



HAL
open science

Implication de la protéine SG1 dans le maintien des épigénomes chez *Arabidopsis thaliana*

Aurélie Deremetz

► **To cite this version:**

Aurélie Deremetz. Implication de la protéine SG1 dans le maintien des épigénomes chez *Arabidopsis thaliana*. Biologie moléculaire. Université Paris Saclay (COMUE), 2015. Français. NNT : 2015SACLS119 . tel-01674180v2

HAL Id: tel-01674180

<https://theses.hal.science/tel-01674180v2>

Submitted on 2 Jan 2018

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.

NNT : 2015SACLS119

THESE DE DOCTORAT
DE L'UNIVERSITE PARIS-SACLAY,
préparée à l'Université Paris-Sud

ÉCOLE DOCTORALE N° 567
Sciences du Végétal : du Gène à l'Ecosystème

Spécialité de doctorat : Biologie

Par

Aurélie DEREMETZ

Implication de la protéine SG1 dans
le maintien des épigénomes chez *Arabidopsis thaliana*.

Thèse présentée et soutenue à Versailles, le 3 décembre 2015 :

Composition du Jury :

M. DERAGON Jean-Marc	Professeur, Université de Perpignan	Rapporteur
M. MATHIEU Olivier	Directeur de Recherche, CNRS, Clermont-Ferrand	Rapporteur
Mme DELARUE Marianne	Professeure, Université Paris-Sud	Examinatrice
M. ROUDIER François	Chargé de Recherche, CNRS, Paris	Examinateur
Mme VITTE Clémentine	Chargée de Recherche, CNRS, Le Moulon	Examinatrice
M. BOUCHE Nicolas	Chargé de Recherche, INRA, Versailles	Directeur de thèse



Titre : Implication de la protéine SG1 dans le maintien des épigénomes chez *Arabidopsis thaliana*.

Mots clés : INCREASE IN BONSAI METHYLATION (IBM), FPA, polyadénylation, épiallèles, *Arabidopsis*.

Résumé : La chromatine est le support de l'information génétique et sa structure, ainsi que son activité transcriptionnelle, peuvent être modulées par des modifications épigénétiques. Le maintien des marques répressives telles que la méthylation de l'ADN et des histones, hors du corps des gènes, est nécessaire pour le bon développement de la plante. Chez *Arabidopsis thaliana*, le mutant *sg1* présente des défauts développementaux sévères caractéristiques de mutants affectés dans des mécanismes épigénétiques. Nous avons montré que le phénotype de *sg1* est causé par une hyperméthylation CHG et H3K9me2 dans de nombreux gènes. En effet, SG1 contrôle la transcription de l'histone déméthylase IBM1 et les modifications de l'épigénome observées chez *sg1* sont dues à une dérégulation de IBM1. Nous avons identifié sept protéines partenaires de SG1, dont certaines se lient aux marques chromatiniennes. Nous avons réalisé un crible suppresseur qui a permis d'identifier FPA, une protéine régulant la polyadénylation de certains transcrits, comme acteur impliqué dans le contrôle des cibles de SG1, dont IBM1. Nos résultats montrent que le complexe SG1 régule la transcription de ses cibles en influençant, par un mécanisme encore inconnu, le choix du site de polyadénylation, en lien avec les marques chromatiniennes présentes aux locus cibles. D'autre part, certaines épimutations induites par la mutation *sg1* peuvent être maintenues pendant plusieurs générations. Pour rechercher un lien entre méthylation des gènes et conséquences phénotypiques, nous avons caractérisé des épimutations liées à un défaut de développement de la fleur et identifié un certain nombre de gènes candidats potentiellement responsables du phénotype. Les résultats obtenus au cours de ma thèse ont contribué à préciser le rôle joué par le complexe SG1 et à comprendre le lien entre celui-ci et les marques épigénétiques.

Title : Involvement of SG1 protein in maintaining the epigenomes in *Arabidopsis thaliana*.

Keywords: INCREASE IN BONSAI METHYLATION (IBM), FPA, polyadenylation, epialleles, *Arabidopsis*.

Abstract : Chromatin is known to contain the genetic information and its structure and transcriptional state can be regulated by epigenetic modifications. Repressive marks such as DNA and histones methylation needs to be kept away from gene bodies to enable the proper development of the plant. In *Arabidopsis thaliana*, *sg1* mutants show a range of severe developmental defects similar to those observed in mutants affected in epigenetic pathways. We have shown that *sg1* mutant phenotype is caused by an increase of CHG and H3K9me2 methylation in many gene bodies. Indeed, SG1 regulates the histone demethylase *IBM1* transcription and the impairment observed in *sg1* mutant epigenomes is caused by *IBM1* misregulation. We found seven proteins interacting with SG1, among which some partners are able to bind chromatin marks. Through a suppressor screen we identified FPA, already known to regulate the polyadenylation of some transcripts, as a player involved in SG1 targets regulation, including *IBM1*. Our results show that the SG1 complex regulates target genes transcription by affecting polyadenylation site choice, in a way that remains to be determined, in a chromatin marks dependent manner. We also found that some of the *sg1*-induced epimutations can be maintained through several generations. To investigate the link between gene body methylation and phenotypic consequences, we have characterized epimutations related to a defect in floral development and identified some candidate genes potentially responsible for the floral phenotype. Thus, our results contributed to clarify the role of SG1 and to understand its connection with epigenetic marks.

THESE DE DOCTORAT
DE L'UNIVERSITE PARIS-SACLAY,
préparée à l'Université Paris-Sud

ÉCOLE DOCTORALE N° 567
Sciences du Végétal : du Gène à l'Ecosystème

Spécialité de doctorat : Biologie

Par

Aurélie DEREMETZ

Implication de la protéine SG1 dans
le maintien des épigénomes chez *Arabidopsis thaliana*.

Thèse présentée et soutenue à Versailles, le 3 décembre 2015 :

Composition du Jury :

M. DERAGON Jean-Marc	Professeur, Université de Perpignan	Rapporteur
M. MATHIEU Olivier	Directeur de Recherche, CNRS, Clermont-Ferrand	Rapporteur
Mme DELARUE Marianne	Professeure, Université Paris-Sud	Examinatrice
M. ROUDIER François	Chargé de Recherche, CNRS, Paris	Examinateur
Mme VITTE Clémentine	Chargée de Recherche, CNRS, Le Moulon	Examinatrice
M. BOUCHE Nicolas	Chargé de Recherche, INRA, Versailles	Directeur de thèse



Titre : Implication de la protéine SG1 dans le maintien des épigénomes chez *Arabidopsis thaliana*.

Mots clés : INCREASE IN BONSAI METHYLATION (IBM), FPA, polyadénylation, épiallèles, *Arabidopsis*.

Résumé : La chromatine est le support de l'information génétique et sa structure, ainsi que son activité transcriptionnelle, peuvent être modulées par des modifications épigénétiques. Le maintien des marques répressives telles que la méthylation de l'ADN et des histones, hors du corps des gènes, est nécessaire pour le bon développement de la plante. Chez *Arabidopsis thaliana*, le mutant *sg1* présente des défauts développementaux sévères caractéristiques de mutants affectés dans des mécanismes épigénétiques. Nous avons montré que le phénotype de *sg1* est causé par une hyperméthylation CHG et H3K9me2 dans de nombreux gènes. En effet, SG1 contrôle la transcription de l'histone déméthylase IBM1 et les modifications de l'épigénome observées chez *sg1* sont dues à une dérégulation de IBM1. Nous avons identifié sept protéines partenaires de SG1, dont certaines se lient aux marques chromatiniennes. Nous avons réalisé un crible suppresseur qui a permis d'identifier FPA, une protéine régulant la polyadénylation de certains transcrits, comme acteur impliqué dans le contrôle des cibles de SG1, dont IBM1. Nos résultats montrent que le complexe SG1 régule la transcription de ses cibles en influençant, par un mécanisme encore inconnu, le choix du site de polyadénylation, en lien avec les marques chromatiniennes présentes aux locus cibles. D'autre part, certaines épimutations induites par la mutation *sg1* peuvent être maintenues pendant plusieurs générations. Pour rechercher un lien entre méthylation des gènes et conséquences phénotypiques, nous avons caractérisé des épimutations liées à un défaut de développement de la fleur et identifié un certain nombre de gènes candidats potentiellement responsables du phénotype. Les résultats obtenus au cours de ma thèse ont contribué à préciser le rôle joué par le complexe SG1 et à comprendre le lien entre celui-ci et les marques épigénétiques.

Title : Involvement of SG1 protein in maintaining the epigenomes in *Arabidopsis thaliana*.

Keywords: INCREASE IN BONSAI METHYLATION (IBM), FPA, polyadenylation, epialleles, *Arabidopsis*.

Abstract : Chromatin is known to contain the genetic information and its structure and transcriptional state can be regulated by epigenetic modifications. Repressive marks such as DNA and histones methylation needs to be kept away from gene bodies to enable the proper development of the plant. In *Arabidopsis thaliana*, *sg1* mutants show a range of severe developmental defects similar to those observed in mutants affected in epigenetic pathways. We have shown that *sg1* mutant phenotype is caused by an increase of CHG and H3K9me2 methylation in many gene bodies. Indeed, SG1 regulates the histone demethylase *IBM1* transcription and the impairment observed in *sg1* mutant epigenomes is caused by *IBM1* misregulation. We found seven proteins interacting with SG1, among which some partners are able to bind chromatin marks. Through a suppressor screen we identified FPA, already known to regulate the polyadenylation of some transcripts, as a player involved in SG1 targets regulation, including *IBM1*. Our results show that the SG1 complex regulates target genes transcription by affecting polyadenylation site choice, in a way that remains to be determined, in a chromatin marks dependent manner. We also found that some of the *sg1*-induced epimutations can be maintained through several generations. To investigate the link between gene body methylation and phenotypic consequences, we have characterized epimutations related to a defect in floral development and identified some candidate genes potentially responsible for the floral phenotype. Thus, our results contributed to clarify the role of SG1 and to understand its connection with epigenetic marks.

Partie I : Introduction	3
(I) La chromatine	5
1) La structure de la chromatine	5
2) L'organisation de la chromatine dans le génome d' <i>Arabidopsis</i>	6
(II) La transcription des gènes chez <i>Arabidopsis thaliana</i>	7
1) La transcription des gènes par l'ARN polymérase II	7
a. <i>Le complexe d'initiation de la transcription</i>	7
b. <i>Elongation du transcrit</i>	8
c. <i>Terminaison de la transcription</i>	9
2) Maturation des ARN	9
a. <i>La pose de la coiffe en 5'</i>	9
b. <i>L'épissage des ARN messagers</i>	10
c. <i>Polyadénylation et choix de son site</i>	11
(III) La dynamique de la chromatine	15
1) Méthylation de l'ADN	15
a. <i>Le maintien de la méthylation de type CG</i>	16
b. <i>Le maintien de la méthylation CHG</i>	17
c. <i>Le maintien de la méthylation CHH et la voie du RdDM</i>	18
d. <i>La déméthylation de l'ADN</i>	20
e. <i>L'empreinte parentale</i>	22
f. <i>Les paramutations</i>	22
2) Variants d'histones	23
3) Modifications post-traductionnelles des histones	25
a. <i>Acétylation, phosphorylation et ubiquitination des histones</i>	25
Acétylation des histones	25
Phosphorylation des histones	26
Ubiquitination des histones	26
b. <i>Méthylation des histones</i>	27
La méthylation de H3K4	28
La méthylation de H3K36	29
La méthylation de H3K27	29
4) Cas de la méthylation H3K9	31
a. <i>Répartition des marques H3K9</i>	32
b. <i>Etablissement de la méthylation H3K9</i>	32
c. <i>Interaction entre H3K9me2 et méthylation de l'ADN</i>	33
d. <i>Régulation de la méthylation H3K9 dans les régions géniques</i>	34
Identification de <i>INCREASE IN BONSAI METHYLATION1 (IBM1)</i>	34
Caractérisation de la fonction de IBM1	35
Régulation du locus <i>IBM1</i>	36
5) Les réseaux de marques épigénétiques	37
(IV) Héritabilité de la méthylation de l'ADN	39
1) Epiallèles naturels	39
2) Variabilité épigénétique naturelle	41
3) Epiallèles induits par mutation	43
(V) Sujet et objectifs de la thèse	45
Partie II : La protéine SG1 permet le maintien de l'épigénome d'<i>Arabidopsis thaliana</i> en régulant <i>IBM1</i>	47
(I) Introduction : identification du locus <i>SG1</i>	47
(II) Article 1 : « SHOOT GROWTH1 maintiens <i>Arabidopsis</i> epigenomes by preventing ectopic DNA and histone methylation of gene bodies » PLoS ONE, 2014.	50

(III) Conclusion de l'article	77
Partie III : La transcription de <i>IBM1</i> est régulée par différentes voies.....	79
(I) Introduction.....	79
(II) Article 2 : « Multiple pathways are involved in the processing of the functional <i>IBM1</i> transcript »	82
(III) Conclusion de l'article	106
Partie IV : Discussion générale sur la fonction de SG1.	108
Les régions hétérochromatiques dans les introns des cibles de SG1 sont-elles un obstacle ou un signal pour la transcription longue ?	108
Quels sont les rôles respectifs des différents membres du complexe SG1 ?	109
Quelle est l'étendue des cibles de SG1 ?	112
Comment le complexe SG1 influence-t-il le choix du site de polyadénylation ?	113
Partie V : Caractérisation d'épimutations apparaissant dans un fond génétique <i>sg1</i>.....	117
(I) Génération des populations d'épimutants <i>sg1</i> induits	118
(II) Etude du phénotype floral des populations d'épimutants.....	119
1) Suivi du phénotype floral	119
2) Sélection des plantes en génération F2	119
3) Phénotypage des populations F3	120
(III) Identification des épimutations associées au phénotype floral	122
1) Stabilité du phénotype floral	122
2) Cartographie des épimutations	123
3) Identification des épimutations par séquençage.....	124
a. <i>Matériel biologique utilisé</i>	124
b. <i>Séquençage et alignement</i>	125
c. <i>Analyse des données de méthylome</i>	126
d. <i>Recoupement des données de cartographie/méthylome : approche gène candidat</i>	128
(IV) Discussion et perspectives	130
Des épimutations peuvent être induites par la perte de fonction de <i>IBM1</i>	130
Le phénotype floral suivi est-il lié à des épimutations ?	131
Héritabilité des épimutations induites	133
Conclusion.....	136
(V) Matériels et méthodes.....	137
Préparation des échantillons d'ADN envoyés au séquençage.	137
Conclusion générale.....	139
Références bibliographiques.....	140

Partie I : Introduction

Depuis sa découverte par Watson et Crick dans les années 50, l'acide désoxyribonucléique, ou ADN, est considéré comme le support de l'information génétique et permet la transmission des caractères de génération en génération (Watson and Crick, 1953). Ainsi, l'ADN est le promoteur de la diversité du vivant. Pourtant, contrairement aux concepts de génétique classique avancés par Mendel, de nombreux travaux ont contribué à démontrer que le phénotype n'est pas déterminé uniquement par la séquence de l'ADN. En effet, la génétique ne suffit pas à expliquer à elle seule la diversité au sein d'une espèce, voire au sein d'un même organisme. Toutes les cellules d'un même organisme partagent pourtant la même séquence d'ADN, néanmoins, elles n'ont pas les mêmes caractéristiques, ni les mêmes fonctions.

