
ND7_L_T30	ND7_L_T30	ctrl	30	LATE	LATE_ctrl	NextSeq	Betermier
ND7_X_Veg	ND7_X_Veg	ctrl	Veg	VEG	VEG_ctrl	NextSeq	Betermier
ND7_X_T2	ND7_X_T2	ctrl	2	EARLY	EARLY_ctrl	NextSeq	Betermier
ND7_X_T7	ND7_X_T7	ctrl	7	INTER	INTER_ctrl	NextSeq	Betermier
ND7_X_T22	ND7_X_T22	ctrl	22	LATE	LATE_ctrl	NextSeq	Betermier
ND7_X_T32	ND7_X_T32	ctrl	32	LATE	LATE_ctrl	NextSeq	Betermier
ND7_X_T42	ND7_X_T42	ctrl	42	LATE	LATE_ctrl	NextSeq	Betermier
PGM_Veg	PGM_Veg	PGM	Veg	VEG	VEG_PGM	HiSeq	Betermier
PGM_T2	PGM_T2	PGM	2	INTER	INTER_PGM	HiSeq	Betermier
PGM_T5	PGM_T5	PGM	5	INTER	INTER_PGM	HiSeq	Betermier
PGM_T10	PGM_T10	PGM	10	INTER	INTER_PGM	HiSeq	Betermier
PGM_T20	PGM_T20	PGM	20	LATE	LATE_PGM	HiSeq	Betermier
PGM_T30	PGM_T30	PGM	30	LATE	LATE_PGM	HiSeq	Betermier
PGM_T40	PGM_T40	PGM	40	LATE	LATE_PGM	HiSeq	Betermier
XRCC4_Veg	XRCC4_Veg	XRCC4	Veg	VEG	VEG_XRCC4	NextSeq	Betermier
XRCC4_T2	XRCC4_T2	XRCC4	2	EARLY	EARLY_XRCC4	NextSeq	Betermier
XRCC4_T7	XRCC4_T7	XRCC4	7	INTER	INTER_XRCC4	NextSeq	Betermier
XRCC4_T22	XRCC4_T22	XRCC4	22	LATE	LATE_XRCC4	NextSeq	Betermier
XRCC4_T32	XRCC4_T32	XRCC4	32	LATE	LATE_XRCC4	NextSeq	Betermier
XRCC4_T42	XRCC4_T42	XRCC4	42	LATE	LATE_XRCC4	NextSeq	Betermier

Tableau 27 - Métadonnées utilisées pour l'analyse des gènes différentiellement exprimés

2.2 CARACTÉRISATION DE L'EFFET D'UN ARNI CONTRE LE COMPLEXE CONDENSINE

ID <i>Nom du gène</i>	Profil d'expression	Dérégulation	TurboID	Motif Position (depuis TSS)
PTET.51.1.G0050262 CPD1-1a	Early repression	C : ↗	-	-
PTET.51.1.G0160194 CPD1-1b	Early repression	-	-	-
PTET.51.1.G0100159 CPD1-1c	Early peak	-	-	-
PTET.51.1.G0160217 CPD1-2a	Early peak	P : ↗	-	-
PTET.51.1.G0050285 CPD1-2b	Early peak	P : ↗	-	-
PTET.51.1.G0430205 CPD2	Intermediate peak	C : ↘ P, K & X : ↗	Pgm & PgmL4	AAATTCTTGCTAATATTT -66 à -49
PTET.51.1.G1350115 CPG1	Early/Intermediate peak	C : ↘ P, K & X : ↗	-	-
PTET.51.1.G0480269 CPH3a	none	-	-	-
PTET.51.1.G0280009 CPH3	Early peak	-	-	-
PTET.51.1.G0260213 CPH4a	Intermediate peak	C : ↘ P, K & X : ↗	Pgm & PgmL4	AAAATCATTAAAATATTT -53 à -36
PTET.51.1.G0090252 CPH4b	Intermediate peak	C : ↘ P, K & X : ↗	-	AAAATCATTAAAATATTT -54 à -37
PTET.51.1.G0450077 SMC2a	Early/Intermediate peak	C : ↘ P, K & X : ↗	PgmL4	-
PTET.51.1.G0330075 SMC2b	Early/Intermediate peak	C : ↘ P, K & X : ↗	PgmL4	-
PTET.51.1.G0410063 SMC4-1	Late repression	-	-	-
PTET.51.1.G0590135 SMC4-2	Intermediate peak	C : ↘ P, K & X : ↗	-	AAAATCATTAAAATATTT -54 à -37

Tableau 28 - Caractéristiques des gènes condensines identifiés chez la paramécie

Ce tableau regroupe le numéro d'accèsion (ID), le groupe de profil d'expression (Arnaiz et al., 2017) de gènes homologues aux gènes des complexes condensines (gènes identifiés par Valerio Vitali). Pour tous les gènes la colonne 'Dérégulation' indique la surexpression (↗) ou la perte d'expression (↘) pour la cinétique CTIP au temps précoce ou intermédiaire (C), les cinétiques PGM (P), KU80c (K) et XRCC4 (X) au temps tardif. La colonne 'TurboID' indique dans quelles turbo-ID les protéines ont été trouvées (communication personnelle Marc Guérineau). Enfin la colonne 'Motif' indique la séquence du motif (5' → 3') identifiée dans le promoteur (si elle est présente) ainsi que sa position relative au TSS du gène.

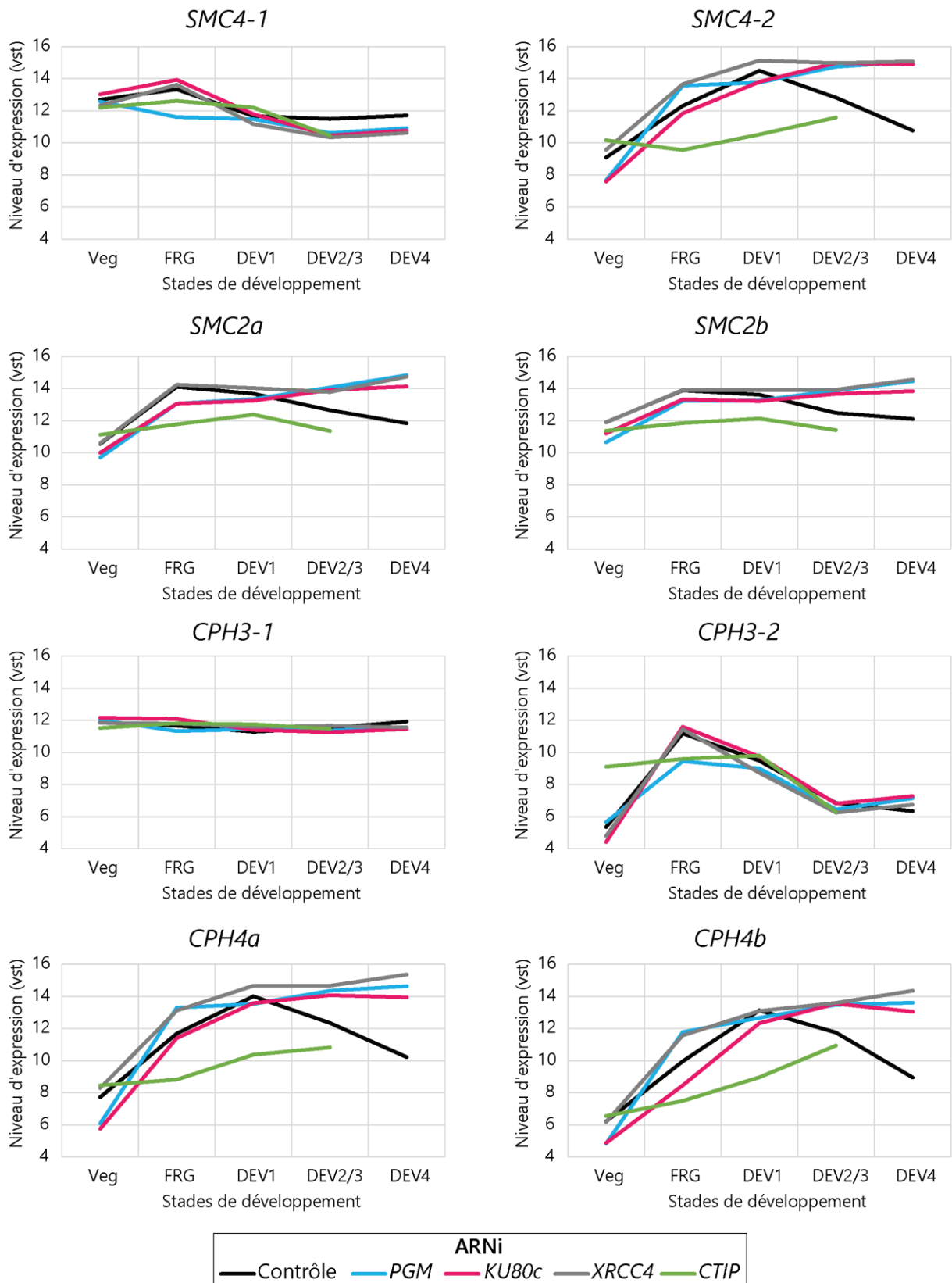


Figure 82 - Profils d'expression des gènes pouvant coder pour les Smc et les kleisines composant les complexes condensine de la paramécie

Profils d'expression au cours de l'autogamie dans différentes cinétiques d'autogamie. Veg = Végétatif, FRG = Fragment (T0-T4), DEV : Développement des nouveau MAC (DEV1 : T5-T10, DEV2/3 : T11-T19, DEV4 : T20-T50)

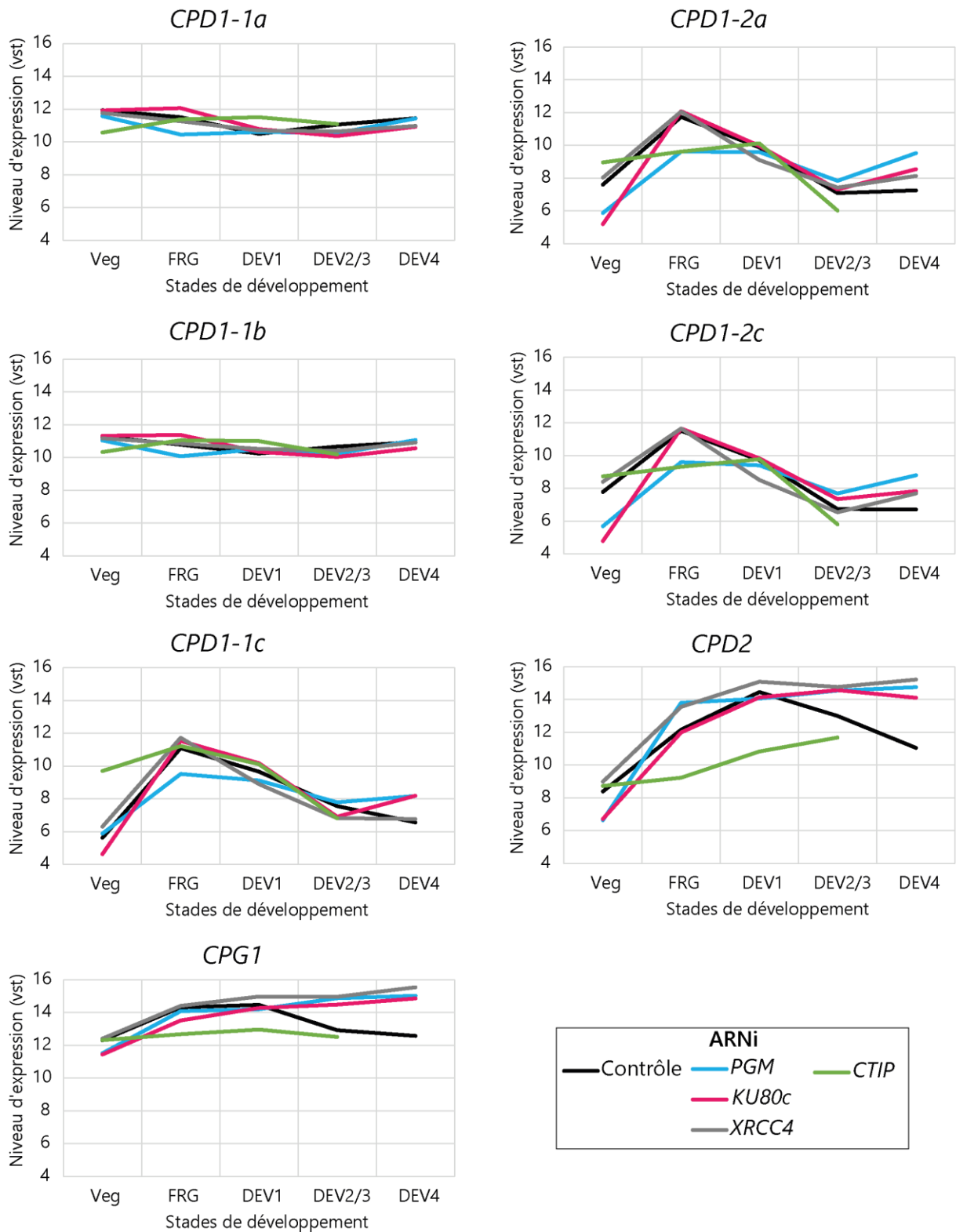


Figure 83 - Profils d'expression des gènes pouvant coder pour les protéines HAWK des complexes condensine de la paramécie

