



HAL
open science

Caractérisation des protéines “ Plant Mobile Domain ” dans la régulation de l’expression des gènes et la répression des éléments transposables chez *Arabidopsis thaliana*

Mélody Nicolau

► **To cite this version:**

Mélody Nicolau. Caractérisation des protéines “ Plant Mobile Domain ” dans la régulation de l’expression des gènes et la répression des éléments transposables chez *Arabidopsis thaliana*. Sciences agricoles. Université de Perpignan, 2020. Français. NNT : 2020PERP0019 . tel-03106200

HAL Id: tel-03106200

<https://theses.hal.science/tel-03106200>

Submitted on 11 Jan 2021

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.

THÈSE

Pour obtenir le grade de
Docteur

Délivré par
UNIVERSITE DE PERPIGNAN VIA DOMITIA

Préparée au sein de l'école doctorale
« Energie et Environnement » ED305
Et de l'unité de recherche
Laboratoire Génome et Développement des Plantes

Spécialité : Biologie

Présentée par Mélody Nicolau

**Caractérisation des protéines « Plant Mobile
Domain » dans la régulation de l'expression des
gènes et la répression des éléments transposables
chez *Arabidopsis thaliana***

Soutenue le 15 octobre 2020 devant le jury composé de

Cécile Raynaud, DR, IPS2, Paris
Hervé Vaucheret, DR, IJPB, INRAE UMR 1318
Christoph Grunau, PR, IHPE, UMR 5244
Christel Carles, MdC, UGA, CNRS UMR 5168
Angélique Déléris, CR, IBENS, Paris
Marie Mirouze, CR, IRD, Montpellier
Guillaume Moissiard, CR, LGDP, UMR 5096

Rapportrice
Rapporteur
Examinateur
Examinatrice
Examinatrice
Examinatrice
Directeur de thèse

REMERCIEMENTS

Je tiens tout d'abord à remercier chaleureusement tous les membres de mon jury : **Cécile Raynaud, Hervé Vaucheret, Christoph Grunau, Christel Carles, Angélique Déléris** et **Marie Mirouze** pour avoir accepté de juger mon travail et pour l'intérêt qu'ils y ont porté.

Je remercie à nouveau **Christel**, mais également **Thierry Lagrange** et **Antoine Martin** pour avoir accepté de faire partie de mon CSI, vos savoirs et vos encouragements ont été d'une grande aide pour faire progresser ce projet jusqu'à aujourd'hui.

Je remercie également **Jean-Marc Deragon** pour m'avoir accueillie au sein du LGDP, mais également pour sa collaboration précieuse dans ce projet. Et je remercie aussi **Olivier Panaud**, qui a su gérer la reprise du laboratoire après le confinement (non sans l'aide précieuse de l'équipe des déconfinées !) et grâce à qui j'ai pu reprendre mes expériences dès que l'on a eu le droit de remettre le nez dehors.

Je remercie nos super agents administratifs **Myriam, Jean-René** et **Elisabeth** sans qui nous serions bien perdus ! Et je remercie tout particulièrement **Elisabeth** qui a dû plus d'une fois venir me harceler pour des papiers à signer ou à remplir, je te remercie pour ta patience et ta bonne humeur.

Un énorme merci à tous les membres de l'équipe MEAC : **Ariadna** (bon courage pour la suite !) pour ton énergie et ta bonne humeur, **Fred** pour tous les papiers que tu as signé pour moi (haha !) mais surtout pour ton aide, que ce soit à la paillasse ou pour la théorie ! Et surtout merci à **Julie, Nathalie** et **Guillaume** sans qui rien n'aurait été possible. Merci **Julie** pour toutes ces qPCR que tu as enchaîné sans faiblir, pour toute ton aide au labo, pour la piscine, les rires... Merci **Nathalie** pour ton expertise bioinfo et tes corrections, sans toi la majorité de ce manuscrit n'existerait pas, merci pour ta gentillesse et ta bonne humeur. Enfin, merci **Guillaume**, merci pour tout ce que tu m'as appris, pour ta patience, merci de m'avoir permis de venir prendre part à ce projet et merci pour ton soutien même à distance !

Merci aussi aux autres personnes du labo : **Michèle, Edouard, Sophie, Christel, Anne, Pascale, Viviane**.. Pour toute l'aide et la bonne humeur que vous m'avez apporté tout au long de ces trois ans. Merci **Dom** et **Jacinthe** pour votre expertise en chromatine et en protéines, je vous dois énormément ! Merci **JJ, Christophe R., Christophe B., Rémy, MC** et tous ceux que j'ai déjà cités pour les conversations animées du midi ! Nombre d'entre elles resteront dans ma mémoire !

Merci à tous les non-permanents du labo : **Celso, Panpan** mais aussi **Jean-Loup, Avilien, Vichet** les « petits nouveaux » du labo même si vous entamez déjà votre deuxième année (le temps passe trop vite), merci pour les conversations plus ou moins intellectuelles, bonne continuation à vous tous, et bon courage ! Merci à **Tommy** pour ta bonne humeur et tes blagues, et merci d'avoir gardé toutes ces images de Klaus sur ta paillasse, ça me fait toujours rire quand je passe devant. Merci à **Emilie** et **Marc-Alexandre** d'avoir été des voisins de bureau géniaux et surtout tolérants à nos délires avec **Alicia**, je sais que ça n'a pas toujours été facile ! La troisième année vous tend les bras mais vous aller y arriver, pas de panique !

Merci à **Laura** et **Alicia**, pour m'avoir fait découvrir plein de nouvelles choses au cours de ces trois années, comme le sport (mais pas trop) ou l'avion ! Mais aussi pour avoir partagé avec moi énormément de choses que ce soient les fous rires, les blagues et les gâteaux ! Merci **Alicia** pour m'avoir accueillie si souvent chez toi (et Zelda pour les câlins) le temps d'un soir ou d'un week-end pour regarder des films en bonne compagnie. Et merci **Laura** pour l'invitation à San Diego, ce voyage a été fou !

Et en parlant de bonne compagnie, un énorme merci à **Eléonore, Mégane, Stephen, Mélina, Chrystelle, Christophe, Nélia, Sylvain** pour toutes ces super soirées, JDR ou autre, je suis tellement heureuse de vous avoir rencontrés ! Vous mériteriez tous un paragraphe entier pour vous remercier mais je n'aurais jamais assez de place ici, mais pas d'inquiétude, ce n'est que partie remise ! Un petit merci spécial tout de même au gang Ecossois **Eléonore, Mégane, Steph** et **Mélina**, on remet ça quand vous voulez !

Merci à mes amis en dehors de Perpignan, des amis de longue date et que je retrouve dès que je peux en Auvergne. Merci **Jennifer** d'être toujours là après 16 ans d'amitié (déjà !), merci à **Romain** et **Jérôme** de répondre toujours présents quand je leur dis que je remonte du sud !

Enfin, je tiens à adresser un merci tout particulier à ma famille pour avoir toujours été là pour moi, même à 430 km de distance, merci à vous tous. Merci **Papa**, merci **Maman**, merci **Maryline** sans vous je ne serais pas là à écrire ces mots (aussi bien au sens figuré qu'au sens propre), c'est à vous que je dois tout, alors merci, merci d'avoir été là pour me soutenir et m'encourager tout au long de ma vie ! Merci aussi au reste de la famille, **Romain, Arthur** et **Logan** bien sûr ! Merci à mes **Tata**, à mon **Tonton**, merci **Lydie** et merci **Mémé** !

Table des matières

INTRODUCTION	1
I) L'organisation de l'ADN au sein du noyau eucaryote	2
I.1) La chromatine.....	2
I.2) Le nucléosome et les protéines histones	2
I.3) Les différents niveaux de compaction de la chromatine	3
I.3.1) L'hétérochromatine constitutive	3
I.3.2) L'euchromatine	3
I.3.3) L'hétérochromatine facultative.....	4
I.3.4) Les paysages épigénétiques	4
II) Les modifications épigénétiques.....	5
II.1) Définition de l'épigénétique.....	5
II.2) Les différents acteurs épigénétiques	5
II.3) La méthylation de l'ADN	6
II.3.1) Les différents contextes de méthylation de l'ADN	6
II.3.2) La répartition de la méthylation de l'ADN	7
II.3.3) La mise en place et le maintien de la méthylation de l'ADN chez les mammifères	7
II.3.4) La mise en place et le maintien de la méthylation de l'ADN chez <i>A. thaliana</i>	9
II.3.5) La déméthylation de l'ADN	11
II.4) Les protéines histones.....	13
II.4.1) Les modifications post-traductionnelles (MPTs) des histones.....	13
II.4.2) L'acétylation des histones	13
II.4.3) La méthylation des histones	14
II.4.5) Les autres MPTs des histones	16
II.4.6) Les variants d'histones	17
II.5) Fonctions biologiques des modifications épigénétiques	24
II.5.1) impact des modifications épigénétiques sur l'organisation de la chromatine dans le noyau	24
II.5.2) La régulation de l'expression des gènes et des TEs	27
II.5.3) Modifications épigénétiques et développement.....	29
II.5.4) modifications épigénétiques et réponse aux stress.....	31
III) Facteurs épigénétiques	31
III.1) Les complexes remodeleurs de la chromatine	32

III.2) MORPHEUS' MOLECULE 1 (MOM1)	33
III.3) La famille des protéines MICRORCHIDIA (MORC)	34
III.4) Les ARN non-codants (ARNnc)	35
III.5) le « mille-feuille » épigénétique.....	36
IV) Les éléments transposables (TEs)	37
IV.1) Découverte des TEs.....	37
IV.2) Les différentes classes d'élément transposables.....	38
IV.2.1) Classe I : les rétrotransposons	38
IV.2.2) Classe II : les transposons	39
IV.3) La contribution des éléments transposables dans le génome	40
V) La génétique classique	41
V.1) Principe général	42
V.2) Le système <i>ATCOPIA28::GFP</i>	42
V.3) Résultats du crible.....	43
VI) La famille des gènes Plant Mobile Domain (PMD).....	44
VI.1) Origine du domaine protéique	44
VI.2) Les gènes PMD chez <i>A. thaliana</i>	45
VI.2.1) Les gènes MAIN et MAIL1	46
VII) Les facteurs de transcription (FTs).....	48
VII.1) Les différents types de FTs.....	49
VII.1.1) Les FTs pionniers	49
VII.1.2) Les FTs spécifiques	50
VII.1.3) Les FTs généraux	52
VII.2) Interaction entre les facteurs de transcription et les modifications épigénétiques	52
VIII) Les protéines phosphatases.....	53
VIII.1) Les sous-familles PP1 à PP6	54
VIII.2) La sous-famille PP7	55
VIII.3) Rôle des phosphatases dans la chromatine.....	56
VIII.3.1) Condensation de la chromatine au cours de la mitose	56
VIII.3.2) La réparation de l'ADN.....	57
VIII.3.3) La réplication de l'ADN.....	57
VIII.3.4) La régulation de la transcription	58
IX) Problématique et objectif de thèse	59
RESULTATS.....	60
CHAPITRE 1	61
CHAPITRE 2	97

I) Le gène <i>MORC1</i> est réprimé dans les mutants <i>pmd</i> et <i>pp7l-2</i>	97
II) Étude des gènes surexprimés dans les mutants <i>pmd</i>	99
II.1) <i>FLOWERING LOCUS C</i> , un répresseur de la floraison	99
II.2) <i>Taf4b</i> , un facteur impliqué dans la formation des crossing-over	101
II.3) Conclusions	102
Chapitre III	103
I) Etude de la localisation subcellulaire de MAIN et MAIL1	103
II) Etude de l'interaction de MAIN et MAIL1 avec la chromatine	103
III) Conclusions.....	105
Chapitre VI.....	106
I) MAIL2	106
II) MAIL3	107
III) Conclusions.....	110
DISCUSSION	111
I) La formation du complexe MAIN/MAIL1/PP7L.....	112
II) Les protéines MAIN, MAIL1 et PP7L régulent un même réseau de gènes.....	114
III) Le mode d'action du complexe MAIN/MAIL1/PP7L.....	116
IV) Interaction génétique et rôle des protéines PP7L et MAIL3.....	118
V) Impact du complexe MAIN/MAIL1/PP7L sur la structure de la chromatine	119
VI) MAIL2 est une protéine essentielle pour le développement de la plante	121
VII) Le rôle du PMD chez les plantes	121
MATERIEL ET METHODES	123
BIBLIOGRAPHIE.....	132

INTRODUCTION

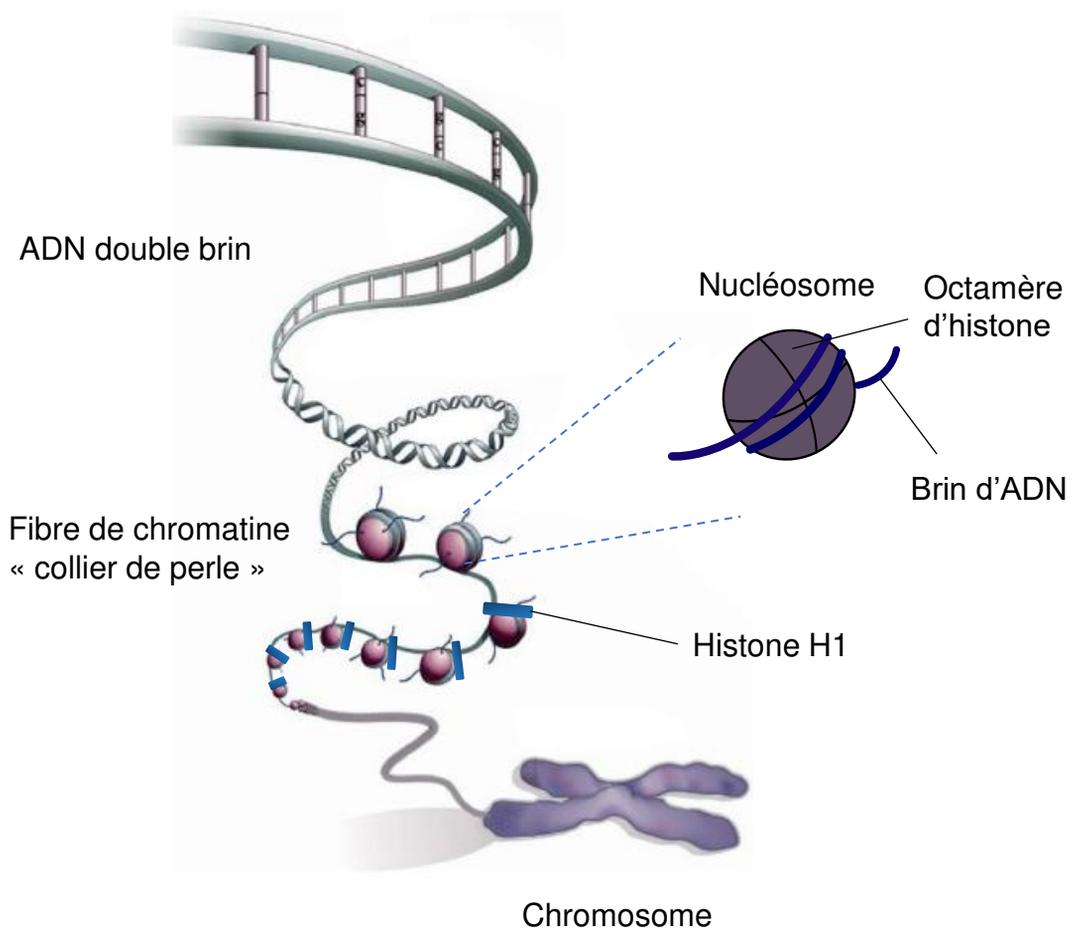


Figure 1 : Les différents niveaux de compaction de la chromatine.

Un brin de 146 pb d'ADN s'enroule autour d'un octamère d'histone (H2A-H2B-H3-H4) pour former le nucléosome. L'assemblage de ces nucléosomes sur l'ADN forme la structure chromatinienne dite en « collier de perles » et la fixation de l'histone H1 sur l'ADN « linker » va permettre la condensation de cette structure. La condensation de la chromatine atteint sa forme la plus condensée lors de la formation des chromosomes au cours des processus de mitose et méiose. Modifiée de Maleszewska and Kaminska, 2013.

I) L'organisation de l'ADN au sein du noyau eucaryote

I.1) La chromatine

Chez les eucaryotes, l'ADN, porteur de l'information génétique, est précisément organisé sous forme de chromatine, un complexe nucléo-protéique permettant la compaction de l'ADN dans le noyau. L'organisation de la chromatine se fait à plusieurs échelles moléculaires (figure 1).

I.2) Le nucléosome et les protéines histones

Le nucléosome constitue le premier niveau d'organisation. Il est composé d'environ 146 paires de base (bp) d'ADN enroulé autour d'un octamère de protéines histones H2A, H2B, H3 et H4, contenant deux copies de chaque protéine (Luger *et al.*, 1997). Le nucléosome est assemblé de manière séquentielle : Les histones H3 et H4 sont déposées en premier sur l'ADN, sous forme de deux dimères ou d'un tétramère, puis les histones H2A et H2B sont déposées sous forme de deux dimères de part et d'autre du tétramère H3-H4 (Arents *et al.*, 1991). Cet assemblage peut se faire *via* deux voies différentes : Une voie couplée à la réplication de l'ADN et une voie indépendante de cette réplication (Tagami *et al.*, 2004). Chaque voie fait intervenir des facteurs spécifiques qui prendront en charge les différentes histones, et seront détaillés plus loin dans ce manuscrit. L'assemblage des nucléosomes le long de l'ADN forme une structure dite en « collier de perles » d'environ 10 nm de diamètre (Fransz and De Jong, 2011). En condition physiologique, cette structure est rapidement condensée en une fibre de 30 nm, grâce notamment à l'histone « linker » H1 qui se lie à l'ADN situé entre deux nucléosomes (appelé ADN « linker »), permettant le rapprochement des nucléosomes (Song *et al.*, 2014). La formation d'une fibre de 30 nm n'est toutefois pas toujours observée *in vivo*, comme cela a été montré chez l'Homme. Dans ce cas-là, les fibres de 10 nm se condensent pour former des structures chromosomiques irrégulières (Nishino *et al.*, 2012).

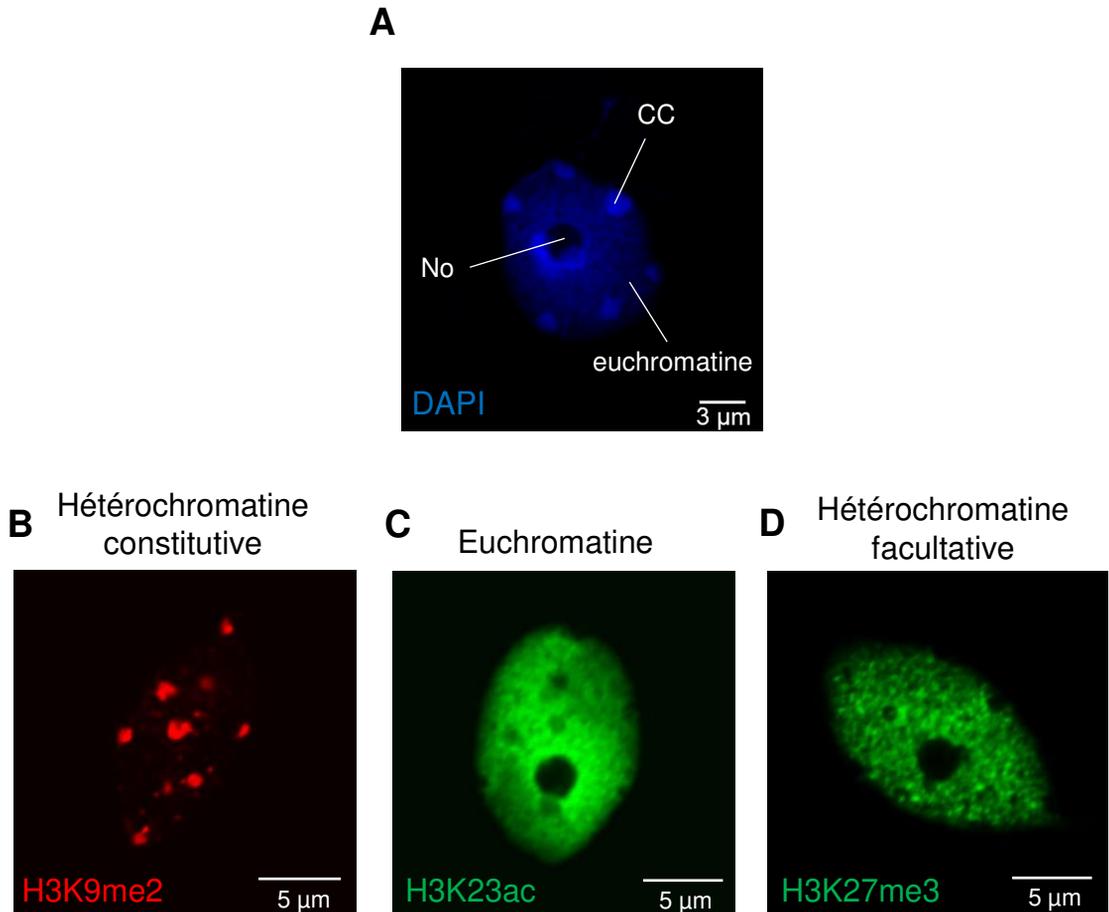


Figure 2 : Les différents états de la chromatine dans le noyau en interphase chez *A. thaliana*.

A. Dans le noyau, l'hétérochromatine constitutive apparaît sous la forme de *foci* d'ADN fortement condensés appelés chromocentres (CCs) qui se concentrent à la périphérie nucléaire ou proche du nucléole (No). L'euchromatine, quant à elle, apparaît plus diffuse au sein du noyau. **B-D.** L'hétérochromatine constitutive est caractérisée par la marque répressive H3K9me2 (**B**), l'euchromatine par des marques permissives comme H3K23ac (**C**) et l'hétérochromatine facultative par la marque répressive H3K27me3 (**D**). DAPI (6'4'-diamidino-2-phénylindole) est un fluorochrome qui permet de détecter l'ADN en se fixant à celui-ci.

I.3) Les différents niveaux de compaction de la chromatine

Dans le noyau en interphase, la chromatine n'est pas uniformément structurée. En microscopie, nous pouvons distinguer deux types principaux de compartiments, ou de régions de la chromatine, montrant un niveau de compaction différent. En 1928, E. Heitz désigne l'hétérochromatine comme les régions chromosomiques restant condensées lors de l'interphase et l'oppose à l'euchromatine qui désigne, quant à elle, les régions décondensées qui deviennent presque invisibles en microscopie (Heitz, 1928; Passarge, 1979). Aujourd'hui, ces termes ont été affinés et se basent à présent sur le niveau de condensation de la chromatine, mais également sur leur composition en gènes et/ou en éléments d'ADN répétés, ainsi que leur enrichissement en marques épigénétiques (voir ci-dessous).

1.3.1) L'hétérochromatine constitutive

L'hétérochromatine constitutive désigne l'état le plus condensé de la chromatine. Chez certains organismes comme la plante modèle *Arabidopsis thaliana*, elle est organisée, pendant l'interphase, sous forme de chromocentres (CCs) (figure 2). L'hétérochromatine constitutive est composée d'éléments d'ADN répétés, incluant les éléments transposables (TEs) qui sont transcriptionnellement réprimés (Grewal and Jia, 2007).

1.3.2) L'euchromatine

L'euchromatine est une forme moins condensée de la chromatine diffuse dans le noyau (figure 2). Les régions euchromatiniennes sont enrichies en gènes et sont plus permissives à la transcription. Ceci se reflète par la présence de marques épigénétiques et facteurs protéiques impliqués dans ce processus cellulaire initiant l'expression des gènes.

1.3.3) L'hétérochromatine facultative

A l'inverse de l'hétérochromatine constitutive, l'hétérochromatine facultative désigne les régions chromatinienne qui ont la capacité de passer d'un état transcriptionnellement réprimé à actif au cours du développement de l'organisme. Cette forme chromatinienne est étroitement liée à la voie Polycomb (PcG) et à des marques épigénétiques répressives spécifiques (figure 2 et voir plus loin). L'hétérochromatine facultative est donc enrichie en gènes transcriptionnellement réprimés dont l'expression est associée à des processus développementaux ou en réponse à des stimuli environnementaux (Trojer and Reinberg, 2007).

1.3.4) Les paysages épigénétiques

Comme nous le verrons tout au long de cette introduction, les nombreuses modifications épigénétiques portées par la chromatine offrent une multitude de combinaisons possibles. Ainsi, au fil des années, cette vision dichotomique hétéro/euchromatine s'est avérée être trop simpliste. Il est à ce jour clairement défini que la chromatine existe sous de nombreuses formes, on parle d'ailleurs de paysages épigénétiques ou « epigenetic landscapes » qui définissent plus subtilement les différents états chromatiniens. Par exemple, chez *Drosophila melanogaster*, l'étude de ces différentes combinaisons a permis de définir neuf paysages épigénétiques (Kharchenko *et al.*, 2011). Le même type d'analyses chez *A. thaliana* a permis de définir d'abord quatre (Roudier *et al.*, 2011), puis neuf paysages épigénétiques (Vergara and Gutierrez, 2017). Ces différents paysages épigénétiques impactent la compaction de la chromatine, ce qui a un effet direct sur l'expression du génome, mais aussi sur tout autre mécanisme cellulaire en relation avec la chromatine.

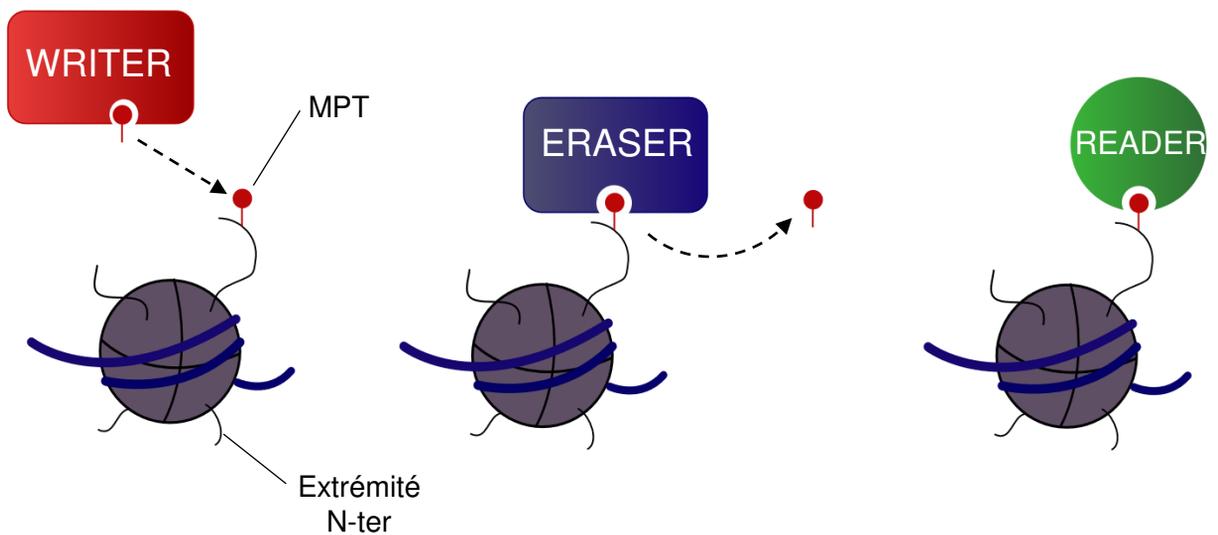


Figure 3 : Les acteurs épigénétiques de la cellule.

Représentation schématiques des protéines « writer », « eraser » et « reader ». Les protéines « writer » et « eraser » sont capables de déposer et de retirer des modifications épigénétiques respectivement. Les protéines « reader » se lient sur les modifications épigénétiques. MPT : Modification post-traductionnelle. Extrémité N-ter : Extrémité N-terminale des histones.

II) Les modifications épigénétiques

II.1) Définition de l'épigénétique

Historiquement, l'épigénétique dérive des termes « epigenesis » et « epigenotype » définis par C.H. Waddington comme une théorie selon laquelle génétique et phénotype sont étroitement interconnectés tout au long du développement d'un organisme et au sein de chaque lignée cellulaire qui le compose. Waddington proposait aussi que l'expression spécifique de gènes était une force majeure promouvant la variation phénotypique et la plasticité cellulaire (Waddington, 1942).

Le terme « épigénétique » a évolué au cours du temps pour faire, à présent, référence à l'étude des mécanismes impliqués dans la modification de l'expression des gènes de manière réversible et héritable tout en maintenant l'intégrité de la séquence d'ADN (absence de mutation). Les principales modifications épigénétiques sont la méthylation de l'ADN et la modification des histones, mais plus globalement, tout mécanisme réversible et héritable impactant la compaction de la chromatine et par extension l'expression des gènes pourrait être défini comme mécanisme épigénétique. Les mécanismes épigénétiques sont essentiels au bon développement de tout organisme, mais ils jouent aussi un rôle important dans toutes les situations impliquant une adaptation rapide de l'expression des gènes, tels qu'en réponse à des stimuli environnementaux et des maladies (Cavalli and Heard, 2019).

II.2) Les différents acteurs épigénétiques

La dynamique des modifications épigénétiques est coordonnée par l'action de protéines spécialisées que l'on peut séparer en trois groupes. Les protéines effectrices dites « writers » sont capables de déposer les marques épigénétiques, et les « erasers » sont, quant à elles, capables de les retirer. Enfin, les protéines dites « readers » reconnaissent et lient spécifiquement les différentes marques épigénétiques (figure 3). Il est courant qu'une protéine « writer » ou « eraser » porte

également un domaine de type « reader ». Cependant, de nombreuses protéines portent uniquement un domaine « reader », jouant ainsi un rôle régulateur des voies épigénétiques en autorisant ou non le recrutement des protéines effectrices.

Dans les parties suivantes, seront décrites les principales modifications épigénétiques connues, impliquant directement la molécule d'ADN et les histones, ainsi que leurs rôles biologiques dans la cellule.

II.3) La méthylation de l'ADN

La méthylation de l'ADN consiste en l'ajout d'un groupement méthyl sur les cytosines à la position du carbone #5 (5mC). Présente chez de nombreux organismes tels que les mammifères, certains insectes (comme le diptère *Apis mellifera*), le poisson zèbre (*Danio rerio*), les plantes à fleurs et les gymnospermes, les algues (comme *Chlamydomonas reinhardtii*), les hépatiques (comme *Marchantia polymorpha*), certains champignons filamenteux (comme *Neurospora crassa*), cette marque épigénétique est absente ou quasiment indétectable chez les levures *Schizosaccharomyces pombe* et *Saccharomyces cerevisiae*, et chez le nématode *Caenorhabditis elegans* (Capuano *et al.*, 2014; Ausin *et al.*, 2016; Montgomery *et al.*, 2020). Finalement, le débat reste ouvert concernant la présence de 5mC chez *Drosophila melanogaster*. En effet, bien que cette marque épigénétique ait pu être détectée, elle reste à des niveaux très faibles en comparaison à d'autres organismes présentant des niveaux de 5mC beaucoup plus élevés. De plus, 5mC apparaît beaucoup moins stable chez *D. melanogaster* ; son potentiel rôle biologique et la machinerie moléculaire impliquée dans la déposition de cette marque restent énigmatique chez ce diptère solitaire (Deshmukh *et al.*, 2018).

II.3.1) Les différents contextes de méthylation de l'ADN

La méthylation de l'ADN peut se faire dans trois contextes de séquence différents : CG, CHG et CHH (où H est soit A, T ou C). Les deux premiers sont dits « symétriques », car les deux brins d'ADN présentent une cytosine méthylée. A l'inverse, la méthylation de type CHH est dite « asymétrique ». Chez les mammifères,

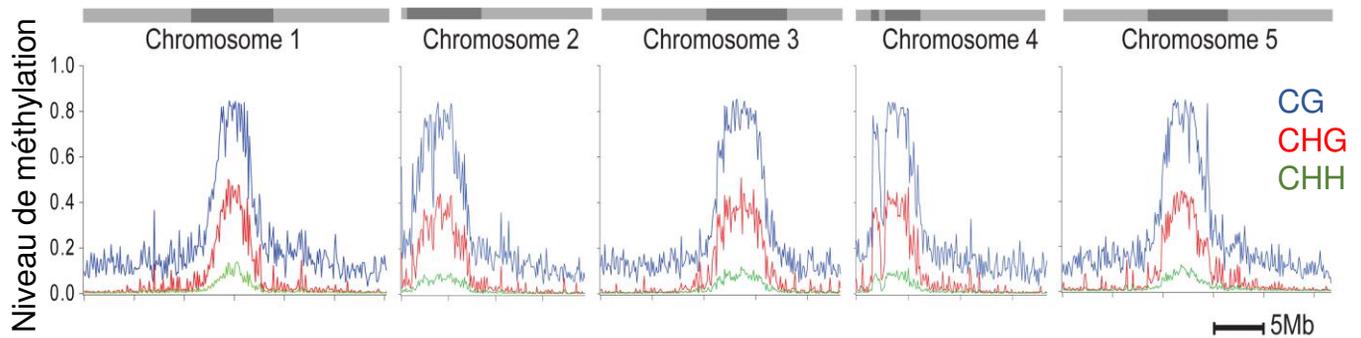


Figure 4 : Répartition de la méthylation de l'ADN chez *A. thaliana*.

Répartition de la méthylation de l'ADN le long des chromosomes chez *A. thaliana*. Les chromosomes sont représentés schématiquement au-dessus des graphiques. Les bras chromosomiques et les régions péri-centromériques sont représentés en gris clair et gris foncé respectivement. Mb : Mégabases. Modifiée de Bernatavichute *et al.*, 2008.

la méthylation de l'ADN se produit presque exclusivement en contexte CG, toutefois de la méthylation de type non-CG a aussi été observée dans les cellules souches embryonnaires (Ramsahoye *et al.*, 2000; Lister *et al.*, 2009). Chez les plantes, l'ADN est méthylé dans les trois contextes de séquence avec respectivement 24% des CG, 6.7% des CHG et 1.7% CHH qui sont méthylés à l'échelle du génome d'*A. thaliana* (Cokus *et al.*, 2008; Lister *et al.*, 2008).

II.3.2) La répartition de la méthylation de l'ADN

La méthylation de l'ADN n'est pas répartie de façon homogène sur les génomes (Cokus *et al.*, 2008 ; Lister *et al.*, 2009). De manière générale, l'ADN hétérochromatinien, composé de séquences répétées comme les TEs, est fortement méthylé, particulièrement au sein des régions péri-centromériques (Cokus *et al.*, 2008 ; Lister *et al.*, 2008, 2009) (figure 4).

Chez les mammifères, la majorité du génome est méthylé dans le contexte CG (Lister *et al.*, 2009). Il existe cependant des régions génomiques non-méthylées, appelées îlots CpG qui sont enrichis dans les régions promotrices de gènes actifs (Lister *et al.*, 2009). Chez *A. thaliana*, différents profils de méthylation peuvent être observés : les régions hétérochromatiniennes péri-centromériques sont fortement méthylées dans les trois contextes de séquences, tandis qu'une méthylation uniquement de type CG peut aussi être observée sur environ 1/3 des gènes actifs. On parle alors de « gene body methylation, GbM » (Zhang *et al.*, 2006; Cokus *et al.*, 2008; Zilberman, 2017).

Le dépôt et la maintenance de 5mC est assuré par plusieurs enzymes spécialisées impliquant des facteurs différents selon les organismes et le contexte de séquence.

II.3.3) La mise en place et le maintien de la méthylation de l'ADN chez les mammifères

Le profil de méthylation du génome des mammifères est établi tôt lors de l'embryogenèse, au moment de l'implantation du blastocyste (Monk *et al.*, 1987; Kafri *et al.*, 1992). Lors de la gamétogenèse, les cellules germinales primordiales (PGCs) subissent deux vagues successives de déméthylation puis de méthylation *de novo* de

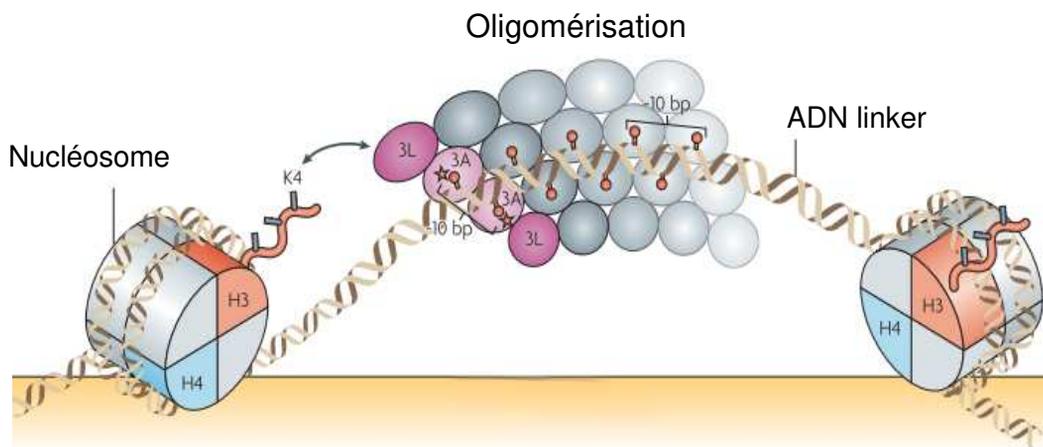


Figure 5 : La méthylation *de novo* chez les mammifères.

Représentation schématique de la méthylation *de novo* chez les mammifères par DNMT3A et DNMT3L. La protéine DNMT3L (3L) est capable d'interagir avec les H3K4 non-méthylées et recrutent DNMT3A (3A) sur l'ADN. DNMT3A et DNMT3L forment alors un tétramère et DNMT3A méthyle l'ADN linker à des intervalles de 10 paires de bases (bp). Les tétramères DNMT3A/3L s'oligomérisent sur l'ADN. Les cercles rouges représentent la méthylation de l'ADN et les étoiles rouges représentent le site actif de DNMT3A. Modifiée de Law and Jacobsen, 2010.

l'ADN (Sasaki and Matsui, 2008; Weber *et al.*, 2012). Cette méthylation *de novo* n'affecte toutefois pas l'intégralité du génome, et certaines régions restent déméthylées, comme les îlots CpG (Bird *et al.*, 1985).

Chez les mammifères, il existe trois méthyltransférases de l'ADN actives : DNA METHYLTRANSFERASE 1 (DNMT1), DNMT3A et DNMT3B. DNMT1 agit principalement dans la maintenance de la méthylation, tandis que DNMT3A et DNMT3B agissent dans la méthylation *de novo* (Bestor *et al.*, 1988; Okano *et al.*, 1998). Cependant, cette séparation des rôles n'est pas absolue, et chaque protéine peut intervenir dans la maintenance ou la méthylation *de novo*, selon le type de séquence cible (Arand *et al.*, 2012). Lors de la méthylation *de novo* des PCGs, DNMT3A et DNMT3B interagissent avec un paralogue inactif DNMT3-like (DNMT3L) (Bourc'his *et al.*, 2001; Ooi *et al.*, 2007). La méthylation de l'ADN serait ciblée sur les gènes et les TEs *via* la reconnaissance de l'histone H3 non-méthylée sur la lysine 4 (H3K4me0) par DNMT3L et DNMT3A (Jia *et al.*, 2007; Ooi *et al.*, 2007; Otani *et al.*, 2009) (figure 5). Ce modèle est soutenu par l'existence d'une exclusion mutuelle entre la méthylation de l'ADN et la méthylation de H3K4 sur certains gènes (Fournier *et al.*, 2002 ; Vu *et al.*, 2004). En plus de ce mécanisme, DNMT3A et DNMT3B peuvent être recrutées sur certains loci par les protéines EUCHROMATIC HISTONE LYSINE METHYLTRANSFERASE 2 (EHMT2) et M-PHASE PHOSPHOPROTEIN 8 (MPP8), des protéines impliquées dans la voie de la méthylation des histones (Epsztejn-Litman *et al.*, 2008; Kokura *et al.*, 2010).

La maintenance de la méthylation de type CG est assurée majoritairement par DNMT1 qui agit au niveau la fourche de réplication de l'ADN pour maintenir la méthylation de l'ADN sur les brins néosynthétisés (Kim *et al.*, 2009). DNMT1 est recrutée par le facteur « Proliferating cell nuclear antigen » (PCNA) et Ubiquitin-like, containing PHD and RING finger domains 1 (UHRF1) (Estève *et al.*, 2006; Bostick *et al.*, 2007; Sharif *et al.*, 2007). UHRF1 reconnaît l'ADN contenant une séquence de type CG hémiméthylée (Bostick *et al.*, 2007). Ainsi, UHFR1 va donc se lier aux brins d'ADN hémiméthylés, produits de la réplication, afin de recruter DNMT1 et permettre la maintenance de la méthylation de l'ADN.

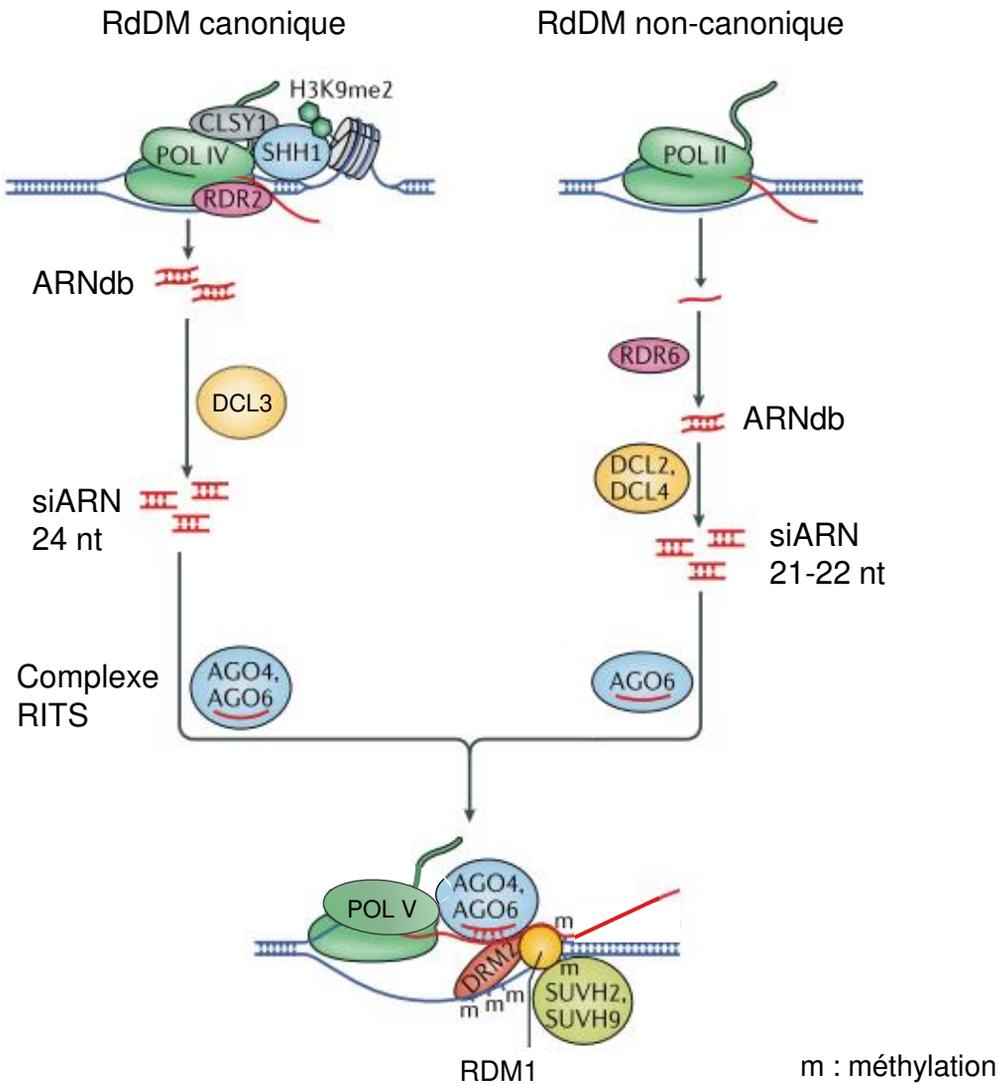


Figure 6 : Les voies du RdDM.

Représentation schématique simplifiée des voies canoniques et non-canoniques du RdDM. Dans la voie canonique, un ARN est transcrit par l'ARN pol IV qui est recrutée par SHH1 et les protéines CLSY1 et 2. L'ARN est ensuite pris en charge par RDR2 pour former un ARN double brin (ARNdb) qui sera clivé par DCL3 pour produire des siARN de 24 nucléotides. Ces siARN seront pris en charge par AGO4 ou AGO6 dans les complexes RITS qui seront recrutés sur les loci cibles grâce aux siARN, à de nombreux acteurs protéiques et conjointement à la transcription d'un long ARN non codant produit par l'ARN pol V. L'ADN des loci cibles sera finalement méthylé par DRM2. Dans la voie non-canonique, l'ARN pol II remplace l'ARN pol IV, RDR6 synthétise les ARNdb qui seront clivés par DCL2/4 pour produire des siARN de 21-22 nucléotides. Ceux-ci seront ensuite pris en charge par le complexe RITS-AGO6 qui sera finalement recruté sur les loci cibles par des mécanismes similaires à ceux de la voie canonique. Modifiée de Zhang *et al.*, 2018.

II.3.4) La mise en place et le maintien de la méthylation de l'ADN chez A. thaliana

La méthylation *de novo* de l'ADN est assurée par la protéine DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), homologue des DNMT3 de mammifères (Cao *et al.*, 2000). Son proche homologue DRM1 contribuerait également à la méthylation *de novo* spécifiquement au cours du développement de l'ovule (Jullien *et al.*, 2012). Plusieurs études ont démontré que la 5mC *de novo* dans les trois contextes de séquence requiert DRM2 et la voie du « RNA-directed DNA methylation (RdDM) » (Law and Jacobsen, 2010; Martínez de Alba *et al.*, 2013; Matzke *et al.*, 2015). Le RdDM est une voie complexe impliquant de nombreux facteurs protéiques dont deux ARN polymérases spécifiques des plantes, l'ARN polymérase IV (ARN Pol IV) et ARN Pol V, conjointement à l'action d'ARNs non codants (ARNnc) tels que les ARNs interférents (siARN) (figure 6). Le RdDM peut être découpé en deux étapes : tout d'abord en amont du RdDM, l'ARN pol IV est recrutée sur ses séquences d'ADN cibles par les protéines SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1), CLASSY 1 (CLSY1) et CLSY2 qui se lient à la marque répressive H3K9me2 (Law *et al.*, 2013; Zhou *et al.*, 2018). Inversement, CLSY3 et CLSY4 agiraient dans une voie indépendante en recrutant l'ARN pol IV par reconnaissance de la méthylation de type CG (Zhou *et al.*, 2018). L'ARN pol IV synthétise alors un ARNnc simple brin utilisé comme matrice par la RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) pour produire un ARN double brin (ARNdb) (Li *et al.*, 2015; Zhai *et al.*, 2015). Cet ARNdb est ensuite clivé par la protéine DICER-LIKE PROTEIN 3 (DCL3) pour donner des siARN de 24 nucléotides (24-nt) (Xie *et al.*, 2004). La seconde étape (en aval du RdDM) implique le recrutement de l'ARN pol V *via* le complexe RNA-induced transcriptional silencing (RITS) composé des siARN 24-nt et de la protéine ARGONAUTE 4 (AGO4) (Zilberman *et al.*, 2003; Law and Jacobsen, 2010). En se liant à de l'ADN méthylé, les protéines SUPPRESSOR OF VARIATION 3-9 HOMOLOG PROTEIN 2 (SUVH2) et SUVH9 jouent également un rôle essentiel dans le recrutement de l'ARN pol V (Johnson *et al.*, 2014). Ensuite, dans le modèle conventionnel, l'ARN Pol V produit alors de longs ARNnc (ARNInc) sur lesquels le complexe RITS serait recruté par similarité de séquence entre les siARN 24-nt et l'ARNInc (Zhang *et al.*, 2018b). Finalement la protéine RNA-DIRECTED DNA METHYLATION 1 (RDM1) servirait de pont entre AGO4 et DRM2 pour permettre le

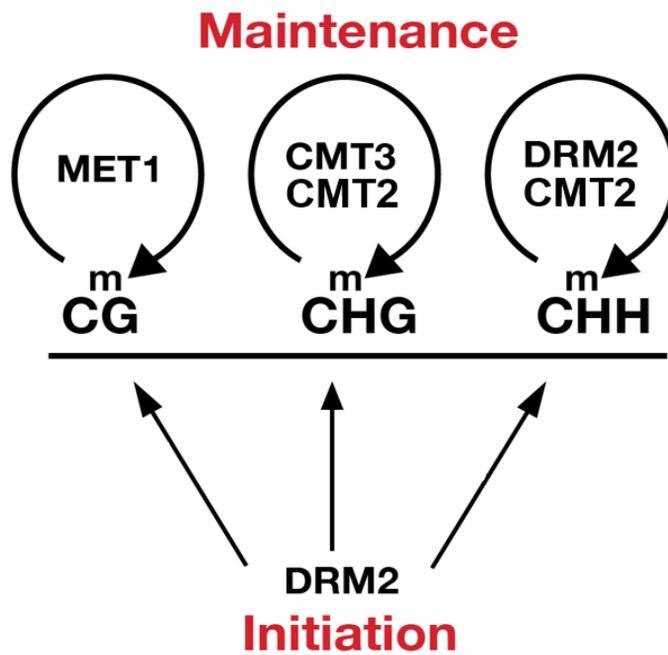


Figure 7 : La maintenance de la méthylation de l'ADN dans les différents contextes de séquence chez *A. thaliana*.

La maintenance de la méthylation de l'ADN est assurée par différentes enzymes selon le contexte de séquence. MET1 pour la méthylation en CG, CMT3 et CMT2 pour CHG et DRM2 et CMT2 pour CHH. Pour ces trois contextes, DRM2 en assure la mise en place (initiation). Modifiée de Cao *et al.*, 2003.

recrutement de cette dernière (Gao *et al.*, 2010; Zhang *et al.*, 2018b). Cependant, des analyses plus récentes réalisées dans l'équipe de Thierry Lagrange suggère un modèle alternatif où le complexe RITS se lierait directement sur l'ADN ouvert de la boucle de transcription réalisée par l'ARN PolIV, ceci afin de recruter directement DRM2 au niveau de l'ADN (Lahmy *et al.*, 2016). En plus de cette voie dite RdDM canonique, il existe une voie non-canonique reposant sur l'action de facteurs différents : l'ARN polymérase II (ARN pol II) produit des ARN potentiellement convertis en ARNdb par RDR6. Les protéines DCL4 et DCL2 produisent ensuite des siARN de 21- ou 22-nt, respectivement, qui seront pris en charge par AGO6 pour recruter les facteurs en aval du RdDM tels que les complexes ARN Pol V et DRM2 (Zhang *et al.*, 2018b). Le RdDM non-canonique cible principalement les TEs transcriptionnellement actifs et les gènes produisant des siARN agissant en *trans* (Nuthikattu *et al.*, 2013; McCue *et al.*, 2015).

La maintenance de la méthylation de l'ADN est assurée par différentes enzymes selon le contexte de séquence (figure 7). La méthylation de type CG est maintenue exclusivement par la protéine METHYLTRANSFERASE 1 (MET1), orthologue de DNMT1 chez les mammifères (Kankel *et al.*, 2003; Cokus *et al.*, 2008). Malgré ce rôle de maintenance, MET1 a aussi montré une légère activité de type *de novo* sur certaines régions géniques actives transcriptionnellement (Zubko *et al.*, 2012). Le recrutement de MET1 se fait par l'action redondante des protéines VARIANT IN METHYLATION 1 (VIM1), VIM2 et VIM3, orthologues de UHRF1, qui comme cette dernière, reconnaissent l'ADN hémiméthylé (Woo *et al.*, 2008; Kim *et al.*, 2014). Ainsi le recrutement des deux homologues MET1 et DNMT1 se fait par des mécanismes moléculaires similaires.

La maintenance de la méthylation de type CHG est catalysée par CHROMOMETHYLASE 3 (CMT3) et CMT2 (Lindroth *et al.*, 2001; Zemach *et al.*, 2013; Stroud *et al.*, 2014). Cette maintenance est fortement liée à la marque H3K9me2, déposée par les histones méthyltransférases (HMTs) SUVH4 ou KRYPTONITE (KYP), SUVH5 et SUVH6 à travers un mécanisme de boucle rétroactive où le domaine « SET and RING-associated » (SRA) de ces HMTs reconnaît l'ADN méthylé et inversement, les domaines « bromo adjacent homology domains » (BAH) et Chromo des CMTs reconnaît H3K9me2 (Du *et al.*, 2015) (figure 8). Ainsi, dans des plantes mutantes *suvh4 suvh5 suvh6* la perte de H3K9me2 est corrélée avec la perte de méthylation

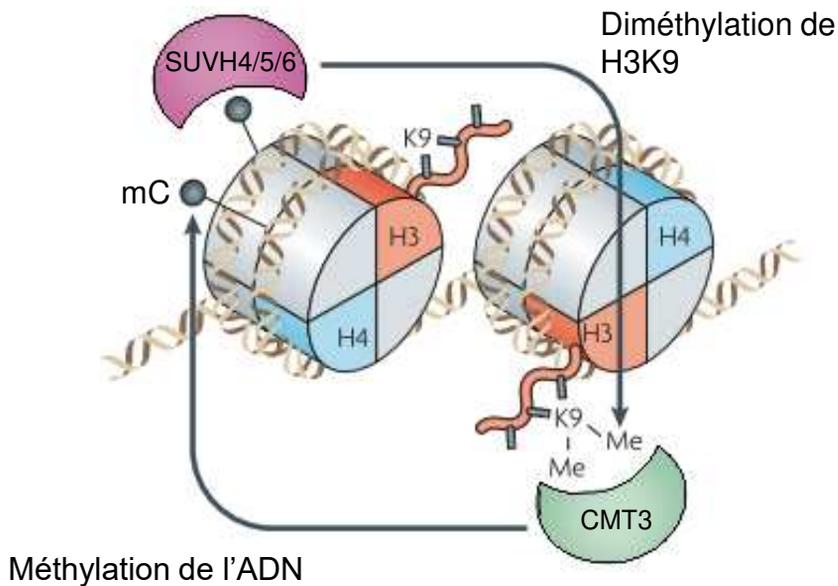


Figure 8 : Boucle d'autorégulation de la méthylation de l'ADN de type CHG et de la marque répressive H3K9me2 chez *A. thaliana*.

Les protéines HKMTs SUVH4/5/6 sont capables de reconnaître la méthylation de l'ADN de type CHG pour déposer la marque H3K9me2 au sein des nucléosomes adjacents. La marque H3K9me2 est alors reconnue par CMT3 qui catalyse le dépôt de la méthylation CHG à ces régions génomiques, maintenant ainsi une répression transcriptionnelle. Modifiée de Law and Jacobsen, 2010

CHG (Stroud *et al.*, 2013). Réciproquement, dans un mutant *drm1 drm2 cmt2 cmt3*, le niveau de H3K9me2 diminue de la même manière que dans un mutant *suvh4 suvh5 suvh6* (Stroud *et al.*, 2014). Ces analyses montrent une boucle de rétrocontrôle entre ces deux marques.

La maintenance de la méthylation CHH est assurée par l'action de DRM2 et CMT2 (Stroud *et al.*, 2013, 2014). DRM2 maintient cette marque *via* la voie du RdDM, et CMT2 agit probablement en se liant à H3K9me2. Ces deux protéines maintiennent la méthylation sur des régions différentes du génome (Stroud *et al.*, 2014; Zhang *et al.*, 2018a).

Toutes ces voies de méthylation de l'ADN, ainsi que la voie H3K9me2 interagissent entre elles à travers un réseau complexe de mécanismes. Par exemple, il a été montré que MET1 et CMT3 peuvent jouer un rôle dans le maintien de la méthylation de type CHH (Zhang *et al.*, 2018b). En effet, une diminution de cette marque épigénétique est observée dans un double mutant *met1 cmt3* à certaines régions génomiques présentant également une diminution de H3K9me2. Le rôle de MET1 et CMT3 dans le maintien de la méthylation de type CHH pourrait cependant être indirect. En effet, La diminution de H3K9me2 due à la mutation *cmt3* impacterait le recrutement de CMT2 et, ainsi, aurait un effet sur le niveau de CHH (Zhang *et al.*, 2018b).

II.3.5) La déméthylation de l'ADN

La méthylation des cytosines est une marque épigénétique généralement maintenue au cours du développement. Toutefois, il est parfois nécessaire de retirer cette méthylation. La déméthylation de l'ADN peut s'effectuer de manière passive via la réplication de l'ADN et l'absence de maintenance. Elle peut également être retirée de manière active par des enzymes spécialisées intervenant dans la voie de réparation par excision de base (BER) (Zhang *et al.*, 2018b). Chez *A. thaliana*, les mC sont directement reconnues par des ADN glycosylases qui peuvent les exciser du brin d'ADN générant un site apyrimidique (AP). Celui-ci est ensuite reconnu par l'endonucléase DNA-(APURINIC OR APYRIMIDINIC SITE) LYASE (APE1L) et l'ADN phosphatase POLYNUCLEOTIDE 3'-PHOSPHATASE (ZDP) qui coupent les liaisons phosphate entre les nucléotides, générant un site vide dans le brin d'ADN (Zhang *et al.*, 2018b). Ce site vide sert de substrat à une ADN polymérase inconnue et à l'ADN

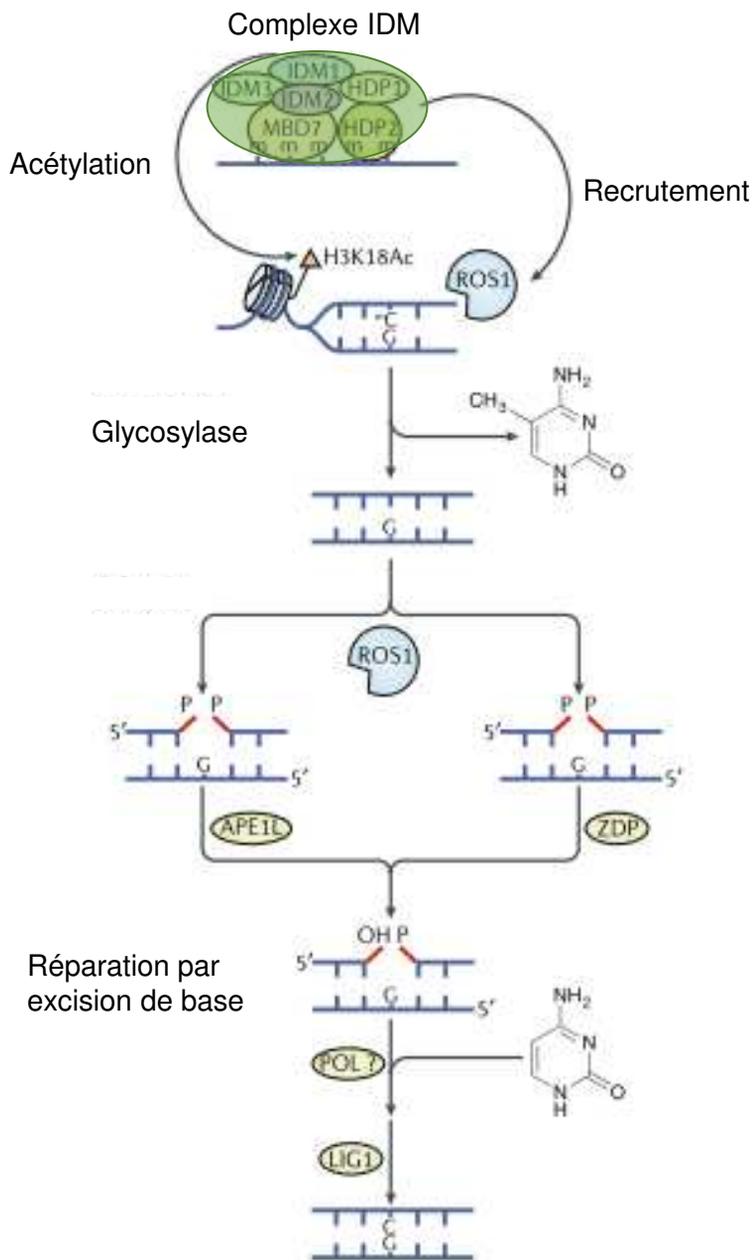


Figure 9 : Déméthylation active de l'ADN par ROS1 chez *A. thaliana*

Le complexe INCREASED DNA METHYLATION (IDM), composé de plusieurs protéines et notamment IDM1 est capable de reconnaître la marque H3K18Ac et de recruter ROS1. Par la suite, ROS1 retire la cytosine méthylée créant un site apyrimidique reconnu par APE1L et ZDP et la voie de réparation par excision de base restaure une cytosine non-méthylée *via* l'action d'une polymérase inconnue et de LIG1. Modifiée de Zhang *et al.*, 2018

ligase DNA LIGASE 1 (LIG1) qui restaurent une cytosine non méthylée (Zhang *et al.*, 2018b) (figure 9). Il existe plusieurs ADN glycosylases chez *A. thaliana*. DEMETER (DME) est essentielle au phénomène de l’empreinte parentale se déroulant dans l’albumen et essentiel pour la viabilité de la graine (Choi *et al.*, 2002). DME est aussi exprimée au cours de la gamétogenèse dans la cellule centrale du gamétophyte femelle afin de permettre l’activation de gènes tels que *MEDEA (MEA)*, *FLOWERING WAGENINGEN (FWA)* et *FERTILIZATION INDEPENDENT SEED 2 (FIS2)* (Huh *et al.*, 2008; Gehring *et al.*, 2009; Hsieh *et al.*, 2009). DME cible de préférence les TEs présents dans les promoteurs de gènes, ce qui impacte leur expression (Huh *et al.*, 2008; Gehring *et al.*, 2009; Hsieh *et al.*, 2009). Dans les tissus végétatifs, les ADN glycosylases REPRESSOR OF SILENCING (ROS1), DEMETER-LIKE 2 (DML2) et DML3 agissent de façon redondante (Choi *et al.*, 2002; Penterman *et al.*, 2007; Ortega-Galisteo *et al.*, 2008). ROS1 cible des TEs proche de gènes et des régions enrichies en marques H3K18ac et H3K27me3 et appauvries en marques H3K27me1 et H3K9me2 (Tang *et al.*, 2016). Son action permettrait d’établir les frontières entre les transposons et les gènes en limitant la propagation de la méthylation de l’ADN en jouant un rôle antagoniste au RdDM (Tang *et al.*, 2016). Le recrutement de ROS1 sur ces cibles peut se faire via le complexe INCREASED DNA METHYLATION (IDM), qui contient notamment le facteur IDM1 capable de lier la méthylation de l’ADN et possédant une activité histone acétyltransférase qui pourrait déterminer les cibles de ROS1 (Qian *et al.*, 2012; Lang *et al.*, 2015; Duan *et al.*, 2017). D’autres protéines ont été identifiées comme jouant un rôle dans la voie ROS1, comme par exemple ROS3, une protéine se liant aux ARNncs et qui ciblerait ROS1 sur certaines de ses cibles, et plus récemment METHYLATION ELEVATED MUTANT 1 (MEM1), dont le mode d’action reste à caractériser (Zheng *et al.*, 2008; Lu *et al.*, 2020).

Chez les animaux, il n’existe pas d’homologue des protéines DME ou ROS1. La déméthylation de l’ADN implique les protéines ten-eleven translocation 1 (TET1), TET2 et TET3 qui modifient les 5mC en 5-hydroxyméthyl cytosine (5hmC), 5-formylcytosine (5fC) et 5-carboxyl cytosine (5caC) (Wu and Zhang, 2017). Ces modifications sont ensuite soit reconnues par des thymine DNA glycosylases (TDG) qui les excisent, et enclenchent la voie de la « base excision repair » (BER), soit éliminées par dilution via la réplication de l’ADN (Wu and Zhang, 2017). Comme chez les plantes, la BER implique une endonucléase, une ADN polymérase et une ligase : respectivement AP endonucléase 1 (APE1), DNA polymerase β (pol β) et ligase 3

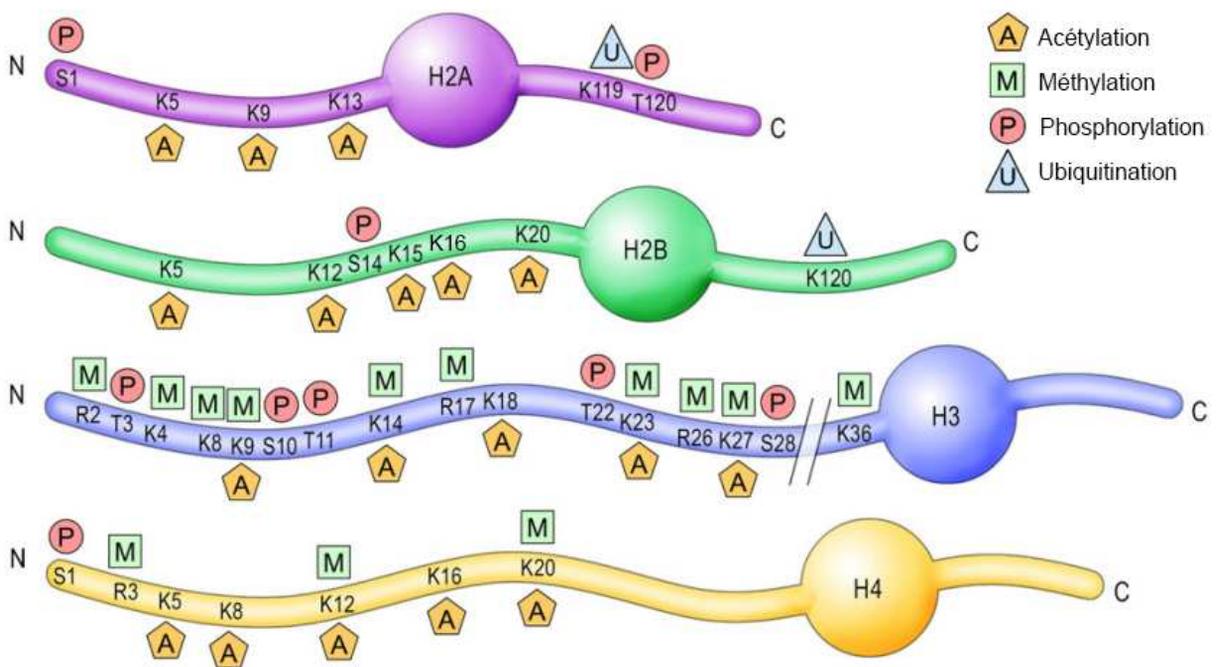


Figure 10 : Les principales modifications post-traductionnelles (MPTs) des histones.

Représentation schématique des sites de MPTs sur les extrémités N et C-terminales des histones composant le nucléosome. S : sérine, K : lysine, T : thréonine ; R : arginine. Modifiée de Graff and Mansuy, 2008

(LIG3) (Wu and Zhang, 2017). Une autre protéine pourrait être impliquée dans la déméthylation de l'ADN chez les animaux : Activation induced deaminase (AID), une cytidine désaminase de la famille des APOBEC, est capable de changer les cytosines en uraciles et les mC en thymines par hydrolyse (Muramatsu *et al.*, 2000). Cette protéine est majoritairement exprimée dans les lymphocytes, et de façon transitoire dans les autres tissus. En cas de cancer, AID est exprimée de façon ectopique dans différents tissus (Muramatsu *et al.*, 1999; Rommel *et al.*, 2013). Le rôle de AID dans la déméthylation est encore débattu (Bochtler *et al.*, 2017), mais un modèle propose que AID agisse de manière indirecte sur la déméthylation de l'ADN en désaminant les mC, la transition C->T étant par la suite réparée grâce à la voie du BER (Santos *et al.*, 2013; Dominguez and Shaknovich, 2014; Grin and Ishchenko, 2016).

II.4) Les histones

II.4.1) Les modifications post-traductionnelles (MPTs) des histones

Les histones peuvent être modifiées par l'ajout de groupements chimiques variés, tels que des groupements acétyl, méthyl et phosphate, majoritairement sur leur extrémité N-terminale par l'activité de plusieurs enzymes spécialisées (figure 10). Les histones peuvent également être modifiées sur leur extrémité C-terminale ainsi qu'au sein de leur domaine globulaire, situé au cœur du nucléosome. La combinaison de ces différentes MPTs constitue le fameux code des histones. En combinaison avec la méthylation de l'ADN, les MPTs des histones jouent un rôle essentiel dans tous les processus épigénétiques en lien avec la chromatine, tels que la régulation de la transcription, la répression des TEs, la division cellulaire, ou encore la réparation de l'ADN (voir partie **II.5**).