Le terme « épigénétique » a été inventé par Conrad Waddington qui, dès 1942, décrit l'épigénétique comme la branche de la biologie qui s'intéresse à l'interaction entre les gènes, ainsi que leurs produits, et qui transforme le phénotype en un être vivant (Waddington, 1942). De nos jours, cette définition se précise et est décrite comme l'étude des changements héréditaires qui régulent l'information contenue dans les gènes, transmissibles d'une génération à l'autre, sans qu'il n'y ait de modification de la séquence de l'ADN. Ainsi, l'épigénétique est impliquée dans divers processus développementaux et va contribuer à permettre à toutes les cellules d'un organisme, qui portent la même séquence d'ADN, à utiliser des programmes d'expression qui vont être différents d'un type cellulaire à l'autre, ou au sein d'une même cellule au cours de sa vie.

Chez les plantes, comme l'organisme modèle *Arabidopsis thaliana*, le manque de mobilité les contraint à faire face à différents stress biotiques et abiotiques et donc à s'adapter rapidement à leur environnement. Certains des processus, mis en place pour répondre à ces stress, se font par l'intermédiaire de mécanismes de régulation épigénétique. De plus, ces mécanismes sont également impliqués dans de nombreux processus développementaux, comme la floraison ou l'architecture de la plante.

Il est désormais établi que les modifications épigénétiques sont inscrites sur la chromatine pour déterminer sa structure, son organisation et son activité transcriptionnelle. L'étude des mécanismes sous-jacents est essentielle pour comprendre comment ces modifications agissent sur la chromatine pour contribuer à la régulation de l'expression des gènes et des programmes développementaux.

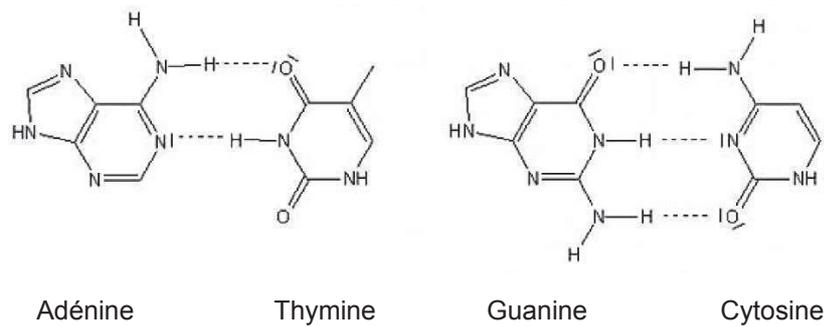


Figure I-1 : Formules des différentes bases azotés de l'ADN et leur appariement.

Les bases complémentaires s'apparient deux à deux par l'intermédiaire de liaisons hydrogènes. Les adénines sont liées aux thymines par des doubles liaisons, alors que les guanines et cytosines sont liées par des triples liaisons. Chacune des bases de la double hélice d'ADN est appariée à son complémentaire.

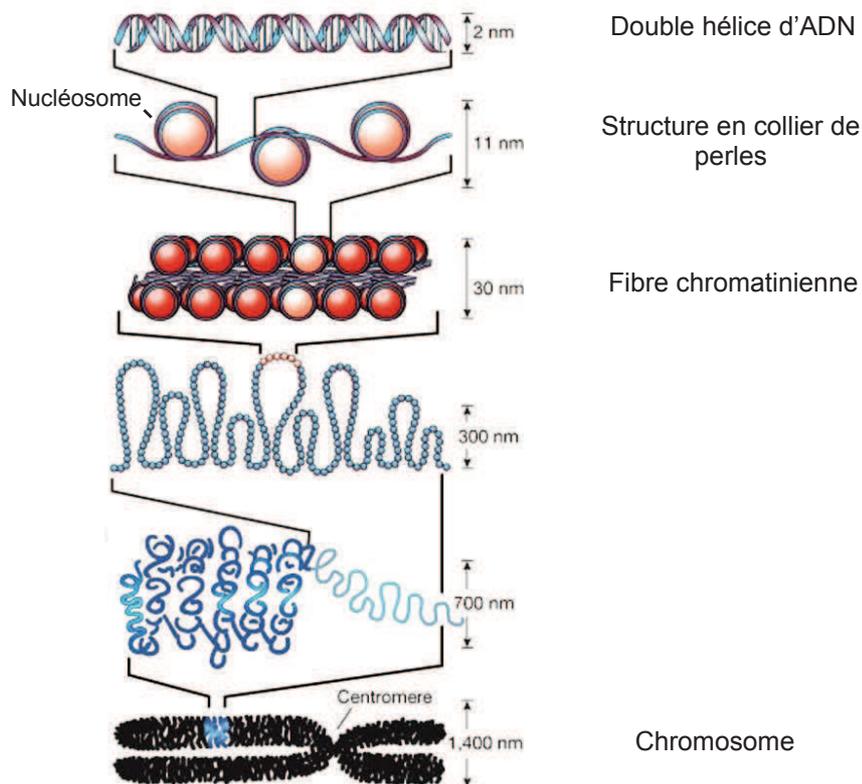


Figure I-2 : Les étapes de compaction de la chromatine, de l'ADN au chromosome. D'après Felsenfeld and Groudine, 2003.

(I) La chromatine

1) La structure de la chromatine

Chez les eucaryotes, l'ADN est contenu dans le noyau sous forme compactée, associé à différentes protéines, le tout formant la chromatine. L'ADN est composé d'une succession d'unités nucléotidiques (**Figure I-1**). Dans les noyaux des cellules eucaryotes, environ 150 pb d'ADN viennent s'enrouler autour d'un octamère constitué de deux copies de chacune des quatre histones H2A, H2B, H3 et H4. Les histones sont des protéines très conservées qui ont une structure secondaire constituée d'une succession de trois hélices alpha. Elles se caractérisent par des extrémités C-terminales qui comportent un domaine « *histone-fold* » conservé et des extrémités N-terminales, exposées à l'extérieur du nucléosome, qui peuvent être modifiées de manière post-traductionnelle (Luger et al., 1997).

Cette succession de nucléosomes, séparés par une région d'ADN de liaison de taille variable selon les organismes, ne constitue que la première étape de compaction de l'ADN. Par la suite, cette structure, dite en « collier de perle » de 11 nm de diamètre se compacte pour former la fibre chromatinienne de 30 nm que vient stabiliser l'histone de liaison H1. Au cours de la métaphase de mitose, cette fibre subit des repliements de plus en plus complexes jusqu'à atteindre son état maximum de compaction sous forme de chromosome (**Figure I-2**) (Fransz et al., 2011).

C'est l'observation cytologique du noyau interphasique qui, dès 1928, a permis la distinction de deux états de la chromatine dans le contenu cellulaire. D'une part l'euchromatine, qui correspond à une chromatine peu condensée et considérée comme transcriptionnellement active. D'autre part l'hétérochromatine, qui est une forme de chromatine très condensée, visible au microscope sous forme de régions denses appelées chromocentres.

Ces deux états chromatinien ne sont pas distribués aléatoirement dans le génome. Chez *Arabidopsis thaliana*, l'euchromatine correspond aux bras des chromosomes, riches en gènes, contrairement à l'hétérochromatine qui peut être distinguée sous deux formes : l'hétérochromatine constitutive, qui regroupe l'ensemble des régions centromériques, péri-centromériques et télomériques, riches en séquences répétées, et l'hétérochromatine

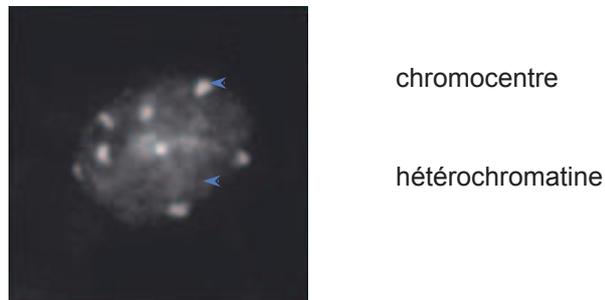


Figure I-3 : Noyau d'*Arabidopsis thaliana* vu en microscopie à fluorescence.

La coloration du noyau au 4,6-diamidino-2-phénylindole (DAPI) qui est un intercalant de l'ADN, montre des régions intensément marquées qui correspondent à l'hétérochromatine, état condensé de la chromatine. Ces régions bien visibles sont appelées « chromocentres ». Les zones d'intensité plus diffuse correspondent à l'euchromatine, plus décondensée (Fransz and Hans de Jong, 2011).

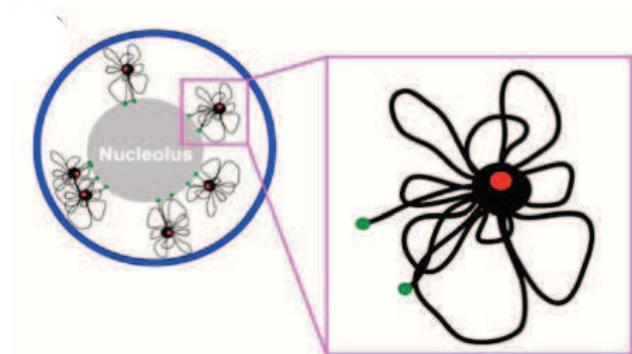


Figure I-4 : Structure de la chromatine en rosette chez *Arabidopsis thaliana*.

Représentation de l'organisation des chromosomes dans le noyau interphasique : les régions hétérochromatiques, dont le centromère (en rouge), forment le chromocentre (en noir) duquel partent les boucles d'euchromatine. Les chromocentres sont localisés à la périphérie de l'enveloppe nucléaire (en bleu). Les télomères (en vert) se placent autour du nucléole (en gris). D'après Pawlowski, 2010.

facultative, qui correspond à des régions euchromatiniennes qui se compactent par hétérochromatisation suite à l'action de protéines, comme les variants d'histones, les enzymes de modification des histones, ou les méthyltransférases (**Figure I-3**).

2) L'organisation de la chromatine dans le génome d'*Arabidopsis*

L'organisation de la chromatine dans le noyau peut varier selon la taille du génome et la structure des chromosomes. Pour des petits génomes (de taille inférieure à 500 Mb), comparables à celui d'*Arabidopsis*, la chromatine s'organise autour d'un chromocentre alors que pour des génomes de taille plus importante, qui comportent des régions hétérochromatiques qui s'étendent au niveau des bras des chromosomes, la chromatine peut s'organiser autour de plusieurs chromocentres, voire être en grande partie condensée de sorte que l'euchromatine et l'hétérochromatine ne peuvent plus être distinguées (van Driel and Fransz, 2004).

La plante modèle *Arabidopsis thaliana* possède un génome de petite taille dont 115,4 Mb ont été séquencées. La taille des composants non séquencés, à savoir les centromères et les séquences d'ADN ribosomiques, est estimée à 10 Mb. Ainsi, il a été déterminé que le génome d'*Arabidopsis* a une taille totale de 125 Mb (The Arabidopsis Genome Initiative, 2000). Il comprend 27 416 gènes, 4 827 pseudogènes et éléments transposables, ainsi que plus de 1 300 ARN non codants d'après les prédictions informatiques (TAIR10). Chez *Arabidopsis thaliana*, l'hétérochromatine est essentiellement localisée dans les régions centromériques et péri-centromériques. Au cours de l'interphase, ces régions se regroupent autour des chromocentres, au nombre de 6 à 10, d'où partent des boucles d'euchromatine. Cette structure de la chromatine est dite « en rosette » (**Figure I-4**) (Fransz et al., 2002). Cette organisation n'est pas contrainte, car la compaction des chromocentres peut être régulée. Ainsi, lorsque les régions péri-centromériques sont décompactées, comme c'est le cas chez certains mutants, les chromocentres sont de taille réduite (Soppe et al., 2002).

(II) La transcription des gènes chez *Arabidopsis thaliana*

Cinq ARN polymérases ont été mises en évidence chez *Arabidopsis thaliana* ; Les ARN polymérases I à III sont décrites dans tous les organismes eucaryotes, alors que les ARN polymérases IV et V sont spécifiques des plantes. L'ARN polymérase I permet la transcription des ARN ribosomiques (ARNr) 28S, 18S et 5,8S. L'ARN polymérase II transcrit une grande majorité d'ARN messagers (ARNm) qui codent pour des protéines, mais transcrit également certains ARN non codants (Zheng et al., 2009). L'ARN polymérase III, transcrit majoritairement des ARN de transfert (ARNt), l'ARNr 5S, et des petits éléments nucléaires intercalés (SINE). L'ARN Polymérase IV est impliquée dans la synthèse des ARN qui servent de matrice pour générer des petits ARN interférents de 24 nucléotides et l'ARN Polymérase V synthétise l'ARN non codant qui sera ciblé par ces petits ARN dans le but de catalyser la méthylation de l'ADN *de novo* (Herr et al., 2005; Kanno et al., 2005).

1) La transcription des gènes par l'ARN polymérase II

La transcription par l'ARN polymérase II est la plus étudiée chez les eucaryotes puisqu'elle est responsable de la synthèse des ARNm à partir des gènes codant des protéines. Elle peut se décomposer en trois étapes majeures qui sont l'initiation de la transcription, l'élongation du transcrit et la terminaison de la transcription. Contrairement à ce qui peut être observé chez les procaryotes, l'ARN polymérase II requiert la présence de diverses protéines pour initier la transcription.

a. *Le complexe d'initiation de la transcription*

Il existe une séquence consensus TATAA (ou TATA box) présente dans les promoteurs des gènes transcrits par l'ARN polymérase II, environ 25 à 30 nucléotides en amont du site d'initiation de la transcription. Cinq facteurs TF sont impliqués dans l'initiation de la transcription par l'ARN polymérase II : TFIIB, TFIID, TFIIE, TFIIIF et TFIIH.

La TATA box est d'abord reconnue par le facteur de transcription TFIID, composé de plusieurs sous-unités, dont TBP (*TATA-Binding Protein*) qui recrute TFIIB pour former ainsi le complexe de pré-initiation. La polymérase vient alors se lier au promoteur par l'intermédiaire du complexe TBP-TFIIB et de TFIIF (Buratowski and Zhou, 1993). La présence ou non de facteurs généraux de transcription peut modifier l'affinité de la polymérase pour le promoteur. La liaison des facteurs TFIIE et TFIIH au complexe de pré-initiation est ensuite nécessaire pour initier la transcription.

Deux des sous-unités de TFIIH sont des hélicases qui sont impliquées dans le déroulement de l'hélice d'ADN et provoquent l'ouverture des deux brins, permettant ainsi l'initiation de la transcription. Une autre de ses sous-unités est une kinase qui participe à la phosphorylation des sérines d'un heptapeptide répété situé sur le domaine C-terminal de la polymérase, appelé CTD (Corden, 1990; Glover-Cutter et al., 2009). La phosphorylation de ces séquences jouerait un rôle dans la libération de la polymérase vis-à-vis du complexe d'initiation et permettrait alors de poursuivre avec l'étape d'élongation. La transition entre l'initiation et l'élongation fait intervenir des réarrangements structuraux importants, notamment des changements de facteurs de transcription liés à la phosphorylation du CTD (Phatnani and Greenleaf, 2006).

b. *Elongation du transcrit*

L'étape d'élongation correspond à la polymérisation de l'ARN du 5' vers le 3', à partir du site d'initiation du transcrit, le long de la séquence du brin ADN anti-sens. Ainsi, la séquence de l'ARN obtenu est identique à celle du brin sens, à la différence près que les Thymines de l'ADN sont remplacées par des Uraciles. Cette polymérisation se fait par incorporation successive de nucléotides triphosphate (NTP), dépendamment de la séquence ADN. Le site d'addition de la polymérase reconnaît le NTP et, s'il est complémentaire à la séquence de l'ADN, il y a formation d'un pont phosphodiester, puis le nucléotide est ajouté à la séquence de l'ARN déjà initié par translocation (Boeger et al., 2005).