Profils d'expression au cours de l'autogamie dans différentes cinétiques d'autogamie. Veg = Végétatif, FRG = Fragment (T0-T4), DEV : Développement des nouveau MAC (DEV1 : T5-T10, DEV2/3 : T11-T19, DEV4 : T20-T50)

2.3 MISE AU POINT DU CUT&TAG

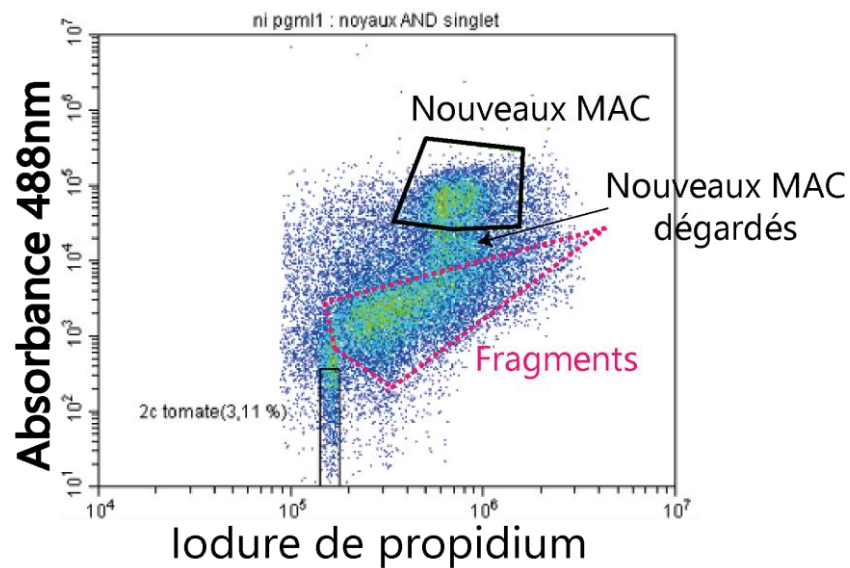


Figure 84 - Exemple de profil de noyaux abimés

Exemple de profil observé par cytométrie en flux sur des noyaux abimés marqués avec un anti-PgmL1 et un anticorps secondaire couplé à un fluorophore AF488. Les noyaux abimés sont moins bien marqués par PgmL1 ce qui se traduit par le drapage observé sous les noyaux correspondants aux nouveaux MAC.

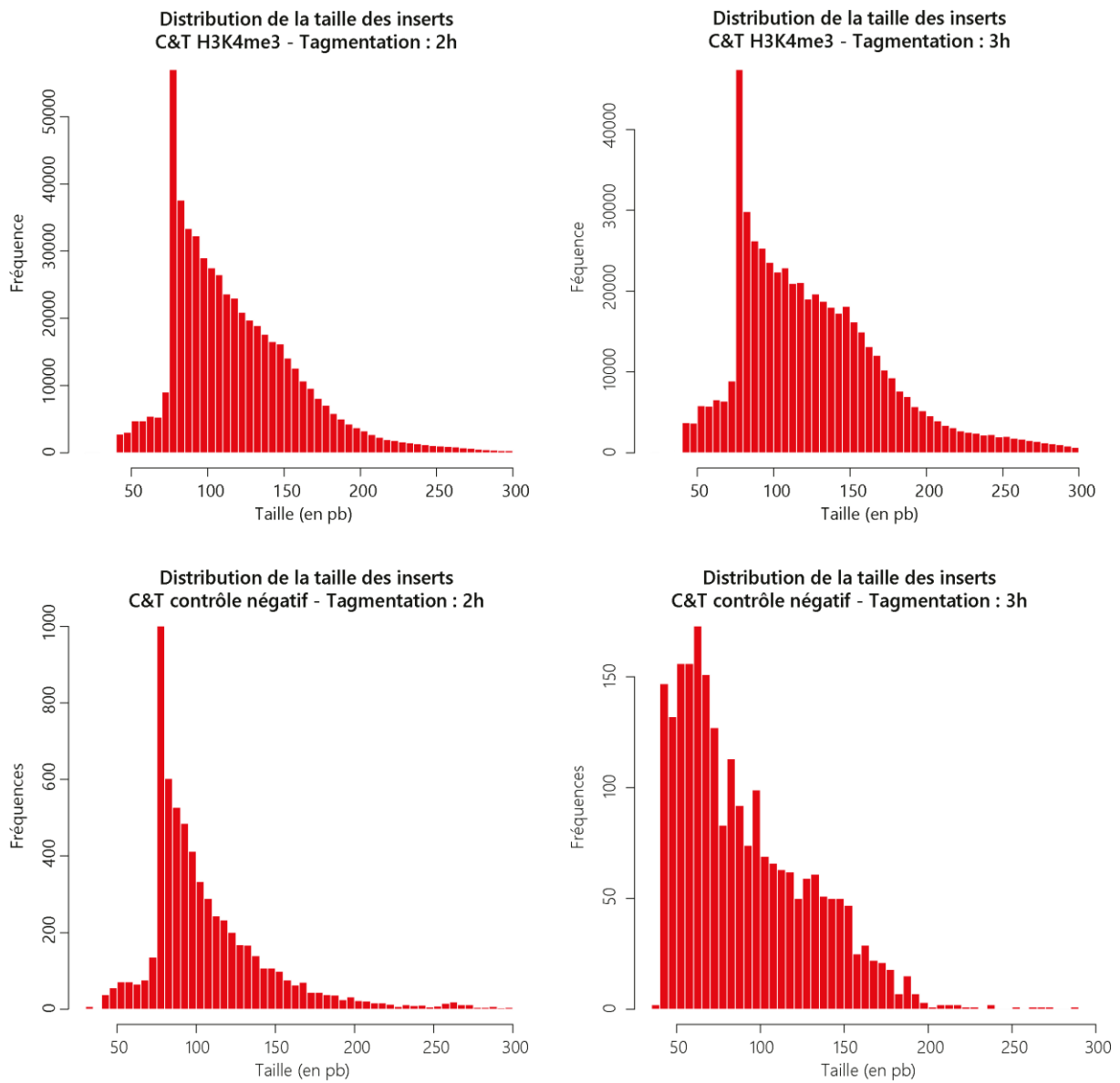


Figure 85 - Distribution de la taille des inserts séquencés pour les banques de Cut&Tag

Les contrôles positifs correspondent aux conditions pour lesquels l'anticorps primaire anti-H3K4me3 n'a pas été ajouté. (Analyse fait par Oliver Arnaiz)

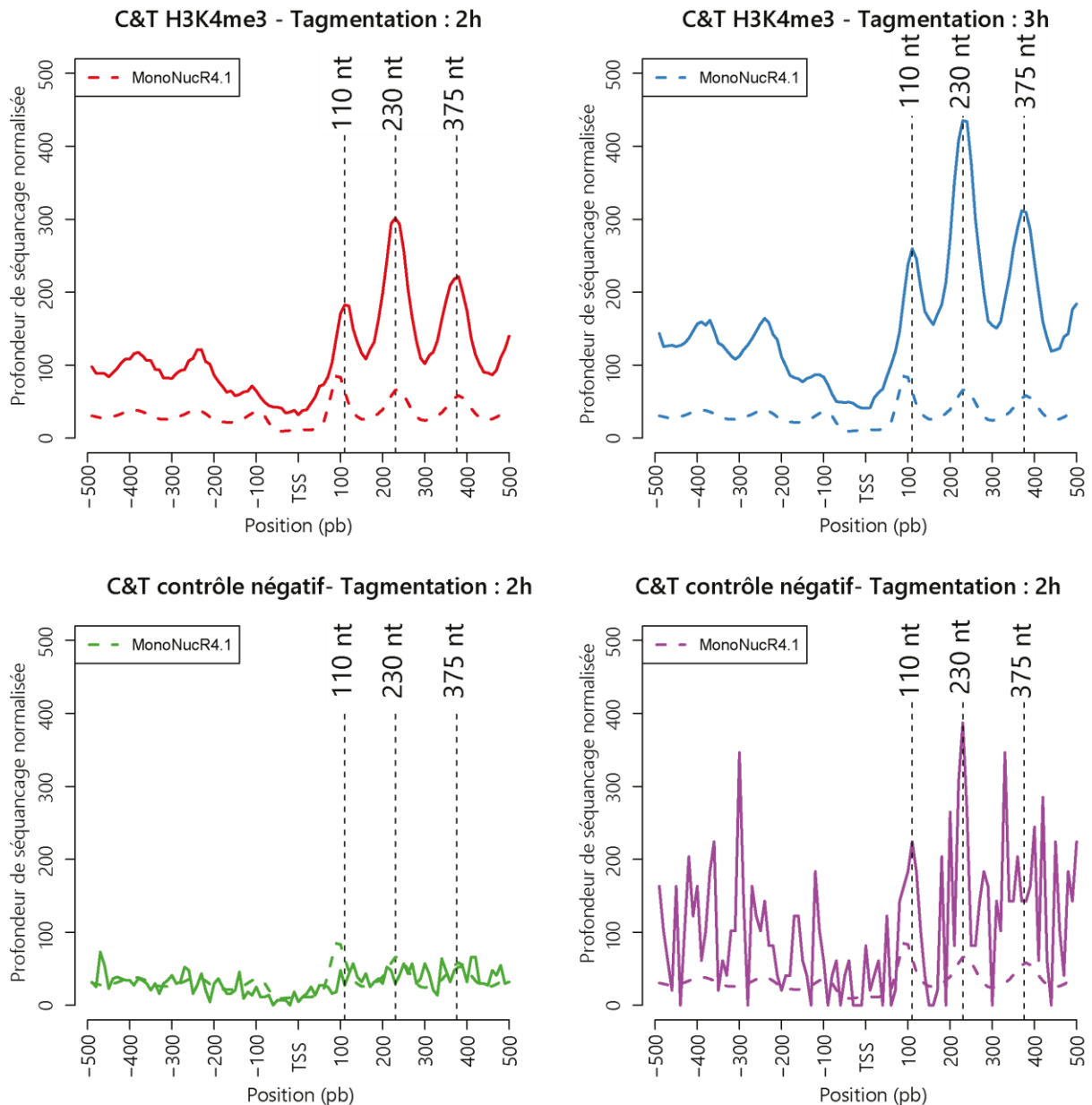


Figure 86 - Comparaison de la position des inserts séquençés pour les Cut&Tag avec 2h ou 3h de tagmentation

Pour chacune des banques et leur contrôle négatif (sans anticorps primaire H3K4me3) les inserts mesurant 50 à 100 pb sont sélectionnés. Puis la couverture normalisée par le nombre total de lecture est calculée pour chaque position (la position d'un insert est déterminée par son centre). Les graphiques représentent ici la couverture autour des TSS des gènes. (Analyse fait par Olivier Arnaiz)

Chapitre 3 - Données supplémentaires de la partie discussion

Noms du gène	Rôle dans les RPG	Profil d'expression	nb IES	CTIP	KU80c	PGM	XRCC4	Turbo Pgm	Turbo PgmL4	EZL1 (1)	EZL1 (2)
<i>PGM</i>	Excision	<i>Intermediate peak</i>	0	Down	Up	ns	Up	Présent	Présent	Up	Up
<i>PGML1</i>	Excision	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>PGML2</i>	Excision	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>PGML3a</i>	Excision	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>PGML3b</i>	Excision	<i>Intermediate peak</i>	0	Down	ns	Up	Up	ns	ns	ns	ns
<i>PGML3c</i>	Excision	<i>Late induction</i>	0	Down	ns	Up	Up	ns	ns	Up	Up
<i>PGML4a</i>	Excision	<i>Intermediate peak</i>	0	Down	Up	Up	Up	Présent	Présent	Up	Up
<i>PGML4b</i>	Excision	<i>Intermediate peak</i>	0	Down	Up	Up	Up	Présent	Présent	Up	Up
<i>PGML5a</i>	Excision	<i>Intermediate peak</i>	1	Down	Up	Up	Up	Présent	Présent	Up	Up
<i>PGML5b</i>	Excision	<i>Intermediate peak</i>	1	Down	Up	Up	Up	ns	Présent	Up	Up
<i>KU70a</i>	Excision & Réparation	<i>Intermediate peak</i>	0	Down	Up	Up	Up	Présent	Présent	Up	Up
<i>KU70b</i>	Excision & Réparation	<i>none</i>	0	ns	ns	Up	Up	Présent	Présent	ns	ns
<i>KU80c</i>	Excision & Réparation	<i>Intermediate peak</i>	0	Down	ns	Up	Up	ns	ns	Up	Up
<i>CERa</i>	Réparation	<i>Early peak</i>	0	Down	ns	ns	ns	ns	ns	ns	ns
<i>CERb</i>	Réparation	<i>none</i>	0	ns	ns	Up	ns	ns	ns	ns	ns
<i>DIE5a</i>	Réparation	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>DIE5b</i>	Réparation	<i>Intermediate peak</i>	0	Down	ns	Up	Up	ns	ns	Up	Up
<i>DNA-PKcs</i>	Réparation	<i>Early peak</i>	4	ns	ns	Up	ns	ns	ns	ns	ns