II.4.2) L'acétylation des histones

L'acétylation des histones est une modification qui se fait sur les résidus lysines. Cette marque est très dynamique et sa régulation se fait par deux familles de protéines antagonistes : les histones acétyl transférases (HATs), qui déposent les groupements acétyl sur la chromatine et les histones désacétylases (HDACs) qui les retirent (Kouzarides, 2007; Bannister and Kouzarides, 2011). Le groupement acétyl neutralise

la charge positive de l'histone, ce qui affaiblit son interaction avec l'ADN (Kouzarides, 2007 ; Bannister and Kouzarides, 2011). L'acétylation des histones est une marque permissive pour la transcription ; elle est enrichie dans l'euchromatine. Il existe deux classes majeures de HAT. La classe de type-A est séparée en 3 familles selon leur séquence et leur conformation : les familles GNAT, MYST et CBP/p300 (Hodawadekar and Marmorstein, 2007). Ces HATs sont souvent en complexe avec d'autres protéines qui contrôlent leur recrutement, activité et spécificité de substrat (Yang and Seto, 2007). La classe de type-B, correspond à des HAT cytoplasmiques qui acétylent les histones libres plutôt que les histones incorporées dans la chromatine (Parthun, 2007). De par leur fonction, les HDACs sont des répresseurs transcriptionnels. Elles sont séparées en trois classes selon leur homologie avec les HDAC de levure. Les HDACs de classe I sont homologues de Reduced Potassium Deficiency 3 (RPD3), celles de classe II sont homologues à l'Histone Deacetylase-A 1 (HADA1), enfin les HDACs de la classe IV montrent des homologies à RPD3 et HADA1 (Chen *et al.*, 2019). Plusieurs études ont montré que les HDACs ont des rôles importants dans de nombreux processus biologiques, comme le développement, la réponse aux stress biotiques et abiotiques et la sénescence chez les plantes (Ueda *et al.*, 2017; Chen *et al.*, 2019; Zheng *et al.*, 2020). Chez *A. thaliana*, HADA6 est impliquée dans des processus de répression des TEs en interagissant avec MET1 et SUVH4/5/6 (Liu *et al.*, 2012 ; Yu *et al.*, 2017) ou en réponse à des stimuli environnementaux pour réprimer l'expression de gènes cibles (Hung *et al.*, 2019).

II.4.3) La méthylation des histones

La méthylation des histones est une modification qui se fait sur les résidus lysine et arginine. Dans ce paragraphe, je me focaliserai sur la méthylation des résidus lysine. Jusqu'à trois groupements méthyl peuvent être ajoutés sur une même lysine. Selon la position de la lysine au sein de la séquence peptidique de l'histone et le nombre de groupement méthyl ajouté, cette modification peut être permissive ou répressive pour la transcription (Fromm and Avramova, 2014; Du *et al.*, 2015). La majorité des histones lysine méthyl transférases (HKMTs) sont caractérisées par un domaine SET, identifié originellement dans les HKMTs Suppressor of variegation 3-9 [su(var)3-9], Enhancer of zeste [E(Z)] et Trithorax de *D. melanogaster* (Elgin and Reuter, 2013).

Chez *A. thaliana*, les HKMTs sont regroupées en cinq classes (Pontvianne *et al.*, 2010) :

- La classe I est composée de CURLY LEAF (CLF), MEDEA (MEA) et SWINGER (SWN), des homologues de E(Z). Ce sont des acteurs essentiels de la voie polycomb (PcG) qui, comme nous l'avons vu plus haut, est nécessaire à l'élaboration de l'hétérochromatine facultative, à travers la déposition de la marque tri-méthylation de la lysine 27 de H3 (H3K27me3). H3K27me3 est enrichie dans les gènes dont l'expression doit être réprimée à certaines étapes du développement (Mozgova and Hennig, 2015).
- La classe II est composée des protéines homologues de Set2, une HKMT de *S. cerevisiae*, telle que SET DOMAIN GROUP 8 (SDG8) (Zhao *et al.*, 2005). Ces HKMTs déposent la tri-méthylation de la lysine 36 de H3 (H3K36me3) pendant la phase d'élongation de la transcription. Chez *S. cerevisiae*, H3K36me3 permet de réprimer la transcription cryptique (Carrozza *et al.*, 2005). Chez *A. thaliana*, elle est associée à l'activation de la transcription, jouant notamment un rôle dans le déclenchement de la floraison (Zhao *et al.*, 2005; Xu *et al.*, 2008).
- La classe III est composée des ARABIDOPSIS TRITHORAX-LIKE (ATX), qui font partie du complexe COMPASS-related (AtCOMPASS). Ces HKMTs, comme SDG2, sont généralement impliquées dans le dépôt de la tri-méthylation de la lysine 4 de H3 (H3K4me3), une marque permissive pour la transcription, enrichie dans l'euchromatine (Fromm and Avramova, 2014). Ces enzymes sont également capables de déposer la marque H3K36me3 (Lee *et al.*, 2017).
- La classe IV est composée des HKMTs ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 (ATXR5) et ATXR6. Ces HKMTs, spécifiques des plantes, déposent un unique groupement méthyl sur la lysine 27 de H3 (H3K27me1). H3K27me1 est enrichie dans l'hétérochromatine constitutive et réprime la transcription. Cet enrichissement de H3K27me1 dans l'hétérochromatine constitutive semble être spécifique des plantes (Jacob and Michaels, 2009; Jacob *et al.*, 2009).
- La classe V est la classe la plus abondante. Elle se compose de la famille des SU(VAR)3-9 HOMOLOGOUS (SUVH) qui comporte 10 membres. Comme décrit précédemment, SUVH4 (ou KYP), SUVH5 et SUVH6

permettent le dépôt de la di-méthylation de la lysine 9 de H3 (H3K9me₂), une marque répressive enrichie dans l'hétérochromatine (Du *et al.*, 2015). Les protéines SUVH2 et SUVH9 jouent, quant à elles, un rôle dans le RdDM, en recrutant l'ARN pol V sur la chromatine (Johnson *et al.*, 2014).

La méthylation des histones a longtemps été considérée comme une marque stable, avant que des enzymes portant une activité d'histone déméthylase soient identifiées (Shi *et al.*, 2004; Tsukada *et al.*, 2006). Chez les animaux et les plantes, le domaine Jumonji C (JmjC) est conservé et porte l'activité déméthylase (Tsukada *et al.*, 2006 ; Lu *et al.*, 2008 ; Pedersen & Helin, 2010). Chez *A. thaliana*, 21 gènes codent des protéines JUMONJI (JMJ) (Lu *et al.*, 2008 ; Chen *et al.*, 2011). Les protéines JMJ14-19 déméthylent spécifiquement H3K4me_{1/2/3} (Iwase *et al.*, 2007 ; Klose & Zhang, 2007 ; Lee *et al.*, 2007). JMJ14 joue un rôle dans la répression de FT et SOC, et permet le maintien de la répression transcriptionnelle *via* le RdDM (Deleris *et al.*, 2010). Il a également été montré que JMJ14 active la voie de la répression post-transcriptionnelle (Le Masson *et al.*, 2012). Les histones déméthylases jouent de nombreux rôles dans la plante, notamment dans le contrôle de la floraison avec RELATIVE OF EARLY FLOWERING 6 (REF6) et EARLY FLOWERING 6 (ELF6) (Noh *et al.*, 2004), ou dans la tolérance aux stress (Shen *et al.*, 2014 ; Huang *et al.*, 2019). REF6 et ELF6 ont également été récemment impliqués dans des processus de reprogrammation et réinitialisation épigénétique au sein du gamète mâle (Borg *et al.*, 2020).

II.4.5) Les autres MPTs des histones

La phosphorylation des histones peut se faire sur les résidus sérine, thréonine et tyrosine (Rossetto *et al.*, 2012). La phosphorylation comme l'acétylation est une marque très dynamique, mise en place par les kinases et retirée par les phosphatases. L'ajout d'un groupement phosphate sur l'extrémité N-terminale des histones augmente les charges négatives des histones, ce qui provoque un relâchement de la structure de la chromatine. Cette modification joue donc un rôle dans le contrôle de la transcription (Rossetto *et al.*, 2012). Elle joue également un rôle important dans la voie de réparation de l'ADN, en effet lors de cassures double brin (DSB), l'histone H2A est

phosphorylée par plusieurs protéines kinases comme ATAXIA TELANGIECTASIA MUTATED (ATM) et ATM AND RAD3 RELATED PROTEIN (ATR) (Friesner et al., 2005). Enfin, cette modification est aussi associée à la condensation des chromosomes lors de la division cellulaire. Par exemple, l'histone H3 est phosphorylée sur la sérine 10 (H3S10ph) au moment de la condensation des chromosomes lors de la prophase (Kaszas and Cande, 2000 ; Goto et al., 2002).

Les histones peuvent également être modifiées par de petits peptides tels que l'ubiquitine ou les small ubiquitin-like modifier (SUMO) (Kouzarides, 2007). L'ubiquitination des histones est réalisée par l'action de trois enzymes E1-activating, E2-conjugating et E3-ligating (Hershko and Ciechanover, 1998). Une histone peut être modifiée par l'ajout d'une seule ou plusieurs ubiquitines. Selon l'histone et/ou le résidu modifié, l'ubiquitination aura un impact négatif ou positif sur la transcription, par exemple l'ubiquitination de la lysine 123 de H2B est une marque permissive, tandis que l'ubiquitination de la lysine 119 de H2A est répressive (Wang *et al.*, 2004; Lee *et al.*, 2007; Kim *et al.*, 2009). La sumoylation peut être portée par toutes les histones nucléosomales, et son dépôt est assuré par un processus enzymatique similaire à celui réalisant l'ubiquitination (Bannister and Kouzarides, 2011). La sumoylation des histones joue un rôle de répresseur de la transcription, en permettant le recrutement de protéines HDAC lorsque H4 est sumoylée (Shiio and Eisenman, 2003). Elle peut agir comme compétiteur de l'acétylation ou de l'ubiquitination à certaines lysines comme par exemple pour H2BK16ac (Nathan *et al.*, 2006).

II.4.6) Les variants d'histones

Chez les eucaryotes, toutes les histones possèdent des variants (Talbert *et al.*, 2012; Long *et al.*, 2019). Ces variants d'histone peuvent différer de quelques acides aminés ou par la présence de domaines protéiques supplémentaires. On différencie deux types de variants d'histone : le premier type correspond à des histones incorporées dans la chromatine lors de la phase S du cycle cellulaire, après le passage de la fourche de réplication, formant ainsi de nouveaux nucléosomes. Ces variants sont appelés des variants d'histone répliatifs, ils correspondent aux versions canoniques des histones. Chez les mammifères et la drosophile, les gènes d'histones canoniques sont organisés en cluster avec une régulation unique, et ils ne portent

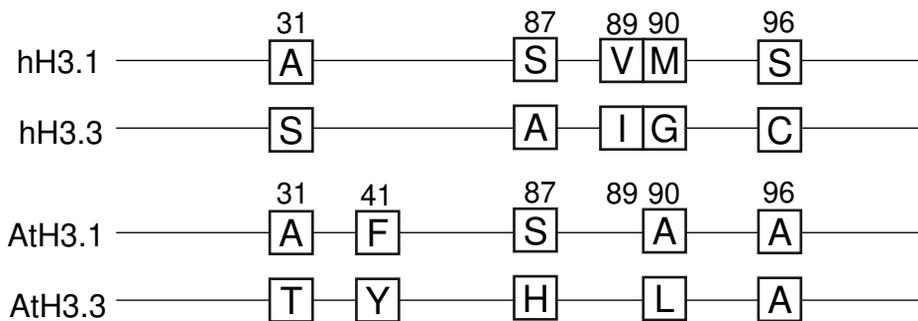


Figure 11 : Différences de séquences peptidiques entre les variants d'histone H3.1 et H3.3 et entre espèces.

Représentation schématique des différences de séquences entre les variants d'histone H3.1 et H3.3 chez l'Homme (h) et *A. thaliana* (At). Les acides aminés qui diffèrent entre H3.1 et H3.3 sont encadrés et leur position dans la séquence est indiqué au-dessus.

généralement pas d'intron (Elsaesser *et al.*, 2010). Chez les plantes, cette organisation en cluster n'est pas observée, et seuls les gènes codant l'histone H3 canonique ne portent pas d'intron (Chaubet *et al.*, 1992). Le deuxième type de variant correspond à des gènes exprimés tout le long du cycle cellulaire. Ils assurent le turn-over des histones dans les nucléosomes, ou leur remplacement après l'éviction des nucléosomes lors de mécanismes comme la transcription (Venkatesh and Workman, 2015). Ces variants correspondent aux histones dites non-canoniques, aussi appelées histones remplaçantes.

II.4.6.1) Les variants H3

a) Les différents types d'histone H3

Il existe trois variants « principaux » d'histone H3, communs aux animaux et aux plantes : L'histone canonique H3.1, et deux histones non-canoniques H3.3 et cenH3. La distribution chromosomique de ces variants diffère grandement. H3.1 est enrichie dans les régions hétérochromatiques, tandis que H3.3 est retrouvée dans les régions euchromatiques. H3.3 est positivement corrélée à la transcription, chez les mammifères, elle est enrichie dans les régions proches du site d'initiation de la transcription (TSS) des gènes actifs et dans leurs régions transcrites (Goldberg *et al.*, 2010). Chez les plantes, H3.3 est retrouvée dans les régions 3' des gènes, mais est déplétée du TSS (Stroud *et al.*, 2012; Wollmann *et al.*, 2012). Elle est aussi enrichie au sein des régions régulatrices, des télomères et des répétitions d'ADN ribosomique (ADNr) (Shi *et al.*, 2011; Shu *et al.*, 2014; Duc *et al.*, 2017). CenH3 (appelée CENP-A chez les mammifères et CID chez *D. melanogaster*) est une histone localisée spécifiquement dans les régions centromériques (Nagaki *et al.*, 2003; Schuh *et al.*, 2007; Black and Cleveland, 2011). H3.1 et H3.3 ne diffèrent que de quelques acides aminés qui sont différents selon les espèces considérées, suggérant que les variants H3 sont apparus indépendamment dans les différents règnes (figure 11) (Waterborg and Robertson, 1996; Talbert *et al.*, 2012). Cependant, la répartition similaire des variants H3.1 et H3.3 le long des chromosomes d'espèces phylogénétiquement éloignées laisse supposer une convergence évolutive de ces histones (Mito *et al.*, 2005; Goldberg *et al.*, 2010). La séquence peptidique de CenH3 diffère, quant à elle,

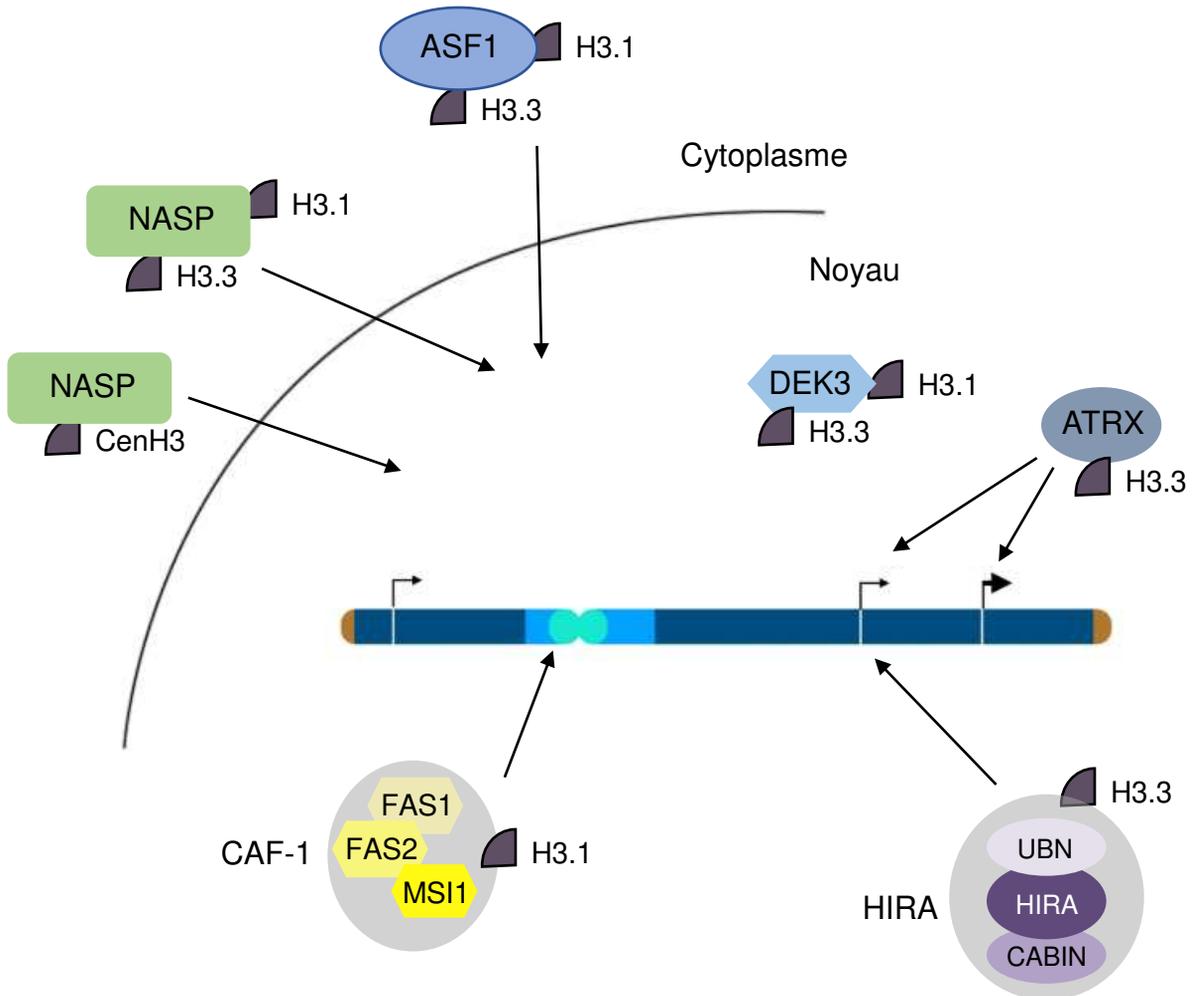


Figure 12 : Les variants d'histone H3 et leurs chaperonnes.

Représentation schématique de la répartition des variants H3.1 et H3.3 le long des chromosomes par leur chaperonnes spécifiques. Les histones H3.1 et H3.3 sont importées dans le noyau par NASP et ASF1 où elles sont ensuite prises en charge par leur chaperonnes respectives. H3.1 est intégrée dans la chromatine par CAF-1 lors de la réplication, et H3.3 est intégrée dans la chromatine par HIRA et ATRX. H3.1 et H3.3 peuvent également être prises en charge par la chaperonne DEK3. Les bras chromosomiques sont représentés en bleu foncé et les régions péri-centromériques en bleu clair. Modifiée de Probst *et al.*, 2020.

grandement de H3.1 et H3.3, ainsi qu'entre espèces de différents règnes, ce qui suppose une adaptation de la protéine aux séquences centromériques sur lesquelles elle se lie (Talbert *et al.*, 2002; Simon *et al.*, 2015).

Chez *A. thaliana*, 15 gènes appelés *HISTONE THREE RELATED (HTR)* 1-15, codent pour les histones H3. Parmi ces gènes, cinq codent l'histone canonique H3.1 (*HTR1*, *HTR2*, *HTR3*, *HTR9* et *HTR13*) trois gènes codent l'histone non-canonique H3.3 (*HTR4*, *HTR5* et *HTR8*) et un gène code l'histone CenH3 : *HTR12* (Chaubet *et al.*, 1992; Okada *et al.*, 2005; Jiang and Berger, 2017). Parmi les 6 gènes restants, *HTR6*, *HTR10*, *HTR11*, *HTR14* et *HTR15* codent respectivement les histones H3.6, H3.10, H3.11, H3.14 et H3.15. Les histones H3.6, H3.10 et H3.14 portent des résidus spécifiques de H3.3, tandis que H3.11 présente des caractéristiques de H3.1. Malgré ces ressemblances, l'impact fonctionnel de ces variants atypiques n'est pas bien compris (Tachiwana *et al.*, 2011; Probst *et al.*, 2020). Enfin, *HTR7* est classé comme un pseudogène.

b) Prise en charge des variants H3

A cause de leur nature basique, les histones peuvent interagir de manière aspécifique avec d'autres protéines et des acides aminés. Afin d'éviter ces interactions non-spécifiques, les histones sont prises en charge par des protéines ou des complexes protéiques spécifiques appelés chaperonnes d'histone.

Les histones H3 nouvellement synthétisées dans le cytoplasme sont prises en charge par les chaperonnes ANTI-SILENCING FACTOR 1 (ASF1) et NUCLEAR AUTOANTIGENIC SPERM PROTEIN (NASP), qui permettent le transport dans le noyau des dimères H3-H4 (Campos *et al.*, 2010). ASF1 va ensuite transmettre les dimères H3-H4 à d'autres chaperonnes spécifiques qui assureront leur dépôt dans la chromatine (English *et al.*, 2006; Natsume *et al.*, 2007). Au niveau de la fourche de réplication, le dépôt *de novo* des histones est assuré par le complexe CAF-1 (Hoek and Stillman, 2003). Chez les plantes, CAF-1 est composé des sous-unités FASCIATA1 (FAS1), FAS2 et MSI1, et son recrutement à la fourche de réplication se fait par l'interaction de FAS1 avec le facteur de réplication PCNA (Jiang and Berger, 2017). Des expériences chez les mammifères et *A. thaliana* ont montré que le complexe CAF-1 lie et incorpore spécifiquement l'histone H3.1 (Tagami *et al.*, 2004; Otero *et al.*, 2016; Jiang and Berger, 2017; Benoit *et al.*, 2019) (figure 12).

Le dépôt des histones H3.3 dans la chromatine est assuré par l'action des complexes HIRA et ATRX-DAXX (Ray-Gallet *et al.*, 2002; Goldberg *et al.*, 2010; Nie *et al.*, 2014; Duc *et al.*, 2015, 2017). Chez les mammifères et *A. thaliana*, le complexe HIRA est composé des sous-unités HIRA, CABIN et UBINUCLEIN (Nie *et al.*, 2014; Duc *et al.*, 2015). La protéine HIRA lie spécifiquement H3.3 (Nie *et al.*, 2014). Chez les mammifères, dans le complexe ATRX-DAXX, la protéine DAXX donne la spécificité de liaison à l'histone H3.3, en préférant la glycine à la méthionine en position 90 sur les histones H3.3 (Elsässer *et al.*, 2012). Chez *A. thaliana*, aucun homologue de DAXX n'a été identifié, et il a été montré qu'ATRX lie directement les histones H3.3 (Duc *et al.*, 2017). Chez *A. thaliana*, le complexe HIRA permettrait donc l'incorporation des histones H3.3 dans les régions hétérochromatiques (Goldberg *et al.*, 2010; Lewis *et al.*, 2010). De plus, toujours chez *A. thaliana*, la protéine DEK-DOMAIN CONTAINING PROTEIN 3 (DEK3) peut aussi interagir avec les dimères H3-H4 et a un effet sur l'occupation des nucléosomes à certains loci (Waidmann *et al.*, 2014). DEK3 est enrichie sur les régions 3' des gènes de la même manière que les histones H3.3, suggérant une préférence de DEK3 pour H3.3 (Brestovitsky *et al.*, 2019). Enfin, au cours de la transcription les histones sont prises en charge par SPT6 et le complexe « FACILITATES CHROMATIN TRANSCRIPTION » (FACT), afin de permettre leur échange et leur recyclage (Duroux *et al.*, 2004; Grasser, 2020) (figure 12). Les voies de recyclage des histones H3.1 et H3.3 dans la chromatine sont très conservées à travers les espèces.

L'histone CenH3 est déposée par des voies différentes en fonction de l'espèce étudiée. Chez l'Homme, elle est déposée par la chaperonne Holliday junction recognition protein (HJURP) (Dunleavy *et al.*, 2009; Foltz *et al.*, 2009). Chez la *D. melanogaster*, la chaperonne chromosome alignment defect 1 (CAL1) joue ce rôle (Phansalkar *et al.*, 2012). Enfin, chez *A. thaliana*, une étude récente a montré que la protéine chaperonne NASP agit dans l'incorporation de CenH3 au sein des centromères (Le Goff *et al.*, 2020).

c) Spécificité et rôle des variants d'histone H3

Chez *A. thaliana*, une réduction de l'expression des gènes codant H3.1 affecte le développement des plantes (Kaya *et al.*, 2001). Ce phénotype ne peut pas être complété par l'expression de gènes *H3.3* sous le contrôle d'un promoteur de gène *H3.1*, ce qui démontre bien les rôles spécifiques de chaque variant (Jiang and Berger, 2017). Combiner des mutations au sein des différents gènes *H3.3* est létal, et une réduction de leur expression provoque des phénotypes développementaux, ainsi qu'une perturbation globale de l'expression génique (Wollmann *et al.*, 2017).

Les MPTs des histones sont enrichies différenciellement sur les variants H3.1 et H3.3 : les marques répressives, comme H3K9me2 et H3K27me1/3, sont enrichies sur H3.1, tandis que celles qui sont permissives à la transcription, comme H3K4me2/3, sont enrichies sur H3.3 (Vergara and Gutierrez, 2017). Ceci est évidemment cohérent avec les localisations sub-chromatiniennes des différentes marques épigénétiques et variants H3 décrites plus haut. Enfin, il a été montré que H3.3 est essentielle pour la mise en place de la marque H3K9me2 sur les télomères (Udugama *et al.*, 2015).

Chez *A. thaliana*, H3.10 est un variant spécifique des gamètes males (Okada *et al.*, 2005). Ce variant est intégré dans la chromatine au cours de la gamétogénèse, au cours de laquelle la marque H3K27me3 est remise à zéro. H3.10 participe à cette reprogrammation épigénétique en empêchant le dépôt de H3K27me3 sur les nucléosomes dans lesquels il est incorporé (Borg *et al.*, 2020).

II.4.6.2) Les variants H2A

a) Les différents types d'histones H2A

Le génome des eucaryotes code pour plusieurs variants H2A : les variants H2A.Z et H2A.X sont présents chez tous les eucaryotes, les variants macroH2A et H2AB (ou H2ABbd) sont spécifiques aux animaux, enfin le variant H2A.W est spécifique des plantes (Yelagandula *et al.*, 2014). Chaque variant est codé par plusieurs gènes. Le génome d'*A. thaliana* code pour quatre H2A canoniques, deux

H2A.X, trois H2A.W et quatre H2A.Z (Probst *et al.*, 2020). Chacun de ces variants jouent des rôles différents au sein de la chromatine.

Chez les animaux, l'histone H2A.B est enrichie sur les gènes activement transcrits (Chen *et al.*, 2014), et les nucléosomes contenant ce variant sont moins stables provoquant ainsi une structure moins compacte de la chromatine (Weber and Henikoff, 2014).

H2A.Z est impliquée dans la régulation de la transcription et peut agir comme un facteur activateur (Santisteban *et al.*, 2000; Mahrez *et al.*, 2016) ou répresseur (Fan *et al.*, 2004; Sura *et al.*, 2017). Chez *A. thaliana*, ce double rôle de H2A.Z dépend de son enrichissement le long des régions transcrites. H2A.Z est associée à l'activation de la transcription si elle est enrichie dans le nucléosome en +1 après le TSS (Zilberman *et al.*, 2008; Mahrez *et al.*, 2016), et à la répression de la transcription si elle est enrichie dans le corps des gènes (Sura *et al.*, 2017). Les dimères H2A.Z-H2B sont remplacés plus rapidement que les dimères contenant d'autres variants d'H2A (Osakabe *et al.*, 2018). Il a également été montré que H2A.Z joue un rôle dans la réparation des DSB. La présence de H2A.Z sur les régions de cassure permet de maintenir une conformation ouverte de la chromatine, et ainsi la mise en place de modifications post-traductionnelles nécessaires à la réparation des DSB (Xu *et al.*, 2012 ; Murr *et al.*, 2006). Le remplacement de l'histone H2A canonique par H2A.Z peut avoir un rôle dans plusieurs mécanismes physiologiques comme le développement de la plante (Jarillo and Piñeiro, 2015), ou la réponse aux stress (Asensi-Fabado *et al.*, 2017). Elle a notamment été montrée comme jouant un rôle dans l'expression de *FLOWERING LOCUS C (FLC)*, un répresseur de la floraison chez *A. thaliana* (Deal *et al.*, 2007).

H2A.X joue un rôle important dans la réparation des DSB chez les eucaryotes. En réponse aux DSB, H2A.X est phosphorylée sur sa région C-terminale par plusieurs protéines kinases comme ATM et ATR (Friesner *et al.*, 2005). La phosphorylation de H2A.X s'accumule sur les régions de DSB et permet le recrutement de plusieurs facteurs essentiels à la réparation de l'ADN (Rogakou *et al.*, 1998 ; Stiff *et al.*, 2006 ; Stucki *et al.*, 2005 ; Wu *et al.*, 2008).

H2A.W est spécifiquement enrichi dans l'hétérochromatine, permettant la répression des TEs indépendamment de la méthylation de l'ADN (Yelagandula *et al.*, 2014 ; Lorkovic *et al.*, 2017). Il a été montré que la phosphorylation du variant H2A.W.7

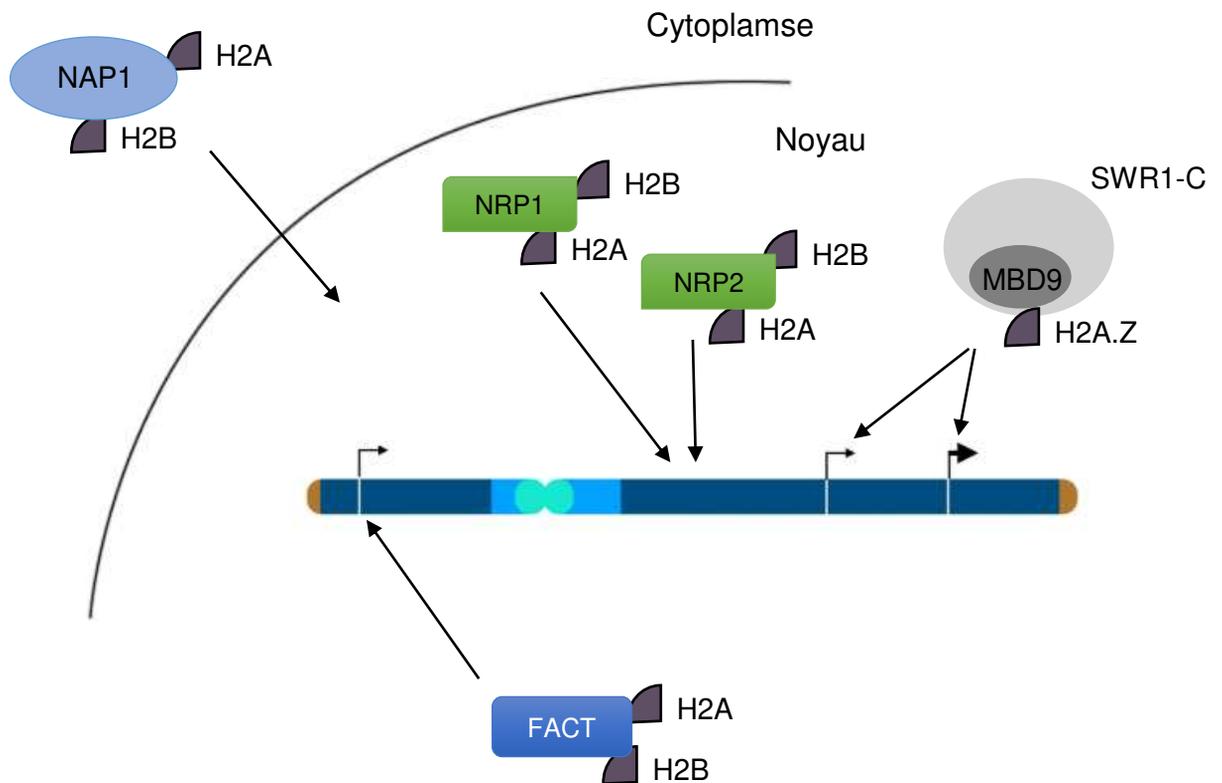


Figure 13 : Les histones H2A et H2B et leurs chaperonnes.

Représentation schématique de la répartition des histones H2A et H2B le long des chromosomes par leurs chaperonnes spécifiques. Les histones H2A et H2B sont importées dans le noyau par NAP1 où elles sont ensuite prises en charge par les chaperonnes NRP1 et NRP2 pour être intégrées à la chromatine. La chaperonne FACT peut déplacer H2A et H2B lors de la transcription. Le variant H2A.Z est intégré à la chromatine par la chaperonne SWR1-C contenant MBD9. Les bras chromosomiques sont représentés en bleu foncé et les régions péricentromériques en bleu clair. Modifiée de Probst *et al.*, 2020.

d'*A. thaliana* est nécessaire à la réparation des dommages de l'ADN dans les régions d'hétérochromatine (Roitinger et al., 2015).

b) La prise en charge des variants H2A

Les dimères H2A-H2B sont échangés plus rapidement que les dimères H3-H4. Nous disposons de moins de données sur leur mode de dépôt sur la chromatine (figure 13). Chez *A. thaliana*, NUCLEOSOME ASSEMBLY PROTEIN 1 (NAP1), NAP1-RELATED PROTEIN 1 (NRP1) et NRP2 sont capables de lier les dimères H2A-H2B (Dong et al., 2005 ; Zhu et al., 2006 ; Zhu et al., 2017) et sont localisées dans le cytoplasme et le noyau respectivement. NAP1 se lie aux dimères H2A-H2B dans le cytoplasme et les transporte dans le noyau (Polo and Almouzni, 2006). Les protéines NRP1 et NRP2 s'associent à la chromatine, et sont impliquées dans la régulation globale de la transcription et la croissance de la racine (Zhu et al., 2006 ; 2017). Au cours de la transcription et de la réplication, les dimères H2A-H2B peuvent être pris en charge par le complexe FACT, afin de permettre leur échange et leur recyclage (Lolas et al., 2010 ; Grasser, 2020). Le dépôt de H2A.Z implique le complexe SWR1-C, un remodeleur de la chromatine, contenant la protéine METYL-CpG-BINDING DOMAIN 9 (MBD9). Cette protéine est capable de reconnaître des régions riches en marques d'histones activatrices (comme H3K4me3 et H3Ac) pour y induire le dépôt de H2A.Z (Sijacic et al., 2019 ; Potok et al., 2019). MBD9 se lie également avec les remodeleurs de la chromatine CHR11 et CHR17, permettant le glissement des nucléosomes et l'expression des gènes (Luo et al., 2020). Il a été montré que le complexe SWR1-C est important pour l'incorporation de H2A.Z sur la région 5'UTR du gène ETHYLENE INSENSITIVE 2 (EIN2) (Zander et al., 2019).

II.4.6.3) Les variants H2B

Il existe plusieurs variants de H2B chez les plantes et les mammifères, mais les connaissances sur leur distribution dans le génome et leur fonction sont limitées. Chez le lys, un variant de H2B serait spécifiquement exprimé dans le pollen (Ueda et al., 2000) et des variants d'H2B portent des modifications post-traductionnelles spécifiques chez *A. thaliana* (Bergmüller et al., 2007). On ne sait pas si ces variants

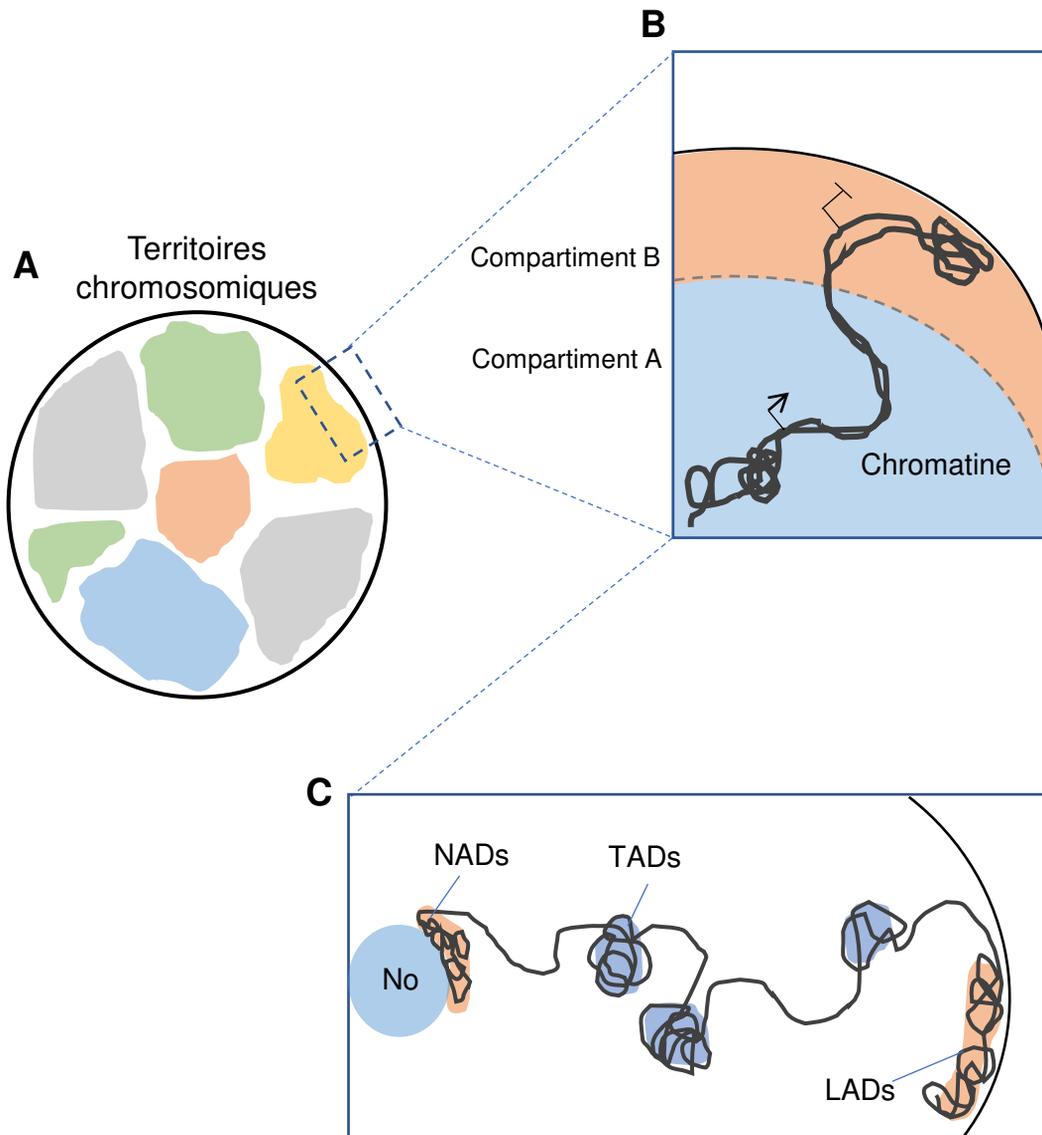


Figure 14 : La chromatine est organisée à plusieurs niveaux dans le noyau.

(A) Au sein du noyau, les chromosomes sont organisés en territoires chromosomiques bien définis. (B) Des compartiments A/B peuvent également être définis, ils correspondent à des régions d'ADN exprimées et réprimées respectivement, avec le compartiment B localisé à la périphérie nucléaire et le compartiment A vers le centre du noyau. (C) Les interactions de la chromatine avec elle-même ainsi qu'avec les composants du noyau définissent plusieurs domaines associés : Les NADs, LADs et TADs. Les deux premiers correspondent à des domaines de chromatine réprimée, et le dernier à des domaines chromatiniens interagissant préférentiellement entre eux. No : Nucléole.

sont préférentiellement enrichis dans certaines régions du génome ou s'ils sont associés préférentiellement à certains variants H2A.

II.5) Fonctions biologiques des modifications épigénétiques

La méthylation de l'ADN et les MPTs des histones sont engagées dans de nombreux processus nucléaires. Elles contribuent à l'organisation de la chromatine, régulent l'expression des gènes et des TEs. Comme déjà évoqué, elles sont aussi essentielles aux processus de réparation de l'ADN et de division cellulaire (Zhang *et al.*, 2018a; Greenberg and Bourc'his, 2019). De manière générale, les modifications épigénétiques sont, en combinaison d'autres facteurs cellulaires, impliquées dans tout processus biologique impliquant la chromatine.

II.5.1) Impact des modifications épigénétiques sur l'organisation de la chromatine dans le noyau

L'organisation 3D de la chromatine dans le noyau en interphase n'est pas aléatoire et plusieurs domaines fonctionnels sont définis à différentes échelles moléculaires par leurs interactions avec le reste du génome et leur paysage épigénétique. De nombreuses espèces montrent un plan d'organisation similaire (figure 14).

II.5.1.1) Les territoires chromosomiques

Chaque chromosome occupe une région spécifique, que l'on appelle des territoires chromosomiques et les interactions interchromosomiques sont limitées aux frontières entre ces territoires (Shardin *et al.*, 1985 ; Tanabe *et al.*, 2002 ; Bolzer *et al.*, 2005 ; Branco & Pombo, 2006). Chez les mammifères, les régions chromosomiques riches en gènes sont proches du centre du noyau ou des pores nucléaires (NPC), tandis que les régions chromosomiques plus riches en séquences d'ADN répétées et TEs sont proches de la périphérie nucléaire ou du nucléole (Croft *et al.*, 1999; Ibarra and Hetzer, 2015; Pontvianne and Liu, 2020). Le niveau d'expression des gènes et

leur réplication a également un impact sur leur localisation dans le noyau, les gènes répliqués précocement et les plus exprimés seront généralement plus proches du centre que les gènes répliqués tardivement et peu ou pas exprimés (Grasser et al. 2008, Takizawa et al. 2008). Chez les plantes, cette répartition des territoires chromosomiques en fonction du niveau d'expression n'est pas observée (Pecinka *et al.*, 2010; Schubert *et al.*, 2012). Les territoires chromosomiques chez *A. thaliana* montrent une répartition aléatoire dans le noyau et entre eux, à l'exception des territoires associés au nucléole (Pecinka *et al.*, 2010; Grob *et al.*, 2014).

II.5.1.2) Les compartiments A/B

Les chromosomes se répartissent en deux compartiments aux propriétés différentes : le compartiment A correspond aux régions euchromatiniennes, dense en gènes fortement exprimés et enrichies en H3K4me2/3. Le compartiment B correspond aux régions hétérochromatiniennes réprimées, riches en séquences d'ADN répétés et marquée par des marques répressives comme H3K9me2 ou 5mC. Le compartiment B est associé aux lamines qui composent la lamina à la périphérie des noyaux de mammifères (Lieberman-Aiden *et al.*, 2009 et voir ci-dessous). Les compartiments A/B sont séparés physiquement dans le noyau. Cependant, par des processus dynamiques, une région génomique peut passer d'un compartiment à l'autre selon les besoins de la cellule en termes d'expression des gènes (Ryba *et al.*, 2010 ; Wang *et al.*, 2016 ; Dong *et al.*, 2017). C'est le cas par exemple lors de la différenciation des cellules souches humaines au cours de laquelle 36% du génome montre un changement de compartiment (Dixon *et al.*, 2015). La présence de tels compartiments et leur séparation peut permettre des enrichissements locaux de la machinerie de transcription et des facteurs épigénétiques nécessaires, et ainsi augmenter leur efficacité (Wang *et al.*, 2016). Chez *A. thaliana*, les chromosomes adoptent une structure en compartiments A/B plus simple avec le compartiment A correspondant aux bras chromosomiques et le compartiment B correspondant aux régions péri-centromériques et aux centromères (Feng *et al.*, 2014; Grob *et al.*, 2014; Dong *et al.*, 2017).

II.5.1.3) Les domaines topologiquement associés (TADs)

Les domaines topologiquement associés (TADs) ont été décrits pour la première fois chez les mammifères, en analysant les interactions entre les régions chromosomiques (Dixon *et al.*, 2012). Ils ont par la suite été identifiés dans de nombreuses espèces d'eucaryotes comme *Drosophila* (Sexton *et al.*, 2012), le poisson-zèbre (Gomez-Marin *et al.*, 2015), *Caenorhabditis elegans* (Crane *et al.*, 2015), la levure (Hsieh *et al.*, 2015, Mizuguchi *et al.*, 2015) et plus récemment chez les plantes comme le maïs ou le riz (Dong *et al.*, 2017). Ce niveau d'organisation chromatinienne est cependant absent chez *A. thaliana*, probablement due à la faible complexité de son génome (Grob *et al.*, 2014). Les TADs correspondent à des régions génomiques interagissant fortement entre elles et plus faiblement avec le reste du génome (Dixon *et al.*, 2012 ; Szabo *et al.*, 2019). Leur position sur le génome est stable et indépendante du niveau d'expression des gènes, toutefois des variations d'interaction interne ont pu être observées dans les TADs au cours de la différenciation de cellules souches embryonnaires humaines (Dixon *et al.*, 2015). Chez les mammifères, les bordures des TADs sont définies par l'enrichissement de plusieurs facteurs tels que les modifications d'histones H3K4me3 et H3K36me3, des sites de début de transcription, des gènes de ménage, des gènes d'ARN de transfert (ARNt) et la protéine CTCF (CCCTC-binding factor). Les protéines CTCF jouent un rôle important dans la mise en place des TADs, en effet on observe une réduction de la formation des TADs lorsque CTCF est absente (Zuin *et al.*, 2014). Chez les plantes, il n'existe pas d'homologue de CTCF, toutefois une séquence liée par un facteur de transcription TEOSINTE BRANCHED1/CINCINNATA/PROLIFERATING CELL FACTOR (TCP) a été identifié dans la bordure des TADs chez le riz (Liu *et al.*, 2017). Un modèle propose que la formation de structures similaires aux TADs chez les plantes à grand génome reposerait donc sur des mécanismes qui pourraient se faire *via* des facteurs encore méconnus ou par la présence de gènes fortement exprimés au niveau des frontières des TADs, comme c'est le cas chez *D. melanogaster* (Sexton *et al.*, 2015 ; Ulianov *et al.*, 2016 ; Dong *et al.*, 2017). Chez les plantes la position des TADs n'est apparemment pas conservée entre les espèces, contrairement aux mammifères (Dixon *et al.*, 2015 ; Dong *et al.*, 2017).

II.5.1.4) Les domaines associés aux lamines

L'enveloppe nucléaire est composée de deux membranes. La membrane interne des animaux est recouverte par la lamina, un réseau protéique composé de lamines en complexe avec d'autres protéines (Turgay *et al.*, 2017). Ces protéines sont capables d'interagir avec la chromatine et plus spécifiquement avec des marques répressives telles que H3K9me3 (Ye *et al.*, 1996 ; Zuleger *et al.*, 2013 ; Solovei *et al.*, 2013). Cette interaction entre lamines et chromatine définit les domaines associés à la lamina (LADs). Les LADs sont composés principalement de régions génomiques réprimées (Solovei *et al.*, 2013 ; Luperchio *et al.*, 2018). Ces domaines peuvent être séparés en deux catégories : les LADs constitutifs, enrichis en éléments répétés et les LADs variables, constitués de gènes qui sont exprimés au cours de la différenciation cellulaire (Meuleman *et al.*, 2013). Les plantes ne produisent pas de protéines lamines. Cependant, il a été montré chez *A. thaliana* que NUCLEOPORIN1 (NUP1), qui fait partie du NPC, s'associe avec l'hétérochromatine, et que les protéines CROWDED NUCLEI (CRWN), des « lamin-like », sont nécessaires à la localisation de l'hétérochromatine à la périphérie du noyau (Bi *et al.*, 2017 ; Hu *et al.*, 2019).

II.5.1.5) Les domaines associés au nucléole

Dans le noyau, l'hétérochromatine est également retrouvée dans les domaines associés au nucléole (NADs) (Pontvianne and Liu, 2020). D'abord identifiés chez les animaux, les NADs existent également chez *A. thaliana* (Pontvianne *et al.*, 2016). Les NADs présentent les mêmes caractéristiques que les LADs : enrichissement en gènes réprimés, TEs et marques répressives comme H3K9me2.

II.5.2) La régulation de l'expression des gènes et des TEs

II.5.2.1) La répression de la transcription

La méthylation de l'ADN et H3K9me2 jouent un rôle dans la répression transcriptionnelle des séquences d'ADN répétées et des TEs, ceci afin de réprimer la transposition de ces derniers qui pourrait menacer l'intégrité du génome (Slotkin and

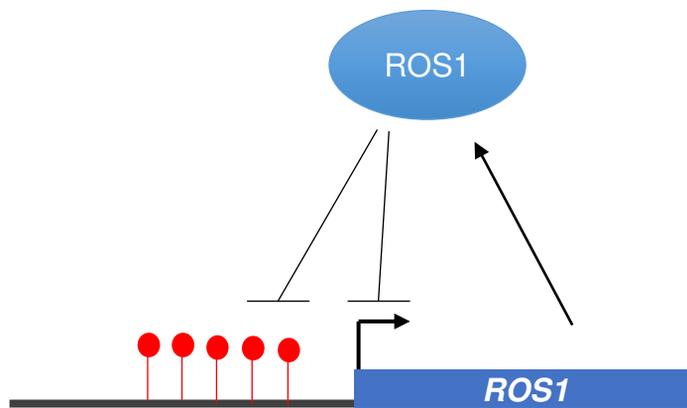


Figure 15 : La régulation de l'expression de *ROS1* par la méthylation de l'ADN.

La méthylation de l'ADN sur le promoteur de *ROS1* active l'expression du gène, la protéine ROS1 retire par la suite cette méthylation, ce qui réprime l'expression de *ROS1*. Ce phénomène est appelé « methylstat ».

Martienssen, 2007). Chez les plantes, l'importance de ces marques épigénétiques a été montrée par l'analyse de plantes combinant les mutations *met1*, *dmr1*, *dmr2* et/ou *cmt3* dans lesquelles une augmentation de la transposition des TEs est observée (Kato *et al.*, 2003 ; Zhang and Jacobsen, 2006).

Ces modifications épigénétiques peuvent également être présentes sur les régions promotrices des gènes, et contribuent globalement, en combinaison avec d'autres marques épigénétiques, à la mise en place d'un environnement épigénétique réprimant la transcription (Weber *et al.*, 2007 ; Domcke *et al.*, 2015 ; Zhu *et al.*, 2016 ; Yin *et al.*, 2017).

II.5.2.2) L'activation de la transcription

Dans les régions d'euchromatine transcriptionnellement active, l'expression des gènes est fortement corrélée à des modifications épigénétiques permissives, comme les marques H3K4me2/3 et H3K36me3 ou les variants d'histone H3.3 et H2A.Z. Ces modifications permettent d'activer la transcription en rendant l'ADN plus accessible aux facteurs de transcription, à la machinerie transcriptionnelle ainsi qu'à une myriade de co-facteurs protéiques ou autres facteurs de remodelage comme le complexe Nucleosome Remodeling Factor (NURF), qui aident à la transcription (Deal and Henikoff, 2011 et partie VII de cette introduction).

Bien que généralement répressive, il a été montré que, dans certains cas, la méthylation de l'ADN pouvait avoir un rôle d'activateur transcriptionnel. Chez *A. thaliana*, la méthylation du promoteur de *ROS1* constitue un « thermostat » régulant positivement l'expression de ce gène, afin de contrôler le niveau global de 5mC du génome (Williams *et al.*, 2015 ; Lei *et al.*, 2015) (figure 15). Enfin, il a également été montré que la méthylation de l'ADN sur certains promoteurs permet le recrutement des protéines SUVH1/3 qui à leur tour recrutent deux protéines DNAJ1 et DNAJ2 afin d'initier la transcription (Harris *et al.*, 2018).

II.5.2.3) La "gene body methylation" (GbM)

Chez *A. thaliana*, un tiers des gènes présente une méthylation de type CG, enrichie au sein de leur partie codante (Zhang *et al.*, 2006). Il a initialement été montré

que la GbM n'avait pas d'effet sur l'expression génique. En effet, dans un mutant *met1*, l'hypométhylation en CG de ces régions géniques ne s'accompagne pas d'une variation de l'expression des gènes (Zhang *et al.*, 2006). En outre, pour un gène donné, une variation de GbM entre différentes populations naturelles d'*A. thaliana* n'est pas corrélée avec des variations de son niveau d'expression (Kawakatsu *et al.*, 2016). Une étude récente chez *A. thaliana* a montré que, conjointement à l'histone « linker » H1, la GbM inhibe la transcription cryptique à l'intérieur de gènes (Choi *et al.*, 2019). Chez les plantes, la GbM est plus fréquente sur les gènes constitutivement exprimés et évoluant lentement ce qui pourrait lier la GbM au rythme d'évolution du génome de différentes espèces (Kim and Zilberman, 2014 ; Bewick and Schmitz, 2017). Cette relation entre GbM et inhibition de la transcription cryptique a aussi été observée chez la souris, de concert avec la marque H3K36me3 qui permet le recrutement de DNMT3B sur ces régions (Neri *et al.*, 2017). De plus, toujours chez les mammifères, la GbM influencerait sur l'épissage des gènes (Gelfman *et al.*, 2013 ; Yang *et al.*, 2014 ; Shayevitch *et al.*, 2018).

Chez les plantes, l'influence de la méthylation sur la maturation des ARN messagers (ARNm) semble aussi assurée par la méthylation, cette fois dans tous les contextes de séquence des introns contenant des TEs ou des séquences répétées (Wang *et al.*, 2013 ; Duan *et al.*, 2017). Par exemple, la formation d'hétérochromatine sur les régions introniques de *INCREASE IN BONSAI METHYLATION 1 (IBM1)* aide au recrutement d'un complexe, contenant notamment IBM2, mettant en place la polyadénylation des ARNm de *IBM1* (Wang *et al.*, 2013; Duan *et al.*, 2017).

II.5.3) Modifications épigénétiques et développement

II.5.3.1) L'empreinte parentale

Au cours du développement de l'organisme, certains gènes sont soumis à l'empreinte parentale, un phénomène au cours duquel l'allèle maternel « maternally expressed gene » (MEG) ou l'allèle paternel « paternally expressed gene » (PEG) est spécifiquement exprimé (Gehring, 2013). Elle peut être maintenue tout au long du développement de l'organisme : au sein de l'embryon et de son tissu nourricier tel que le placenta chez les placentaires, ou au sein de l'individu adulte (Gehring, 2013). Les

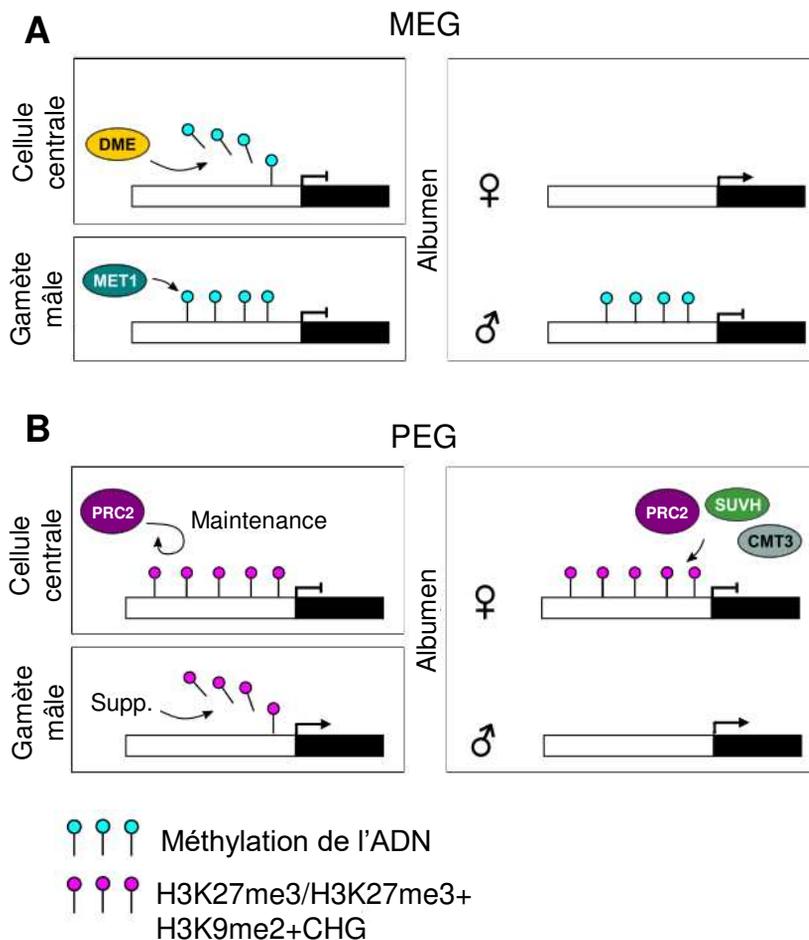


Figure 16 : Le mécanisme d’empreinte parentale chez les plantes.

Représentation de deux modèles de régulation épigénétique des gènes maternellement (A) ou paternellement (B) exprimés. (A) L’allèle maternel est activement déméthylé par DME dans la cellule centrale tandis que l’allèle paternel est maintenu méthylé par MET1, résultant en l’expression spécifique de l’allèle maternel dans l’albumen. (B) L’allèle maternel est marqué par H3K27me3 déposé par PRC2, tandis que cette marque est retirée de l’allèle paternel, résultant en l’expression spécifique de l’allèle paternel dans l’albumen. L’allèle maternel est maintenu réprimé par l’action combinée de PRC2, SUVH et CMT3. Adaptée de Batista and Köhler, 2020.

modifications épigénétiques jouent un rôle important dans ce phénomène d’empreinte parentale (Gehring, 2013).

Chez les plantes à fleurs, l’empreinte parentale est principalement retrouvée dans l’endosperme (ou albumen), tissu nourricier de la graine, avec seulement quelques gènes soumis à l’empreinte parentale dans l’embryon (Satyaki and Ghering, 2017). L’empreinte parentale est mise en place au cours de la gamétogenèse, par plusieurs voies épigénétiques. La méthylation de l’ADN et la marque H3K27me3 permettent la répression des allèles paternels et maternels respectivement (Batista and Köhler, 2020) (figure 16). Pour certains MEGs spécifiques de l’endosperme, leur expression repose sur l’action antagoniste de DME et MET1 (Hsieh *et al.*, 2011 ; Satyaki and Ghering, 2017 ; Batista and Köhler, 2020). En effet, la mutation de l’un ou de l’autre provoque des défauts d’empreinte sur ces MEGs (Kinoshita *et al.*, 2004 ; Hsieh *et al.*, 2011 ; Vu *et al.*, 2013). La mise en place correcte de l’empreinte parentale est essentielle pour le développement de la graine. En effet, des défauts d’empreinte parentale peuvent provoquer l’avortement des graines chez *A. thaliana* (Choi *et al.*, 2002).

II.5.3.2) La transition florale chez A. thaliana

Chez *A. thaliana*, le gène *FLOWERING LOCUS C (FLC)* code un facteur de transcription de type « MADS-Box » inhibant la floraison, si bien que la répression de *FLC* est nécessaire pour accomplir la transition florale (Michaels and Amasino, 1999). Cette répression est réalisée par deux voies permettant la mise en place de modifications épigénétiques : la voie de la vernalisation et la voie autonome (Sharma *et al.*, 2020; Wu *et al.*, 2020) (figure 17). Dans la voie de la vernalisation, une exposition prolongée au froid permet l’activation du gène *VERNALIZATION INSENSITIVE3 (VIN3)*, codant pour un remodeleur de la chromatine, et la transcription du long ARNnc (ARNnc) *COLD AIR*. Ces deux composants agissent ensuite de concert avec le complexe polycomb repressive complex 2 (PRC2) pour permettre le dépôt de H3K27me3 et l’éviction de H3K4me3 et de H3K36me3, conduisant à la répression de *FLC* (Sharma *et al.*, 2020). Dans la voie autonome, la polyadénylation de l’ARNnc antisens *COOL AIR* par un complexe protéique composé de *FLOWERING CONTROL LOCUS A (FCA)*, *FPA* et *FY* permet le recrutement de *FLOWERING LOCUS D (FLD)*,

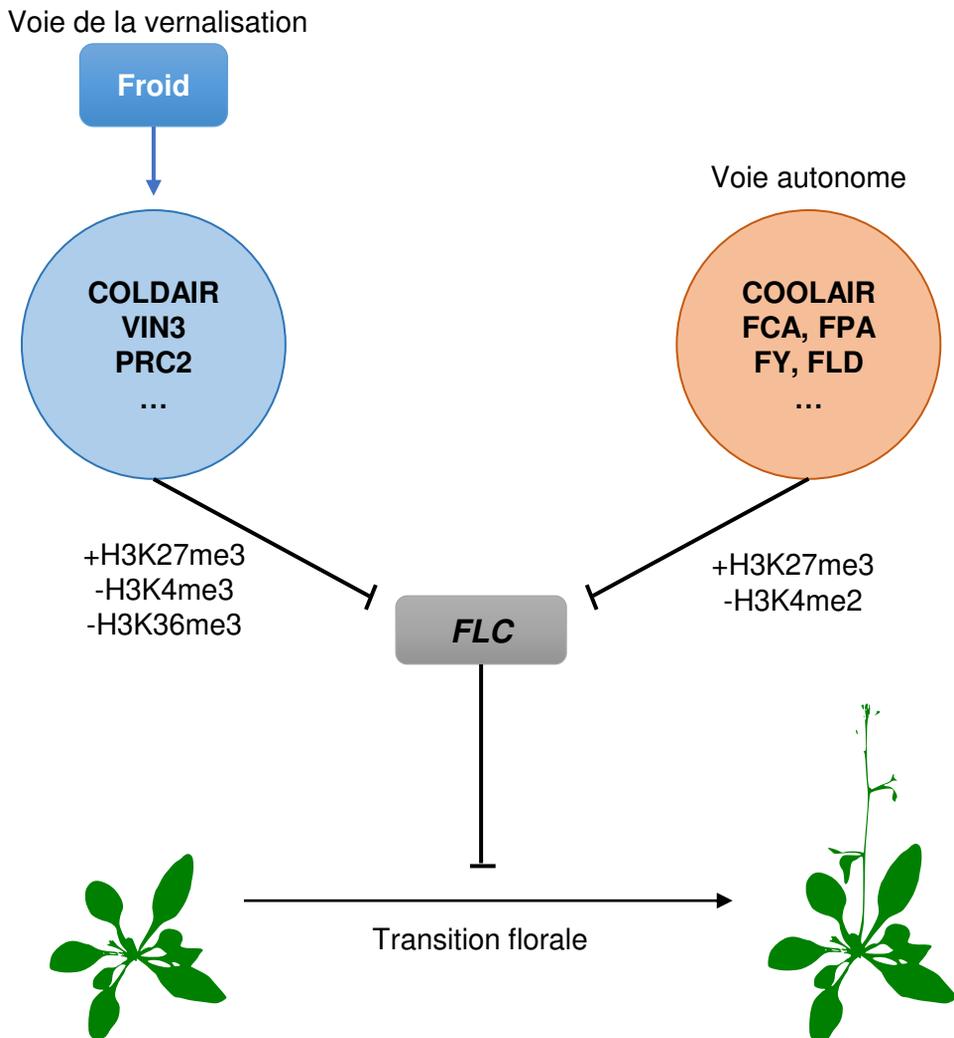


Figure 17 : La répression de *FLC* est assurée par plusieurs voies indépendantes.

La voie de la vernalisation et la voie autonome convergent pour réprimer *FLC* et permettre la transition florale. La voie de la vernalisation, activée par le froid, réprime *FLC* via le dépôt de H3K27me3 et l'éviction de H3K4me3 et H3K36me3. La voie autonome, quand a elle, permet la répression de *FLC* via le dépôt de H3K27me3 et l'éviction de H3K4me2 grâce à, entre autres, l'activité du long ARN non codant COOLAIR et des protéines « RNA binding » FLOWERING CONTROL LOCUS A (FCA) et FPA .

une H3K4me2 déméthylase. L'action de FLD se produit en parallèle de l'ajout de H3K27me3 par PRC2 conduisant à la répression de *FLC* et ainsi au processus de floraison (Wu *et al.*, 2020).

II.5.4) modifications épigénétiques et réponse aux stress

Chez les plantes, les modifications épigénétiques jouent également un rôle dans la réponse aux stress biotiques et abiotiques. Chez *A. thaliana* et le soja, une hypométhylation de l'ADN peut être observée suite à une infection par des nématodes *Heterodera schachtii* et *Heterodera glycines*, respectivement (Rambani *et al.*, 2015 ; Hewezi *et al.*, 2017). Chez *Medicago truncatula*, la formation des nodules nécessite DME (Satge *et al.*, 2016). Chez le riz, une hyperméthylation de gènes de réponse au stress est observée en cas de carence en phosphate (Secco *et al.*, 2015).

Les modifications d'histone et la méthylation de l'ADN agissent dans la « mémoire » épigénétique du stress. Par exemple, Chez *A. thaliana*, les changements dans la méthylation de l'ADN causés par un stress salin peuvent être partiellement transmis à la descendance (Wibowo *et al.*, 2016). Toujours lors d'un stress salin, une diminution de la marque H3K27me3 sur certains gènes a été observée, conférant une meilleure résistance au stress (Sani *et al.*, 2013).

III) Facteurs épigénétiques

En plus de la méthylation de l'ADN et des modifications d'histones, d'autres facteurs sont impliqués dans la régulation de la transcription, la répression des TEs et l'organisation de la chromatine. Ces facteurs agissent de concert avec les modifications épigénétiques pour, par exemple, réguler l'accessibilité de la machinerie transcriptionnelle à l'ADN. De par leur fonctionnement, ces facteurs sont généralement désignés comme des facteurs épigénétiques.

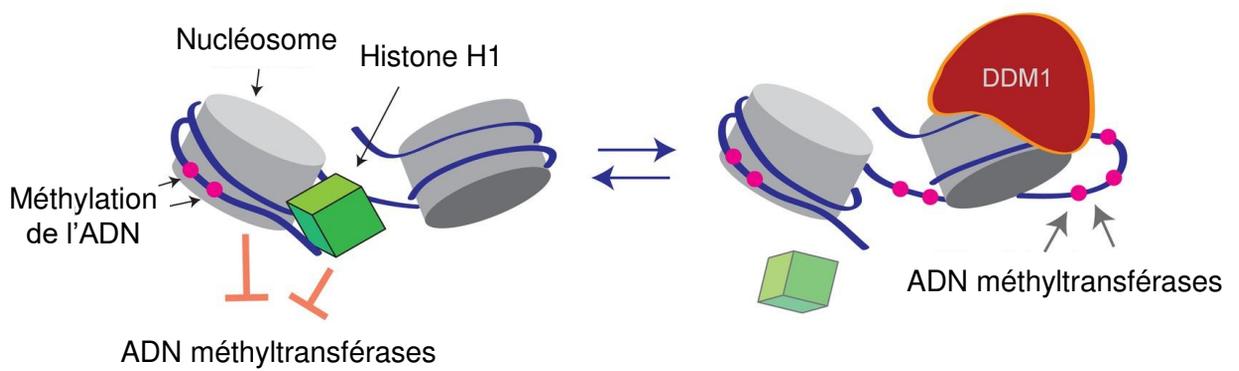


Figure 18 : Mode d'action de DDM1 chez *A. thaliana*.

En condition normale, la méthylation de l'ADN est inhibée par les histones H1. La fixation de DDM1 sur la chromatine permet l'éviction des histones H1 et ainsi l'accès aux ADN méthyltransférases. Lyons and Zilberman, 2017.