Au fur et à mesure de l'avancée de la polymérase, des mécanismes sont mis en place pour que celle-ci puisse progresser malgré la barrière que forment les nucléosomes, autour desquels est enroulé l'ADN. En effet, pour faciliter l'élongation, les nucléosomes peuvent être déplacés ou

partiellement désassemblés, impliquant différents facteurs comme le complexe FACT (*Facilitates Chromatin Transcription*) et des facteurs de remodelage de la chromatine (Belotserkovskaya et al., 2003).

c. Terminaison de la transcription

La terminaison de la transcription est déclenchée lorsque la polymérase rencontre un ou plusieurs facteurs de terminaison, comme le signal de polyadénylation conservé AAUAAA. Lorsque la polymérase transcrit ce signal, il est reconnu par des facteurs protéiques de terminaison qui accompagnent la polymérase et qui se lient à l'ARN néosynthétisé. Celui-ci est alors clivé au niveau du signal et toute la séquence synthétisée en aval est dégradée. La polymérase se sépare de la matrice ADN et l'ARN libéré sera par la suite polyadénylé à son extrémité 3'.

2) Maturation des ARN

A l'issue de la transcription par l'ARN polymérase II, les transcrits ne sont pas matures et ces préARNm doivent subir un certain nombre de modifications avant de pouvoir être traduits en protéine. Le contrôle se fait par des ribonucléoprotéines qui vérifient la maturation correcte de l'ARNm avant de l'exporter pour la traduction. Il existe, chez les eucaryotes, trois étapes de maturation du transcrit qui aboutissent à un ARNm fonctionnel : la pose de la coiffe à l'extrémité 5', l'épissage, et la polyadénylation en 3'.

a. La pose de la coiffe en 5'

La pose de la coiffe en 5' ou « capping » est la première étape de maturation du transcrit. Elle permet de stabiliser l'ARN avant qu'il ne soit exporté du noyau et de le protéger ainsi de la dégradation. Elle se déroule dès le début de la polymérisation du transcrit, lorsque celui-ci n'est long que de quelques dizaines de nucléotides.

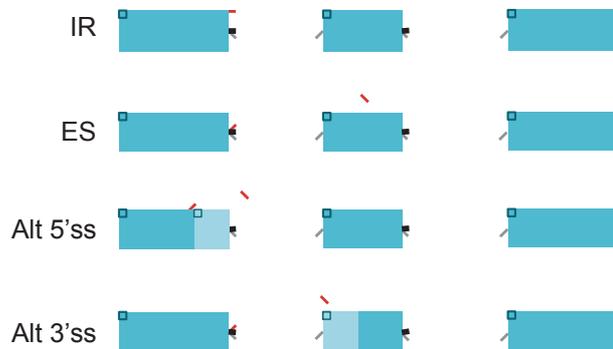


Figure I-5 : Les différents types d'événements d'épissage alternatif.

Représentation des quatre types principaux d'épissage alternatif observés chez les plantes : la rétention d'intron (IR) qui est le plus fréquent et qui inclut un intron dans la séquence codante, le saut d'exon (ES) qui entraîne l'exclusion d'un exon, l'épissage à extrémité 5' ou 3' alternative (Alt 5' ou 3'ss) où des sites 5' et 3' différents sont sélectionnés pour l'épissage. Les cassettes bleues correspondent aux exons et les lignes noires aux introns. Les traits gris représentent l'épissage et les traits rouges marquent les différents types d'épissage alternatif.

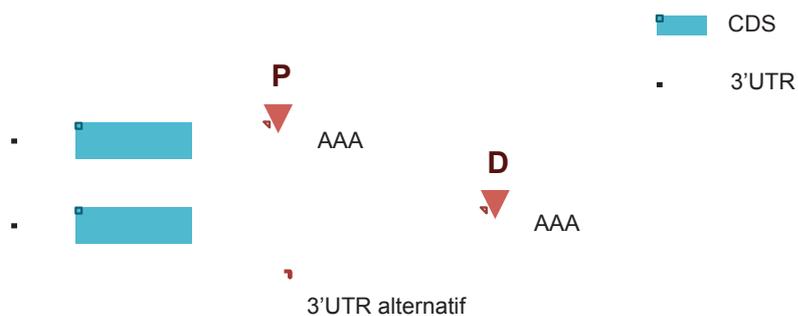


Figure I-6 : La polyadénylation alternative en 3'-UTR.

La polyadénylation alternative dans la région 3'-UTR permet de générer des isoformes de transcrits qui diffèrent par leur 3'-UTR. La région située entre les sites de polyadénylation proximal (P) et distal (D) est appelée UTR alternatif.

Des facteurs enzymatiques, dont une guanyl-transférase et une méthylase, se fixent sur le CTD phosphorylé de la polymérase. La guanyl-transférase catalyse l'ajout d'une guanine sur l'ARN en cours de transcription (Shuman, 1995). Ceci marque le début de la pose de la coiffe. Par la suite, la méthyltransférase ajoute un groupement méthyle en position 5 de la guanine de la coiffe. Puis, un autre groupement méthyle sera ajouté sur le ribose du premier nucléotide. Une fois la coiffe ajoutée au préARNm, le facteur guanyl-transférase se détache du CTD et sera par la suite remplacé par d'autres facteurs impliqués dans le recrutement du spliceosome, qui est le complexe ribonucléoprotéique qui catalyse l'excision des introns lors de l'épissage.

b. L'épissage des ARN messagers

Chez les eucaryotes, les gènes sont constitués d'une alternance entre les exons, parties codantes du gène, et les introns, non codants. Les introns sont délimités par des séquences spécifiques (GU en 5' et AG en 3'). Ces séquences spécifiques qui bordent les introns sont conservées chez les plantes. La polymérase transcrit l'intégralité du gène, puis deux réactions de transestérification se succèdent, les introns présents dans le transcrit sont excisés et les exons sont ligaturés entre eux pour produire l'ARNm mature. Cette étape de maturation de l'ARN a lieu au fur et à mesure de la progression de la machinerie de transcription et fait intervenir le grand complexe ribonucléoprotéique du spliceosome.

Certains préARNm subissent un épissage alternatif qui permet de produire des isoformes d'ARNm qui peuvent coder des protéines de caractéristiques et de fonctions différentes à partir d'un même gène. De cette façon, l'épissage alternatif a pour conséquence majeure d'augmenter la diversité des protéines en incluant ou excluant des séquences ou domaines protéiques.

Quatre types principaux d'épissages alternatifs ont été décrits chez les plantes : la rétention d'intron (IR), le saut d'exon (ES), et les épissages à extrémités 5' ou 3' alternatives (Alt 5'ss et Alt 3'ss) (**Figure I-5**). La survenue d'un événement d'épissage alternatif et son type sont déterminés par des signaux portés par la séquence du transcrit et par leur reconnaissance par des facteurs d'épissage. Un grand nombre de ces événements a lieu dans la séquence codante et peuvent générer l'apparition de codons stop précoces. Ces transcrits tronqués et non

fonctionnels sont par la suite dégradés. L'épissage alternatif des transcrits apporte également un degré de régulation supplémentaire à l'expression des gènes (Kalyna et al., 2012). Quand ils ont lieu dans les régions 5'- ou 3'-UTR, ils peuvent influencer sur l'export des ARNm, ou faire apparaître des codons d'initiation et des sites de polyadénylation différents.

Chez *Arabidopsis thaliana*, environ 60% des gènes qui comportent des introns sont soumis à un épissage alternatif (Marquez et al., 2012). Des analyses à l'échelle du génome montrent que parmi les différents types d'épissage alternatif connus, la rétention d'intron est l'événement le plus fréquemment observé (Filichkin et al., 2010).

c. Polyadénylation et choix de son site

La polyadénylation de l'extrémité 3' du préARNm est un processus qui est couplé à la terminaison de la transcription. En effet, l'ARN polymérase ne s'arrête pas précisément au niveau du signal de terminaison, mais poursuit la transcription plusieurs centaines de nucléotides après le site de polyadénylation. Ce signal de polyadénylation, dont la séquence est conservée chez les eucaryotes, est reconnu par le facteur CPSF (*Cleavage and Polyadenylation Specific Factor*). Le facteur CPSF est constitué de différentes sous-unités capables de reconnaître et lier des éléments de séquence spécifiques en cis. Ces sous-unités, ainsi que toutes celles qui composent la machinerie qui prend en charge l'extrémité 3' des ARN, sont relativement conservées chez les eucaryotes (Mandel et al., 2008). Chez la levure *S. cerevisiae* le signal de polyadénylation est composé de trois éléments : un « *Efficiency Element* » (EE) riche en AU, un « *Positioning Element* » (PE) riche en A et des régions riches en U de part et d'autre du site de clivage (cf figure 1, Mandel et al., 2008). Chez *A. thaliana*, ces éléments sont appelés « *Near Upstream Element* » (NUE), « *Far Upstream Element* » (FUE) et « *Cleavage Element* » (CE). Par la suite, d'autres facteurs vont venir se lier à CPSF pour cliver la partie 3'-UTR. La fixation de ces facteurs au cours de la transcription va plier le transcrit qui sera clivé par l'activité endonucléase de la sous-unité CPSF30. Celle-ci génère un transcrit dont l'extrémité 3' porte un groupement hydroxyle libre (Addepalli and Hunt, 2007). Ce clivage entraîne une déstabilisation du complexe formé par la polymérase II avec l'ARN, ce qui aboutit à la terminaison de la transcription. A l'extrémité 3' vont être ajoutées successivement plusieurs adénosines monophosphate par une Poly(A) Polymérase. Il s'agit d'une polymérase particulière capable de synthétiser la queue poly(A),

sans utiliser de matrice ADN, qui stabilise l'ARNm, dernière étape de sa maturation, essentielle à son export, à sa régulation et à sa traduction.

De nombreux gènes codants possèdent plusieurs sites de polyadénylation. Ainsi, ces gènes ont la capacité de coder plusieurs isoformes d'ARNm qui diffèrent d'un point de vue de la taille de leurs 3'-UTR (**Figure I-6**). Deux études ont été menées à l'échelle du génome pour déterminer la position des sites polyadénylation chez *Arabidopsis*. Elles montrent qu'une grande majorité des sites de polyadénylation sont retrouvés dans les séquences codantes et le 3'-UTR, alors qu'ils sont plus rarement présents au niveau du 5'-UTR et des introns (Shen et al., 2011; Wu et al., 2011).

Le choix du site de polyadénylation est un mécanisme qui est activement régulé. Il a été montré chez la souris, que la taille du 3'-UTR est négativement corrélée au niveau d'expression de certains facteurs qui prennent en charge l'extrémité 3' du transcrit. Ainsi, des altérations de l'activité de la machinerie responsable de la polyadénylation pourraient moduler le choix du site de polyadénylation (Ji et al., 2011). Par ailleurs, il a été montré que des interactions entre les machineries de transcription et de polyadénylation augmenterait la probabilité d'utilisation d'un site de polyadénylation proximal (Di Giammartino et al., 2011). De même, la sous-unité U1 du spliceosome semble jouer un rôle dans le choix du site de polyadénylation puisqu'un niveau plus faible de cette protéine favoriserait les sites proximaux, suggérant ainsi un lien entre le choix du site de polyadénylation et l'épissage des ARNm (Berg et al., 2012).

Les sites de polyadénylation sont particulièrement pauvres en nucléosomes alors que les régions en aval sont enrichies (Spies et al., 2009). Pour les gènes qui présentent plusieurs sites de polyadénylation, cette répartition des nucléosomes est d'autant plus vraie pour le site de polyadénylation majoritaire, ce qui suggère que la chromatine et d'éventuelles modifications épigénétiques pourraient influencer le choix du site de polyadénylation.

Le choix du site de polyadénylation, selon s'il est distal ou proximal, va inclure ou exclure certaines régions de l'ARNm, et peut donc modifier la séquence de la protéine qui sera traduite. De plus, des variations de la région 3'-UTR peuvent masquer ou révéler des sites régulateurs, comme par exemple des sites de liaison pour des microARN. Par conséquent, le choix du site de polyadénylation est important car il conditionne la fonctionnalité de l'ARNm.

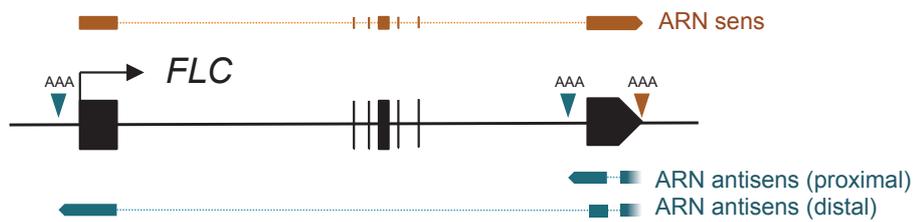


Figure I-7 : Transcriptions sens et antisens du locus *FLC*.

Ce schéma illustre le gène *FLC*, avec ses exons schématisés par les boîtes noires. Le transcrit sens est représenté en marron, et les deux transcrits alternatifs antisens sont représentés en bleu. Le transcrit antisens généré dépend du site de polyadénylation utilisé. Les sites de polyadénylation (AAA) sont indiqués sur la séquence du gène par des flèches en accord avec le code couleur des transcrits.

Chez *Arabidopsis thaliana*, la protéine FPA est connue pour jouer un rôle dans la formation des extrémités 3' des ARNm, en particulier dans la régulation du choix du site de polyadénylation (Hornyik et al., 2010b). Le gène *FPA* a été initialement identifié lors d'un crible génétique pour rechercher des gènes impliqués dans la floraison (Koornneef et al., 1991; Schomburg et al., 2001). En effet, il a été montré que la mutation *fpa* entraîne un retard de floraison, indépendamment de la photopériode (Schomburg et al., 2001). La protéine FPA comprend trois motifs RRM (*RNA-Recognition Motif*) répétés à son extrémité N-terminale et à son extrémité C-terminale un domaine SPOC (*Spen Paralog and Ortholog C-terminal*), de fonction inconnue, mais homologue aux protéines de la famille Spen qui regroupe des protéines connues pour jouer un rôle crucial dans le déterminisme cellulaire et le développement chez les animaux (Ariyoshi and Schwabe, 2003).