Partie V – Annexes | Données supplémentaires de la partie discussion

Noms du gène	Rôle dans les RPG	Profil d'expression	nb IES	CTIP	KU80c	PGM	XRCC4	Turbo Pgm	Turbo PgmL4	EZL1 (1)	EZL1 (2)
<i>LIG4a</i>	Réparation	<i>Early peak</i>	1	ns	ns	ns	ns	ns	Présent	ns	ns
<i>POLXa</i>	Réparation	<i>Intermediate peak</i>	0	Down	Up	Up	Up	Présent	Présent	Up	Up
<i>POLXb</i>	Réparation	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>XRCC4</i>	Réparation	<i>Early peak</i>	0	ns	ns	Up	ns	ns	Présent	ns	ns
<i>CPD2</i>	Condensine D	<i>Intermediate peak</i>	0	Down	Up	Up	Up	Présent	Présent	Up	Up
<i>CPG1</i>	Condensine D	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>CPH4a</i>	Condensine D	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	Présent	Up	Up
<i>CPH4b</i>	Condensine D	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>SMC2a</i>	Condensine D	<i>Early peak</i>	0	Down	Up	Up	Up	ns	Présent	Up	Up
<i>SMC2b</i>	Condensine D	<i>Early peak</i>	0	Down	Up	Up	Up	ns	Présent	Up	Up
<i>SMC4-2</i>	Condensine D	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>CAF1</i>	PRC2	<i>Early peak</i>	0	ns	ns	ns	ns	ns	ns	ns	ns
<i>EAP1</i>	PRC2	<i>Early peak</i>	1	ns	Up	Up	ns	ns	ns	ns	ns
<i>EED</i>	PRC2	<i>Early peak</i>	0	ns	ns	ns	ns	ns	ns	ns	ns
<i>EZL1</i>	PRC2	<i>Early peak</i>	0	ns	ns	Up	ns	ns	ns	ns	ns
<i>RF2</i>	PRC2	<i>Early peak</i>	0	ns	ns	Up	Up	ns	ns	Up	Up
<i>RF4</i>	PRC2	<i>Early peak</i>	1	ns	ns	ns	ns	ns	ns	ns	ns
<i>SUZ12</i>	PRC2	<i>Early peak</i>	1	ns	ns	ns	ns	ns	ns	Up	Up
<i>mtFA</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	ns	Up	Up	ns	ns	Up	Up
<i>mtFL1A</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL2A</i>	Reconnaissance ?	<i>Late induction</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL3A</i>	Reconnaissance ?	<i>Late induction</i>	0	Down	ns	Up	Up	ns	ns	Up	Up
<i>mtFL4A</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL5A</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL5B</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL6A</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL6B</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL7A</i>	Reconnaissance ?	<i>Late induction</i>	0	Down	ns	Up	Up	ns	ns	Up	Up
<i>mtFL7B</i>	Reconnaissance ?	<i>Late induction</i>	0	Down	ns	Up	Up	ns	ns	Up	Up
<i>mtFL8A</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL8B</i>	Reconnaissance ?	<i>Late induction</i>	0	Down	ns	Up	Up	ns	ns	Up	Up

Partie V – Annexes | Données supplémentaires de la partie discussion

Noms du gène	Rôle dans les RPG	Profil d'expression	nb IES	CTIP	KU80c	PGM	XRCC4	Turbo Pgm	Turbo PgmL4	EZL1 (1)	EZL1 (2)
<i>mtFL9A</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL10A</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL11A</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	ns	ns	Up	ns	ns	Up	Up
<i>mtFL11B</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	ns	ns	Up	ns	ns	Up	Up
<i>mtFL12A</i>	Reconnaissance ?	<i>Late induction</i>	0	Down	ns	ns	ns	ns	ns	Up	Up
<i>mtFL12B</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL13A</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL13B</i>	Reconnaissance ?	<i>Late induction</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>mtFL14A</i>	Reconnaissance ?	<i>Late induction</i>	1	ns	ns	Down	Down	ns	ns	ns	ns
<i>mtFL15A</i>	Reconnaissance ?	<i>Intermediate peak</i>	1	Down	ns	ns	ns	ns	ns	Up	Up
<i>mtFL16A</i>	Reconnaissance ?	<i>Late induction</i>	0	Down	ns	Up	Up	ns	ns	Up	Up
<i>mtFL17A</i>	Reconnaissance ?	<i>Intermediate peak</i>	0	Down	ns	Up	ns	ns	ns	Up	Up
<i>mtFL18A</i>	Reconnaissance ?	<i>Late induction</i>	0	Down	ns	ns	ns	ns	ns	Up	Up
<i>DCL5</i>	Reconnaissance ARNies	<i>Intermediate peak</i>	0	Down	ns	Up	Up	ns	ns	Up	Up
<i>PDSG2</i>	Reconnaissance ARNies	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>PTWI10</i>	Reconnaissance ARNies	<i>Late induction</i>	1	Down	Down	Down	Down	ns	ns	ns	ns
<i>PTWI11</i>	Reconnaissance ARNies	<i>Late induction</i>	1	Down	ns	Down	Down	ns	ns	ns	Up
<i>DCL2</i>	Reconnaissance ARNscn	<i>Early peak</i>	0	ns	ns	ns	ns	ns	ns	ns	ns
<i>DCL3</i>	Reconnaissance ARNscn	<i>Early peak</i>	1	ns	ns	ns	ns	ns	ns	ns	ns
<i>NOWA1</i>	Reconnaissance ARNscn	<i>Early peak</i>	0	ns	Up	Up	Up	ns	ns	Up	Up
<i>NOWA2</i>	Reconnaissance ARNscn	<i>Early peak</i>	0	ns	Up	Up	Up	ns	ns	Up	Up
<i>PDSG1</i>	Reconnaissance ARNscn	<i>Early peak</i>	0	ns	ns	ns	ns	ns	ns	ns	ns
<i>PTWI01</i>	Reconnaissance ARNscn	<i>Early peak</i>	2	ns	ns	ns	ns	ns	ns	ns	ns
<i>PTWI09</i>	Reconnaissance ARNscn	<i>Early peak</i>	1	ns	ns	ns	ns	ns	ns	ns	ns
<i>PTMB.220</i>	Reconnaissance ARNscn	<i>Early peak</i>	2	ns	Up	Up	ns	ns	ns	ns	Up
<i>SPT4mA</i>	Reconnaissance ARNscn	<i>Early peak</i>	0	ns	Up	Up	Up	ns	ns	Up	Up

Noms du gène	Rôle dans les RPG	Profil d'expression	nb IES	CTIP	KU80c	PGM	XRCC4	Turbo Pgm	Turbo PgmL4	EZL1 (1)	EZL1 (2)
<i>SPT4mB</i>	Reconnaissance ARNscn	<i>Early peak</i>	0	ns	ns	ns	ns	ns	ns	ns	ns
<i>SPT4vA</i>	Reconnaissance ARNscn	<i>Late repression</i>	0	ns	ns	Up	Up	ns	ns	ns	ns
<i>SPT4vC</i>	Reconnaissance ARNscn	<i>none</i>	0	ns	ns	ns	ns	ns	ns	ns	ns
<i>SPT5m</i>	Reconnaissance ARNscn	<i>Early peak</i>	1	ns	ns	ns	ns	ns	ns	ns	ns
<i>SPT5v</i>	Reconnaissance ARNscn	<i>Early peak</i>	0	ns	Up	Up	Up	ns	ns	Up	Up
<i>TFIIS4</i>	Reconnaissance ARNscn	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>ISWI1a</i>	Remodelage	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up
<i>SPT16</i>	Remodelage	<i>Intermediate peak</i>	0	Down	Up	Up	Up	ns	ns	Up	Up

Tableau 29 - Dérégulation des gènes impliqués dans les RPG

Ce tableau regroupe les gènes impliqués dans les RPG ainsi que leur rôle ou le complexe auquel ils appartiennent, le groupe de profil d'expression de gènes (Arnaiz et al., 2017), le nombre d'IES présents dans leur séquence codante (nb IES). Pour chacun des gènes est indiqué s'ils sont significativement dérégulés au temps précoce ou intermédiaire dans la cinétique CTIP, au temps tardif dans les cinétiques PGM, KU80c et XRCC4 ; mais également s'ils sont enrichis lors d'expérience de TurboID ciblant Pgm ou PgmL4 (respectivement 'Turbo Pgm' et 'Turbo PgmL4', Communication personnelle Marc Guérineau) et s'ils sont significativement dérégulés au temps tardif lors d'un ARNi contre EZL1 (EZL1 (1) : données publiées (Frapporti et al., 2019), EZL1 (2) : analyse refaite avec les réplicats techniques).

Up = surexprimée, Down = expression réduite

Gras = Surexprimé au temps tardif lors d'ARNi contre *PGM*, *KU80c* et *XRCC4*, Gras bleu = comme pour les gras mais ayant également une expression réduite au temps précoce ou intermédiaire lors d'un ARNi contre *CTIP*, Rouge = Expression réduite au temps tardif lors d'un ARNi contre *PGM*, *KU80c* ou *XRCC4*, ns = non-significatif.

ID	Nom du gène	Profil d'expression	Référence
PTET.51.1.G0660190	<i>mtGa</i>	Intermediate peak	(Bhullar et al., 2018)
PTET.51.1.G0710056	<i>mtGb</i>	Intermediate peak	(Bhullar et al., 2018)
PTET.51.1.G0590028	<i>mtF</i>	Intermediate peak	(Meyer and Keller, 1996)
PTET.51.1.G0050231	<i>mtFL1A</i>	Intermediate peak	Données non-publiées de l'équipe d'Éric Meyer
PTET.51.1.G0380022	<i>mtFL2A</i>	Late induction	
PTET.51.1.G0350167	<i>mtFL3A</i>	Late induction	
PTET.51.1.G1530110	<i>mtFL4A</i>	Intermediate peak	
PTET.51.1.G1300067	<i>mtFL5A</i>	Intermediate peak	
PTET.51.1.G1400105	<i>mtFL5B</i>	Intermediate peak	
PTET.51.1.G0350166	<i>mtFL6A</i>	Intermediate peak	
PTET.51.1.G0370136	<i>mtFL6B</i>	Intermediate peak	
PTET.51.1.G0010451	<i>mtFL7A</i>	Late induction	
PTET.51.1.G0080368	<i>mtFL7B</i>	Late induction	
PTET.51.1.G0920155	<i>mtFL8A</i>	Intermediate peak	
PTET.51.1.G0910136	<i>mtFL8B</i>	Late induction	
PTET.51.1.G0610198	<i>mtFL9A</i>	Intermediate peak	
PTET.51.1.G0030168	<i>mtFL10A</i>	Intermediate peak	
PTET.51.1.G0360089	<i>mtFL11A</i>	Intermediate peak	
PTET.51.1.G0210213	<i>mtFL11B</i>	Intermediate peak	
PTET.51.1.G0020335	<i>mtFL12A</i>	Late induction	
PTET.51.1.G0220140	<i>mtFL12B</i>	Intermediate peak	
PTET.51.1.G0230222	<i>mtFL13A</i>	Intermediate peak	
PTET.51.1.G0120245	<i>mtFL13B</i>	Late induction	
PTET.51.1.G1150114	<i>mtFL14A</i>	Late induction	
PTET.51.1.G1200062	<i>mtFL15A</i>	Intermediate peak	
PTET.51.1.G0260051	<i>mtFL16A</i>	Late induction	
PTET.51.1.G0350154	<i>mtFL17A</i>	Intermediate peak	
PTET.51.1.G0450225	<i>mtFL18A</i>	Late induction	

Tableau 30 - Liste d'autres gènes impliqués dans les PRG

Ce tableau regroupe le numéro d'accèsion (ID), le groupe de profil d'expression de gènes (Arnaiz et al., 2017) impliqués (ou pouvant être impliqués) dans les PRG et de leurs orthologues.