III.1) Les complexes remodeleurs de la chromatine

Les remodeleurs de la chromatine sont des complexes enzymatiques capables de catalyser des modifications de la chromatine comme le déplacement des nucléosomes le long de l'ADN, des changements de conformation de l'ADN dans le nucléosome, ou encore l'incorporation de variants d'histones au sein du nucléosome (Rippe et al., 2007). Tous ces processus permettent de moduler l'accessibilité de différents facteurs à l'ADN (Clapier and Cairns, 2009 ; Tyagi et al., 2016 ; Clapier et al., 2017). Ils peuvent être séparés en quatre familles. La famille « chromodomain-helicase-DNA binding » (CHD) est caractérisée par un « chromatin organisation » (chromo) domaine qui lie spécifiquement les histones modifiées et un domaine hélicase ATP-dépendent, facilitant le déplacement des nucléosomes (Marfella and Imbalzano 2007 ; Mills, 2017). La famille « inositol requiring 80 » (INO80) est caractérisée par la présence d'un anneau hexamérique de protéines ATPases RuvB-like (Willhoft and Wigley, 2020). La capacité de remodelage de la famille INO80 dépend de l'action de protéines « actin-related » (Arps), qui sont capables de se lier préférentiellement sur certains types d'histones (Harata et al., 1999 ; Shen et al., 2003 ; Downs et al., 2004). La famille « switch/sucrose-non-fermenting » (SWI/SNF) agit en déplaçant, en remplaçant les nucléosomes, ou en retirant le nucléosome *via* son activité ATPase (Zhou et al., 2016 ; Tang et al., 2010). Certains membres de cette famille sont également caractérisés par un bromodomaine, un domaine « reader » capable de se lier aux histones acétylées (Ribeiro-Silva et al., 2019). Chez *A. thaliana*, la protéine SWI/SNF DECREASE IN DNA METHYLATION 1 (DDM1) est requise pour la maintenance de la méthylation de l'ADN, la condensation de l'hétérochromatine et la répression des TEs (Jeddeloh et al., 1999 ; Probst et al., 2003 ; Soppe et al., 2002 ; Feng et al., 2014). Il a été proposé que DDM1 est capable de retirer les histones H1 de l'ADN pour permettre l'accès aux ADN méthyltransférases et ainsi assurer la maintenance de la méthylation de l'ADN (Zemach et al., 2013) (figure 18). Enfin, la famille « imitation switch » (ISWI) possède un domaine ATPase proche de la famille SWI/SNF (Elfring et al., 1994). Les remodeleurs de cette famille portent un domaine ATPase SWI2/SNF2 et un domaine « HANDSANT-SLIDE » capable de se lier à l'ADN (Petty and Pillus, 2013). Cette famille permet l'assemblage de la chromatine en altérant la position des nucléosomes, ce qui amène à la répression de la transcription. Les

remodeleurs ISWI sont également impliqués dans l'assemblage *de novo* des nucléosomes (Tyagi et al., 2016).

En plus de ces remodeleurs, d'autres ATPases jouent un rôle dans l'accessibilité et la compaction de l'ADN. Les protéines Structural Maintenance of Chromosomes (SMC) forment trois complexes protéiques. SMC1 et SMC3 forment, en s'associant à d'autres protéines, les complexes « cohésine » qui assurent la cohésion entre les chromatides sœurs pendant la mitose ou la méiose (Peters et al., 2008). Ces complexes peuvent aussi réguler la transcription (Birkenbihl and Subramani, 1995 ; Guacci et al., 1997). SMC2 et SMC4 forment les complexes « condensine » qui interviennent dans la condensation des chromosomes et la répression des gènes (Hirano et al., 1997 ; Ono et al., 2003). Enfin, SMC5 et SMC6 constituent un complexe intervenant dans la réparation de l'ADN (Fousteri and Lehmann, 2000).

III.2) MORPHEUS' MOLECULE 1 (MOM1)

MORPHEUS' MOLECULE 1 (MOM1) est une protéine spécifique des plantes, nécessaire pour la répression des TEs (Amedeo *et al.*, 2000). MOM1 porte un domaine ATPase proche de celui de CHD3 mais qui n'est pas nécessaire pour réprimer les TEs (Čaikovski *et al.*, 2008). Inversement, son domaine conservé « Conserved MOM1 Motif 2 (CMM2) » est essentiel pour son activité de répression des TEs, ainsi que pour la dimérisation de MOM1 (Nishimura *et al.*, 2012). De plus, CMM2 permet l'interaction avec les protéines PROTEIN INHIBITOR OF ACTIVATED STAT (PIAS)-type SUMO E3 ligase-like 1 (PIAL1) et PIAL2, elles-mêmes nécessaires pour réprimer les TEs (Nishimura *et al.*, 2012; Han *et al.*, 2016).

MOM1 n'est pas requise pour la maintenance de la méthylation de l'ADN, ni pour la condensation des chromocentres (Probst et al., 2003, Feng et al., 2014). Elle agit en synergie avec la voie du RdDM pour réprimer certains loci hétérochromatiques (Yokthongwattana *et al.*, 2010). La protéine MOM1 définit une voie de répression des TEs dont le mode d'action est encore mal connu.

III.3) La famille des protéines MICRORCHIDIA (MORC)

Les protéines MICRORCHIDIA (MORC) sont des ATPases de type GHKL (Gyrase, HSP90, histidine kinase, MutL) présentent chez les procaryotes et les eucaryotes (Iyer et al., 2008). Ces protéines sont impliquées dans des processus de régulation de la transcription et de répression des TEs, dans l'organisation de la chromatine et la réparation de l'ADN (Iyer et al., 2008 ; Dong et al., 2018 ; Li et al., 2013 ; Koch et al., 2017).

Chez la souris, la protéine MORC1 est requise pour la fertilité mâle ; une souris mâle mutante *morc1* présente des petits testicules (Watson PNAS 1998, Inoue et al., 1999), le nom *microrchidia* est donc depuis utilisé comme terme médical désignant des petits testicules anormaux (Inoue et al., 1999). Bien plus tard, et suite à des découvertes obtenues chez *A. thaliana* et *C. elegans*, il a été démontré que MORC1 est requise pour la répression des TEs et la méthylation de l'ADN dans la lignée germinale de souris mâle (Pastor *et al.*, 2014).

C'est donc initialement chez *A. thaliana*, qu'il a été découvert que MORC1, MORC6 (et dans une moindre mesure MORC2) sont nécessaires pour la répression des TEs et des gènes dont l'ADN est méthylé, ainsi que pour la condensation de l'hétérochromatine (Moissiard et al., 2012 ; 2014 ; Lorkovic et al, 2012 ; Brabbs et al., 2013). Les protéines MORC1 et MORC6 sont localisées à la périphérie des chromocentres et forment des hétérodimères de type MORC1/6 ou MORC2/6 (Moissiard *et al.*, 2012, 2014). Bien que ces protéines ne soient pas nécessaires pour le maintien de la méthylation de l'ADN ou de H3K9me2, plusieurs études ont montré qu'elles interagissent avec des acteurs du RdDM (Lorkovic et al, 2012 ; Liu et al., 2014 ; Liu et al., 2016). Ainsi, il a été proposé que le complexe MORC1/6 agirait en amont du RdDM, et que son interaction avec certains facteurs du RdDM permettrait de renforcer le recrutement de cette voie épigénétique au niveau de ces loci réprimés (Gallego-Bartolome et al., 2019). De plus, les protéines MORC1/6 agissent en synergie avec MOM1 dans la répression des TEs (Moissiard et al., 2014). Les protéines MORC4 et MORC7, quant à elles, agissent de manière redondante et forment des homodimères pour réguler l'expression des gènes indépendamment de

MORC1/6. MORC4 et MORC7 sont impliquées dans la réponse contre le pathogène *Hyaloperonospora arabidopsidis* (Harris et al., 2016).

Le génome de *C. elegans* code une seule protéine Morc1 qui est aussi requise pour la répression des gènes (Moissiard et al., 2012). Récemment, il a été montré que la protéine Morc1 forme des complexes multimériques capables de piéger l'ADN sous forme de boucle, permettant ainsi la compaction de la chromatine (Kim et al., 2019).

Étant donné que chez tous ces organismes les protéines MORCs ont une fonction conservée, nous pouvons donc supposer que le mode d'action de Morc1 chez *C. elegans* soit également conservé chez les autres espèces.

III.4) Les ARN non-codants (ARNnc)

Chez les eucaryotes, les ARNnc interviennent dans de nombreux processus cellulaires, comme le contrôle de l'expression des gènes et la répression des TEs. Les ARNnc peuvent agir à deux niveaux pour moduler l'expression génique : au niveau post-transcriptionnel (post-transcriptional gene silencing ou PTGS), en jouant un rôle dans la dégradation des transcrits ou dans la répression de la traduction (Filipowicz et al., 2005), ou au niveau transcriptionnel (transcriptional gene silencing ou TGS), en contribuant à la mise en place de marques épigénétiques répressives (Kawasaki et al., 2005 ; Morris, 2005). Il existe plusieurs types d'ARNnc, qui sont classés selon leur taille et origine.

Les siARNs, d'une taille comprise entre 21- et 24-nt, sont issus du clivage par les protéines DICER d'un ARNdb ou d'une structure d'ARN de type « tige-boucle ». Ils sont pris en charge par les protéines AGO spécialisées au sein des « RNA-induced silencing complex » (RISC) ou « RNA-induced transcriptional silencing (RITS) complex » (Matzke and Mosher, 2014). Chez les plantes, les siARNs sont classés en plusieurs types, selon leur origine et leur mode d'action (Vaucheret et al., 2006 ; Borges et Martienssen, 2015). Comme évoqué précédemment, la voie du RdDM canonique implique majoritairement des siARNs de 24-nt, bien que des siARN de 21- et 22-nt aient été impliqués dans des voies RdDM non-canoniques (Zheng *et al.*, 2009 ; Nuthikattu *et al.*, 2013 ; Wu *et al.*, 2012 ; Matzke and Mosher, 2014).

Les micro-ARNs (miARNs) sont généralement d'une taille de 21- ou 22-nt. Ils sont codés par des gènes spécifiques, les gènes *MIR* (Bartel, 2004 ; Naqvi et al., 2012). Impliqués dans la voie du PTGS en dégradant des ARNm cibles ou en réprimant la traduction, les miARNs jouent des rôles essentiels dans le développement de l'organisme, notamment dans la formation des feuilles, la croissance racinaire et la transition florale chez les plantes (Souret et al., 2004 ; Branscheid et al., 2015 ; Yu et al., 2017 ; Vaucheret et al., 2004).

Les piwi-interacting ARNs (piARNs) sont de petits ARN absents des plantes. Ils sont impliqués dans la répression des TEs et jouent un rôle important dans la gamétogenèse chez les animaux (Aravin et al., 2007 ; Brennecke et al., 2007 ; Siomi et al., 2011). Ils peuvent induire le dépôt de H3K9me3 sur les clusters piARNs chromatinien pour permettre la transcription des précurseurs des piARNs, constituant ainsi une boucle d'autorégulation (Le Thomas et al., 2014). Chez *C. elegans*, les piARNs et les siARNs endogènes (endo-siARNs), exprimés dans les cellules germinales, jouent un rôle important dans l'hérédité de marques épigénétiques telles que H3K9me3, permettant ainsi de maintenir les régions d'hétérochromatine entre les générations (Weiser and Kim, 2019).

Finalement, les longs ARNs non codants (ARNlncs), sont des ARNs de plus de 200-nt, majoritairement transcrits par l'ARN Pol II (Ulitsky and Bartel, 2013). Cependant, ils peuvent aussi être transcrits par l'ARN Pol III et, chez les plantes, par l'ARN Pol IV et l'ARN Pol V (Dieci et al., 2007 ; Wierzbicki et al., 2008 ; 2009). Ils sont impliqués dans de nombreux mécanismes de régulation de la transcription *via* des modifications épigénétiques, mais aussi au niveau post-transcriptionnel (Zhu and Wang, 2012 ; Marín-Béjar and Huarte, 2015). Chez les mammifères, l'ARNlnc *Xist* participe à la répression du chromosome X en permettant le recrutement de HDAC3 et de PRC2 (McHugh et al., 2015). Chez *A. thaliana*, les ARNlnc COOLAIR et COLDAIR agissent avec les facteurs FCA, FLD et VIN3 pour réprimer le répresseur floral *FLC* *via* le dépôt de H3K27me3 et l'éviction de H3K4me2 (Csorba et al., 2014 ; Wu et al., 2016 ; Tian et al., 2019).

III.5) le « mille-feuille » épigénétique

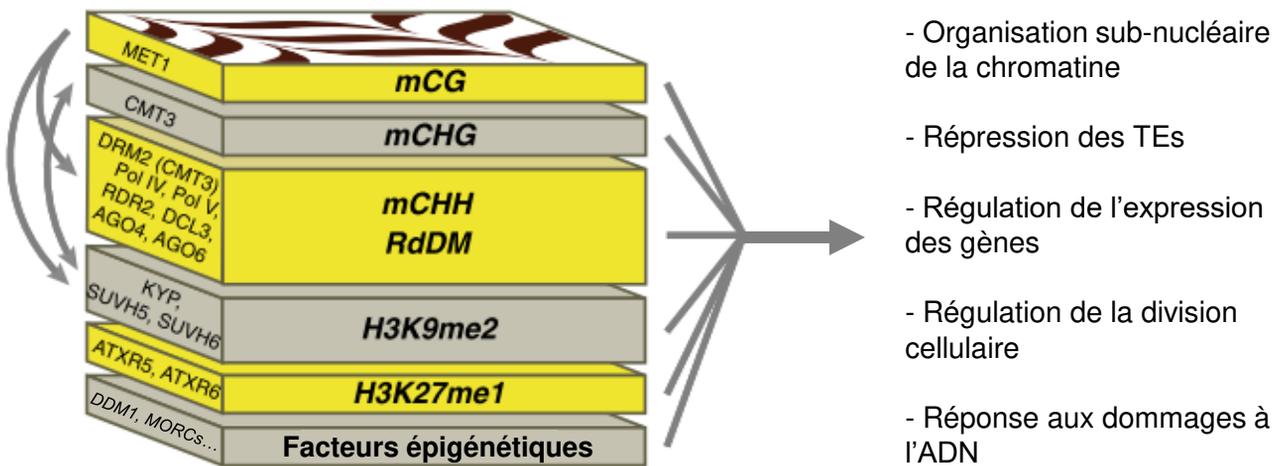


Figure 19 : Le « mille-feuille » épigénétique.

Représentation des différents niveaux de régulation composant le « mille-feuille » épigénétique. Les flèches à gauche représentent les interactions entre les différentes voies épigénétiques. Modifiée de Rigal and Mathieu, 2011.

En parallèle à la méthylation de l'ADN et aux MPTs des histones, les voies épigénétiques présentées dans cette partie sont toutes impliquées dans la régulation de l'expression des gènes, la répression des TEs, ainsi que de nombreux autres mécanismes chromatinien tels que la réparation de l'ADN ou la division cellulaire. Ensemble, ces voies forment une succession de couches, dite « mille-feuille » épigénétique, qui convergent pour maintenir une bonne organisation de la chromatine et pour réguler l'expression du génome (Rigal and Mathieu, 2011) (figure 19). Comme nous l'avons vu, ces nombreuses voies et marques épigénétiques sont capables de s'autoréguler entre elles, comme cela a, par exemple, été décrit avec la boucle d'autorégulation de la méthylation en CHG et de la marque H3K9me2 chez *A. thaliana* (Du *et al.*, 2015).

IV) Les éléments transposables (TEs)

Les éléments transposables (TEs) sont des séquences d'ADN répétées ayant la capacité de se déplacer au sein du génome. Ce sont des composants majeurs des génomes eucaryotes, constituant jusqu'à 80% du génome du maïs et 90% du génome du blé (Devos *et al.*, 2005). Les TEs ne sont pas répartis aléatoirement au sein des génomes, en outre, ils sont enrichis dans les régions hétérochromatiques. Chez *A. thaliana*, les TEs représentent environ 20% du génome et sont principalement enrichis dans les régions péricentromériques et exclus des bras chromosomiques (Cokus *et al.*, 2008 ; Joly-Lopez and Bureau, 2014) (figure 20).

IV.1) Découverte des TEs

Les TEs ont été décrits pour la première fois par Barbara McClintock chez le maïs (McClintock, 1950). En s'intéressant au bras court du chromosome 9, elle observe qu'un locus particulier est très souvent clivé, et que le clivage de ce locus est associé à l'apparition de phénotypes récessifs. Ce locus est nommé *dissociation* (*Ds*),

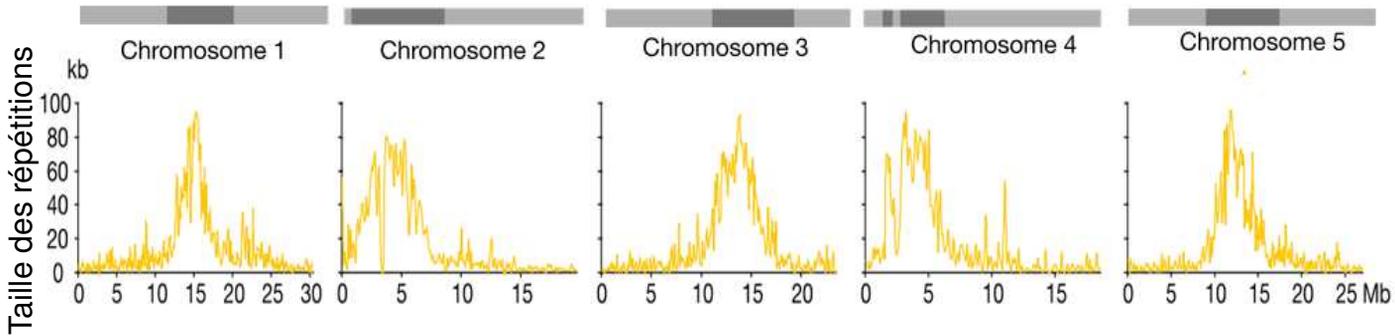


Figure 20 : Répartition des éléments d'ADN répétés et TEs le long des chromosomes chez *A. thaliana*.

Représentation graphique de la répartition des éléments répétés et TEs le long des cinq chromosomes d'*A. thaliana*. Les chromosomes sont représentés schématiquement au-dessus des graphiques. Les bras chromosomiques et les régions péri-centromériques sont représentés en gris clair et gris foncé respectivement. Mb : Mégabases. Modifiée de Cokus *et al.*, 2008.

et elle montre que *Ds* est capable de transposition en se déplaçant dans le génome. C'est le déplacement de *Ds* proche de gènes qui provoque l'apparition des phénotypes récessifs, cela va définir la notion de « controlling elements », des éléments qui ne font pas partie de la région génique mais qui sont capables de contrôler l'expression d'un gène (McClintock, 1950; Jones, 2005). Lors de cette étude, elle démontre également que *Ds* nécessite un autre locus pour se déplacer, elle nomme ce locus *Activator* (*Ac*) et montre qu'il est également capable de transposer dans le génome (McClintock, 1950). Ces deux loci seront par la suite décrits comme des éléments transposables (Fedoroff *et al.*, 1984; Jones, 2005). Suite à la découverte des TEs et du phénomène de transposition, Barbara McClintock reçut en 1983 le prix Nobel de physiologie ou médecine.

IV.2) Les différentes classes d'élément transposables

Les TEs sont divisés en deux grandes classes basées sur leur mécanisme de transposition, et chaque classe est elle-même divisée en superfamilles basées sur la structure des TEs. Dans chaque superfamille, les TEs sont répartis en TEs autonomes, c'est-à-dire qu'ils codent toutes les protéines nécessaires à leur transposition, et en TEs non-autonomes, qui utilisent la machinerie d'autres TEs pour se déplacer (Wicker *et al.*, 2007) (figure 21).

IV.2.1) Classe I : les rétrotransposons

Les TEs de classe I, ou rétrotransposons, transposent par un mécanisme dit de « copié-collé », *via* un intermédiaire ARN. Une partie des ARNm issue du TE seront traduits pour permettre la synthèse des protéines nécessaires à la formation d'une pseudo-particule virale dans laquelle un ARNm sera intégré et rétrotranscrit en ADN double brin. Cet ADN sera ensuite intégré à une autre position du génome (Boeke *et al.*, 1985) (figure 22).

Les rétrotransposons sont séparés en deux sous-classes : Les rétrotransposons à LTR (Long Terminal Repeat) et les rétrotransposons sans LTR.

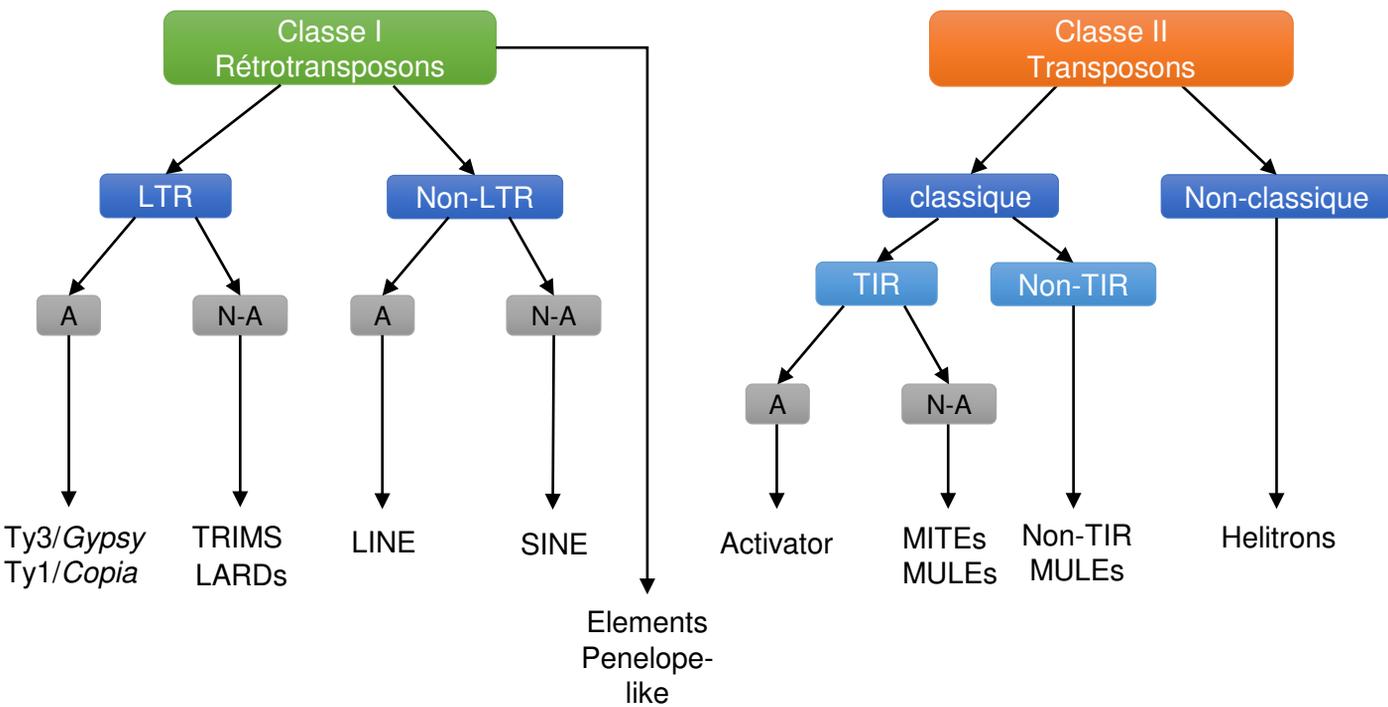


Figure 21 : Les TEs sont répartis en plusieurs superfamilles réparties en deux classes. Les TEs sont séparés en deux groupes selon leur mode de transposition : la classe I transpose selon le mode du « copié-collé » et la classe II transpose selon le mode du « coupé-collé ». Ils sont à nouveau séparés en sous-classes selon leur structure et enfin selon leur nature autonomes (A) ou non (N-A).

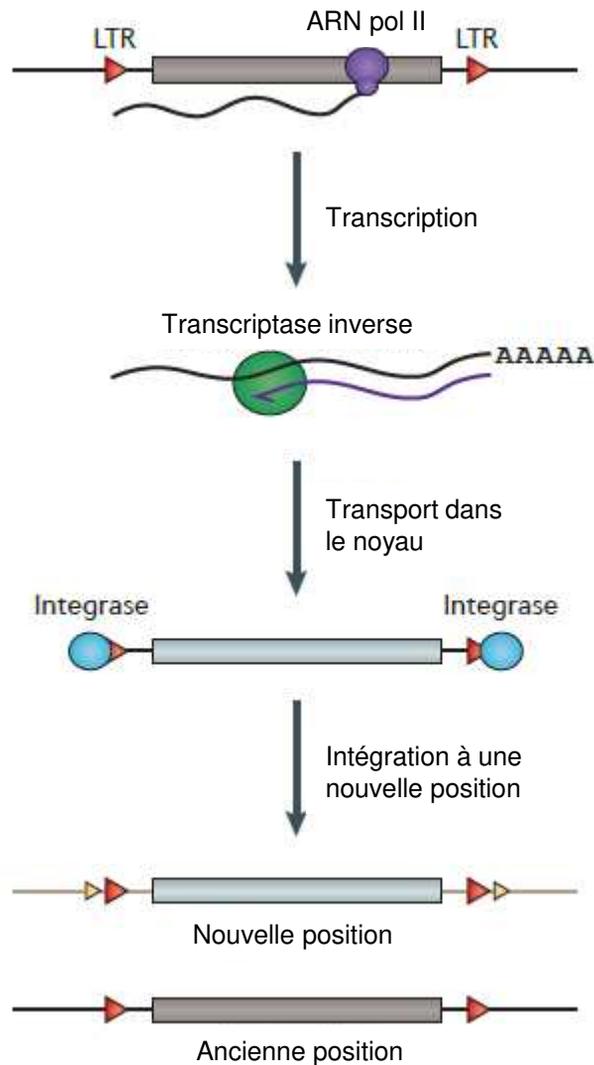


Figure 22 : Mode de transposition des rétrotransposons.

Les rétrotransposons se déplacent selon la méthode du « copié-collé ». Le rétrotransposon est transcrit par l'ARN pol II et son ARN est ensuite rétrotranscrit par une transcriptase inverse. Le rétrotransposon peut alors s'intégrer à une nouvelle position dans le génome grâce à l'activité de sa protéine intégrase. Modifiée de Lisch, 2013.

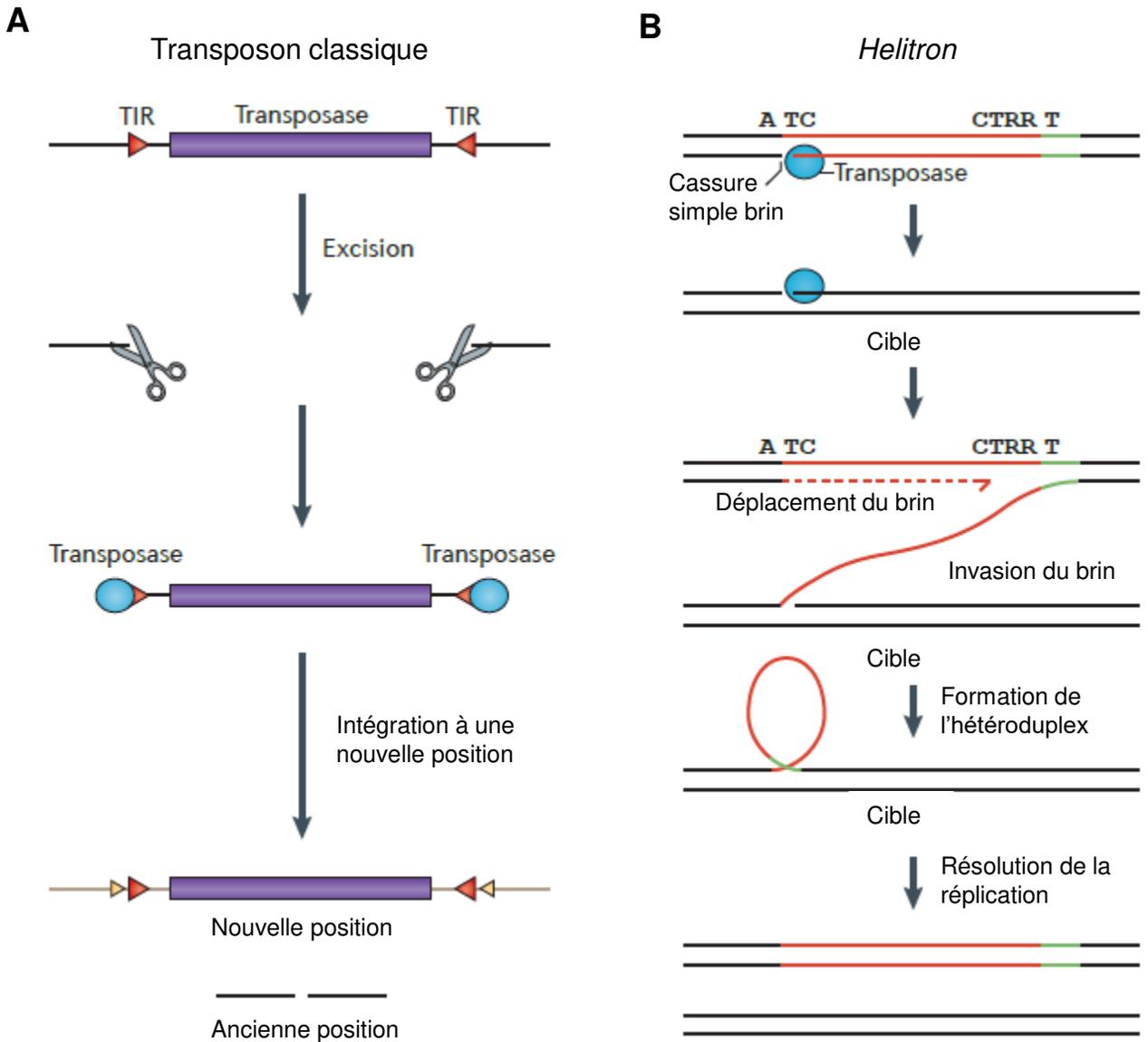


Figure 23 : Mode de transposition des transposons.

(A) Les transposons se déplacent selon la méthode du « coupé-collé ». Le transposon est excisé de sa position initiale par une transposase et il est ensuite intégré à une nouvelle position. (B) Les *Helitrons* se déplacent selon un mécanisme impliquant une cassure simple brin provoquée par la transposase, sur le site du TE et le site cible. Le site cible est ensuite envahi par le TE qui vient former une boucle d'ADN qui sera résolue par des mécanismes de réparation de l'ADN. Modifiée de Lisch, 2013.

Les rétrotransposons à LTR sont prédominants dans le génome des plantes et moins abondants dans celui des animaux (Wicker et al., 2007). D'un point de vue évolutif, ces TEs sont proches des rétrovirus (Frankel and Young, 1998 ; Seelamgari et al., 2004). Au sein de cette sous-classe, les superfamilles Ty3/*Gypsy* et Ty1/*Copia* sont les plus largement représentées au sein des génomes (Bennetzen et al., 2005). Les rétrotransposons sans LTR sont répartis en deux superfamilles : les LINEs (Long Interspersed Nuclear Elements) et les SINEs (Short Interspersed Nuclear Elements). Les LINEs sont présents chez tous les eucaryotes, et sont majoritaires au sein des génomes animaux (Wicker et al., 2007). Les SINEs, quant à eux, sont des TEs dérivés de la transposition accidentelle de transcrits issus de l'ARN polymérase III (Kramerov and Vassetzky, 2005). Le SINE *Alu* est un des mieux caractérisé et le plus abondant dans le génome humain (Rowold and Herrera, 2000 ; Batzer and Deininger, 2002).

IV.2.2) Classe II : les transposons

Les TEs de classe II, ou transposons, transposent par un mécanisme dit de « coupé-collé ». Des protéines spécifiques sont capables d'exciser le TE d'un site donneur et de permettre son intégration à un autre endroit du génome (Greenblatt and Brink, 1963 ; Rubin et al., 1982) (figure 23).

Les transposons sont séparés en deux sous-classes : Les transposons dits classiques, et les transposons non-classiques. La première sous-classe comporte les transposons à TIR (Terminal Inverted Repeats), dont notamment les *Mutator-like* (MULE), des transposons autonomes ou non. Certains MULEs portent des fragments de gènes issus du génome hôte, on parle alors de « pack-MULEs » (Jiang et al., 2004). La première sous-classe contient également les transposons sans TIR. Tous ces TEs sont capables de se multiplier dans le génome en transposant au cours de la réplication de l'ADN, depuis une région déjà répliquée (Greenblatt and Brink, 1962). Ils peuvent également utiliser le système de réparation suite à une excision pour se multiplier (Nassif et al., 1994). La sous-classe des transposons non-classiques comporte la superfamille des *Helitrons*. Les TEs de cette superfamille transposent en utilisant un mécanisme particulier de réplication circulaire dit « rolling circle », au cours duquel un seul brin d'ADN est clivé (Kapitonov and Jurka, 2001 ; Grabundzija et al.,

2016). Chez le maïs, les *Helitrons* portent des fragments de gènes capturés lors de la transposition, on parle du phénomène de transduplication (Morgante et al., 2005).

IV.3) La contribution des éléments transposables dans le génome

Historiquement, les TEs ont longtemps été considérés comme des éléments parasites, dit ADN « poubelle », des génomes (Ohno, 1972). De par leur capacité à transposer, ils sont généralement vus comme des éléments fortement mutagènes qui doivent être réprimés par leur hôte, afin de maintenir son intégrité. Cette vision a cependant évolué au fil des années, grâce à des études démontrant l'impact positif des TEs sur les génomes (Volf, 2006 ; Oliver and Greene, 2009 ; Fedoroff, 2013). Les TEs sont en effet une importante source de variabilité, d'adaptation et d'innovation génétique. Ils jouent également un rôle essentiel dans la structure des génomes, en effet les régions péri-centromériques et sub-télomériques des chromosomes sont enrichies en TEs qui permettent le maintien de l'hétérochromatine sur ces régions du fait qu'ils soient visés par les voies épigénétiques répressives (Slotkin and Martienssen, 2007).

La présence d'un TE dans une région régulatrice -comme le promoteur- d'un gène peut influencer l'expression de ce dernier. Cette influence peut être répressive, comme c'est le cas chez la vigne où l'insertion d'un rétrotransposon dans le promoteur du gène responsable de la couleur noire des grains réprime ce dernier, donnant des grains blancs dans la variété Chardonnay (Kobayashi *et al.*, 2004; Lisch, 2013). Cette influence peut aussi être activatrice, par exemple chez le phalène du boulot, chez qui l'insertion du TE *carbonaria* dans un intron du gène *cortex* active fortement son expression, donnant une couleur noire au papillon, ce qui lui permet d'échapper aux prédateurs (Hof *et al.*, 2016). Un autre exemple est celui de l'orange sanguine, chez laquelle l'insertion d'un TE dans le promoteur du gène *Ruby* active son expression, après une exposition au froid et donne sa couleur caractéristique au fruit (Butelli *et al.*, 2012; Lisch, 2013).

En parallèle, de nombreuses études menées chez les eucaryotes ont montré que des protéines (ou domaines protéiques) issus de TEs pouvaient jouer des rôles essentiels dans de nombreux processus cellulaires de l'organisme hôte. Par exemple,

chez les vertébrés, les protéines RAGs sont des transposases impliquées dans le système immunitaire (Jangam et al., 2017). De même, chez la drosophile le maintien des télomères est assuré par les TEs *TART* et *heT-A*, apparentés aux LINEs (Jangam et al., 2017). Des exemples ont aussi été décrits chez les plantes, comme les protéines FAR-RED IMPAIRED RESPONSE 1 (FAR1) et FAR1-RELATED SEQUENCES (FRS) qui sont des facteurs de transcription dérivés d'un transposon *MULE* (Lin et al., 2007). Mais aussi les protéines SLEEPER et les protéines MUSTANG (MUG) sont dérivés de transposons *hAT* et *MULE* respectivement et sont essentielles pour le développement d'*A. thaliana* (Bundock and Hooykaas, 2005 ; Joly-lopez et al., 2012). Enfin, les protéines Plant Mobile Domain (PMD) sont possiblement issues d'un rétrotransposon *Ty3/Gypsy* (Ikeda et al., 2017).

D'un point de vue évolutif, il est maintenant clairement accepté que les TEs peuvent avoir un effet positif sur la « fitness » sélective de l'organisme hôte, en particulier en réponse à un changement environnemental important. Ainsi, la cellule hôte peut détourner à son profit une séquence de type TE pour l'utiliser comme séquence régulatrice de ses propres gènes. Elle peut également domestiquer un gène de TE (comme par exemple un gène codant une transposase) pour le détourner de sa fonction originelle. On parle alors de domestication de gène ou de séquence de TEs, un phénomène appelé aussi « co-option » ou « exaptation ». Dans tous les cas, ces processus impliquent une notion d'adaptation de l'organisme à son environnement, à travers la sélection d'un caractère remplissant une nouvelle fonction et ayant un avantage pour la « fitness » sélective de l'organisme hôte (Chuong et al., 2017; Jangam et al., 2017)

V) La génétique classique

L'approche dite de génétique classique (ou crible génétique) est une méthode très largement utilisée chez la plante modèle *A. thaliana* dont le but est d'identifier le gène muté qui est responsable du phénotype mutant étudié. A l'opposé, la génétique inverse étudie la conséquence phénotypique due à la mutation d'un gène qui est connu (Page and Grossniklaus, 2002).

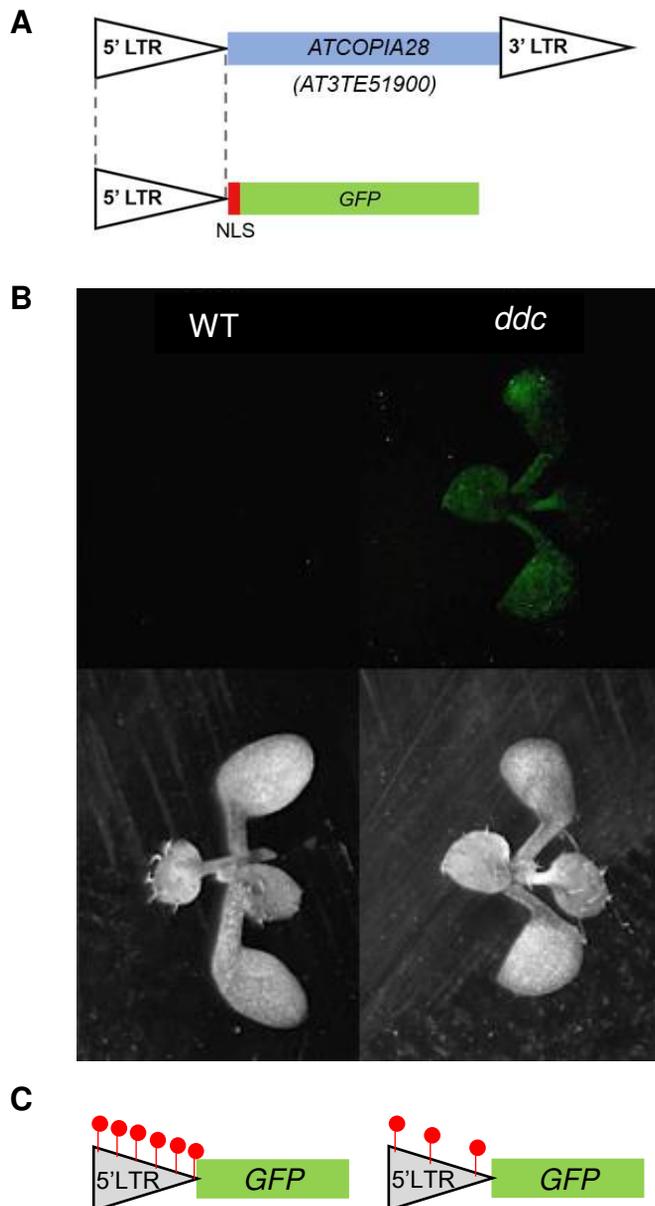


Figure 24 : Construction et matériel végétal utilisé lors du crible génétique.

A. Représentation schématique de la construction *ATCOPIA28::GFP*. La GFP est placée sous le contrôle du 5'LTR du rétrotransposon *AT3TE51900*. Un NLS est ajouté pour optimiser l'accumulation de la GFP dans le noyau et ainsi optimiser la détection de la protéine sous lumière UV. **B.** Photos représentatives des deux lignées *ATCOPIA28::GFP* WT et *ATCOPIA28::GFP ddc*. L'expression de la GFP est observée sous lumière UV. Les mêmes plantes photographiées sous lumière blanche sont montrées en dessous. **C.** Représentation schématique de la construction *ATCOPIA28::GFP* et de son niveau de méthylation de l'ADN dans les deux lignées. Les cercles rouges représentent la méthylation de l'ADN. Ce niveau est corrélé au niveau d'expression de la GFP. Portion de figure extraite de Nicolau *et al.*, 2020.

V.1) Principe général

Le concept du crible génétique repose tout d'abord sur la création de mutations aléatoires au sein du génome. Ces mutations peuvent être provoquées par un traitement à l'éthyle méthanesulfonate (EMS), un agent chimique provoquant principalement des mutations par substitution (C en T et A en G), par traitement aux rayons gamma, ou encore par l'insertion d'ADN de transfert (ADN-T) ou de TEs (par exemple le TE *Ac/Ds* du maïs) (Page and Grossniklaus, 2002). Les populations mutantes obtenues après mutagenèse sont ensuite criblées pour le phénotype d'intérêt. Cela peut être l'apparition d'un phénotype mutant chez une plante sauvage, mais il peut également s'agir d'un phénotype mutant qui est restauré (crible suppresseur), ou d'un phénotype mutant qui est aggravé (crible « enhancer »). Le crible peut aussi se baser sur l'expression d'un gène rapporteur, comme ceux codant la β -glucoronidase (GUS) ou la green fluorescent protein (GFP), placé sous le contrôle du promoteur d'un gène d'intérêt. La génétique classique a permis d'identifier de nombreux gènes impliqués dans différentes voies épigénétiques, notamment *CMT3*, *MORC1* et *MORC6* (Bartee *et al.*, 2001; Lindroth *et al.*, 2001; Moissiard *et al.*, 2012).

V.2) Le système *ATCOPIA28::GFP*

Un des projets développés par l'équipe mécanismes épigénétiques et architecture de la chromatine (MEAC) consiste en l'identification de nouveaux facteurs épigénétiques en utilisant une approche de génétique classique. Le crible se base sur une construction appelée *ATCOPIA28::GFP*, où l'expression de la GFP est sous le contrôle du 5'LTR d'un rétrotransposon de la famille *ATCOPIA28* (*AT3TE51900*) (figure 24.A). Ainsi, grâce à ses régions promotrices, le 5'LTR contrôle la transcription de la GFP. Dans un fond génétique sauvage, *AT3TE51900* est réprimé par plusieurs voies épigénétiques, telles que, entre autres, la méthylation de l'ADN et H3K9me2. De manière similaire, l'ADN du promoteur 5'LTR d'*ATCOPIA28::GFP* est méthylé ; le transgène est donc également réprimé dans le fond sauvage (*ATCOPIA28::GFP* WT) (figure 24.B). Dans le fond mutant *drm1 drm2 cmt3 (ddc)*, l'ADN de *AT3TE51900* est

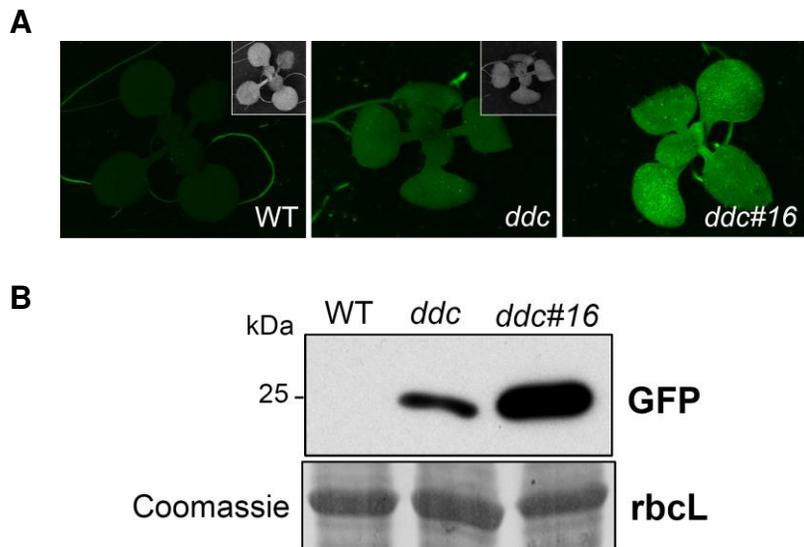


Figure 25 : La lignée *ddc#16* montre une surexpression de la GFP.

(**A**) Photos représentatives de plantes sauvages (WT), *ddc* et *ddc#16* de 10 jours sous lumière UV. Des photos sous lumière blanche des plantes WT et *ddc* sont montrées en encart. (**B**) Western blot des lignées WT, *ddc* et *ddc#16* réalisé avec un anticorps GFP. La grande sous-unité de la Rubisco (*rbcL*) est utilisée comme témoin de charge. La lignée *ddc#16* montre une surexpression de la GFP par rapport à la lignée *ddc* de base. Figure extraite de Nicolau *et al.*, 2020.

partiellement hypométhylé (figure 24.C), ce qui a pour conséquence d'induire une légère activation transcriptionnelle. En outre, *AT3TE51900* peut aussi être activé dans les fonds *morc6*, *morc1* ou *mom1* (Moissiard et al., 2014). L'introduction de *ATCOPIA28::GFP* dans un fond *ddc* a permis de récapituler les phénotypes moléculaires observés pour *AT3TE51900*. Ainsi, une plante *ATCOPIA28::GFP ddc* exprime la GFP à un niveau basal (figure 24.B).

Le matériel *ATCOPIA28::GFP ddc* constitue donc un fond génétique dit « sensibilisé » où les voies épigénétiques DRM2 (RdDM) et CMT3 sont affectées. Ce matériel a été utilisé pour réaliser une mutagenèse à l'EMS dans le but d'identifier de nouvelles voies épigénétiques impliquées dans le silencing des TEs.

V.3) Résultats du crible

Après mutagenèse, les populations mutantes *ATCOPIA28::GFP ddc* ont été criblées sous microscope couplé à une lumière UV afin d'identifier des populations mutantes présentant des individus surexprimant la GFP, comme par exemple la lignée *ddc #16* (figure 25). Des analyses de transcription inverse suivie de PCR quantitative (RT-qPCR) ont ensuite été réalisées afin de s'assurer qu'en plus d'*ATCOPIA28::GFP*, plusieurs TEs dont les deux *ATCOPIA28* endogènes *AT3TE51900* et *AT5TE41315* étaient aussi surexprimés.

Il a alors fallu générer les populations de cartographie en croisant les individus mutants positifs avec la lignée parentale utilisée pour la mutagenèse (*ATCOPIA28::GFP ddc*). Puis, des approches de reséquençage du génome (NGS) couplé à des analyses de type « Bulk segregant analysis » ont permis de cartographier, pour chaque population mutante d'intérêt, la région génomique où se trouve le gène muté potentiellement responsable du phénotype de surexpression de la GFP. Des analyses de génétique traditionnelle ont ensuite permis de confirmer ces résultats : test de complémentation par croisement de l'allèle mutant EMS avec un allèle mutant T-DNA du gène candidat, utilisation de marqueurs de types « derived Cleaved Amplified Polymorphic Sequences » (dCAPS).

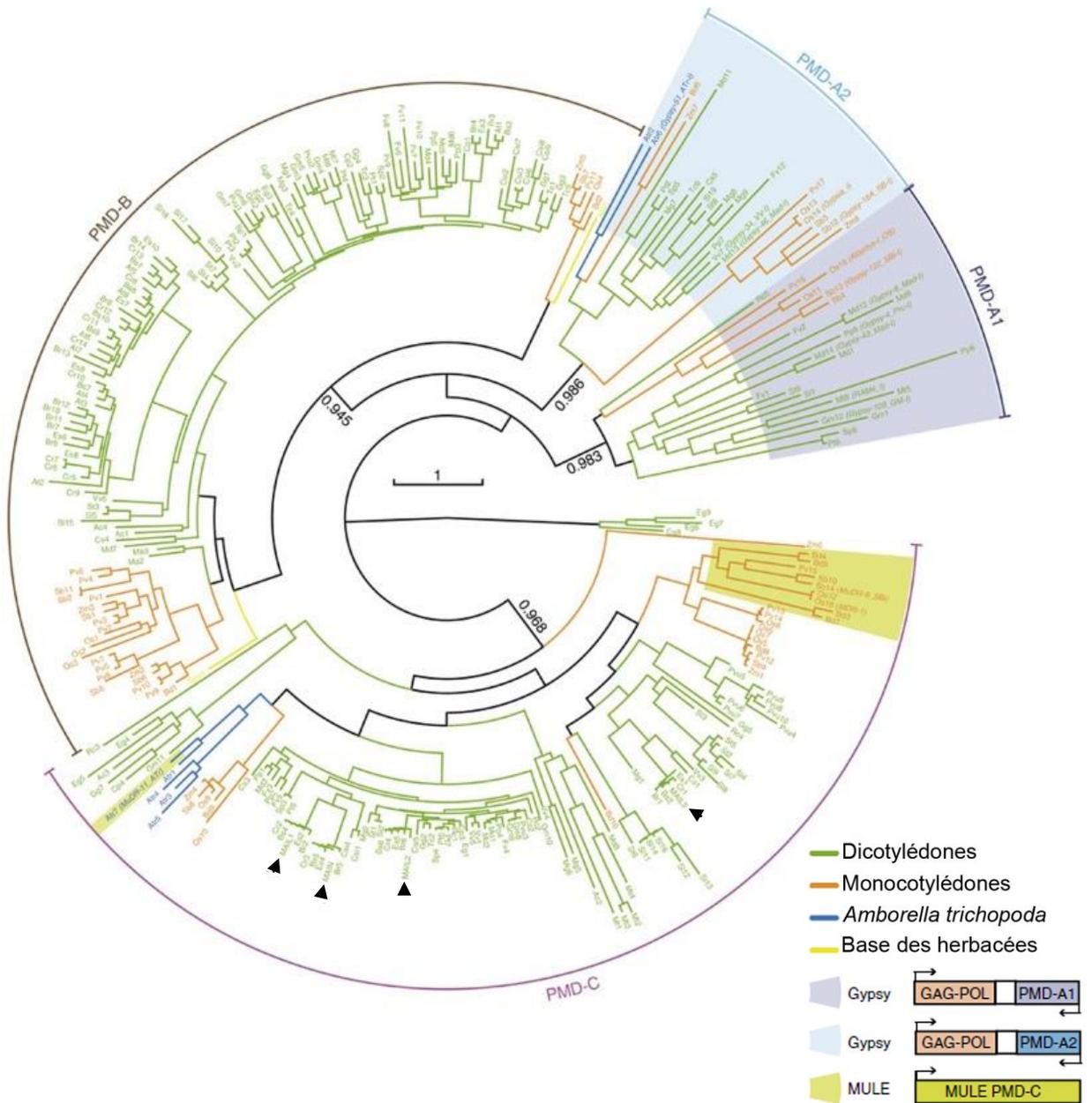


Figure 26 : Arbre phylogénique des PMD chez les angiospermes.

Un arbre phylogénique a été construit à partir de 308 PMD de 33 espèces d'angiospermes différentes. Les trois clades PMD-A, -B et -C sont indiqués et les flèches indiquent les protéines MAIN, MAIL1, MAIL2 et MAIL3 d'*A. thaliana*. Figure extraite de Ikeda *et al.*, 2017.

Suite à ces analyses, il a été démontré que la lignée *ddc #16* était mutée pour le gène *AT1G17930* codant la protéine Plant Mobile Domain (PMD) MAINTENANCE OF MERISTEM (MAIN), et que c'était bien cette mutation qui était responsable du phénotype mutant. Le mutant *ddc #16* a alors été renommé *ddc main-3*. C'est à ce moment-là que j'ai rejoint l'équipe MEAC pour commencer ma thèse. Les résultats découlant de l'identification de la population *ddc main-3* seront présentés dans le chapitre 1 de la partie RESULTATS (Nicolau *et al.*, 2020).

VI) La famille des gènes Plant Mobile Domain (PMD)

VI.1) Origine du domaine protéique

Le Plant Mobile Domain (PMD) est un domaine protéique uniquement présent chez les Angiospermes (Ikeda *et al.*, 2017). Il peut être de nature génique et très conservé entre espèces d'une même famille, ou associé à des TEs de type *MULE* ou *Ty3/Gypsy* dans de nombreux génomes de plantes (Babu *et al.*, 2006; Steinbauerová *et al.*, 2011; Ikeda *et al.*, 2017). Cette association TE-PMD n'est pas retrouvée chez les *Brassicaceae*.

Une analyse phylogénique précise a permis de séparer le domaine PMD en trois clades distincts largement distribués chez les Angiospermes (Ikeda *et al.*, 2017) (figure 26) : Le clade PMD-A, lui-même séparé en 2 sous-groupes PMD-A-1 et PMD-A-2, regroupe des PMD associés uniquement aux rétrotransposons *Ty3/Gypsy* (Ikeda *et al.*, 2017). Le clade PMD-B regroupe des PMD de type génique. Enfin, le clade PMD-C, auquel appartient le gène *MAIN* d'*A. thaliana*, regroupe des PMD soit de type génique, soit associés aux transposons *MULE*, ceci au moins chez plusieurs espèces de monocotylédones et chez *Amborella trichopoda* (Ikeda *et al.*, 2017). Il semblerait cependant que certaines espèces de *Solanaceae*, comme la tomate et la pomme de terre, possèdent également des TEs *MULE* portant un domaine PMD divergent.

Deux scénarios ont été proposés pour expliquer la distribution particulière des PMD-A, -B et -C au sein des angiospermes. Le premier scénario, privilégié, propose

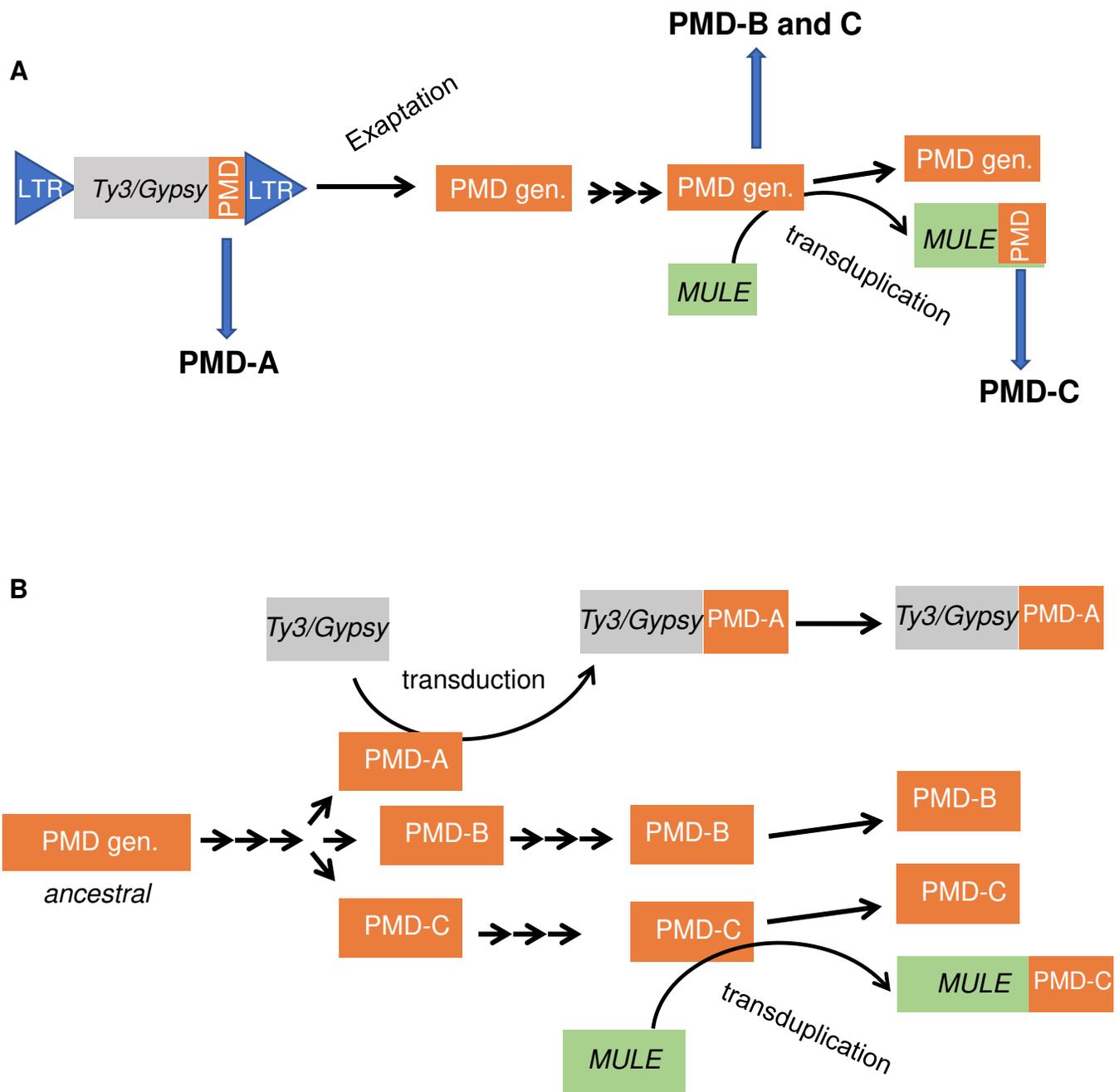


Figure 27 : Deux scénarios évolutifs proposés pour expliquer la répartition des différentes versions de PMD au sein des génomes d'angiospermes.

(A) Le premier scénario propose l'existence d'un PMD-A ancestral, associé à un TE de type *Ty3/Gypsy*, qui aurait été co-opté par l'ancêtre commun des angiospermes pour créer une version PMD générique (PMD gen.). Ce PMD générique aurait évolué et se serait diversifié, créant ainsi les versions PMD-B et PMD-C. Plus tard au cours de l'évolution des angiospermes, au moins deux événements indépendants de transduplication auraient permis à des TEs de type *MULE* de capturer une version PMD-C. (B) Le deuxième scénario propose plutôt l'existence d'un PMD générique ancestral qui aurait évolué pour donner les versions PMD-A, PMD-B et PMD-C. Par la suite, la version PMD-A aurait tout d'abord été capturée par transduction par un TE de type *Ty3/Gypsy*. Plus tard au cours de l'évolution des angiospermes, et de manière indépendante chez plusieurs espèces d'angiospermes, des versions PMD-C auraient été capturées par un TE de type *MULE* par transduplication.

qu'un PMD-A ancestral, associé à un rétrotransposon *Ty3/Gypsy*, aurait été co-opté par l'organisme hôte, ancêtre commun des angiospermes, par le processus de domestication de gène de TE. Au cours de l'évolution, des versions géniques PMD-B et -C seraient apparues, puis, des versions PMD-C auraient été capturées par des transposons MULEs lors, d'au moins, deux événements indépendants de transduplication (Ikeda *et al.*, 2017) (figure 27.A). Le deuxième scénario propose l'existence d'un PMD génique ancestral qui aurait évolué pour donner les versions PMD-A, -B et -C. Assez tôt au cours de l'évolution des angiospermes, la version PMD-A aurait été capturée par un rétrotransposon *Ty3/gypsy* via le processus de transduction. Plus tardivement, des versions PMD-C auraient été capturées par des transposons MULEs, ceci à plusieurs reprises et de manières indépendantes chez plusieurs espèces d'angiospermes, via le processus de transduplication (figure 27.B).

Du fait de son association avec des TEs, le domaine PMD peut être très représenté chez certains angiospermes (Babu *et al.*, 2006). Par exemple, il existe 427 copies de PMD chez la tomate (*Solanum lycopersicum*) et 668 chez le riz (*Oryza sativa*). Inversement, il existe 13 copies de PMD chez *A. thaliana*.

La structure 3D du PMD n'a pas encore été résolue, mais une structure secondaire composée de 8 feuillets β et plus de 10 hélices α a été prédite, et une similarité avec des protéines de type aminotransférases a été proposée (Babu *et al.*, 2006, Uhlken *et al.*, 2014, domaine PFAM = PF10536). En outre, une prédiction de structure protéique réalisée au sein de notre équipe avec l'outil I-TASSER (Iterative Threading ASSEmbly Refinement) (Zhang *et al.*, 2017) suggère, avec une très faible certitude, que le PMD aurait une structure similaire à des protéines de type « transférase ». En conclusion, la structure 3D du PMD reste énigmatique.

Les versions PMD géniques (non associées aux TEs) existent sous deux configurations : La majorité d'entre elles correspond à un PMD unique flanqué de régions non structurées, on parle alors de version « standalone ». Certaines versions sont associées à d'autres domaines protéiques comme par exemple un domaine phosphatase.

AT1G17930	MAIN		478
AT2G25010	MAIL1		509
AT2G04865	MAIL2		667
AT1G48120	MAIL3		1340

Figure 28 : La famille PMD-C chez *A. thaliana*.

Représentations schématiques des protéines de la famille PMD-C. A gauche sont indiqués les « locus identifier » et le nom des gènes. A droite est indiquée la taille en acide aminé de chaque protéine. Les représentations schématiques des protéines sont issues du site pfam, et les rectangles bleu et gris représentent des régions désorganisées de la protéine. Les flèches représentent la position du NLS dans la protéine.

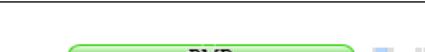
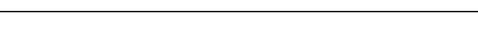
AT1G51538	PMD5		696
AT1G50830	PMD6		768
AT4G16050	PMD7		666
AT5G18510	PMD8		702
AT1G50820	PMD9		528
AT1G32120	PMD10		978
AT1G50790	PMD11		812
AT1G50770	PMD12		632
AT1G50750	PMD13		416

Figure 29 : La famille PMD-B chez *A. thaliana*.

Représentations schématiques des protéines de la famille PMD-B. A gauche sont indiqués les numéros ATG et le nom des gènes. A droite est indiqué la taille en acide aminé de chaque protéine. Les représentations schématiques des protéines sont issues du site pfam, et les rectangles bleu et gris représentent des régions désorganisées de la protéine.

VI.2) Les gènes PMD chez *A. thaliana*

Le génome d'*A. thaliana* code 13 protéines PMD, dont 9 PMD-B « standalone » et 4 PMD-C (Ikeda *et al.*, 2017). Parmi les 4 PMD-C, MAIN, MAIN-LIKE 1 (MAIL1) et MAIL2 sont des « standalone », alors que MAIL3 possède également un domaine de type sérine/thréonine phosphatase (PPP) (Ühlken *et al.*, 2014; Nicolau *et al.*, 2020). De plus, tous les membres de cette famille portent un signal de localisation nucléaire (NLS) dans leur séquence (Wenig *et al.*, 2013; Ühlken *et al.*, 2014) (figure 28). Parmi les protéines PMD-B, certaines d'entre elles présentent un domaine PMD partiel ou tronqué (figure 29). Alors que les protéines MAIN et MAIL1 ont été impliquées dans plusieurs processus cellulaires et développementaux, les rôles de MAIL2, MAIL3 ou des PMD-B restent à ce jour totalement inconnus.

VI.2.1) Les gènes MAIN et MAIL1

Des analyses utilisant les gènes rapporteurs GUS et GFP ont montré que les gènes *MAIN* et *MAIL1* sont fortement exprimés dans les tissus méristématiques apicaux caulinaire et racinaire, ainsi que dans les organes reproducteurs de la fleur (Wenig *et al.*, 2013; Ühlken *et al.*, 2014). Par des analyses de RTqPCR utilisant les parties aériennes de plantes âgées de 3 semaines (avant floraison), nous avons pu constater que *MAIN* et *MAIL1* étaient exprimés à des niveaux similaires (figure 30). *MAIN* et *MAIL1* sont les deux gènes les mieux caractérisés de la famille et plusieurs études ont montré leur rôle dans le développement, la stabilité du génome et la répression des TEs.

VI.2.1.1) Rôle dans le développement et la stabilité du génome.

MAIN a initialement été identifié dans un crible génétique basé sur un défaut de distribution sub-racinaire de la GFP, en utilisant le matériel *pSUC2::GFP* (Wenig *et al.*, 2013).

Les premières études fonctionnelles de *MAIN* et *MAIL1* ont montré leur importance dans la croissance de la plante et la stabilité du génome (Wenig *et al.*, 2013; Ühlken *et al.*, 2014). Les plantes mutantes KO *main-2* et *mail1-1* montrent un développement fortement affecté : elles sont plus petites que les plantes sauvages, la croissance de leur racine primaire est drastiquement réduite et les feuilles sont

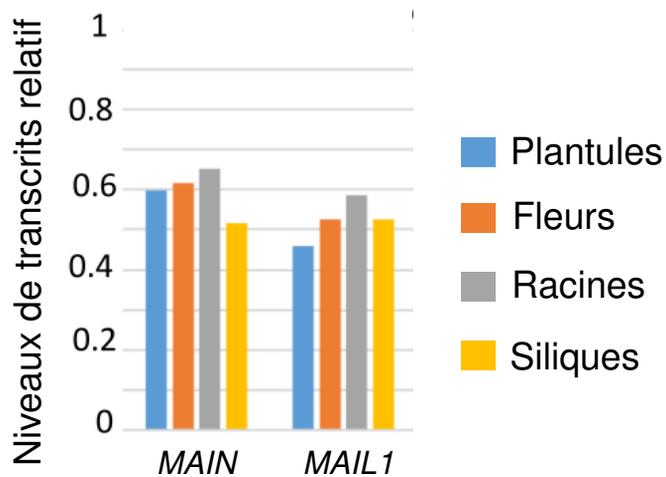


Figure 30 : Niveaux d'expression de *MAIN* et *MAIL1* dans les différents organes chez *A. thaliana*.

Le niveau d'expression de *MAIN* et *MAIL1* sont testés par RT-qPCR. Les niveaux de transcrits sont exprimés relativement au gène de ménage *RHIP1*.

déformées (Wenig *et al.*, 2013; Ühlken *et al.*, 2014) (figure 31). Ce phénotype développemental s'explique probablement par un méristème apical racinaire (MAR) altéré avec une diminution de la division cellulaire et une différenciation cellulaire précoce, ainsi qu'une perte de la totipotence des cellules du centre quiescent du MAR. Les zones d'expression des gènes de « patterning » de la racine sont altérées, avec certains gènes exprimés de façon ectopique et d'autres dont la zone d'expression est diminuée. Dans les plantes *main-2*, le méristème apical caulinaire (MAC) est également désorganisé avec une division aberrante des cellules (Wenig *et al.*, 2013). Dans les plantes *mail1-1*, l'expression de gènes embryonnaires reste active longtemps au cours du développement provoquant parfois la formation de cals (Ühlken *et al.*, 2014). En plus de ces dérèglements, les plantes *main-2* et *mail1-1* montrent aussi une mort cellulaire accrue dans le MAR qui s'accompagne d'une augmentation de l'expression des gènes de réponse aux dommages à l'ADN, et une sensibilité aux traitements génotoxiques plus élevée que des plantes sauvages. Tous ces phénotypes indiquent un rôle de *MAIN* et *MAIL1* dans le maintien de la stabilité du génome, en particulier dans les méristèmes apicaux. Ces études montrent aussi qu'une épistasie existe entre les deux gènes (Wenig *et al.*, 2013; Ühlken *et al.*, 2014).

VI.2.1.2) Rôle de *MAIN* et *MAIL1* dans la répression des TEs

Une étude plus récente a montré l'implication de *MAIN* et *MAIL1* dans la répression des TEs, principalement péracentromériques (Ikeda *et al.*, 2017). Il a été suggéré que *MAIN* et *MAIL1* agissent dans la même voie de répression, et indépendamment des différentes voies déjà connues (méthylation de l'ADN, modifications d'histone, siARN, *MOM1* et *AtMORC6*). Enfin, ces protéines joueraient un rôle dans la condensation de l'hétérochromatine (Ikeda *et al.*, 2017). Tous ces résultats seront plus amplement discutés plus tard dans ce manuscrit.