Il a été montré que FPA a une influence sur la floraison en régulant *FLOWERING LOCUS C* (*FLC*). Le gène *FLC* est connu pour être un répresseur majeur de la floraison. Il a été montré, par des analyses d'expression à l'échelle du génome, qu'il existe des ARN antisens correspondant au locus *FLC* (Yamada et al., 2003). Deux types de transcrits antisens de *FLC* ont été mis en évidence: l'un est long et correspond au choix du site de polyadénylation distal et l'autre est plus court, s'arrêtant au niveau du site de polyadénylation proximal situé dans le sixième intron de *FLC* (**Figure I-7**)(Liu et al., 2010). C'est la production du transcrit antisens court (proximal) qui entraîne la réduction d'expression du transcrit sens. La protéine FPA permet de limiter l'expression sens de *FLC* en favorisant l'utilisation du site de polyadénylation proximal (Hornyik et al., 2010a; Liu et al., 2010). La floraison tardive observée chez le mutant *fpa* (Michaels and Amasino, 2001) est donc due à l'accumulation du transcrit sens *FLC* qui n'est plus dégradé. Le choix du site proximal du transcrit antisens est corrélé à l'activité de FPA (Hornyik et al., 2010a; Liu et al., 2010). Sans FPA, le transcrit antisens le plus court n'est plus produit. De la même façon que FPA, FCA possède des motifs RRM et est impliquée dans la régulation du délai de floraison (Michaels and Amasino, 2001). Il a été montré que FCA est capable de s'autoréguler en influençant le choix de son propre site de polyadénylation (Quesada et al., 2003). FCA est également connue pour réguler le locus *FLC* en interaction avec FY (Simpson et al., 2003). FY a été identifié dans le même crible de floraison que FPA et affecte la formation de l'extrémité 3' des ARNm (Koornneef et al., 1991; Simpson et al., 2003). La protéine FY présente des homologies avec d'autres

protéines, qui sont également impliquées dans la formation de l'extrémité 3' des ARN, identifiées chez les animaux, mais aussi chez *S. cerevisiae* (Ohnacker et al., 2000).

Au delà de la floraison et de la régulation de *FLC*, FPA semble être impliqué dans d'autres processus du développement de la plante. Ainsi, il a été montré que FPA régule également l'état de la chromatine à un certain nombre de locus, de manière redondante avec FCA (Baurle et al., 2007). En effet, FPA est impliqué dans le maintien de la chromatine dans un état silencieux, comme c'est le cas pour le rétroélément *AtSNI* qui est réactivé et déméthylé dans un fond mutant *fpa* (Baurle et al., 2007). FPA est impliqué dans le contrôle du choix du site de polyadénylation d'autres ARNm que *FLC*, dont son propre transcrit (Duc et al., 2013; Hornyik et al., 2010b). En effet, FPA favorise, à son propre locus, l'utilisation du site de polyadénylation proximal, au niveau du premier intron, ce qui lui permet de s'autoréguler négativement (Hornyik et al., 2010b). Il a été montré également que FPA est capable d'influencer l'expression de certains gènes (Duc et al., 2013).

Même si l'implication de FPA, de manière redondante avec FCA, dans la régulation du locus *FLC* a été démontrée, des analyses de d'expression à l'échelle du génome montrent que FPA joue un rôle beaucoup plus étendu. Le mutant *fpa* et le double mutant *fpa fca* présentent des profils d'expression modifiés au niveau de nombreux gènes, éléments transposables et régions intergéniques non-annotées (Duc et al., 2013; Sonmez et al., 2011). En effet, dans un fond génétique *fpa fca* de nombreux segments intergéniques, qui ne s'expriment normalement pas dans un fond sauvage, sont réactivés et cette réactivation de la transcription s'accompagne d'une réduction de méthylation. Il a pu être noté que ces régions intergéniques identifiées sont situées généralement à l'extrémité 3' de gènes ou pseudogènes, et seraient transcrites dans la continuité de ceux-ci en l'absence de FPA et FCA (Sonmez et al., 2011). Ce résultat suggère que FPA et FCA sont impliqués dans un même processus de détermination du site de polyadénylation et de terminaison de la transcription à plus grande échelle. Les mécanismes sous-jacents à cette action conjointe de FPA et FCA restent pourtant à élucider. En effet, même si l'interaction entre FCA et FY permet d'expliquer l'implication de FCA dans ce processus, aucun lien direct ne permet encore de relier FPA à des composants de la machinerie de polyadénylation.

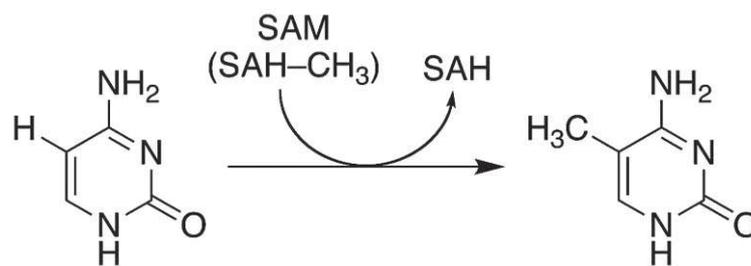


Figure I-8 : Réaction chimique de méthylation des cytosines.

La méthylation des cytosines se fait par l'intermédiaire d'enzymes qui catalysent le transfert d'un groupement méthyle (CH₃) du substrat, qui est la S-adénosylméthionine (SAM), vers le carbone en position 5 de la cytosine. La déméthylation de la SAM aboutit à la formation de S-adénosylhomocystéine (SAH).

(III) La dynamique de la chromatine

La chromatine est une structure dynamique qui va être modulée de différentes façons pour pouvoir assurer le bon développement de l'organisme. En effet, l'état de la chromatine peut être modulé par le remplacement des histones canoniques par des variants d'histones, mais aussi par l'intermédiaire de marques épigénétiques comme les modifications post-traductionnelles des histones, et la méthylation de l'ADN. Ces modifications vont permettre de remanier localement la structure de la chromatine et ses propriétés pour permettre à la machinerie de transcription d'accéder à l'ADN (Shu et al., 2012).

1) Méthylation de l'ADN

La méthylation de l'ADN est une marque épigénétique très étudiée et répandue au sein des eucaryotes. Son établissement et sa maintenance au cours de la réplication de l'ADN sont essentiels au bon déroulement de nombreux processus biologiques, en particulier dans les processus développementaux et l'extinction des transposons et séquences répétées.

Chez les eucaryotes, la méthylation de l'ADN se fait quasi exclusivement au niveau des cytosines. Elle consiste en l'ajout, sur le carbone cyclique en position 5 des cytosines, d'un groupement méthyle, catalysé par des ADN méthyltransférases, à partir d'un substrat donneur, la S-adenosylméthionine (Roberts and Cheng, 1998) (**Figure I-8**).

Chez les mammifères, l'étude de cette méthylation a montré qu'elle intervient majoritairement lorsque les cytosines sont dans un contexte symétrique CG. Il a été estimé que 70 à 80% des cytosines du génome en contexte CG sont méthylées (Ehrlich et al., 1982). Les profils de méthylation sont maintenus au cours des divisions cellulaires par une ADN méthyltransférase, appelée Dnmt1 (*DNA methyltransferase 1*), capable de reconnaître les cytosines hémiméthylées, c'est-à-dire méthylées sur un seul des deux brins d'ADN (Goll and Bestor, 2005). Néanmoins, il a été montré récemment que les cytosines au niveau des cellules souches embryonnaires humaines peuvent être méthylées dans un contexte non-CG, à savoir CHG ou CHH, où H peut être n'importe quel nucléotide sauf G (Lister et al., 2009).

Chez les plantes, la méthylation des cytosines se fait dans tous les contextes, symétriques (CG et CHG) et non-symétrique (CHH). Une étude s'intéressant à la méthylation du génome d'*Arabidopsis thaliana* (accession Columbia) a estimé les niveaux de méthylation des cytosines à 24%, 7% et 2% dans les contextes CG, CHG et CHH respectivement (Cokus et al., 2008). Ces valeurs peuvent aussi varier en fonction du type de tissus, des conditions environnementales et, plus généralement, entre différentes accessions, révélant l'existence d'une variabilité épigénétique au sein de l'espèce (Hsieh et al., 2009; Vaughn et al., 2007)

La méthylation chez les plantes, contrairement à ce qui a été observé chez les mammifères, n'est pas répartie sur l'intégralité du génome. Elle est plus particulièrement retrouvée au niveau des régions péri-centromériques, des séquences répétées et des éléments transposables (Cokus et al., 2008; Zhang et al., 2006). En effet, chez *Arabidopsis thaliana*, la méthylation de l'ADN est très répandue dans les régions intergéniques où elle a notamment pour rôle de limiter la transcription et la prolifération des éléments transposables (Kato et al., 2003; Lippman et al., 2004). Les cytosines au niveau de ces éléments mobiles et des séquences répétées sont lourdement méthylées dans les trois contextes (CG, CHG et CHH), alors que dans les régions codantes, la méthylation des cytosines se limite aux sites CG (Cokus et al., 2008).

Lors de la réplication de l'ADN, la méthylation des cytosines doit être maintenue et cette maintenance se fait selon des mécanismes distincts pour les trois contextes de méthylation, qui font intervenir différents acteurs.

a. Le maintien de la méthylation de type CG

Le gène *DNA METHYLTRANSFERASE1 (MET1)* est connu depuis longtemps pour son implication dans la méthylation de l'ADN dans un contexte CG. *MET1* est essentiel car il permet le maintien de la méthylation au moment de la gamétogenèse. L'étude de mutants a montré que la perte de la fonction de *MET1* entraîne une réduction de la méthylation dans le contexte CG (Saze et al., 2003). Plus récemment, il a été montré que trois gènes codant des membres de la famille VARIANT IN METHYLATION (VIM), *VIM1*, *VIM2* et *VIM3*, participent également au maintien de la méthylation de type CG (Woo et al., 2008). Ces gènes, comme *MET1*, codent des protéines qui correspondent à des homologues décrits chez

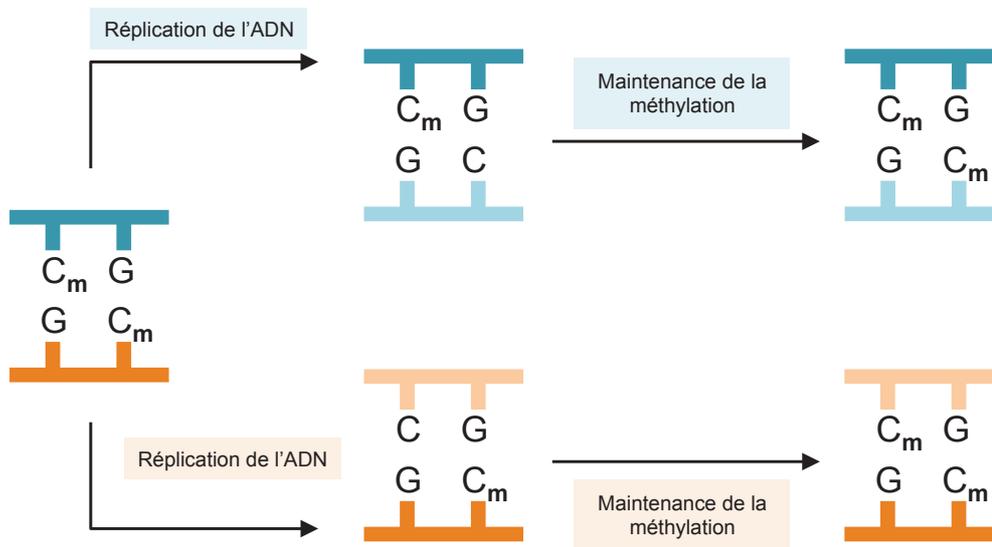


Figure I-9 : Maintenance de la méthylation CG.

Après la répliation des cytosines méthylées en contexte symétrique, les deux brins fils sont hémiméthylés. Ils servent de matrice aux ADN méthyltransférases, comme par exemple MET1, qui vont maintenir la méthylation de type CG en catalysant la méthylation du brin néosynthétisé.

les mammifères, ce qui suggère que les mécanismes de maintenance de la méthylation de type CG pourraient être communs. La maintenance des profils de méthylation CG se fait lors de la réplication de l'ADN. Après son cycle de réplication, l'ADN est hémiméthylé. Il est formé du brin néosynthétisé non méthylé et du brin parental qui porte la méthylation et va être pris en charge par l'ADN méthylase MET1 et par les protéines VIM. MET1 va ainsi catalyser la méthylation du brin néosynthétisé de façon symétrique au profil de méthylation du brin parental (**Figure I-9**).

Environ un tiers des gènes sont méthylés dans leur partie codante. Pour un certain nombre de gènes, cette méthylation est présente dans le corps du gène, mais pas dans les régions promotrices en 5' et les régions 3'. Ce profil de méthylation est spécifique des gènes et n'est pas retrouvé dans les gènes non exprimés et pseudogènes (Zhang et al., 2006). Ceci suggère une interaction entre la méthylation en 5' et en 3' d'une part, et l'initiation ainsi que la terminaison de la transcription d'autre part. De plus, il a été montré que les gènes méthylés dans leur région transcrite ont tendance à être davantage surexprimés dans un mutant *met1* en comparaison aux gènes qui ne sont pas méthylés (Zilberman et al., 2007). Ces résultats viennent appuyer l'hypothèse qui considère que la méthylation dans le corps des gènes a une influence sur la transcription. De la même façon, les pseudogènes et transposons sont réactivés dans un mutant *met1*, suggérant ainsi que la méthylation de type CG est suffisante à l'extinction de l'activité transcriptionnelle au niveau de certains éléments transposables (Lippman et al., 2004; Zilberman et al., 2007).

b. Le maintien de la méthylation CHG

Chez *Arabidopsis thaliana*, 91% des transposons et 58% des pseudogènes sont méthylés (Zilberman et al., 2007). Leur méthylation est dense et se fait dans les trois contextes de cytosines pour empêcher leur transcription et donc leur remobilisation. Ainsi, la maintenance correcte de la méthylation dans ces régions est essentielle.

La méthylation en contexte CHG est maintenue par une boucle de renforcement qui implique d'une part CHROMOMETHYLASE3 (CMT3), qui est une méthyltransférase spécifique des plantes, et la diméthylation de l'histone H3 au niveau de la lysine 9 (appelée H3K9me2 ; voir paragraphe 4) c.) d'autre part (**Figure I-10**). Une étude réalisée à l'échelle du génome a

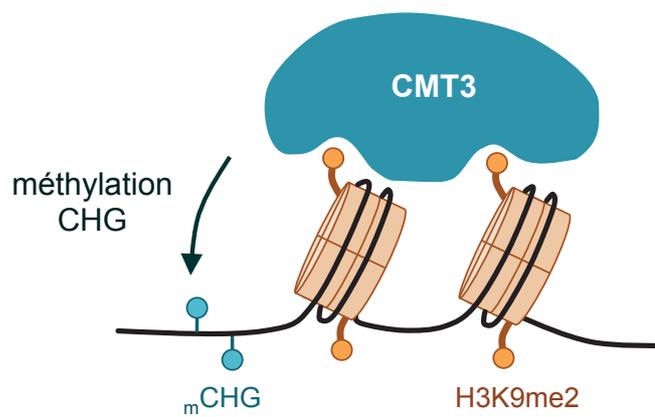


Figure I-10 : Maintenance de la méthylation CHG.

Le brin d'ADN correspond au trait en noir qui s'enroule autour des octamères d'histones, représentés par les cylindres marrons. Le maintien de la méthylation en contexte CHG (boules bleues) se fait par une boucle de renforcement qui fait intervenir la marque H3K9me2 (boules oranges). CMT3 reconnaît ces marques et méthyle les sites CHG à proximité du nucléosome.

permis de mettre en évidence un lien entre ces deux marques. En effet, il y a une corrélation importante entre la méthylation de l'ADN dans un contexte CHG et la modification d'histone H3K9me2 (Bernatavichute et al., 2008). CMT3 est capable de reconnaître cette modification d'histone, et de catalyser la méthylation des sites CHG à proximité (Du et al., 2012). En effet, il a été montré que la perte de CMT3 entraîne une réduction importante de la méthylation CHG à l'échelle du génome (Lindroth et al., 2001).