ID	Nom du gène	1 ^{er} pb	Orientation	Séquence
PTET.51.1.G0430205	<i>CPD2</i>	-66	-	AAATTCTTGCTAATATTT
PTET.51.1.G0260213	<i>CPH4a</i>	-53	-	AAAATCATTAAAATATTT
PTET.51.1.G0090252	<i>CPH4b</i>	-54	-	AAAATCATTAAAATATTT
PTET.51.1.G0990073	<i>DCL3</i>	-122	+	AATATCCTTAATTTAAAT
PTET.51.1.G0070121	<i>DCL5</i>	-66	+	ATAATCAAGCAATTTTTT
PTET.51.1.G0860013	<i>DIE5a</i>	-56	-	AATATCATGCTTATTATT
PTET.51.1.G0950081	<i>DIE5b</i>	-54	-	AATATCATGCTTATTATT
PTET.51.1.G0140243	<i>ISWI1a</i>	-43	-	TTAATCATCAAATATTT
PTET.51.1.G0140243	<i>ISWI1a</i>	-138	+	GAAAGCATTAAATAATTT
PTET.51.1.G0150242	<i>KU70a</i>	-61	+	ATTATCATTCCAATAATT
PTET.51.1.G0150242	<i>KU70a</i>	-50	-	AATATCATTATAATTATT
PTET.51.1.G0590028	<i>mtFA</i>	-111	-	AAAAGCATTGATTTAATT
PTET.51.1.G0030168	<i>mtFL10A</i>	-54	+	TAAATCCCTATAATTTTT
PTET.51.1.G0030168	<i>mtFL10A</i>	-140	-	TAAATAATTATTTTATTT
PTET.51.1.G0360089	<i>mtFL11A</i>	-56	+	CAAATCCTTGTAATAATTT
PTET.51.1.G0210213	<i>mtFL11B</i>	-56	+	CAAATCCTTGTAATAATTT
PTET.51.1.G0220140	<i>mtFL12B</i>	-63	+	AATTTCTGTTATTTATT
PTET.51.1.G1200062	<i>mtFL15A</i>	-84	+	ATTTTCATTTTATTTTT
PTET.51.1.G0260051	<i>mtFL16A</i>	-53	-	AAATTCATTGTAATATTT
PTET.51.1.G0350154	<i>mtFL17A</i>	-56	-	ATAATCCTTCTAATTATT
PTET.51.1.G0050231	<i>mtFL1A</i>	-75	-	TATATCATGGTAATAAAT
PTET.51.1.G0350167	<i>mtFL3A</i>	-91	+	AAAATCATTTTATTTATT
PTET.51.1.G1530110	<i>mtFL4A</i>	-102	-	ATTTTCATTATATATTT
PTET.51.1.G0370136	<i>mtFL6B</i>	-35	+	ATTTTCCTTCTATTAATT
PTET.51.1.G0080368	<i>mtFL7B</i>	-149	-	TATATCCTGAATTTATTT
PTET.51.1.G0080368	<i>mtFL7B</i>	-31	+	TAATTCATTCTAAAAATT
PTET.51.1.G0920155	<i>mtFL8A</i>	-53	+	ATAATCATTGAATTCATT
PTET.51.1.G0910136	<i>mtFL8B</i>	-27	+	ATAATCATTGTATTCATT
PTET.51.1.G0910136	<i>mtFL8B</i>	-100	+	ATATTCCATGTATTTATT
PTET.51.1.G0170077	<i>NOWA1</i>	-29	+	TAAATCAATCTAAAAATT
PTET.51.1.G0490206	<i>NOWA2</i>	-29	+	TAAATCAATCTAAAAATT
PTET.51.1.G0490162	<i>PGM</i>	-55	+	AAAATCATTAAAATATTT
PTET.51.1.G0110267	<i>PGML1</i>	-62	+	AAAATCATTATAATATTT
PTET.51.1.G0010374	<i>PGML3a</i>	-78	-	ATTATCCTTTTATTATTT
PTET.51.1.G0080308	<i>PGML3b</i>	-88	-	AATATCCTTTTATTATTT
PTET.51.1.G0020217	<i>PGML3c</i>	-75	-	GATATCCTTTTATTTTTT
PTET.51.1.G0570051	<i>PGML5a</i>	-52	+	AAAATCCAGGAAATAAATT
PTET.51.1.G0510172	<i>PGML5b</i>	-55	+	AAAATCCAGGAAATAAATT
PTET.51.1.G0110289	<i>PDSG2</i>	-54	-	TAAATCCTTATAAAATTT
PTET.51.1.G0710112	<i>PTIWI01</i>	-116	-	AAATTAATTTAAATAAATT
PTET.51.1.G0240239	<i>PTIWI10</i>	-56	+	AAAATCCTTAGATTTTTT
PTET.51.1.G0590135	<i>SMC4-2</i>	-54	-	AAAATCATTAAAATATTT
PTET.51.1.G0710091	<i>SPT16</i>	-53	+	AAATGCATTGAAATAAATT

ID	Nom du gène	1 ^{er} pb	Orientation	Séquence
PTET.51.1.G0380048	<i>SPT5v</i>	-135	-	AATGTCCTTAAAAATTTT
PTET.51.1.G0900102	<i>TFIIS4</i>	-54	-	AAAAGCCTTTTAATAATT
PTET.51.1.G1110086	<i>XRCC4</i>	-41	-	TAAATCAATGAAAATATT
PTET.51.1.G1110086	<i>XRCC4</i>	-38	+	ATTTTCATTGATTTAATT

Tableau 31 - Gènes connus porteurs du motif dans leur promoteur

Ce tableau regroupe le numéro d'accèsion (ID), des gènes impliqués dans les RPG porteurs d'un motif dans leur promoteur. La position de la 1^{ère} base du motif par rapport au TSS du gène est indiquée dans la colonne '1^{er} pb'. L'orientation et la séquence du motif sont donnés dans les deux colonnes suivantes. En gras et bleu sont indiqués les gènes surexprimés au temps tardif lors d'ARNi contre *PGM*, *KU80c* et *XRCC4* ayant également une expression réduite au temps précoce ou intermédiaire lors d'un ARNi contre *CTIP*.

Chapitre 4 - Données supplémentaires de la partie matériel et méthodes

4.1 PROTOCOLES DE PARAMÉCIES

Gene cible	Nom du plasmide	Séquence de l'insert
ND7	p78_p0ND7C (Garnier et al., 2004)	agacatttgggtgcaactgaggagtaaacgtgctaggaggaaatgcaagagatgga gcattttttaacagttgacgacatttgggtaagtggatcagagtgctcctgaagagt cataatataaattagataagaaggttagttataaggatgaattggaggattatta taattaaccagatttgggaattagattagaaattatacttagcatcctaagcaaaga gatatgatttaattgaaaaattaaattgaaagtagaagagttgttattgattaaa tgataagtatagaagaacttgaaatagaggagcatataaattattgatgaatt cgcaaagaataaagaatatgatggagaatttttagcaaaaattgttgatgaggat gctgaaggttaatgggtacc
PGM	p79_pPBL49HN (Baudry et al., 2009)	ccatggttctgtgttgagtgctcctcatttacaacatagctctatttgaatc gttattccacatagctcaaacatgatacaaccttaattggaatgtgaagtgttg gaagcatcacaagtagatattctttgattccccatttagtggtttatctggat tatacactttgaattgaaaccttctgcaaaaggaatcacaacctcatcaatagt gatgaattctccagatagtaataacttagcaaatctcataattcatttgatttaa aagtctctgattttccagatgggatctctgcccctcaattttaatgcctcagccc cactgatattcaacaatttcttatagctctaaacttttcataactgacgataga actcattctttgagtgacccaatgagcataacgatccatcttttcttaggccaat tctctcaactttctctccagctctcagccttctttataggtctctaccttgca aagacttccaataatcccttttagtttgataatctaacaatacaagaatatttg aacagcaaaaatagctt
KU80c	p135_pKU80C-2 (Marmignon et al., 2014)	ctagtatttcattttgtcctccatgatttaatccactggaaccagattcctacca tattagatocctcttattatggtgtccttttctattggagtcataattgggatcct ctggattgtaaatagatgagatcattcctcacatgattcttattggtatcagtttg aaattctccaagggtagaatgttttctaaagataggcaatctctcttcaaacat cgtttatataattagaactgtacagagatctcatcattaatttggaaatcccccc tgaatgatgctctagcttacaagaatctggatgaaattgggaatacaattcaaa ggcaacatttgcctgggaatattctcatggtgttctcaattcttggaccatcaaa aaaacataacgagaattgatagattggttcttaacattgatatgaagctgtaagca ttctattat
XRCC4	p250_pXR-496-a (Kapusta et al., 2011)	aaatctccttgcaaaccttgtaaaaatgggtctaattggtgaaatgatggatttc aatcaaatagactcgtaacaatattacctcgttcaatttgggtactttctcagac aatatcaaaacataaatcgcaaggaaacttgctcattacctcacttcttttaaag aataactatataataataaattataataacctaataatctttatctggtctttct aacaggtgataattttgtcaaaaatatttttaatcccagtcattctcttgattta agcattccatgctattcatttcaacttttgaacatttaattccctattttctcaat aaacacattaatctatccaaatgcaaatttaatttattatcgattggatgatca aataacttttggaatataataaataatctatccttcccttgattcaattttaaagca ggctgttttgggtcaatggagcctattcgccatctgtttaaatttctttttattc

Gene cible	Nom du plasmide	Séquence de l'insert
<i>CtIPa</i>	p76_L4440::CtIP-AS+ (Godau et al., 2019)	cttcttgataatcaacccaaatcaaccaaagttttacaaccaaattttcttgaaa tatttgaagacaaataaaactcaagcatccaaaataaaaaaccattgaaatgcaag cgaagaaaaacaaaaaatttcaatataaacagaaattgatcttccaaaagaaaa gtgttcagagaaaactgttataaacagaaaggaaagacaataaatcaatgctcatg agtgcgaagaatgtgaacaattttataaaagcttgcgcaattcgaattaagcaga aaaattaaaacaggatttttcaaggcataggattaatcataagttaaatcaagaa aaattactattaactccagaatga
<i>CtIPb</i>	p77_L4440::CtIP-BS+ (Godau et al., 2019)	cttatcgataatcaacccaaatcagctaaaattttataaccaaatttttttgaaa tattcgaagataggtaaaactcaatcatccaaaaccataaattataaaaaactaa tgaagaatagaagaagttttctatttagactgaggctgcaacttccaaaggagaag gtgtttagagaagtcgttaaaaaatagaaaagaaagataataaatggatgccatt aatgcgaggaatgtgaaagattttacaaagccttgcgcaataataataaacgga aaaattaaaacaggattactcgaggcatcgaatgaatcataagataaatcaagaa aaattattattaactccagaatga
<i>SMC4-1</i>	p424_L4440_SMC4-1	ataaaaaataaaagacattggataaggatgaagtgaagattaaagatgacttaca aaaaatacaatctaaaattgaataaatgaattaagaatttaaaaaagctgaagtc taaatgatcgaaactatcgccttgtataaaaattgcaaggaggctgaagaaaatg cagaaaaagaatacaaggagactcaagcaaggataaagaatttgatgaattgat ttctcggttgaaagagaagactaataaattaaagtaggaaaaggaggaatgtag tcactatttaagaaagcctaagaagaattgagtgaaggagtaaccaaattagaat aaattagattgaactataaagaaatgtaaaacgattatgaatttttgatgaact agaagagatagctaattggaaactctaagatt
<i>SMC4-2</i>	p421_L4440_SMC4-2	tgcaagagatattggaataggctaaagaaaagagtggtaacttttagatggaaaa ctagtagagtaaagtggagtgatgtcaggtgggtccagaacagaggaaaggaggaa tgtctagtaaattttatgaggaactttctaatgattaaagagaataattaatcag gaaagaaaccttcttaagagagattggatcaactaaatttgaaatcttagca cttgagaaaaagtttaagagaaaccttgcggatgaggaatggccaattctaac acaaaataagattagttgataaaaaatgctctaagaataaataatgataataa aaagcaaatggaatccttttagattcaatattaagaatgccattataattaatag aaaaattaagaagttaaaaagagattgaagttgtgtaagagaaatagagaaaa tacaaaataaaatttgtttagatttaataagatctgcaagcgatagatgata

Tableau 32 - Plasmides utilisé pour induire les ARNi

Les plasmides utilisés ne diffèrent que pas l'intert intégré entre les promoteurs T7 convergeant du plasmide L4440.

4.2 CUT&TAG

ID	Index	Séquences
OMB1369	TAGATCGC	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTCAGATGTGTAT
OMB1370	CTCTCTAT	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTTCGTCGGCAGCGTCAGATGTGTAT
OMB1371	TATCCTCT	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTCAGATGTGTAT
OMB1372	AGAGTAGA	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTCAGATGTGTAT
OMB1373	GTAAGGAG	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTCAGATGTGTAT
OMB1374	ACTGCATA	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTCAGATGTGTAT
OMB1375	AAGGAGTA	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTCAGATGTGTAT
OMB1376	CTAAGCCT	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTCAGATGTGTAT
OMB1377	TGGAATC	AATGATACGGCGACCACCGAGATCTACACTGGAATCTCGTCGGCAGCGTCAGATGTGTAT
OMB1378	AACATGAT	AATGATACGGCGACCACCGAGATCTACACAACATGATTCGTCGGCAGCGTCAGATGTGTAT
OMB1379	TAAGGCGA	CAAGCAGAAGACGGCATAACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGTG
OMB1380	CGTACTAG	CAAGCAGAAGACGGCATAACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGTG
OMB1381	AGGCAGAA	CAAGCAGAAGACGGCATAACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGTG
OMB1382	TCCTGAGC	CAAGCAGAAGACGGCATAACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGTG
OMB1383	GGACTCCT	CAAGCAGAAGACGGCATAACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGTG
OMB1384	TAGGCATG	CAAGCAGAAGACGGCATAACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGTG
OMB1385	CTCTCTAC	CAAGCAGAAGACGGCATAACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGTG
OMB1386	CAGAGAGG	CAAGCAGAAGACGGCATAACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGTG
OMB1387	GCTACGCT	CAAGCAGAAGACGGCATAACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGTG
OMB1388	CGAGGCTG	CAAGCAGAAGACGGCATAACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGTG
OMB1501	TGATGAAA	AATGATACGGCGACCACCGAGATCTACACTGATGAAATCGTCGGCAGCGTCAGATGTGTAT
OMB1502	GTCGGACT	AATGATACGGCGACCACCGAGATCTACACGTCGGACTTCGTCGGCAGCGTCAGATGTGTAT
OMB1503	TTTCTAGC	AATGATACGGCGACCACCGAGATCTACACTTTCTAGCTCGTCGGCAGCGTCAGATGTGTAT
OMB1504	TAACCAAG	AATGATACGGCGACCACCGAGATCTACACTAACCAAGTCGTCGGCAGCGTCAGATGTGTAT
OMB1505	GTGTATCG	AATGATACGGCGACCACCGAGATCTACACGTGTATCGTCGTCGGCAGCGTCAGATGTGTAT
OMB1506	TCCATCAA	AATGATACGGCGACCACCGAGATCTACACTCCATCAATCGTCGGCAGCGTCAGATGTGTAT
OMB1507	TTCGTGCA	AATGATACGGCGACCACCGAGATCTACACTTCGTGCATCGTCGGCAGCGTCAGATGTGTAT
OMB1508	AGGTTGCC	AATGATACGGCGACCACCGAGATCTACACAGGTTGCCTCGTCGGCAGCGTCAGATGTGTAT
OMB1509	CCTTATGT	AATGATACGGCGACCACCGAGATCTACACCCTTATGTTTCGTCGGCAGCGTCAGATGTGTAT
OMB1510	CAGCAACG	AATGATACGGCGACCACCGAGATCTACACCAGCAACGTCGTCGGCAGCGTCAGATGTGTAT
OMB1511	AAGAGGCA	CAAGCAGAAGACGGCATAACGAGATTGCCTCTGTCTCGTGGGCTCGGAGATGTG
OMB1512	GTAGAGGA	CAAGCAGAAGACGGCATAACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGTG
OMB1513	TGGATCTG	CAAGCAGAAGACGGCATAACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGTG
OMB1514	CCGTTTGT	CAAGCAGAAGACGGCATAACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGTG
OMB1515	TGCTGGGT	CAAGCAGAAGACGGCATAACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGTG
OMB1516	AGGTTGGG	CAAGCAGAAGACGGCATAACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGTG
OMB1517	GTGTGGTG	CAAGCAGAAGACGGCATAACGAGATCACCACACGTCTCGTGGGCTCGGAGATGTG
OMB1518	TGGGTTTC	CAAGCAGAAGACGGCATAACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGTG
OMB1519	TGGTCACA	CAAGCAGAAGACGGCATAACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGTG
OMB1520	TTGACCCT	CAAGCAGAAGACGGCATAACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGTG