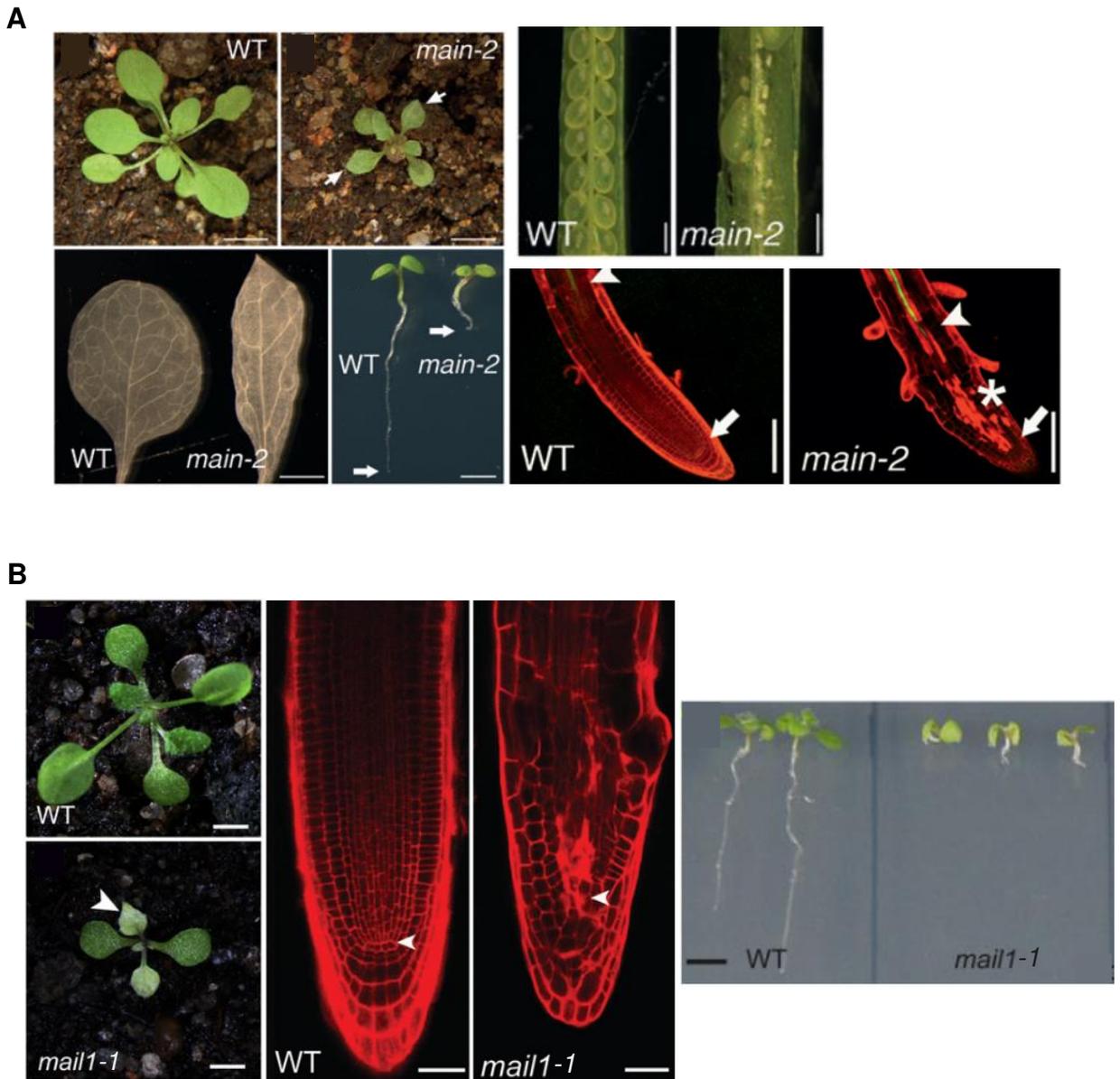


Figure 31 : Phénotypes développementaux des mutants *main-2* et *mail1-1*

(A) Photos représentatives de plantes sauvages (WT) et *main-2* montrant les différents phénotypes développementaux des mutants *main-2*. Les plantes *main-2* ont une taille de la rosette et des feuilles diminuée, ainsi qu'une déformation des feuilles. Elles ont également un nombre de graines par silique diminué et un défaut de croissance de la racine. Des observations en microscopie de la racine montrent une différenciation précoce des cellules (indiqué par les flèches) et une coloration à l'iodure de propidium (IP) montre de la mort cellulaire dans l'extrémité de la racine (indiqué par l'astérisque). (B) Photos représentatives de plantes WT et *mail1-1*. Les plantes *mail1-1* montrent également une diminution de la taille des feuilles et de la rosette, ainsi qu'une diminution de la croissance de la racine. Des observations en microscopie de l'extrémité de la racine montrent une désorganisation du MAR avec de la mort cellulaire montrée par coloration à l'IP. Le centre quiescent est indiqué par la flèche blanche (extrait de Wenig *et al.*, 2013 et Ühlken *et al.*, 2014).

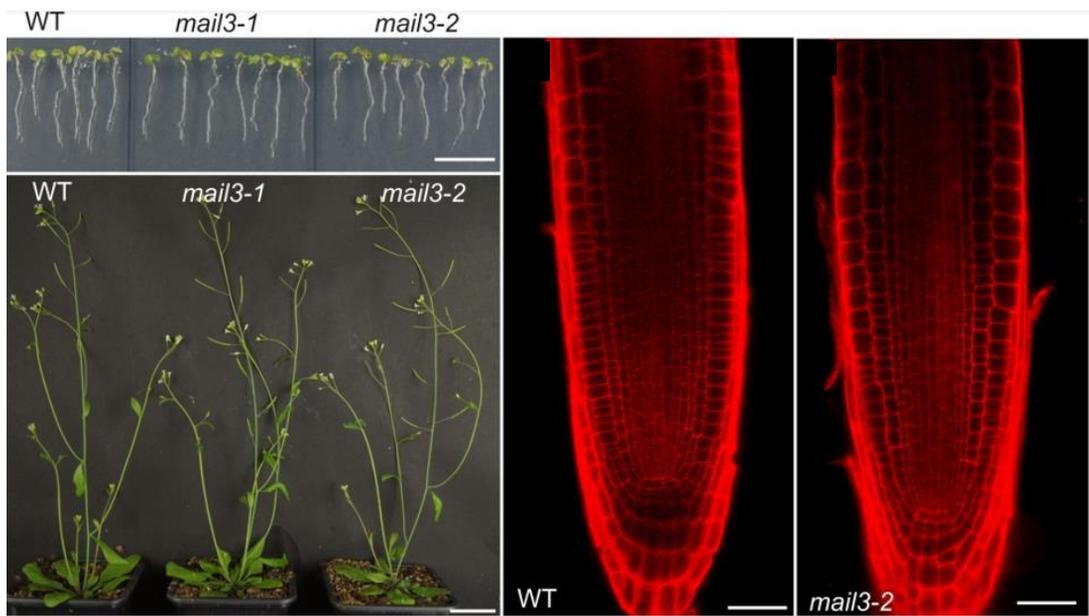


Figure 32 : Phénotypes des plantes *mail3-1* et *mail3-2*.

Photos représentatives de plantes sauvages WT, *mail3-1* et *mail3-2*. Ces plantes ne montrent aucun phénotype développemental. Des observations en microscopie de l'extrémité de la racine ne montrent pas de différence entre les plantes WT et *mail3-2* (extrait de Ühlken *et al.*, 2014).

Superclasse	Type de domaine	Exemples de FTs
1	Domaines basiques	bZIP, facteurs hélice-boucle-hélice (bHLH)
2	DBD dépendents du zinc (à doigt de zinc)	Facteurs à doigt de zinc (znf) cys4 znf, Cys2His2 znf
3	Hélice-tour-hélice	FT à homéodomaine, facteurs "heat shock"
4	Facteurs "beta-Scaffold Factors with Minor Groove Contacts"	p53, groupe High mobility (HMG), protéines "TATA-binding"
0	Autres FTs	Protéines "copper fist", domaines "pocket"

Figure 33 : Les superclasses de FTs chez les eucaryotes.

Les FTs sont répartis en plusieurs superclasses selon la structure de leur domaine de liaison à l'ADN. Chaque superclasse regroupe plusieurs familles de FTs.

VI.2.2) Les gènes *MAIL2* et *MAIL3*

Les gènes *MAIL2* et *MAIL3* montrent un patron d'expression similaire à celui de *MAIL1* et *MAIN* (Ühlken *et al.*, 2014).

MAIL2 n'a jamais été étudié. *MAIL3*, quant à lui, a été étudié à l'aide de deux allèles mutants *mail3-1* et *mail3-2*. Les plantes mutantes *mail3-1* et *mail3-2* ne montrent aucun phénotype développemental (figure 32), et les plantes *mail3-2* ne semblent pas montrer de surexpression des TEs (Ühlken *et al.*, 2014 ; Ikeda *et al.*, 2017). *MAIL3* ne semble donc pas impliqué dans le développement de la plante ni la répression des TEs dans les conditions de culture contrôlées des laboratoires (Ühlken *et al.*, 2014; Ikeda *et al.*, 2017; de Luxán-Hernández *et al.*, 2019).

Les rôles physiologiques et moléculaires de *MAIL2* et *MAIL3* sont encore inconnus et des analyses approfondies seront nécessaires pour comprendre la fonction de ces gènes.

VII) Les facteurs de transcription (FTs)

Une régulation précise et coordonnée de la transcription est essentielle pour le bon fonctionnement des cellules. De nombreux mécanismes cellulaires sont impliqués dans cette régulation (Lelli *et al.*, 2012). Les facteurs de transcription (FTs) sont des protéines spécialisées qui régulent la transcription. Ces facteurs peuvent agir comme activateurs, via un domaine d'activation transcriptionnel, ou répresseurs, via un domaine de répression ou en bloquant les facteurs activateurs (Latchman, 1997). Les FTs portent également un domaine de liaison à l'ADN (« DNA binding domain » DBD) qui leur permet de reconnaître des séquences d'ADN spécifiques, appelée éléments cis-régulateurs (Latchman, 1997 ; Istrail and Davidson, 2005). Les FTs sont séparés en cinq classes selon la nature de leur DBD (Matys *et al.*, 2006) (figure 33) :

VII.1) Les différents types de FTs

Les FTs peuvent être séparés en trois catégories selon leur mode de fonctionnement et leur position au sein des processus de régulation et d'initiation de la transcription.

VII.1.1) Les FTs pionniers

Les FTs pionniers sont présents chez les animaux et les plantes. Ils ont la capacité de se lier à l'hétérochromatine afin de promouvoir sa décondensation et permettre à d'autres FTs de se lier sur ces régions, conduisant potentiellement à l'expression des gènes ciblés (Lai et al., 2018). Aussi bien chez les animaux que chez les plantes, ces facteurs sont essentiels pour le développement et la reprogrammation cellulaire. Par exemple chez les animaux, les FTs Forkhead Box A1 (FOXA1) et GATA binding protein 4 (GATA4) sont capable d'induire l'organogenèse à partir des cellules endodermes (Lai et al., 2018). Chez les plantes, LEAFY COTYLEDON 1 (LEC1) est un FT impliqué dans le développement embryonnaire, LEAFY (LFY) dans la mise en place du méristème floral, enfin APETALA 1 (AP1) et SEPALLATA 3 (SEP3) sont impliqués dans le développement de la fleur (Tao et al., 2017 ; Moyroud et al., 2010 ; Pajoro et al., 2014).

Plusieurs modes d'action des FTs pionniers ont été observés (Lai et al., 2018) :

- Des FTs portent un DBD « histone-like » leur permettant de déplacer les histones et de rendre la chromatine plus accessible pour d'autres FTs. C'est le cas des FTs FOXA1 et LEAFY COTYLEDON1 (LEC1) qui portent des DBDs similaire à ceux des histones H1 et H2A-H2B respectivement (Iwafuchi-Doi et al., 2016 ; Lee et al., 2003).
- Des FTs ont la capacité de se lier sur l'ADN nucléosomal sans avoir besoin de déplacer les histones, c'est le cas des FTs OCTAMER-BINDING TRANSCRIPTION FACTOR 4 (OCT4) et LFY (Soufi et al., 2012 ; Sayou et al., 2016).
- Des FTs sont capables de recruter des remodeleurs de la chromatine permettant l'ouverture de la chromatine. C'est le cas de LFY et SEP3 chez les plantes

qui recrutent les remodeleurs ISWI SPLAYED (SYD) et BRAHMA (BRM) (Wu et al., 2012)

- Enfin, des FTs sont capables de se lier sur l'ADN méthylé comme PAIRED BOX 7 (PAX7) chez les mammifères (Mayran et al., 2018).

Les FTs pionniers permettent la mise en place d'une structure ouverte de la chromatine pour les autres facteurs de transcription, et favorisent le dépôt de marques permissives et le retrait de marques répressives telles que H3K27me3 (Johnson et al., 2018 ; Tao et al., 2017).

VII.1.2) Les FTs spécifiques

Les FTs spécifiques sont des activateurs ou des répresseurs transcriptionnels. Ils agissent en se liant à des séquences d'ADN spécifiques situées en amont du TSS, et sont le plus souvent activés par des voies de signalisation spécifiques ou des signaux environnementaux (Lelli et al., 2012). Ces FTs sont séparés en différentes familles selon leur mode de fonctionnement et les séquences qu'ils lient. Chez *A. thaliana*, il existe plusieurs grandes familles de FTs. Dans cette partie, je me concentrerai sur trois grandes familles de FTs spéciaux.

VII.1.2.1) La famille WRKY

La famille de FTs WRKY est caractérisée par la présence d'un domaine WRKY, un DBD à doigt de zinc (Babu et al., 2006 ; Rushton et al., 2010). Ces FTs sont capables de reconnaître une séquence spécifique, appelée boîte W, et de s'y lier. Les FTs WRKY sont capables d'activer ou réprimer la transcription, et certains portent même les deux fonctions. C'est le cas de WRKY53 qui est capable d'activer ou de réprimer la transcription selon le contexte du promoteur sur lequel il se lie (Miao et al., 2008). Une des particularités des FTs WRKY est leur capacité d'autorégulation ou de régulation croisée, en effet le promoteur des gènes *WRKY* porte également une boîte W, ce qui permet la liaison des FTs WRKY sur leur propre promoteur ou le promoteur d'autres WRKY (Eulgem and Somssich, 2007 ; Eulgem et al., 1999 ; Robatzek and

Somssich, 2002). Les protéines WRKY sont impliquées dans de nombreux mécanismes cellulaires comme la réponse aux stress biotiques et abiotiques, le développement de la graine ou encore la senescence (Grunewald, et al., 2008 ; Jiang and Deyholos, 2009 ; Li et al., 2009 ; Luo et al., 2005 ; Guo et al., 2004 ; Robatzek and Somssich, 2002).

VII.1.2.2) La famille basic (région) leucine zipper (bZIP)

Les FTs bZIP existent chez tous les eucaryotes. Ces FTs possèdent un DBD basique conservé et un domaine de glissière à leucine qui permet leur hétérodimérisation, ce qui leur confère une grande flexibilité d'interaction et de liaison à leurs ADN cibles (Dröge-Laser et al., 2018 ; Rodriguez-Martinez et al., 2017). Chez *A. thaliana*, la famille bZIP compte 78 membres, répartis en 13 groupes (Jakoby et al., 2002). Ces FTs sont impliqués dans de nombreux processus comme la réponse aux stress biotiques et abiotiques, le métabolisme énergétique, la floraison ou la senescence (Alves et al., 2013 ; Fujita et al., 2005 ; Baena-González et al., 2007 ; Abe et al., 2005 ; Smykowski et al., 2010).

VII.1.2.1) La famille FAR1-RELATED SEQUENCE (FRS)

À l'instar de son membre fondateur FAR-RED IMPAIRED RESPONSE 1 (FAR1), les FTs de cette famille dérivent de transposase de TEs de type MULE (Lin et al., 2007). La majorité des membres de cette famille portent trois domaines distincts : un DBD FAR1 de type doigt de zinc, un domaine transposase putatif dérivé de MULE et un domaine SWIM (SWI2/SNF2 & MuDR transposase) de type zinc finger (Lin and Wang, 2004). Les membres les mieux décrits de cette famille sont FAR1 et FAR-RED ELONGATION HYPOCOTYLS 3 (FHY3) qui sont impliqués dans de nombreux processus cellulaires comme la détection de la lumière, le développement du méristème apical et floral et l'immunité de la plante (Ma and Li, 2018). En plus de ces deux protéines, cette famille comporte douze protéines FAR1-RELATED SEQUENCE (FRS) et quatre protéines FRS-RELATED FACTOR (FRF) chez *A. thaliana* (Lin and

Wang, 2004 ; Aguilar-Martinez et al., 2015). Contrairement à FAR1 et FHY3, les fonctions physiologiques et moléculaire des protéines FRS et FRF sont largement méconnues. Il a néanmoins été montré que FRS6 et FRS8 jouent un rôle dans la régulation du timing de floraison, et que FRF1 est capable de se lier sur le promoteur de *SHOOT MERISTEMLESS (STM)* un gène requis pour la maintenance du MAC (Aguilar-Martínez et al., 2015).

VII.1.3) Les FTs généraux

Les FTs généraux sont impliqués dans le recrutement et la formation du complexe de pré-initiation de l'ARN polymérase II (ARN pol II). Il existe 6 FTs généraux, appelés « transcription factor II » A (TFIIA), TFIIB, TFIID, TFIIE, TFIIF, TFIIG et TFIIH. Ils agissent généralement via la reconnaissance de séquences d'ADN régulatrices spécifiques telle que la boîte TATA, présente dans le promoteur des gènes transcrits par l'ARN pol II et proche du TSS (Orphanides et al., 1996). La reconnaissance de la boîte TATA est assurée, en autres, par TFIID, un complexe protéique composé de la « TATA box-binding protein » (TBP) et de plusieurs « TBP-associated factors » (TAF) (Sainsbury et al., 2015). TBP est requis pour l'initiation de la transcription à partir des promoteurs contenant une boîte TATA, tandis que les TAF apportent une spécificité de séquence (Sainsbury et al., 2015). Les FTs généraux agissent sur la chromatine en aval des FTs spécifiques, grâce à des ponts moléculaires impliquant de nombreuses protéines telles que les facteurs du complexe Mediator ou encore le complexe SAGA impliqué dans l'acétylation des histones et le recrutement de l'ARN Pol II (Sainsbury et al., 2015).

VII.2) Interaction entre les facteurs de transcription et les modifications épigénétiques

Les modifications épigénétiques, par leur capacité à modifier l'accessibilité de l'ADN, peuvent influencer la liaison des FTs. Ainsi, les marques répressives comme la méthylation de l'ADN, H3K9me2 ou H3K27me3 peuvent représenter un obstacle pour les FTs, mais certaines d'entre elles peuvent, dans certaines conditions, agir comme

plateforme de recrutement (Zhu et al., 2016 ; Lai et al. 2018). C'est le cas par exemple lors du recrutement du FT pionnier OCT4 au niveau de ses sites de liaison à l'ADN (Yin et al., 2017). Les marques épigénétiques peuvent également influencer la reconnaissance de certaines séquences par des FTs (Zhu et al., 2016). La fixation des FTs sur l'ADN va également moduler la dynamique des marques épigénétiques afin de permettre la transcription. Ainsi certains FTs sont capables de recruter des remodeleurs de la chromatine, provoquant le déplacement des histones et l'ouverture de la chromatine (Iwafuchi-Doi et al., 2016), tandis que d'autres FTs recrutent des protéines modifiant les marques épigénétiques. Par exemple, les FTs bZIP28 et bZIP60 sont capable d'interagir avec le complexe protéique AtCOMPASS catalysant le dépôt des marques H3K4me2/3 (Song et al., 2015), ce qui permet la transcription des gènes ciblés par ces FTs. De plus, il a été montré que l'acétylation des histones permet le recrutement et l'activation de la RNA pol II (Stasevich et al., 2014), et plusieurs HAT jouent un rôle dans l'activation de la transcription (Bauerand Martin, 2017 ; Stasevich et al., 2014).

VIII) Les protéines phosphatases

La phosphorylation des protéines est un processus très dynamique qui module l'activité, la conformation, la localisation ou encore la stabilité des protéines cibles (Hunter, 1995). La phosphorylation est assurée par les kinases et la déphosphorylation par les phosphatases.

Les phosphatases sont classées en quatre familles basées sur leur séquence et leur mécanisme catalytique (Moorhead *et al.*, 2007; Shi, 2009; Heroes *et al.*, 2013). Deux familles catalysent spécifiquement la déphosphorylation des résidus sérine et thréonine : la famille des « phosphoprotein phosphatases » (PPP) et la famille des « Mg²⁺- or Mn²⁺-dependent protein phosphatases » (PPM). Les deux autres familles catalysent la déphosphorylation des résidus tyrosine mais également sérine et thréonine : la famille des phosphotyrosine phosphatase (PTP) et la famille des «

aspartate (Asp)-dependent enzymes ». Dans cette partie, je m'intéresserai uniquement à la famille des PPP, et plus particulièrement chez les plantes.

Chez les eucaryotes, la famille des PPP est composée de sept sous-familles : Ser/Thr protein phosphatase type one (PP1), PP2 (ou PP2A), PP3 (ou PP2B, absente chez les plantes), PP4, PP5, PP6 et PP7 (spécifique des plantes) (Farkas *et al.*, 2007). Ces protéines partagent des séquences et structures proches, ainsi qu'un mécanisme catalytique identique (Moorhead *et al.*, 2007; Shi, 2009; Heroes *et al.*, 2013). Elles interagissent avec d'autres protéines, sous-unités régulatrices, qui leur confèrent leur spécificité de substrat. Par exemple, plusieurs centaines de sous-unités régulatrices de PP1 ont été identifiées chez l'homme (Moorhead *et al.*, 2007; Heroes *et al.*, 2013). De même chez les plantes, plusieurs sous-unités régulatrices ont été identifiées (Takemiya *et al.*, 2009, 2013; Templeton *et al.*, 2011).

VIII.1) Les sous-familles PP1 à PP6

PP1 est une protéine conservée chez les eucaryotes. Chez *A. thaliana*, neuf gènes codent une PPP de type PP1 (Kerk *et al.*, 2008). Au niveau cellulaire, les protéines PP1 sont localisées dans le noyau et le cytosol, chez l'homme comme chez *A. thaliana* (Takemiya *et al.*, 2009; Templeton *et al.*, 2011). L'activité de PP1 est modulée par ses sous-unités régulatrices qui recrutent les cibles de PP1 ou contrôlent l'accessibilité de son domaine catalytique (Peti *et al.*, 2013). PP1 est impliquée dans de nombreux processus cellulaires, comme la transcription, la croissance et la différenciation cellulaire (Rebelo *et al.* 2015). Les protéines PP1 jouent également un rôle dans la défense contre les pathogènes chez *A. thaliana* (Liu *et al.*, 2019).

Les sous-familles PP2A, PP4 et PP6 sont les plus proches phylogénétiquement (Uhrig *et al.*, 2013). PP2A, PP4 et PP6 sont toutes capables d'interagir avec la protéine TAP46, un homologue de la protéine Tap42 de la levure, impliquée dans la croissance cellulaire, l'autophagie et la synthèse protéique (Ahn *et al.*, 2011). Les enzymes PP2A sont composées de trois sous-unités : une sous-unité catalytique (PP2Ac), une sous-unité régulatrice (PP2Ab) et une sous-unité structurelle (PP2Aa). PP2A est impliquée dans plusieurs processus comme la voie de signalisation de l'éthylène, la régulation de la croissance et le développement (Zhou *et al.*, 2004; Blakeslee *et al.*, 2008; Skottke

et al., 2011; Charpentier *et al.*, 2014). Chez les animaux, PP4 est impliquée dans la régulation des interactions entre les microtubules et le centrosome (Brewis *et al.*, 1993). Chez *A. thaliana*, PP4 est codée par deux gènes, qui sont exprimés faiblement dans plusieurs organes de la plante. Récemment, il a été montré que PP4 est impliquée dans la biogenèse des miARN, en recrutant Pol II au niveau des gènes MIR (Wang *et al.*, 2019). PP6 est aussi codée par deux gènes chez *A. thaliana* (Farkas *et al.*, 2007). Des études ont montré que les protéines PP6 sont impliquées dans plusieurs voies de signalisation, notamment les voies hormonales de l'auxine et de l'acide abscissique en régulant la phosphorylation des protéines PIN-FORMED (PIN) (Dai *et al.*, 2012). Les PP6 jouent également un rôle dans le contrôle de la photomorphogénèse de la plante (Kim *et al.*, 2002; Haga and Sakai, 2018; Yu *et al.*, 2019).

Chez les mammifères, PP5 fait partie des complexes HEAT SHOCK PROTEIN 90 (HSP90) et des cryptochromes (CRY1 et CRY2), et elle est impliquée dans la régulation des canaux à ions, la voie de signalisation des hormones, le cycle cellulaire et l'apoptose (Chinkers, 2001 ; Zeke *et al.*, 2005). Chez les plantes, PP5 est impliquée dans la résistance aux pathogènes, via une interaction avec HSP90 et AGO1 (Iki *et al.*, 2012 ; de la Fuente van Bentem *et al.*, 2005), la thermotolérance grâce à une activité chaperonne de son domaine tetratricopeptide repeat (TPR) (Kharchenko *et al.*, 2011) et la perception de la lumière (Ryu *et al.*, 2005).

VIII.2) La sous-famille PP7

Chez *A. thaliana*, cette sous-famille est composée de trois membres : PP7, PP7-like (PP7L, désignée dans un premier temps comme inactive PP7) et MAIL3 (ou long-PP7) (Farkas *et al.*, 2007).

PP7 est la protéine la plus étudiée de cette sous-famille. Elle est impliquée dans de nombreux mécanismes cellulaires, en interagissant avec divers facteurs. Il a été montré que la mutation *pp7* diminue la thermotolérance chez *A. thaliana* et que son expression est induite après un stress thermique à 37°C (Liu *et al.*, 2007). PP7 est aussi impliquée dans la voie de la perception de la lumière et l'ouverture des stomates en interagissant avec les protéines NUCLEOTIDEDIPHOSPHATE KINASE 2 (NDPK2)

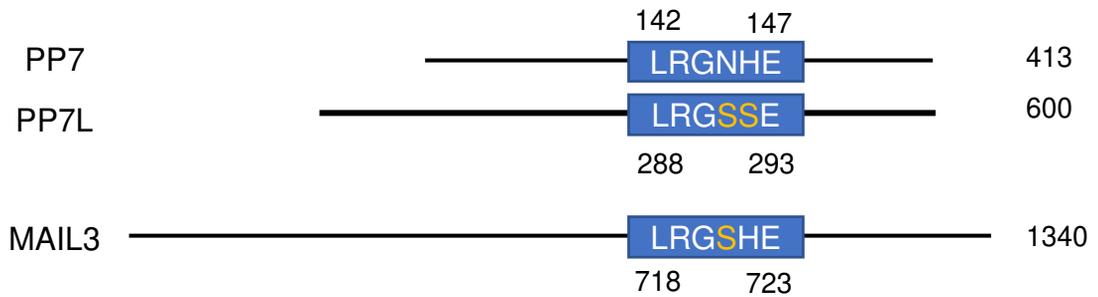


Figure 34 : Le domaine phosphatase de PP7L est muté sur son site catalytique.

Comparaison du site catalytique de PP7, PP7L et MAIL3 (comme défini dans de Luxán-Hernández *et al.*, 2019). Les positions du premier et du dernier acide aminé du site sont indiquées au-dessus ou au dessous de la séquence protéique. Les acides aminés mutés dans PP7L et MAIL3 sont indiqués en jaune. La taille des protéines est donnée à droite des alignements.

et HYPERSENSITIVE TO RED AND BLUE 1 (HRB1) respectivement (Genoud *et al.*, 2008; Sun *et al.*, 2012).

PP7L est considérée comme une phosphatase non fonctionnelle car plusieurs acides aminés essentiels sont mutés au sein de son domaine catalytique (Goldberg *et al.*, 1995; Farkas *et al.*, 2007) (figure 34). Récemment, il a été montré que PP7L est localisée dans le noyau et le cytosol (de Luxán-Hernández *et al.*, 2019; Xu *et al.*, 2019). PP7L est importante pour l'expression de protéines chloroplastiques, un mutant *pp7l* est donc impacté dans le développement de ses chloroplastes. Il présente des défauts de tolérance aux stress salin et lumineux (Xu *et al.*, 2019).

Au cours de ma thèse, j'ai pu montrer l'existence d'une interaction entre PP7L et les protéines MAIN et MAIL1. J'ai aussi montré l'implication de PP7L dans la répression des TEs (Nicolau *et al.*, 2020). Ces résultats seront décrits dans le chapitre 1 de la partie RESULTATS de ce manuscrit. En outre, des résultats similaires ont été décrits dans une autre étude récente (de Luxán-Hernández *et al.*, 2019). Les mutants *pp7l* montrent sensiblement le même phénotype développemental que des mutants *main* ou *mail1*, avec un développement racinaire et foliaire fortement affecté et un retard de floraison (Nicolau *et al.*, 2020, de Luxán-Hernández *et al.*, 2019).

VIII.3) Rôle des phosphatases dans la chromatine

Les processus dynamiques de phosphorylation/déphosphorylation des histones, des différents facteurs épigénétiques ou des FTs jouent un rôle central dans la régulation des mécanismes moléculaires associés à la chromatine (Gil and Vagnarelli, 2019).

VIII.3.1) Condensation de la chromatine au cours de la mitose

Lors de la mitose, les complexes condensines participent à la condensation de la chromatine en chromosome. La protéine PP2 joue deux rôles antagonistes dans la régulation des complexes condensine : d'une part elle permet le recrutement de ces complexes sur la chromatine indépendamment de son activité phosphatase

(Takemoto et al., 2009), d'autre part PP2 peut être recrutée sur les promoteurs des gènes par la protéine TBP et déphosphoryler les complexes condensine, ce qui les désactive (Xing et al., 2008). La condensation de la chromatine est également assurée par une voie parallèle aux condensines dans laquelle intervient le facteur REGULATOR OF CHROMOSOME ARCHITECTURE (RCA). Ce facteur est rendu inactif à la fin de l'anaphase par PP1 qui le déphosphoryle et permet ainsi la décondensation de la chromatine (Vagnarelli et al., 2006 ; Samejima et al., 2018). Au cours de la mitose, l'histone H3 est phosphorylée sur les sérines 10 et 28 et notamment sur la thréonine 3 (H3T3Ph). Dans la phase précoce de l'anaphase, PP1 est recrutée sur les chromosomes où elle va déphosphoryler H3T3Ph, ce qui permet la ségrégation correcte des chromosomes (Gil and Vagnarelli, 2019).

VIII.3.2) La réparation de l'ADN

En réponse aux cassures de l'ADN, les histones H2A.X sont phosphorylées par deux kinases ATM et ATR, qui vont également phosphoryler d'autres facteurs activant la voie de réponse aux dommages à l'ADN (DDR). PP2 interagit avec ATM et l'inhibe (Goodarzi *et al.*, 2004). Lors d'une irradiation, l'interaction en PP2 et ATM est perdue et ATM peut initier le DDR (Goodarzi *et al.*, 2004). Il a également été montré que PP5 joue un rôle dans l'activation de ATM (Ali *et al.*, 2004). Une fois la cassure de l'ADN réparée, les histones H2A.X sont déphosphorylées par plusieurs phosphatases comme PP2, PP6 et PP4 chez les mammifères (Gil and Vagnarelli, 2019).

VIII.3.3) La réplication de l'ADN

Peu de choses sont connues sur le rôle des phosphatases dans la réplication de l'ADN. Il a été montré chez *S. cerevisiae* que PP2 est importante pour l'initiation de la réplication de l'ADN en permettant la formation du complexe de pré-réplication (Gil and Vagnarelli, 2019). PP1 joue également un rôle dans la réplication de l'ADN en stabilisant l'ORIGIN REPLICATION COMPLEX (ORC) lors de la phase G1 et en l'inhibant lors de la phase S (Gil and Vagnarelli, 2019).

VIII.3.4) La régulation de la transcription

Les phosphatases peuvent réguler la transcription à plusieurs niveaux en régulant le niveau de phosphorylation de l'ARN pol II, des FTs ou des histones.

L'ARN pol II porte un domaine C-terminal (CTD) qui peut être différentiellement phosphorylé au cours des différentes étapes de la transcription. Selon son état de phosphorylation, le CTD sert de plateforme pour recruter des facteurs impliqués dans la mise en place de la coiffe, l'élongation de la transcription et l'épissage des ARN pré-messagers (Gil and Vagnarelli, 2019). La déphosphorylation du CTD est nécessaire pour terminer la transcription et permettre le détachement de l'ARN pol II de l'ADN (Gil and Vagnarelli, 2019).

Les FTs sont régulés par de nombreuses MPTs, dont la phosphorylation. Un exemple est la phosphorylation des FTs au cours de la mitose, qui les exclut de la chromatine et permettant la mise en pause de la transcription (Delcuve et al., 2008). La réactivation de la transcription après la mitose est assurée par la déphosphorylation des FTs par des phosphatases qui n'ont pas encore été bien décrites (Palozola et al., 2017). Chez *A. thaliana*, PP6 déphosphoryle les FTs PHYTOCHROME INTERACTING FACTORS (PIFs), provoquant leur dégradation et ainsi régulant la photomorphogénèse (Yu et al., 2019).

Enfin, la phosphorylation des histones joue également un rôle dans la régulation de la transcription. Par exemple, la phosphorylation de H3S10 est liée à la transcription de gènes de croissance (Chadee et al., 1999). Cette MPT est retirée par les phosphatases PLECKSTRIN HOMOLOGY DOMAIN LEUCINE-RICH REPEAT PROTEIN PHOSPHATASE (PHLPP) et MAP KINASE PHOSPHATASE 1 (MKP-1) (Kinney et al., 2009 ; Reyes et al., 2014). La phosphorylation de H3T11 et H3T6 régule aussi la transcription en permettant notamment l'éviction de H3K9me des gènes qui seront transcrits (Metzger et al., 2008, 2010). Lors de dommages à l'ADN, H3T6 est déphosphorylée, ce qui est corrélé à une diminution de l'expression des gènes du cycle cellulaire (Shimada et al., 2008).

IX) Problématique et objectif de thèse

Mon projet de thèse s'inscrit dans la continuité de l'identification de la lignée *ddc #16 (ddc main-3)* par génétique directe (partie VI de cette introduction). Lors du commencement de ma thèse, un article décrivant le rôle des protéines MAIN et MAIL1 en tant que nouveaux facteurs réprimant les TEs venait d'être publié (Ikeda *et al.*, 2017). Cette étude suggérait que MAIN et MAIL1 agiraient en grande partie dans une même voie génétique qui est indépendante de la méthylation de l'ADN, des siARN, et des voies MORC6 et MOM1. Elle a également montré que la condensation de l'hétérochromatine semblerait être affectée dans les mutants *main* et *mail1*.

Deux études plus anciennes avaient également mis en évidence la localisation nucléaire de ces deux protéines PMD, ainsi que leur rôle essentiel pour le bon développement de la plante (Wenig *et al.*, 2013; Ühlken *et al.*, 2014).

En m'appuyant sur toutes ces données bibliographiques, mon projet de thèse a eu pour objectifs :

- 1) De définir l'interactome des protéines MAIN et MAIL1 et plus particulièrement de savoir si ces deux protéines interagissent physiquement ensemble. Pour cela, des expériences d'immunoprécipitation suivie de spectrométrie de masse (IP-MS) et de co-immunoprécipitation (co-IP) ont été réalisées. Les résultats ont été publiés dans Nicolau *et al.*, 2020 (chapitre 1 de la partie RESULTATS).
- 2) De mieux décrire l'impact global de MAIN et MAIL1 sur la régulation de l'expression des gènes et la répression des TEs. Pour cela, des approches de séquençage haut débit d'ARN (RNA-seq) ont été effectuées et les résultats sont également publiés dans Nicolau *et al.*, 2020 (chapitres 1 et 2 de la partie RESULTATS).
- 3) De déterminer le mode d'action des protéines MAIN et MAIL1, en testant si ces protéines interagissent avec la chromatine. Des analyses de fractionnement cellulaire ont été réalisées et les résultats seront présentés dans le chapitre 3 de la partie RESULTATS.
- 4) Finalement, de définir l'impact des protéines MAIL2 et MAIL3 sur la régulation de la transcription et le développement de la plante. Les résultats préliminaires obtenus seront présentés dans le chapitre 4 de la partie RESULTATS.

RESULTATS

CHAPITRE 1

Caractérisation du rôle de MAIN et MAIL1 dans le contrôle de la transcription et identification d'un partenaire protéique.

Le premier chapitre est consacré aux analyses visant à caractériser les mutants KO *main-2* et *mail1-1*. Ces travaux ont fait l'objet d'un article récemment publié dans la revue PLoS Genetics (Nicolau *et al.*, 2020). Dans cet article, nous avons caractérisé la lignée *ddc#16* issue du crible génétique *ATCOPIA28::GFP*, et nous avons montré que le gène muté dans cette population est *MAIN*. En utilisant la lignée *ddc#16*, nous avons étudié l'interaction génétique entre *MAIN* et la voie de la méthylation de l'ADN, montrant que ces deux voies agissent synergiquement pour la répression des TEs. Par la suite, nous avons caractérisé les mutants KO *main-2* et *mail1-1* au niveau moléculaire, montrant l'impact de leur mutation sur la régulation de l'expression des gènes et la répression des TEs. Nous avons également montré que la méthylation de l'ADN est très peu affectée dans ces mutants. En utilisant des lignées exprimant des protéines MAIN et MAIL1 étiquetées, nous avons pu montrer que ces protéines interagissent ensemble, mais aussi avec PP7L, une phosphoprotéine phosphatase. Suite à cette analyse, des plantes mutantes *mail1-1 pp7l-2* nous ont permis de montrer que les deux protéines agissent dans la même voie génétique pour la régulation de la transcription et la répression des TEs. Nous avons aussi identifié une séquence enrichie dans le promoteur des gènes sous-exprimés dans les différentes lignées mutantes, que nous avons nommé « DOWN motif ». Enfin, nous avons réalisé une étude phylogénique des domaines PMD et PPP, et nous avons montré que ces deux domaines ont co-évolués chez les Eudicotylédons.

RESEARCH ARTICLE

The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in *Arabidopsis thaliana*

Melody Nicolau^{1,2}, Nathalie Picault^{1,2}, Julie Descombin^{1,2}, Yasaman Jami-Alahmadi³, Suhua Feng⁴, Etienne Bucher⁵, Steven E. Jacobsen^{4,6}, Jean-Marc Deragon^{1,2,7}, James Wohlschlegel³, Guillaume Moissiard^{1,2*}

1 LGDP-UMR5096, CNRS, Perpignan, France, **2** LGDP-UMR5096, Université de Perpignan, France, **3** Department of Biological Chemistry, University of California at Los Angeles, Los Angeles, California, United States of America, **4** Department of Molecular, Cell and Developmental Biology, University of California at Los Angeles, Los Angeles, California, United States of America, **5** Plant Breeding and Genetic Resources, Agroscope, Nyon, Switzerland, **6** Howard Hughes Medical Institute, University of California at Los Angeles, Los Angeles, California, United States of America, **7** Institut Universitaire de France, Paris, France

* guillaume.moissiard@univ-perp.fr



OPEN ACCESS

Citation: Nicolau M, Picault N, Descombin J, Jami-Alahmadi Y, Feng S, Bucher E, et al. (2020) The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in *Arabidopsis thaliana*. PLoS Genet 16(4): e1008324. <https://doi.org/10.1371/journal.pgen.1008324>

Editor: Claudia Köhler, Swedish University of Agricultural Sciences (SLU), SWEDEN

Received: July 17, 2019

Accepted: February 28, 2020

Published: April 14, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1008324>

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: Nucleotide sequencing data generated in this study have been

Abstract

Transposable elements (TEs) are DNA repeats that must remain silenced to ensure cell integrity. Several epigenetic pathways including DNA methylation and histone modifications are involved in the silencing of TEs, and in the regulation of gene expression. In *Arabidopsis thaliana*, the TE-derived plant mobile domain (PMD) proteins have been involved in TE silencing, genome stability, and control of developmental processes. Using a forward genetic screen, we found that the PMD protein MAINTENANCE OF MERISTEMS (MAIN) acts synergistically and redundantly with DNA methylation to silence TEs. We found that MAIN and its close homolog MAIN-LIKE 1 (MAIL1) interact together, as well as with the phosphoprotein phosphatase (PPP) PP7-like (PP7L). Remarkably, *main*, *mail1*, *pp7l* single and *mail1 pp7l* double mutants display similar developmental phenotypes, and share common subsets of upregulated TEs and misregulated genes. Finally, phylogenetic analyses of PMD and PP7-type PPP domains among the Eudicot lineage suggest neo-association processes between the two protein domains to potentially generate new protein function. We propose that, through this interaction, the PMD and PPP domains may constitute a functional protein module required for the proper expression of a common set of genes, and for silencing of TEs.

Author summary

The plant mobile domain (PMD) is a protein domain of unknown function that is widely spread in the angiosperm plants. Although most PMDs are associated with repeated DNA sequences called transposable elements (TEs), plants have domesticated the PMD to

deposited in European Nucleotide Archive (ENA) under the accession number PRJEB33240 (<http://www.ebi.ac.uk/ena/data/view/PRJEB33240>). The proteomics data have been deposited to the MassIVE data repository (<https://massive.ucsd.edu>) with the dataset identifier MSV000084089.

Funding: This study was supported by the French Laboratory of Excellence project "TULIP" (ANR-10-LABX-41; ANR-11-IDEX-0002-02), by a UPVD BQR grant, and by a grant from the French CNRS GDR EPIPLANT consortium. Work in the Jacobsen laboratory was supported by NIH grant 1R35GM130272 and a grant from the W.M. Keck Foundation. S. E. J. is an Investigator of the Howard Hughes Medical Center. E.B. was supported by ERC 725701 BUNGEE. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

produce genic versions that play important roles within the cell. In *Arabidopsis thaliana*, MAINTENANCE OF MERISTEMS (MAIN) and MAIN-LIKE 1 (MAIL1) are genic PMDs that are involved in genome stability, developmental processes, and silencing of TEs. The mechanisms involving MAIN and MAIL1 in these cellular processes remain elusive. Here, we show that MAIN, MAIL1 and the phosphoprotein phosphatase (PPP) named PP7-like (PP7L) interact to form a protein complex that is required for the proper expression of genes, and the silencing of TEs. Phylogenetic analyses revealed that PMD and PP7-type PPP domains are evolutionary connected, and several plant species express proteins carrying both PMD and PPP domains. We propose that interaction of PMD and PPP domains would create a functional protein module involved in mechanisms regulating gene expression and repressing TEs.

Introduction

In eukaryotes, DNA methylation and post-translational modifications of histones are epigenetic marks involved in chromatin organization, regulation of gene expression and silencing of DNA repeats such as transposable elements (TEs) [1–3]. Constitutive heterochromatin is highly condensed and enriched in silenced TEs that are targeted by DNA methylation and histone H3 lysine 9 dimethylation (H3K9me₂). Euchromatin is more relaxed and composed of genes that are more permissive to transcription, depending on the recruitment of transcription factors (TFs), cofactors and RNA polymerases [1, 4]. In plants, DNA methylation occurs in three different cytosine contexts: CG, CHG and CHH (where H = A, T or C), involving specialized DNA methyltransferases [5]. In *Arabidopsis thaliana*, DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) and DRM1 mediate de novo DNA methylation in all sequence contexts through the RNA-directed DNA methylation (RdDM) pathway, which involves among other components, RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) and DICER-LIKE 3 (DCL3) for the production of short interfering (si)RNAs [6, 7]. The maintenance of CG methylation is specifically performed by METHYLTRANSFERASE 1 (MET1), while CHROMOMETHYLASE 2 (CMT2) and CMT3 are involved in the maintenance at CHG sites [8, 9]. CMT2 can also be involved in the deposition of CHH methylation at specific genomic location [10, 11]. Finally, DRM2 is mostly required for the maintenance of CHH methylation through the RdDM pathway [6, 7, 9]. Together with DNA methylation, additional pathways play important roles in TE silencing. The MICRORCHIDIA 1 (MORC1) and MORC6 ATPases interact together, and are required for heterochromatin condensation and repression of TEs, acting mostly downstream of DNA methylation and RdDM pathway [12–14].

More recently, the *A. thaliana* plant mobile domain (PMD) proteins MAINTENANCE OF MERISTEM (MAIN) and MAIN-LIKE 1 (MAIL1) were identified as new factors required for TE silencing [15]. In addition, these two proteins have been involved in genome stability, and regulation of developmental processes such as cell division and differentiation [16, 17]. The PMD is a large protein domain of unknown function that is widely represented among the angiosperms, predominantly associated with TEs [15, 18]. It has been proposed that genic PMD versions, such as the MAIN and MAIL1 proteins derived from TEs after gene domestication [15, 18, 19]. Previous studies suggested that genic PMDs could act as cellular factors related to transcription, possibly acting as transcription factor (TF)-like, co-factor or repressor proteins regulating this cellular process [16, 18]. Nevertheless, the role of PMD proteins in the regulation of transcription remains elusive. Most of genic PMD proteins are standalone

versions, however, in some cases, the PMD is fused to another protein domain, such as protease, kinase or metallo-phosphatase (MPP) domains. For instance in *A. thaliana*, the MAIL3 protein carries a PMD, which is fused to a putative serine/threonine-specific phosphoprotein phosphatase (PPP) domain phylogenetically related to the plant-specific protein phosphatase 7 (PP7) [20]. PP7 is a calmodulin-binding PPP that has been related to cryptochrome (CRY)-mediated blue-light signaling, and to the control of stomatal aperture [20–22]. PP7 is also involved in the perception of red/far red light by controlling the phytochrome pathway [23, 24]. In addition to PP7 and MAIL3 (also known as “long PP7”), the protein PP7-like (PP7L) belongs to the same phylogenetic clade [20]. PP7L was recently identified as a nuclear protein involved in chloroplast development and abiotic stress tolerance [25]. The *pp7l* mutant plants showed photosynthetic defects and strong developmental phenotype associated with misregulation of several genes [25].

In this study, we described a forward genetic screen based on a GFP reporter gene that allowed us to identify a mutant population in which *MAIN* is mutated, leading to GFP overexpression. We then deciphered the genetic interaction between the *DRM2*, *CMT3* and *MAIN*, showing that these proteins are part of different epigenetic pathways that act redundantly or synergistically to repress TEs. Biochemical analyses indicated that *MAIN* and *MAIL1* physically interact together. These analyses also identified *PP7L* as a robust interactor of *MAIN* and *MAIL1* proteins. In addition, the characterization of developmental and molecular phenotypes of *pmd* and *pp7l* single and double mutant plants strongly suggest that these proteins interact together to silence TEs, and regulate the expression of a common set of genes. Finally, phylogenetic analyses allowed us to determine the distribution of PMD and PP7/PP7L domains among the Eudicots. Based on these analyses, we have evidences of co-evolution linked to the neo-association of the PMD and PP7-type PPP domains on single proteins in several Eudicot species, suggesting a convergent evolution between these two protein domains.

Results

Mutation in *MAIN* is responsible for TE silencing defects

The *ATCOPIA28* retrotransposon *AT3TE51900* (hereafter called *ATCOPIA28*) is targeted by several epigenetic pathways such as DNA methylation and the *MORC1/6* complex, which altogether contribute to its repression. We engineered a construct in which the 5' long terminal repeat (LTR) promoter region of *ATCOPIA28* controls GFP transcription (Fig 1A). While the *ATCOPIA28::GFP* transgene is fully silenced in wild type (WT) plants, it is weakly expressed in the DNA methylation-deficient *drm1 drm2 cmt3 (ddc)* triple mutant background (Fig 1B) [26]. We performed an ethyl methane sulfonate (EMS) mutagenesis using the *ATCOPIA28::GFP ddc* plants as sensitized genetic material, and screened for mutant populations showing GFP overexpression. Among, the selected populations, we retrieved two new mutant alleles of *MORC6* carrying missense mutations in either the GHKL or S5 domains of the protein (S1A–S1C Fig). We also identified the population *ddc #16* showing strong overexpression of GFP and misregulation of several endogenous TEs, including *ATCOPIA28* (Fig 1B–1D). Mapping experiments based on whole genome resequencing and bulk segregant analysis indicated that *ddc #16* carries a missense point mutation (C230Y) in the gene *AT1G17930*, previously named *MAIN* (S1D and S1E Fig). Genetic complementation analyses by crossing the *ddc #16* EMS mutant with the knock-out (KO) transferred DNA (T-DNA) insertion line *main-2* generated F1 *ddc #16 x main-2* plants that did not express the GFP (S1F Fig). Transcriptional profiling analyses showed, however, that endogenous TEs, including *ATCOPIA28*, were upregulated in F1 *ddc #16 x main-2* plants, but not in F1 control plants generated from the backcross of *ddc #16* with WT Columbia (Col) plants (S1G Fig). Self-fertilization of F1 *ddc #16 x main-2* plants

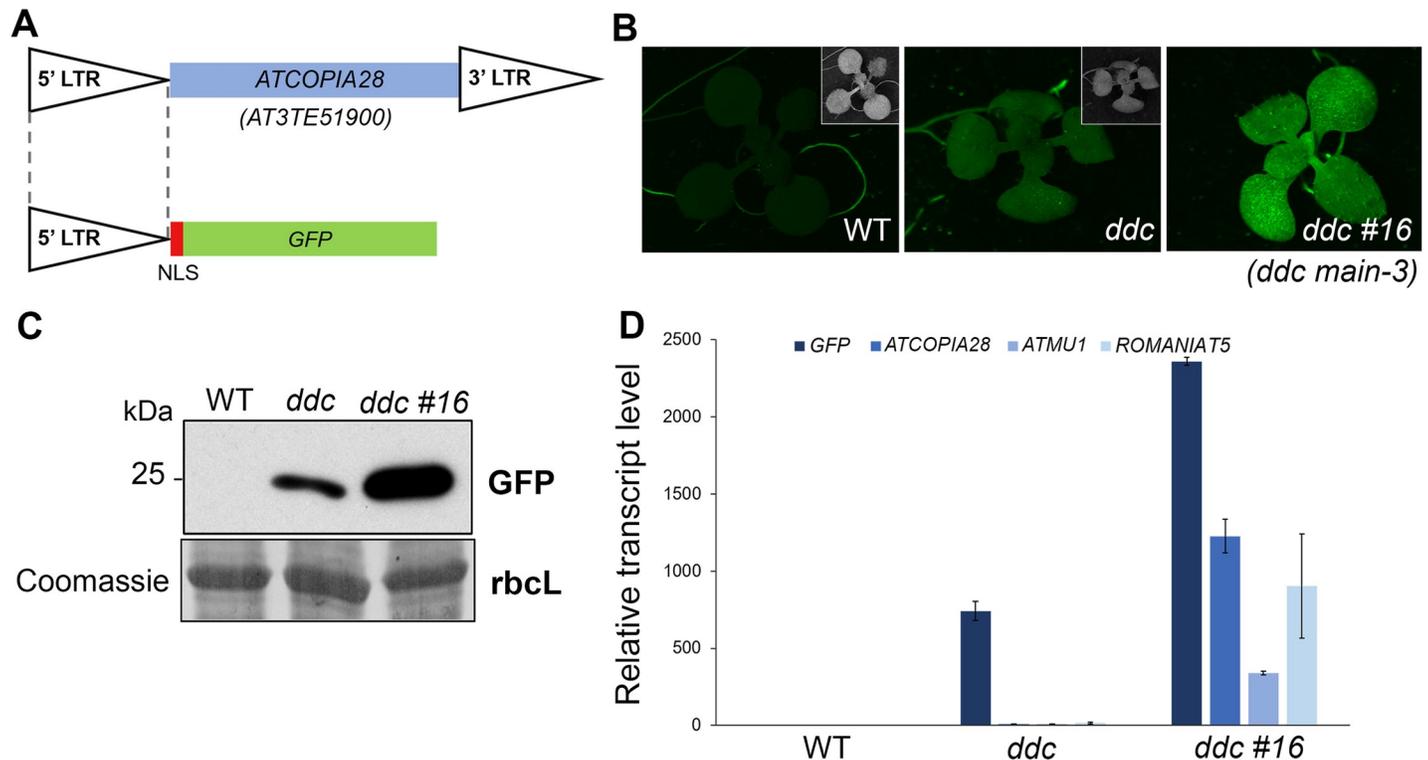


Fig 1. The *ddc #16* EMS population shows overexpression of *ATCOPIA28::GFP* and upregulation of endogenous TEs. (A) Schematic representation of the *ATCOPIA28::GFP* transgene. The 5' long terminal repeat (LTR) promoter region of an *ATCOPIA28* LTR-retrotransposon (AT3TE51900) is used to control the expression of GFP. The construct carries a Nuclear Localization Signal (NLS) to target the GFP in the nucleus. (B) WT and *drm1 drm2 cmt3* (*ddc*) triple mutant plants carrying the *ATCOPIA28::GFP* transgene showed no and weak GFP fluorescence under UV light, respectively. By comparison, the *ddc #16* EMS mutant showed strong GFP fluorescence. Insets show plants under white light. (C) Western blot using anti-GFP antibody confirmed *ATCOPIA28::GFP* overexpression in *ddc #16*. Coomassie staining of the large Rubisco subunit (rbcL) is used as a loading control. kDa: kilodalton. (D) Relative expression analyses of *ATCOPIA28::GFP* (*GFP*) and three endogenous TEs in *ddc* and *ddc #16* assayed by Real-Time quantitative PCR (RT-qPCR). RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the mutants are represented relative to WT. Error bars indicate standard deviation based on three independent biological replicates. Screening of EMS mutant populations was done on MS plates to allow for visualization of GFP-positive individuals under UV light.

<https://doi.org/10.1371/journal.pgen.1008324.g001>

allowed us to retrieve several F2 *ddc #16* x *main-2* plants overexpressing the GFP (S1F Fig). Among these GFP positive F2 plants, we identified individuals that were either homozygote for the EMS mutation in the *MAIN* gene, or plants carrying both the EMS and T-DNA *main-2* mutant alleles (S1F Fig). Moreover, while all these plants were homozygote for the *drm2* mutation, half of them segregated the *cmt3* mutation. Thus, altogether, these analyses suggested that *ATCOPIA28::GFP* silencing is more DRM2- than CMT3-dependent. More importantly, they confirmed that *MAIN* was the mutated gene causing the upregulation of *ATCOPIA28::GFP* and several endogenous TEs. Therefore, *ddc #16* was renamed *ddc main-3*.

The MAIN, DRM2 and CMT3 pathways act synergistically to repress TEs and DNA-methylated genes

To determine the genetic interaction of *ddc* and *main-3* mutations on TE silencing, we carried out two independent RNA sequencing (RNA-seq) experiments in the hypomorphic *main-3* single, *ddc* triple and *ddc main-3* quadruple mutant plants (Fig 2A and S2A Fig). As previously described, the *ddc* mutant showed upregulation of several TEs spread over the five chromosomes (Fig 2B–2D and S2B Fig and S1 Table) [11]. Loss of TE silencing was also observed to a milder degree in the *main-3* mutant, with the significant enrichment of pericentromeric TEs

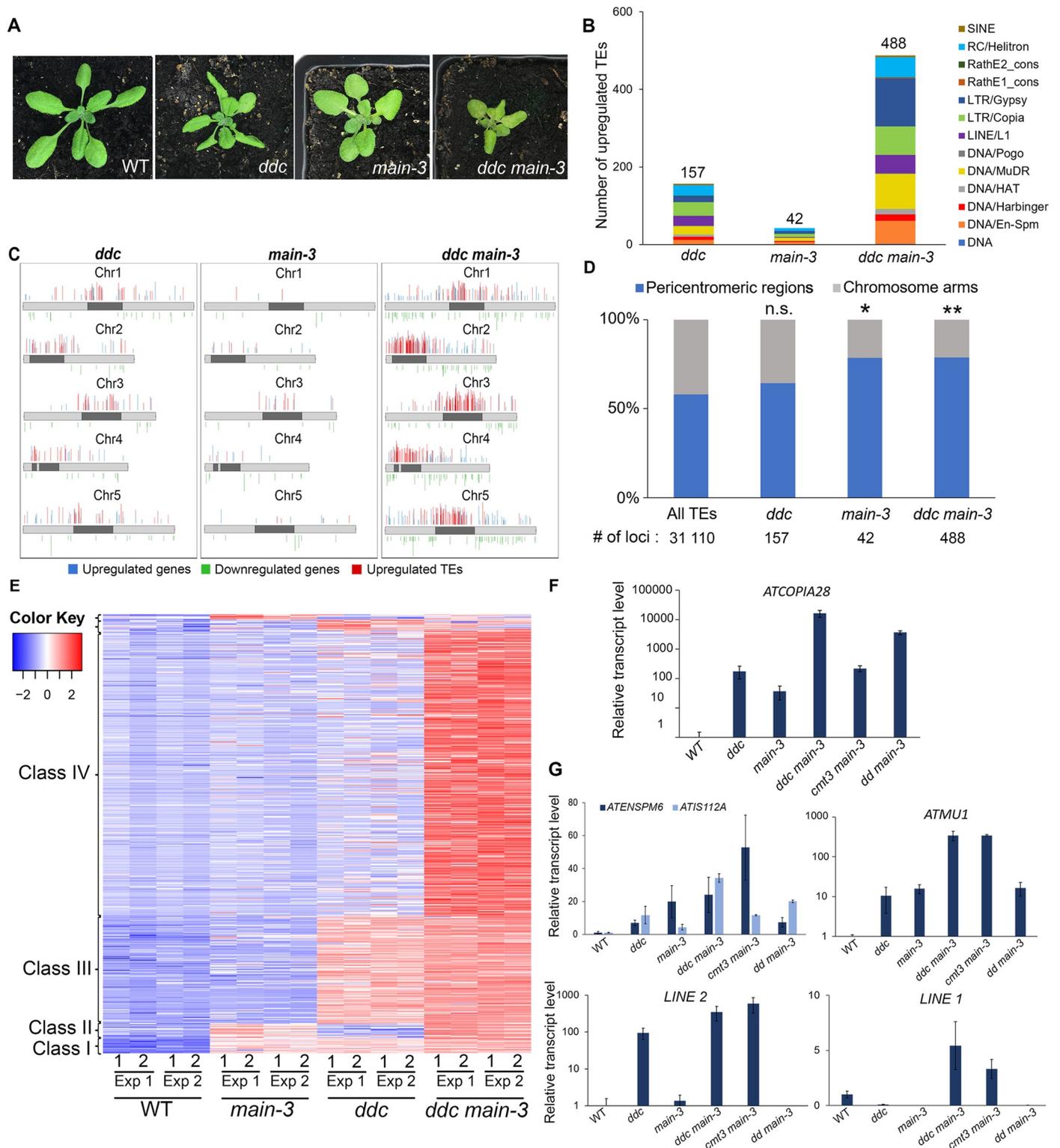


Fig 2. MAIN, DRM2 and CMT3 act synergistically to repress TEs. (A) Representative pictures showing the developmental phenotype of 3-week-old *ddc*, *main-3* and *ddc main-3* mutants in comparison to WT plant. (B) Number of upregulated TEs in *ddc*, *main-3* and *ddc main-3* over WT, and classified by TE superfamily. (C) Chromosomal distributions of misregulated loci in *ddc*, *main-3* and *ddc main-3* over WT. Chromosome arms are depicted in light grey, pericentromeric regions in dark grey as defined in [50]. Upregulated genes and TEs are represented in blue and red, respectively; downregulated genes are represented in green. (D) Fraction of upregulated TEs in *ddc*, *main-3* and *ddc main-3* located in chromosome arms or in pericentromeric regions as defined in [50]. Asterisks indicate statistically

significant enrichments of TEs in pericentromeric regions in comparison to the genomic distribution of all *A. thaliana* TEs (Chi-Square test, *: p-value \leq 0.05, **: p-value \leq 0.01 n.s: not significant). (E) Heatmap showing upregulated TEs in *ddc*, *main-3* and *ddc main-3* mutants in comparison to WT plants. (F-G) Relative expression analyses of *ATCOPIA28* (F) and several endogenous TEs (G) in *ddc*, *main-3*, *ddc main-3*, *cmt3 main-3* and *drm1 drm2 (dd) main-3* assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHPI1* gene, and transcript levels in the mutants are represented relative to WT. Error bars indicate standard deviation based on three independent biological replicates. RNA-seq threshold: $\log_2 \geq 2$, or $\log_2 \leq -2$; p-adj $<$ 0.01.

<https://doi.org/10.1371/journal.pgen.1008324.g002>

among the upregulated TEs (Fig 2B–2D and S2B Fig and S1 Table). The *ddc main-3* mutant showed an exacerbation of TE silencing defects, with a large number of pericentromeric TEs being specifically upregulated in this mutant background (Fig 2B–2D and S2B Fig and S1 Table). Comparative analyses revealed that upregulated TEs cluster into four distinct classes (Fig 2E and S2C Fig). Class I TEs are upregulated in *ddc*, *main-3* and *ddc main-3* mutants (Fig 2E and S2C and S2D Fig). Class II and class III TEs are targeted by the MAIN and DRM2/CMT3 pathways, respectively (Fig 2E and S2C and S2D Fig). However, the upregulation of class II and class III TEs is further enhanced in *ddc main-3*, which suggests that the MAIN and DRM2/CMT3 pathways can partially compensate each other at these genomic locations (S2D Fig). Finally, the most abundant class IV TEs are only misregulated in *ddc main-3*, which implies that the MAIN and DRM2/CMT3 pathways act redundantly to silence these TEs (Fig 2E and S2C and S2D Fig).

Several genes were also misregulated in the three mutant backgrounds (S1 Table). Among these genes, a subset was commonly upregulated in *ddc*, *main-3* and *ddc main-3* (S2E Fig). Remarkably, genes that were upregulated in *ddc*, *main-3* or *ddc main-3* were significantly enriched in pericentromeric regions of chromosomes, where constitutive heterochromatin resides (S2F Fig). This is consistent with the fact that, among these upregulated genes, we identified a large proportion of genes that were DNA-methylated (in the three cytosine contexts) and targeted by H3K9me2 (S2F Fig). Conversely, we could only identify one gene commonly downregulated in *ddc*, *main-3* and *ddc main-3* (S2F Fig). Furthermore, downregulated genes in *ddc*, *main-3* or *ddc main-3* were rather enriched in chromosome arms, and most of them were not DNA-methylated genes (S2F Fig).

To further dissect the genetic interaction between the DRM2, CMT3 and MAIN pathways, we generated the *drm1 drm2 main-3 (dd main-3)* and *cmt3 main-3* mutants (S2G Fig). We then analyzed the expression level of several TEs previously identified as misregulated in *ddc*, *main-3* and/or *ddc main-3*. The endogenous *ATCOPIA28* was the most expressed in *ddc main-3* and *dd main-3*, and to a lesser extent, in *cmt3 main-3* (Fig 2F). This is consistent with the fact that all the F2 *ddc* #16 x *main-2* plants overexpressing *ATCOPIA28::GFP* were *drm2* homozygote, although they segregated the *cmt3* mutation (S1F Fig). Further analyses showed that most of the tested TEs tend to be more expressed in *cmt3 main-3* than in *dd main-3*, with the exception of *ATIS112A* that was more upregulated in *dd main-3* than in *cmt3 main-3* (Fig 2G). In conclusion, these analyses showed complex genetic interactions between the DRM2, CMT3 and MAIN pathways, suggesting that MAIN and DNA methylation pathways act synergistically to repress TEs and DNA-methylated genes.

MAIN and MAIL1 are required for the proper expression of a common set of genes and TEs

Beside a role of MAIN in TE and gene silencing, our transcriptomic analyses using the hypomorphic *main-3* mutant suggested that MAIN would be required for the expression of several genes that are not controlled by the DRM2 and CMT3 pathway (S2E Fig). To further study the role of MAIN and MAIL1 in the regulation of gene expression and TE silencing, we performed

two independent RNA-seq experiments in the *main-2* and *mail1-1* null mutants (RNA-seq Exp1 and Exp3), and combined these experiments with the reanalysis of previously published RNA-seq datasets (RNA-seq Exp2) [15]. Principal component analyses (PCA) showed that for each RNA-seq experiment, *main-2* and *mail1-1* mutant samples tend to cluster together, and away from the WT samples (S3A Fig). Analyzing these three RNA-seq experiments together allowed to identify large numbers of genes and TEs that were misregulated in the *main-2* and *mail1-1* null mutants (Fig 3A and 3B and S2 Table).

We then compared the transcriptomes of *main-2* and *mail1-1* mutants, together with the *main-3* mutant allele (Fig 3A and 3B, S1 and S2 Tables). As expected by the fact that *main-2* and *mail1-1* are null mutants while *main-3* is a hypomorphic mutant allele, we identified greater numbers of misregulated loci in *main-2* and *mail1-1* in comparison to *main-3* (Fig 3A and 3B). Fractions of these loci were specifically misregulated in each mutant background (Fig 3C and 3D). In addition, we identified subsets of genes and TEs that were only misregulated in *main-2* and *mail1-1* null mutants, but not in the hypomorphic *main-3* mutant (Fig 3C and 3D and S3 Table). Finally, these analyses revealed subsets of loci that were commonly misregulated in the three mutant backgrounds (Fig 3C and 3D, S3B–S3D Fig and S3 Table).

The biggest overlaps between misregulated loci in *main-2*, *mail1-1* and *main-3* mutants were among the downregulated genes and upregulated TEs, whereas only a small proportion of genes commonly upregulated in *main-2* and *mail1-1* were also upregulated in *main-3* (Fig 3D). As observed in *main-3* (S2F Fig), upregulated TEs in *main-2* and *mail1-1* were enriched in pericentromeric regions, and genes that were downregulated in *main-2* and *mail1-1* were not targeted by DNA methylation, and mostly located in the chromosome arms (Fig 3E). However, unlike in *main-3*, the upregulated genes in *main-2* and *mail1-1* were not enriched in pericentromeric regions, and only small fractions of them were DNA-methylated genes (Fig 3E). This discrepancy can be explained by the fact that *main-2* and *mail1-1* null mutations have a much greater impact on the misregulation of gene expression than the hypomorphic *main-3* mutant allele.

Finally, we compared the sets of misregulated loci in *main-2*, *mail1-1*, *ddc* and *ddc main-3* (S1 and S2 Tables). We found significant overlaps among upregulated genes and TEs between *main-2*, *mail1-1*, *ddc* and *ddc main-3* (S3E Fig). This suggests that MAIN, MAIL1, DRM2 and CMT3 cooperate to silence these subsets of genes and TEs. However, we could not find significant overlaps among downregulated genes between *main-2*, *mail1-1* and *ddc* (S3E Fig). Instead, a significant overlap was identified only by comparing the lists of downregulated genes in *main-2*, *mail1-1* and *ddc main-3*, three genetic backgrounds carrying a mutation in either MAIN or MAIL1 (S3E Fig). Thus, this suggests that MAIN and MAIL1 are required for the expression of specific genes, in a DRM2- and CMT3-independent manner.

In conclusion, these comparative analyses allowed to precisely define the loci that were misregulated in *main-2* and *mail1-1* in comparison to *main-3*, *ddc* and *ddcmain-3* mutants. Among these loci, several TEs and DNA-methylated genes are commonly targeted by the MAIN, MAIL1, DRM2 and CMT3 pathways, which suggests that MAIN, MAIL1 and DNA methylation pathways cooperate to silence these TEs and DNA-methylated genes. Besides, several genes are downregulated in *main-2* and *mail1-1*, and subsets of these genes are also downregulated in *main-3*, and *ddcmain-3* but not in *ddc*. This suggests that the MAIN and MAIL1 act independently of DRM2 and CMT3 to ensure the expression of these genes. Finally, these results revealed important overlaps between the misregulated loci in *main-2* and *mail1-1* null mutants, which strongly suggests that the two proteins act in the same pathway to regulate the expression of common sets of loci.