Pourtant, dans le mutant *cmt3*, certains sites méthylés en contexte CHG ne sont pas affectés. Ces sites sont le plus souvent dépendants de l'activité de DOMAINS REARRANGED METHYLTRANSFERASE1 et 2 (DRM1 et DRM2) puisqu'ils sont presque entièrement déméthylés dans un triple mutant *drm1 drm2 cmt3* (Cao and Jacobsen, 2002; Stroud et al., 2014). En effet, il a été montré que DRM1 et DRM2, même s'ils sont majoritairement impliqués dans la voie de méthylation CHH, sont aussi responsables de la méthylation de certains sites CHG (Cao and Jacobsen, 2002).

Récemment, une étude s'intéressant à la méthylation de l'ADN à l'échelle du génome a permis de montrer que la perte de CMT2, une autre chromométhylase, affecte également la méthylation de type CHG. En effet, même si CMT2 catalyse préférentiellement la méthylation dans un autre contexte, elle est pourtant impliquée dans la méthylation CHG, de manière redondante avec CMT3 (Stroud et al., 2014; Zemach et al., 2013).

c. Le maintien de la méthylation CHH et la voie du RdDM

Le maintien de la méthylation en contexte CHH doit se faire par l'intermédiaire de mécanismes différents de ceux qui maintiennent la méthylation en contexte symétrique. En effet, l'asymétrie du contexte entraîne une perte de l'information sur l'un des deux brins lors de la réplication de l'ADN. Le maintien de la méthylation CHH se fait par la voie de « méthylation de l'ADN dirigée par les ARN », appelée aussi RdDM, qui fait intervenir des petits ARN interférents (siARN). Au-delà du maintien de la méthylation CHH, cette voie de méthylation guidée par les petits ARN est essentielle à la mise en place de la méthylation dite *de novo*, au niveau de sites non méthylés, quel que soit leur contexte. Chez *Arabidopsis thaliana*, la méthylation par la voie RdDM représente environ 30% de la méthylation totale

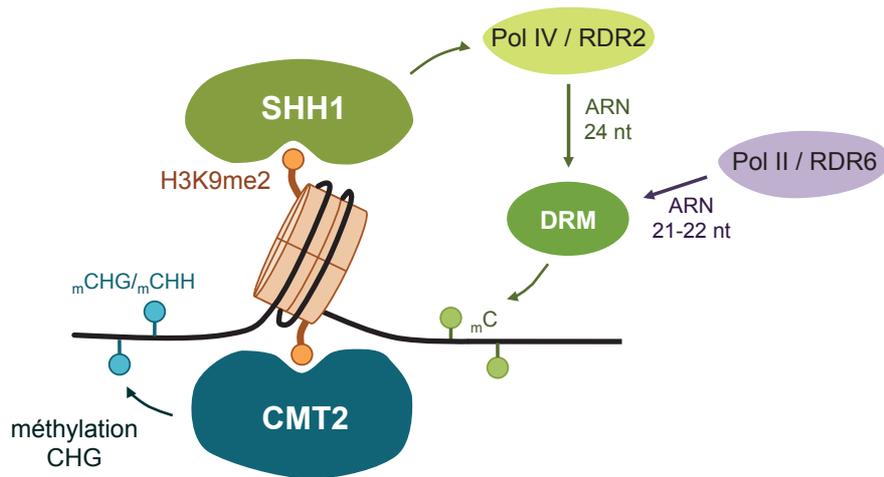


Figure I-11 : La voie de méthylation *de novo* de l'ADN.

Représentation de différentes voies qui permettent la méthylation *de novo* chez Arabidopsis. Ces mécanismes se mettent en place suite à la reconnaissance de la marque H3K9me2 (boules oranges). Les voies qui font intervenir les petits ARN et les DRM interviennent quel que soit le contexte de méthylation (boules vertes), alors que la voie CMT2 ne catalyse la méthylation qu'en contexte non-CG (boules bleues). D'après Kawashima and Berger, 2014.

(Lister et al., 2008). Plusieurs voies sont aujourd'hui connues pour être impliquées dans le contrôle de la méthylation *de novo*.

La voie de méthylation *de novo* la plus décrite est catalysée par les enzymes DRM1 et DRM2. Même si le profil d'expression de DRM1 se limite aux gamètes femelles, il a été montré que les deux DRM ont la capacité de méthyler les cytosines dans les trois contextes (Stroud et al., 2013; Zhong et al., 2014). La voie du RdDM fait intervenir l'ARN polymérase IV, spécifique du règne végétal, qui transcrit les régions hétérochromatiques, comme les éléments transposables. Celle-ci est recrutée au niveau de ces régions cibles par l'intermédiaire de la protéine SAWADEE HOMEODOMAIN HOMOLOGUE1 (SHH1) qui reconnaît et lie la chromatine riche en marques d'histone H3K9me2 (Law et al., 2013). La polymérase IV fonctionne de concert avec RNA-DEPENDANT RNA POLYMERASE2 (RDR2), avec laquelle elle interagit physiquement, et qui copie le transcrit synthétisé par la polymérase IV en double brin d'ARN (Haag et al., 2012). Cette structure double brin est reconnue et clivée en duplexes de petits ARN interférents de 24 nucléotides par la ribonucléase DICER-LIKE3 (DCL3). Ces siARN sont par la suite pris en charge par la protéine ARGONAUTE4 (AGO4) (Zilberman et al., 2004). AGO4 dirige le siARN sur l'ARN non codant synthétisé par la polymérase V et qui correspond à la région cible. Cette interaction permet le recrutement de DRM2 qui va catalyser la méthylation *de novo* du brin d'ADN (**Figure I-11**).

Plus récemment, une méthylation CHH résiduelle a pu être observée à l'échelle du génome dans le triple mutant *drm1 drm2 cmt3*, ce qui suggérait l'implication d'une autre méthyltransférase (Cokus et al., 2008). En effet, l'enzyme CMT2, dont l'activité a désormais été démontrée *in vitro*, intervient dans la méthylation *de novo* au niveau des sites CHH par reconnaissance de la marque histone H3K9me2. (**Figure I-11**) (Stroud et al., 2014).

La méthylation de l'ADN est également influencé par des facteurs de remodelage de la chromatine, comme c'est le cas pour *DECREASE IN DNA METHYLATION1 (DDMI)*. *DDMI* a été identifié lors d'un crible génétique visant à mettre en évidence des mutants hypométhylés à l'aide d'une enzyme de digestion sensible aux cytosines méthylées et le nombre de cytosines méthylées était réduit de 70% dans le mutant correspondant (Vongs et al., 1993). *DDMI* code une protéine de la famille SWI2/SNF2, qui regroupe des protéines impliquées dans le remodelage de la chromatine et dont l'action permet la restructuration du nucléosome afin de permettre à d'autres protéines d'accéder à l'ADN (Jeddeloh et al., 1999;

Ryan and Owen-Hughes, 2011). La mutation *ddm1* montre une perte globale et importante de la méthylation, notamment au niveau des éléments transposables et séquences répétées et cette perte s'accompagne de la réactivation de certains transposons (Jeddeloh et al., 1999; Lippman et al., 2004). Ainsi, DDM1 est essentiel au maintien des profils de méthylation. De plus, il a été montré qu'une partie de la méthylation CHH dans les régions hétérochromatiques est liée à l'action de DDM1, indépendamment de la voie du RdDM (Zemach et al., 2013).

Une variante de la voie de méthylation *de novo* a été mise en évidence lors de l'étude du mutant *ddm1*. Celle-ci fait intervenir la polymérase II et RDR6 à la place de la polymérase IV et de RDR2, et se fait par l'intermédiaire de siARN de 21 à 22 nucléotides (**Figure I-11**). Cette voie de RdDM, dépendante de RDR6, serait davantage impliquée dans l'initiation de la méthylation au niveau des éléments transposables que dans le maintien de celle-ci. Dans la mesure où la polymérase IV est spécifique des plantes, il est possible que cette voie ait été héritée d'un mécanisme plus ancien, partagé avec d'autres organismes (Nuthikattu et al., 2013).

d. La déméthylation de l'ADN

La méthylation de l'ADN est une marque épigénétique considérée comme relativement stable puisqu'elle doit être transmise de cellule à cellule. Pour la plupart des gènes, les profils de méthylation sont à la fois conditionnés par la méthylation, mais aussi par la déméthylation des cytosines. Lors du développement, d'éventuelles reprogrammations épigénétiques qui nécessitent l'activation d'un ou plusieurs gènes supposent l'action de mécanismes qui permettent la déméthylation des cytosines. Au-delà de son implication dans l'activation des gènes, la déméthylation de l'ADN aurait également pour rôle de compenser l'effet du RdDM et ainsi empêcher la méthylation des transposons et séquences répétées de venir envahir d'éventuels gènes à proximité (Penterman et al., 2007). L'existence de la déméthylation de l'ADN a longtemps été sujette à controverse, mais de nombreuses études montrent aujourd'hui qu'elle peut se dérouler selon deux processus différents, passif ou actif.

La déméthylation passive correspond à une perte de la méthylation des cytosines par défaut de maintenance au moment de la réplication de l'ADN. Alors, le brin d'ADN néosynthétisé est

conservé dans son état non méthylé et la méthylation peut être perdue progressivement au cours des cycles cellulaires.

Chez les plantes, la déméthylation active est assurée par des enzymes qui possèdent des activités glycosylase et lyase qui s'accompagnent d'une réparation de la base excisée. Lors du processus de déméthylation, l'ADN glycosylase va reconnaître spécifiquement la cytosine méthylée et l'exciser. Par la suite, l'activité lyase va cliver le brin d'ADN qui sera réparé avec incorporation d'une cytosine non méthylée (Agius et al., 2006). REPRESSOR OF SILENCING1 (ROS1), DEMETER (DME), DEMETER-LIKE2 et 3 (DML2 et DML3) sont quatre ADN glycosylases qui ont été identifiées chez *Arabidopsis thaliana* et qui ont la capacité de retirer les cytosines méthylées (Morales-Ruiz et al., 2006; Ortega-Galisteo et al., 2008). Il a été montré qu'à l'inverse de *DME*, les trois DML (*ROS1*, *DML2* et *DML3*) sont plutôt exprimées dans les tissus végétatifs (Gong et al., 2002; Ortega-Galisteo et al., 2008; Penterman et al., 2007). L'étude des mutants correspondants montre que ROS1, DML2 et DML3 peuvent avoir une activité redondante.

Les protéines DML ont initialement été identifiées comme des enzymes essentielles au phénomène d'empreinte parentale (Choi et al., 2002). En effet, DME, en particulier, est responsable de la déméthylation massive de l'endosperme lors du développement de la graine (Gehring et al., 2009; Hsieh et al., 2009). Pourtant, des études plus récentes suggèrent que les DML jouent un rôle beaucoup plus étendu. L'analyse des profils de méthylation des mutants *dml2*, *dml3*, ainsi que du double mutant montre d'importants changements des profils de méthylation. En effet, des locus fortement méthylés dans un fond génétique sauvage sont hypométhylés chez ces mutants et à l'inverse, des positions qui n'étaient pas méthylées le deviennent (Ortega-Galisteo et al., 2008). De plus, il a été montré que ROS1 est impliquée dans la limitation de la méthylation de l'ADN au niveau des promoteurs de certains gènes (Zhu et al., 2007). Néanmoins, les simples, doubles ou triples mutants *ros1*, *dml2* et *dml3* ne montrent pas de défauts développementaux importants, ce qui suggère que la déméthylation active de l'ADN chez *Arabidopsis* pourrait ne pas être un aspect critique du développement (Gong et al., 2002; Penterman et al., 2007).

e. L’empreinte parentale

Alors que *ROSI* est globalement exprimé tout au long du développement, il a été montré que *DME* a un profil d’expression limité à la gamétogenèse (Choi et al., 2002; Gong et al., 2002). *DME* est connu en particulier pour réguler l’activation transcriptionnelle du gène *MEDEA* (*MEA*) en excisant les cytosines méthylées de l’allèle maternel spécifiquement, et ainsi permettre le bon développement de la graine (Gehring et al., 2006). Après fécondation, il n’y a dans l’endosperme que l’allèle maternel de *MEA* qui est exprimé, alors que l’allèle paternel est maintenu dans un état silencieux. Ce processus qui aboutit à l’expression différentielle des allèles en fonction de leur origine parentale est appelé l’empreinte parentale.

Chez les plantes, l’empreinte parentale a lieu dans l’endosperme et doit être établie à nouveau à chaque génération. Les gènes soumis à empreinte parentale ne sont généralement pas exprimés dans les stades végétatifs de la plante. De nombreuses études se sont intéressées à l’origine parentale des allèles exprimés dans l’endosperme (Gehring et al., 2011; Luo et al., 2011; McKeown et al., 2011). L’une d’entre elles a permis d’identifier chez *Arabidopsis thaliana* 208 gènes soumis à empreinte parentale. Parmi eux, une majorité se sont révélés être à empreinte maternelle, c’est-à-dire que, comme c’est le cas pour le gène *MEA*, l’allèle paternel n’est pas exprimé (Gehring et al., 2011).

f. Les paramutations

Le phénomène de paramutation correspond à une interaction en *trans* entre différents allèles, ou locus présentant des homologies de séquence, et qui aboutit à la modification de l’expression d’un de ces allèles : Un des allèles, qui est actif d’un point de vue transcriptionnel, peut être éteint suite à l’interaction avec un allèle inactif ou faiblement transcrit qui va imposer son état transcriptionnel. L’allèle inactif est dit « paramutagène » et va inactiver l’allèle qui est actif appelé « paramutable ». Ce dernier devient alors « paramuté ».

Ce phénomène a été initialement observé chez le maïs au locus *R* et décrit l’interaction de deux allèles *R* et *R-st* (*stippled*), ce dernier étant responsable de la formation de graines tachetées de noir (Brink, 1958; Chandler et al., 2000). Lorsqu’une plante portant l’allèle *R*

paramutable est croisée avec une plante portant l'allèle *R-st* paramutagène, l'expression de l'allèle *R* est fortement réduite dans les F1. L'allèle *R*, paramuté suite à son interaction avec l'allèle paramutagène, est alors appelé *R'*. En F2, les plantes portant les allèles *R'/R'* développent alors des graines tachetées, comme si elles portaient l'allèle *R-st* (Chandler et al., 2000).

Un autre exemple connu chez le maïs est celui du locus *b1* (*booster1*) qui régulent des gènes de la voie de biosynthèse des flavonoïdes (Coe, 1959). L'allèle paramutable *B-I*, qui entraîne une coloration anthocyanée de la tige, est paramuté en F1 par l'allèle *B'*, qui n'entraîne pas de coloration, lors d'un croisement entre une plante portant *B-I* et une plante portant l'allèle paramutagène *B'*. Par la suite, toutes les générations présentent des tiges non colorées, et l'allèle *B-I* paramuté a alors la capacité d'être paramutagène à son tour (Chandler et al., 2000).

Pour expliquer l'interaction en *trans* de locus comme *b1* ou *R*, un modèle impliquant la voie du RdDM a été proposé : des petits ARN de 24 nucléotides seraient produits à partir de l'allèle paramutagène et entraîneraient la méthylation *de novo* de l'allèle paramutable (Arteaga-Vazquez et al., 2010; Erhard and Hollick, 2011).

La méthylation des allèles en *trans* a également été observée chez *Arabidopsis thaliana* chez des hybrides issus d'un croisement entre les accessions Landsberg *erecta* et C24 (Greaves et al., 2012). Ce résultat suggère que l'interaction et la méthylation entre locus homologues en *trans* existe plus largement à l'échelle du génome. Chez *Arabidopsis* également, la présence d'ARN de 24 nucléotides au niveau de certains locus méthylés en *trans* est un argument en faveur de l'implication d'un mécanisme comme le RdDM (Groszmann et al., 2011).