Tableau 33 - Séquences des amorces disponibles pour Cut&Tag

Partie VI – Bibliographie

- Abello, A., Régnier, V., Arnaiz, O., le Bars, R., Bétermier, M. and Bischerour, J.** (2020). Functional diversification of *Paramecium* Ku80 paralogs safeguards genome integrity during precise programmed DNA elimination. *PLoS Genet* **16**, e1008723. doi/10.1371/journal.pgen.1008723.
- Adl, S. M., B Simpson, A. G., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., et al.** (2012). The Revised Classification of Eukaryotes Eukaryotic Microbiology. *J. Eukaryot. Microbiol* **59**, 429–493. doi/10.1111/j.1550-7408.2012.00644.x.
- Allen, S. E. and Nowacki, M.** (2017). Necessity Is the Mother of Invention: Ciliates, Transposons, and Transgenerational Inheritance. *Trends in Genetics* **33**, 197–207. doi/10.1016/j.tig.2017.01.005.
- Allen, S. E., Hug, I., Pabian, S., Rzeszutek, I., Hoehener, C. and Nowacki, M.** (2017). Circular Concatemers of Ultra-Short DNA Segments Produce Regulatory RNAs. *Cell* **168**, 990-999.e7. doi/10.1016/j.cell.2017.02.020.
- Alzohairy, A. M., Gyulai, G., Jansen, R. K. and Bahieldin, A.** (2013). Transposable elements domesticated and neofunctionalized by eukaryotic genomes. *Plasmid* **69**, 1–15. doi/10.1016/j.plasmid.2012.08.001.
- Arambasic, M., Sandoval, P. Y., Hoehener, C., Singh, A. and Swart, E. C.** (2014). Pdsg1 and Pdsg2, Novel Proteins Involved in Developmental Genome Remodelling in *Paramecium*. *PLoS One* **10**, doi/10.1371/journal.pone.0112899.
- Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.-M. M., Denby Wilkes, C., Garnier, O., Labadie, K., Lauderdale, B. E., le Mouël, A., et al.** (2012). The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLoS Genet* **8**, e1002984. doi/10.1371/journal.pgen.1002984.
- Arnaiz, O., van Dijk, E., Bétermier, M., Lhuillier-Akakpo, M., de Vanssay, A., Duharcourt, S., Sallet, E., Gouzy, J. and Sperling, L.** (2017). Improved methods and resources for *paramecium* genomics: transcription units, gene annotation and gene expression. *BMC Genomics* **18**, 483. doi/10.1186/s12864-017-3887-z.

- Arnould, C., Rocher, V., Finoux, A.-L., Clouaire, T., Li, K., Zhou, F., Caron, P., Mangeot, Philippe. E., Ricci, E. P., Mourad, R., et al. (2021). Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature* **590**, 660–665. doi/10.1038/s41586-021-03193-z.
- Aufderheide, K. J. (1986). Clonal aging in *Paramecium tetraurelia*. II. Evidence of functional changes in the macronucleus with age. *Mech Ageing Dev* **37**, 265–279. doi/10.1016/0047-6374(86)90044-8.
- Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B. M., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., et al. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**, 171–178. doi/10.1038/nature05230.
- Bailey, T. L. (2021). STREME: Accurate and versatile sequence motif discovery. *Bioinformatics* **37**, 2834–2840. doi/10.1093/BIOINFORMATICS/BTAB203.
- Baroin, A., Prat, A. and Caron, F. (1987). Telomeric site position heterogeneity in macronuclear DNA of *Paramecium primaurelia*. *Nucleic Acids Res* **15**, 1717–1728. doi/10.1093/nar/15.4.1717.
- Baudry, C., Malinsky, S., Restituto, M., Kapusta, A., Rosa, S., Meyer, E. and Bétermier, M. (2009). PiggyMac, a domesticated *piggyBac* transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev* **23**, 2478–2483. doi/10.1101/gad.547309.
- Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J. R., et al. (2010). Silencing Specific *Paramecium tetraurelia* Genes by Feeding Double-Stranded RNA. *Cold Spring Harb Protoc* **2010**, pdb.prot5363. doi/10.1101/pdb.prot5363.
- Bellush, J. M. and Whitehouse, I. (2017). DNA replication through a chromatin environment. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20160287. doi/10.1098/rstb.2016.0287.
- Berger, J. D. (1973). *Nuclear Differentiation and Nucleic Acid Synthesis in Well-fed Exconjugants of Paramecium aurelia*.

- Berger, J. D. and Schmidt, H. J.** (1978). Regulation of macronuclear DNA content in *Paramecium tetraurelia*. *Journal of Cell Biology* **76**, 116–126. doi/10.1083/jcb.76.1.116.
- Betermier, M. and Duharcourt, S.** (2014). Programmed Rearrangement in Ciliates: *Paramecium*. *Microbiol Spectr* **2**, MDNA-0035-2014. doi/10.1128/microbiolspec.MDNA3-0035-2014.
- Bétermier, M., Duharcourt, S., Seitz, H. and Meyer, E.** (2000). Timing of Developmentally Programmed Excision and Circularization of *Paramecium* Internal Eliminated Sequences. *Mol Cell Biol* **20**, 1553–1561. doi/10.1128/MCB.20.5.1553-1561.2000.
- Bétermier, M., Bertrand, P. and Lopez, B. S.** (2014). Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet* **10**, e1004086. doi/10.1371/journal.pgen.1004086.
- Bétermier, M., Borde, V. and de Villartay, J.-P.** (2020). Coupling DNA Damage and Repair: an Essential Safeguard during Programmed DNA Double-Strand Breaks? *Trends Cell Biol* **30**, 87–96. doi/10.1016/j.tcb.2019.11.005.
- Bhullar, S., Denby Wilkes, C., Arnaiz, O., Nowacki, M., Sperling, L. and Meyer, E.** (2018). A mating-type mutagenesis screen identifies a zinc-finger protein required for specific DNA excision events in *Paramecium*. *Nucleic Acids Res* **46**, 9550–9562. doi/10.1093/nar/gky772.
- Bischerour, J., Bhullar, S., Denby Wilkes, C., Régnier, V., Mathy, N., Dubois, E., Singh, A., Swart, E., Arnaiz, O., Sperling, L., et al.** (2018). Six domesticated PiggyBac transposases together carry out programmed DNA elimination in *Paramecium*. *Elife* **7**, 1–24. doi/10.7554/eLife.37927.
- Borodin, P., Chen, A., Forstmeier, W., Fouché, S., Malinovskaya, L., Pei, Y., Reifová, R., Ruiz-Ruano, F. J., Schlebusch, S. A., Sotelo-Muñoz, M., et al.** (2022). Mendelian nightmares: the germline-restricted chromosome of songbirds. *Chromosome Research* **30**, 255–272. doi/10.1007/s10577-022-09688-3.

- Bouallègue, M., Rouault, J.-D. D., Hua-Van, A., Makni, M. and Capy, P. (2017).** Molecular Evolution of piggyBac Superfamily: From Selfishness to Domestication. *Genome Biol Evol* **9**, 323–339. doi/10.1093/gbe/evw292.
- Bouhouche, K., Gout, J.-F. F., Kapusta, A., Bétermier, M. and Meyer, E. (2011).** Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res* **39**, 4249–64. doi/10.1093/nar/gkq1283.
- Bouhouche, K., le Borgne, P., Lemullois, M. and Tassin, A. M. (2021).** La paramécie, un organisme modèle pour étudier la ciliogenèse et les maladies ciliaires. *médecine/sciences* **37**, 632–638. doi/10.1051/MEDSCI/2021087.
- Bourgain, F. M. and Katinka, M. D. (1991).** Telomeres inhibit end to end fusion and enhance maintenance of linear DNA molecules injected into the *Paramecium primaurelia* macronucleus. *Nucleic Acids Res* **19**, 1541–1547. doi/10.1093/nar/19.7.1541.
- Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. M. and Sherlock, G. (2004).** GO::TermFinder - Open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20**, 3710–3715. doi/10.1093/BIOINFORMATICS/BTH456.
- Bruguière, J. G. (1791).** *Tableau encyclopédique et méthodique des trois règnes de la nature*. Paris: Panckoucke doi/tableauencyclopd1791brug.
- Burki, F., Roger, A. J., Brown, M. W. and Simpson, A. G. B. (2020).** The New Tree of Eukaryotes. *Trends Ecol Evol* **35**, 43–55. doi/10.1016/j.tree.2019.08.008.
- Callen, A. M., Adoutte, A., Andrew, J. M., Baroin-Tourancheau, A., Bré, M. H., Ruiz, P. C., Clérot, J. C., Delgado, P., Fleury, A. and Jeanmaire-Wolf, R. (1994).** Isolation and characterization of libraries of monoclonal antibodies directed against various forms of tubulin in *Paramecium*. *Biol Cell* **81**, 95–119. doi/10.1016/s0248-4900(94)80002-2.
- Capy, P. (2021).** Taming, Domestication and Exaptation: Trajectories of Transposable Elements in Genomes. *Cells* **10**, 3590. doi/10.3390/CELLS10123590.

- Caron, F. and Meyer, E.** (1985). Does *Paramecium primaurelia* use a different genetic code in its macronucleus? *Nature* **314**, 185–188. doi/10.1038/314185a0.
- Catania, F., Wurmser, F., Potekhin, A. A., Przybos, E. and Lynch, M.** (2009). Genetic Diversity in the *Paramecium aurelia* Species Complex. *Mol Biol Evol* **26**, 421–431. doi/10.1093/molbev/msn266.
- Ceccaldi, R., Rondinelli, B. and D’Andrea, A. D.** (2016). Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol* **26**, 52–64. doi/10.1016/j.tcb.2015.07.009.
- Cervantes, M. D., Coyne, R. S., Xi, X. and Yao, M.-C.** (2006). The Condensin Complex Is Essential for Amitotic Segregation of Bulk Chromosomes, but Not Nucleoli, in the Ciliate *Tetrahymena thermophila*. *Mol Cell Biol* **26**, 4690–4700. doi/10.1128/MCB.02315-05.
- Chalker, D. L. and Yao, M. C.** (1996). Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of *Tetrahymena thermophila* with germ line-limited DNA. *Mol Cell Biol* **16**, 3658–3667. doi/10.1128/MCB.16.7.3658.
- Chen, Q., Luo, W., Veach, R. A., Hickman, A. B., Wilson, M. H. and Dyda, F.** (2020). Structural basis of seamless excision and specific targeting by piggyBac transposase. *Nat Commun* **11**, 3446. doi/10.1038/s41467-020-17128-1.
- Cheng, C.-Y., Vogt, A., Mochizuki, K. and Yao, M.-C.** (2010). A Domesticated piggyBac Transposase Plays Key Roles in Heterochromatin Dynamics and DNA Cleavage during Programmed DNA Deletion in *Tetrahymena thermophila*. *Mol Biol Cell* **21**, 1753–1762. doi/10.1091/mbc.E09.
- Cheng, C.-Y. Y., Young, J. M., Lin, C.-Y. G. Y. G., Chao, J.-L. L., Malik, H. S. and Yao, M.-C. C.** (2016). The piggyBac transposon-derived genes TPB1 and TPB6 mediate essential transposon-like excision during the developmental rearrangement of key genes in *Tetrahymena thermophila*. *Genes Dev* **30**, 2724–2736. doi/10.1101/gad.290460.116.
- Christie, S. M., Fijen, C. and Rothenberg, E.** (2022). V(D)J Recombination: Recent Insights in Formation of the Recombinase Complex and Recruitment of DNA