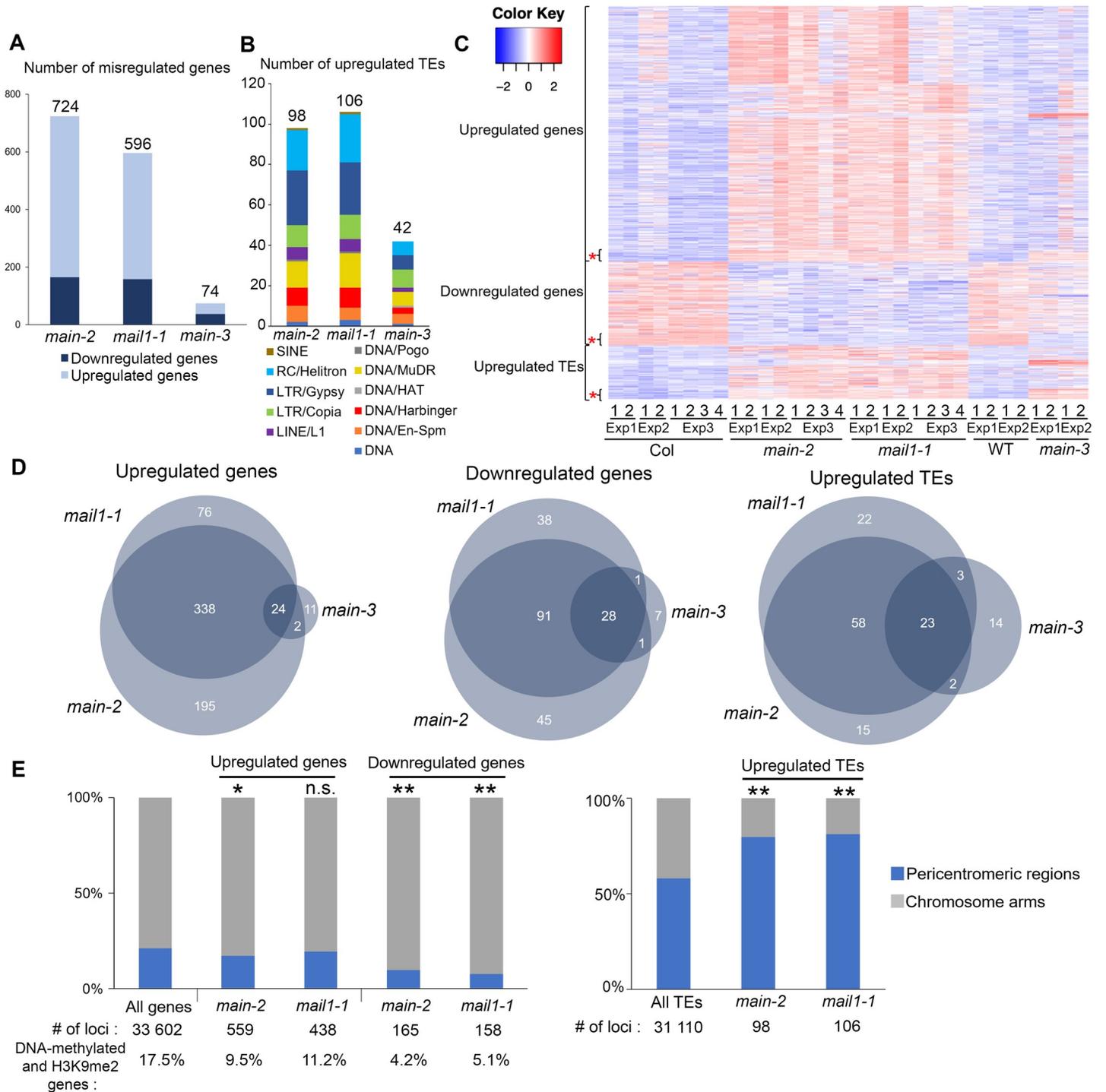


Fig 3. MAIN and MAIL1 are required for the proper expression of similar genes, and for TE silencing. (A-B) Number of misregulated genes (A) and upregulated TEs (B) in *main-2*, *mail1-1* and *main-3* mutants in comparison to WT Col plants. TEs are classified by superfamily. (C) Heatmap showing misregulated loci in *main-2*, *mail1-1* and *main-3* in comparison to Col and WT controls, respectively. Asterisks represents loci that are commonly misregulated in the three mutant backgrounds. (D) Venn diagrams analyses representing the overlaps between misregulated loci in *main-2*, *mail1-1* and *main-3*. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value < 2.2.10e-16). (E) Fraction of misregulated loci in *main-2* and *mail1-1* located in chromosome arms or in pericentromeric regions as defined in [50]. Asterisks indicate statistically significant enrichments of downregulated genes and upregulated genes and TEs in chromosome arms and pericentromeric regions, respectively, in comparison to the genomic distributions of all *A. thaliana* genes and TEs (Chi-Square test, *: p-value ≤ 0.05, **: p-value ≤ 0.01, n.s: not significant). Percentages of genes targeted by DNA methylation and H3K9me2 were calculated based on enrichment in heterochromatin states 8 and 9 as defined in [51]. RNA-seq threshold: log₂ ≥ 2, or log₂ ≤ -2; p-adj < 0.01.

<https://doi.org/10.1371/journal.pgen.1008324.g003>

Slight increase in non-CG methylation in the *main-2* mutant does not correlate with changes in gene expression and TE silencing defect

Whole genome bisulfite sequencing (BS-seq) analyses showed that, at the chromosome scale, DNA methylation level is mostly unchanged in *main-2* in comparison to WT, with the exception of a slight increase in CHG methylation in pericentromeric regions (Fig 4A). Subtle but statistically significant CHG hypermethylation was further confirmed in pericentromeric TEs and genes, which are mostly TE genes (Fig 4B and 4C). Slight CHG and CHH hypermethylation was also detected in TEs located in chromosome arms (Fig 4D). Conversely, genes located

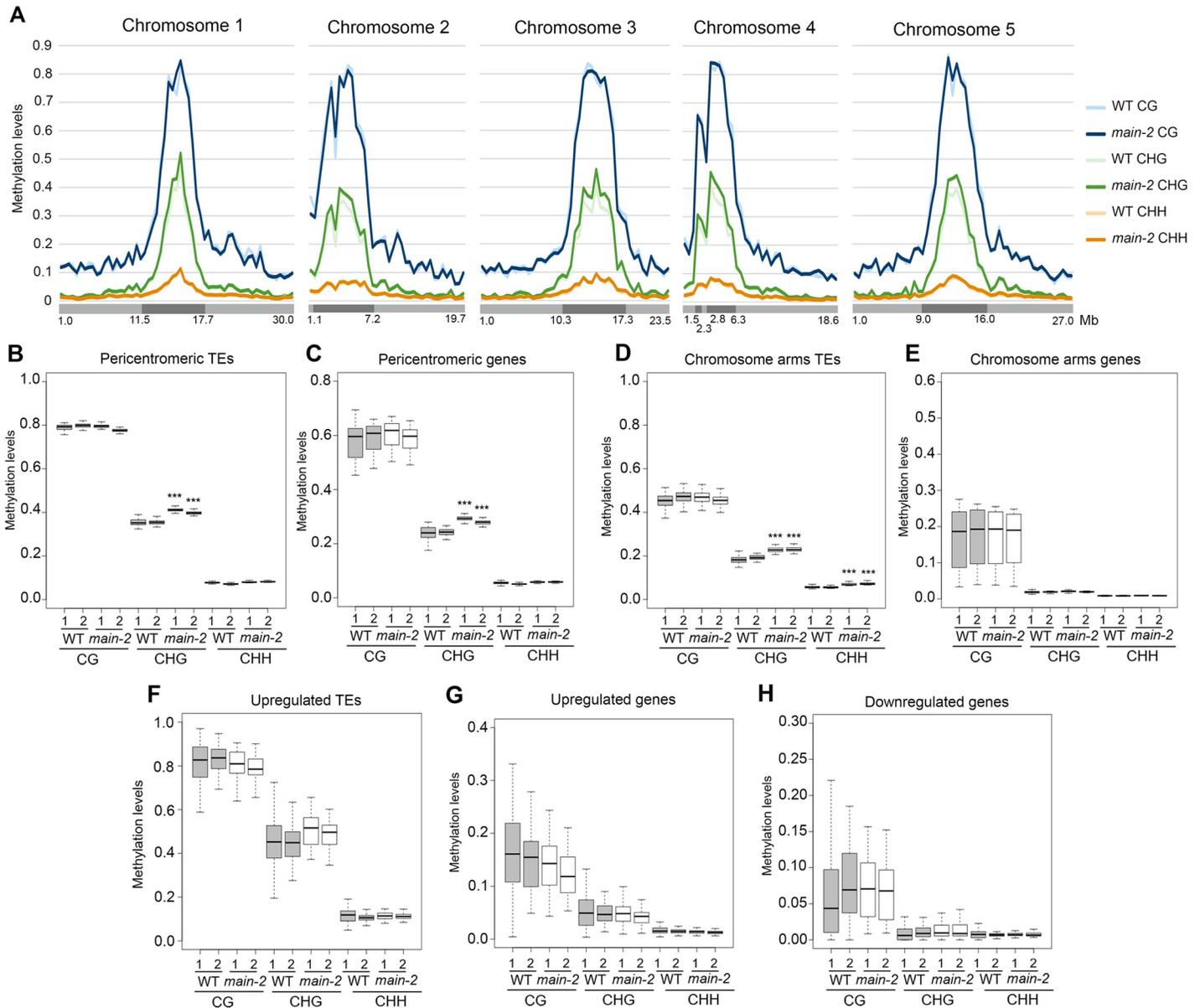


Fig 4. The *main-2* mutation has a slight effect on non-CG DNA methylation levels. (A) Genome-wide DNA methylation levels along the five *Arabidopsis* chromosomes in *main-2* versus WT Col plants. Chromosome arms are depicted in light grey, pericentromeric regions in dark grey as defined in [50]. Mb: megabase. (B-H) Boxplot analyses in two *main-2* and WT Col biological replicates showing the DNA methylation levels of all pericentromeric TEs (B) and genes (C), all chromosome arms TEs (D) and genes (E), TEs that are upregulated in *main-2* (F), and genes that are upregulated (G) and downregulated (H) in *main-2*. p-values were calculated using a Wilcoxon test. ***: p-value < 2.10e-16.

<https://doi.org/10.1371/journal.pgen.1008324.g004>

in chromosome arms did not show significant changes in DNA methylation level in *main-2* (Fig 4E). Identical results were obtained by analyzing the DNA methylation level at upregulated TEs and misregulated genes in *main-2* (Fig 4F–4H). We then analyzed the DNA methylation level at genomic locations previously defined as differentially hypomethylated regions (hypo DMRs) at CHG and CHH sites in *cmt3* and *drm1drm2* (*dd*) mutants, respectively [26]. The *cmt3* and *dd* hypo DMRs are mostly located in TEs. As observed with pericentromeric genes and all TEs (Fig 4B–4D), we found slight increases in CHG and CHH methylation at *cmt3* and *dd* hypo DMRs, respectively, in *main-2* (S4A and S4B Fig). Finally, DMR calling in *main-2* using stringent parameters only identified a few DMRs (S4C Fig). Thus, DNA methylation is mostly unaffected in *main-2*, with the exception of a slight increase in non-CG methylation at pericentromeric genes and all TEs. Moreover, this subtle non-CG hypermethylation does not correlate with changes in gene and TE expression observed in *main-2* because DNA methylation level in *main-2* is unchanged at these misregulated loci (Fig 4F–4H).

MAIN, MAIL1 and the metallo-phosphatase PP7L physically interact together

The *main-2* and *mail1-1* null mutants display similar molecular and developmental phenotypes (Fig 3 and Fig 5A). Thus, we hypothesized that MAIN and MAIL1 proteins may act in the same pathway, possibly by interacting together. To test this hypothesis, we generated transgenic lines expressing FLAG- and MYC-tagged genomic PMD versions driven by their endogenous promoters. We confirmed that epitope-tagged MAIN and MAIL1 proteins were produced at the expected sizes, and they could complement the respective developmental phenotypes of null mutant plants (Fig 5A and 5B). Importantly, they could also efficiently rescue the TE silencing and gene expression defects observed in *main-2* and *mail1-1* mutants, implying that epitope-tagged MAIN and MAIL1 are functional proteins (Fig 5C–5E). Using FLAG-tagged MAIN and MAIL1 expressing plants, immunoprecipitation followed by mass spectrometry (IP-MS) analyses were carried out to determine potential protein interactors. Mass spectrometry (MS) analyses indicated that MAIL1 was strongly immunoprecipitated with MAIN-FLAG and *vice versa* (Fig 5F). To validate IP-MS results, we crossed the MAIN-FLAG and MAIL1-MYC lines together. We then performed co-immunoprecipitation (co-IP) experiments using F1 hybrid plants co-expressing the two transgenes, and confirmed that MAIN and MAIL1 interact together (Fig 5G). MS analyses of MAIN-FLAG and MAIL1-FLAG IP also identified the metallo-phosphatase PP7L as putative interactor (Fig 5F). MAIN, MAIL1 and PP7L were the only three proteins reproducibly enriched across multiple replicates (Fig 5F). Co-IP experiments using plants co-expressing either PP7L-FLAG together with MAIN-MYC or MAIL1-MYC constructs confirmed the interaction between PP7L and each PMD protein (Fig 5H and 5I). Thus, the three proteins MAIN, MAIL1 and PP7L physically interact together.

The *main*, *mail1* and *pp7l* mutants display similar developmental and molecular phenotypes

PP7L is a putative metallo-phosphatase that was recently identified as a nuclear protein required for photosynthesis [20, 25]. The *pp7l-2* null mutant displays abnormal developmental phenotype reminiscent of *main-2* and *mail1-1* mutant plants, and 3-week-old *mail1-1 pp7l-2* double mutant plants do not show exacerbation of this phenotype (Fig 6A). To determine the genetic interaction between PMD and PP7L, we compared the transcriptomes of *main-2*, *mail1-1*, *pp7l-2* single and *mail1-1 pp7l-2* double mutants (S5A Fig and S2 and S4 Tables). We identified large numbers of misregulated loci in *pp7l-2* and *mail1-1 pp7l-2* (S5B and S5C Fig).

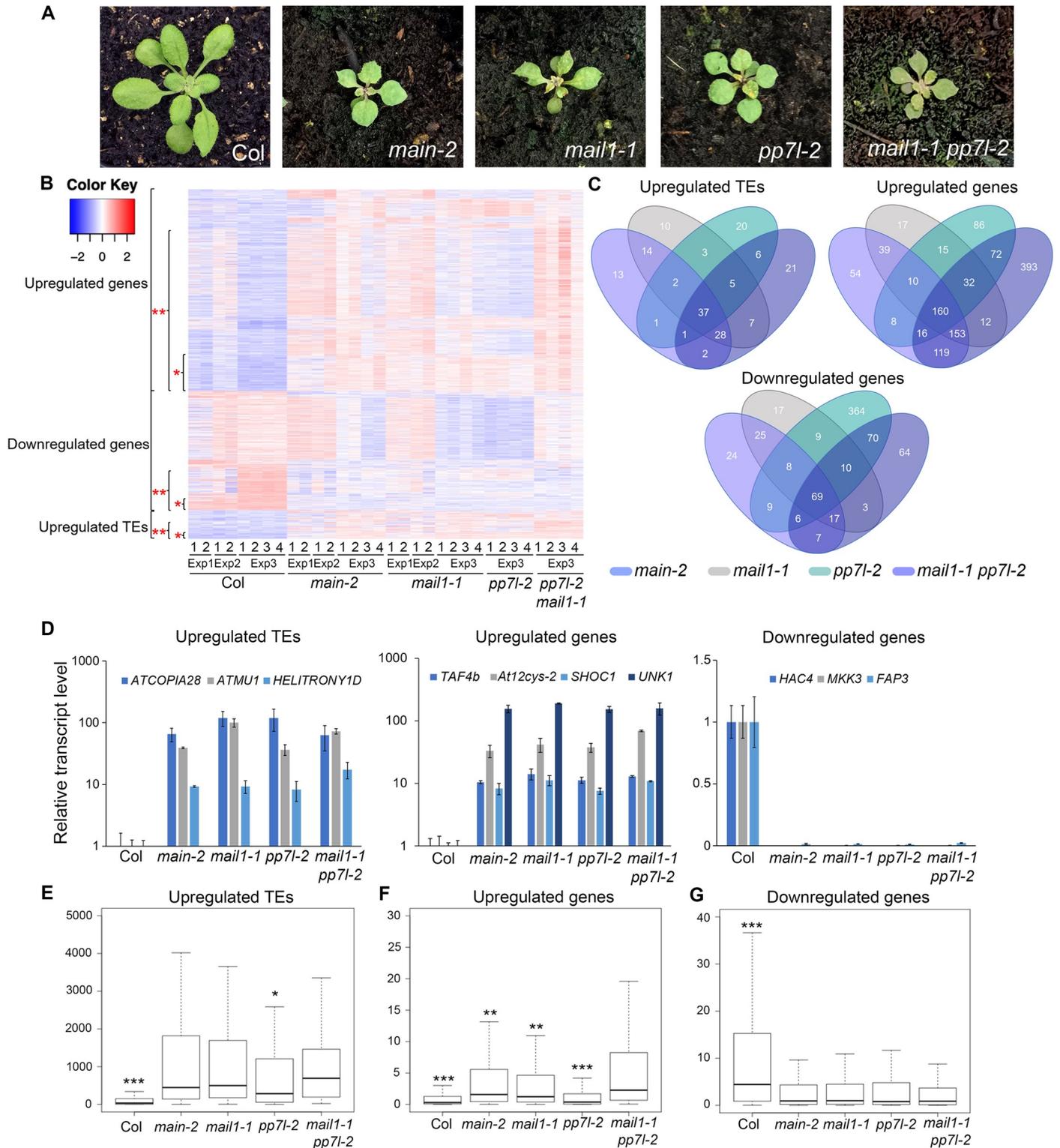


Fig 6. *main-2*, *mail1-1*, *pp7l-2* single and *mail1-1 pp7l-2* double mutants display similar developmental and molecular phenotypes. (A) Representative pictures of 3-week-old *main-2*, *mail1-1*, *pp7l-2* single and *mail1-1 pp7l-2* double mutants in comparison to WT *Col* plant. (B) Heatmap showing misregulated loci in *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2* mutants in comparison to WT *Col* plants using the datasets of RNA-seq Exp1, Exp2 and Exp3 (S2 and S4 Tables). One asterisk defines the loci that are commonly misregulated in all mutant backgrounds. Two asterisks define the loci that are misregulated in the *mail1-1 pp7l-2* double mutant. (C) Venn diagrams analyses representing the overlaps between misregulated loci in *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2*. Fisher's exact test statistically

confirmed the significance of Venn diagram overlaps (p-value < 2.2.10e-16). (D) Relative expression analyses of upregulated TEs, genes and downregulated genes in the different genotypes assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different mutants are represented relative to WT Col. Error bars indicate standard deviation based on three independent biological replicates. (E-G) Boxplots analyses showing average RPKM values of upregulated TEs (E), upregulated genes (F) and downregulated genes (G) in *mail1-1 pp7l-2* in the indicated genotypes of RNA-seq Exp3. These analyses are based on the misregulated loci datasets defined by ** in panel B. P-values were calculated using a Wilcoxon test, and only significant p-values are shown. *: p-value < 1.10e-3; **: p-value < 3.10e-6; ***: p-value < 2.10e-16.

<https://doi.org/10.1371/journal.pgen.1008324.g006>

As observed in *main-2* and *mail1-1*, TEs upregulated in *pp7l-2* and *mail1-1 pp7l-2* were enriched in pericentromeric regions, while up- and downregulated genes were mostly located in the chromosome arms, and not targeted by DNA-methylation (S5D Fig).

Comparative analyses revealed that significant proportions of loci were commonly misregulated in *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2* mutants, which is consistent with the fact that MAIN, MAIL1 and PP7L interact together to possibly regulate gene expression and silence TEs (Fig 6B–6D and S5 Table). These analyses also identified loci that were specifically misregulated in *main-2*, *mail1-1* or *pp7l-2*, which suggests that each protein is independently required for the proper expression of subsets of loci (Fig 6B and 6C). Besides, these analyses revealed loci that were exclusively misregulated in the *mail1-1 pp7l-2* double mutant, which implies that PP7L and MAIL1 may act redundantly to ensure the proper expression of these loci (Fig 6B and 6C). Further analyses showed that, among the loci that were misregulated in *mail1-1 pp7l-2*, upregulated genes were significantly more expressed in the double mutant than in each single mutant, and upregulated TEs were significantly differentially expressed only between *mail1-1 pp7l-2* and *pp7l-2* mutants (Fig 6E and 6F). Conversely, there was no significant difference of expression between the double mutant and single mutants for the downregulated genes (Fig 6G). Thus, these analyses suggest that combining the *pp7l-2* and *mail1-1* mutations may lead to synergistic defects mostly at genes that are upregulated in the double mutant.

We then performed in silico analyses to identify enriched DNA motif within a 1kb promoter region upstream of start codon of genes that were up- or downregulated in the different mutant backgrounds. We could not detect any enrichment of a DNA motif among any lists of upregulated genes (including overlapping lists). Likewise, we could not identify a DNA motif enriched in the lists of downregulated genes in *pp7l-2* or *ddc*. However, we identified a discrete DNA motif (hereafter called ‘DOWN’ motif) that was partially enriched in the promoter of genes that were downregulated in *main-2*, *mail1-1* and *mail1-1 pp7l-2* mutants (S5E Fig). The *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2* null mutants display strong developmental phenotype, and large numbers of misregulated loci (Fig 6A). Therefore, it is likely that some of the gene misregulation observed in these mutants might be due to side effects of the mutations. To overcome this issue and refine our analysis, we investigated the proportion of the ‘DOWN’ motif among downregulated genes in the hypomorphic *main-3* and *ddc main-3* mutants, as well as in the different overlapping lists of genes commonly downregulated (S3 and S5 Tables). The ‘DOWN’ motif was strongly enriched among the downregulated genes in *main-3*, and to a lesser extent in *ddc main-3* (S5E Fig). It was also significantly enriched in the overlapping lists of commonly downregulated genes in *main-2*, *mail1-1* and *main-3* as well as in the *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2* overlap (S5E Fig). It was further enriched in the promoters of genes commonly downregulated in all the mutant backgrounds—except *ddc*—analyzed in this study: twenty-five out of twenty-six genes, 96% of enrichment (S5E Fig, S6 and S7 Tables). We analyzed the DNA methylation level of the ‘DOWN’ motif in the promoters of these twenty-five genes in WT and *main-2*, and found that this DNA motif was not targeted by DNA methylation. Besides, further analyses showed that only a small fraction of all *Arabidopsis* genes carried the ‘DOWN’ motif in their promoter (12,46%, S5E Fig). Finally, random test

analyses based on twenty-six randomly picked genes strongly suggested that the enrichment of the 'DOWN' motif in the promoter of downregulated genes was substantial (S7 Table).

Thus, altogether, these analyses showed that MAIN, MAIL1 and PP7L are equally required for the repression of several genes and TEs. The three proteins are also required for the proper expression of a common set of genes that are downregulated in each single mutant as well as in *mail1-1 pp7l-2* double mutant, and significant fractions of these downregulated genes carry the 'DOWN' DNA motif in their promoter. Furthermore, the 'DOWN' DNA motif is strongly enriched among the genes that are always identified as downregulated in every mutant background carrying the *main-2*, *mail1-1*, *pp7l-2* or *main-3* mutant alleles. This suggests that transcriptional activation of this subset of loci equally requires MAIN, MAIL1 and PP7L activity, and possibly the recognition of the 'DOWN' DNA motif.

PP7L is not required for heterochromatin condensation

WT *Arabidopsis* nuclei at interphase exhibit condensed DNA foci called chromocenters that are composed of constitutive heterochromatin, and are enriched in H3K9me2 [27]. In several epigenetic mutants, decondensation of constitutive heterochromatin correlates with disruption of chromocenters, and loss or diffusion of H3K9me2 in the nucleoplasm [27]. Thus, analyzing H3K9me2 subnuclear distribution by immunofluorescence (IF) experiments has been reproducibly used as a cytological approach to assay for heterochromatin decondensation [12, 27, 28]. A previous study showed that subnuclear distributions of chromocenters and H3K9me2 were unchanged in *main-2* and *mail1-1* mutants [15]. However, fluorescent in situ hybridization (FISH) experiments using a DNA probe for the 106B pericentromeric repeats suggested that heterochromatin was decondensed in the *main-2* and *mail1-1* in comparison to WT plants [15]. We performed IF experiments to analyze the subnuclear distribution of H3K9me2 in the *pp7l-2* mutant. These analyses did not show any change in the condensation level of chromocenters in *pp7l-2* nuclei in comparison to WT (Fig 7). Instead, we observed that *pp7l-2* nuclei were proportionally more condensed than WT nuclei (Fig 7). This is likely due to the fact that *pp7l-2* mutant displays abnormal phenotype and growth delay in comparison to WT plants that are entering the floral transition stage, a developmental stage where partial chromocenter decondensation has been documented [29]. In conclusion, based on the H3K9me2 IF experiments, we can conclude that *pp7l-2* is not impaired in chromocenter condensation.

The PMD and PP7 domains have co-evolved among the Eudicots

Among the Angiosperms, most of the genic PMDs, like MAIN and MAIL1, are standalone versions [18]. However, some genic PMDs can associate with other protein domains, such as for instance a PPP domain. In *A. thaliana*, the protein MAIL3, which carries a PMD fused to a PPP domain, is a close homolog of both MAIN/MAIL1 and PP7/PP7L through its PMD and PPP domains, respectively. Considering that the PMD proteins MAIN and MAIL1 interact with PP7L, and are required for the expression of similar set of loci, we decided to determine the distribution of related genic PMD and PPP domains, and to retrace their evolutionary history among plant species. The *A. thaliana* MAIN, MAIL1 and MAIL3 are all members of the PMD-C family that also includes MAIL2 [15]. Since our objective is to retrace the evolution of genic (and not TE-containing) PMD-C, we have decided to retrace our search to Eudicots. Indeed, Eudicot species contain mainly genic PMD-C, while other angiosperms may contain variable numbers of closely related genic and TE-associated PMD-C motifs that would be difficult to distinguish in our analysis. To retrace the evolution history of the genic PMD-C family, we used *A. thaliana* PMD-C genes to search and collect their relatives (paralogues and

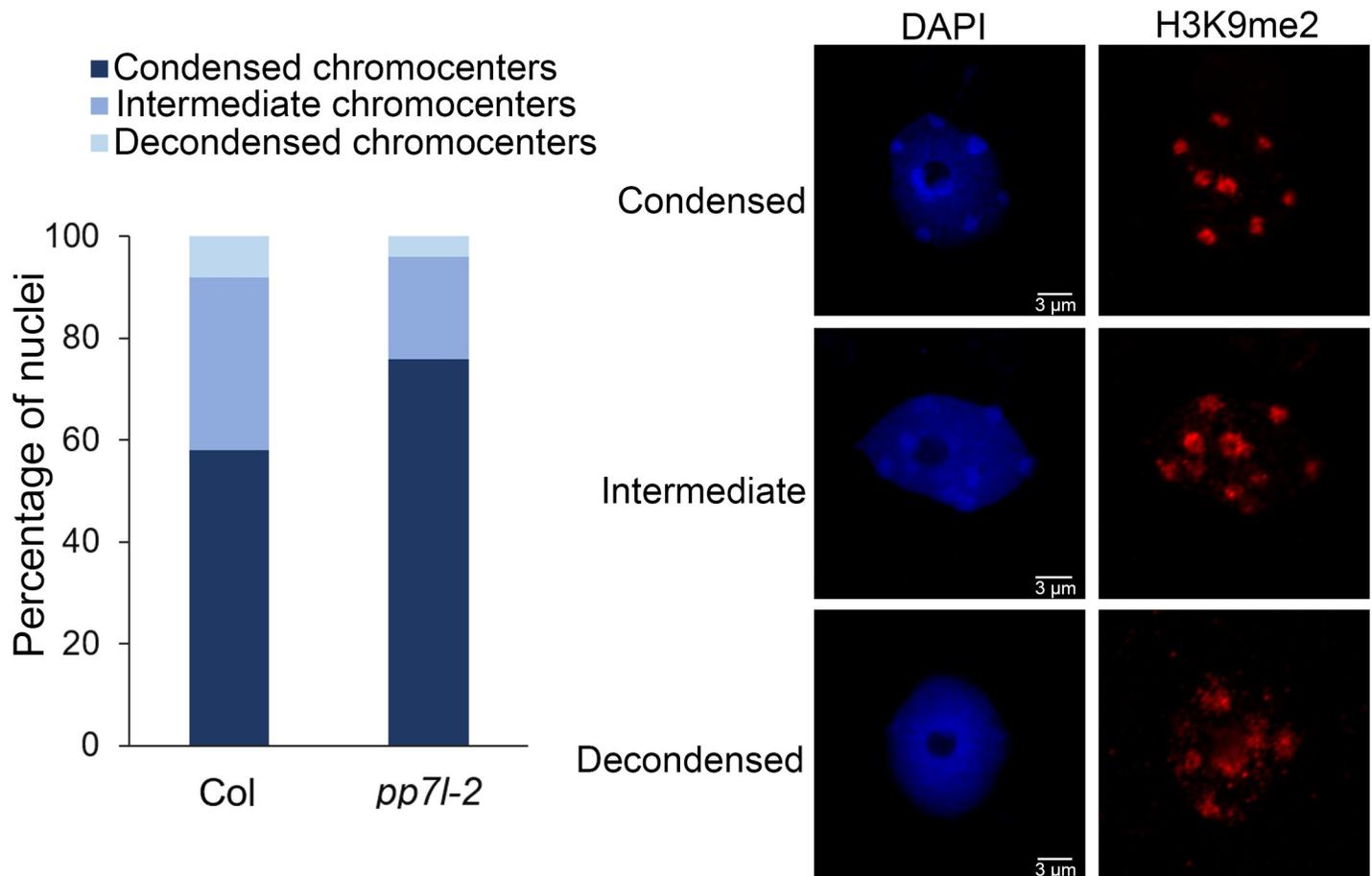


Fig 7. Constitutive heterochromatin appears unaltered in *pp7l-2* mutant. Proportion of nuclei showing condensed, partially decondensed (intermediate), or decondensed chromocenters in the *pp7l-2* mutant in comparison to WT control (Col) based on H3K9me2 immunostaining of nuclei. Representative pictures of nuclei displaying condensed, partially decondensed or decondensed chromocenters. DAPI: DNA stained with 4',6-diamidino-2-phenylindole.

<https://doi.org/10.1371/journal.pgen.1008324.g007>

orthologues) in 30 genomes representative of the Eudicot diversity (see [S8 Table](#) for a list of species and their corresponding codes used in [Fig 8](#) and [S9 Table](#) for motif sequences).

In our phylogenetic analysis, the genic PMD-C family can be clearly separated in two major clades. The first clade is composed of orthologues of *A. thaliana* MAIL2, MAIL1 and MAIN, while the second one includes orthologues of *A. thaliana* MAIL3 ([Fig 8A](#)). MAIL2 orthologues were found in all species tested, forming a closely related group, which suggests that they are under strong purifying selection (see the very short branch lengths linking most MAIL2 genes in [Fig 8A](#)). In several species, additional MAIL2 paralogues were also detected. They were either imbedded in the major MAIL2 group, or forming independent and more divergent subgroups, like in the case of MAIL1 and MAIN that are Brassicaceae-specific MAIL2 paralogues. By comparison, MAIL3 orthologues were not found in all Eudicot species tested, and, except in Brassicaceae, MAIL3 genes appear to be under much weaker purifying selection compare to MAIL2 and MAIL2-like genes (see the longer branch lengths in the tree of [Fig 8A](#)). Brassicaceae MAIL3 genes contrast with other MAIL3, by forming a closely related group in the phylogenetic tree. This suggests a clear change in selection pressure, typical of a neofunctionalization event that could correlate with the acquisition of the PPP motif by these genes ([Fig 8B](#) and see below). Remarkably, another fusion event between PMD-C and PPP motifs

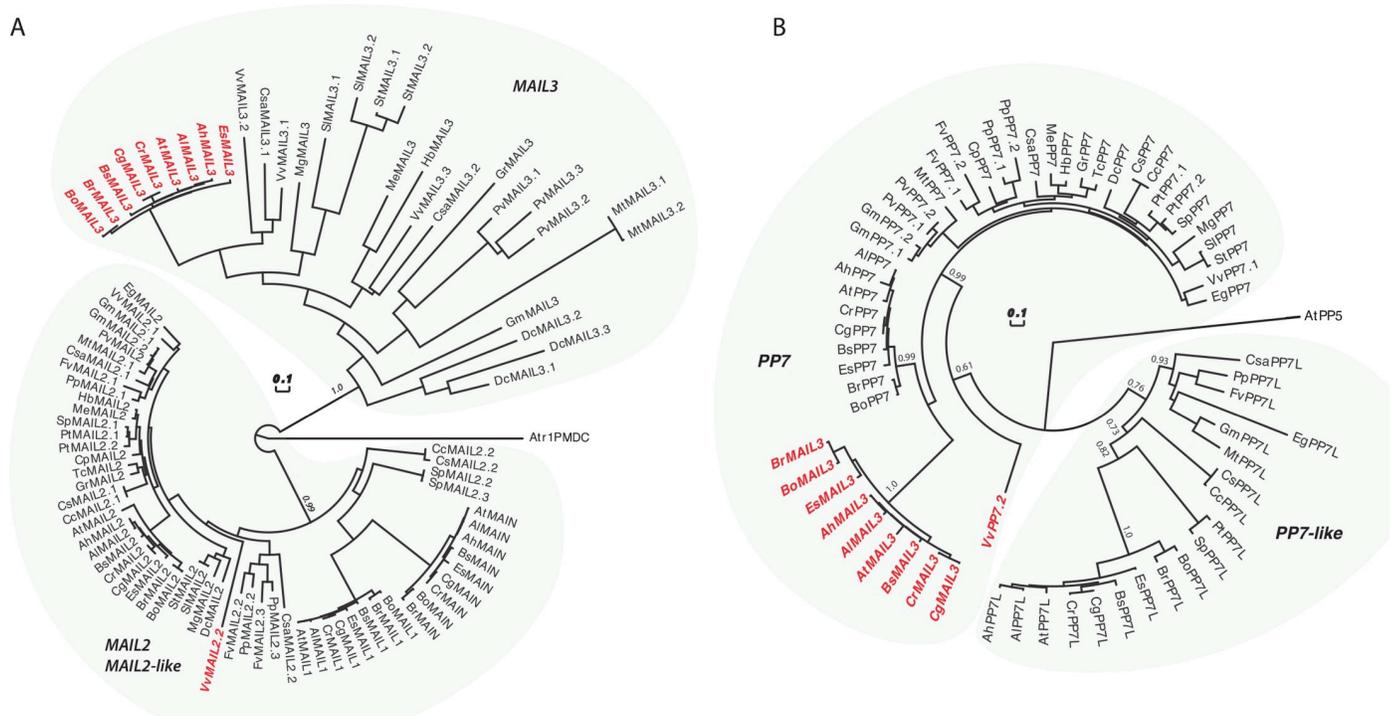


Fig 8. Evolutionary history of PMD-C and PP7 proteins in plants. (A) An alignment of the PMD-C motifs from 30 representative Eudicot species was used to construct a phylogenetic tree. The two major clades (MAIL2/MAIL2-like and MAIL3) are indicated. The species codes are given in [S11 Table](#), and corresponding protein sequences in [S12 Table](#). In red are genes presenting a fusion between a PMD-C and a PP7 motif. Statistical supports of key nodes calculated with the approximate likelihood-ratio test are indicated. Scale bar indicates one substitution/site. The tree was rooted using the *Amborella trichopoda* PMD-C motif (Atr1PMDC). (B) Phylogenetic tree constructed using an alignment of the PP7 motif from the same species as in (A). The two major clades (PP7 and PP7L) are indicated. In red are genes presenting a fusion between a PP7 and a PMD-C motif. Statistical supports of key nodes calculated with the approximate likelihood-ratio test are indicated. Scale bar indicates one substitution/site. The tree was rooted using the *A. thaliana* PP5 motif (AtPP5).

<https://doi.org/10.1371/journal.pgen.1008324.g008>

occurred independently in grapevine, but this time involving a MAIL2 paralogue (VvMAIL2.2, [Fig 8A](#)).

We then used the PPP motif found in *A. thaliana* MAIL3, to collect orthologous genes and retrace the evolution history of this motif in the same Eudicot species used above. We confirmed that these genes can be clearly separated in two distinct clades: PP7 and PP7-like (PP7L) ([Fig 8B](#)). All tested species present one or several closely related PP7 paralogues. Although the Brassicaceae MAIL3 PPP motif belongs to the PP7 clade, it diverged significantly compared to other standalone PP7 paralogues ([Fig 8B](#)). Same observation was made regarding the PP7 domain of VvMAIL2.2. Thus, as described for the PMD of Brassicaceae MAIL3 and grapevine VvMAIL2.2, this suggests a fast-evolving period and neofunctionalization of the PP7 domain in these species, subsequently to the PMD-C/PP7 fusion. Conversely, PP7L orthologues were not found in all species tested and, accordingly, these genes are under weaker purifying selection compare to genes belonging to the PP7 subfamily. In conclusion, phylogenetic analyses showed that, in at least Brassicaceae and grapevine, neo-association of PMD-C and PP7 domains have potentially create new protein functions that were maintained through evolution.

Discussion

In *A. thaliana*, MAIN and MAIL1 are standalone PMD proteins that have been involved in genome integrity, regulation of cell division and differentiation, and silencing of TEs [[15–17](#)].

In this study, we show that TE silencing is widely impaired in the *ddc main-3* higher order mutant, which is both partially defective in DNA methylation and MAIN activity. We also identify the putative phosphatase protein PP7L as MAIN and MAIL1 protein interactor, and show that among the loci that are commonly misregulated in *pmd* and *pp7l* single and double mutants, a substantial fraction of downregulated genes carries the ‘DOWN’ DNA motif in their promoter. Finally, phylogenetic analyses among Eudicots suggest a mechanism of neo-functionalization between the PMD and PP7-type PPP, to potentially acquire a functional module that requires the two protein domains.

The PMD MAIN protein acts independently of DRM2- and CMT3 pathways to silence TEs and DNA-methylated genes

Previous analyses showed that some TEs were synergistically upregulated in the *mail1 rdr2* double mutant plants, suggesting that MAIL1 acts independently of RdDM pathway [15]. In our whole genome transcriptomic analyses, we show that several TEs and DNA-methylated genes are upregulated in both *main-3* and *ddc* mutants, as well as in the *ddc main-3* quadruple mutant (Fig 2 and S2 Fig). We also identify TEs that are upregulated in either *ddc* or *main-3* mutants, but display stronger misregulation in the *ddc main-3* higher order mutant (Fig 2 and S2 Fig). Finally, we identify a large class of TEs that are only upregulated in *ddc main-3* (Fig 2 and S2 Fig). Altogether, these analyses reveal complex genetic interaction between the MAIN, DRM2 and CMT3 proteins to silence TE. Previous work showed that DNA methylation is not impaired in *mail1-1* [15]. We found that DNA methylation is mostly unaffected in the *main-2* null mutant. However, we detected a mild but significant hypermethylation at non-CG sites in TEs and pericentromeric genes (Fig 4). One hypothesis is that CHG and CHH hypermethylation observed in *main-2* is a backup mechanism to compensate for MAIN loss of function, and to dampen TE silencing defects. Although further studies will be required to test this hypothesis, it is consistent with the fact that combining the *main-3* and *ddc* mutations leads to an exacerbation of TE silencing defects. Thus MAIN, DRM2 and CMT3 pathways cooperate to silence TE. Synergistic effects between different epigenetic pathways have already been described. For instance, it has been shown that MORPHEUS MOLECULE 1 (MOM1) and MORC1/MORC6 proteins, or MOM1 and the RdDM pathway act synergistically to efficiently silence TEs [13, 30]. Altogether, these observations contribute to the “mille-feuille” (i.e. “multiple layers”) model, in which different epigenetic pathways converge towards the silencing of TEs [31].

The putative phosphatase PP7L interacts with the PMD MAIN and MAIL1 protein to regulate a similar set of genes and TEs

Recently, the putative phosphoprotein phosphatase PP7L was involved in the biogenesis of chloroplasts and plant response upon abiotic stress [25]. Here, we show that PP7L interact with MAIN and MAIL1, and *main-2*, *pp7l-2*, *mail1-1* single and *mail1-1 pp7l-2* double mutant plants display similar developmental and molecular phenotypes (Figs 5 and 6). We also show that, as described for *main-2* and *mail1-1* [15], the subnuclear distribution of chromocenters and H3K9me2 are unaltered in *pp7l-2* (Fig 7). The 106B pericentromeric repeats appeared decondensed in *main-2* and *mail1-1* mutants [15], future work will determine if similar phenotype is observed in *pp7l-2*. Although MAIN, MAIL1 and PP7L interact together, we cannot exclude that an additional protein is required for the interaction. In addition, PP7L may have additional partners independently of MAIN and MAIL1. Further biochemical studies such as IP-MS analyses using the FLAG-tagged PP7L line will contribute to addressing these points.

Transcriptomic analyses revealed complex genetic interaction between MAIN, MAIL1 and PP7L; the three proteins acting either independently or together to ensure the proper expression of genes, and to perform TE silencing. Moreover, transcriptome profiling of *mail1-1 pp7l-2* double mutant revealed that the two mutations may have synergistic effects, specifically at genes that are upregulated in the mutant. To further study the genetic interaction between the three proteins, it will be important to analyze the transcriptome of *main-2 mail1-1 pp7l-2* triple mutant. Altogether and considering that *i*) MAIN, DRM2 and CMT3 pathways cooperate to silence TEs, and *ii*) the *main-2* mutant show a slight increase in DNA methylation at CHG and CHH sites, we cannot rule out that MAIN is playing a dual role: regulating gene expression through its interaction with MAIL1 and PP7L, and involved in TE silencing through its genetic interaction with DNA methylation. In the future, it will be important to analyze DNA methylation in *pp7l-2*, but also in *pmd pp7l-2* higher order mutants. In parallel, studying the *ddc pp7l-2* mutant will allow to further decipher the genetic interaction between the PP7L and DNA methylation pathways.

A fraction of genes that are commonly downregulated in *main*, *mail1* and *pp7l* mutants carry the ‘DOWN’ motif in their promoters

A substantial fraction of genes that are commonly downregulated in *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2* carry the ‘DOWN’ motif in their promoter (S5E Fig and S7 Table). Furthermore, twenty-five out of twenty-six genes commonly downregulated in all the mutant backgrounds analyzed in this study—except *ddc*—carry the ‘DOWN’ DNA motif in their promoter (S5E Fig and S7 Table). The ‘DOWN’ motif is also enriched in fractions of downregulated genes in *main-2*, *mail1-1*, *mail1-1 pp7l-2*, *main-3* and *ddc main-3*. However, it is not enriched among downregulated genes in *pp7l-2* mutant. One explanation for this discrepancy is that too many loci were identified as downregulated in *pp7l-2*, which created a dilution of the loci carrying the ‘DOWN’ motif in their promoter.

Based on our results, we hypothesize that the ‘DOWN’ motif may act as a putative cis-regulatory element (CRE) recognized by an unidentified TF, which would be required for the transcription of genes identified as downregulated in *pmd* and *pp7l* mutants. This unknown TF could be recruited or activated by the MAIN/MAIL1/PP7L protein complex. Another hypothesis is that the ‘DOWN’ motif is directly recognized by the MAIN/MAIL1/PP7L protein complex. Further study will be required to test if MAIN/MAIL1/PP7L protein complex interact with chromatin, and bind the ‘DOWN’ motif. In parallel, further biochemical analyses may allow to identify an uncharacterized putative TF as MAIN/MAIL1/PP7L protein interactor.

Altogether, these analyses suggest that MAIN, MAIL1 and PP7L are involved in three distinct activities. First, they are required for the silencing of TEs and DNA-methylated genes, cooperating with canonical epigenetic factors such as DRM2 and CMT3 to efficiently repress these loci. Second, they are required for the repression of subsets of genes that are not targeted by DNA methylation. For this category of loci, one hypothesis is that MAIN, MAIL1 and PP7L may act as transcriptional repressor. Third, MAIN, MAIL1 and PP7L are required for the transcriptional activation of several genes, and fractions of those genes carry the ‘DOWN’ motif in their promoter. In the future, it will be important to determine the molecular mechanisms that are involved in these three activities of MAIN, MAIL1 and PP7L.

The association of PMD-C and PP7/PP7L domains creates a functional protein module

In this study, we identified PP7L as a protein partner of the two standalone PMDs MAIN and MAIL1, and showed that these proteins are required for the proper expression of a common

set of genes, and for TE silencing. Besides, we showed that the Brassicaceae MAIL3 and the grapevine VvMAIL2.2 proteins carry a PMD fused to a PP7 domain. Based on these results, we hypothesize that depending on the configuration, the association of PMD-C and PP7/PP7L domains would create a functional protein module in trans or in cis. It is likely that the cis-association of PMD and PP7 found in the Brassicaceae MAIL3 proteins occurred in the common ancestors of this Eudicot lineage, possibly through the process of gene duplication. Since then, the MAIL3 PMD/PP7 fusion was maintained under strong purifying selection, arguing for a neofunctionalization of the fusion protein. It is likely that a similar process happened in grapevine, and possibly, in closely related Vitaceae species. To some extent, the two distinct events that occurred in Brassicaceae and grapevine are reminiscent of convergent evolution processes leading to the production of a functional PMD/PP7 module.

The occurrence of PMD and PP7/PP7L protein fusion in several Brassicaceae and grapevine is reminiscent of the concept of Rosetta stone chimera proteins, which describes that two proteins interacting together in one organism can be found fused together in another species to facilitate enzymatic activity [32]. There are several examples of Rosetta stone proteins, described for instance with different subunits of DNA topoisomerase or RNA polymerase [32]. Here, we show that, at least in *A. thaliana*, the Rosetta stone chimera MAIL3 coexist with its close homologs MAIN/MAIL1 and PP7L that interact together. The fact that the PMD and PP7 domains are fused together in MAIL3 may be a strategy to optimize protein activity. Conversely, the enzymatic activity of the MAIN/MAIL1/PP7L protein complex could be further regulated by allowing, or not, the three proteins to interact together. Nevertheless, in both scenarios, it is likely that PMD and PP7/PP7L association creates a functional protein module, which might be specialized in distinct biological processes depending on its composition. Thus, we hypothesize that the MAIL3 and MAIN/MAIL1/PP7L protein complexes play different role in the plant. This is consistent with the fact that, unlike *main-2*, *mail1-1* and *pp7l-2* mutant, the *mail3-2* mutant does not show abnormal developmental phenotype [17]. Further studies will be required to describe the role of MAIL3 in the plants.

In conclusion, we show here that the two *A. thaliana* PMD MAIN and MAIL1 proteins interact with PP7L, and are involved in the regulation of a common set of genes and TEs. In addition, we show that distinct events of PMD-C and PP7 fusions have occurred among the Eudicots (among several Brassicaceae species and in grapevine), suggesting some convergent evolution processes and a potential neofunctionalization of PMD/PP7 module in cis. The biological significance of PMD/PP7 fusion proteins will be investigated in the future by studying the role of MAIL3 in *A. thaliana*. In addition, it will be important to determine whether the PMD proteins play important roles in other plant species with agronomic value.

Materials and methods

Plant material and growing conditions

All the plant material is in the Columbia (Col) ecotype. Col = Non-transgenic WT Columbia ecotype. The *drm1-2* (SALK_031705), *drm2-2* (SALK_150863), *cmt3-11* (SALK_148381), *ddc* triple, *main-2* (GK-728H05), *mail1-1* (GK-840E05) and *pp7l-2* (SALK_003071) null mutant lines were previously described [15–17, 25, 26], and obtained from The Nottingham Arabidopsis Stock Centre. The *mail1-1 pp7l-2* double mutant was obtained by crossing the respective single mutants. T-DNA insertions were confirmed by PCR-based genotyping and RT-qPCR analyses. The *ATCOPIA28::GFP* WT line (WT) carries the transgene in WT Col ecotype. The *ATCOPIA28::GFP ddc* line (*ddc*) carries the transgene in *ddc*. The *ATCOPIA28::GFP ddc main-3* line (*ddc main-3* = *ddc* #16) carries the transgene in the *ddc main-3* background. The *ATCOPIA28::GFP main-3* line (*main-3*) was obtained by backcrossing *ddc main-3* with WT,

F1 plants were self-fertilized, and F2 plants were screened by PCR-based genotyping to identify plants homozygote for the *main-3* mutation and WT for *DRM2* and *CMT3*. The *main-3* mutant allele was scored by derived cleaved amplified polymorphic sequences (dCAPS) using the restriction enzyme FokI. Primer sequences are described in [S10 Table](#). All the WT Col and T-DNA mutant plants were grown on soil under a 16h-light/8h-dark cycle. When experiments required to screen for GFP expression under UV light, plants carrying the *ATCOPIA28::GFP* transgene were first grown on Murashige and Skoog (MS) plates under continuous light, 10-day old plants were then screened for GFP expression under UV light, and subsequently transferred onto soil. For *in vitro* plant culture, seeds were surface-sterilized and sowed on solid MS medium containing 0.5% sucrose (w/v).

Cloning of *ATCOPIA28::GFP*

The pCambia3300-NLS-GFP-T35S vector was previously described [12]. The 5'LTR promoter corresponding to a region of ~1 kb upstream of *ATCOPIA28* (*AT3TE51900*) was PCR amplified from WT genomic DNA, and cloned into pCR2.1 TOPO vector (Invitrogen). Quikchange site-directed mutagenesis (Stratagene) was performed according to Manufacturer's instruction to create a polymorphism site (MfeI→NdeI) within the 5'LTR promoter, which was subsequently mobilized into pCambia3300 upstream of NLS-GFP-T35S sequence. *ddc* triple mutant plants were transformed with the *ATCOPIA28::GFP* construct using the *Agrobacterium*-mediated floral dip method [33]. Transgenic plants showing GFP fluorescence were backcrossed with a WT plant to promote the silencing of *ATCOPIA28::GFP* in the F1 generation. F1 plants were self-crossed and their F2 progenies were screened for GFP fluorescence, and PCR-based genotyped to obtain *ATCOPIA28::GFP* WT and *ATCOPIA28::GFP ddc* plants. Primer sequences used for *ATCOPIA28::GFP* cloning and PCR genotyping are described in [S10 Table](#).

EMS mutagenesis, GFP screening and mapping analyses

Five thousand seeds of *ATCOPIA28::GFP ddc* were mutagenized in 0.26% EMS solution for 12 hours with rotation. Seeds were subsequently washed with water and sown on soil. Fifteen hundred M2 populations were collected, and subsequently screened for GFP fluorescence under UV light using a SMZ18 Nikon Fluorescence Stereomicroscope coupled with the C-HGFI intensilight fluorescence filter. Pictures were taken using the DS Qi1MC digital camera kit. Mapping and identification of the EMS mutation responsible for the phenotype were performed by bulk segregant analysis coupled with deep genome re-sequencing as previously described [12], with the following differences. Reads were mapped against the reference genome (*Arabidopsis* TAIR10) and single nucleotide polymorphisms called in Geneious (Biomatters). Using R, single nucleotide polymorphisms were filtered for EMS mutations (G: C→A:T) and zygosity called based on the variant frequency provided by Geneious (≥80% homozygous mutation, ≥45%, and ≤55% heterozygous mutation). Plots were then created by calculating the ratio of the number of homozygous and heterozygous and mutations in a 500-kb window as previously described [34].

Cloning of epitope-tagged versions of PMD and PP7L proteins

MAIN, *MAIL1* and *PP7L* genomic regions were PCR amplified and FLAG or MYC epitopes were added to the C-terminus of each protein as previously described [12]. Each time, the amplified region includes a ~1Kb promoter sequence upstream of the respective transcriptional start site. Within the *MAIN* promoter, a MluI site was modified to allow LR reaction without changing the sequence integrity of the gene. *main-2* and *mail1-1* mutant plants were transformed with the *MAIN-FLAG* or *MAIN-MYC* and *MAIL1-FLAG* or *MAIL1-MYC*

constructs, respectively, using the *Agrobacterium*-mediated floral dip method [33]. *MAIN-MYC* and *MAIL1-MYC* lines were subsequently supertransformed with the *PP7L-FLAG* construct to perform co-IP experiments. Primer sequences are described in [S10 Table](#). The *PP7L* DNA and protein sequences used in this study are described in [S12 Table](#).

IP and MS analysis

Ten grams of 3-week-old seedling tissue were ground in liquid nitrogen and resuspended in 50mL ice-cold IP buffer [50mM Tris HCl pH 7.6, 150mM NaCl, 5mM MgCl₂, 0.1% Nonidet P-40, 10% glycerol (v/v), 0.5mM DTT, 1x Protease Inhibitor Mixture (Roche)] and centrifuged 2 times for 15 min at 4°C at 15 350g. 400μL of M2 magnetic FLAG-beads (Sigma, M8823) were added to the supernatants, and incubated for 90 min rotating at 4°C. M2 magnetic FLAG-beads were washed seven times in ice-cold IP buffer for 5 min rotating at 4°C, and immunoprecipitated proteins were eluted 3 times with 150μL 3x-FLAG peptides (Sigma, F4799) for 25 min each at 25°C. The eluted protein complexes were precipitated by trichloroacetic acid and subjected to MS analyses as previously described [13]. Peptide and protein-level false discovery rates were calculated by the DTASelect algorithm using the decoy database approach. Based on a peptide PSM level p-value filter of less than 0.01 and a requirement for at least two peptides per protein, the protein-level false discovery rate was less than 1% for all proteins detected.

Co-IP and immunoblotting

0.5 g of 3-week-old seedling tissue were ground in liquid nitrogen, resuspended in 1.5mL ice-cold IP buffer [50mM Tris pH 7.6, 150mM NaCl, 5mM MgCl₂, 0.1% Nonidet P-40, 10% glycerol, 0.5 mM DTT, 1x Protease Inhibitor Mixture (Roche)], and centrifuged 2 times for 15 min at 4°C, 16 000g. 50μL M2 magnetic FLAG-beads (Sigma, M8823) were added to the supernatants and incubated for 2 hour rotating at 4°C. Beads were washed 3 times in ice-cold IP buffer for 10 min rotating at 4°C. Immunoprecipitated proteins were denatured in Laemmli buffer for 5min at 95°C. 10μL of input and bead elution were run on 10% SDS-PAGE gels, and proteins were detected by western blotting using either Anti-FLAG M2 monoclonal antibody-peroxidase conjugate (Sigma, A8592) at a dilution of 1:10000, or c-Myc rat monoclonal antibody (Chromotek, 9E1-100) at a dilution of 1:1000 followed by goat anti-rat IgG horseradish peroxidase (Abcam, ab205720) used at a dilution of 1:20000 as secondary antibody. Western blots were developed using Substrat HRP Immobilon Western (Merck Millipore, WBKLS0500).

RNA extraction

Total RNA was extracted from aerial parts of 3-week-old seedlings grown on soil using either RNeasy Plant Mini Kit (Qiagen, 74904) or Monarch Total RNA Miniprep Kit (NEB, T2010) according to the manufacturer's protocols.

RNA sequencing

RNA-seq libraries were generated from 1μg of input RNA using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, E7490) according to the manufacturer's protocols. Libraries were sequenced on an Illumina HiSeq 4000 or NextSeq 550 machines. Reads were trimmed using Trimmomatic [35], and mapped to the *A. thaliana* genome (*Arabidopsis* TAIR10 genome) using HISAT2 [36]. The sequence alignment files were sorted by name and indexed using SAMtools [37]. Files were converted to BAM files and number of reads mapped onto a gene calculated using HTSeq-count [38]. Differentially expressed genes were obtained with DESeq2 [39], using a log₂ fold-change ≥ 2 (up-regulated genes) or ≤ -2 (down-regulated

genes) with an adjusted p -value of 0.01. Batch effects were modeled within the DESeq2 study design. For PCA, we removed the batch effect using limma's 'removeBatchEffect' function [40]. Heat map visualizations were realized using the heatmap2 function from the R gplots package. Boxplots were realized using boxplot function from R. Re-analyses of previously published RNA-seq datasets from *main-2* and *mail1-1* (PRJEB15202) [15] were performed as described above.

RT-qPCR

1 μ g of input RNA was converted to cDNA using GoScript Reverse Transcriptase (Promega A501C) according to the manufacturer's protocol. The final reaction was diluted 6 times with RNase free water. RT-qPCR experiments were performed with 4 μ L of cDNA combined to the Takyon No Rox SYBR MasterMix (Eurogentec, UF-NSMT-B0701), using a LightCycler 480 instrument (Roche). Amplification conditions were as follows: 95°C 5 min; 45 cycles, 95°C 15s, 60°C 15s, 72°C 30s; melting curves. RT-qPCR analyses used the $2^{-\Delta\Delta C_t}$ method. For each analysis, ΔC_t was first calculated based on the housekeeping *RHIP1* gene Ct value [41]. $\Delta\Delta C_t$ were then obtained by subtracting the wt ΔC_t from the ΔC_t of each sample. Values were represented on bar charts relative to WT. Three technical replicates were performed per biological replicate, and 3 biological replicates were used in all experiments, unless otherwise stated. Primer sequences are described in S10 Table. RT-qPCR raw data are described in S13 Table.

DNA motif detection

The motifs for enhancer sequences (1kb upstream the TSS) were discovered using MEME (Multiple Em for Motif Elicitation). MEME represents motifs as position-dependent letter-probability matrices which describe the probability of each possible letter at each position in the pattern [42].

Bisulfite sequencing

Genomic DNA was extracted from aerial parts of 3-week-old seedlings using Quick-DNA Plant/Seed Miniprep Kit (Zymo research, D6020) according to the manufacturer's protocol. Whole genome bisulfite sequencing (WGBS) library was prepared from 50 ng genomic DNA using NuGen Ovation Ultralow Methyl-Seq kit. Bisulfite treatment was carried out by Qiagen Epitect bisulfite kit. WGBS libraries were sequenced on an Illumina HiSeq 4000 machine. The raw reads (single end) were trimmed using Trimmomatic in order to remove adapter sequences [35]. The remaining sequences were aligned against the *A. thaliana* genome TAIR10 version using Bismark [43]. Duplicated reads were collapsed into one read. For meta-plot and boxplot visualization, we used ViewBS [44]. Boxplots were realized using boxplot function from R. DMRs (differentially methylated regions) were defined comparing methylation in wildtype with the *main-2* mutant analyzed using the R package "DMRcaller" [45]. We used "noise filter" method to compute CpG, CpHpG and CpHpH DMRs. We selected bins where the p -value was less than 0.01, the difference in methylation level was at least 40% in the CG context, 20% in the CHG context or 10% in the CHH context, with at least four cytosines; each cytosine had on average at least four reads.

Sequence selection, multiple sequences alignments and phylogenetic reconstruction

Blast searches (blastp) were performed starting from known *A. thaliana* PMD-C and PP7/PP7L motifs on the thirty species representing the diversity of the Eudicot lineages. When

necessary tblastn searches were also used to obtain complete protein sequences. To build the phylogenetic trees, PMD-C or PP7/PP7L motifs were aligned using the multiple sequence comparison by log-expectation (MUSCLE v3.7) software [46]. Trees were reconstructed using the fast-maximum likelihood tree estimation program PHYML [47] using the LG amino acids replacement matrix [48]. Statistical support for the major clusters were obtained using the approximate likelihood-ratio test (aLRT) [49].

Immunofluorescence and DAPI-staining

Leaves from 3-week-old plants, were fixed for 20 min rotating at 4°C in 2% formaldehyde in Tris buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 100 mM NaCl), washed two times for 10 min rotating at 4°C in cold Tris buffer and subsequently chopped in LB01 buffer (15 mM Tris-HCl pH 7.5, 2 mM EDTA, 0.5 mM spermine, 80 mM KCl, 20mM NaCl and 0.1% Triton- X-100). Nuclei were filtered through a 30 µm cell strainer cap (Sysmex, 04-0042-2316) and 5µl of the nuclei solution was diluted in 10 µl of sorting buffer (100mM Tris-HCl pH 7.5, 50 mM KCl, 2 mM MgCl₂, 0.05% Tween-20 and 5% sucrose). 20µl of the nuclei dilution were spread onto a polylysine slide and air-dried for 40 min. Slides were post-fixed in 2% formaldehyde in 1X PBS for 5 min and washed 2 times with water. Slides were incubated 15 min in 1X PBS, 0.5% Triton X-100 at RT and washed 3 times with 1X PBS for 5 min. For detection, slides were incubated over night with a mouse anti-H3K9me2 monoclonal antibody (Abcam, Ab 1220) at 1:500 in 3% BSA, 0.05% Tween in 1X PBS at 4°C in a moist chamber. After 3 washes in 1X PBS for 5 min, slides were incubated 2h with a goat anti-mouse antibody coupled to Alexa fluor 568 (Invitrogen, A11004) at 1:1000 in 3% BSA, 0.05% Tween in 1X PBS in a moist chamber. Slides were washed 1 time 5 min with 1X PBS, 1 time 10 min with 1X PBS, 1µg/mL DAPI, and 1 time 5 min with 1X PBS. DNA was counterstained with 1µg/mL DAPI in Vectashield mounting medium (Vector Laboratories). Observation and imaging were performed using a LSM 700 epifluorescence microscope (Zeiss).

Supporting information

S1 Fig. MAIN is the mutated gene responsible for *ATCOPIA28::GFP* and TE overexpression in the *ddc* #16 mutant. (A) Representative pictures of *ddc* #18 (*ddc morc6-8*) and *ddc* #344 (*ddc morc6-9*) mutants in comparison to *ATCOPIA28::GFP* WT and *ddc* control plants under UV light. Insets show plants under white light. (B) Enrichment in homozygote/heterozygote ratio of EMS over WT single nucleotide polymorphisms (SNPs), defining the linkage intervals for the populations *ddc* #18 and *ddc* #344. Mb: megabase. Gray-shaded rectangles delimit the mapping intervals. (C) Location of the point mutations corresponding to the *morc6-8* and *morc6-9* alleles within the *MORC6* genomic sequence. Nucleotide and corresponding amino acid changes are indicated above the gene. Positions of the mutations are indicated relative to the transcription start site (+1). Grey boxes represent 5' and 3' UTR, blue boxes and lines represent exons and introns, respectively. (D) Enrichment in homozygote/heterozygote ratio of EMS over WT single nucleotide polymorphisms (SNPs), defining the linkage intervals for the population *ddc* #16. Gray-shaded rectangle delimits the mapping interval. (E) Location of the point mutation corresponding to the *main-3* mutant allele within the *MAIN* genomic sequence. (F) Genetic complementation analyses using the KO T-DNA insertion line *main-2*. *ddc* #16 plants were crossed with *main-2* plants. F1 plants were self-crossed, and F2 plants were screened under UV light to select GFP-overexpressing plants. Western blotting using anti-GFP antibodies confirmed GFP overexpression in selected F2 plants. Coomassie staining of the large Rubisco subunit (*rbcL*) is used as a loading control. KDa: kilodalton. Among the selected F2 plants, the presence of *main-3* EMS and *main-2* T-DNA mutant

alleles were determined by dCAPS-PCR and PCR analyses, respectively. *DRM2* and *CMT3* genotyping were determined by PCR analyses. WT: Wild type, Ho: Homozygote mutant. He: Heterozygote. (G) Relative expression analyses of several TEs in the indicated genotypes assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are represented relative to WT. Error bars indicate standard deviation based on two independent biological replicates. Screening of EMS mutant populations was done on MS plates to allow for visualization of GFP-positive individuals under UV light.

(TIF)

S2 Fig. Combining the *drm2*, *cmt3* and *main-3* mutations exacerbate TE silencing defects.

(A) Principal component analysis (PCA) performed after batch correction for first two components of the sixteen samples described in RNA-seq EMS Exp1 and Exp2. (B) Relative expression analyses of *ATCOPIA28* and *HELITRONY1D* (*AT5TE35950*) in *ddc*, *main-3* and *ddc main-3* assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are represented relative to WT. Error bars indicate standard deviation based on three independent biological replicates. (C) Venn diagrams analysis showing the overlaps between reproducibly upregulated TEs in *ddc*, *main-3* and *ddc main-3*. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value <2.2.10e-16). (D) Same as panel B for TEs defined as class I-IV TEs. Frames of RT-qPCR graphs are using the same color code as shown in panel C. (E) Venn diagrams analyses defining the overlaps between up- and downregulated genes in the different genotypes. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value <2.2.10e-16). (F) Fraction of misregulated genes in *ddc*, *main-3* and *ddc main-3* located in chromosome arms or in pericentromeric regions as defined in [50]. Asterisks indicate statistically significant enrichments of misregulated genes in chromosome arms or pericentromeric regions in comparison to the genomic distributions of all *A. thaliana* genes (Chi-Square test, **: p-value ≤ 0.01). Percentages of genes targeted by DNA methylation and H3K9me2 were calculated based on enrichment in heterochromatin states 8 and 9 as defined in [51]. (G) Relative expression analyses of *DRM2* and *CMT3* in *ddc*, *main-3*, *ddc main-3*, *cmt3 main-3* and *dd main-3* assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are represented relative to WT. Error bars indicate standard deviation based on three independent biological replicates. Screening of EMS mutant populations was done on MS plates to allow for visualization of GFP-positive individuals under UV light.

(TIF)

S3 Fig. Identification of reproducibly misregulated loci in *main-2*, *mail1-1* and *main-3*. (A)

Principal component analysis (PCA) performed after batch correction for first two components of the twenty-four *main-2*, *mail1-1* and WT Col samples described in RNA-seq Exp1, Exp2 and Exp3. (B-D) Relative expression analyses of several upregulated TEs (B), upregulated genes (C), and downregulated genes (D) in *main-2*, *mail1-1* and *main-3* assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are represented relative to respective WT controls. Error bars indicate standard deviation based on three independent biological replicates. (E) Venn diagrams analyses representing the overlaps between misregulated loci in *main-2*, *mail1-1*, *ddc* and *ddc main-3*. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value <0.005).

(TIF)

S4 Fig. DNA methylation analyses in the *main-2* mutant. (A-B) Boxplot analyses in two *main-2* and WT Col biological replicates showing the DNA methylation levels at genomic sites previously defined as hypo CHG differentially methylated regions (DMR) in *cmt3* (A) and hypo CHH DMR in *drm1 drm2* (B) based on [26]. p-values were calculated using a Wilcoxon test. *: p-value < 5.10e-7, **: p-value < 5.10e-10, ***: p-value < 2.10e-16.

(TIF)

S5 Fig. MAIN, MAIL1 and PP7L are required for the proper expression of similar loci, and commonly downregulated genes carry the ‘DOWN’ DNA motif in their promoter. (A)

Principal component analysis (PCA) performed after batch correction for first two components of the thirty-two samples described in RNA-seq Exp1, Exp2 and Exp3. (B) Number of misregulated genes in the different genotypes in comparison to WT Col plants from RNA-seq Exp3 (four biological replicates, S3 and S6 Tables). (C) Number of upregulated TEs in *pp7l-2* and *mail1-1 pp7l-2*, and classified by TE superfamily. (D) Fraction of misregulated loci in *pp7l-2* and *mail1-1 pp7l-2* located in chromosome arms or in pericentromeric regions as defined in [50]. Asterisks indicate statistically significant enrichments of downregulated genes, upregulated genes and TEs in chromosome arms and pericentromeric regions, respectively, in comparison to the genomic distributions of all *A. thaliana* genes and TEs (Chi-Square test, *: p-value ≤ 0.05, **: p-value ≤ 0.01, n.s: not significant). Percentages of genes targeted by DNA methylation and H3K9me2 were calculated based on enrichment in heterochromatin states 8 and 9 as defined in [51]. (E) Identification and proportions of the ‘DOWN’ DNA motif among the promoters of downregulated genes and all *Arabidopsis* genes using the MEME software. Promoter regions are defined as 1kb upstream of ATG. The list of all *Arabidopsis* genes used to determine genomic distributions is based on the TAIR file: TAIR10_upstream_1000_translation_start_20101028. RNA-seq threshold: $\log_2 \geq 2$, or $\log_2 \leq -2$; p-adj < 0.01.

(TIF)

S6 Fig. Full size images of Western blot panels described in Fig 1C, Fig 5B and Fig 5G–5I.

(TIF)

S1 Table. Lists of differentially expressed loci in *ddc*, *main-3* and *ddc main-3*.

(XLSX)

S2 Table. Lists of differentially expressed loci in *main-2* and *mail1-1*.

(XLSX)

S3 Table. Lists of loci commonly misregulated in *main-2*, *mail1-1* and *main-3*.

(XLSX)

S4 Table. Lists of differentially expressed loci in *pp7l-2* and *mail1-1 pp7l-2*.

(XLSX)

S5 Table. Lists of loci commonly misregulated in *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2*.

(XLSX)

S6 Table. Lists of loci commonly misregulated in all mutant backgrounds (except *ddc*) analyzed in this study.

(XLSX)

S7 Table. Lists of commonly downregulated genes displaying the “DOWN” motif in their promoter and random test analyses.

(XLSX)

S8 Table. List of species used to construct the two trees of Fig 8, their codes and the presence/absence of the different PMD-C and PP7 motifs.

(XLSX)

S9 Table. (A) PMD-C and (B) PP7/PP7L motifs used to construct the two phylogenetic trees of Fig 8.

(XLSX)

S10 Table. List of primers used in this study.

(XLSX)

S11 Table. Next Generation Sequencing (NGS) mapping and coverage statistics.

(XLSX)

S12 Table. Genomic DNA and protein sequences of *PP7L (AT5G10900)* used in this study.

(XLSX)

S13 Table. Raw data of RT-qPCR experiments described in this study.

(XLSX)

Acknowledgments

The authors want to thank Thierry Lagrange, Frederic Pontvianne and other team members for fruitful discussions, and all the LGDP platform members for their outstanding technical assistance and plant care. The authors also thank Remy Merret and Michele Laudie for their precious help as part of the UPVD Bio-environment facility.

Author Contributions

Conceptualization: Guillaume Moissiard.

Data curation: Nathalie Picault, Yasaman Jami-Alahmadi, Suhua Feng, Etienne Bucher, James Wohlschlegel.

Funding acquisition: Guillaume Moissiard.

Investigation: Melody Nicolau, Nathalie Picault, Julie Descombin, Yasaman Jami-Alahmadi, Suhua Feng, Etienne Bucher, Jean-Marc Deragon, James Wohlschlegel, Guillaume Moissiard.

Resources: Steven E. Jacobsen.

Supervision: Guillaume Moissiard.

Validation: Melody Nicolau, Julie Descombin.

Visualization: Melody Nicolau, Nathalie Picault, Etienne Bucher, Jean-Marc Deragon.

Writing – original draft: Jean-Marc Deragon, Guillaume Moissiard.

Writing – review & editing: Melody Nicolau, Nathalie Picault, Julie Descombin, Steven E. Jacobsen, Jean-Marc Deragon, Guillaume Moissiard.