2) Variants d'histones

Chez les plantes, comme chez les mammifères, les histones dites canoniques H1, H2A, H2B, H3 et H4 sont chacune codées par plusieurs gènes. Il existe pourtant dans le génome d'autres gènes qui codent des histones dont la séquence varie par rapport à celle des histones canoniques. Ils vont être incorporés au nucléosome de façon indépendante à la réplication de l'ADN pour permettre une modulation de la structure et de la fonction de la chromatine, alors

que les gènes des histones canoniques s'expriment au moment de la phase S du cycle cellulaire (Li et al., 2007).

Le nombre de variants existants dépend de l'histone concernée. En effet, plusieurs variants ont été mis en évidence pour les histones H2A et H3 alors qu'aucun variant de l'histone H4 n'est connu à ce jour. Les variants d'histone peuvent différer des histones canoniques au niveau de leurs queues N-terminales, de leurs domaines « *histone-fold* », ou ne varier que des quelques acides aminés, comme l'histone de remplacement H3.3.

Au-delà des variations de séquence protéique par rapport aux histones canoniques, ces variants d'histones peuvent présenter des modifications post-traductionnelles différentes. Ainsi, il a été montré que l'histone H3.3 peut contenir plus de marques que l'histone H3 canonique, en particulier des modifications associées à la chromatine active (McKittrick et al., 2004).

L'histone H2A.Z est un des variants de H2A les plus étudiés. Ce variant est associé à des régions qui sont transcriptionnellement actives et est impliqué dans le maintien de l'état euchromatinien afin d'éviter que ces régions ne se condensent. En effet, des études *in vitro* ont montré que les modifications des domaines de H2A.Z par rapport à l'histone H2A permettraient de réduire les interactions entre les nucléosomes (Fan et al., 2002). Il a été démontré plus récemment que le variant H2A.Z contribuerait à l'activation de la transcription en empêchant la méthylation de s'établir au niveau des promoteurs des gènes (Zilberman et al., 2008).

Certains variants possèdent des caractéristiques qui vont modifier les propriétés du nucléosomes, comme le variant H2A.Z, et d'autres vont être spécifiques d'une région du génome, comme par exemple le variant CenH3 qui confère son identité au centromère et qui est impliqué dans la ségrégation des chromosomes (Palmer et al., 1987). Ces variants d'histone permettent d'apporter un degré supplémentaire à la dynamique de la chromatine.

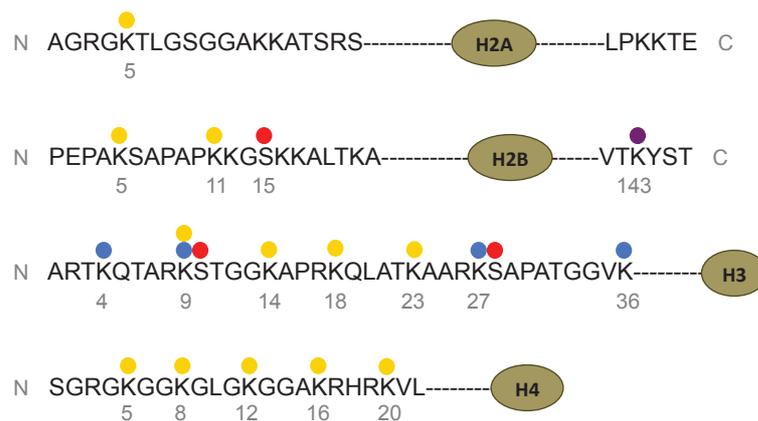


Figure I-12 : Les modifications post-traductionnelles d’histones identifiées chez les plantes.

Les histones H2A, H2B, H3 et H4 sont représentées schématiquement de leur partie N à C terminales. Les modifications post-traductionnelles sont indiquées : phosphorylation (en rouge), acétylation (en jaune), mono-ubiquitination (en violet) et méthylation (en bleu). D’après les données obtenues par spectrométrie de masse (Zhang et al., 2007).

3) Modifications post-traductionnelles des histones

Dès 1997, l'étude de la structure du nucléosome par cristallographie a permis de mettre en relation des modifications post-traductionnelles au niveau de la queue N-terminale des histones et la structure de la chromatine (Luger et al., 1997). En effet, les extrémités N-terminales des histones sont la cible de nombreuses modifications post-traductionnelles comme l'acétylation, la phosphorylation, l'ubiquitination, la méthylation, la sumoylation et l'ADP-ribosylation (**Figure I-12**). Au total, plus de 60 résidus d'histones sont connus pour leur capacité à pouvoir être modifiés et ce nombre est sans doute largement sous-estimé.

Potentiellement, chacune de ces modifications peut avoir des conséquences d'un point de vue fonctionnel. Notamment, elles peuvent modifier le contexte chromatinien et avoir un impact sur divers processus comme la transcription, la réparation et la réplication de l'ADN ainsi que la condensation de la chromatine. Ce contexte chromatinien va dépendre des différentes modifications post-traductionnelles présentes sur les histones. La combinaison de ces modifications d'histones à un instant donné est aussi appelée le « code histone ».

a. Acétylation, phosphorylation et ubiquitination des histones

Acétylation des histones

L'acétylation des histones se fait au niveau de certaines lysines et a pour conséquence la neutralisation de la charge positive de l'acide aminé. En effet, ces résidus lysines chargés interagissent par liaison électrostatique avec les charges négatives portées par l'ADN. L'acétylation des histones pourrait ainsi avoir une influence sur la stabilité de l'interaction entre l'histone et l'ADN et générer une structure chromatinienne relaxée (Shahbazian and Grunstein, 2007). En effet, cette modification post-traductionnelle est généralement associée à une activation de la transcription des gènes.

L'acétylation des histones est catalysée par une famille d'enzymes appelées histone acétyltransférases (HAT) qui transfèrent un groupement acétyle sur la chaîne latérale de la lysine de l'histone à partir de l'acétylcoenzyme A. Cette acétylation est réversible et peut être

retirée par des histones désacétylases (HDAC). La balance entre les activités HAT et HDAC permet donc de maintenir et de réguler les niveaux d'acétylation des histones.

Chez *Arabidopsis thaliana*, l'acétylation peut avoir lieu au niveau des lysines 5 de l'histone H2A, des lysines 6, 11, 27 et 32 des histones H2B, des lysines 9, 14, 18 et 23 de l'histone H3 ainsi que les lysines 5, 8, 12, 16 et 20 des histones H4 (Zhang et al., 2007a).

Phosphorylation des histones

La phosphorylation des histones peut se faire sur les chaînes latérales des résidus sérine ou thréonine. Elle consiste en l'ajout d'un groupement phosphate provenant de l'adénosine triphosphate (ATP) catalysé par une kinase. La phosphorylation des histones est un processus très dynamique. Les résidus phosphorylés par les kinases peuvent être déphosphorylés par des phosphatases (Oki et al., 2007).

Il s'agit d'un processus très régulé et la phosphorylation des résidus des différentes histones se fait en réponse à diverses voies de signalisation. Chez la levure par exemple, l'histone H2A est phosphorylée par une kinase particulière suite à des dommages sur l'ADN (Foster and Downs, 2005). De même, les histones sont le plus souvent phosphorylées à des sites très spécifiques au cours de la division cellulaire (Barber et al., 2004).

Ubiquitination des histones

L'ubiquitination des histones se fait par liaison covalente d'une ubiquitine, qui est une petite protéine de 76 acides aminés très conservée, au niveau de résidus lysine. Cette liaison nécessite l'action de trois enzymes E1 à E3. En premier lieu, l'ubiquitine est activée par E1, puis conjuguée à l'enzyme E2. Cette conjugaison recrute la ligase E3 qui va catalyser la liaison covalente de l'histone avec la glycine C-terminale de l'ubiquitine.

La poly-ubiquitination, c'est à dire l'ajout d'une succession de résidus ubiquitine, est connu pour être un signal d'adressage pour la dégradation des protéines. En revanche la mono-ubiquitination affecte plutôt l'activité biochimique de la protéine, voire sa localisation subcellulaire. Combinée avec d'autres modifications post-traductionnelles d'histone, la mono-

ubiquitination permet de réguler la structure du nucléosome, ainsi que l'accessibilité de la chromatine.

La mono-ubiquitination de l'histone H2B (H2Bub1) se fait sur un résidu lysine conservé qui se trouve en position 143 de l'histone H2B d'*Arabidopsis thaliana* (Zhang et al., 2007a). Il a été montré que la marque H2Bub1 est une modification d'histone associée à des gènes transcriptionnellement actifs et qui est retrouvée en combinaison avec la triméthylation des lysines 4 et/ou 36 de l'histone H3 (Roudier et al., 2011).

b. Méthylation des histones

La méthylation des histones correspond à l'ajout d'un à trois groupements méthyle sur les queues N-terminales des histones en utilisant la S-adénosylméthionine comme substrat. Contrairement à l'acétylation ou la phosphorylation, cette modification ne s'accompagne pas d'une modification de la charge globale de l'histone. La méthylation des histones se fait au niveau des lysines essentiellement, mais également certains résidus arginine. La méthylation des histones est catalysée par différentes histones méthyltransférases (HMT) qui sont spécifiques de certains résidus et dont l'activité est influencée par d'éventuels groupements méthyles déjà présents (Bannister and Kouzarides, 2011).

La méthylation des histones a longtemps été considérée comme une marque très stable, voire irréversible, notamment à cause de la nature de sa liaison carbone-azote. Ainsi, jusqu'à la découverte chez l'Homme de la protéine LSD1 (*Lysine-Specific Demethylase1*), première enzyme à activité histone déméthylase (HDMT) décrite, le seul mécanisme permettant la déméthylation des histones connu était indirect, par échange d'histones. LSD1-LIKE1 (LDL1), LDL2, LDL3 et FLOWERING LOCUS D (FLD) sont les quatre homologues de LSD1 décrits à ce jour chez *Arabidopsis thaliana*. Comme LSD1, cette famille d'homologues n'est pas capable de catalyser la déméthylation des résidus lysine triméthylés (Shi et al., 2004). Différentes protéines appartenant à une seconde famille appelée JumonjiC (JmjC) ont été identifiées plus récemment et possèdent une activité HDMT qui leur permet de déméthyliser également les résidus triméthylés (Kooistra and Helin, 2012). Différentes combinaisons de profils de méthylation des lysines et arginines sont possibles et celles-ci confèrent une diversité de fonction pour ces résidus, pouvant aller par exemple d'un rôle activateur de

transcription pour certaines, à la formation d'hétérochromatine pour d'autres (Santos-Rosa et al., 2002; Zhang and Reinberg, 2001).

Les marques d'histone les plus majoritairement décrites chez *Arabidopsis thaliana*, sont la méthylation de H3K4, H3K36, H3K27 et H3K9.

La méthylation de H3K4

Les gènes de la famille *Trithorax* (*Trx*) codent des facteurs qui régulent la structure de la chromatine par leur capacité à méthyler les H3K4. Ils ont d'abord été décrits chez les animaux, notamment pour leur implication dans la régulation des gènes homéotiques Hox. C'est une famille qui est également retrouvée chez les plantes. Le gène *ARABIDOPSIS TRITHORAX1* (*ATX1*) code une protéine homologue de l'histone méthyltransférase TRX qui est responsable spécifiquement de la méthylation de H3K4 (Alvarez-Venegas et al., 2003). L'étude des niveaux de di- et triméthylation de H3K4 chez le mutant *atx1* montre que d'autres acteurs interviennent dans la mise en place de ces marques (Alvarez-Venegas and Avramova, 2005). En effet, des homologues de *ATX1* ont pu être identifiés chez *Arabidopsis* dont un paralogue de *ATX1*, appelé *ATX2* qui est issu d'un événement de duplication. Alors que chez d'autres organismes la même enzyme est responsable à la fois de la diméthylation et de la triméthylation de H3K4, chez *Arabidopsis*, *ATX1* et *ATX2* semblent catalyser préférentiellement la triméthylation et la diméthylation respectivement (Saleh et al., 2008). Des analyses bioinformatiques et phylogénétiques ont révélé l'existence de nombreuses autres H3K4 méthyltransférases putatives chez *Arabidopsis thaliana*. Parmi elles, *ARABIDOPSIS TRITHORAX RELATED7* (*ATXR7*), *SET DOMAIN PROTEIN25* (*SDG25*), *SDG14* et *SDG16* (Baumbusch et al., 2001; Springer et al., 2003). Plus récemment, des homologues des composants du complexe COMPASS (*COMplex Proteins ASSociated with Set1*) responsable de la triméthylation H3K4 chez mammifères ont également été identifiés. Ainsi, il a été montré que les protéines Ash2, RbBP5 et WDR5 qui forment ce complexe peuvent triméthyler, mais pas mono-, ni diméthyler les H3K4 (Jiang et al., 2011).

Les régions géniques présentent un enrichissement en mono-, di- et triméthylation H3K4, contrairement aux régions péricentromériques, aux éléments transposables et aux séquences répétées. En comparant la distribution de ces trois marques H3K4 avec des données d'expression génique, il a été montré que H3K4me3 n'était présente qu'au niveau des gènes

transcriptionnellement actifs alors que H3K4me1 et H3K4me2 sont retrouvées autant dans les gènes actifs qu'inactifs. De plus, la marque H3K4me3 est particulièrement corrélée avec la présence de transcrits, ce qui confirme son rôle important dans l'activation de la transcription (Zhang et al., 2009).

La méthylation de H3K36

Il a été montré que chez la drosophile et les mammifères, H3K36 est méthylé spécifiquement par ABSENT, SMALL, OR HOMEOTIC DISCS1 (ASH1) qui est une histone méthyltransférase (Berger, 2007). Des protéines présentant des similitudes avec ASH1 de la drosophile ont été mises en évidence chez *Arabidopsis*. Quatre de ces protéines possèdent un domaine SET dans leur région centrale, comme ASH1, et ont donc été nommées ASH1 HOMOLOG1 ou SET DOMAIN PROTEIN26 (ASHH1/SDG26), ASHH2/SDG8, ASHH3/SDG7 et ASHH4/SDG24. De plus, un cinquième homologue a pu être identifié, ASH1 RELATED3 (ASHR3/SDG4), qui présente en revanche un domaine SET à son extrémité C-terminale (Baumbusch et al., 2001; Springer et al., 2003).

Les protéines SDG26, SDG8, et SDG4 ont montré une activité histone méthyltransférase *in vitro* ou *in vivo* (Cartagena et al., 2008; Xu et al., 2008). Chez *Arabidopsis*, la perte de SDG8 entraîne des défauts dans de nombreux processus biologiques, comme la floraison ou la défense contre des pathogènes (Berr et al., 2012; Xu et al., 2008). En effet, l'étude de la méthylation globale de H3K36 dans un fond mutant *sdg8* montre que SDG8 est responsable de la majorité de la méthylation H3K36me2 et H3K36me3 (Dong et al., 2008). De même, les niveaux de H3K36me2 et H3K36me3 sont réduits dans les noyaux végétatifs en l'absence de SDG4 (Cartagena et al., 2008). Globalement, les protéines SDG sont impliquées dans la régulation de gènes spécifiques et de différents processus biologiques par l'intermédiaire de la méthylation des histones.