- Repair Machinery. *Front Cell Dev Biol* **10**, 886718. doi/10.3389/fcell.2022.886718.
- Chuong, E. B., Elde, N. C. and Feschotte, C.** (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* **18**, 71–86. doi/10.1038/nrg.2016.139.
- Cosby, R. L., Chang, N. C. and Feschotte, C.** (2019). Host–transposon interactions: Conflict, cooperation, and cooption. *Genes Dev* **33**, 1098–1116. doi/10.1101/gad.327312.119.
- Coyne, R. S., Lhuillier-Akakpo, M. and Duharcourt, S.** (2012). RNA-guided DNA rearrangements in ciliates: Is the best genome defence a good offence? *Biol Cell* **104**, 309–325. doi/10.1111/boc.201100057.
- Davidson, I. F. and Peters, J. M.** (2021). Genome folding through loop extrusion by SMC complexes. *Nat Rev Mol Cell Biol* **22**, 445–464. doi/10.1038/s41580-021-00349-7.
- Deep, A., Gu, Y., Gao, Y.-Q., Ego, K. M., Herzik, M. A., Zhou, H. and Corbett, K. D.** (2022). The SMC-family Wadjet complex protects bacteria from plasmid transformation by recognition and cleavage of closed-circular DNA. *Mol Cell* **82**, 4145-4159.e7. doi/10.1016/j.molcel.2022.09.008.
- Déléris, A., Berger, F. and Duharcourt, S.** (2021). Role of Polycomb in the control of transposable elements. *Trends in Genetics* **37**, 882–889. doi/10.1016/j.tig.2021.06.003.
- de Vanssay, A., Touzeau, A., Arnaiz, O., Frapporti, A., Phipps, J. and Duharcourt, S.** (2020). The Paramecium histone chaperone Spt16-1 is required for Pgm endonuclease function in programmed genome rearrangements. *PLoS Genet* **16**, e1008949. doi/10.1371/journal.pgen.1008949.
- di Stefano, L.** (2022). All Quiet on the TE Front? The Role of Chromatin in Transposable Element Silencing. *Cells* **11**, 2501. doi/10.3390/cells11162501.
- Drews, F., Salhab, A., Karunanithi, S., Cheaib, M., Jung, M., Schulz, M. H. and Simon, M.** (2022a). Broad domains of histone marks in the highly compact

- Paramecium* macronuclear genome. *Genome Res* **32**, 710–725. doi/10.1101/gr.276126.121.
- Drews, F., Boenigk, J. and Simon, M. (2022b). *Paramecium* epigenetics in development and proliferation. *Journal of Eukaryotic Microbiology* **69**, 1–25. doi/10.1111/jeu.12914.
- Drotos, K. H. I., Zagoskin, M. V., Kess, T., Gregory, T. R. and Wyngaard, G. A. (2022). Throwing away DNA: programmed downsizing in somatic nuclei. *Trends in Genetics* **38**, 483–500. doi/10.1016/j.tig.2022.02.003.
- Dubois, E., Bischerour, J., Marmignon, A., Mathy, N., Régnier, V. and Bétermier, M. (2012). Transposon Invasion of the *Paramecium* Germline Genome Countered by a Domesticated PiggyBac Transposase and the NHEJ Pathway. *Int J Evol Biol* **2012**, 1–13. doi/10.1155/2012/436196.
- Dubois, E., Mathy, N., Régnier, V., Bischerour, J., Baudry, C., Trouslard, R. and Bétermier, M. (2017). Multimerization properties of PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements. *Nucleic Acids Res* **45**, gkw1359. doi/10.1093/nar/gkw1359.
- Duharcourt, S. and Sperling, L. (2018). The Challenges of Genome-Wide Studies in a Unicellular Eukaryote With Two Nuclear Genomes. In *Methods in Enzymology*, pp. 101–126. Elsevier Inc. doi/10.1016/bs.mie.2018.08.012.
- Duharcourt, S., Butler, A. and Meyer, E. (1995). Epigenetic self-regulation of developmental excision of an internal eliminated sequence on *Paramecium tetraurelia*. *Genes Dev* **9**, 2065–2077. doi/10.1101/gad.9.16.2065.
- Duharcourt, S., Keller, A.-M. and Meyer, E. (1998). Homology-Dependent Maternal Inhibition of Developmental Excision of Internal Eliminated Sequences in *Paramecium tetraurelia*. *Mol Cell Biol* **18**, 7075. doi/10.1128/MCB.18.12.7075.
- Duret, L., Cohen, J., Jubin, C., Dessen, P., Goût, J.-F., Mousset, S., Aury, J.-M., Jaillon, O., Noël, B., Arnaiz, O., et al. (2008). Analysis of sequence variability in the macronuclear DNA of *Paramecium tetraurelia*: A somatic view of the germline. *Genome Res* **18**, 585–596. doi/10.1101/gr.074534.107.

- Feng, L., Wang, G., Hamilton, E. P., Xiong, J., Yan, G., Chen, K., Chen, X., Dui, W., Plemens, A., Khadr, L., et al. (2017). A germline-limited piggyBac transposase gene is required for precise excision in *Tetrahymena* genome rearrangement. *Nucleic Acids Res* **45**, 9481–9502. doi/10.1093/nar/gkx652.
- Forney, J. D. and Blackburn, E. H. (1988). Developmentally controlled telomere addition in wild-type and mutant paramecia. *Mol Cell Biol* **8**, 251–258. doi/10.1128/mcb.8.1.251-258.1988.
- Forney, J. D., Epstein, L. M., Preer, L. B., Rudman, B. M., Widmayer, D. J., Klein, W. H. and Preer, J. R. (1983). Structure and expression of genes for surface proteins in *Paramecium*. *Mol Cell Biol* **3**, 466–474. doi/10.1128/mcb.3.3.466-474.1983.
- Frapporti, A., Miró Pina, C., Arnaiz, O., Holoch, D., Kawaguchi, T., Humbert, A., Eleftheriou, E., Lombard, B., Loew, D., Sperling, L., et al. (2019). The Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in *Paramecium*. *Nat Commun* **10**, 2710. doi/10.1038/s41467-019-10648-5.
- Furrer, D. I., Swart, E. C., Kraft, M. F., Sandoval, P. Y. and Nowacki, M. (2017). Two Sets of Piwi Proteins Are Involved in Distinct sRNA Pathways Leading to Elimination of Germline-Specific DNA. *Cell Rep* **20**, 505–520. doi/10.1016/j.celrep.2017.06.050.
- Galvani, A. and Sperling, L. (2002). RNA interference by feeding in *Paramecium*. *Trends in Genetics* **18**, 11–12. doi/10.1016/S0168-9525(01)02548-3.
- Ganji, M., Shaltiel, I. A., Bisht, S., Kim, E., Kalichava, A., Haering, C. H. and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. *Science (1979)* **360**, 102–105. doi/10.1126/science.aar7831.
- Garnier, O., Serrano, V., Duhaucourt, S. and Meyer, E. (2004). RNA-Mediated Programming of Developmental Genome Rearrangements in *Paramecium tetraurelia*. *Mol Cell Biol* **24**, 7370–7379. doi/10.1128/MCB.24.17.7370-7379.2004.

- Gilbert, C., Peccoud, J. and Cordaux, R.** (2021). Transposable Elements and the Evolution of Insects. *Annu. Rev. Entomol.* 2021 **66**, 355–372. doi/10.1146/annurev-ento-070720.
- Gilley, D. and Blackburn, E. H.** (1994). Lack of telomere shortening during senescence in Paramecium. *Proceedings of the National Academy of Sciences* **91**, 1955–1958. doi/10.1073/pnas.91.5.1955.
- Gnan, S., Matelot, M., Weiman, M., Arnaiz, O., Guérin, F., Sperling, L., Bétermier, M., Thermes, C., Chen, C.-L. and Duharcourt, S.** (2022). GC content, but not nucleosome positioning, directly contributes to intron splicing efficiency in *Paramecium*. *Genome Res* **32**, 699–709. doi/10.1101/gr.276125.121.
- Godau, J., Ferretti, L. P., Trenner, A., Dubois, E., von Aesch, C., Marmignon, A., Simon, L., Kapusta, A., Guérois, R., Bétermier, M., et al.** (2019). Identification of a miniature Sae2/Ctp1/CtIP ortholog from *Paramecium tetraurelia* required for sexual reproduction and DNA double-strand break repair. *DNA Repair (Amst)* **77**, 96–108. doi/10.1016/j.dnarep.2019.03.011.
- Godiska, R., Aufderheide, K. J., Gilley, D., Hendrie, P., Fitzwater, T., Preer, L. B., Polisky, B. and Preer, J. R.** (1987). Transformation of *Paramecium* by microinjection of a cloned serotype gene. *Proceedings of the National Academy of Sciences* **84**, 7590–7594. doi/10.1073/pnas.84.21.7590.
- Goerner-Potvin, P. and Bourque, G.** (2018). Computational tools to unmask transposable elements. *Nat Rev Genet* **19**, 688–704. doi/10.1038/s41576-018-0050-x.
- Gotz, U., Marker, S., Cheaib, M., Andresen, K., Shrestha, S., Durai, D. A., Nordstrom, K. J., Schulz, M. H. and Simon, M.** (2016). Two sets of RNAi components are required for heterochromatin formation in trans triggered by truncated transgenes. *Nucleic Acids Res* **44**, 5908–5923. doi/10.1093/nar/gkw267.
- Gout, J. F., Kahn, D. and Duret, L.** (2010). The Relationship among Gene Expression, the Evolution of Gene Dosage, and the Rate of Protein Evolution. *PLoS Genet* **6**, 20. doi/10.1371/JOURNAL.PGEN.1000944.

- Gout, J.-F., Johri, P., Arnaiz, O., Doak, T. G., Bhullar, S., Couloux, A., Guérin, F., Malinsky, S., Sperling, L., Labadie, K., et al. (2019). Universal trends of post-duplication evolution revealed by the genomes of 13 *Paramecium* species sharing an ancestral whole-genome duplication. *bioRxiv* 573576. doi/10.1101/573576.
- Grandchamp, S. and Beisson, J. (1981). Positional control of nuclear differentiation in paramecium. *Dev Biol* **81**, 336–341. doi/10.1016/0012-1606(81)90297-9.
- Grant, C. E., Bailey, T. L. and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017. doi/10.1093/BIOINFORMATICS/BTR064.
- Gratias, A. and Bétermier, M. (2003). Processing of Double-Strand Breaks Is Involved in the Precise Excision of *Paramecium* Internal Eliminated Sequences. *Mol Cell Biol* **23**, 7152–7162. doi/10.1128/MCB.23.20.7152-7162.2003.
- Grattepanche, J., Walker, L. M., Ott, B. M., Paim Pinto, D. L., Delwiche, C. F., Lane, C. E. and Katz, L. A. (2018). Microbial Diversity in the Eukaryotic SAR Clade: Illuminating the Darkness Between Morphology and Molecular Data. *BioEssays* **40**, 1700198. doi/10.1002/bies.201700198.
- Gruchota, J., Denby Wilkes, C., Arnaiz, O., Sperling, L. and Nowak, J. K. (2017). A meiosis-specific Spt5 homolog involved in non-coding transcription. *Nucleic Acids Res* **45**, 4722–4732. doi/10.1093/nar/gkw1318.
- Guérin, F., Arnaiz, O., Boggetto, N., Denby Wilkes, C., Meyer, E., Sperling, L. and Duharcourt, S. (2017). Flow cytometry sorting of nuclei enables the first global characterization of *Paramecium* germline DNA and transposable elements. *BMC Genomics* **18**, 327. doi/10.1186/s12864-017-3713-7.
- Guérineau, M., Bessa, L., Moriau, S., Lescop, E., Bontems, F., Mathy, N., Guittet, E., Bischerour, J., Bétermier, M. and Morellet, N. (2021). The unusual structure of the PiggyMac cysteine-rich domain reveals zinc finger diversity in PiggyBac-related transposases. *Mob DNA* **12**, 12. doi/10.1186/s13100-021-00240-4.