References

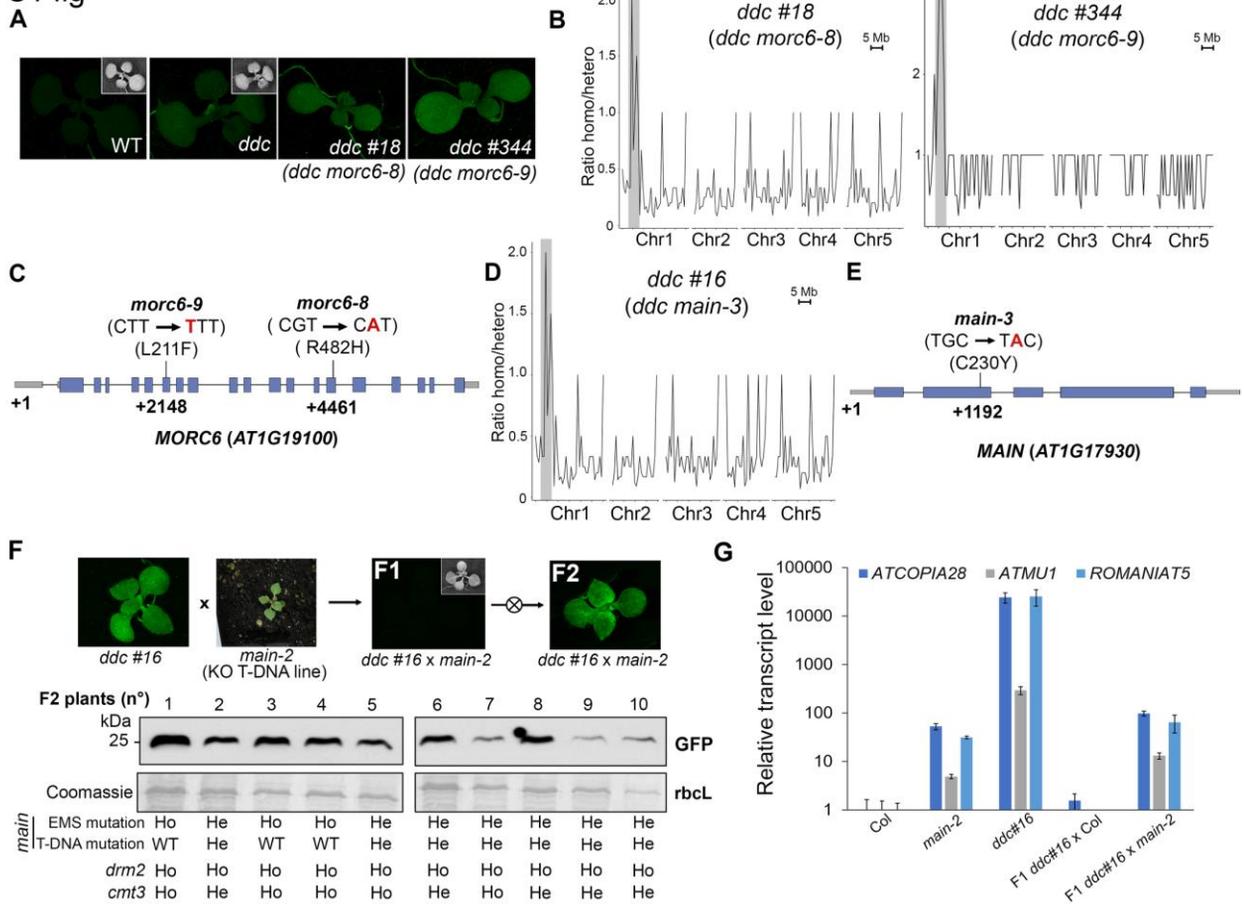
1. Grewal SI, Jia S. Heterochromatin revisited. *Nat Rev Genet.* 2007; 8(1):35–46. <https://doi.org/10.1038/nrg2008> PMID: 17173056
2. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 2007; 8(4):272–85. <https://doi.org/10.1038/nrg2072> PMID: 17363976

3. Deniz O, Frost JM, Branco MR. Regulation of transposable elements by DNA modifications. *Nat Rev Genet.* 2019; 20(7):417–31. <https://doi.org/10.1038/s41576-019-0106-6> PMID: 30867571
4. Dergai O, Hernandez N. How to Recruit the Correct RNA Polymerase? Lessons from snRNA Genes. *Trends Genet.* 2019; 35(6):457–69. <https://doi.org/10.1016/j.tig.2019.04.001> PMID: 31040056
5. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet.* 2010; 11(3):204–20. <https://doi.org/10.1038/nrg2719> PMID: 20142834
6. Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet.* 2014; 15(6):394–408. <https://doi.org/10.1038/nrg3683> PMID: 24805120
7. Wendte JM, Pikaard CS. The RNAs of RNA-directed DNA methylation. *Biochim Biophys Acta.* 2016.
8. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol.* 2015; 16(9):519–32. <https://doi.org/10.1038/nrm4043> PMID: 26296162
9. Zhang H, Lang Z, Zhu JK. Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol.* 2018; 19(8):489–506. <https://doi.org/10.1038/s41580-018-0016-z> PMID: 29784956
10. Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell.* 2013; 153(1):193–205. <https://doi.org/10.1016/j.cell.2013.02.033> PMID: 23540698
11. Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol.* 2014; 21(1):64–72. <https://doi.org/10.1038/nsmb.2735> PMID: 24336224
12. Moissiard G, Cokus SJ, Cary J, Feng S, Billi AC, Stroud H, et al. MORC family ATPases required for heterochromatin condensation and gene silencing. *Science.* 2012; 336(6087):1448–51. <https://doi.org/10.1126/science.1221472> PMID: 22555433
13. Moissiard G, Bischof S, Husmann D, Pastor WA, Hale CJ, Yen L, et al. Transcriptional gene silencing by Arabidopsis microRNA homologues involves the formation of heteromers. *Proc Natl Acad Sci U S A.* 2014; 111(20):7474–9. <https://doi.org/10.1073/pnas.1406611111> PMID: 24799676
14. Lorkovic ZJ, Naumann U, Matzke AJ, Matzke M. Involvement of a GHKL ATPase in RNA-directed DNA methylation in Arabidopsis thaliana. *Curr Biol.* 2012; 22(10):933–8. <https://doi.org/10.1016/j.cub.2012.03.061> PMID: 22560611
15. Ikeda Y, Pelissier T, Bourguet P, Becker C, Pouch-Pelissier MN, Pogorelcnik R, et al. Arabidopsis proteins with a transposon-related domain act in gene silencing. *Nat Commun.* 2017; 8:15122. <https://doi.org/10.1038/ncomms15122> PMID: 28466841
16. Wenig U, Meyer S, Stadler R, Fischer S, Werner D, Lauter A, et al. Identification of MAIN, a factor involved in genome stability in the meristems of Arabidopsis thaliana. *Plant J.* 2013; 75(3):469–83. <https://doi.org/10.1111/tpj.12215> PMID: 23607329
17. Uhlken C, Horvath B, Stadler R, Sauer N, Weingartner M. MAIN-LIKE1 is a crucial factor for correct cell division and differentiation in Arabidopsis thaliana. *Plant J.* 2014; 78(1):107–20. <https://doi.org/10.1111/tpj.12455> PMID: 24635680
18. Babu MM, Iyer LM, Balaji S, Aravind L. The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res.* 2006; 34(22):6505–20. <https://doi.org/10.1093/nar/gkl888> PMID: 17130173
19. Steinbauerova V, Neumann P, Novak P, Macas J. A widespread occurrence of extra open reading frames in plant Ty3/gypsy retrotransposons. *Genetica.* 2011; 139(11–12):1543–55. <https://doi.org/10.1007/s10709-012-9654-9> PMID: 22544262
20. Farkas I, Dombradi V, Miskei M, Szabados L, Koncz C. Arabidopsis PPP family of serine/threonine phosphatases. *Trends Plant Sci.* 2007; 12(4):169–76. <https://doi.org/10.1016/j.tplants.2007.03.003> PMID: 17368080
21. Sun X, Kang X, Ni M. Hypersensitive to red and blue 1 and its modification by protein phosphatase 7 are implicated in the control of Arabidopsis stomatal aperture. *PLoS Genet.* 2012; 8(5):e1002674. <https://doi.org/10.1371/journal.pgen.1002674> PMID: 22589732
22. Liu HT, Li GL, Chang H, Sun DY, Zhou RG, Li B. Calmodulin-binding protein phosphatase PP7 is involved in thermotolerance in Arabidopsis. *Plant Cell Environ.* 2007; 30(2):156–64. <https://doi.org/10.1111/j.1365-3040.2006.01613.x> PMID: 17238907
23. Genoud T, Santa Cruz MT, Kulisic T, Sparla F, Fankhauser C, Metraux JP. The protein phosphatase 7 regulates phytochrome signaling in Arabidopsis. *PLoS One.* 2008; 3(7):e2699. <https://doi.org/10.1371/journal.pone.0002699> PMID: 18628957
24. Uhrig RG, Labandera AM, Moorhead GB. Arabidopsis PPP family of serine/threonine protein phosphatases: many targets but few engines. *Trends Plant Sci.* 2013; 18(9):505–13. <https://doi.org/10.1016/j.tplants.2013.05.004> PMID: 23790269

25. Xu D, Marino G, Klingl A, Enderle B, Monte E, Kurth J, et al. Extrachloroplastic PP7L Functions in Chloroplast Development and Abiotic Stress Tolerance. *Plant Physiol.* 2019.
26. Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell.* 2013; 152(1–2):352–64. <https://doi.org/10.1016/j.cell.2012.10.054> PMID: 23313553
27. Fransz P, de Jong H. From nucleosome to chromosome: a dynamic organization of genetic information. *Plant J.* 2011; 66(1):4–17. <https://doi.org/10.1111/j.1365-313X.2011.04526.x> PMID: 21443619
28. Zhao S, Cheng L, Gao Y, Zhang B, Zheng X, Wang L, et al. Plant HP1 protein ADCP1 links multivalent H3K9 methylation readout to heterochromatin formation. *Cell Res.* 2019; 29(1):54–66. <https://doi.org/10.1038/s41422-018-0104-9> PMID: 30425322
29. Tessadori F, Schulkes RK, van Driel R, Fransz P. Light-regulated large-scale reorganization of chromatin during the floral transition in Arabidopsis. *Plant J.* 2007; 50(5):848–57. <https://doi.org/10.1111/j.1365-313X.2007.03093.x> PMID: 17470059
30. Yokthongwattana C, Bucher E, Caikovski M, Vaillant I, Nicolet J, Mittelsten Scheid O, et al. MOM1 and Pol-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing. *Embo J.* 29(2):340–51. <https://doi.org/10.1038/emboj.2009.328> PMID: 19910926
31. Rigal M, Mathieu O. A "mille-feuille" of silencing: epigenetic control of transposable elements. *Biochim Biophys Acta.* 2011; 1809(8):452–8. <https://doi.org/10.1016/j.bbagr.2011.04.001> PMID: 21514406
32. Galperin MY, Koonin EV. Who's your neighbor? New computational approaches for functional genomics. *Nat Biotechnol.* 2000; 18(6):609–13. <https://doi.org/10.1038/76443> PMID: 10835597
33. Clough SJ, Bent AF. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* 1998; 16(6):735–43. <https://doi.org/10.1046/j.1365-313x.1998.00343.x> PMID: 10069079
34. Hristova E, Fal K, Klemme L, Windels D, Bucher E. HISTONE DEACETYLASE6 Controls Gene Expression Patterning and DNA Methylation-Independent Euchromatic Silencing. *Plant Physiol.* 2015; 168(4):1298–308. <https://doi.org/10.1104/pp.15.00177> PMID: 25918117
35. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
36. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015; 12(4):357–60. <https://doi.org/10.1038/nmeth.3317> PMID: 25751142
37. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
38. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015; 31(2):166–9. <https://doi.org/10.1093/bioinformatics/btu638> PMID: 25260700
39. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15(12):550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
40. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 2015; 43(7):e47. <https://doi.org/10.1093/nar/gkv007> PMID: 25605792
41. Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.* 2005; 139(1):5–17. <https://doi.org/10.1104/pp.105.063743> PMID: 16166256
42. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009; 37(Web Server issue):W202–8. <https://doi.org/10.1093/nar/gkp335> PMID: 19458158
43. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics.* 2011; 27(11):1571–2. <https://doi.org/10.1093/bioinformatics/btr167> PMID: 21493656
44. Huang X, Zhang S, Li K, Thimmapuram J, Xie S, Wren J. ViewBS: a powerful toolkit for visualization of high-throughput bisulfite sequencing data. *Bioinformatics.* 2018; 34(4):708–9. <https://doi.org/10.1093/bioinformatics/btx633> PMID: 29087450
45. Catoni M, Tsang JM, Greco AP, Zabet NR. DMRcaller: a versatile R/Bioconductor package for detection and visualization of differentially methylated regions in CpG and non-CpG contexts. *Nucleic Acids Res.* 2018; 46(19):e114. <https://doi.org/10.1093/nar/gky602> PMID: 29986099
46. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32(5):1792–7. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147

47. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 2003; 52(5):696–704. <https://doi.org/10.1080/10635150390235520> PMID: [14530136](https://pubmed.ncbi.nlm.nih.gov/14530136/)
48. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol.* 2008; 25(7):1307–20. <https://doi.org/10.1093/molbev/msn067> PMID: [18367465](https://pubmed.ncbi.nlm.nih.gov/18367465/)
49. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol.* 2006; 55(4):539–52. <https://doi.org/10.1080/10635150600755453> PMID: [16785212](https://pubmed.ncbi.nlm.nih.gov/16785212/)
50. Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS ONE.* 2008; 3(9): e3156. <https://doi.org/10.1371/journal.pone.0003156> PMID: [18776934](https://pubmed.ncbi.nlm.nih.gov/18776934/)
51. Vergara Z, Gutierrez C. Emerging roles of chromatin in the maintenance of genome organization and function in plants. *Genome Biol.* 2017; 18(1):96. <https://doi.org/10.1186/s13059-017-1236-9> PMID: [28535770](https://pubmed.ncbi.nlm.nih.gov/28535770/)

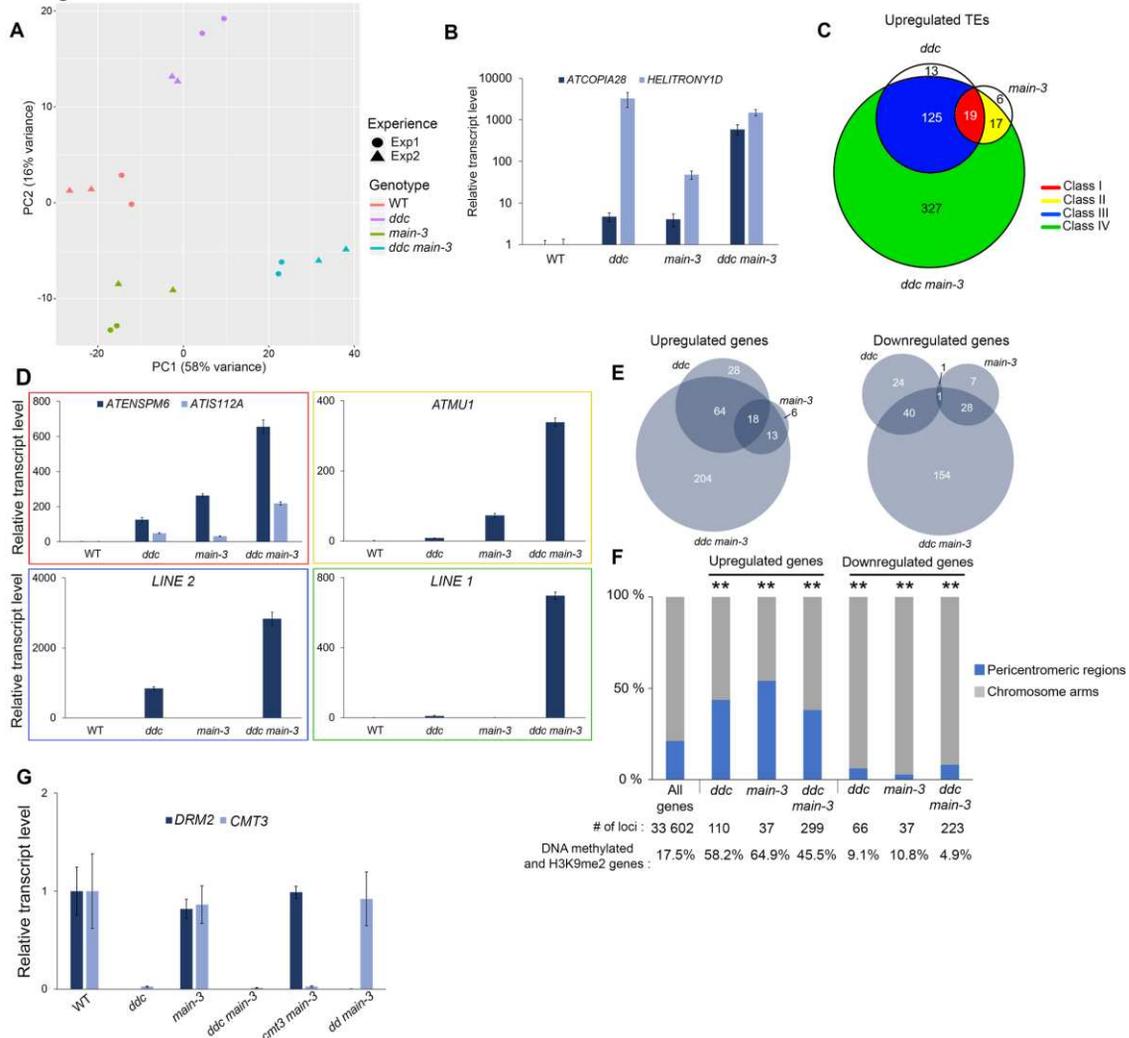
S1 fig



S1 Fig. MAIN is the mutated gene responsible for *ATCOPIA28::GFP* and TE overexpression in the *ddc #16* mutant.

(A) Representative pictures of *ddc #18* (*ddc morc6-8*) and *ddc #344* (*ddc morc6-9*) mutants in comparison to *ATCOPIA28::GFP* WT and *ddc* control plants under UV light. Insets show plants under white light. (B) Enrichment in homozygote/heterozygote ratio of EMS over WT single nucleotide polymorphisms (SNPs), defining the linkage intervals for the populations *ddc #18* and *ddc #344*. Mb: megabase. Gray-shaded rectangles delimit the mapping intervals. (C) Location of the point mutations corresponding to the *morc6-8* and *morc6-9* alleles within the *MORC6* genomic sequence. Nucleotide and corresponding amino acid changes are indicated above the gene. Positions of the mutations are indicated relative to the transcription start site (+1). Grey boxes represent 5' and 3' UTR, blue boxes and lines represent exons and introns, respectively. (D) Enrichment in homozygote/heterozygote ratio of EMS over WT single nucleotide polymorphisms (SNPs), defining the linkage intervals for the population *ddc #16*. Gray-shaded rectangle delimits the mapping interval. (E) Location of the point mutation corresponding to the *main-3* mutant allele within the *MAIN* genomic sequence. (F) Genetic complementation analyses using the KO T-DNA insertion line *main-2*. *ddc #16* plants were crossed with *main-2* plants. F1 plants were self-crossed, and F2 plants were screened under UV light to select GFP-overexpressing plants. Western blotting using anti-GFP antibodies confirmed GFP overexpression in selected F2 plants. Coomassie staining of the large Rubisco subunit (*rbcL*) is used as a loading control. KDa: kilodalton. Among the selected F2 plants, the presence of *main-3* EMS and *main-2* T-DNA mutant alleles were determined by dCAPS-PCR and PCR analyses, respectively. *DRM2* and *CMT3* genotyping were determined by PCR analyses. WT: Wild type, Ho: Homozygote mutant. He: Heterozygote. (G) Relative expression analyses of several TEs in the indicated genotypes assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are represented relative to WT. Error bars indicate standard deviation based on two independent biological replicates. Screening of EMS mutant populations was done on MS plates to allow for visualization of GFP-positive individuals under UV light.

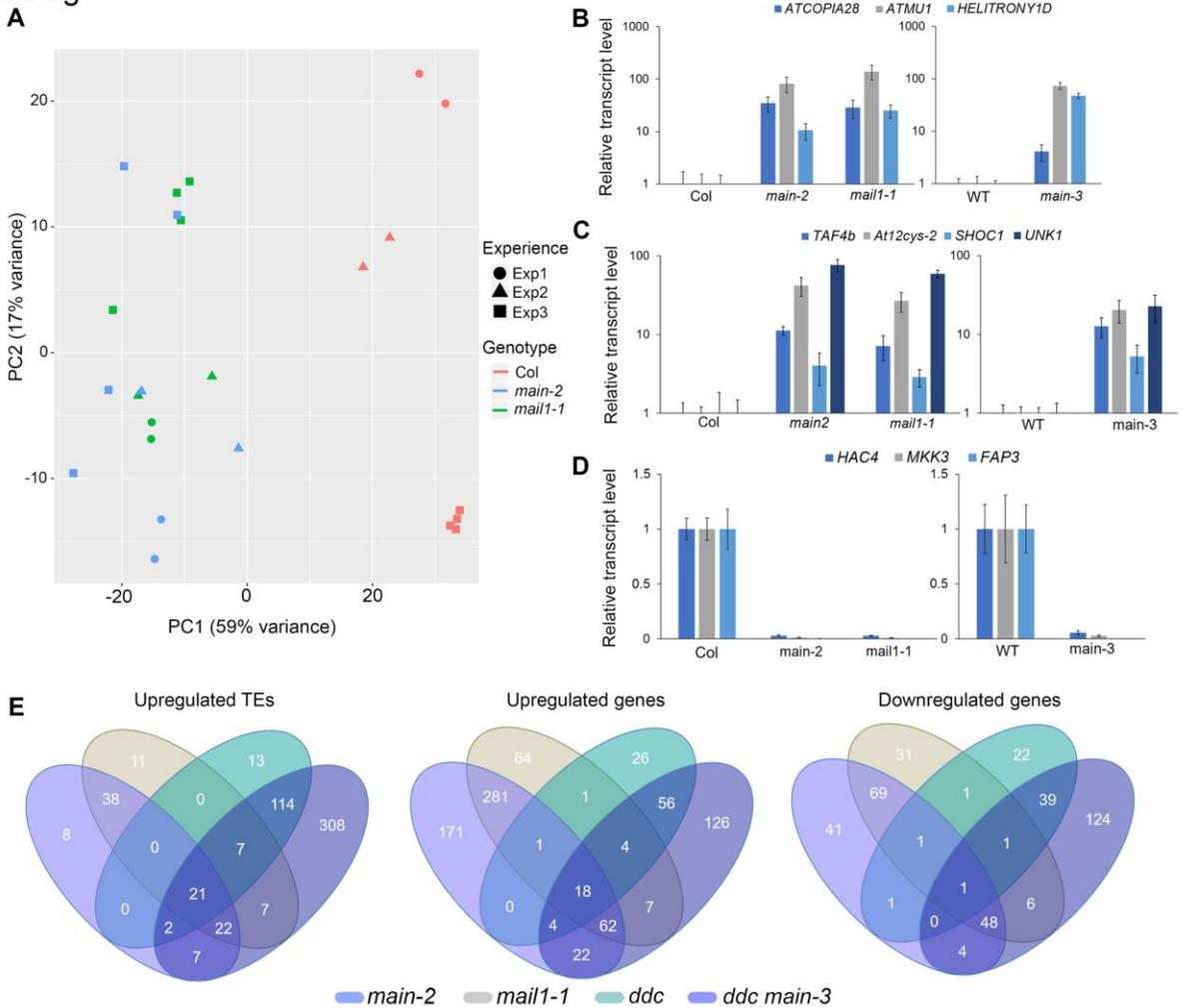
S2 fig



S2 Fig. Combining the *drm2*, *cmt3* and *main-3* mutations exacerbate TE silencing defects.

(A) Principal component analysis (PCA) performed after batch correction for first two components of the sixteen samples described in RNA-seq EMS Exp1 and Exp2. (B) Relative expression analyses of *ATCOPIA28* and *HELITRONY1D* (*AT5TE35950*) in *ddc*, *main-3* and *ddc main-3* assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are represented relative to WT. Error bars indicate standard deviation based on three independent biological replicates. (C) Venn diagrams analysis showing the overlaps between reproducibly upregulated TEs in *ddc*, *main-3* and *ddc main-3*. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value <2.2.10e-16). (D) Same as panel B for TEs defined as class I-IV TEs. Frames of RT-qPCR graphs are using the same color code as shown in panel C. (E) Venn diagrams analyses defining the overlaps between up- and downregulated genes in the different genotypes. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value <2.2.10e-16). (F) Fraction of misregulated genes in *ddc*, *main-3* and *ddc main-3* located in chromosome arms or in pericentromeric regions as defined in [50]. Asterisks indicate statistically significant enrichments of misregulated genes in chromosome arms or pericentromeric regions in comparison to the genomic distributions of all *A. thaliana* genes (Chi-Square test, **: p-value ≤ 0.01). Percentages of genes targeted by DNA methylation and H3K9me2 were calculated based on enrichment in heterochromatin states 8 and 9 as defined in [51]. (G) Relative expression analyses of *DRM2* and *CMT3* in *ddc*, *main-3*, *ddc main-3*, *cmt3 main-3* and *ddc main-3 cmt3 main-3* assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are represented relative to WT. Error bars indicate standard deviation based on three independent biological replicates. Screening of EMS mutant populations was done on MS plates to allow for visualization of GFP-positive individuals under UV light.

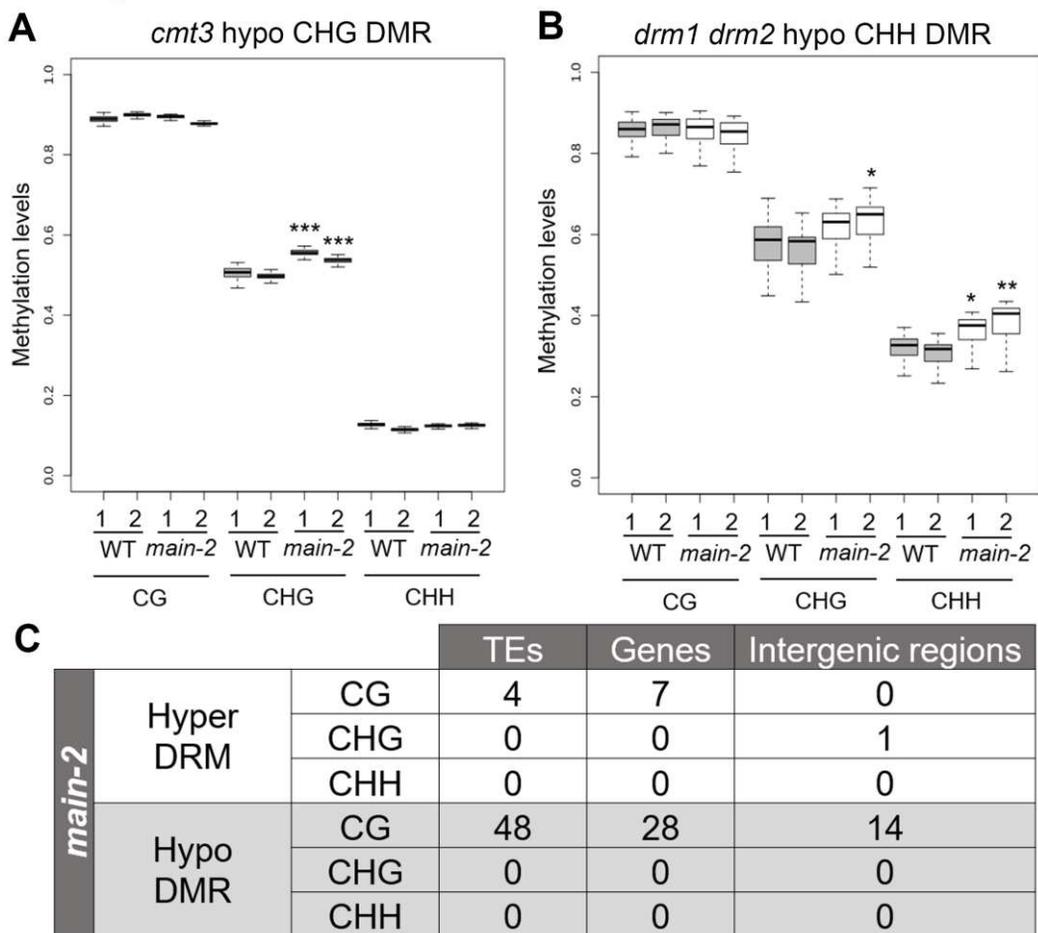
S3 fig



S3 Fig. Identification of reproducibly misregulated loci in *main-2*, *mail1-1* and *main-3*.

(A) Principal component analysis (PCA) performed after batch correction for first two components of the twenty-four *main-2*, *mail1-1* and WT Col samples described in RNA-seq Exp1, Exp2 and Exp3. (B-D) Relative expression analyses of several upregulated TEs (B), upregulated genes (C), and downregulated genes (D) in *main-2*, *mail1-1* and *main-3* assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are represented relative to respective WT controls. Error bars indicate standard deviation based on three independent biological replicates. (E) Venn diagrams analyses representing the overlaps between misregulated loci in *main-2*, *mail1-1*, *ddc* and *ddc main-3*. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value <0.005).

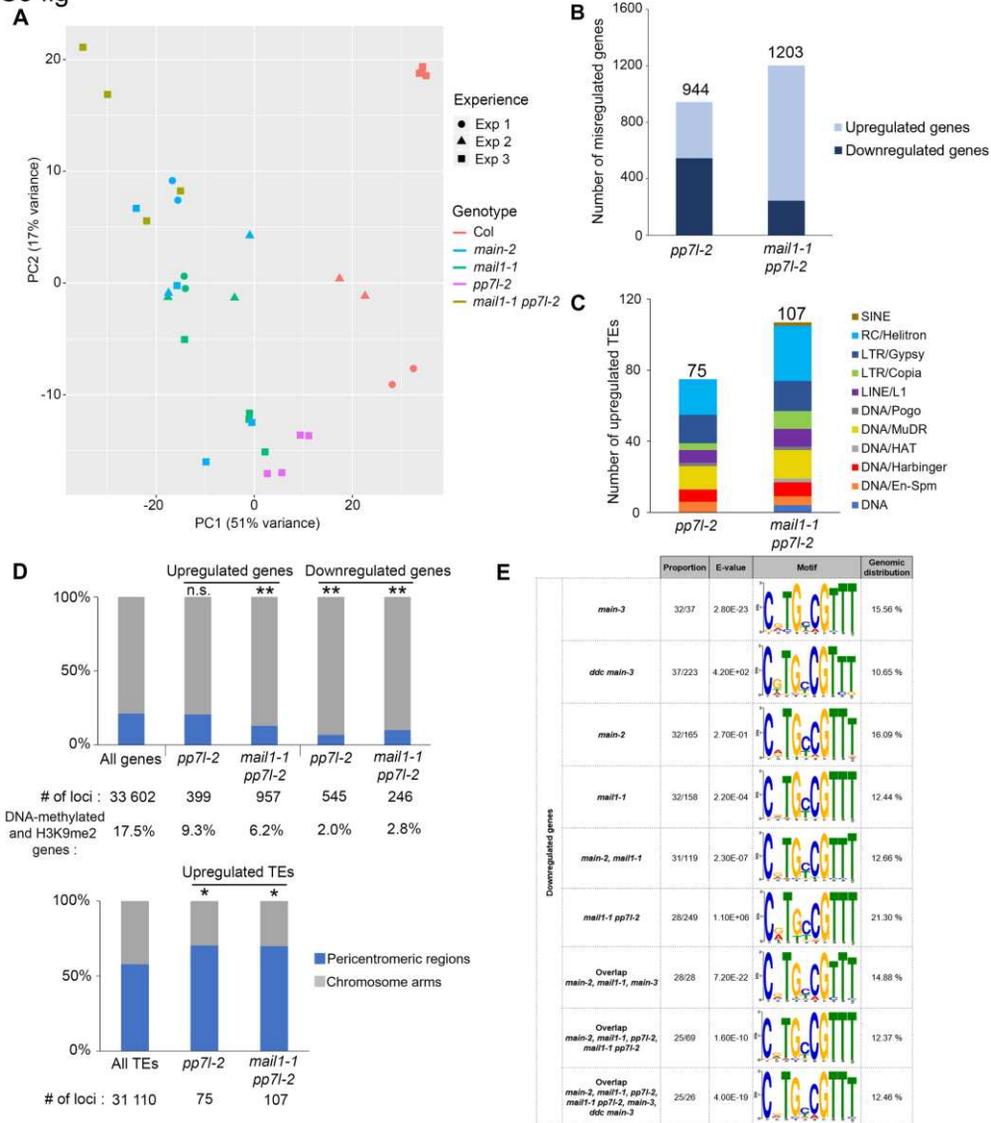
S4 fig



S4 Fig. DNA methylation analyses in the *main-2* mutant

(A-B) Boxplot analyses in two *main-2* and WT Col biological replicates showing the DNA methylation levels at genomic sites previously defined as hypo CHG differentially methylated regions (DMR) in *cmt3* (A) and hypo CHH DMR in *drm1 drm2* (B) based on [26]. p-values were calculated using a Wilcoxon test. *: p-value $< 5.10e-7$, **: p-value $< 5.10e-10$, ***: p-value $< 2.10e-16$.

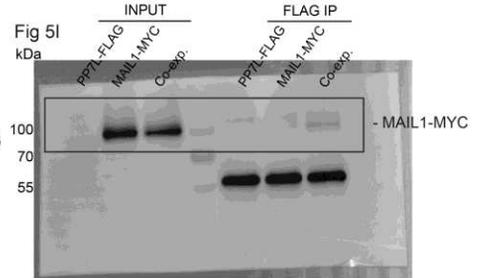
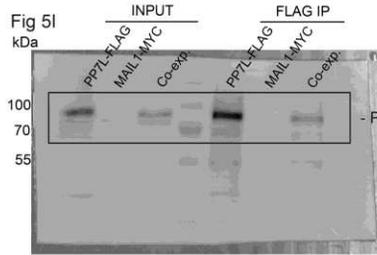
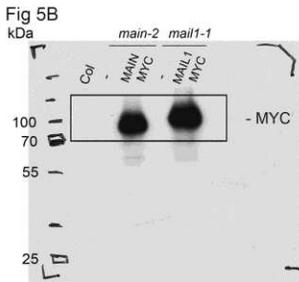
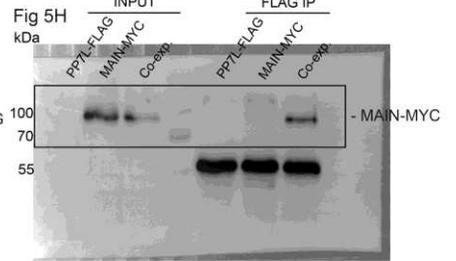
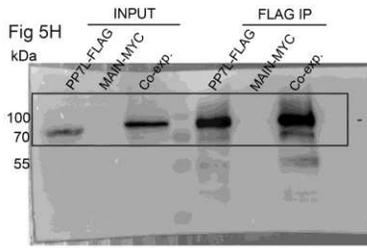
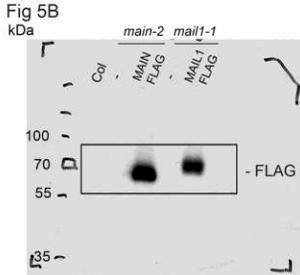
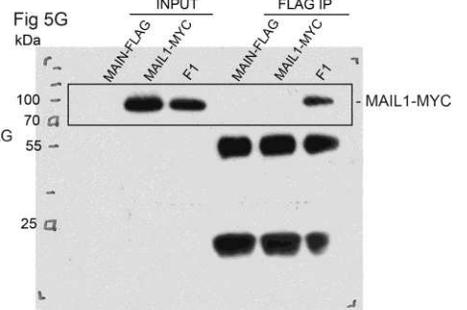
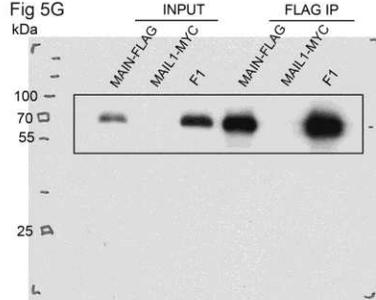
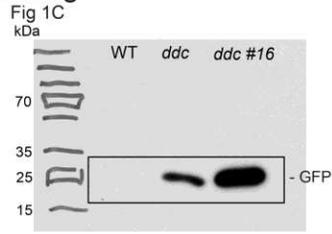
S5 fig



S5 Fig. MAIN, MAIL1 and PP7L are required for the proper expression of similar loci, and commonly downregulated genes carry the ‘DOWN’ DNA motif in their promoter.

(A) Principal component analysis (PCA) performed after batch correction for first two components of the thirty-two samples described in RNA-seq Exp1, Exp2 and Exp3. (B) Number of misregulated genes in the different genotypes in comparison to WT Col plants from RNA-seq Exp3 (four biological replicates, S3 and S6 Tables). (C) Number of upregulated TEs in *pp7l-2* and *mail1-1 pp7l-2*, and classified by TE superfamily. (D) Fraction of misregulated loci in *pp7l-2* and *mail1-1 pp7l-2* located in chromosome arms or in pericentromeric regions as defined in [50]. Asterisks indicate statistically significant enrichments of downregulated genes, upregulated genes and TEs in chromosome arms and pericentromeric regions, respectively, in comparison to the genomic distributions of all *A. thaliana* genes and TEs (Chi-Square test, *: p-value \leq 0.05, **: p-value \leq 0.01, n.s.: not significant). Percentages of genes targeted by DNA methylation and H3K9me2 were calculated based on enrichment in heterochromatin states 8 and 9 as defined in [51]. (E) Identification and proportions of the ‘DOWN’ DNA motif among the promoters of downregulated genes and all *Arabidopsis* genes using the MEME software. Promoter regions are defined as 1kb upstream of ATG. The list of all *Arabidopsis* genes used to determine genomic distributions is based on the TAIR file: TAIR10_upstream_1000_translation_start_20101028. RNA-seq threshold: $\log_2 \geq 2$, or $\log_2 \leq -2$; p-adj < 0.01.

S6 fig



S6 Fig. Full size images of Western blot panels described in Fig 1C, Fig 5B and Fig 5G-I

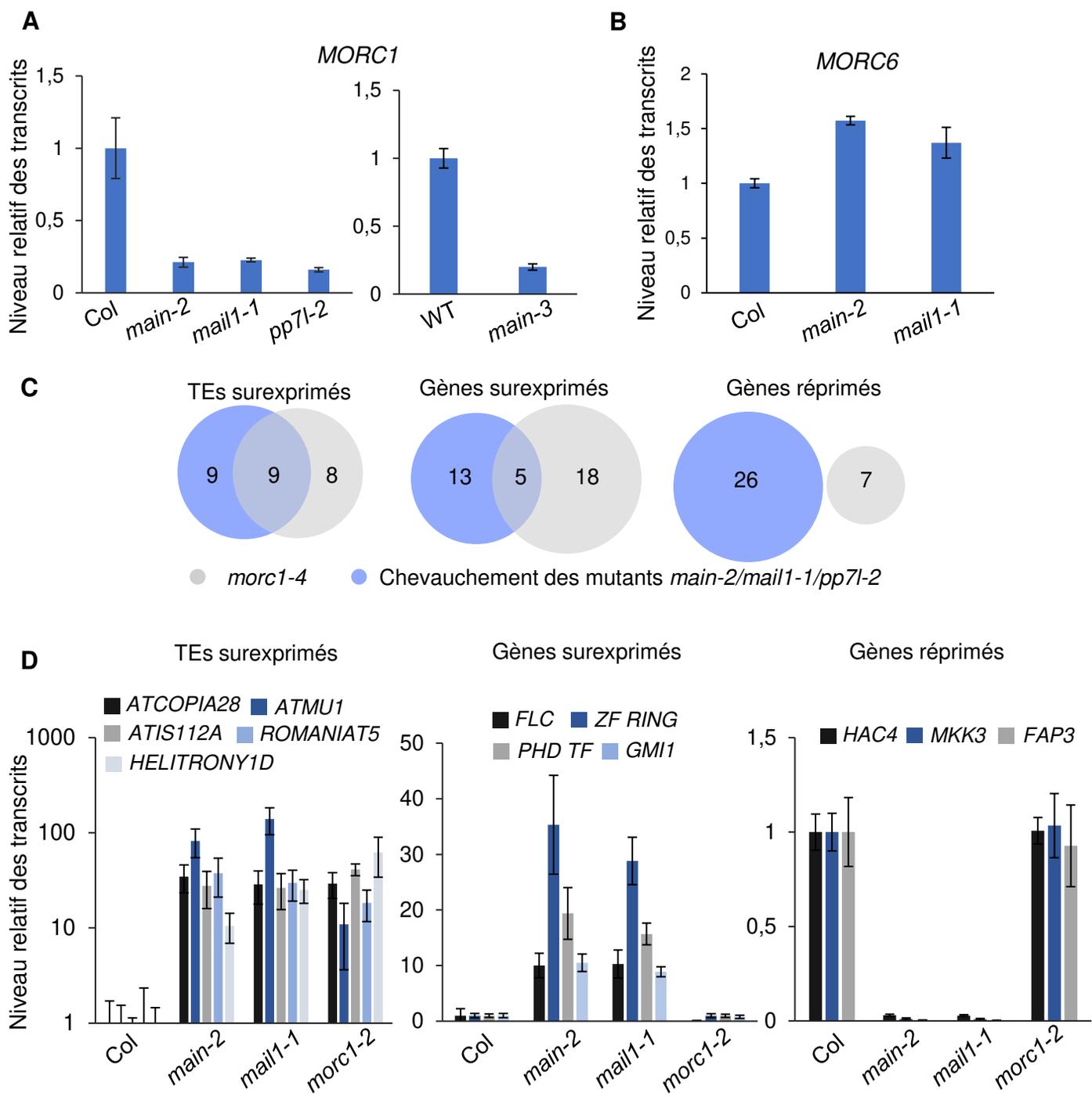


Figure 35 : Les voies MAIN/MAIL1/PP7L et MORC1 régulent un ensemble de TEs commun.

(A-B) Analyses de l'expression relative de *MORC1* (A) dans les mutants *main-2*, *mail1-1*, *pp7l-2* et *main-3* et de *MORC6* (B) dans les mutants *main-2* et *mail1-1* par RT-qPCR. (C) Diagramme de Venn représentant les chevauchements des loci dérégulés dans les mutants *main-2*, *mail1-1*, *pp7l-2* (tables S3 et S4 Nicolau *et al.*, 2020) et *morc1-4*. (D) Analyses de l'expression relative des loci dérégulés dans les mutants *main-2* et *mail1-1* par RT-qPCR. Les analyses de RT-qPCR sont normalisées en utilisant le gène de ménage *RHIP1*, et le niveau des transcrits dans les mutants est représenté de façon relative à la plante sauvage écotype Columbia (Col) pour *main-2*, *mail1-1*, *pp7l-2* et *morc1-2*, ou à *ATCOPIA28::GFP* WT (WT) pour *main-3*. Les barres d'erreur indiquent l'écart type basé sur trois répliquas biologiques indépendants.

CHAPITRE 2

Etude des gènes dérégulés dans les mutants *main-2*, *mail1-1* et *pp7l-2*.

Les résultats présentés dans le premier chapitre nous ont permis de définir un réseau de gènes dérégulés dans les mutants *main-2*, *mail1-1* (désignés tout au long de ce chapitre mutants *pmd*) et *pp7l-2*. En utilisant ces données, nous avons déterminé les gènes communément dérégulés dans ces mutants. Dans ce deuxième chapitre, sont décrits plusieurs gènes communément dérégulés qui ont attiré notre attention car impliqués dans des mécanismes cellulaires en lien avec la chromatine. Dans une première partie, je m'intéresserai plus particulièrement au gène *MORC1*. Dans une seconde partie, je m'intéresserai à deux gènes surexprimés, et notamment *FLC*.

I) Le gène *MORC1* est réprimé dans les mutants *pmd* et *pp7l-2*.

Parmi les gènes réprimés, nous avons identifié *MICRORCHIDIA1* (*MORC1*). Nous avons confirmé sa répression dans les mutants *pmd*, *pp7l-2*, ainsi que dans le mutant hypomorphe *main-3* par des expériences de transcription inverse suivie de PCR quantitative (RT-qPCR) (figure 35.A). Nous avons également testé le niveau d'expression de *MORC6*, partenaire protéique de *MORC1*, dans les mutants *pmd* par RT-qPCR et constaté que le niveau d'ARNm de ce gène n'est pas affecté dans ces mutants (figure 35.B). *MAIN*, *MAIL1* et *PP7L* sont donc nécessaires pour l'expression correcte de *MORC1*, mais pas de *MORC6*.

Étant donné que la protéine *MORC1* joue un rôle dans la répression des TEs, et que nous observons une surexpression de plusieurs TEs dans les mutants *pmd* et *pp7l-2* (Moissiard *et al.*, 2012, 2014; Ikeda *et al.*, 2017; Nicolau *et al.*, 2020), nous avons émis l'hypothèse que la répression de *MORC1* pourrait être responsable, du moins en partie, de la surexpression des TEs. Pour tester cette hypothèse, nous avons, dans un premier temps, comparé les listes de loci dérégulés dans les mutants *pmd* et *pp7l-2* avec ceux dérégulés dans le mutant *morc1-4* (Moissiard *et al.*, 2012).

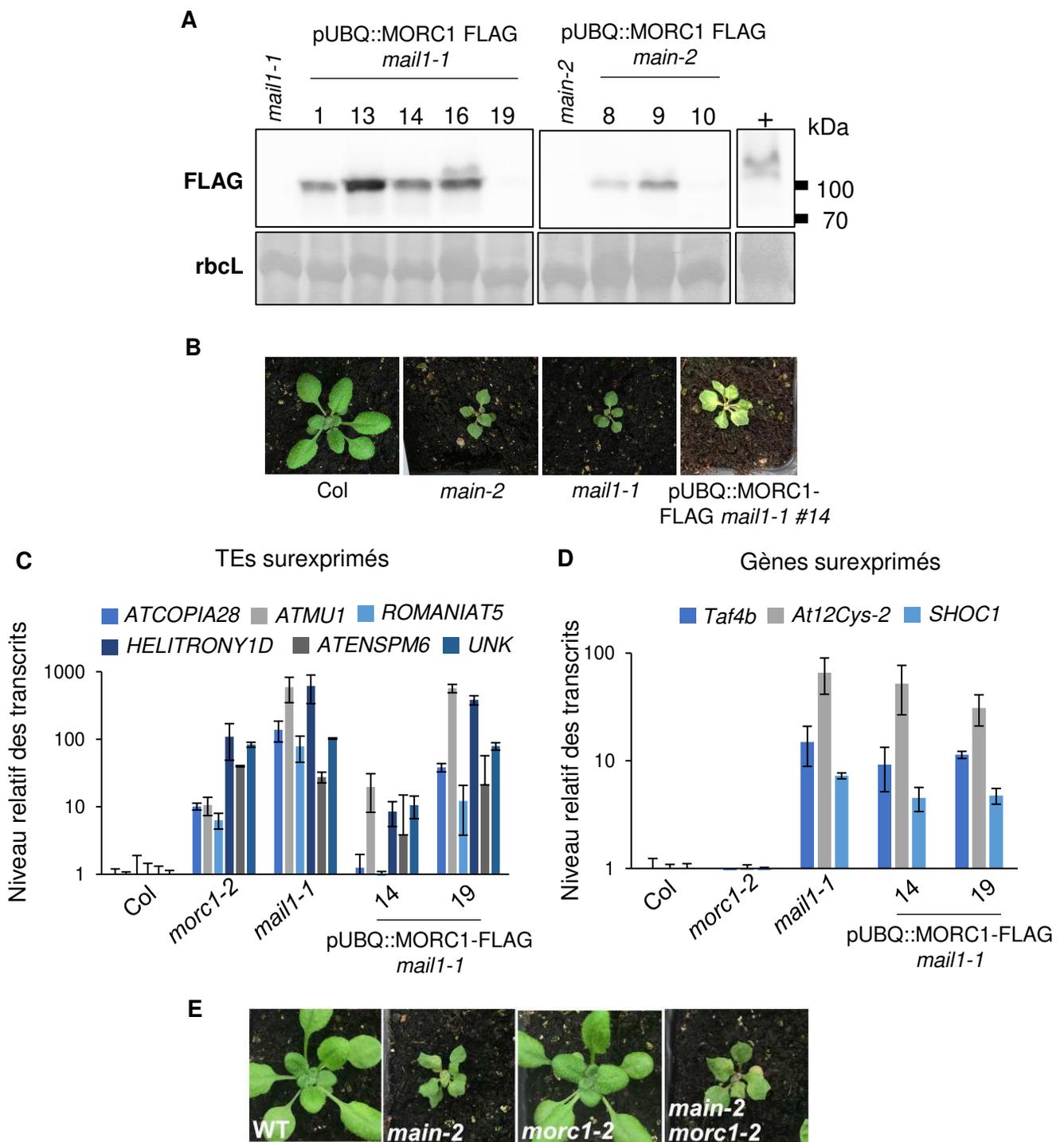


Figure 36 : Le rétablissement de l'expression de *MORC1* restaure en partie la répression des TEs.

(A) Niveau d'expression de MORC1-FLAG dans plusieurs lignées pUBQ::MORC1-FLAG analysées par western blot avec un anticorps anti-FLAG. La coloration au bleu de Coomassie de la grande sous-unité de la Rubisco (*rbcL*) est utilisée comme contrôle de charge. kDa : kilodalton. + : contrôle positif.

(B) Photos représentatives de plantes sauvages Col, *main-2*, *mail1-1* et de la plante pUBQ::MORC1-FLAG *mail1-1* #14 âgées de 3 semaines.

(C-D) Analyses par RT-qPCR dans les lignées pUBQ::MORC1-FLAG *mail1-1* 14 et 19 des niveaux d'expression des TEs (C) et des gènes (D) préalablement identifiés comme étant surexprimés dans les mutants *main-2* et *mail1-1* (Nicolau *et al.*, 2020). Les analyses de RT-qPCR sont normalisées en utilisant le gène de ménage *RHIP1*, et le niveau des transcrits est représenté de façon relative à la plante sauvage Col. Les barres d'erreur indiquent l'écart type basé sur trois répliquas biologiques indépendants.

(E) Photos représentant des plantes sauvages Col, et mutantes *main-2*, *morc1-2* et *main-2 morc1-2* de trois semaines.

Dans cette analyse, nous avons utilisé les loci communément dérégulés dans les mutants KO *pmd* et le mutant *pp7l-2* (figure 35.C). Les chevauchements obtenus montrent que 50% (9 sur 18) des TEs et 28% (5 sur 18) des gènes surexprimés dans les fonds mutants *pmd* et *pp7l-2* sont partagés avec le mutant *morc1-4*, en revanche aucun gène réprimé n'est partagé (figure 35.C). En outre, une analyse plus précise des gènes communément surexprimés dans les mutants *pmd*, *pp7l-2* et *morc1-4* nous a permis de constater que ces cinq gènes sont ciblés par la méthylation de l'ADN dans les trois contextes de séquence et transcriptionnellement réprimés dans un fond sauvage. Ces analyses comparatives ont été validées pour plusieurs loci grâce à des analyses de RT-qPCR (figure 35.D). Ces résultats confortent donc l'hypothèse selon laquelle la surexpression des TEs et de gènes réprimés par la méthylation de leur ADN observée dans les mutants *pmd* et *pp7l-2* serait l'effet indirect d'une répression transcriptionnelle de *MORC1*. Il faut cependant nuancer ces résultats qui s'appuient sur la comparaison de données transcriptomiques issues d'expériences indépendantes. D'autres analyses sont donc nécessaires pour valider notre hypothèse. Ainsi, nous avons développé deux approches afin de tester l'implication « indirecte ou en aval » de *MORC1* dans la répression des TEs via la voie MAIN/MAIL1/PP7L. La première approche est une approche de type complémentarité, dans laquelle l'expression de *MORC1* est rétablie dans des plantes mutantes *main-2* ou *mail1-1*, ceci dans le but de restaurer la répression des TEs. La deuxième approche consiste à déterminer les interactions génétiques entre les voies MAIN/MAIL1/PP7L et *MORC1* par l'analyse de plantes double mutants *main-2 morc1-2*, *mail1-1 morc1-2* et *pp7l-2 morc1-2*.

Pour la première approche, des lignées *main-2* et *mail1-1* ont été transformées avec une construction *pUBQUITIN10::MORC1-FLAG* (*pUBQ::MORC1-FLAG*) dans le but de restaurer l'expression de ce gène grâce au promoteur du gène de ménage *UBQUITIN10*. L'accumulation de la protéine *MORC1-FLAG* a été analysée par western blot, et plusieurs transformants *pUBQ::MORC1-FLAG main-2* et *pUBQ::MORC1-FLAG mail1-1* ont été sélectionnés selon leur niveau d'accumulation de la protéine (figure 36.A). En outre, il est important de noter que tous ces individus accumulant la protéine *MORC1-FLAG* continuent de montrer un phénotype de type mutant *pmd* (figure 36.B), l'accumulation de la protéine *MORC1-FLAG* n'est donc pas suffisante pour rétablir un phénotype développemental sauvage. Par la suite, les niveaux d'expression de plusieurs loci ont été analysés par RT-qPCR. Ces résultats

qui restent préliminaires montrent que, pour la lignée pUBQ::MORC1-FLAG *mail1-1* #14, une complémentation partielle de la répression des TEs serait observée (figure 36.C). Ceci est également le cas pour le gène *UNK* qui est un gène dont l'ADN est fortement méthylé et donc réprimé dans une plante sauvage (Kurihara *et al.*, 2008; Moissiard *et al.*, 2012) (figure 36.D). En revanche, pour les autres gènes ayant été identifiés comme spécifiquement dérégulés dans les mutants *pmd* et *pp7l-2* (mais pas dans *morc1-4*), aucune complémentation n'est observée (figure 36.D). Ainsi, bien que ces résultats soient préliminaires et davantage d'analyses soient nécessaires, il semblerait que rétablir l'expression de MORC1 dans les mutants *pmd* aurait au moins un effet partiel sur le rétablissement de la répression des TEs. Les lignées pUBQ::MORC1-FLAG *main-2* et *mail1-1* sélectionnées seront par la suite utilisées pour des expériences de RNA-seq afin d'analyser à l'échelle du génome l'impact de l'expression de *MORC1* sur les niveaux d'expression des TEs dans les mutants *pmd*.

La deuxième approche consiste à étudier l'interaction génétique entre les voies MAIN/MAIL1/PP7L et MORC1. Pour cela, des plantes doubles mutantes *main-2 morc1-2*, *mail1-1 morc1-2* et *pp7l-2 morc1-2* ont été, ou sont en cours de génération. À ce jour, nous avons seulement obtenu le double mutant *main-2 morc1-2*. Il est intéressant de noter que comme précédemment mentionné pour la lignée pUBQ::MORC1-FLAG *mail1-1* #14, les phénotypes développementaux du double mutant *main-2 morc1-2* et du simple mutant *main-2* sont identiques (figure 36.E). Inversement, une plante *morc1-2* présente, dans nos conditions de culture, un phénotype développemental similaire à celui d'une plante sauvage (figure 36.E). Suite à l'obtention des autres doubles mutants, des expériences de RNA-seq seront réalisées sur ce matériel afin d'évaluer les interactions génétiques entre MAIN/MAIL1/PP7L et MORC1.

II) Étude des gènes surexprimés dans les mutants *pmd*

II.1) *FLOWERING LOCUS C*, un répresseur de la floraison

Le gène surexprimé le plus notable est *FLOWERING LOCUS C (FLC)*. Comme cela a été évoqué en introduction, ce gène code un FT de type MADS-box dont

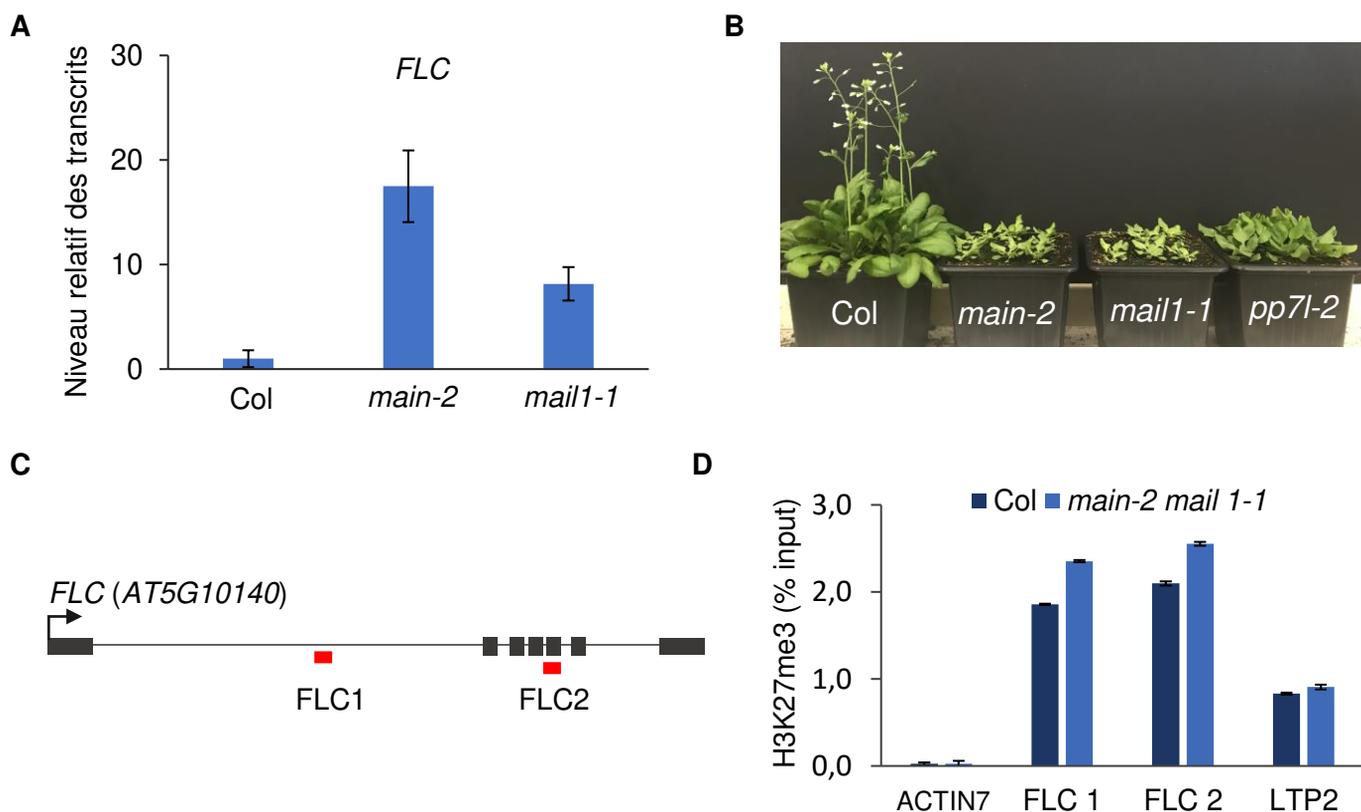


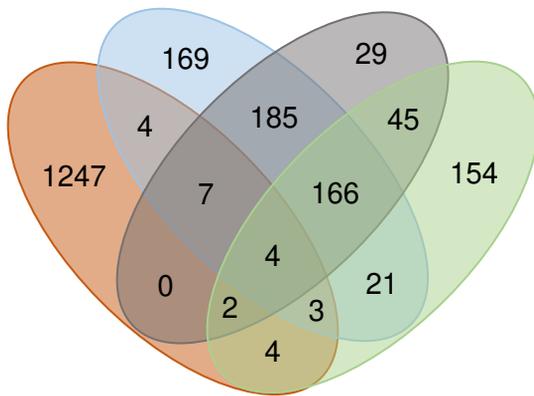
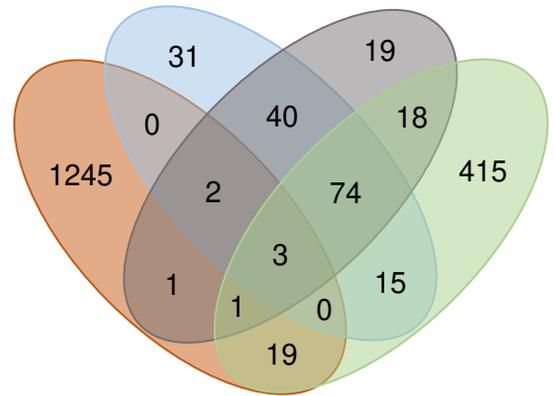
Figure 37 : La surexpression de *FLC* dans les mutants *main-2* et *mail1-1* n'est pas corrélée à une diminution de la marque H3K27me3.

(A) Analyses de l'expression relative de *FLC* dans les mutants *main-2*, *mail1-1* par RT-qPCR. Les analyses de RT-qPCR sont normalisées en utilisant le gène de ménage *RHIP1*, et le niveau des transcrits est représenté de façon relative à la plante sauvage Col. (B) Photos représentatives de plantes sauvages Col, et mutants *main-2*, *mail1-1* et *pp7l-2* à un stade d'environ 4 semaines. (C) Représentation schématique du gène *FLC*. Les exons et les introns sont représentés par des boîtes noires et des traits respectivement. Les couples d'amorces utilisées en (D) amplifient les régions géniques représentées par les boîtes rouges en-dessous du gène (Wu et al., 2016). (D) Analyses ChIP-qPCR de la marque H3K27me3 sur les gènes *FLC* et *LTP2* dans la plante sauvage Col et le double mutant *main-2 mail 1-1*. Les niveaux de H3K27me3 sont représentés en pourcentage d'input et le gène *ACTIN7* est utilisé comme contrôle négatif. Les barres d'erreur indiquent l'écart type (SEM) basé sur deux répliquats indépendants.

l'expression retarde la floraison. *FLC* est réprimé par plusieurs voies épigénétiques dont la voie Polycomb à travers le dépôt de H3K27me3 (Crevillén and Dean, 2011; Yang *et al.*, 2014; Wu *et al.*, 2016). Nous avons tout d'abord validé la surexpression de *FLC* dans les mutants *pmd* et *pp7l-2* par RT-qPCR (figure 37.A). En outre, cette surexpression pourrait expliquer le retard de floraison observé dans ces mutants (figure 37.B). Nous avons ensuite cherché à déterminer si la surexpression de *FLC* dans les mutants *pmd* et *pp7l-2* était due à une diminution de la marque H3K27me3. J'ai donc réalisé des expériences d'immunoprécipitation de la chromatine combinée à des analyses de PCR quantitative (ChIP-qPCR) dans le fond double mutant *main-2 mail1-1* en comparaison à la plante sauvage (figure 37.C, tableau 1). Les résultats obtenus montrent que le double mutant *main-2 mail1-1* a des niveaux de H3K27me3 similaires à ceux observés dans le fond sauvage (figure 37.D). La surexpression de *FLC* dans les mutants *pmd* ne peut donc pas s'expliquer par une diminution de la marque H3K27me3. J'ai également réalisé des ChIP-qPCR en utilisant des amorces sur le gène *LIPID TRANSFERT PROTEIN 2 (LTP2)* (tableau 1), un autre gène ciblé par le complexe polycomb et surexprimé dans les mutants *pmd* et *pp7l-2* (figure 37.E). Comme pour *FLC*, aucune différence entre les niveaux de H3K27me3 sur *LTP2* entre les plantes sauvage et mutante n'a été observée (figure 37.D). Bien qu'une analyse globale par ChIP couplée à du séquençage NGS (ChIP-seq) soit nécessaire pour s'en assurer, nous pouvons malgré tout supposer que le dépôt de la marque H3K27me3 n'est pas perturbé dans les mutants *pmd*, au moins aux régions *FLC* et *LTP2*.

Les plantes mutantes *pmd* et *pp7l-2* semblent présenter un retard de floraison. Ce phénotype développemental peut réellement être un retard de floraison (« late flowering phenotype ») qui, chez *A. thaliana*, s'évalue en comptant le nombre de feuilles de rosette au moment de la floraison de la plante. Il pourrait cependant être tout simplement la conséquence d'un retard de croissance de la plante, comme lors de la perturbation du plastochron (Wang *et al.*, 2008). Le plastochron correspond à l'intervalle de temps entre l'émergence de deux paires de feuilles. Il permet de mesurer l'âge des plantes en terme morphologique plutôt que chronologique.

Le test de comptage de feuilles de rosette à floraison a été effectué sur des plantes mutantes *pp7l-1* et *pp7l-3* et celles-ci ne présentent pas le phénotype de retard de floraison (de Luxán-Hernández *et al.*, 2019). En extrapolant, nous pouvons supposer que cela serait également le cas pour les mutants *pmd*. Il faudra cependant confirmer cette hypothèse en réalisant un test de comptage de feuilles de rosette à

AGènes surexprimés dans
main-2, *mail1-1* et *pp7l-2***B**Gènes réprimés dans *main-2*,
mail1-1 et *pp7l-2*

● Gènes réprimés dans *taf4b* ● *main-2* ● *mail1-1* ● *pp7l-2*

Figure 38 : La surexpression de *Taf4b* dans les mutants *main-2*, *mail1-1* et *pp7l-2* n'est pas suffisante pour induire l'expression de gènes cibles de *Taf4b*

(A-B) Diagrammes de Venn représentant les chevauchements des gènes réprimés dans *taf4b* avec les gènes surexprimés **(A)** ou réprimés **(B)** dans les mutants *main-2*, *mail1-1* et *pp7l-2*.

floraison avec les mutants *pmd* en comparaison à la plante sauvage. En se basant sur tous ces résultats, notre hypothèse privilégiée est donc que les mutants *pmd* et *pp7l-2* présenteraient un retard de croissance, et donc de floraison, du fait de la perturbation de leur plastochron.

II.2) *Taf4b*, un facteur impliqué dans la formation des crossing-over

Le gène *Taf4b* (pour *TATA binding protein-associated factor 4b*) code une sous-unité du facteur de transcription général TFIID, il est fortement exprimé dans les méiocytes mâles, cellules spécialisées dont découlent les microspores (Lawrence *et al.*, 2019). Il a été montré que la protéine joue un rôle important dans la formation des « crossing-over » lors de la méiose (Lawrence *et al.*, 2019).

Dans un mutant *taf4b*, de nombreux gènes sont réprimés, dont une majorité correspondant à des gènes plus fortement exprimés dans les méiocytes mâles en comparaison aux feuilles, cela suggère que *Taf4b* jouerait spécifiquement un rôle dans la régulation de la transcription au sein de ces cellules spécialisées (Lawrence *et al.*, 2019). Considérant que *Taf4b* est impliqué dans le recrutement de la pol II, nous avons émis l'hypothèse que sa surexpression stochastique dans les mutants *pmd* et *pp7l-2* pourrait perturber l'expression des gènes observée dans ces mutants. Afin de tester cette hypothèse, nous avons comparé la liste de gènes qui sont fortement exprimés dans les méiocytes mâles et réprimés dans le mutant *taf4b* avec celles des gènes réprimés ou surexprimés dans les mutants *pmd* et *pp7l-2* (figure 38). Les résultats de ces analyses montrent des chevauchements très faibles entre nos données et celles du mutant *taf4b* (figure 38). La surexpression de *taf4b* dans les mutants *pmd* et *pp7l-2* n'est donc pas suffisante pour induire l'expression de gènes spécifiques des méiocytes mâles dans leurs cellules somatiques de feuilles. Des analyses complémentaires seront très probablement nécessaires pour étudier les interactions génétiques entre les voies MAIN/MAIL1/PP7L et *Taf4b*, comme par exemple, l'étude de plantes combinant les mutations *pmd/pp7l-2* et *taf4b*.

II.3) Conclusions

De nombreux gènes sont dérégulés dans les mutants *pmd* et *pp7l-2*, ceci montrant le rôle probable des protéines MAIN/MAIL1/PP7L dans la régulation de la transcription chez *A. thaliana*. Parmi les gènes qui sont communément dérégulés dans ces mutants, certains ont retenu notre attention étant donné leur rôle dans des processus chromatinien. Les plus notables sont *MORC1*, *FLC* et *Taf4b*. En particulier, concernant *MORC1*, nous avons montré qu'il appartient à un réseau de gènes présentant un motif d'ADN (appelé le « DOWN motif ») qui est enrichi dans leur promoteur (Nicolau *et al.*, 2020). Dans le chapitre 3 des résultats et la discussion, je présenterai des analyses et des hypothèses concernant le mode d'action des protéines MAIN/MAIL1/PP7L dans le processus de régulation de l'expression de ces gènes. Je discuterai également du rôle potentiel du « DOWN motif » dans la régulation génique dépendante de MAIN, MAIL1 et PP7L.