La méthylation de H3K27

Les résidus H3K27 peuvent être mono-, di- ou triméthylés. L'étude de la distribution chez *Arabidopsis thaliana* de la marque H3K27me2 à l'échelle du génome montre qu'elle est à la fois présente dans l'hétérochromatine et dans l'euchromatine, dispersée au niveau de gènes codants et de transposons (Park et al., 2012; Roudier et al., 2011). La marque H3K27me1

présente un enrichissement au niveau de l'hétérochromatine constitutive, c'est-à-dire aux régions centromériques, péri-centromériques et télomériques (Mathieu et al., 2005). A l'inverse, la triméthylation de H3K27, qui est une marque très abondante et très étudiée, est retrouvée plutôt dans l'euchromatine au niveau de gènes impliqués dans des processus développementaux, faiblement exprimés, voire réprimés (Lindroth et al., 2004; Mathieu et al., 2005; Zhang et al., 2007b). Beaucoup de gènes développementaux sont réprimés par l'intermédiaire de la marque H3K27me3, dont *FLC* (Jiang et al., 2008). Plus généralement, la méthylation des résidus H3K27 est considérée comme une modification qui réprime l'expression des gènes.

La marque H3K27me1 est déposée au niveau des régions hétérochromatiques par ATXR5 et ATXR6, qui sont deux des histones méthyltransférases à domaine SET identifiées chez *Arabidopsis thaliana* (Baumbusch et al., 2001; Jacob et al., 2009). En effet, il a été montré que les activités de ATXR5 et ATXR6 sont nécessaires pour la maintenance de l'hétérochromatine (Jacob et al., 2010). De plus, la protéine SU(VAR)3-9 HOMOLOG2 (SUVH2) qui possède également un domaine SET, participe à la monométhylation de H3K27 *in vivo* (Naumann et al., 2005).

La marque H3K27me3 est très étudiée, chez les animaux comme chez les plantes et les mécanismes sous-jacents ont été décrits chez certains organismes. La méthylation H3K27me3 est mise en place chez les animaux par un complexe protéique appelé Polycomb Repressor Complex 2 (PRC2), très conservé entre les espèces. Le complexe PRC2, identifié en premier lieu chez la drosophile, est composé de 4 sous-unités : Extra Sex Comb (ESC), Enhancer of Zeste E(Z), Suppressor of Zeste 12 (SU(Z)12) et la protéine p55 (Tie et al., 2001). La triméthylation de H3K27 est catalysée par E(Z) qui possède un domaine SET associé à une activité histone méthyltransférase (Czermin et al., 2002). Chez la drosophile, cette marque va être reconnue par un second complexe, PRC1 et cette reconnaissance des H3K27me3 se fait par l'intermédiaire du chromodomaine de la protéine Polycomb (PC), sous-unité du complexe PRC1 (Cao et al., 2002). Quatre autres sous-unités protéiques composent ce complexe : Posterior Sex Combs (PSC), Polyhomeotic (PH), RING, et dans une moindre mesure Sex Comb on Midleg (SCM) (Francis et al., 2001; Saurin et al., 2001). Il a été montré que PRC1 est impliqué dans la compaction de la chromatine, entraînant ainsi une réduction de l'activité transcriptionnelle (King et al., 2002). Chez les mammifères, l'ensemble des sous-unités des complexes PRC1 et PRC2 est conservé (Levine et al., 2002).

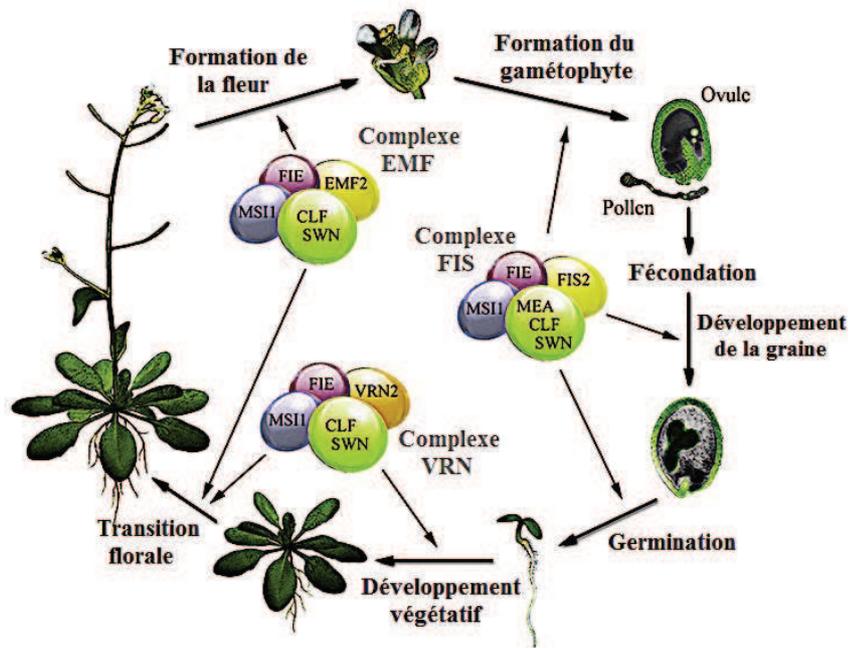


Figure I-13 : Régulation du développement chez *Arabidopsis* par les complexes PRC2.

Les différents complexes PRC2 agissent à des étapes précises du développement de la plante, et font intervenir des acteurs qui sont communs (comme FIE, MSI, CLF/SWN) ou spécifiques d'un complexe (EMF2, FIS2, VRN2). D'après Butenko and Ohad, 2011.

Chez *Arabidopsis thaliana*, 12 orthologues des différentes sous-unités de PRC2 ont été identifiées. Ainsi, CURLY LEAF (CLF), SWINGER (SWN) et MEDEA (MEA) sont les trois orthologues de E(z) ; VERNALIZATION2 (VRN2), FERTILIZATION INDEPENDANT SEEDS2 (FIS2) et EMBRYONIC FLOWER2 (EMF2) sont les orthologues de Su(z)12 ; FERTILIZATION INDEPENDANT ENDOSPERM (FIE) est l'orthologue de ESC et p55 possède 5 orthologues chez *Arabidopsis* appelés MULTICOPY SUPPRESSOR OF IRA1 à 5 (MSI1 à 5). Des études à la fois génétiques et moléculaires, montrent que ces homologues interviennent dans trois complexes PRC2 impliqués à différents stades du développement (Butenko and Ohad, 2011; Hennig and Derkacheva, 2009) (**Figure I-13**).

En revanche, seuls des homologues de PSC et RING, respectivement AtBM1a/AtBM1b et AtRING1a/AtRING1b, ont pu être identifiés chez *Arabidopsis thaliana*, ce qui suggère un mécanisme différent de celui observé chez les animaux (Zheng and Chen, 2011). La protéine LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) est un analogue fonctionnel de la sous-unité PC du complexe PRC1. En effet, elle possède un chromodomaine capable de lier les marques H3K27me3 *in vitro* et sa localisation au niveau du génome montre qu'elle colocalise avec les régions riches en H3K27me3 (Turck et al., 2007; Zhang et al., 2007c). Il a récemment été montré que LHP1 est capable d'interagir avec le complexe PRC2 *in vivo* par l'intermédiaire de MSI1 (Derkacheva et al., 2013).

La famille des PcG est plutôt bien conservée entre les animaux et les plantes. Même si les protéines de la famille Trx mises en évidence chez la drosophile ne sont pas toutes retrouvées chez les plantes, ces deux familles de protéines participent à une balance entre le maintien d'un état actif des gènes par les Trx et d'un état répressif par les PcG (Alvarez-Venegas, 2010; Avramova, 2009).

4) Cas de la méthylation H3K9

La méthylation de H3K9, résidu qui peut être mono-, di- ou triméthylé, est connue pour son implication, chez les animaux comme chez les plantes, dans la répression de la transcription.

En effet, ces marques répressives sont le plus souvent associées à l'hétérochromatine pour maintenir les éléments transposables et répétés dans un état silencieux.

a. Répartition des marques H3K9

Malgré une grande conservation de la méthylation des lysines entre les eucaryotes, leur distribution et leur rôle biologique peuvent varier d'un organisme à un autre (Zhang et al., 2007a). En effet, chez les animaux, les résidus H3K9 peuvent être mono-, di- ou triméthylés, alors que les marques trouvées majoritairement chez *Arabidopsis thaliana* sont les marques H3K9me1 et H3K9me2, même si une petite proportion de H3K9me3 peut cependant être détectée (Jackson et al., 2004). Les marques H3K9me1 et H3K9me2 sont enrichies dans les chromocentres chez *Arabidopsis*, alors qu'elles ne sont pas retrouvées dans les chromocentres chez la souris, par exemple (Fransz et al., 2006). De la même façon, les marques H3K9me3 sont localisées chez la souris dans les régions hétérochromatiques, alors qu'elles sont associées chez *Arabidopsis* à l'euchromatine, et dans certains cas à des gènes transcriptionnellement actifs (Charron et al., 2009).

b. Etablissement de la méthylation H3K9

Les protéines de la famille des Su(var)3-9, capables de méthyler les H3K9, ont été les premières histones méthyltransférases décrites chez la drosophile (Rea et al., 2000). Des homologues de ces protéines ont été mis en évidence chez d'autres organismes, dont *Arabidopsis thaliana* qui possède 10 homologues appelés SU(VAR)3-9 HOMOLOG (SUVH). Parmi eux, SUVH4, appelé également KRYPTONITE (KYP), qui n'a pas de chromodomaine comme son homologue Su(var)3-9 chez la drosophile, mais qui possède un domaine SRA (*SET and RING Associated*) capable de se lier à la chromatine. KYP a été identifié lors d'un crible suppresseur du défaut de nombre des organes floraux observé dans l'allèle *clark kent (clk)* du gène *SUPERMAN (SUP)* (Jackson et al., 2002). En effet, *clk* est un des allèles de *SUP* qui est rendu silencieux par méthylation ectopique du locus (Jacobsen and Meyerowitz, 1997). La protéine KYP a la capacité de catalyser la mono- et la diméthylation des résidus H3K9, mais n'a pas la capacité de les trimétyler. L'étude du mutant *kyp* montre

que sa fonction est essentielle pour l'établissement des marques H3K9me2 au niveau de l'hétérochromatine (Jackson et al., 2004).

Pourtant, la méthylation H3K9 résiduelle observée dans un fond mutant *kyp* suggère que d'autres acteurs sont impliqués dans la méthylation H3K9 chez *Arabidopsis thaliana*. Les dix membres de la famille SUVH partagent une grande homologie de séquence et des domaines conservés (Baumbusch et al., 2001). Il a été montré que SUVH6 possède une activité méthyltransférase, avec différentes affinités *in vitro* pour les marques H3K9me1 et H3K9me2, mais pas pour H3K9me3 (Jackson et al., 2004). De la même façon, l'activité H3K9 méthyltransférase de SUVH5 a été démontrée *in vitro* alors que les sept autres membres de la famille SUVH ne montrent pas d'activité H3K9 méthyltransférase (Ebbs and Bender, 2006). La forte déméthylation observée chez le triple mutant suggère que KYP, SUVH5 et SUVH6 agissent ensemble et de façon redondante au niveau de certains locus (Ebbs and Bender, 2006). KYP semble pourtant assurer le maintien d'une grande partie de la méthylation H3K9 puisque la marque H3K9me2 est fortement réduite en son absence (Jackson et al., 2004). KYP est le seul membre de la famille SUVH à contenir des introns dans sa séquence codante. Ainsi, il pourrait être un membre ancestral de la famille alors que les autres membres seraient des paralogues issus de sa duplication.

c. Interaction entre H3K9me2 et méthylation de l'ADN

Chez les animaux, la formation de l'hétérochromatine est en partie dépendante de l'état de méthylation des H3K9, mais également de la protéine HETEROCHROMATIN PROTEIN1 (HP1) qui est capable d'interagir spécifiquement avec les H3K9 méthylés (Bannister et al., 2001). En effet, il a été montré que la marque H3K9me2 permet le recrutement de HP1, qui est une protéine conservée, associée à l'hétérochromatine (Bannister et al., 2001; Lachner et al., 2001). Cette interaction se fait par l'intermédiaire du chromodomaine de HP1, très conservé également, et connu pour son implication dans la formation d'hétérochromatine (Platero et al., 1995).

Il a été montré que LHP1, qui est l'homologue chez *Arabidopsis* de HP1 reconnaît et co-localise avec les marques H3K27me3 (voir paragraphe 3) b.), mais pas H3K9me2 (Turck et al., 2007). En effet, les plantes possèdent un autre mécanisme pour la maintenance de la

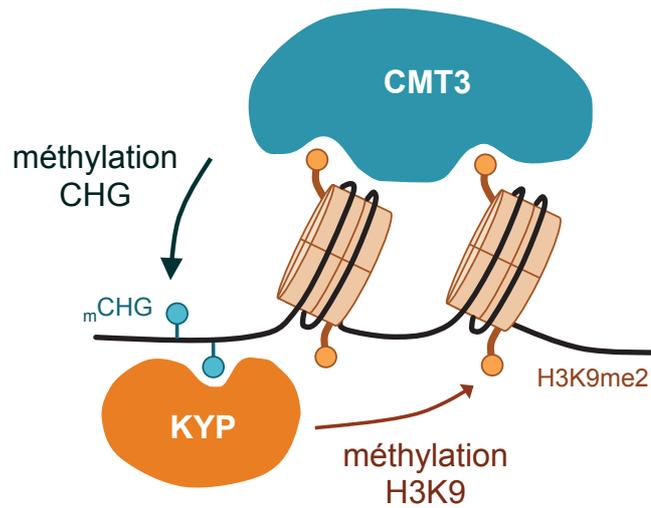


Figure I-14 : Boucle de rétrocontrôle positif CMT3-KYP.

Cette boucle de rétrocontrôle impliquant méthylation de l'ADN et méthylation H3K9me2 permet d'expliquer le maintien de la méthylation CHG. CMT3 maintient la méthylation en contexte CHG (boules bleues), qui est reconnue par le domaine SRA de KYP et KYP catalyse la diméthylation des résidus H3K9 (boules oranges), marque nécessaire à la maintenance de la méthylation CHG et à la reconnaissance par le chromodomaine de CMT3.

méthylation des résidus H3K9. Chez le mutant *kyp*, la diminution de la marque H3K9me2 ainsi que de la méthylation CHG au niveau des régions hétérochromatiques suggère que les deux marques sont liées et que KYP est également impliqué dans le contrôle de la méthylation des sites CHG (Lindroth et al., 2004; Malagnac et al., 2002). De plus, il a été montré que KYP est capable de se lier à l'ADN méthylé en contexte CHG par l'intermédiaire de son domaine SRA (Johnson et al., 2007). Aussi, CMT3, qui méthyle les cytosines dans un contexte CHG, est capable de reconnaître et d'interagir avec deux marques H3K9me2, par l'intermédiaire de son domaine BAH et son chromodomaine qui sont tous deux essentiels à son activité *in vivo*, tandis que KYP peut lier les cytosines méthylées en contexte CHG et méthyle les H3K9 en retour (**Figure I-14**) (Du et al., 2014; Du et al., 2012). Même si CMT3 et KYP n'interagissent pas directement, ils sont tous deux partenaires dans cette boucle d'auto-renforcement qui permet de maintenir efficacement à la fois la méthylation CHG et la marque H3K9me2 au niveau des régions hétérochromatiques.

d. Régulation de la méthylation H3K9 dans les régions géniques

Malgré une perte importante de la méthylation H3K9me2 et CHG dans des mutants *kyp* et *cmt3*, seulement une faible réactivation des transposons est observée (Lippman et al., 2003). Ces résultats suggèrent que cette boucle impliquant KYP et CMT3 serait davantage impliquée dans la répression de la transcription des gènes que dans l'inactivation des éléments transposables. Ceci implique alors qu'il existe des mécanismes capables de reconnaître et de protéger les gènes qui doivent être transcrits vis-à-vis de ces marques répressives.