- Hamilton, E. P., Kapusta, A., Huvos, P. E., Bidwell, S. L., Zafar, N., Tang, H., Hadjithomas, M., Krishnakumar, V., Badger, J. H., Caler, E. v, et al. (2016). Structure of the germline genome of *Tetrahymena thermophila* and relationship to the massively rearranged somatic genome. *Elife* **5**, e19090. doi/10.7554/eLife.19090.
- Hammond, C. M., Strømme, C. B., Huang, H., Patel, D. J. and Groth, A. (2017). Histone chaperone networks shaping chromatin function. *Nat Rev Mol Cell Biol* **18**, 141–158. doi/10.1038/nrm.2016.159.
- Hardy, A., Matelot, M., Touzeau, A., Klopp, C., Lopez-Roques, C., Duharcourt, S. and Defrance, M. (2021). DNAModAnnot: a R toolbox for DNA modification filtering and annotation. *Bioinformatics* **37**, 2738–2740. doi/10.1093/bioinformatics/btab032.
- Hartl, T. A., Sweeney, S. J., Knepler, P. J. and Bosco, G. (2008a). Condensin II Resolves Chromosomal Associations to Enable Anaphase I Segregation in *Drosophila* Male Meiosis. *PLoS Genet* **4**, 1000228. doi/10.1371/journal.pgen.1000228.
- Hartl, T. A., Smith, H. F. and Bosco, G. (2008b). Chromosome alignment and transvection are antagonized by condensin II. *Science (1979)* **322**, 1384–1387. doi/10.1126/SCIENCE.1164216/SUPPL_FILE/HARTL.SOM.PDF.
- Henikoff, S., Henikoff, J. G., Kaya-Okur, H. S. and Ahmad, K. (2020). Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. *Elife* **9**, e63274. doi/10.7554/eLife.63274.
- Hill, J. (1752). *An history of animals: descriptions of the birds, beasts, fishes, and insects, of the several parts of the world, and including accounts of the several classes of animalcules, visible only by the assistance of microscopes*. London: Printed for Thomas Osborne doi/historyanimals00hill.
- Hirano, T. (2012). Condensins: universal organizers of chromosomes with diverse functions. *Genes Dev* **26**, 1659–1678. doi/10.1101/gad.194746.112.
- Hof, A. E. van't, Campagne, P., Rigden, D. J., Yung, C. J., Lingley, J., Quail, M. A., Hall, N., Darby, A. C. and Saccheri, I. J. (2016). The industrial melanism

- mutation in British peppered moths is a transposable element. *Nature* **534**, 102–105. doi/10.1038/nature17951.
- Howard-Till, R. and Loidl, J.** (2018). Condensins promote chromosome individualization and segregation during mitosis, meiosis, and amitosis in *Tetrahymena thermophila*. *Mol Biol Cell* **29**, 466–478. doi/10.1091/mbc.E17-07-0451.
- Howard-Till, R., Tian, M. and Loidl, J.** (2019). A specialized condensin complex participates in somatic nuclear maturation in *Tetrahymena thermophila*. *Mol Biol Cell* **30**, 1326–1338. doi/10.1091/mbc.E18-08-0487.
- Ignarski, M., Singh, A., Swart, E. C., Arambasic, M., Sandoval, P. Y. and Nowacki, M.** (2014). Paramecium tetraurelia chromatin assembly factor-1-like protein PtCAF-1 is involved in RNA-mediated control of DNA elimination. *Nucleic Acids Res* **42**, 11952–64. doi/10.1093/nar/gku874.
- Jones, J. M. and Gellert, M.** (2004). The taming of a transposon: V(D)J recombination and the immune system. *Immunol Rev* **200**, 233–248. doi/10.1111/j.0105-2896.2004.00168.x.
- Kapusta, A., Matsuda, A., Marmignon, A., Ku, M., Silve, A., Meyer, E., Forney, J. D., Malinsky, S. and Bétermier, M.** (2011). Highly Precise and Developmentally Programmed Genome Assembly in Paramecium Requires Ligase IV-Dependent End Joining. *PLoS Genet* **7**, e1002049. doi/10.1371/journal.pgen.1002049.
- Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K. and Henikoff, S.** (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930. doi/10.1038/s41467-019-09982-5.
- Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K. and Henikoff, S.** (2020). Efficient low-cost chromatin profiling with CUT&Tag. *Nat Protoc* **15**, 3264–3283. doi/10.1038/s41596-020-0373-x.
- Klobutcher, L. A. and Herrick, G.** (1995). Consensus inverted terminal repeat sequence of Paramecium IESs: resemblance to termini of Tc1-related and

- Euplotes Tec transposons. *Nucleic Acids Res* **23**, 2006–13. doi/10.1093/nar/23.11.2006.
- Klobutcher, L. A. and Herrick, G.** (1997). Developmental genome reorganization in ciliated protozoa: the transposon link. *Prog Nucleic Acid Res Mol Biol* **56**, 1–62. doi/10.1016/s0079-6603(08)61001-6.
- Kloc, M., Kubiak, J. Z. and Ghobrial, R. M.** (2022). Natural genetic engineering: A programmed chromosome/DNA elimination. *Dev Biol* **486**, 15–25. doi/10.1016/j.ydbio.2022.03.008.
- Koizumi, S. and Kobayashi, S.** (1989). Microinjection of plasmid DNA encoding the A surface antigen of *Paramecium tetraurelia* restores the ability to regenerate a wild-type macronucleus. *Mol Cell Biol* **9**, 4398–4401. doi/10.1128/mcb.9.10.4398-4401.1989.
- Kong, M., Cutts, E. E., Pan, D., Beuron, F., Kaliyappan, T., Xue, C., Morris, E. P., Musacchio, A., Vannini, A. and Greene, E. C.** (2020). Human Condensin I and II Drive Extensive ATP-Dependent Compaction of Nucleosome-Bound DNA. *Mol Cell* **79**, 99-114.e9. doi/10.1016/j.molcel.2020.04.026.
- Kono, N. and Arakawa, K.** (2019). Nanopore sequencing: Review of potential applications in functional genomics. *Dev Growth Differ* **61**, 316–326. doi/10.1111/dgd.12608.
- Kurosaki, T., Popp, M. W. and Maquat, L. E.** (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol* **20**, 406–420. doi/10.1038/s41580-019-0126-2.
- le Mouël, A., Butler, A., Caron, F. and Meyer, E.** (2003). Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia. *Eukaryot Cell* **2**, 1076–90. doi/10.1128/ec.2.5.1076-1090.2003.
- Lepère, G., Bétermier, M., Meyer, E. and Duharcourt, S.** (2008). Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev* **22**, 1501–1512. doi/10.1101/gad.473008.

- Lepère, G., Nowacki, M., Serrano, V., Gout, J.-F., Guglielmi, G., Duharcourt, S. and Meyer, E. (2009). Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res* **37**, 903–915. doi/10.1093/nar/gkn1018.
- Lhuillier-Akakpo, M., Frapporti, A., Denby Wilkes, C., Matelot, M., Vervoort, M., Sperling, L. and Duharcourt, S. (2014). Local effect of enhancer of zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements. *PLoS Genet* **10**, e1004665. doi/10.1371/journal.pgen.1004665.
- Lhuillier-Akakpo, M., Guérin, F., Frapporti, A. and Duharcourt, S. (2016). DNA deletion as a mechanism for developmentally programmed centromere loss. *Nucleic Acids Res* **44**, 1553–65. doi/10.1093/nar/gkv1110.
- Lin, I.-T., Chao, J.-L. and Yao, M.-C. (2012). An essential role for the DNA breakage-repair protein Ku80 in programmed DNA rearrangements in *Tetrahymena thermophila*. *Mol Biol Cell* **23**, 2213–2225. doi/10.1091/mbc.e11-11-0952.
- Lis, M. and Walther, D. (2016). The orientation of transcription factor binding site motifs in gene promoter regions: does it matter? *BMC Genomics* **17**, 185. doi/10.1186/s12864-016-2549-x.
- Lisch, D. (2013). How important are transposons for plant evolution? *Nat Rev Genet* **14**, 49–61. doi/10.1038/nrg3374.
- Liu, C., Zhang, Y., Liu, C. C. and Schatz, D. G. (2022a). Structural insights into the evolution of the RAG recombinase. *Nat Rev Immunol* **22**, 353–370. doi/10.1038/s41577-021-00628-6.
- Liu, H. W., Roisné-Hamelin, F., Beckert, B., Li, Y., Myasnikov, A. and Gruber, S. (2022b). DNA-measuring Wadjet SMC ATPases restrict smaller circular plasmids by DNA cleavage. *Mol Cell* **82**, 4727-4740.e6. doi/10.1016/j.molcel.2022.11.015.
- Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550. doi/10.1186/s13059-014-0550-8.

- Luo, Z., Hu, T., Jiang, H., Wang, R., Xu, Q., Zhang, S., Cao, J. and Song, X.** (2020). Rearrangement of macronucleus chromosomes correspond to TAD-like structures of micronucleus chromosomes in *Tetrahymena thermophila*. *Genome Res* **30**, 406–414. doi/10.1101/gr.241687.118.
- Maliszewska-Olejniczak, K., Gruchota, J., Gromadka, R., Denby Wilkes, C., Arnaiz, O., Mathy, N., Duharcourt, S., Bétermier, M. and Nowak, J. K.** (2015). TFIIIS-Dependent Non-coding Transcription Regulates Developmental Genome Rearrangements. *PLoS Genet* **11**, e1005383. doi/10.1371/journal.pgen.1005383.
- Marmignon, A., Bischerour, J., Silve, A., Fojcik, C., Dubois, E., Arnaiz, O., Kapusta, A., Malinsky, S. and Bétermier, M.** (2014). Ku-Mediated Coupling of DNA Cleavage and Repair during Programmed Genome Rearrangements in the Ciliate *Paramecium tetraurelia*. *PLoS Genet* **10**, e1004552. doi/10.1371/journal.pgen.1004552.
- Matityahu, A. and Onn, I.** (2021). Hit the brakes – a new perspective on the loop extrusion mechanism of cohesin and other SMC complexes. *J Cell Sci* **134**, jcs.247577. doi/10.1242/jcs.247577.
- Matsuda, A., Shieh, A. W.-Y., Chalker, D. L. and Forney, J. D.** (2010). The Conjugation-Specific Die5 Protein Is Required for Development of the Somatic Nucleus in both *Paramecium* and *Tetrahymena*. *Eukaryot Cell* **9**, 1087–1099. doi/10.1128/EC.00379-09.
- Maupetit-Mehouas, S. and Vaury, C.** (2020). Transposon Reactivation in the Germline May Be Useful for Both Transposons and Their Host Genomes. *Cells* **9**, 1172. doi/10.3390/cells9051172.
- May, D. G., Scott, K. L., Campos, A. R. and Roux, K. J.** (2020). Comparative Application of BioID and TurboID for Protein-Proximity Biotinylation. *Cells* **9**, 1070. doi/10.3390/cells9051070.
- McCormick-Graham, M.** (1997). Variable telomeric repeat synthesis in *Paramecium tetraurelia* is consistent with misincorporation by telomerase. *EMBO J* **16**, 3233–3242. doi/10.1093/emboj/16.11.3233.

- McCormick-Graham, M. and Romero, D. P.** (1996). A single telomerase RNA is sufficient for the synthesis of variable telomeric DNA repeats in ciliates of the genus *Paramecium*. *Mol Cell Biol* **16**, 1871–1879. doi/10.1128/MCB.16.4.1871.
- McGrath, C. L., Gout, J.-F., Doak, T. G., Yanagi, A. and Lynch, M.** (2014a). Insights into Three Whole-Genome Duplications Gleaned from the *Paramecium caudatum* Genome Sequence. *Genetics* **197**, 1417–1428. doi/10.1534/genetics.114.163287.
- McGrath, C. L., Gout, J.-F., Johri, P., Doak, T. G. and Lynch, M.** (2014b). Differential retention and divergent resolution of duplicate genes following whole-genome duplication. *Genome Res* **24**, 1665–1675. doi/10.1101/gr.173740.114.
- Meers, M. P., Bryson, T. D., Henikoff, J. G. and Henikoff, S.** (2019). Improved CUT&RUN chromatin profiling tools. *Elife* **8**, 1–16. doi/10.7554/eLife.46314.
- Meyer, E. and Keller, A. M.** (1996). A Mendelian mutation affecting mating-type determination also affects developmental genomic rearrangements in *Paramecium tetraurelia*. *Genetics* **143**, 191–202.
- Miller, R. v, Neme, R., Clay, D. M., Pathmanathan, J. S., Lu, M. W., Yerlici, V. T., Khurana, J. S. and Landweber, L. F.** (2021). Transcribed germline-limited coding sequences in *Oxytricha trifallax*. *G3 Genes|Genomes|Genetics* **11**, jkab092. doi/10.1093/g3journal/jkab092.
- Miró-Pina, C., Charmant, O., Kawaguchi, T., Holoch, D., Michaud, A., Cohen, I., Humbert, A., Jaszczyszyn, Y., Chevreux, G., del Maestro, L., et al.** (2022). *Paramecium* Polycomb repressive complex 2 physically interacts with the small RNA-binding PIWI protein to repress transposable elements. *Dev Cell* **57**, 1037-1052.e8. doi/10.1016/j.devcel.2022.03.014.
- Mitra, R., Fain-Thornton, J. and Craig, N. L.** (2008). piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J* **27**, 1097–1109. doi/10.1038/emboj.2008.41.
- Mochizuki, K., Fine, N. A., Fujisawa, T. and Gorovsky, M. A.** (2002). 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.

- Müller, O. F.** (1773). *Vermivm terrestrium et fluviatilium, seu, Animalium infusoriorum, helminthicorum et testaceorum, non marinorum, succincta historia.* et Lipsiae: apud Heineck et Faber doi/10.5962/bhl.title.46299.
- Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J. and Chang, H. Y.** (2016). HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* **13**, 919–922. doi/10.1038/nmeth.3999.
- Nanney, D. L.** (1957). Mating Type Inheritance at Conjugation in Variety 4 of *Paramecium aurelia*. *J Protozool* **4**, 89–95. doi/10.1111/j.1550-7408.1957.tb02492.x.
- Noto, T. and Mochizuki, K.** (2017). Whats, hows and whys of programmed DNA elimination in Tetrahymena. *Open Biol* **7**, 170172. doi/10.1098/rsob.170172.
- Nowacki, M., Zagorski-Ostojka, W. and Meyer, E.** (2005). Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in *Paramecium tetraurelia*. *Curr Biol* **15**, 1616–28. doi/10.1016/j.cub.2005.07.033.
- Nowak, J. K., Gromadka, R., Juszczyk, M., Jerka-Dziadosz, M., Maliszewska, K., Mucchielli, M.-H., Gout, J.-F., Arnaiz, O., Agier, N., Tang, T., et al.** (2011). Functional Study of Genes Essential for Autogamy and Nuclear Reorganization in *Paramecium*. *Eukaryot Cell* **10**, 363–372. doi/10.1128/EC.00258-10.
- Ohno, S.** (1970). *Evolution by Gene Duplication*. Springer Berlin Heidelberg doi/10.1007/978-3-642-86659-3.
- Oldenkamp, R. and Rowland, B. D.** (2022). A walk through the SMC cycle: From catching DNAs to shaping the genome. *Mol Cell* **82**, 1616–1630. doi/10.1016/j.molcel.2022.04.006.
- Orias, E., Singh, D. P. and Meyer, E.** (2017). Genetics and Epigenetics of Mating Type Determination in *Paramecium* and *Tetrahymena*. *Annu Rev Microbiol* **71**, 133–156. doi/10.1146/annurev-micro-090816-093342.
- Owsian, D., Gruchota, J., Arnaiz, O. and Nowak, J. K.** (2022). The transient Spt4-Spt5 complex as an upstream regulator of non-coding RNAs during development. *Nucleic Acids Res* **50**, 2603–2620. doi/10.1093/NAR/GKAC106.