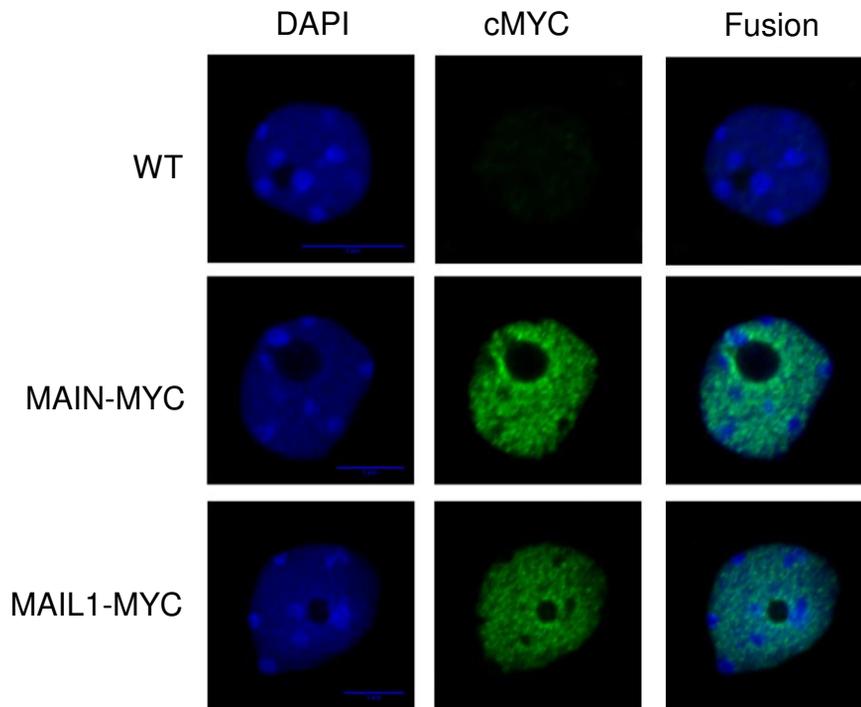


Figure 39 : MAIN et MAIL1 sont des protéines nucléaires.

Images d'immunofluorescence couplée à de la microscopie confocale de noyaux de plantes exprimant les constructions MAIN-MYC ou MAIL1-MYC. Des noyaux de plantes sauvages (WT) sont utilisés comme contrôle négatif. La détection des protéines étiquetées a été réalisée avec un anticorps anti-MYC et l'ADN des noyaux est coloré au DAPI.

Chapitre III

Étude du mode d'action du complexe MAIN/MAIL1/PP7L

Dans les deux chapitres précédents, nous avons décrit l'interaction entre MAIN, MAIL1 et PP7L, ainsi que leur rôle dans la régulation de l'expression de gènes et la répression des TEs. Dans ce chapitre, nous nous intéressons au mode d'action de ces trois protéines.

I) Etude de la localisation subcellulaire de MAIN et MAIL1

Dans la littérature, des observations microscopiques réalisées en cellule d'épiderme de tabac ont montré que les protéines MAIN et MAIL1 sont localisées dans le noyau et que leur NLS est nécessaire pour cette localisation subcellulaire (Wenig *et al.*, 2013 ; Uhlken *et al.*, 2014). Au sein de l'équipe, nous avons produit des lignées complémentées exprimant MAIN-MYC ou MAIL1-MYC, sous le contrôle de leur promoteur endogène respectif (Nicolau *et al.*, 2020). Des expériences de microscopie confocale ont été réalisées avec ce matériel, confirmant ainsi la localisation nucléaire des protéines MAIN et MAIL1. En outre, ces analyses montrent que MAIN et MAIL1 sont uniformément réparties au sein du noyau, à l'exception des chromocentres où elles sont exclues (figure 39).

II) Etude de l'interaction de MAIN et MAIL1 avec la chromatine

Étant donné que les protéines MAIN et MAIL1 sont localisées dans le noyau et que nous avons identifié le motif d'ADN « DOWN » enrichi au sein des promoteurs de plusieurs gènes réprimés, nous avons émis deux hypothèses :

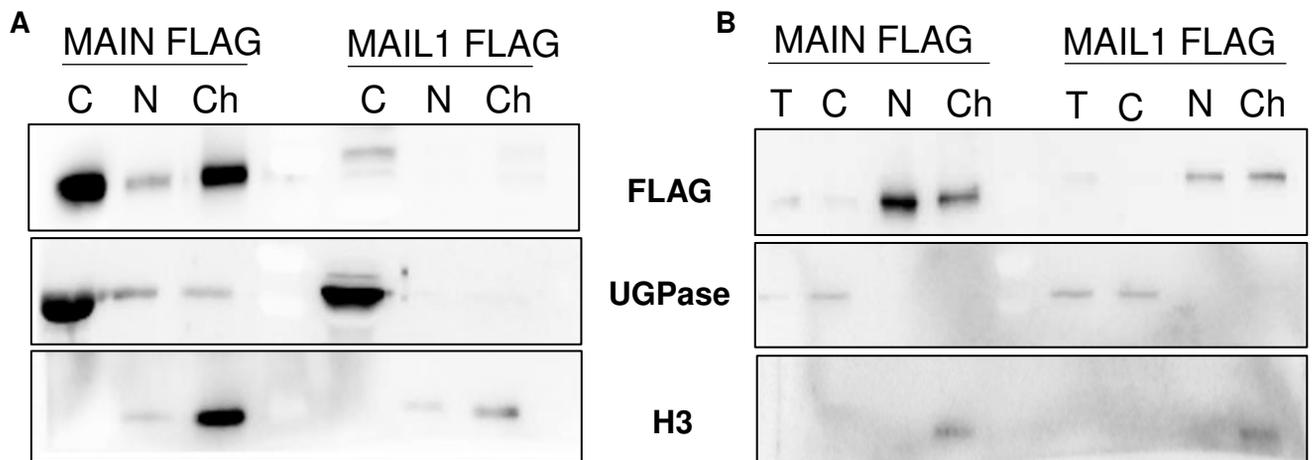


Figure 40 : Les protéines MAIN et MAIL1 sont enrichies dans les fractions nucléoplasmiques et chromatiniennes.

(A-B) Analyses des fractions subcellulaires de plantes exprimant MAIN-FLAG ou MAIL1-FLAG par western blot en utilisant un anticorps anti-FLAG. La bonne séparation et l'absence de contamination entre les différentes fractions est vérifiée en détectant des protéines contrôles enrichies spécifiquement dans ces fractions : UGPase pour la fraction cytoplasmique et H3 pour la fraction chromatinienne en utilisant respectivement des anticorps anti-UGPase et anti-H3. Pour les western blots le même volume d'échantillon (A) ou la même quantité de protéines (B) a été déposé pour chaque fraction. T = total, N = nucléoplasme, C = cytoplasme, Ch = chromatine.

- 1) Le complexe MAIN/MAIL1/PP7L pourrait interagir directement avec la chromatine, potentiellement en reconnaissant le motif « DOWN ».
- 2) Le complexe MAIN/MAIL1/PP7L pourrait recruter un facteur de transcription inconnu, lui-même reconnaissant le motif « DOWN ».

En outre, ces deux hypothèses ne sont pas mutuellement exclusives. Afin de tester notre 1^{ère} hypothèse, nous avons tout d'abord réalisé des expériences de fractionnement cellulaire pour analyser l'enrichissement des protéines MAIN et MAIL1 dans les différents compartiments cellulaires que sont le cytoplasme, le nucléoplasme et la chromatine (figure 40). Deux expériences de fractionnement cellulaire indépendantes ont été effectuées sur des plantes de 3 semaines et des plantes de 10 jours respectivement. Pour la première, des volumes équivalents d'échantillon ont été utilisés pour le Western blot (figure 40.A). Pour la seconde expérience, les échantillons protéiques ont été dosés, et des quantités équivalentes de protéines ont été chargées sur gel (figure 40.B). Dans les deux cas, ces analyses montrent que les protéines MAIN et MAIL1 sont enrichies dans la fraction chromatinienne, ainsi qu'un enrichissement de MAIN dans la fraction cytoplasmique. Pour MAIL1, cet enrichissement dans le cytoplasme n'est observé que lors de la première expérience. Ceci peut s'expliquer par le fait que le cytoplasme est un compartiment très concentré en protéines, dont MAIL1 ne représente qu'une très faible proportion. En déposant la même quantité de protéine pour toutes les fractions, MAIL1 devient alors indétectable par western blot dans le cytoplasme. Lors de cette 2^{ème} expérience, où de mêmes quantités de protéines ont déposées entre toutes les fractions, nous avons constaté que MAIN est nettement plus enrichie dans les fractions chromatinienne et nucléoplasmique en comparaison avec la fraction cytoplasmique (figure 40.B). Ces résultats confortent donc l'hypothèse d'une probable interaction entre les protéines MAIN, MAIL1 et la chromatine. Cependant à ce stade, nous ne savons pas si cette interaction est directe, ou si les protéines MAIN/MAIL1 ont besoin d'un autre facteur pour interagir avec la chromatine. Ce dernier pourrait être PP7L ou un facteur inconnu.

Afin de tester les propriétés biochimiques et l'activité enzymatique des protéines MAIN, MAIL1 et PP7L, nous avons récemment initié la production *in vitro* de protéines en utilisant le système TnT® Coupled Wheat germ extract (Promega), qui couple les processus de transcription et traduction à partir d'extrait de germes de blé afin d'optimiser la synthèse protéique. Les protéines ainsi produites sont ensuite purifiées grâce à une fusion avec le peptide HaloTag (Los *et al.*, 2008). Pour une expérience

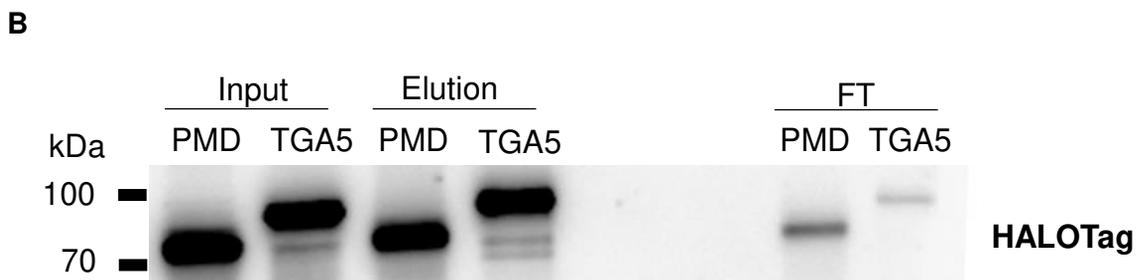
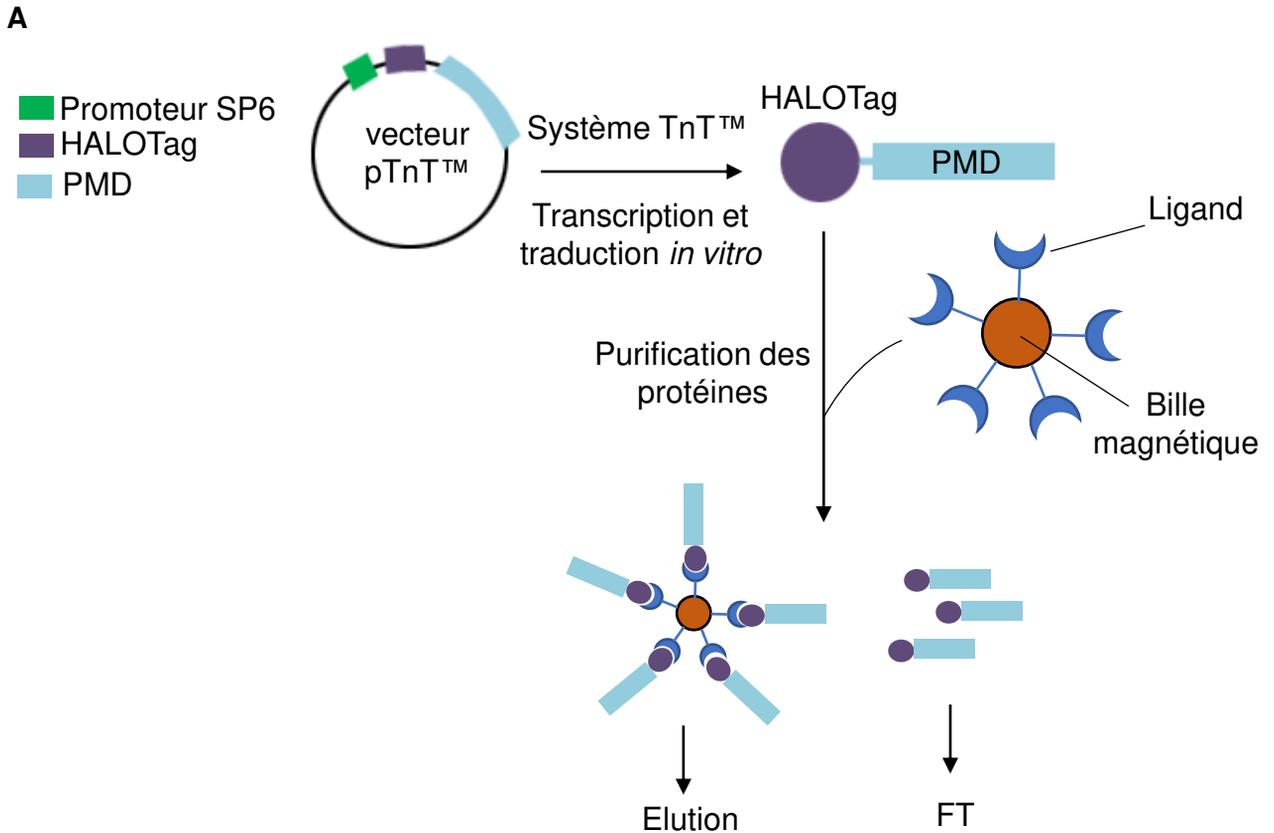


Figure 41 : Production *in vitro* de protéines HaloTag-PMD par le système TnT™ Wheat Germ Extract. (A) Représentation schématique de la production de protéines *in vitro* par le système TnT™. La transcription de la construction HaloTag-PMD est sous le contrôle du promoteur SP6, et l'ARNm est traduit en protéine en système *in vitro*. Les protéines produites sont ensuite purifiées par des billes recouvertes de ligand HaloTag. Elution : fraction des protéines liées aux billes. FT: « flowthrough » fraction des protéines non-liées sur les billes magnétiques (B) Vérification par western blot de la purification des HaloTag-PMD en utilisant un anticorps anti-HaloTag. TGA5 : contrôle positif de la synthèse *in vitro*. Input : échantillon avant purification.

préliminaire, le domaine PMD de MAIN fusionné au HaloTag a été produit et purifié, puis la présence de la protéine a été vérifiée par western blot (figure 41). Par une même approche, les protéines HaloTag-MAIL1 et HaloTag-PP7L seront produites. Puis nous utiliserons ce matériel pour répondre à différentes questions :

- 1) Les protéines MAIN, MAIL1 et PP7L sont-elles capables d'interagir directement avec l'ADN, et notamment avec le motif « DOWN » ?
- 2) Les protéines MAIN, MAIL1 et PP7L interagissent-elles directement avec les protéines histones ?
- 3) ou avec les nucléosomes contenant ou non des séquences d'ADN spécifiques comme le motif « DOWN » ?

III) Conclusions

Dans ce chapitre, nous avons montré que MAIN et MAIL1 sont localisées dans le noyau, ce qui a aussi été décrit pour PP7L (de Luxán-Hernández *et al.*, 2019; Xu *et al.*, 2019). Le complexe MAIN/MAIL1/PP7L est donc probablement fonctionnel dans le noyau. Par des expériences de fractionnement cellulaire, nous avons montré que MAIN et MAIL1 sont enrichies dans la fraction chromatinienne, ce qui suggère fortement que ces protéines interagissent (directement ou indirectement) avec la chromatine. Dans la partie discussion, nous proposons des expériences à réaliser dans la continuité de ces résultats, ceci afin de mieux comprendre le mode d'action des protéines MAIN et MAIL1.

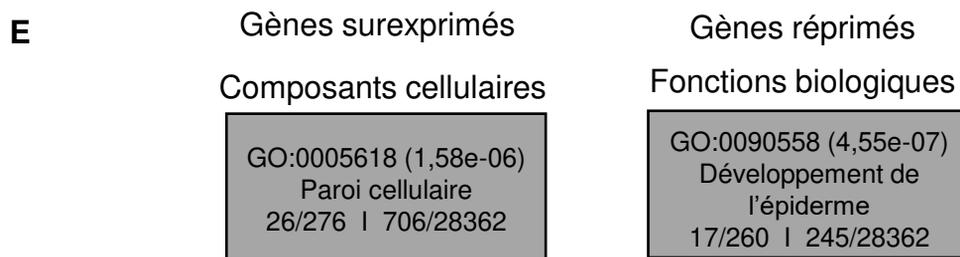
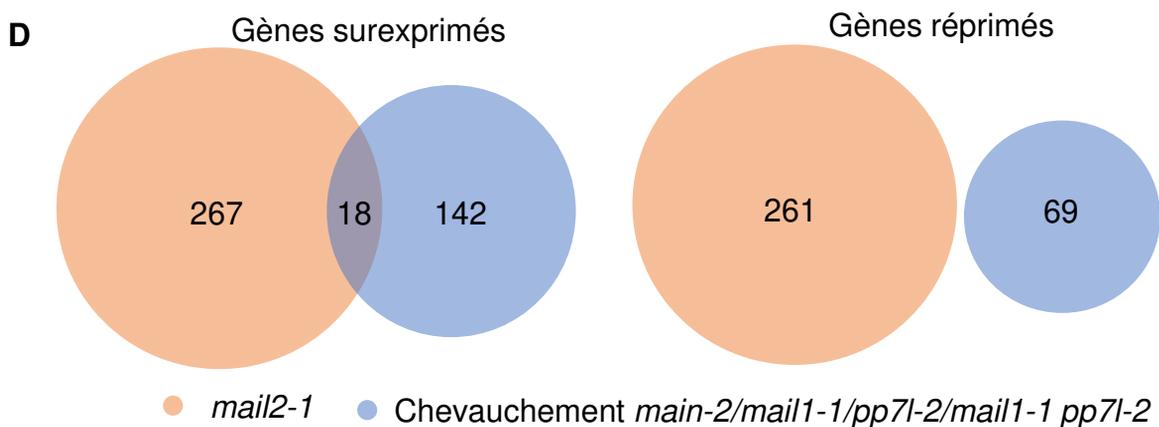
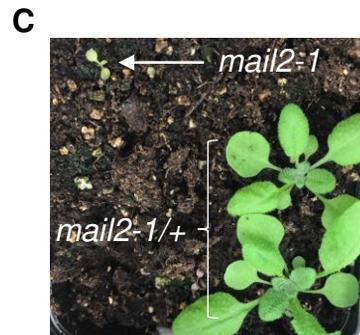
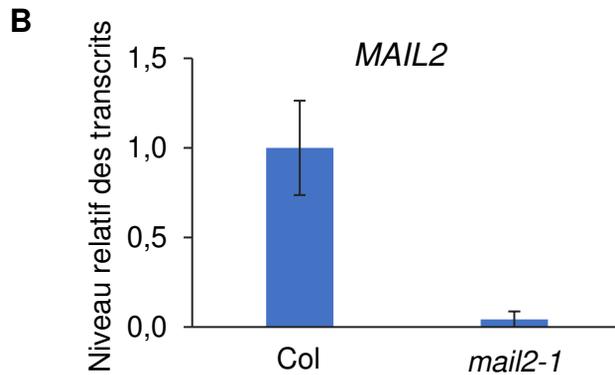
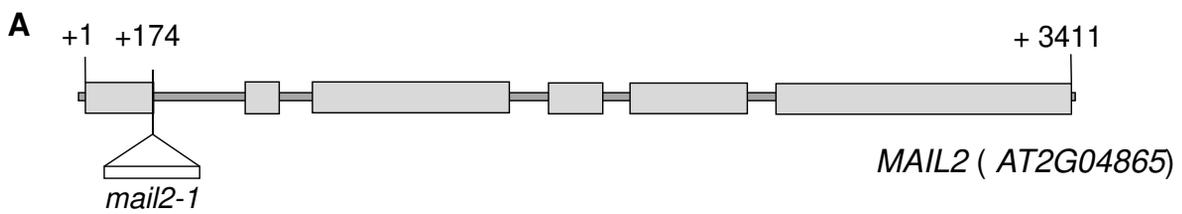


Figure 42 : Caractérisation du mutant *mail2-1*

(A) Localisation de l'insertion d'ADN-T correspondant au mutant *mai2-1* dans le gène *MAIL2*. La position de l'insertion est indiquée relativement au site de début de transcription (+1). Les exons et les introns sont représentés par des boîtes gris clair et des traits respectivement. (B) Analyses de l'expression relative de *MAIL2* dans le mutant *mail2-1* par RT-qPCR. Les analyses de RT-qPCR sont normalisées en utilisant le gène de ménage *RHIP1*, et le niveau des transcrits est représenté de façon relative à la plante sauvage Col. (C) Photos représentatives de plantes *mail2-1* et de plantes hétérozygotes *mail2-1/+* âgées de 3 semaines. (D) Diagramme de Venn représentant les chevauchements des loci dérégulés dans les mutants *main-2*, *mail1-1*, *pp7l-2*, *mail1-1 pp7l-2* et *mail2-1*. (E) Analyses d'ontologie génétique pour les gènes dérégulés dans *mail2-1* réalisées sur le site agriGOv2 (<http://systemsbiology.cau.edu.cn/agriGOv2/>). Seuil : $\log_2 \geq 2$, or $\log_2 \leq -2$; $p\text{-adj} < 0.01$

Chapitre VI

Etude des protéines MAIL2 et MAIL3

Chez *A. thaliana*, la famille PMD-C est composée de quatre membres : MAIN, MAIL1, MAIL2 et MAIL3. Nous avons vu l'importance de MAIN et MAIL1 dans la régulation de l'expression des gènes, la répression des TEs, et plusieurs processus cellulaires et développementaux. Au sein de l'équipe, nous nous intéressons également aux autres protéines de la famille : MAIL2 et MAIL3. Ces protéines ont été très peu étudiées, leur fonction au sein de la plante est donc inconnue. Dans ce chapitre sont résumés les résultats préliminaires obtenus sur ces protéines.

I) MAIL2

MAIL2 est un gène portant le PMD comme seul domaine. Nous avons identifié un mutant ADN-T *mail2-1* (figure 42.A) et nous avons validé la perte d'expression de *MAIL2* par RT-qPCR (figure 42.B). Après germination, les plantes mutantes *mail2-1* sont incapables de développer de vraies feuilles, et finalement meurent à un stade très précoce du développement de la plantule (figure 42.C). *MAIL2* pourrait donc être une protéine essentielle pour le développement de la plante. Toutefois, cette observation reste très préliminaire et devra être validée avec d'autres lignées portant un allèle mutant *mail2* différent. En effet, à ce stade, nous ne pouvons pas exclure que le phénotype des plantes mutantes *mail2-1* soit dû à une autre mutation au sein du génome.

Nous avons tout de même réalisé des analyses moléculaires sur la lignée *mail2-1*, afin de mieux comprendre quelles voies biologiques peuvent être affectées en l'absence de *MAIL2*. Un RNA-seq a été réalisé sur le mutant *mail2-1*, montrant que de nombreux gènes sont dérégulés. Nous avons ensuite comparé les listes de gènes dérégulés dans le mutant *mail2-1* avec celles des gènes communément dérégulés dans les mutants *main-2*, *mail1-1*, *pp7l-2* et *mail1-1 pp7l-2* (Nicolau *et al.*, 2020) (figure

42.D). La majorité des gènes dérégulés dans le mutant *mail2-1* n'est pas affectée dans les mutants *pmd* et *pp7l-2*, ce qui suggère que MAIL2 agirait dans une voie différente de celle de MAIN/MAIL1/PP7L.

Nous avons réalisé des analyses d'ontologie génétique, afin d'étudier l'enrichissement des gènes dérégulés dans le mutant *mail2-1* dans différentes catégories ontologiques. Nous avons constaté qu'une partie des gènes surexprimés est enrichie significativement dans la catégorie des composants cellulaires « Paroi cellulaire ». Quant aux gènes réprimés, on remarque un enrichissement significatif dans la catégorie des fonctions biologiques « développement de l'épiderme » (figure 42.E). En s'intéressant de plus près aux gènes dérégulés dans le mutant *mail2-1*, nous avons identifié plusieurs gènes intéressants. Certains gènes comme *ROS1*, *HDA3*, *CLSY3* et *NRPE5*, une sous-unité de la pol V sont impliqués dans la dynamique des marques épigénétiques. D'autres sont quant à eux impliqués dans la mise en place des différents organes de la plante. C'est le cas de *WUSCHEL-RELATED HOMEBOX 11 (WOX11)*, qui agit dans l'émergence de la racine, et de *AGAMOUS-LIKE 24 (AGL24)*, qui est impliqué dans le déclenchement correct de la floraison (Yu *et al.*, 2002; Liu *et al.*, 2014).

Bien que tous ces résultats soient très préliminaires et nécessitent confirmation, particulièrement pour ce qui concerne la causalité de la mutation du gène *MAIL2* sur les phénotypes observés, il se pourrait que MAIL2 soit une protéine importante pour l'expression de facteurs critiques à la mise en place de marques épigénétiques et pour assurer une anatomie correcte de la plante. Ceci sera plus amplement discuté dans la partie discussion.

Si l'on suppose que c'est bien la mutation de *MAIL2* qui est responsable des phénotypes observés, l'étude de ce gène reste toutefois limitée à cause de la létalité de l'allèle *mail2-1*. Des méthodes alternatives nous permettant de mieux comprendre le rôle de MAIL2 seront décrites dans la partie discussion.

II) MAIL3

Contrairement aux autres protéines PMD-C, MAIL3 porte, en plus de son PMD, un domaine PPP, apparenté à celui de PP7L. Pour déterminer le rôle de MAIL3 au sein

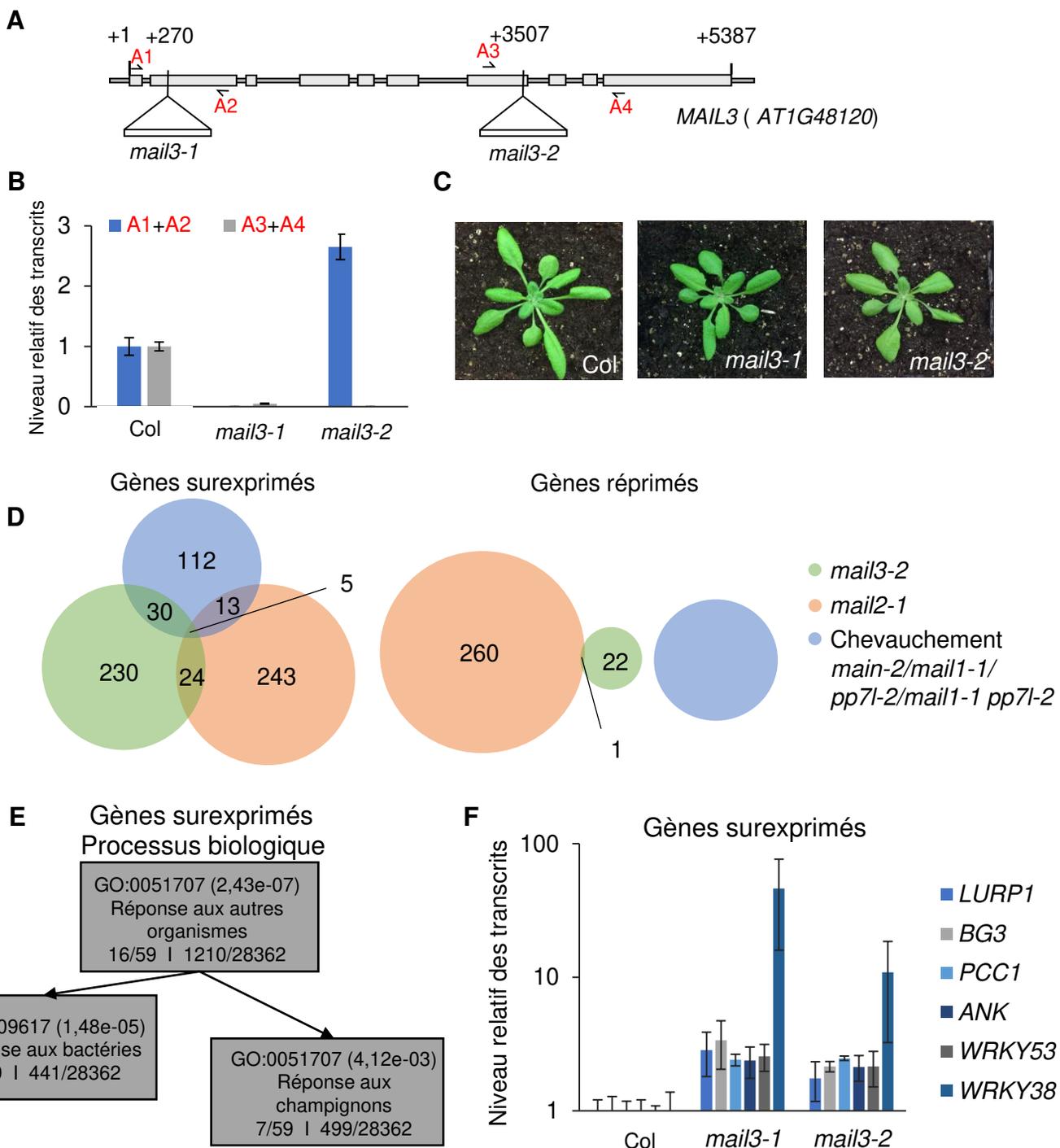


Figure 43 : Caractérisation des allèles mutants *mail3-1* et *mail3-2*

(A) Localisation des insertions d'ADN-T correspondant aux allèles mutants *mail3-1* et *mail3-2* dans le gène *MAIL3*. La position de l'insertion est indiquée relativement au site de début de transcription (+1). Les exons et les introns sont représentés par des boîtes gris clair et des traits respectivement. Les couples d'amorces utilisées en (B) sont représentées par des flèches (Ühlken *et al.*, 2014) (B) Analyses de l'expression relative de *MAIL3* dans les mutants *mail3-1* et *mail3-2* par RT-qPCR. (C) Photos représentatives de plantes sauvages Col, *mail3-1* et *mail3-2* âgées de 3 semaines. (D) Diagramme de Venn représentant les chevauchements des loci dérégulés dans les mutants *main-2*, *mail1-1*, *pp7l-2*, *mail1-1 pp7l-2* et *mail3-2*. (E) Analyses d'ontologie génétique pour les gènes dérégulés dans *mail3-2* réalisées sur le site agriGOv2 (<http://systemsbiology.cau.edu.cn/agriGOv2/>). Seuil : $\log_2 \geq 2$, or $\log_2 \leq -2$; $p\text{-adj} < 0.01$. (F) Analyses de l'expression relative de gènes surexprimés dans les mutants *mail3-1* et *mail3-2* par RT-qPCR. Les analyses de RT-qPCR sont normalisées en utilisant le gène de ménage *RHIP1*, et le niveau des transcrits est représenté de façon relative à la plante sauvage Col.

de la plante, nous avons utilisé deux allèles mutants : *mail3-1* et *mail3-2*. Ces allèles diffèrent par le lieu d'insertion de l'ADN-T dans la séquence du gène MAIL3. Pour l'allèle *mail3-1*, l'ADN-T est inséré dans la région 5' du gène, en amont du domaine PMD, tandis que pour *mail3-2* l'ADN-T est inséré dans le domaine PPP (figure 43.A). Par RT-qPCR, nous avons analysé l'expression du gène *MAIL3* dans ces deux fonds mutants. Nous avons vu qu'il n'y a pas de production d'ARNm dans le mutant *mail3-1*, tandis que le mutant *mail3-2* exprime une version tronquée de l'ARNm, dans lequel la partie 3' de l'ARNm est absente (figure 43.B). La possibilité qu'une version tronquée de MAIL3, contenant uniquement son domaine PMD, s'accumule dans le fond mutant *mail3-2* n'a pas été analysée. Les résultats que nous avons obtenus dans le mutant *mail 3-1* diffèrent de ceux publiés par Ühlken *et al.* en 2014 qui montrent l'accumulation d'une version tronquée de l'ARNm de *MAIL3* dans cet allèle mutant. Comme cela a déjà été décrit, les plantes mutantes *mail3-1* et *mail3-2* ne montrent pas de phénotype développemental différent des plantes sauvages, ce qui suggère que MAIL3 n'est pas nécessaire pour le développement de la plante (figure 43.C, Ühlken *et al.*, 2014). Malgré cela, nous avons réalisé un RNA-seq sur les plantes *mail3-2* et mené des études comparatives entre les listes de gènes dérégulés dans *mail3-2*, *mail2-1* et ceux communément dérégulés dans *main-2*, *mail1-1*, *pp7l-2* et *mail1-1 pp7l-2* (Nicolau *et al.*, 2020) (figure 43.D). Ces analyses n'ont identifié qu'un faible chevauchement entre ces listes, ce qui semble démontrer que MAIL3 agirait dans des voies différentes des autres protéines PMD-C. Par la suite, nous avons analysé les catégories d'ontologie génétique auxquelles appartiennent les gènes dérégulés dans le mutant *mail3-2*. Nous avons observé un enrichissement des gènes surexprimés dans les catégories « réponse aux bactéries » et « réponse aux champignons ». La surexpression de certains gènes impliqués dans la réponse aux pathogènes a été validé par RT-qPCR dans les deux fonds génétiques (figure 43.E). Ces observations nous ont amené à émettre l'hypothèse que les plantes *mail3-1* et *mail3-2* pourraient être plus résistantes aux pathogènes que des plantes sauvages. Pour tester cette hypothèse, deux collaborations ont été initiées avec des équipes spécialisées dans les interactions plante-pathogène. L'équipe de Lionel Navarro, de l'ENS de Paris et l'équipe de Harald Keller, de l'institut Sophia agrobiotech de l'INRA testent la résistance des mutants à la bactérie *Pseudomonas syringae* et au champignon *Hyaloperonospora arabidopsidis* respectivement. Les résultats préliminaires de ces tests montrent que les plantes *mail3-1* semblent plus résistantes à une infection à *Pseudomonas syringae* (figure

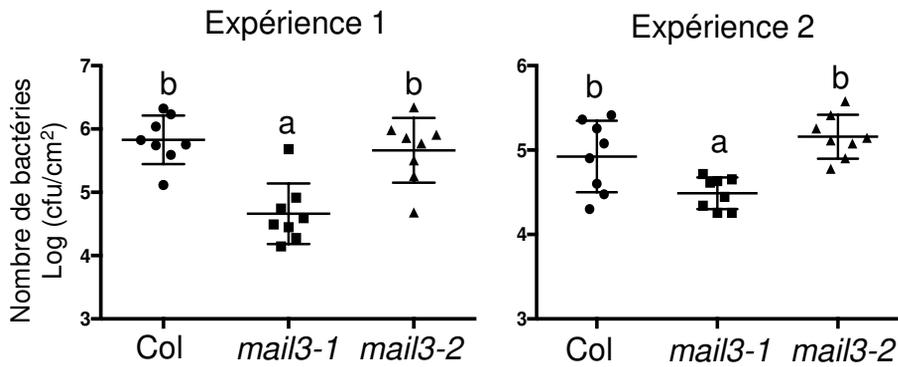
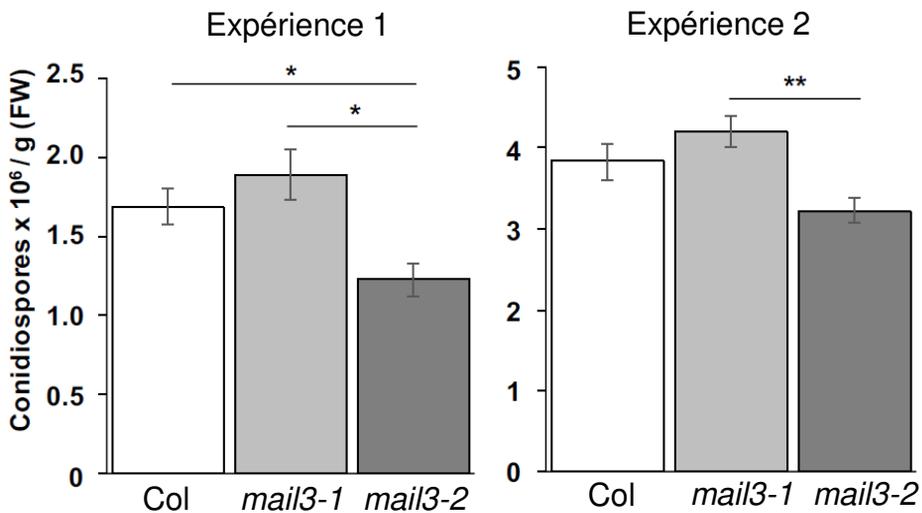
A*Pseudomonas syringae***B***Hyaloperonospora arabidopsidis*

Figure 44 : Résistance des plantes mutantes *mail3-1* et *mail3-2* à des infections bactériennes et fongiques

(A-B) Tests d'infections de plantes sauvages Col, *mail3-1* et *mail3-2* avec la bactérie *Pseudomonas syringae* (A) et le champignon *Hyaloperonospora arabidopsidis* (B). (A) Les infections sont réalisées sur des feuilles. Trois jours après infection, les feuilles sont broyées et étalées sur boîte à 28°C, le nombre de bactéries est déterminé par le comptage des colonies après 36h d'incubation. (B) Les infections sont réalisées sur des cotylédons de plantes de 10 jours et les spores sont comptées sous microscope. * : p<0,05; ** : p<0,01.

44.A), mais pas à *Hyaloperonospora arabidopsidis* (figure 44.B). En revanche les plantes *mail3-2* ne semblent pas être significativement plus résistantes à l'infection par ces deux pathogènes (figure 44). La surexpression des gènes dans le mutant ne semble donc pas suffisante pour conférer une résistance aux pathogènes testés.

III) Conclusions

Au cours de ce chapitre, nous avons présenté des résultats préliminaires suggérant que les protéines MAIL2 et MAIL3 agissent dans des voies génétiques différentes de celles des protéines MAIN et MAIL1. MAIL2 est une protéine essentielle au bon développement de la plante, et elle semble être nécessaire pour l'expression correcte de nombreux gènes. La protéine *MAIL3*, quant à elle, est une protéine à part dans la famille PMD-C. C'est la seule protéine de la famille portant un autre PPP en plus de son domaine PMD, et ne présentant pas de phénotype développemental. Dans des mutants *mail3-1* et *mail3-2*, des gènes de réponse aux pathogènes sont surexprimés, mais cette surexpression ne semble pas conférer une meilleure résistance aux pathogènes que nous avons testés. Les rôles et le mode d'action de ces protéines restent encore à élucider et des approches similaires à celles utilisées pour MAIN et MAIL1 pourront être envisagées.

DISCUSSION

I) La formation du complexe MAIN/MAIL1/PP7L

Nos expériences d'immunoprécipitation suivie de spectrométrie de masse (IP-MS) et de co-immunoprécipitation (co-IP) ont permis de montrer des interactions entre les protéines MAIN, MAIL1 et PP7L (Nicolau *et al.*, 2020). Aux vues de ces interactions, nous supposons que ces protéines forment un complexe protéique dont il faudra valider l'existence grâce à des expériences de chromatographie par filtration sur gel. Ces expériences permettront également de déterminer plus finement la stœchiométrie du complexe protéique, grâce à une séparation des protéines en fonction de leur poids moléculaire. La taille de chaque protéine étant connue, on pourra alors déduire la composition du complexe.

Au-delà de la formation d'un complexe, il est important de comprendre la façon dont ces protéines interagissent entre elles. Dans un premier temps, nous pourrions étudier le rôle de PP7L dans la formation du complexe, en introduisant des versions étiquetées de MAIN et MAIL1 dans un mutant *pp7l*. Par des expériences de Co-IP et d'immunofluorescence, ceci permettra de déterminer si PP7L est nécessaire pour l'interaction entre MAIN et MAIL1 et pour leur localisation nucléaire. Nous pourrions également tester par des approches similaires si MAIN et MAIL1 sont réciproquement nécessaires pour leur interaction avec PP7L.

Dans un deuxième temps, il semble important d'identifier les régions protéiques et acides aminés essentiels à la formation et/ou l'activité du complexe. Des résultats récemment publiés suggèrent que la région C-terminale (C-ter) de MAIL1 semble être essentielle pour son interaction avec PP7L (de Luxán-Hernández *et al.*, 2019). La région C-ter de MAIL1 est une région désorganisée, portant un NLS (Ühlken *et al.*, 2014). Dans leur analyse les auteurs ont utilisé une version de MAIL1 tronquée pour sa région C-ter (MAIL1 Δ C), donc sans son NLS et ainsi probablement exclue du noyau. Par conséquent la perte d'interaction entre MAIL1 Δ C et PP7L suggère que cette interaction est exclusivement nucléaire. Ceci devrait néanmoins être confirmé en réalisant une version mutante MAIL1 Δ NLS qui serait uniquement dépourvue de son NLS.

En parallèle, il sera important de déterminer la fonction de plusieurs acides aminés essentiels et conservés au sein du PMD de MAIN et MAIL1. Par exemple, les mutants hypomorphes *main-3* et *mail1-2* sont respectivement mutés au niveau de la

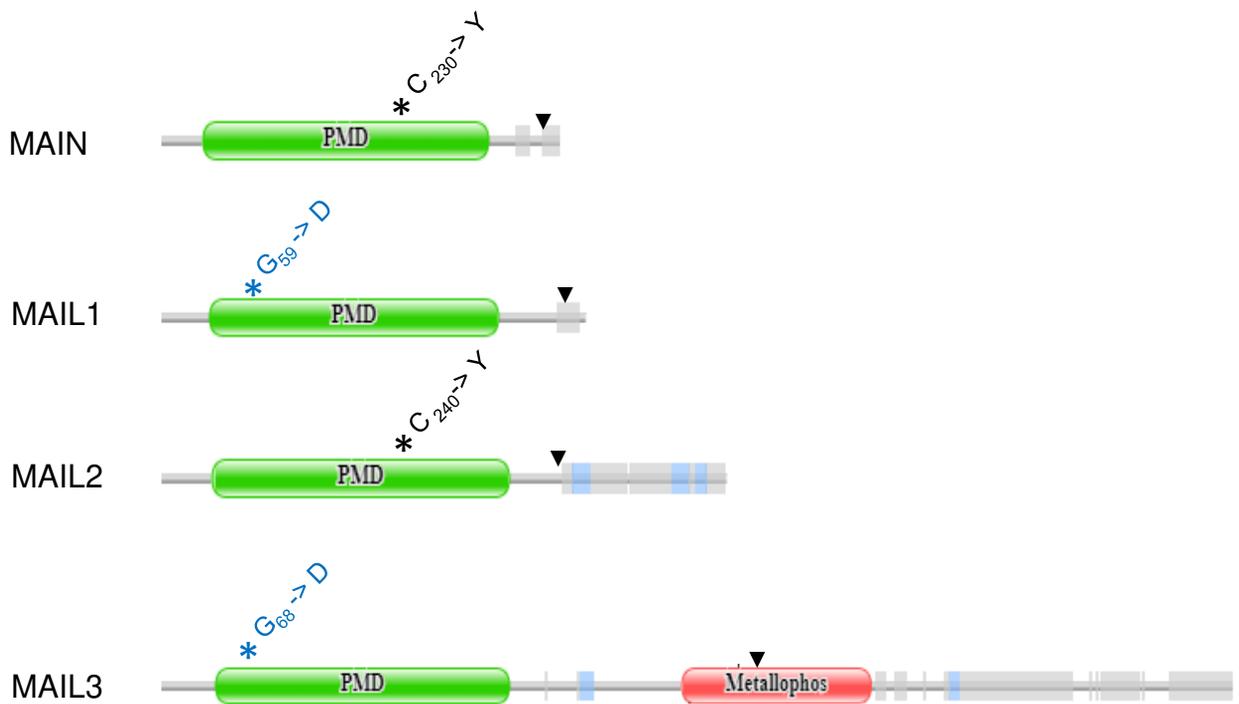


Figure 45 : Mutations ponctuelles générées sur les protéines PMD-C

Représentations des mutations ponctuelles générées sur les protéines PMD-C en se basant sur les positions d'acides aminés conservés. C : cystéine, Y : tyrosine, G : glycine, D : acide aspartique. Les représentations schématiques des protéines sont issues du site pfam, et les rectangles bleus et gris représentent des régions désorganisées de la protéine. Les triangles noirs représentent le NLS.

cystéine 230 (C230) et de la glycine 59 (G59) (Ikeda *et al.*, 2017; Nicolau *et al.* 2020). Ces deux acides aminés conservés sont essentiels pour la répression des TEs et la régulation de l'expression génique (Ikeda *et al.*, 2017; Nicolau *et al.* 2020). Il apparaît donc important de déterminer si ces deux acides aminés sont par exemple nécessaires pour l'interaction MAIN/MAIL1/PP7L. Des lignées exprimant des versions mutantes MAIN C230Y ou MAIL1 G59D, étiquetées -FLAG ou -MYC ont été générées afin de tester cette hypothèse grâce à des expériences d'IP-MS et de Co-IP (figure 45).

Bien que des analyses de prédiction de structure aient montré que le PMD partagerait des similitudes avec des protéines de type « transférase », sa structure tridimensionnelle reste à ce jour non résolue. Connaître la structure tridimensionnelle du PMD nous aidera sûrement à comprendre l'activité de ces protéines au sein de la cellule. Une collaboration avec l'équipe de Jiamu Du (SUSTech, Chine) a été initiée afin de résoudre la structure du PMD par des approches de cristallographie. Ceci nous permettra aussi de positionner les acides aminés essentiels tels que C230 de MAIN et G59 de MAIL1.

Grâce à des approches de biochimie et de phylogénie, nous avons pu décrire chez *A. thaliana*, les associations des domaines PMD et PPP en *trans* à travers l'interaction MAIN/MAIL1/PP7L, mais aussi en *cis* comme c'est le cas pour la protéine MAIL3 (Nicolau *et al.*, 2020). L'association en *cis* de type MAIL3 est, en outre, très conservée chez les Brassicaceae, ce qui suggère une sélection positive suite à une potentielle néo-fonctionnalisation de cette association au sein de cette famille. Il est aussi important de noter qu'une association indépendante PMD/PPP en *cis* est également retrouvée au moins chez la vigne avec la protéine VvMAIL2.2 (Nicolau *et al.*, 2020). Ces associations des domaines PMD et PPP en *trans* et/ou en *cis* rappellent le concept des protéines chimères « Rosetta stone » qui décrit que lorsque deux protéines interagissent ensemble pour former un complexe enzymatique au sein d'un organisme, il est possible de les retrouver sous forme de domaines protéiques fusionnés, ceci afin de former une unique protéine et ainsi optimiser son activité enzymatique (Galperin and Koonin, 2000). Dans le cas de l'association en *cis* PMD/PPP, cette fusion pourrait être une stratégie pour optimiser l'activité enzymatique des protéines MAIL3 ou VvMAIL2.2. Ainsi, dans notre modèle, les associations PMD/PPP en *cis* ou en *trans* permettrait la formation d'un module fonctionnel dont l'activité pourrait être régulée de manière différentielle (figure 46).

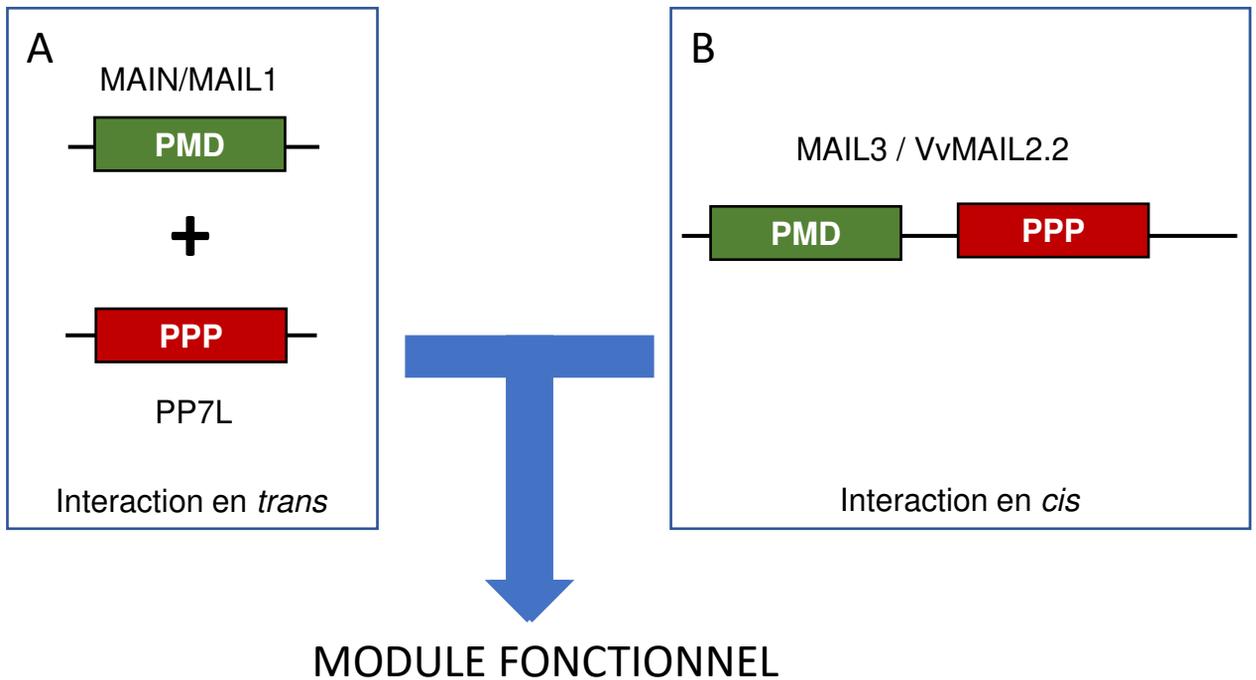


Figure 46 : Modèles hypothétiques décrivant la formation d'un module fonctionnel entre les domaines PMD et PPP.

Deux scénarios aboutissant à la formation d'un module fonctionnel PMD/PPP à travers l'interaction en *trans* (A) des protéines MAIN, MAIL1 et PP7L ou en *cis* (B) des domaines protéiques PMD et PPP au sein de MAIL3 ou VvMAIL2.2.

II) Les protéines MAIN, MAIL1 et PP7L régulent un même réseau de gènes

Les analyses transcriptomiques des mutants *main*, *mail1* et *pp7l* (incluant le double mutant *mail1 pp7l*) ont démontré que les protéines MAIN, MAIL1 et PP7L sont essentielles pour la bonne expression de gènes communs et la répression de nombreux TEs (Nicolau *et al.*, 2020).

Parmi les gènes communément dérégulés dans ces mutants, deux gènes nous ont particulièrement intéressés : *MORC1* et *FLC*. La protéine MORC1 est nécessaire pour la répression des TEs (Moissiard *et al.*, 2012, 2014), et nos résultats préliminaires suggèrent que sa répression dans le mutant *mail1-1* (et potentiellement *main-2*, et *pp7l-2*) pourrait être responsable de la surexpression des TEs que l'on observe (figure 36). Dans le futur, des analyses de RNA-seq seront réalisées sur les lignées pUBQ :: MORC1-FLAG *main-2* et *mail1-1*, ainsi que sur des lignées doubles mutantes *main-2 morc1-2* et *mail1-1 morc1-2* afin de déterminer précisément les interactions génétiques entre les voies MAIN/MAIL1/PP7L et MORC1. Dans les lignées doubles mutantes *main-2 morc1-2* et *mail1-1 morc1-2*, suivant notre hypothèse, nous nous attendons, théoriquement, à retrouver une proportion importante de TEs communément surexprimés dans les simples mutants *main-2*, *mail1-1* et *morc1-2*. Toutefois, il est important de noter que dans les mutants *main-2* et *mail1-1* le gène *MORC1* n'est pas totalement éteint (environ 20% d'expression en comparaison au niveau sauvage, figure 35). A contrario, l'allèle *morc1-2* utilisé est un mutant KO d'insertion d'ADN-T. Les mutants simples et doubles portant cet allèle présentent donc un niveau d'expression de *MORC1* qui est « nul ». Cette différence d'expression entre les différentes lignées mutantes pourrait donc conduire à des variations notables en termes de nombre de TEs dérégulés dans les doubles mutants en comparaison aux simples mutants *main-2* et *mail1-1*. De plus, il pourrait être envisageable d'observer des effets synergétiques sur les niveaux d'expression de certains TEs en combinant les mutations *main*, *mail1* et *morc1-2*.

FLC est surexprimé dans les mutants *main*, *mail1* et *pp7l* et nous observons un retard de floraison chez ces mutants. Nous avons donc émis l'hypothèse que ce retard de floraison pourrait être dû à la surexpression de *FLC*. Toutefois, des résultats de comptage de feuille de rosette à floraison sur un mutant *pp7l-1* montrent que ce mutant

ne présente pas le phénotype dit de « retard de floraison » (de Luxán-Hernández *et al.*, 2019). Ainsi, la surexpression de FLC, observée dans des plantes mutantes *main-2*, *mail1-1* et *pp7l-2* âgées de trois semaines en comparaison à la plante sauvage, pourrait être simplement la conséquence d'un retard de croissance des mutants. Celui-ci pourrait être dû à un défaut du plastochron des mutants qui se développeraient alors plus lentement que les plantes sauvages. Des analyses complémentaires seront nécessaires pour confirmer cette hypothèse.

En plus de *MORC1* et *FLC*, nous avons également identifié d'autres gènes, dérégulés dans les mutants *main*, *mail1* et *pp7l*, qui ont attiré notre attention.

Tout d'abord parmi les gènes réprimés, le gène *HISTONE ACETYLTRANSFERASE OF THE CBP FAMILY 4 (HAC4)*. HAC4 fait partie d'une famille multigénique, composée de cinq membres (Li *et al.*, 2014). Cette famille est homologue à la famille des gènes CREB-binding protein (CBP)/p300 des mammifères. Les protéines CBP/p300 sont capables d'agir comme des cofacteurs de la transcription, en permettant l'interaction entre des facteurs de transcription généraux et la machinerie transcriptionnelle (Eckner *et al.*, 1994; Goodman and Smolik, 2000). Elles portent également une activité HAT, et sont aussi capables d'acétyler des protéines non-histones (Manning *et al.*, 2001; Fu *et al.*, 2004). Ces caractéristiques font des protéines HAC de potentiels régulateurs de la transcription. Une plante *hac4* ne montre pas de phénotype développemental anormal, en revanche *hac4* en combinaison avec des mutations *hac1* et *hac5* atténue partiellement les phénotypes mutants *hac1 hac5*, HAC4 jouerait donc un rôle antagoniste à HAC1 et HAC5 (Li *et al.*, 2014).

FAR1-RELATED SEQUENCES-RELATED FACTOR 2 (FRF2) est également réprimé dans les mutants *main*, *mail1* et *pp7l*. FRF2 fait partie d'une famille multigénique, composée de quatre membres. Comme évoqué en introduction, les protéines FRF portent un domaine de liaison à l'ADN, similaire à celui du facteur de transcription FAR1 (Aguilar-Martínez *et al.*, 2015). La protéine FRF2 est donc potentiellement capable de lier l'ADN. Tout comme un mutant *hac4*. Une plante mutante *frf2* ne présente pas de phénotype développemental anormal (Aguilar-Martínez *et al.*, 2015). Les gènes *HAC4* et *FRF2* ont donc attiré notre attention du fait de leur potentiel rôle dans le contrôle de la transcription. Cependant, étant donné que les mutants *hac4* et *frf2* ne montrent pas de phénotype développemental anormal et qu'il n'existe aucune donnée transcriptomique sur ces mutants, établir des connexions entre les voies MAIN/MAIL1/PP7L, HAC4 ou FRF2 reste difficile. En outre, nous pouvons très

probablement exclure l'hypothèse selon laquelle la répression de HAC4 ou FRF2 dans les mutants *main*, *mail1* et *pp7l* jouerait un rôle important dans la mise en place des phénotypes observés.

Parmi les gènes surexprimés, deux gènes ont également attiré notre attention : *Taf4b* qui a déjà été décrit dans le chapitre 2 des résultats et *SHORTAGE IN CHIASMATA 1* (*SHOC1*). Chez *A. thaliana*, la protéine SHOC1 qui est homologue de l'endonucléase XPF, est impliquée dans la formation des crossing over (CO) de type I au cours de la méiose (Macaisne *et al.*, 2008, 2011). Dans un mutant *shoc1* le taux de CO est réduit, ce qui entraîne une diminution de la fertilité des plantes. Il apparaît donc remarquable que les gènes *Taf4b* et *SHOC1* qui sont tous deux impliqués dans le processus de méiose (Macaisne *et al.*, 2008, 2011; Lawrence *et al.*, 2019), soient surexprimés dans les tissus somatiques des mutants *main*, *mail1* et *pp7l* (Nicolau *et al.*, 2020). Il pourrait donc être intéressant de tester si les mutants *main*, *mail1* et *pp7l* présentent des défauts de division cellulaire (mitose) au sein de leurs méristèmes racinaire et caulinaire qui seraient associés à la présence de CO sporadiques, ou bien un défaut de CO au cours de la méiose. Ceci est d'autant plus important étant donné que le mutant *main* a une fertilité diminuée (Wenig *et al.*, 2013). Sachant que *SHOC1* est impliqué dans la fertilité de la plante, il peut être ainsi intéressant de tester la contribution de *SHOC1* dans la diminution de la fertilité du mutant *main*, en produisant par exemple des double mutants *main shoc1*. Nous pourrions également tester si la surexpression de *Taf4b* est corrélée aux phénotypes développementaux des mutants *main*, *mail1* et *pp7l* en produisant des doubles mutants *main taf4b*, *mail1 taf4b* et *pp7l taf4b*.

III) Le mode d'action du complexe MAIN/MAIL1/PP7L

Suite à l'identification d'un réseau de gènes communs dérégulés dans les mutants *main*, *mail1* et *pp7l*, et à la présence du motif d'ADN « DOWN motif » dans le promoteur d'une fraction significative des gènes réprimés, nous avons émis deux hypothèses principales : 1) le complexe MAIN/MAIL1/PP7L régulerait l'expression des gènes contenant le « DOWN motif » en interagissant directement avec la chromatine. 2) le complexe protéique régulerait ces gènes grâce au recrutement d'un FT inconnu. J'ai testé l'interaction de MAIN et MAIL1 avec la chromatine en réalisant des fractionnements cellulaires. Ces expériences ont montré que les deux protéines sont enrichies dans la fraction chromatinienne, ce qui suggère qu'elles sont

capables d'interagir avec la chromatine de façon directe ou non. Des expériences de ChIP-seq seront réalisées afin de confirmer ces résultats, mais aussi pour identifier les régions génomiques avec lesquelles se lient les deux protéines. Plus particulièrement, des approches de ChIP-qPCR testeront si MAIN/MAIL1/PP7L se lient aux régions promotrices portant le motif « DOWN », telles que le promoteur de MORC1.

En parallèle, des protéines HaloTag-MAIN, HaloTag-MAIL1 et HaloTag-PP7L sont en cours de production. Ces protéines seront utilisées pour tester de façon non-biaisée si elles sont capables de se lier directement à l'ADN en utilisant des approches de « DNA Affinity Purification sequencing » (DAP-seq) ou de « Systematic Evolution of Ligands by EXponential enrichment » (SELEX) (Stoltenburg *et al.*, 2007; Bartlett *et al.*, 2017). Ces deux approches nous permettront potentiellement d'identifier les séquences d'ADN préférentiellement liées par des protéines. Les résultats obtenus seront ensuite validés par des expériences de retard sur gel (EMSA). L'interaction des protéines avec le motif « DOWN » pourra également être testée directement par des expériences d'EMSA. Nous utiliserons également les protéines étiquetées HaloTag pour tester *in vitro* si elles sont capables d'interagir avec des histones portant différentes MPTs ou avec les nucléosomes, en utilisant des kits commerciaux ou des protocoles déjà publiés (Luger *et al.*, 1999; Shim *et al.*, 2012). Ces différentes approches nous permettront de comprendre l'interaction du complexe MAIN/MAIL1/PP7L avec la chromatine.

En parallèle nous pourrions tester de façon non-biaisée quelles protéines sont capables de se lier sur le « DOWN motif » par des analyses de « pull down ». Pour cela des sondes biotinylées portant le « DOWN motif », ou un motif aléatoire, pourront être utilisées et les protéines immunoprécipitées seront identifiées par des analyses MS.

Les protéines MAIN, MAIL1 et PP7L portent toutes un NLS, en C-ter pour MAIN et MAIL1, et en N-ter pour PP7L. De plus, des expériences de microscopie effectuées en cellule de tabac et protoplastes d'*A. thaliana* surexprimant les protéines fusionnées en C-ter à la GFP ont décrit MAIN-GFP et MAIL1-GFP comme des protéines nucléaires, et PP7L-GFP comme une protéine à la fois nucléaire et cytoplasmique (Wenig *et al.*, 2013 ; Ühlken *et al.*, 2014 ; Xu *et al.*, 2019).

De mon côté, par des expériences de fractionnement cellulaire, j'ai pu constater que les protéines MAIN-3xFLAG et MAIL1-3xFLAG sont majoritairement présentes dans le noyau, avec cependant une fraction minoritaire qui serait cytoplasmique -du moins pour MAIN, ceci restant à confirmer pour MAIL1 (figure 40). Pour expliquer cette contradiction entre mes résultats et ceux publiés, je propose deux hypothèses. La première, la moins parcimonieuse, serait que, lorsque les protéines sont surexprimées, la quantité importante de protéines produites dans le noyau puisse saturer le signal, rendant ainsi la détection cytoplasmique

difficile. La deuxième hypothèse se base sur l'utilisation de la GFP et sur la position des différents NLS au sein de MAIN, MAIL1 et PP7L. La GFP est une étiquette beaucoup plus grande que le 3xFLAG que j'ai utilisé (26,9 kilodaltons contre environ 10 kilodaltons). Dans le cas de MAIN et MAIL1, la GFP située à proximité de leur NLS pourrait favoriser l'exposition et la reconnaissance des NLS, ce qui aurait pour effet de maintenir les constructions MAIN-GFP et MAIL1-GFP dans le noyau. A l'inverse, pour PP7L dont le NLS est en N-ter, la GFP située en C-ter n'aurait pas ou peu d'effet, ce qui se traduirait par la détection de PP7L-GFP à la fois dans le noyau et le cytoplasme. Il est en outre intéressant de noter que des expériences de « bimolecular fluorescence complementation » (BiFC), où des portions de protéines fluorescentes ont été fusionnées à MAIN, MAIL1 ou PP7L, montrent que le complexe MAIN/MAIL1 est uniquement nucléaire, alors que les complexes MAIN/PP7L et MAIL1/PP7L, bien que majoritairement nucléaires, sont aussi cytoplasmiques (de Luxán-Hernández *et al.*, 2019). Ceci suggère donc, comme observé par fractionnement cellulaire, que la protéine MAIN -et très probablement aussi MAIL1- peuvent aussi être minoritairement localisé dans le cytoplasme. Il sera important dans l'avenir d'analyser le rôle de PP7L dans cette localisation subcellulaire et aussi de déterminer si ce processus est un moyen de réguler l'activité des protéines MAIN et MAIL1.

IV) Interaction génétique et rôle des protéines PP7L et MAIL3

PP7L est décrite comme une phosphatase non-fonctionnelle (Farkas *et al.*, 2007), il a cependant été montré que cette protéine joue un rôle important dans le développement des chloroplastes et dans la régulation de la transcription (Xu *et al.*, 2019; de Luxán-Hernández *et al.*, 2019; Nicolau *et al.*, 2020). A l'inverse, MAIL3 possède un domaine PPP potentiellement fonctionnel (Farkas *et al.*, 2007). Cependant, dans les deux cas, ces analyses *in silico* n'ont pas été confirmées expérimentalement. Dans le futur, il sera important de tester les activités enzymatiques de PP7L et MAIL3 par des expériences *in vitro*. Concernant PP7L, nous pourrions envisager de complexifier PP7L avec MAIN et MAIL1 afin de tester si ces dernières stimulent son activité enzymatique. Dans le cas de résultats positifs, des expériences de phospho-protéomique sur des plantes sauvages et mutantes *pp7l* pourront être réalisées.

Bien que des acides aminés essentiels du site catalytique de PP7L soient mutés (Farkas *et al.*, 2007), il semblerait que son site de fixation du phosphate soit intègre. Il

a déjà été montré que des pseudoenzymes, des enzymes ne portant pas d'activité catalytique, jouent des rôles importants dans la cellule (Adrain and Freeman, 2012; Reiterer *et al.*, 2014). Par exemple, les pseudophosphatases, tout comme les pseudokinases sont impliquées dans de nombreux processus cellulaires. Ces protéines peuvent jouer le rôle de compétiteur ou de co-facteur pour les enzymes actives, d'ancre spatiale pour leur substrat ou encore d'intégrateur de voies de signalisation (Reiterer *et al.*, 2014). Ainsi, une hypothèse pourrait être que PP7L peut agir comme un compétiteur, notamment de MAIL3. C'est l'hypothèse qui a été testé par de Luxán-Hernández *et al.* en 2019, pour cela ils ont analysé la répression de six TEs dans un double mutant *pp7l-3 mail3-2*. Parmi ces TEs, quatre sont significativement réprimés dans le double mutant *pp7l-3 mail3-2* en comparaison au simple mutant *pp7l-3*. Suite à ce résultat, les auteurs proposent deux hypothèses : la perte de MAIL3 permet une meilleure activité répressive de MAIN et MAIL1 à ces TEs ou elle permet l'activation d'une voie répressive alternative. Cette analyse se base sur un échantillon de TEs très limité, afin de tester l'impact de MAIL3 et l'interaction entre MAIL3 et PP7L dans la répression des TEs de façon globale, des analyses RNA-seq pourraient être réalisées sur des plantes doubles mutantes *pp7l mail3* en comparaison à des simples mutants *pp7l* et *mail3*. Nous pourrions également réaliser des expériences de IP-MS de MAIL3-FLAG dans un fond *pp7l* afin de tester si MAIL3 en l'absence de PP7L est capable de se lier aux mêmes protéines et notamment MAIN et MAIL1 *via* son domaine PPP. Ainsi, ces expériences permettraient de tester si PP7L et MAIL3 agissent comme des compétiteurs.

Une autre hypothèse pour l'activité de PP7L est que cette protéine serait capable de se lier à des protéines phosphorylées, notamment à des histones phosphorylées. En effet, en plus des protéines MAIN et MAIL1, nous ne savons pas si PP7L interagit avec d'autres protéines. Les interacteurs de MAIL3 sont également inconnus. Leurs partenaires protéiques potentiels pourront être identifiés par des expériences d'IP-MS.

V) Impact du complexe MAIN/MAIL1/PP7L sur la structure de la chromatine

Suite à l'observation de TEs surexprimés dans les mutants *main*, *mail1* et *pp7l*

nous nous sommes demandé si les protéines MAIN/MAIL1/PP7L étaient nécessaires pour la compaction de la chromatine. Par observation microscopique de la marque H3K9me2 nous avons constaté que, dans nos conditions expérimentales, les CCs ne sont pas significativement décondensés dans le mutant *pp7l-2*. Une même observation avait été faite concernant les mutants *main* et *mail1* (Ikeda *et al.*, 2017). Cependant, par des expériences d'hybridation *in situ* en fluorescence (FISH) sur les répétitions d'ADN centromérique 106B, les auteurs montraient par la suite une décondensation de ces séquences d'ADN centromériques dans les mutants *main* et *mail1* (Ikeda *et al.*, 2017). Cette différence d'observation reste énigmatique et requiert des validations expérimentales. Récemment, en mesurant la taille relative des CCs au sein de noyaux, il a été suggéré que le mutant *pp7l-1* présentait aussi une décondensation des CCs (de Luxán-Hernández *et al.*, 2019). Aux vues de ces résultats contradictoires aux nôtres, il sera essentiel de réaliser des expériences de FISH dirigée contre les ADN répétés 106B ou Cen180 sur des plantes *pp7l-2* mais aussi *main-2* et *mail1-1* afin de déterminer si nous observons le même phénotype de décondensation des CCs. Si cela est confirmé par FISH, nous pourrions donc conclure que les protéines MAIN/MAIL1/PP7L sont nécessaires pour la condensation des CCs.

Il reste néanmoins important de remettre cette hypothèse dans le contexte qui est que *MORC1* est réprimé dans *main*, *mail1* et *pp7l*. En effet, Il a été montré qu'un mutant *morc1* nul présente une décondensation des CCs (Moissiard *et al.*, 2012). Ainsi, une possibilité serait que cette décondensation des CCs soit la conséquence indirecte de la répression de *MORC1*. En outre, il est nécessaire de rappeler qu'en comparaison à un mutant *morc1* nul, *MORC1* est seulement partiellement réprimé dans les mutants *main*, *mail1* et *pp7l* (~20% moins exprimé qu'une plante sauvage, figure 35). Par conséquent, ceci pourrait expliquer des variations entre les niveaux de décondensation des CCs dans les mutants *main*, *mail1* et *pp7l*, en comparaison au mutant *morc1* nul. Dans tous les cas, si nous confirmons une décondensation -même partielle- des CCs dans les mutants *main*, *mail1* et *pp7l*, il pourra être envisageable de tester les niveaux de condensation des CCs dans les lignées pUBQ::MORC1-FLAG *main* ou *mail1*, ainsi que dans les doubles mutants *main morc1*, *mail1 morc1* et *pp7l morc1*.

VI) MAIL2 est une protéine essentielle pour le développement de la plante

La mutation nulle de MAIL2 pourrait être létale après la germination chez *A. thaliana*, ce qui indiquerait que cette protéine pourrait être essentielle pour le développement de la plante. Les analyses de RNA-seq réalisées sur les plantes mutantes *mail2-1* montrent également que MAIL2 agirait dans une voie génétique indépendante des autres protéines PMD-C d'*A. thaliana*. Ces résultats ont été obtenus en utilisant un allèle ADN-T issu d'une collection de mutants. Il a déjà été montré que des réarrangements chromosomiques peuvent se produire dans ces mutants (Clark and Krysan, 2010), pouvant conduire à des phénotypes développementaux sévères. Dans le futur, nous devons donc nous assurer que les phénotypes observés dans la lignée *mail2-1* sont bien dues à la mutation de MAIL2. Pour cela, des approches de complémentation seront réalisées et d'autres allèles mutants seront analysés, notamment en utilisant le système CRISPR-Cas9. Si la létalité de la mutation de MAIL2 se confirme, cela en fera un matériel difficile à utiliser. Afin de résoudre ce problème, des versions mutantes hypomorphes pourront être produites en utilisant un système CRISPR-Cas9-cytidine base editor (CBE). Ce système pourrait par exemple nous permettre de muter la cystéine en position 240 qui est conservée entre les protéines MAIN, MAIL1 et MAIL2. Ce mutant hypomorphe pourrait nous permettre alors de potentiellement mieux comprendre le rôle de MAIL2 dans le développement de la plante.

VII) Le rôle du PMD chez les plantes

Comme nous l'avons vu tout au long de ce manuscrit avec les mutants *main*, *mail1* et *mail2*, le PMD est un domaine essentiel pour le développement d'*A. thaliana*. Les protéines MAIN, MAIL1, MAIL2 et MAIL3 jouent toutes un rôle dans la régulation de la transcription et/ou la répression des TEs. Le PMD est également largement répandu chez les angiospermes. Dans le futur, il sera sûrement intéressant d'utiliser comme fondations les résultats obtenus au cours de ma thèse afin de transposer nos connaissances sur les PMD chez des espèces d'intérêt agronomique, comme le riz ou

la tomate. Ceci permettra très certainement de mieux comprendre le rôle de ces protéines au cours du développement des plantes et potentiellement, lors de processus moléculaires complexes tels que, par exemple, la maturation des fruits. Conjointement, les nombreuses expériences proposées dans cette discussion permettront vraisemblablement de décortiquer les mécanismes moléculaires impliquant les protéines PMD dans les processus chromatinien.

MATERIEL ET METHODES

I. Matériel biologique

Lignées végétales utilisées :

Les travaux présentés dans cette thèse sont réalisés sur la plante modèle *Arabidopsis thaliana*. L'écotype sauvage Colombia (Col) et les mutants T-DNA knock-out (KO) *main-2* (GK-728H05), *mail1-1* (GK-840E05), *pp7l-2* (SALK_003071) et *morc1-2* (SAIL_893_B06) sont utilisés dans cette étude. Le mutant knock-down *main-3* (Nicolau *et al.*, 2020) est également utilisé dans notre étude. Chaque lignée complétée étudiée est introgressée dans le mutant KO correspondant. Les lignées doubles mutantes sont obtenues par croisement des simples mutants correspondants.

Conditions de culture :

En culture sur terreau, les graines sont semées sans stérilisation sur un mélange de 5 volumes de terreau pour 1 volume de vermiculite. La stratification est faite en plaçant les graines à 4°C à l'obscurité pendant 48h. La croissance se fait dans une chambre de culture Percival à une température constante de 22°C en condition de jours longs (16h de jour, 8h de nuit), avec une intensité lumineuse de 150 $\mu\text{mol.m}^{-2}.\text{s}^{-1}$ et un pourcentage d'humidité à 55% constant. Lors du cycle de reproduction, les hampes florales sont tuteurées puis ensachées. Les plantes sont ensuite placées en salle de séchage pour une période minimum de 2 semaines, et les graines sont récoltées à l'aide d'un tamis.

En culture *in vitro*, les graines sont stérilisées à l'éthanol 70% sous agitation pendant 15 minutes et rincées à l'éthanol 100% avant d'être séchées sous hotte à flux laminaire horizontal. Les graines sont semées sur du milieu MS 0,5% sucrose (4,41 g/L de Murashige and Skoog, 5 g/L sucrose et 8 g/L de plant agar, ajusté à un pH de 5,7 avec une solution de KOH) et sont stratifiées à 4°C à l'obscurité pendant 48h. Les plantules poussent dans une armoire de culture en LED en condition de jours longs (16h/8h), avec une intensité lumineuse de 120 $\mu\text{E.m}^{-2}.\text{s}^{-1}$, à une température constante de 20°C.

Génération des lignées transgéniques :

Clonage des gènes dans le vecteur de transformation :

Les clonages sont réalisés en utilisant les amorces présentées dans le tableau 1, dans la souche bactérienne Top10 d'*Escherichia coli*. Les bactéries sont cultivées à 37°C sur la nuit dans un milieu LB (Lysogeny Broth) liquide (25 g/L LB) sous agitation, ou un milieu LB agar solide (35 g/L LB agar). Pour les lignées complémentées, l'ADNg de *MAIN* (AT1G17930), *MAIL1* (AT2G25010) et *PP7L* (AT5G10900) est amplifié par PCR et inséré dans le plasmide pENTR (Invitrogen) selon les recommandations du fabricant, à chaque fois environ 1 kb en amont de la séquence codante est également amplifié comme promoteur endogène. Un site de restriction *MluI* a été modifié dans le promoteur de *MAIN* à l'aide d'un kit de mutagenèse dirigée QuickChange II (Agilent), afin de permettre la réaction de LR, sans modifier l'intégrité de la séquence du gène. Pour chaque protéine, les épitopes FLAG et Myc sont ajoutés en C-terminal comme suit : le plasmide pENTR est digéré avec l'enzyme de restriction *MluI* et le produit de digestion est recombiné dans un plasmide de destination pEG en utilisant la clonase LR (Invitrogen) selon les recommandations du fabricant (Law *et al.*, 2011). La même approche est utilisée pour les lignées surexprimant MORC1-FLAG : de l'ADNc de MORC1 (AT436290) sans codon stop est amplifié par PCR et inséré dans le plasmide pENTR. L'épitope FLAG est ensuite ajouté en C-terminal par sous-clonage en utilisant l'enzyme de restriction *AscI*. MORC1-FLAG est ensuite mobilisé dans le plasmide pDEST pUBQ10 en utilisant la clonase LR (Invitrogen).

Transformation des plantes et sélection des transformants :

Le vecteur pEG est ensuite transformé dans la souche bactérienne GV3101 d'*Agrobacterium tumefaciens* par choc thermique. Les bactéries sont cultivées à 28°C sur la nuit dans un milieu YEP (10 g/L Yeast Extract, 10 g/L bactopeptone et 4 g/L NaCl) liquide sous agitation, ou un milieu YEP agar solide (ajout de 10 g/L d'agar). Les plantes sont transformées en utilisant la méthode du « floral dip » (Clough and Bent, 1998). Les transformants sont ensuite sélectionnés par traitement à l'herbicide BASTA (glufosinate), et contrôlés par Western blot pour l'expression des protéines étiquetées. Les plantes des générations suivantes sont à nouveau sélectionnées par traitement

BASTA jusqu'à obtention de lignées homozygotes pour une insertion unique par ségrégation au cours des générations.

II. Méthodes de biologie moléculaire

Extraction de l'ADN génomique pour génotypage :

L'extraction d'ADN génomique est réalisée à partir de poudre de plantes, préalablement broyées dans de l'azote liquide. La poudre est rapidement resuspendue dans 200 µl de tampon d'extraction (25mM EDTA, 25mM NaCl, 2,5% SDS, 200mM Tris pH 8) et les échantillons sont vortexés puis incubés pendant 10 minutes à 60°C. Les échantillons sont ensuite centrifugés à vitesse maximale pendant 5 minutes puis 120 µl de surnageant sont récupérés et mélangés avec 1 volume d'isopropanol. Les échantillons sont vortexés puis centrifugés à vitesse maximale pendant 10 minutes. Le surnageant est ensuite jeté et le culot est nettoyé avec 200 µl d'éthanol 80% et centrifugé à vitesse maximale pendant 5 minutes à 4°C. Le surnageant est jeté et le culot est séché à l'air libre en incubant les tubes à 65°C pendant 10 à 15 minutes. Le culot est ensuite réhydraté dans 50 µl d'eau.

Extraction d'ADN plasmidique à partir de culture liquide :

Avant extraction, les bactéries sontensemencées dans 4 mL de milieu LB liquide additionné d'antibiotique et incubées à 37°C sous agitation sur la nuit. Le lendemain l'extraction d'ADN plasmidique est réalisé sur la totalité de la culture e utilisant le kit Wizard plus miniprep DNA purification system (Promega) selon les recommandations du fabricant. A la fin de l'extraction, l'ADN plasmidique est repris dans 50 µl d'eau DNase-free et dosé au Nanodrop (ThermoFisher Scientific).

Amplification d'ADN par Polymerase chain reaction (PCR) :

Les PCR sont réalisées à partir de 2 µl d'ADN génomique, en utilisant la polymérase GoTaq® (Promega) dans un mix constitué de 1X de tampon GoTaq, 0,2 mM dNTP, 1µM d'amorces spécifiques et 5 unités de polymérase GoTaq dans un volume final de 20 µl. La réaction de PCR est réalisée dans un thermocycleur Biometra® Tprofessional et les résultats sont analysés par électrophorèse dans un gel d'agarose 1%, additionné de GelRed® (Biotium). Les amorces utilisées sont présentées dans le tableau 1.

Extraction de l'ARN :

L'extraction d'ARN est réalisée à partir de poudre de plantes, préalablement broyées dans de l'azote liquide, en utilisant le kit Monarch Total RNA Miniprep kit (New England Biolabs) selon les recommandations du fabricant. A la fin de l'extraction, les ARN sont repris dans 100 µl d'eau RNase-free et dosés au Nanodrop (ThermoFisher Scientific).

Transcription inverse :

La transcription inverse des ARN en ADNc est réalisée avec le kit GoScript Reverse Transcriptase (Promega) selon les recommandations du fabricant, à partir de 1 µg d'ARN. Les ADNc sont dilués 6 fois avant l'étape de PCR quantitative.

PCR quantitative (qPCR) :

Les qPCR sont réalisées à partir de 4 µl d'ADNc, en utilisant le Takyon No Rox SYBR MasterMix (Eurogentec). Les amorces utilisées sont présentées dans le tableau 1. Les qPCR sont réalisées dans un LightCycler 480 (Roche), et les résultats sont analysés en utilisant la méthode $2^{-\Delta\Delta Ct}$: Le ΔCt pour chaque échantillon est calculé par rapport aux valeurs Ct du gène RHIP1 (AT4G26410) et le $\Delta\Delta Ct$ est calculé en soustrayant les ΔCt de chaque échantillon par le ΔCt des échantillons sauvages. Trois réplicats biologiques sont utilisés par expérience.

Séquençage de l'ARN messager :

Les banques de séquençage ont été générées à partir de 1 µg d'ARN en utilisant le kit NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, E7490) selon les recommandations du fabricant. Les banques ont été séquencées sur une machine NextSeq 550 (Illumina).

Production *in vitro* et purification de protéines HALOtag :

La production de protéines *in vitro* est réalisée avec le kit TnT® SP6 High-Yield Wheat Germ protein expression system (Promega) : environ 1 µg de plasmide pIX-HALO-MAIN PMD ou pIX-HALO-TGA5 est incubé avec 15 µl de TnT® SP6 High-Yield Wheat Germ Master Mix pendant 2 heures à 25°C. Après l'incubation, 1 µl d'échantillon est récupéré pour analyses par western blot. La purification des protéines est ensuite

réalisée avec des billes Magne® HaloTag® (Promega). 50 µl de billes sont lavés deux fois avec 100 µl de PBS/IGEPAL CA-630 0.005% en rotation douce, puis sont resuspendus dans 140 µl du même tampon. Le volume de billes est ensuite divisé en deux entre les échantillons pIX-HALO-MAIN PMD et pIX-HALO-TGA5, soit 70 µl de billes par échantillon. Le mélange est incubé pendant 1 heure à 25°C et après la réaction, 1 µl du surnageant est récupéré pour analyses par western blot. Les billes sont lavées trois fois avec 200 µl de PBS/IGEPAL CA-630 0.005% en rotation douce et elles sont ensuite resuspendues dans 100 µl de PBS et 10 µl de tampon Laemmli pour analyses par western blot.

II. Méthodes biochimiques

Fractionnement cellulaire :

Les expériences de fractionnement cellulaire sont réalisées à partir de 1 g plantules cultivées *in vitro* âgées de 12 jours. Les plantes sont broyées dans un mortier avec de l'azote liquide, et la poudre est resuspendue dans du tampon Honda (25 mM Tris-HCl pH 7,5, 400 mM sucrose, 2,5% Ficoll, 5% Dextran T40, 10 mM MgCl₂, 0,5% Triton X 100, 0,5 mM PMSF, 10 mM β-mercaptoéthanol, cocktail d'inhibiteur de protéase [Roche]). Les échantillons sont ensuite filtrés 2 fois à travers une double couche de Miracloth (Merck Millipore), et la fraction totale est prélevée. Après une étape de centrifugation à 10 000g, le surnageant est prélevé en tant que fraction cytoplasmique, et le culot est lavé 3 fois dans du tampon Honda et 1 fois dans du tampon PBS (2 mM KH₂PO₄ pH 7,4, 1 mM EDTA, 137 mM NaCl, 2,7 mM KCl, 10 mM Na₂HPO₄). Le culot est ensuite resuspendu dans du tampon glycérol (20 mM Tris-Hcl pH 8,0, 75 mM NaCl, 0,5 mM EDTA, 50 % glycérol, 0,125 mM PMSF, 0,85 mM DTT, 10 mM β-mercaptoéthanol, cocktail d'inhibiteur de protéase [Roche]) et du tampon de lyse nucléaire (10 mM HEPES pH 7,6, 7,5 mM MgCl₂, 0,2 mM EDTA, 300 mM NaCl, 1 M urée, 1% NP-40, 1 mM DTT, 0,5 mM PMSF, 10 mM β-mercaptoéthanol, cocktail d'inhibiteur de protéase [Roche]) à volumes égaux. Les échantillons sont incubés 5 minutes sur glace, et centrifugés à 13 200g pendant 2 minutes. Le surnageant est ensuite prélevé en tant que fraction nucléoplasmique, et le culot est lavé 1 fois avec du tampon PBS, avant d'être resuspendu dans le même tampon et conservé en tant que fraction chromatiniennne.

Dosage des protéines à l'Amido black :

Le dosage des protéines est réalisé à température ambiante, 5 µl d'échantillon sont ajusté à 200 µl avec de l'eau, puis 800 µl de solution amido black (10% acide acétique, 90% méthanol, 0,1% amido black) est ajouté. L'échantillon est vortexé et incubé pendant 2 minutes. Il est ensuite centrifugé à vitesse max pendant 15 minutes. Le surnageant est jeté et le culot est lavé avec 1 ml de tampon de lavage (10% acide acétique, 90% éthanol), en inversant le tube deux fois. L'échantillon est de nouveau centrifugé à vitesse maximale pendant 15 minutes. Le surnageant est jeté et le culot est laissé à sécher à l'air libre. Une fois le culot séché, il est repris dans 1 ml d'une solution à 0,2 N NaOH. L'absorbance de l'échantillon est mesurée dans un spectrophotomètre (Shimadzu) à 620 nm, et la concentration de l'échantillon est déterminé à l'aide d'une courbe étalon.

Western blot :

Les expériences de Western blot sont réalisées avec des gels dénaturants SDS-PAGE, composé de 2 couches : Une couche dite « stacking » à une concentration de 5% d'Acrylamide (29 :1 Acrylamide/ Bisacrylamide 30%, Euromedex), qui permet la concentration des protéines avant leur séparation par la deuxième couche dite « running » dont la concentration en acrylamide varie entre 10% et 15% en fonction des protéines étudiées. Avant dépôt sur gel, les protéines sont diluées dans du tampon Laemli 2X (100 mM Tris pH 6,8, 200 mM DTT, 4% SDS, 0,2% Bleu de bromophénol, 20% glycérol) et chauffées à 95°C pendant 5 minutes pour les dénaturer. Les protéines sont séparées par un courant de 100 V constant et un marqueur de poids moléculaire coloré PageRuler Prestained Protein Ladder Plus (Thermo Scientific) est déposé sur gel afin d'estimer le poids moléculaire des protéines. La migration se fait dans un tampon contenant 25 mM Tris, 2 mM Glycine et 0,1% SDS. Le temps de migration dépend de la taille des protéines étudiées et varie entre 2 et 3 heures.

Après la migration, les protéines sont transférées sur une membrane de nitrocellulose 0,45 µm (Merck Millipore). Les membranes sont activées par une incubation dans de l'éthanol 100% pendant quelques secondes. Le transfert des protéines est réalisé en milieu liquide, avec un tampon contenant 25 mM Tris, 2 mM Glycine et 20% éthanol, pendant 1h30 à 4°C.

Une fois le transfert effectué, les membranes sont saturées dans une solution de PBS, 0,1% Tween20, 5% lait pendant minimum 1 heure à température ambiante. Des

anticorps polyclonaux de lapin dirigés contre H3 (1 :10 000), AGO4 (1 :3 000) et UGPase (1 :5 000) et un anticorps polyclonal de rat dirigé contre l'épitope Myc (1 :1 000) sont utilisés. Des anticorps de chèvre anti-lapin (1 :15 000) ou anti-rat (1 :10 000) conjugués à la HRP ont été utilisés comme anticorps secondaires. Enfin, un anticorps monoclonal commercial dirigé contre l'épitope FLAG déjà conjugué à la HRP est utilisé à une dilution de 1 :10 000. Les membranes sont incubées avec les anticorps primaires sur la nuit à 4°C sous agitation, et avec les anticorps secondaires pendant 45 minutes à température ambiante sous agitation. La révélation est réalisée avec des réactifs de chimiluminescence (Immobilon Western chimiluminescent HRP Substrate, Merck Milipore), et la lecture est réalisée sur un instrument Vilber fusion solo S (Vilber).

IV. Méthodes de microscopie :

Visualisation de noyaux et immunolocalisation de protéines nucléaires au microscope confocal :

L'immunofluorescence (IF) est réalisée sur des feuilles de plantes âgées d'environ 2 semaines, avant la croissance de la tige. Les feuilles sont tout d'abord fixées dans du tampon Tris froid (10 mM Tris-HCl pH 7,5, 10 mM EDTA, 100 mM NaCl) contenant 4% de formaldéhyde, pendant 20 minutes. Elles sont ensuite lavées 2 fois dans du tampon Tris froid, pendant 10 minutes. Les feuilles sont hachées finement avec une lame de rasoir dans du tampon LB01 (15 mM Tris-HCl pH 7,5, 2 mM EDTA, 0,5 mM spermine, 80 mM KCl, 20 mM NaCl, 0,1% Triton X-100), puis les échantillons sont filtrés à travers un filtre de 30 µm (CellTrics, Sysmex), et du tampon de tri (100 mM Tris-HCl pH 7,5, 50 mM KCl, 2 mM MgCl₂, 0,05% Tween20, 5% saccharose) est ajouté. Les échantillons sont ensuite déposés sur une lame en polysine et séchés à température ambiante. Les échantillons sont fixés sur la lame en les incubant avec du tampon PBS (137 mM NaCl, 2,7 mM KCl, 10 mM Na₂HPO₄, 1,8 mM KH₂PO₄) contenant 2% de formaldéhyde pendant 5 minutes, ils sont ensuite rincés 3 fois avec du tampon PBS et sont séchés à température ambiante.

Les lames sont ensuite incubées avec du tampon PBS contenant 0,5% de Triton X 100, pendant 15 minutes à température ambiante dans une chambre humide. Les lames sont lavées 3 fois avec du tampon PBS pendant 5 minutes, puis les lames sont incubées avec des anticorps primaires sur la nuit à 4°C dans une chambre humide à l'obscurité. Les anticorps sont dilués dans du tampon PBS contenant 3% d'albumine

de sérum bovine (BSA) et 0,05% de Tween20. Un anticorps de rat dirigé contre l'épitope Myc (1 :500, Chromotek) est utilisé. Le lendemain, les lames sont lavées 3 fois avec du tampon PBS à température ambiante, puis sont incubées pendant 2 à 3 heures à température ambiante dans une chambre humide à l'obscurité, avec un anticorps secondaire anti-rat (1 :1 000) couplé à un fluorochrome Alexa Fluor 488 (Invitrogen) émettant dans le vert. L'anticorps est dilué dans du tampon PBS contenant 3% d'albumine de sérum bovine (BSA) et 0,05% de Tween20. Les lames sont ensuite lavées dans du tampon PBS, puis dans du tampon PBS additionné de 1 µg/mL de DAPI (4',6-diamidino-2-phénylindole), avant un dernier lavage dans du tampon PBS, à chaque fois pendant 5 minutes à température ambiante. Les lames sont montées avec du milieu Vectashield (Vector Laboratories) contenant 1 µg/mL de DAPI. Les observations et les images sont réalisées avec un microscope confocal à balayage laser Zeiss LSM (Carl Zeiss) et le logiciel ZEN (Carl Zeiss).

BIBLIOGRAPHIE

Adrain C, Freeman M. 2012. New lives for old: Evolution of pseudoenzyme function illustrated by iRhoms. *Nature Reviews Molecular Cell Biology* **13**, 489–498.

Aguilar-Martínez JA, Uchida N, Townsley B, West DA, Yanez A, Lynn N, Kimura S, Sinha N. 2015. Transcriptional, posttranscriptional, and posttranslational regulation of SHOOT MERISTEMLESS gene expression in arabidopsis determines gene function in the shoot apex. *Plant Physiology* **167**, 424–442.

Ahn CS, Han JA, Lee HS, Lee S, Pai HS. 2011. The PP2A regulatory subunit Tap46, a component of the TOR signaling pathway, modulates growth and metabolism in plants. *Plant Cell* **23**, 185–209.

Ali A, Zhang J, Bao S, Liu I, Otterness D, Dean NM, Abraham RT, Wang XF. 2004. Requirement of protein phosphatase 5 in DNA-damage-induced ATM activation. *Genes and Development* **18**, 249–254.

Amedeo P, Habu Y, Afsar K, Mittelsten Scheid O, Paszkowski J. 2000. Disruption of the plant gene MOM releases transcriptional silencing of methylated genes. *Nature* **405**, 203–206.

Arand J, Spieler D, Karius T, et al. 2012. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genetics* **8**.

Arents G, Burlingame RW, Wang BIC, Love WE, Moudrianakis EN. 1991. The nucleosomal core histone octamer at 3.1 Å resolution: A tripartite protein assembly and a left-handed superhelix. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 10148–10152.

Ausin I, Feng S, Yu C, et al. 2016. DNA methylome of the 20-gigabase Norway spruce genome. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E8106–E8113.

Babu MM, Iyer L, Balaji S, Aravind L. 2006. The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Research* **34**, 6505–6520.

Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Research* **21**, 381–395.

Bartee L, Malagnac F, Bender J. 2001. Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes and Development* **15**, 1753–1758.

Bartlett A, O'Malley RC, Huang SSC, Galli M, Nery JR, Gallavotti A, Ecker JR. 2017. Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nature Protocols* **12**, 1659–1672.

Batista RA, Köhler C. 2020. Genomic imprinting in plants-revisiting existing models. *Genes & development* **34**, 24–36.

Benoit M, Simon L, Desset S, Duc C, Cotterell S, Poulet A, Le Goff S, Tatout C, Probst A V. 2019. Replication-coupled histone H3.1 deposition determines nucleosome composition and heterochromatin dynamics during Arabidopsis seedling development. *New Phytologist* **221**, 385–398.

Bestor T, Laudano A, Mattaliano R, Ingram V. 1988. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of Molecular Biology* **203**, 971–983.

Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**, 91–99.

Black BE, Cleveland DW. 2011. Epigenetic Centromere Propagation and the Nature of CENP-A Nucleosomes. *Cell* **144**, 471–479.

Blakeslee JJ, Zhou HW, Heath JT, Skottke KR, Rodriguez Barrios JA, Liu SY, DeLong A. 2008. Specificity of RCN1-mediated protein phosphatase 2A regulation in meristem organization and stress response in roots. *Plant Physiology* **146**, 539–553.

Bochtler M, Kolano A, Xu GL. 2017. DNA demethylation pathways: Additional players and regulators. *BioEssays* **39**, 1–13.

Borg M, Jacob Y, Susaki D, et al. 2020. Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin. *Nature Cell Biology* **22**, 621–629.

Bostick M, Kim JK, Estève P-O, Clark A, Pradhan S, Jacobsen SE. 2007. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science* **317**.

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. 2001. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536–2539.

Brestovitsky A, Ezer D, Waidmann S, Maslen SL, Balcerowicz M, Cortijo S, Charoensawan V, Rhodes D, Jonak C, Wigge PA. 2019. DEK influences the trade-off between growth and 1 arrest via H2A.Z-nucleosomes in *Arabidopsis* 2. *bioRxiv* **73170**.

Brewis ND, Street AJ, Prescott AR, Cohen PTW. 1993. PPX, a novel protein serine/threonine phosphatase localized to centrosomes. *EMBO Journal* **12**, 987–996.

Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C. 2012. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* **24**, 1242–1255.

Čaikovski M, Yokthongwattana C, Habu Y, Nishimura T, Mathieu O, Paszkowski J. 2008. Divergent evolution of CHD3 proteins resulted in MOM1 refining epigenetic control in vascular plants. *PLoS Genetics* **4**.

Campos EI, Fillingham J, Li G, et al. 2010. The program for processing newly synthesized histones H3.1 and H4. *Nature Structural and Molecular Biology* **17**, 1343–1351.

Cao X, Springer NM, Muszynski MG, Phillips RL, Kaeppler S, Jacobsen SE. 2000. Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 4979–4984.

Capuano F, Mülleder M, Kok R, Blom HJ, Ralser M. 2014. Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. *Analytical Chemistry* **86**, 3697–3702.

Carrozza MJ, Li B, Florens L, et al. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581–592.

Cavalli G, Heard E. 2019. Advances in epigenetics link genetics to the environment and disease. *Nature* **571**, 489–499.

Charpentier M, Sun J, Wen J, Mysore KS, Oldroyd GED. 2014. Abscisic acid promotion of arbuscular mycorrhizal colonization requires a component of the PROTEIN PHOSPHATASE 2A complex. *Plant Physiology* **166**, 2077–2090.

Chaubet N, Clement B, Gigot C. 1992. Genes encoding a histone H3.3-like variant in *Arabidopsis* contain intervening sequences. *Journal of Molecular Biology* **225**, 569–574.

Chen Y, Chen Q, McEachin RC, Cavalcoli JD, Yu X. 2014. H2A.B facilitates transcription elongation at methylated CpG loci. *Genome Research* **24**, 570–579.

Chen X, Ding AB, Zhong X. 2019. Functions and mechanisms of plant histone deacetylases. *Science China Life Sciences* **62**, 206–216.

Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL. 2002. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. *Cell* **110**, 33–42.

Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics* **18**, 71–86.

Clough SJ, Bent AF. 1998. Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal* **16**, 735–743.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**, 215–219.

Crevillén P, Dean C. 2011. Regulation of the floral repressor gene FLC: The complexity of transcription in a chromatin context. *Current Opinion in Plant Biology* **14**, 38–44.

Dai M, Zhang C, Kania U, et al. 2012. A PP6-type phosphatase holoenzyme directly regulates PIN phosphorylation and auxin efflux in Arabidopsis. *Plant Cell* **24**, 2497–2514.

Deshmukh S, Ponnaluri VKC, Dai N, Pradhan S, Deobagkar D. 2018. Levels of DNA cytosine methylation in the Drosophila genome. *PeerJ* **2018**, 1–13.

Dixon JR, Jung I, Selvaraj S, et al. 2015. Chromatin Architecture Reorganization during Stem Cell Differentiation. *Nature*, 331–336.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–80.

Dominguez PM, Shaknovich R. 2014. Epigenetic function of activation-induced cytidine deaminase and its link to lymphomagenesis. *frontiers in immunology* **5**, 1–10.

Dong P, Tu X, Chu P-Y, Lü P, Zhu N, Grierson D, Du B, Li P, Zhong S. 2017. 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Molecular Plant* **10**, 1497–1509.

Du J, Johnson LM, Jacobsen SE, Patel DJ. 2015. DNA methylation pathways and their crosstalk with histone methylation. *Nature Reviews Molecular Cell Biology* **16**, 519–532.

Duan CG, Wang X, Xie S, et al. 2017. A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation. *Cell Research* **27**, 226–240.

Duc C, Benoit M, Détourné G, et al. 2017. Arabidopsis ATRX modulates H3.3 occupancy and fine-tunes gene expression. *Plant Cell* **29**, 1773–1793.

Duc C, Benoit M, Le Goff S, Simon L, Poulet A, Cotterell S, Tatout C, Probst A V. 2015. The histone chaperone complex HIR maintains nucleosome occupancy and counterbalances impaired histone deposition in CAF-1 complex mutants. *the Plant journal*, 707–722.

Dunleavy EM, Roche D, Tagami H, Lacoste N, Ray-Gallet D, Nakamura Y, Daigo Y, Nakatani Y, Almouzni-Pettinotti G. 2009. HJURP Is a Cell-Cycle-Dependent Maintenance and Deposition Factor of CENP-A at Centromeres. *Cell* **137**, 485–497.

Duroux M, Houben A, Růžička K, Friml J, Grasser KD. 2004. The chromatin remodelling complex FACT associates with actively transcribed regions of the Arabidopsis genome. *Plant Journal* **40**, 660–671.

Eckner R, Ewen ME, Newsome D, Gerdes M, DeCaprio JA, Lawrence JB, Livingston DM. 1994. Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. *Genes and Development* **8**, 869–884.

- Elgin SCR, Reuter G.** 2013. Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harbor Perspectives in Biology* **5**, 1–26.
- Elsaesser SJ, Goldberg AD, Allis CD.** 2010. New functions for an old variant: No substitute for histone H3.3. *Current Opinion in Genetics and Development* **20**, 110–117.
- Elsässer SJ, Huang H, Lewis PW, Chin JW, Allis CD, Patel DJ.** 2012. DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition. *Nature* **491**, 560–565.
- English CM, Adkins MW, Carson JJ, Churchill MEA, Tyler JK.** 2006. Structural Basis for the Histone Chaperone Activity of Asf1. *Cell* **4**, 495–508.
- Epsztejn-Litman S, Feldman N, Abu-Remaileh M, et al.** 2008. De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nature Structural & Molecular Biology* **15**, 1176–1183.
- Estève PO, Hang GC, Smallwood A, Feehery GR, Gangisetty O, Karpf AR, Carey MF, Pradhan S.** 2006. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes and Development* **20**, 3089–3103.
- Fan JY, Rangasamy D, Luger K, Tremethick DJ.** 2004. H2A.Z alters the nucleosome surface to promote HP1 α -mediated chromatin fiber folding. *Molecular Cell* **16**, 655–661.
- Farkas I, Dombrádi V, Miskei M, Szabados L, Koncz C.** 2007. Arabidopsis PPP family of serine/threonine phosphatases. *Trends in Plant Science* **12**, 169–176.
- Fedoroff N V., Furtek DB, Nelson OE.** 1984. Cloning of the bronze locus in maize by a simple and generalizable procedure using the transposable controlling element Activator (Ac). *Proceedings of the National Academy of Sciences* **81**, 3825–3829.
- Feng S, Cokus SJ, Schubert V, Zhai J, Pellegrini M, Jacobsen SE.** 2014. Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. *Molecular Cell* **55**, 694–707.
- Foltz DR, Jansen LET, Bailey AO, Yates JR, Bassett EA, Wood S, Black BE, Cleveland DW.** 2009. Centromere-Specific Assembly of CENP-A Nucleosomes Is Mediated by HJURP. *Cell* **137**, 472–484.
- Fransz P, De Jong H.** 2011. From nucleosome to chromosome: A dynamic organization of genetic information. *Plant Journal* **66**, 4–17.
- Fromm M, Avramova Z.** 2014. ATX1 / AtCOMPASS and the H3K4me3 marks : how do they activate Arabidopsis genes ? *Current Opinion in Plant Biology* **21**, 75–82.
- Fu M, Wang C, Zhang X, Pestell RG.** 2004. Acetylation of nuclear receptors in cellular growth and apoptosis. *Biochemical Pharmacology* **68**, 1199–1208.
- Gao Z, Liu H-L, Daxinger L, et al.** 2010. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* **465**, 106–109.
- Gehring M.** 2013. Genomic Imprinting: Insights From Plants. *Annual Review of Genetics* **47**, 187–208.
- Gehring M, Bubb KL, Henikoff S.** 2009. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* **324**, 1447–1451.
- Genoud T, Cruz MTS, Kulisic T, Sparla F, Fankhauser C, Métraux J-P.** 2008. The Protein Phosphatase 7 Regulates Phytochrome Signaling in Arabidopsis. *PLoS ONE* **3**.
- Gil RS, Vagnarelli P.** 2019. Protein phosphatases in chromatin structure and function. *Biochimica et Biophysica Acta - Molecular Cell Research* **1866**, 90–101.
- Le Goff S, Keçeli BN, Jeřábková H, et al.** 2020. The H3 histone chaperone NASPSIM3 escorts CenH3 in Arabidopsis. *Plant Journal* **101**, 71–86.
- Goldberg AD, Banaszynski LA, Noh KM, et al.** 2010. Distinct Factors Control

Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell* **140**, 678–691.

Goldberg J, Huang H Bin, Kwon YG, Greengard P, Nairn AC, Kuriyan J. 1995. Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature* **376**, 745–753.

Goodarzi AA, Jonnalagadda JC, Douglas P, Young D, Ye R, Moorhead GBG, Lees-Miller SP, Khanna KK. 2004. Autophosphorylation of ataxia-telangiectasia mutated is regulated by protein phosphatase 2A. *EMBO Journal* **23**, 4451–4461.

Goodman RH, Smolik S. 2000. CBP/p300 in cell growth, transformation, and development. *Genes and Development* **14**, 1553–1577.

Grasser KD. 2020. The FACT Histone Chaperone: Tuning Gene Transcription in the Chromatin Context to Modulate Plant Growth and Development. *Frontiers in Plant Science* **11**, 1–8.

Greenberg MVC, Bourc'his D. 2019. The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology* **20**, 590–607.

Grewal SIS, Jia S. 2007. Heterochromatin revisited. *Nature Reviews Genetics* **8**, 35–46.

Grin I, Ishchenko AA. 2016. An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation. *Nucleic Acids Research* **44**, 3713–3727.

Grob S, Schmid MW, Grossniklaus U. 2014. Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. *Molecular Cell* **55**, 678–693.

Haga K, Sakai T. 2018. Involvement of PP6-type protein phosphatase in hypocotyl phototropism in Arabidopsis seedlings. *Plant Signaling & Behavior* **13**, 1–3.

Han Y-F, Zhao Q-Y, Dang L-L, et al. 2016. The SUMO E3 ligase-like proteins PIAL1 and PIAL2 interact with MOM1 and form a novel complex required for transcriptional silencing. *The Plant Cell* **2**, tpc.00997.2015.

Heitz E. 1928. Das Heterochromatin der Moose. *Jahrbücher für Wissenschaftliche Botanik* **69**, 762–818.

Heroes E, Lesage B, Görnemann J, Beullens M, Van Meervelt L, Bollen M. 2013. The PP1 binding code: A molecular-lego strategy that governs specificity. *FEBS Journal* **280**, 584–595.

Hershko A, Ciechanover A. 1998. The ubiquitin system. *Annual review of biochemistry* **67**, 425–79.

Hodawadekar SC, Marmorstein R. 2007. Chemistry of acetyl transfer by histone modifying enzymes: Structure, mechanism and implications for effector design. *Oncogene* **26**, 5528–5540.

Hoek M, Stillman B. 2003. Chromatin assembly factor 1 is essential and couples chromatin assembly to DNA replication in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12183–12188.

Hof AE van't, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri lik J. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**, 102–105.

Hsieh T-F, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D. 2009. Genome-Wide Demethylation of Arabidopsis Endosperm. *Science* **324**, 1451–1455.

Huh JH, Bauer MJ, Hsieh TF, Fischer RL. 2008. Cellular Programming of Plant Gene Imprinting. *Cell* **132**, 735–744.

Hung FY, Chen FF, Li C, Chen C, Chen JH, Cui Y, Wu K. 2019. The LDL1/2-HDA6 histone modification complex interacts with TOC1 and regulates the core circadian

clock components in Arabidopsis. *Frontiers in Plant Science* **10**, 1–10.

Hunter T. 1995. Protein kinases and phosphatases: The Yin and Yang of protein phosphorylation and signaling. *Cell* **80**, 225–236.

Ikeda Y, Pelissier T, Bourguet P, Becker C, Pouch-Pelissier M-N, Pogorelcnik R, Weingartner M, Weigel D, Deragon J, Mathieu O. 2017. Arabidopsis proteins with a transposon-related domain act in gene silencing. *Nature Communications*, 1–10.

Jacob Y, Feng S, Leblanc CA, Bernatavichute Y V, Stroud H, Cokus S, Johnson LM, Pellegrini M, Jacobsen SE, Michaels SD. 2009. ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nature Structural & Molecular Biology* **16**, 763–768.

Jacob Y, Michaels SD. 2009. H3K27me1 is E(z) in animals, but not in plants. *Epigenetics* **2294**, 3–7.

Jangam D, Feschotte C, Betrán E. 2017. Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. *Trends in Genetics* **33**, 817–831.

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. 2007. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449**, 1–6.

Jiang D, Berger F. 2017. DNA replication-coupled histone modification maintains Polycomb gene silencing in plants. *Science* **357**, 1146–1149.

Johnson LM, Du J, Hale CJ, et al. 2014. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507**, 124–128.

Jones RN. 2005. McClintock's controlling elements: The full story. *Cytogenetic and Genome Research* **109**, 90–103.

Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F. 2012. DNA Methylation Dynamics during Sexual Reproduction in Arabidopsis thaliana. *Current Biology* **22**, 1825–1830.

Kafri T, Ariel M, Brandeis M, Shemer R, Urven L, McCarrey J, Cedar H, Razin A. 1992. Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes and Development* **6**, 705–714.

Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddloh JA, Riddle NC, Verbsky ML, Richards EJ. 2003. Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics* **163**, 1109–1122.

Kaya H, Shibahara K ichi, Taoka K ichiro, Iwabuchi M, Stillman B, Araki T. 2001. FASCIATA genes for chromatin assembly factor-1 in Arabidopsis maintain the cellular organization of apical meristems. *Cell* **104**, 131–142.

Kharchenko P V., Alekseyenko AA, Schwartz YB, et al. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**, 480–486.

Kim D, Kang J, Yang S, Chung K, Song P, Park C-M. 2002. A Phytochrome-Associated Protein Phosphatase 2A Modulates Light Signals in Flowering Time Control in Arabidopsis. *The Plant cell* **14**, 3043–3056.

Kim J, Kim JH, Richards EJ, Chung KM, Woo HR. 2014. Arabidopsis VIM proteins regulate epigenetic silencing by modulating DNA methylation and histone modification in cooperation with MET1. *Molecular Plant* **7**, 1470–1485.

Kim JK, Samaranayake M, Pradhan S. 2009. Epigenetic mechanisms in mammals. *Cellular and Molecular Life Sciences* **66**, 596–612.

Kobayashi S, Goto-Yamamoto N, Hirochika H. 2004. Retrotransposon-Induced Mutations in Grape Skin Color. *Science* **304**, 982.

Kokura K, Sun L, Bedford MT, Fang J. 2010. Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion. *The EMBO Journal* **29**, 3673–3687.

Kouzarides T. 2007. Chromatin Modifications and Their Function. *Cell* **128**, 693–705.

Kurihara Y, Matsui A, Kawashima M, et al. 2008. Identification of the candidate genes regulated by RNA-directed DNA methylation in Arabidopsis. *Biochemical and Biophysical Research Communications* **376**, 553–557.

Lahmy S, Pontier D, Bies-Etheve N, et al. 2016. Evidence for ARGONAUTE4–DNA interactions in RNA-directed DNA methylation in plants. *Genes and Development* **30**, 2565–2570.

Lang Z, Lei M, Wang X, et al. 2015. The Methyl-CpG-Binding Protein MBD7 Facilitates Active DNA Demethylation to Limit DNA Hyper-Methylation and Transcriptional Gene Silencing. *Molecular Cell* **57**, 971–983.

Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, Strahl BD, Patel DJ, Jacobsen SE. 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**, 385–389.

Law JA, Jacobsen SE. 2010. Establishing , maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**, 204–220.

Lawrence EJ, Gao H, Tock AJ, Lambing C, Blackwell AR, Feng X, Henderson IR. 2019. Natural Variation in TBP-ASSOCIATED FACTOR 4b Controls Meiotic Crossover and Germline Transcription in Arabidopsis. *Current Biology* **29**, 2676-2686.e3.

Lee K, Park O, Seo PJ. 2017. Arabidopsis ATXR2 deposits H3K36me3 at the promoters of LBD genes to facilitate cellular dedifferentiation. *Science Signaling* **10**, 1–11.

Lee JS, Shukla A, Schneider J, Swanson SK, Washburn MP, Florens L, Bhaumik SR, Shilatifard A. 2007. Histone Crosstalk between H2B Monoubiquitination and H3 Methylation Mediated by COMPASS. *Cell* **131**, 1084–1096.

Lewis PW, Elsaesser SJ, Noh K-M, Stadler SC, Allis CD. 2010. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14075–80.

Li S, Vandivier LE, Tu B, Gao L, Won SY, Li S, Zheng B, Gregory BD, Chen X. 2015. Detection of pol IV/RDR2-dependent transcripts at the genomic scale in arabidopsis reveals features and regulation of sirna biogenesis. *Genome Research* **25**, 235–245.

Li C, Xu J, Li J, Li Q, Yang H. 2014. Involvement of arabidopsis histone acetyltransferase HAC family genes in the ethylene signaling pathway. *Plant and Cell Physiology* **55**, 426–435.

Lindroth AM, Cao X, Jackson JP, Jacobsen SE. 2001. Requirement of CHROMOMETHYLASE3 for Maintenance of CpXpG Methylation. *Science* **292**, 2077–2081.

Lisch D. 2013. How important are transposons for plant evolution? *Nature Reviews Genetics* **14**, 49–61.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell* **133**, 523–536.

Lister R, Pelizzola M, Downen RH, et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322.

Liu H, Li G, Chang HUI, Sun D, Zhou R, Li B. 2007. Calmodulin-binding protein phosphatase PP7 is involved in thermotolerance in Arabidopsis. *Plant, cell & environment* **30**, 156–164.

Liu J, Sheng L, Xu Y, Li J, Yang Z, Huang H, Xu L. 2014. WOX11 and 12 Are Involved in the First-Step Cell Fate Transition during de Novo Root Organogenesis in

Arabidopsis. *The Plant Cell* **26**, 1081–1093.

Liu Y, Yan J, Qin Q, Zhang J, Chen Y, Zhao L, He K, Hou S. 2019. Type one protein phosphatases (TOPPs) contribute to the plant defense response in Arabidopsis. *Journal of Integrative Plant Biology* **00**.

Long M, Sun X, Shi W, et al. 2019. A novel histone H4 variant H4G regulates rDNA transcription in breast cancer. *Nucleic acids research* **47**, 8399–8409.

Los G V, Encell LP, Mcdougall MG, et al. 2008. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS chemical biology* **3**, 373–382.

Lu Y, Dai J, Yang L, et al. 2020. Involvement of MEM1 in DNA demethylation in Arabidopsis. *Plant Molecular Biology*.

Luger K, Mäder a W, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260.

Luger K, Rechsteiner TJ, Richmond TJ. 1999. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods in molecular biology* (Clifton, N.J.) **119**, 1–16.

de Luxán-Hernández C, Lohmann J, Hellmeyer W, et al. 2019. PP7L is essential for MAIL1-mediated transposable element silencing and primary root growth. *Plant Journal*, 703–717.

Macaisne N, Novatchkova M, Peirera L, Vezon D, Jolivet S, Froger N, Chelysheva L, Grelon M, Mercier R. 2008. SHOC1, an XPF Endonuclease-Related Protein, Is Essential for the Formation of Class I Meiotic Crossovers. *Current Biology* **18**, 1432–1437.

Macaisne N, Vignard J, Mercier R. 2011. SHOC1 and PTD form an XPF-ERCC1-like complex that is required for formation of class I crossovers. *Journal of Cell Science* **124**, 2687–2691.

Mahrez W, Arellano MST, Moreno-Romero J, Nakamura M, Shu H, Nanni P, Köhler C, Gruissem W, Hennig L. 2016. H3K36ac is an evolutionary conserved plant histone modification that marks active genes. *Plant Physiology* **170**, 1566–1577.

Manning ET, Ikehara T, Ito T, Kadonaga JT, Kraus WL. 2001. p300 Forms a Stable, Template-Committed Complex with Chromatin: Role for the Bromodomain. *Molecular and Cellular Biology* **21**, 3876–3887.

Martínez de Alba AE, Elvira-Matelot E, Vaucheret H. 2013. Gene silencing in plants: A diversity of pathways. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1829**, 1300–1308.

Matzke MA, Kanno T, Matzke AJM. 2015. RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annual Review of Plant Biology* **66**, 243–267.

McClintock B. 1950. The origin and behavior of mutable loci in maize. *Genetics* **36**.

McCue AD, Panda K, Nuthikattu S, Choudury SG, Thomas EN, Slotkin RK. 2015. ARGONAUTE 6 bridges transposable element m RNA -derived si RNA s to the establishment of DNA methylation . *The EMBO Journal* **34**, 20–35.

Michaels SD, Amasino RM. 1999. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**, 949–956.

Mito Y, Henikoff JG, Henikoff S. 2005. Genome-scale profiling of histone H3.3 replacement patterns. *Nature Genetics* **37**, 1090–1097.

Moissiard G, Bischof S, Husmann D, et al. 2014. Transcriptional gene silencing by Arabidopsis microorchidia homologues involves the formation of heteromers. *Proceedings of the National Academy of Sciences* **111**, 7474–7479.

Moissiard G, Cokus SJ, Cary J, et al. 2012. MORC Family ATPases Required for Heterochromatin Condensation and Gene Silencing. *Science* **336**, 1448–1452.

- Monk M, Boubelik M, Lehnert S.** 1987. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371–382.
- Montgomery SA, Tanizawa Y, Galik B, et al.** 2020. Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. *Current Biology* **30**, 1–16.
- Moorhead GBG, Trinkle-Mulcahy L, Ulke-Lemée A.** 2007. Emerging roles of nuclear protein phosphatases. *Nature Reviews Molecular Cell Biology* **8**, 234–244.
- Mozgova I, Hennig L.** 2015. The Polycomb Group Protein Regulatory Network. *Annual Review of Plant Biology* **66**, 269–296.
- Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T.** 2000. Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. *Cell* **102**, 553–563.
- Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T.** 1999. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *Journal of Biological Chemistry* **274**, 18470–18476.
- Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jiang J.** 2003. Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. *Genetics* **163**, 1221–1225.
- Nathan D, Ingvarsdottir K, Sterner DE, et al.** 2006. Histone sumoylation is a negative regulator in *Saccharomyces cerevisiae* and shows dynamic interplay with positive-acting histone modifications. *Gene & Development* **20**, 966–976.
- Natsume R, Eitoku M, Akai Y, Sano N, Horikoshi M, Senda T.** 2007. Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. *Nature* **446**, 1–4.
- Nicolau M, Picault N, Descombin J, Jami-alahmadi Y, Feng S, Bucher E, Jacobsen SE, Deragon J, Wohlschlegel J, Moissiard G.** 2020. The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in *Arabidopsis thaliana*. *PLoS Genetics*, 1–29.
- Nie X, Wang H, Li J, Holec S.** 2014. The HIRA complex that deposits the histone H3 . 3 is conserved in *Arabidopsis* and facilitates transcriptional dynamics. *Biology open*, 1–9.
- Nishimura T, Molinard G, Petty TJ, Broger L, Gabus C, Halazonetis TD, Thore S, Paszkowski J.** 2012. Structural basis of transcriptional gene silencing mediated by *Arabidopsis* MOM1. *PLoS Genetics* **8**.
- Nishino Y, Eltsov M, Joti Y, et al.** 2012. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO Journal* **31**, 1644–1653.
- Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, Keith Slotkin R.** 2013. The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiology* **162**, 116–131.
- Okada T, Endo M, Singh MB, Bhalla PL.** 2005. Analysis of the histone H3 gene family in *Arabidopsis* and identification of the male-gamete-specific variant AtMGH3. *Plant Journal* **44**, 557–568.
- Okano M, Xie S, Li E.** 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases Non-invasive sexing of preimplantation stage mammalian embryos. *Nature America Inc.* **19**, 219–220.

- Ooi SKT, Qiu C, Bernstein E, et al.** 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717.
- Ortega-Galisteo AP, Morales-Ruiz T, Ariza RR, Roldán-Arjona T.** 2008. Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Molecular Biology* **67**, 671–681.
- Osakabe A, Lorković ZJ, Kobayashi W, Tachiwana H, Yelagandula R, Kurumizaka H, Berger F.** 2018. Histone H2A variants confer specific properties to nucleosomes and impact on chromatin accessibility. *Nucleic Acids Research* **46**, 7675–7685.
- Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M.** 2009. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. *EMBO reports* **10**, 1235–1241.
- Otero S, Desvoyes B, Peiro R, Gutierrez C.** 2016. Histone H3 Dynamics Reveal Domains with Distinct Proliferation Potential in the Arabidopsis Root. *The Plant cell*.
- Page DR, Grossniklaus U.** 2002. The art and design of genetic screens: Arabidopsis thaliana. *Nature Reviews Genetics* **3**, 124–136.
- Parthun MR.** 2007. Hat1: The emerging cellular roles of a type B histone acetyltransferase. *Oncogene* **26**, 5319–5328.
- Passarge E.** 1979. Emil Heitz and the Concept of Heterochromatin: Longitudinal Chromosome Differentiation was Recognized Fifty Years Ago. *American Journal of Human Genetics* **31**, 106–115.
- Pastor WA, Stroud H, Nee K, et al.** 2014. MORC1 represses transposable elements in the mouse male germline. *Nature Communications* **5**, 5795.
- Pecinka A, Dinh HQ, Baubec T, Rosa M, Lettner N, Mittelsten Scheid O.** 2010. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *The Plant cell* **22**, 3118–3129.
- Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL.** 2007. DNA demethylation in the Arabidopsis genome. *PNAS* **104**, 6752–6757.
- Phansalkar R, Lapierre P, Mellone BG.** 2012. Evolutionary insights into the role of the essential centromere protein CAL1 in Drosophila. *Chromosome Research* **20**, 493–504.
- Pontvianne F, Blevins T, Pikaard CS.** 2010. Arabidopsis Histone Lysine Methyltransferases. *Advances in Botanical Research* **53**.
- Pontvianne F, Carpentier M-C, Durut N, et al.** 2016. Identification of Nucleolus-Associated Chromatin Domains Reveals a Role for the Nucleolus in 3D Organization of the A. thaliana Genome. Pontvianne, F., Carpentier, M.-C., Durut, N., Pavlišťová, V., Jaške, K., Schořová, Š., Parrinello, H., Rohmer, M., Pikaard. *Cell reports* **16**, 1574–1587.
- Pontvianne F, Liu C.** 2020. Chromatin domains in space and their functional implications. *Current Opinion in Plant Biology* **April**, 1–10.
- Probst A V., Desvoyes B, Gutierrez C.** 2020. Similar yet Critically Different: The distribution, dynamics and function of Histone Variants. *Journal of experimental botany*, 1–34.
- Qian W, Miki D, Zhang H, et al.** 2012. A Histone Acetyltransferase Regulates Active DNA Demethylation in Arabidopsis. *Science* **336**, 1445–1448.
- Ramsahoye BH, Biniszkievicz D, Lyko F, Clark V, Bird AP, Jaenisch R.** 2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5237–5242.
- Ray-Gallet D, Quivy J, Scamps C, Martini EM, Lipinski M, Cedex V.** 2002. HIRA Is Critical for a Nucleosome Assembly Pathway Independent of DNA Synthesis.

Molecular cell, 1091–1100.

Reiterer V, Eyers PA, Farhan H. 2014. Day of the dead: Pseudokinases and pseudophosphatases in physiology and disease. *Trends in Cell Biology* **24**, 489–505.

Rigal M, Mathieu O. 2011. A “ mille-feuille ” of silencing: Epigenetic control of transposable elements. *BBA - Gene Regulatory Mechanisms* **1809**, 452–458.

Rommel PC, Bosque D, Gitlin AD, Croft GF, Heintz N, Casellas R, Nussenzweig MC, Kriaucionis S, Robbiani DF. 2013. Fate Mapping for Activation-Induced Cytidine Deaminase (AID) Marks Non-Lymphoid Cells During Mouse Development. *PLoS ONE* **8**, 1–9.

Rossetto D, Avvakumov N, Côté J. 2012. Histone phosphorylation A chromatin modification involved in diverse nuclear events. *Epigenetics* **7**, 1098–1108.

Roudier F, Ahmed I, Bérard C, et al. 2011. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *The EMBO Journal* **30**, 1928–1938.

Sani E, Herzyk P, Perrella G, Colot V, Amtmann A. 2013. Hyperosmotic priming of Arabidopsis seedlings establishes a long-term somatic memory accompanied by specific changes of the epigenome. *Genome biology* **14**, R59.

Santisteban MS, Kalashnikova T, Smith MM. 2000. Histone H2A.Z regulates transcription and is partially redundant with nucleosome remodeling complexes. *Cell* **103**, 411–422.

Santos F, Peat J, Burgess H, Rada C, Reik W, Dean W. 2013. Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics and Chromatin* **6**, 1–12.

Sasaki H, Matsui Y. 2008. Epigenetic events in mammalian germ-cell development: Reprogramming and beyond. *Nature Reviews Genetics* **9**, 129–140.

Schubert V, Berr A, Meister A. 2012. Interphase chromatin organisation in Arabidopsis nuclei: Constraints versus randomness. *Chromosoma* **121**, 369–387.

Schuh M, Lehner CF, Heidmann S. 2007. Report Incorporation of Drosophila CID / CENP-A and CENP-C into Centromeres during Early Embryonic Anaphase. *Current Biology* **17**, 237–243.

Sharif J, Muto M, Takebayashi SI, et al. 2007. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**, 908–912.

Sharma N, Geuten K, Giri BS, Varma A. 2020. The molecular mechanism of vernalization in Arabidopsis and Cereals : Role of Flowering Locus C and its homologs. *Physiologia Plantarum*.

Shi Y. 2009. Serine/Threonine Phosphatases: Mechanism through Structure. *Cell* **139**, 468–484.

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y. 2004. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **119**, 941–953.

Shi L, Wang J, Hong F, Spector DL, Fang Y. 2011. Four amino acids guide the assembly or disassembly of Arabidopsis histone H3.3-containing nucleosomes. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10574–10578.

Shiio Y, Eisenman RN. 2003. Histone sumoylation is associated with transcriptional repression. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 13225–13230.

Shim Y, Duan MR, Chen X, Smerdon MJ, Min JH. 2012. Polycistronic coexpression and nondenaturing purification of histone octamers. *Analytical Biochemistry* **427**, 190–192.

Shu H, Nakamura M, Siretskiy A, Borghi L, Moraes I, Wildhaber T, Grussem W,

- Hennig L.** 2014. Arabidopsis replacement histone variant H3.3 occupies promoters of regulated genes. *Genome biology* **15**, R62.
- Simon L, Voisin M, Tatout C, Probst A V.** 2015. Structure and Function of Centromeric and Pericentromeric Heterochromatin in *Arabidopsis thaliana*. *Frontiers in Plant Science* **6**, 1–8.
- Skottke KR, Yoon GM, Kieber JJ, DeLong A.** 2011. Protein phosphatase 2A controls ethylene biosynthesis by differentially regulating the turnover of ACC synthase isoforms. *PLoS Genetics* **7**, 1–20.
- Slotkin RK, Martienssen R.** 2007. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* **8**, 272–285.
- Song F, Chen P, Sun D, Wang M, Dong L, Liang D, Xu RM, Zhu P, Li G.** 2014. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* **344**, 376–380.
- Steinbauerová V, Neumann P, Novák P, Macas J.** 2011. A widespread occurrence of extra open reading frames in plant Ty3/gypsy retrotransposons. *Genetica* **139**, 1543–1555.
- Stoltenburg R, Reinemann C, Strehlitz B.** 2007. SELEX-A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering* **24**, 381–403.
- Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE.** 2014. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nature Structural & Molecular Biology* **21**, 64–72.
- Stroud H, Greenberg MVC, Feng S, Bernatavichute Y V., Jacobsen SE.** 2013. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**, 352–364.
- Stroud H, Otero S, Desvoyes B, Ramirez-parra E, Jacobsen SE.** 2012. Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis thaliana*. *PNAS*.
- Sun X, Kang X, Ni M.** 2012. Hypersensitive to red and blue 1 and its modification by protein phosphatase 7 are implicated in the control of *Arabidopsis* stomatal aperture. *PLoS Genetics* **8**.
- Sura W, Kabza M, Karlowski WM, Bieluszewski T, Kus-Slowinska M, Pawełszek Ł, Sadowski J, Ziolkowski PA.** 2017. Dual role of the histone variant H2A.Z in transcriptional regulation of stress-response genes. *Plant Cell* **29**, 791–807.
- Tachiwana H, Osakabe A, Shiga T, Miya Y, Kimura H, Kagawa W, Kurumizaka H.** 2011. Structures of human nucleosomes containing major histone H3 variants. *Acta Crystallographica Section D: Biological Crystallography* **67**, 578–583.
- Tagami H, Ray-gallet D, Nakatani Y.** 2004. Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis. *Cell*, 51–61.
- Takemiya A, Ariyoshi C, Shimazaki KI.** 2009. Identification and functional characterization of inhibitor-3, a regulatory subunit of protein phosphatase 1 in plants. *Plant Physiology* **150**, 144–156.
- Takemiya A, Yamauchi S, Yano T, Ariyoshi C, Shimazaki KI.** 2013. Identification of a regulatory subunit of protein phosphatase 1 which mediates blue light signaling for stomatal opening. *Plant and Cell Physiology* **54**, 24–35.
- Talbert PB, Ahmad K, Almouzni G, et al.** 2012. A unified phylogeny-based nomenclature for histone variants. *Epigenetics & Chromatin* **5**, 7.
- Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S.** 2002. Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *The Plant cell* **14**, 1053–66.
- Tang K, Lang Z, Zhang H, Zhu JK.** 2016. The DNA demethylase ROS1 targets

genomic regions with distinct chromatin modifications. *Nature Plants* **2**, 1–10.

Templeton GW, Nimick M, Morrice N, Campbell D, Goudreault M, Gingras AC, Takemiya A, Shimazaki KI, Moorhead GBG. 2011. Identification and characterization of AtI-2, an Arabidopsis homologue of an ancient protein phosphatase 1 (PP1) regulatory subunit. *Biochemical Journal* **435**, 73–83.

Trojer P, Reinberg D. 2007. Review Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell* **28**, 1–13.

Tsukada Y, Fang J, Erdjument-bromage H, Warren ME, Borchers CH, Tempst P, Zhang Y. 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811–816.

Udugama M, M Chang FT, Chan FL, Tang MC, Pickett HA, R McGhie JD, Mayne L, Collas P, Mann JR, Wong LH. 2015. Histone variant H3.3 provides the heterochromatic H3 lysine 9 tri-methylation mark at telomeres. *Nucleic acids research* **43**, 10227–37.

Ueda M, Matsui A, Tanaka M, et al. 2017. The distinct roles of class I and II RPD3-like histone deacetylases in salinity stress response. *Plant Physiology* **175**, 1760–1773.

Ühlken C, Horvath B, Stadler R, Sauer N, Weingartner M. 2014. MAIN-LIKE1 is a crucial factor for correct cell division and differentiation in Arabidopsis thaliana. *Plant Journal* **78**, 107–120.

Uhrig RG, Labandera AM, Moorhead GB. 2013. Arabidopsis PPP family of serine/threonine protein phosphatases: Many targets but few engines. *Trends in Plant Science* **18**, 505–513.

Venkatesh S, Workman JL. 2015. Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews. Molecular Cell Biology* **16**, 178–189.

Vergara Z, Gutierrez C. 2017. Emerging roles of chromatin in the maintenance of genome organization and function in plants. *Genome biology* **18**, 1–12.

Waddington CH. 1942. The Epigenotype. *Endeavour*, 18–20.

Waidmann S, Kusenda B, Mayerhofer J, Mechtler K, Jonak C. 2014. A deK Domain-Containing protein modulates chromatin structure and function in Arabidopsis. *Plant Cell* **26**, 4328–4344.

Wang S, Quan L, Li S, et al. 2019. The PROTEIN PHOSPHATASE4 Complex Promotes Transcription and Processing of Primary microRNAs in Arabidopsis. *Plant Cell* **31**, 486–501.

Wang JW, Schwab R, Czech B, Mica E, Weigel D. 2008. Dual effects of miR156-targeted SPL genes and CYP78A5/KLUH on plastochron length and organ size in Arabidopsis thaliana. *Plant Cell* **20**, 1231–1243.

Wang S, Su J, Beliveau BJ, Bintu B, Moffitt JR, Wu C, Zhuang X. 2016. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **544**, 2955–2960.

Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, Zhang Y. 2004. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 862–868.

Waterborg JH, Robertson AJ. 1996. Common features of analogous replacement histone H3 genes in animals and plants. *Journal of Molecular Evolution* **43**, 194–206.

Weber M, Guibert S, Forne T. 2012. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Research* **22**, 633–641.

Weber CM, Henikoff S. 2014. Histone variants: dynamic punctuation in transcription. *Genes & development* **28**, 672–82.

Wenig U, Meyer S, Stadler R, Fischer S, Werner D, Lauter A, Melzer M, Hoth S, Weingartner M, Sauer N. 2013. Identification of MAIN, a factor involved in genome

stability in the meristems of *Arabidopsis thaliana*. *Plant Journal* **75**, 469–483.

Wibowo A, Becker C, Marconi G, et al. 2016. Hyperosmotic stress memory in *Arabidopsis* is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity. *eLife* **5**, e13546.

Wollmann H, Holec S, Alden K, Clarke ND. 2012. Dynamic Deposition of Histone Variant H3 . 3 Accompanies Developmental Remodeling of the *Arabidopsis* Transcriptome. *Plos genetics*, 28–31.

Wollmann H, Stroud H, Yelagandula R, et al. 2017. The histone H3 variant H3.3 regulates gene body DNA methylation in *Arabidopsis thaliana*. *Genome Biology* **18**, 94.

Woo HR, Dittmer TA, Richards EJ. 2008. Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*. *PLoS Genetics* **4**.

Wu Z, Fang X, Zhu D, Dean C. 2020. Autonomous pathway: Flowering locus c repression through an antisense-mediated chromatin-silencing mechanism. *Plant Physiology* **182**, 27–37.

Wu Z, Ietswaart R, Liu F, Yang H, Howard M, Dean C. 2016. Quantitative regulation of *FLC* via coordinated transcriptional initiation and elongation. *Proceedings of the National Academy of Sciences* **113**, 218–223.

Wu X, Zhang Y. 2017. TET-mediated active DNA demethylation: Mechanism, function and beyond. *Nature Reviews Genetics* **18**, 517–534.

Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC. 2004. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biology* **2**, 642–652.

Xu D, Marino G, Klingl A, Enderle B, Monte E, Kurth J, Hiltbrunner A, Leister D, Kleine T. 2019. Extrachloroplastic PP7L Functions in Chloroplast Development and Abiotic Stress Tolerance. *Plant Physiology* **180**, 323–341.

Xu L, Zhao Z, Dong A, Soubigou-taconnat L, Renou J, Steinmetz A, Shen W. 2008. Di- and Tri- but Not Monomethylation on Histone H3 Lysine 36 Marks Active Transcription of Genes Involved in Flowering Time Regulation and Other Processes in *Arabidopsis thaliana*. *Molecular and Cellular Biology* **28**, 1348–1360.

Yang H, Howard M, Dean C. 2014. Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at *Arabidopsis* *FLC*. *Current Biology* **24**, 1793–1797.

Yang XJ, Seto E. 2007. HATs and HDACs: From structure, function and regulation to novel strategies for therapy and prevention. *Oncogene* **26**, 5310–5318.

Yelagandula R, Stroud H, Holec S, et al. 2014. The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in *Arabidopsis*. *Cell* **158**, 98–109.

Yokthongwattana C, Bucher E, Caikovski M, Vaillant I, Nicolet J, Scheid OM, Paszkowski J. 2010. MOM1 and Pol-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing. *The EMBO journal* **29**, 340–351.

Yu X, Dong J, Deng Z, Jiang Y, Wu C, Qin X, Terzaghi W, Chen H, Daia M, Deng XW. 2019. *Arabidopsis* PP6 phosphatases dephosphorylate PIF proteins to repress photomorphogenesis. *PNAS* **116**, 20218–20225.

Yu H, Xu Y, Tan EL, Kumar PP. 2002. AGAMOUS-LIKE 24, a dosage-dependent mediator of the flowering signals. *PNAS* **99**.

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. 2013. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205.

- Zhai J, Bischof S, Wang H, et al.** 2015. A one precursor one siRNA model for pol IV-dependent siRNA biogenesis. *Cell* **163**, 445–455.
- Zhang C, Freddolino PL, Zhang Y.** 2017. COFACTOR: Improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Research* **45**, W291–W299.
- Zhang Y, Harris CJ, Liu Q, et al.** 2018a. Large-scale comparative epigenomics reveals hierarchical regulation of non-CG methylation in Arabidopsis. *Proceedings of the National Academy of Sciences*.
- Zhang H, Lang Z, Zhu JK.** 2018b. Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology* **19**, 489–506.
- Zhang X, Yazaki J, Sundaresan A, et al.** 2006. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. *Cell* **126**, 1189–1201.
- Zhao Z, Yu Y, Meyer D, Wu C, Shen W.** 2005. Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. *nature cell biology* **7**, 1256–1260.
- Zheng Y, Ge J, Bao C, Chang W, Liu J, Shao J, Liu X, Su L, Pan L, Zhou DX.** 2020. Histone Deacetylase HDA9 and WRKY53 Transcription Factor Are Mutual Antagonists in Regulation of Plant Stress Response. *Molecular Plant*, 1–14.
- Zheng X, Pontes O, Zhu J, Miki D, Zhang F, Li WX, Iida K, Kapoor A, Pikaard CS, Zhu JK.** 2008. ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis. *Nature* **455**, 1259–1262.
- Zhou HW, Nussbaumer C, Chao Y, DeLong A.** 2004. Disparate roles for the regulatory A subunit isoforms in Arabidopsis protein phosphatase 2A. *Plant Cell* **16**, 709–722.
- Zhou M, Palanca AMS, Law JA.** 2018. Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. *Nature Genetics* **50**, 865–873.
- Zilberman D.** 2017. An evolutionary case for functional gene body methylation in plants and animals. *Genome Biology* **18**, 17–19.
- Zilberman D, Cao X, Jacobsen SE.** 2003. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**, 716–719.
- Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S.** 2008. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**, 125–129.
- Zubko E, Gentry M, Kunova A, Meyer P.** 2012. De novo DNA methylation activity of METHYLTRANSFERASE 1 (MET1) partially restores body methylation in Arabidopsis thaliana. *Plant Journal* **71**, 1029–1037.