- Philosophical transactions** (1704). London: Royal Society of London
doi/philosophicaltra2317royal.
- Picelli, S., Björklund, A. K., Reinius, B., Sagasser, S., Winberg, G. and Sandberg, R.** (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24**, 2033–40. doi/10.1101/gr.177881.114.
- Preer, J. R.** (1976). Quantitative predictions of random segregation models of the ciliate macronucleus. *Genet Res* **27**, 227–238. doi/10.1017/S0016672300016426.
- Preer, J. R., Preer, L. B., Rudman, B. M. and Barnett, A. J.** (1985). Deviation from the universal code shown by the gene for surface protein 51A in *Paramecium*. *Nature* **314**, 188–190. doi/10.1038/314188a0.
- Preer, L. B., Hamilton, G. and Preer, J. R.** (1992). Micronuclear DNA from *Paramecium tetraurelia*: Serotype 51 A Gene Has Internally Eliminated Sequences. *J. Protozool* **39**, 678–682.
- Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., et al.** (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680. doi/10.1016/j.cell.2014.11.021.
- Rogers, C. H., Mielczarek, O. and Corcoran, A. E.** (2021). Dynamic 3D Locus Organization and Its Drivers Underpin Immunoglobulin Recombination. *Front Immunol* **11**, 633705. doi/10.3389/fimmu.2020.633705.
- Ruehle, M. D., Orias, E. and Pearson, C. G.** (2016). *Tetrahymena* as a Unicellular Model Eukaryote: Genetic and Genomic Tools. *Genetics* **203**, 649–665. doi/10.1534/genetics.114.169748.
- Sandoval, P. Y., Swart, E. C., Arambasic, M. and Nowacki, M.** (2014). Functional diversification of Dicer-like proteins and small RNAs required for genome sculpting. *Dev Cell* **28**, 174–88. doi/10.1016/j.devcel.2013.12.010.
- Sawka-Gądek, N., Potekhin, A., Singh, D. P., Grevtseva, I., Arnaiz, O., Penel, S., Sperling, L., Tarcz, S., Duret, L., Nekrasova, I., et al.** (2021). Evolutionary

- Plasticity of Mating-Type Determination Mechanisms in *Paramecium aurelia* Sibling Species. *Genome Biol Evol* **13**, evaa258. doi/10.1093/gbe/evaa258.
- Sellis, D., Guérin, F., Arnaiz, O., Pett, W., Lerat, E., Boggetto, N., Krenek, S., Berendonk, T., Couloux, A., Aury, J.-M., et al. (2021). Massive colonization of protein-coding exons by selfish genetic elements in *Paramecium* germline genomes. *PLoS Biol* **19**, e3001309. doi/10.1371/journal.pbio.3001309.
- Senti, K.-A., Jurczak, D., Sachidanandam, R. and Brennecke, J. (2015). piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. *Genes Dev* **29**, 1747–62. doi/10.1101/gad.267252.115.
- Shieh, A. W. Y. and Chalker, D. L. (2013). LIA5 Is Required for Nuclear Reorganization and Programmed DNA Rearrangements Occurring during *Tetrahymena* Macronuclear Differentiation. *PLoS One* **8**, e75337. doi/10.1371/journal.pone.0075337.
- Singh, D. P., Saudemont, B., Guglielmi, G., Arnaiz, O., Goût, J.-F. F., Prajer, M., Potekhin, A., Przybòs, E., Aubusson-Fleury, A., Bhullar, S., et al. (2014). Genome-defence small RNAs exapted for epigenetic mating-type inheritance. *Nature* **509**, 447–52. doi/10.1038/nature13318.
- Singh, A., Vancura, A., Woycicki, R. K., Hogan, D. J., Hendrick, A. G. and Nowacki, M. (2018). Determination of the presence of 5-methylcytosine in *Paramecium tetraurelia*. *PLoS One* **13**, e0206667. doi/10.1371/journal.pone.0206667.
- Singh, A., Maurer-Alcalá, X. X., Solberg, T., Häußermann, L., Gisler, S., Ignarski, M., Swart, E. C. and Nowacki, M. (2022). Chromatin remodeling is required for sRNA-guided DNA elimination in *Paramecium*. *EMBO J* **41**, e111839. doi/10.15252/embj.2022111839.
- Skouri, F. and Cohen, J. (1997). Genetic Approach to Regulated Exocytosis Using Functional Complementation in *Paramecium*: Identification of the ND7 Gene Required for Membrane Fusion. *Mol Biol Cell* **8**, 1063–1071.

- Solé-Soler, R. and Torres-Rosell, J.** (2020). Smc5/6, an atypical SMC complex with two RING-type subunits. *Biochem Soc Trans* **48**, 2159–2171. doi/10.1042/BST20200389.
- Sonneborn, T. M.** (1937). Sex, Sex Inheritance and Sex Determination in Paramecium Aurelia. *Proceedings of the National Academy of Sciences* **23**, 378–385. doi/10.1073/pnas.23.7.378.
- Sonneborn, T. M.** (1974). Paramecium aurelia. In *Handbook of Genetics*, pp. 469–594. Boston, MA: Springer US doi/10.1007/978-1-4684-2994-7_20.
- Steele, C. J., Barkocy-Gallagher, G. A., Preer, L. B. and Preer, J. R.** (1994). Developmentally excised sequences in micronuclear DNA of Paramecium. *Proceedings of the National Academy of Sciences* **91**, 2255–2259. doi/10.1073/pnas.91.6.2255.
- Swart, E. C., Wilkes, C. D., Sandoval, P. Y., Arambasic, M., Sperling, L. and Nowacki, M.** (2014). Genome-wide analysis of genetic and epigenetic control of programmed DNA deletion. *Nucleic Acids Res* **42**, 8970–8983. doi/10.1093/NAR/GKU619.
- Swart, E. C., Denby Wilkes, C., Sandoval, P. Y., Hoehener, C., Singh, A., Furrer, D. I., Arambasic, M., Ignarski, M. and Nowacki, M.** (2017). Identification and analysis of functional associations among natural eukaryotic genome editing components. *F1000Res* **6**, 1374. doi/10.12688/f1000research.12121.1.
- Tikhonenkov, D. v., Strassert, J. F. H., Janouškovec, J., Mylnikov, A. P., Aleoshin, V. v., Burki, F. and Keeling, P. J.** (2020). Predatory colponemids are the sister group to all other alveolates. *Mol Phylogenet Evol* **149**, 106839. doi/10.1016/J.YMPEV.2020.106839.
- Timmons, L., Court, D. L. and Fire, A.** (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**, 103–112. doi/10.1016/S0378-1119(00)00579-5.
- Trancoso, I., Morimoto, R. and Boehm, T.** (2020). Co-evolution of mutagenic genome editors and vertebrate adaptive immunity. *Curr Opin Immunol* **65**, 32–41. doi/10.1016/J.COI.2020.03.001.

- Tucker, J. B., Beisson, J., Roche, D. L. and Cohen, J.** (1980). Microtubules and control of macronuclear 'amitosis' in *Paramecium*. *J Cell Sci* **44**, 135–151. doi/10.1242/jcs.44.1.135.
- Van Leewenhoek, A.** (1676). Observations, communicated to the publisher concerning little animals by him observed in Rain-Well-Sea- and Snow water; as also in water wherein pepper had lain infused. In *Philosophical Transactions* , pp. 821–831. The Royal Society doi/10.1098/RSTL.1677.0003.
- Vijayanathan, M., Trejo-Arellano, M. G. and Mozgová, I.** (2022). Polycomb Repressive Complex 2 in Eukaryotes—An Evolutionary Perspective. *Epigenomes* **6**, 3. doi/10.3390/epigenomes6010003.
- Vogt, A. and Mochizuki, K.** (2013). A Domesticated PiggyBac Transposase Interacts with Heterochromatin and Catalyzes Reproducible DNA Elimination in *Tetrahymena*. *Tetrahymena* **9**, 1004032. doi/10.1371/journal.pgen.1004032.
- Wang, J. and Davis, R. E.** (2014). Programmed DNA elimination in multicellular organisms. *Curr Opin Genet Dev* **27**, 26–34. doi/10.1016/j.gde.2014.03.012.
- Wang, Y., Wang, Y., Sheng, Y., Huang, J., Chen, X., Al-Rasheid, K. A. S. and Gao, S.** (2017). A comparative study of genome organization and epigenetic mechanisms in model ciliates, with an emphasis on *Tetrahymena*, *Paramecium* and *Oxytricha*. *Eur J Protistol* **61**, 376–387. doi/10.1016/j.ejop.2017.06.006.
- Wang, Y., Wu, L. and Yuen, K. W. Y.** (2022a). The roles of transcription, chromatin organisation and chromosomal processes in holocentromere establishment and maintenance. *Semin Cell Dev Biol* **127**, 79–89. doi/10.1016/j.semcdb.2022.01.004.
- Wang, C., Solberg, T., Maurer-Alcalá, X. X., Swart, E. C., Gao, F. and Nowacki, M.** (2022b). A small RNA-guided PRC2 complex eliminates DNA as an extreme form of transposon silencing. *Cell Rep* **40**, 111263. doi/10.1016/j.celrep.2022.111263.
- Wignall, S. M., Deehan, R., Maresca, T. J. and Heald, R.** (2003). The condensin complex is required for proper spindle assembly and chromosome

- segregation in *Xenopus* egg extracts. *Journal of Cell Biology* **161**, 1041–1051. doi/10.1083/JCB.200303185/VIDEO-6.
- Xu, X., Zha, S., Ma, T., Zhou, P.-K., Yue, X., Bai, C. and Xie, D. (2020). DNA-PKcs: A Multi-Faceted Player in DNA Damage Response. *Front. Genet* **11**, 607428. doi/10.3389/fgene.2020.607428.
- Yao, M.-C., Chao, J.-L. and Cheng, C.-Y. (2014). Programmed Genome Rearrangements in *Tetrahymena*. *Microbiol Spectr* **2**, MDNA3-0061–2014. doi/10.1128/microbiolspec.MDNA3-0012-2014.
- Yatskevich, S., Rhodes, J. and Nasmyth, K. (2019). Organization of Chromosomal DNA by SMC Complexes. *Annu Rev Genet* **53**, 445–482. doi/10.1146/annurev-genet-112618-043633.
- Yerlici, V. T. and Landweber, L. F. (2014). Programmed Genome Rearrangements in the Ciliate *Oxytricha*. *Microbiol Spectr* **2**, 389–407. doi/10.1128/microbiolspec.MDNA3-0025-2014.
- Yi, Z., Sanjeev, M. and Singh, G. (2021). The Branched Nature of the Nonsense-Mediated mRNA Decay Pathway. *Trends in Genetics* **37**, 143–159. doi/10.1016/j.tig.2020.08.010.
- Zahid, S., Seif, M., Dahan, E., Iehl, F., Fernandez-Varela, P., le Du, M.-H., Ropars, V., Charbonnier, J. B., Timmins, J. and Ravanat, J.-L. (2021). Molecular Sciences The Multifaceted Roles of Ku70/80. *Int. J. Mol. Sci* **22**, 4134. doi/10.3390/ijms22084134.
- Zangarelli, C., Arnaiz, O., Bourge, M., Gorrichon, K., Jaszczyszyn, Y., Mathy, N., Escoriza, L., Bétermier, M. and Régnier, V. (2022). Developmental timing of programmed DNA elimination in *Paramecium tetraurelia* recapitulates germline transposon evolutionary dynamics. *Genome Res* **32**, 2028–2042. doi/10.1101/gr.277027.122.
- Zhang, Y., Cheng, T. C., Huang, G., Lu, Q., Surleac, M. D., Mandell, J. D., Pontarotti, P., Petrescu, A. J., Xu, A., Xiong, Y., et al. (2019). Transposon molecular domestication and the evolution of the RAG recombinase. *Nature* **569**, 79–84. doi/10.1038/s41586-019-1093-7.

Zhang, Y., Parmigiani, G. and Johnson, W. E. (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* **2**, 1–10. doi/10.1093/nargab/lqaa078.

Zhao, X., Xiong, J., Mao, F., Sheng, Y., Chen, X., Feng, L., Dui, W., Yang, W., Kapusta, A., Feschotte, C., et al. (2019). RNAi-dependent Polycomb repression controls transposable elements in *Tetrahymena*. *Genes Dev* **33**, 348–364. doi/10.1101/gad.320796.118.