Identification de *INCREASE IN BONSAI METHYLATION1 (IBMI)*

Afin de déterminer le mécanisme impliqué dans l'exclusion de ces marques répressives au niveau des régions géniques, un crible sur une population mutagénisée a été réalisé. Celui-ci est basé sur un phénotype qui consiste en l'augmentation de la méthylation au niveau du locus *BONSAI (BNS)* (Saze et al., 2008). Ce locus contient un gène codant pour APC13 qui est un régulateur du cycle cellulaire. Le gène *APC13* est bordé par un élément transposable LINE (*Long Interspersed Nuclear Element*) qui est fortement méthylé chez le sauvage. Le locus a été nommé *BONSAI* en référence au phénotype nain observé chez les plantes lorsque la

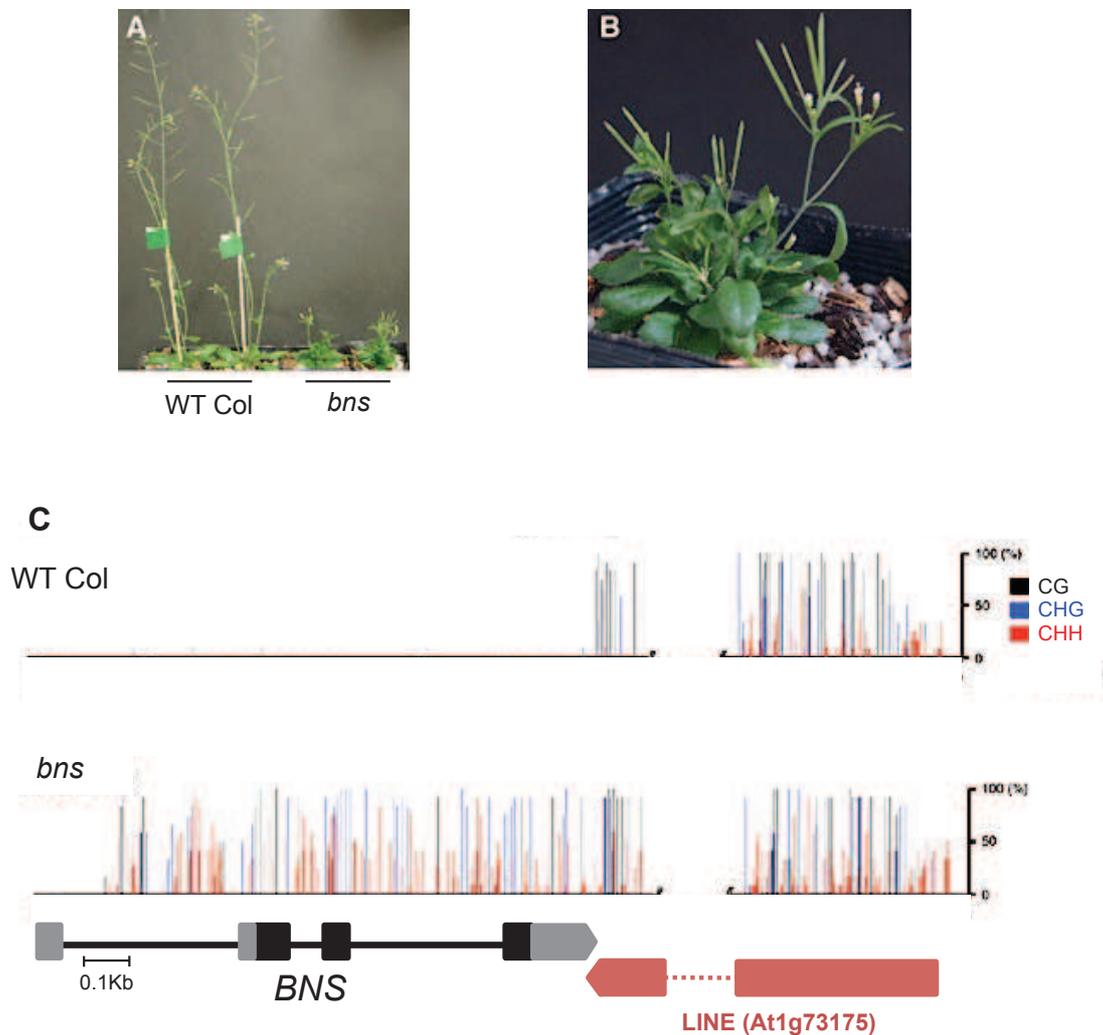


Figure I-15 : Illustrations du phénotype *bonsai*.

A: Comparaison des phénotypes de deux plantes Col à gauche et deux plantes *bns* à droite, âgées de 6 semaines. **B:** Photographie rapprochée de *bns*. **C:** Représentation de la méthylation des cytosines au locus *BNS* pour Col et *bns*. Les barres verticales donnent une indication sur le pourcentage de méthylation de chaque cytosine pour les trois contextes de méthylation (CG en noir, CHG en bleu et CHH en rouge). Dans la partie basse est placée une représentation du gène *BNS* : Les exons sont représentés par des cassettes noires, les introns par des traits et les UTR par des cassettes grises. L'élément LINE placé à proximité est schématisé en rouge. D'après Saze and Kakutani, 2007.

méthylation associée au LINE vient envahir le gène *APC13* (Saze and Kakutani, 2007) (**Figure I-15**).

Ce crible génétique a permis de mettre en évidence l'implication du gène *IBM1* dans la restriction de la méthylation à l'élément SINE. En effet, le mutant *ibm1* identifié, qui correspond à une substitution entraînant un changement d'acide aminé dans la protéine, montre une méthylation CHG ectopique au niveau du locus *BNS* qui est présente dès la première génération où *ibm1* est à l'état homozygote (Saze et al., 2008).

Caractérisation de la fonction de IBM1

IBM1 code une protéine à domaine JumonjiC (JmjC) qui appartient à la famille JHDM2. Cette famille est constituée de protéines à fonction H3K9 méthylase qui sont retrouvées également chez les animaux (Klose et al., 2006). Une expérience s'intéressant à la déméthylation *in vivo* et utilisant des anticorps capables de reconnaître spécifiquement les modifications des lysines de l'histone H3 a permis de démontrer que IBM1 possède bien une activité histone méthylase, comme d'autres membres de la famille JHDM2 identifiés chez les animaux. En effet, il a pu être montré que IBM1 est capable de retirer spécifiquement les H3K9me1 et H3K9me2 (Inagaki et al., 2010).

L'étude du mutant *ibm1* a permis de montrer qu'en plus de causer la méthylation ectopique de l'ADN au niveau du locus *BNS*, la perte de fonction IBM1 entraîne un phénotype développemental sévère, tant au niveau des organes végétatifs que reproducteurs. En effet, le mutant *ibm1-1* présente des défauts de forme de la feuille, de croissance, de formation des grains de pollen ainsi que des siliques (**Figure I-16**). Des observations, au niveau des feuilles notamment, ont permis de montrer que le phénotype nain associé à *BNS* est perdu lorsque *ibm1* est associé aux mutations *kyp* ou *cmt3*. Autrement dit, les doubles mutants *cmt3 ibm1* et *kyp ibm1* retrouvent un phénotype sauvage. Ceci montre que les défauts de développement observés chez le mutant *ibm1* se font par l'intermédiaire de la méthylation non-CG et H3K9. La mutation *kyp* est également capable de supprimer le phénotype d'hyperméthylation du locus *BNS* du mutant *ibm1*, ce qui montre bien que IBM1 a une activité capable de contrer celles de KYP et de CMT3, au moins au locus *BNS* (Saze et al., 2008).



Figure I-16 : Phénotype du mutant *ibm1-1*.

Des différences de phénotype en terme de croissance ainsi que de forme foliaire peuvent être observées entre Col à gauche et *ibm1-1* à droite, ici 20 jours après semis en serre.

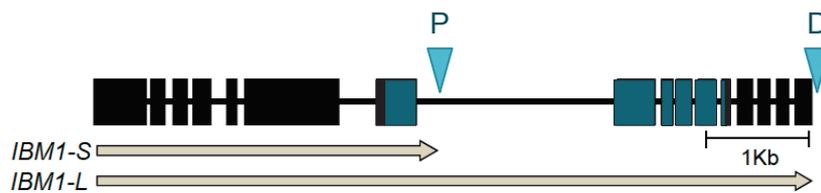


Figure I-17 : Représentation des deux formes du transcrit *IBM1*.

Les cassettes noires correspondent aux exons, et les cassettes bleues schématisent le domaine JmjC présent uniquement quand *IBM1* est transcrit sous sa forme longue (*IBM1-L*). Les sites de polyadénylation proximal (P) et distal (D) sont indiqués par les triangles bleus.

Afin de déterminer l'étendue des gènes ciblés par IBM1, des analyses de la méthylation à l'échelle du génome ont été réalisées. Ainsi, une hyperméthylation globale de l'ADN a pu être mise en évidence chez le mutant *ibm1*, en particulier au niveau de milliers de gènes. Cette méthylation ectopique concerne majoritairement les cytosines en contexte CHG et est localisée dans les gènes (Miura et al., 2009). Cette augmentation de la méthylation induite par *ibm1* s'accompagne d'une accumulation de marques H3K9me2 dans ces mêmes régions codantes (Inagaki et al., 2010). Par conséquent, l'activité de IBM1 protège les gènes vis-à-vis de ces marques répressives, habituellement responsables du maintien des transposons et séquences répétées dans un état silencieux.

Les mécanismes qui permettent à IBM1 de reconnaître et cibler certains gènes sont encore mal connus. Pour mieux comprendre ces mécanismes, des mutants portant des insertions T-DNA dans des gènes cible de IBM1, dans le but de bloquer leur transcription, ont été étudiés. Ces analyses montrent que la méthylation CHG ectopique apparaît dans ces mutants de manière similaire à ce qui est observé en l'absence de IBM1. Ce résultat suggère que la transcription pourrait être un prérequis pour la reconnaissance par IBM1 de ses cibles. De plus, il a été montré que les transposons, lorsqu'ils sont réactivés dans les mutants *kyp* et *cmt3*, sont alors ciblés par IBM1. Ces résultats suggèrent que la fonction de IBM1 pourrait être liée à l'activité transcriptionnelle des gènes cibles (Inagaki et al., 2010).

Régulation du locus *IBM1*

Les prédictions indiquent que le gène *IBM1* est capable de produire deux transcrits différents. Un transcrit court, appelé *IBM1-S*, et un transcrit long, *IBM1-L*, qui comporte le domaine catalytique JmjC complet (**Figure I-17**). Des analyses de complémentation du mutant *ibm1*, avec l'un et l'autre des transcrits, a permis de montrer que *IBM1-L* est le transcrit fonctionnel, et qu'il permet à la fois de restaurer le phénotype morphologique du mutant, ainsi que le profil de méthylation de l'ADN au locus *BNS* (Rigal et al., 2012).

La transcription de l'ARN fonctionnel *IBM-L* est dépendante de l'épissage du septième intron de *IBM1* qui est long comparé à la majorité des introns de *Arabidopsis* (Rigal et al., 2012). En effet, celui-ci mesure un peu plus de 2 kb alors que, chez *Arabidopsis*, la taille moyenne d'un intron est inférieure à 200 pb et que seulement moins de 1% des introns ont une taille supérieure à 1 kb (Wang and Brendel, 2006). Cet intron contient une région

hétérochromatique de plusieurs centaines de paires de bases présentant une méthylation dense de l'ADN, ainsi que des marques d'histone H3K9me2. Dans les fonds mutants *met1* ou *kyp* et *cmt3*, qui présentent des défauts de méthylation en contexte CG et CHG respectivement, la transcription de *IBM1-L* était impactée (Rigal et al., 2012). Ainsi, les méthylations de type CG et CHG sont essentielles à la transcription correcte et fonctionnelle de *IBM1*. En revanche, dans les fonds mutants *nrpd1* et *nrpe2*, qui sont respectivement les sous-unités des ARN polymérase IV et V, la transcription de *IBM1-L* est similaire à ce qui peut être observé chez le contrôle, démontrant ainsi que la transcription de *IBM1* n'est pas liée à la voie du RdDM (Rigal et al., 2012).

Même si l'hyperméthylation CHG et H3K9 dans les gènes, induite par la perte de fonction de *IBM1*, n'est pas systématiquement corrélée avec une répression de l'expression des gènes, les modifications importantes de l'épigénome et le phénotype développemental sévère du mutant *ibm1* témoignent de l'importance de ce mécanisme qui permet de protéger les gènes de ces marques répressives.

5) Les réseaux de marques épigénétiques

Outre ce lien dynamique qui existe entre la méthylation de l'ADN catalysée par CMT3 et la marque histone H3K9me2 déposée par KYP, d'autres marques épigénétiques s'avèrent être également étroitement liées les unes aux autres. En effet, le nombre de modifications des résidus d'histone et les combinaisons possibles de celles-ci permettent de contrôler finement la structure de la chromatine et l'interaction entre elles des différentes marques apporte un degré supplémentaire de complexité. Plusieurs études chez les eucaryotes montrent que les marques épigénétiques interagissent entre elles en un réseau complexe qui permet de réguler l'état de la condensation de la chromatine (Filion et al., 2010; Roudier et al., 2011)

Il a été montré que certaines modifications d'histone sont antagonistes, en particulier celles qui modifient les lysines puisqu'il peut y avoir compétition entre méthylation, acétylation et ubiquitination pour un même résidu. D'autres s'excluent pour des raisons d'encombrement stérique, comme par exemple chez les mammifères où la phosphorylation de la sérine 10 de l'histone H3 empêche l'accès de HP1, protéine impliquée dans la régulation et l'intégrité de l'hétérochromatine, à la marque H3K9me2 (Canzio et al., 2014; Fischle et al., 2005).

Par ailleurs, certaines marques peuvent être dépendantes l'une de l'autre et fonctionner par coopération. Ainsi, deux études menées chez *S. cerevisiae* et chez l'humain montrent que la méthylation des résidus H3K9 et H3K79 sont toutes deux dépendantes de la modification H2BK123ub (Kim et al., 2009; Lee et al., 2007).

Chez *Arabidopsis thaliana*, l'analyse intégrative des différentes marques épigénétiques à l'échelle du génome ont permis de mettre en évidence quatre états transcriptionnels de la chromatine qui sont associés chacun à des combinaisons particulières de marques (Roudier et al., 2011). Ces états chromatiniens appelés CS pour « chromatin state » ont été déterminés suite à l'étude de la localisation de 12 marques épigénétiques : les marques d'histone H3K4me_{2/3}, H3K9me_{2/3}, H3K27me_{1/2/3}, H3K36me₃, H3K56ac, H4K20me₁, H2Bub et la méthylation de l'ADN.

Le premier état de la chromatine appelé CS1 correspond à l'euchromatine, qui contient les gènes actifs d'un point de vue transcriptionnel. Ces gènes, qui correspondent à 54% des gènes totaux sont enrichis en marques H3K4me_{2/3}, H3K9me₃, H3K36me₃, H3K56ac et H2Bub. Les états CS2 et CS3 sont deux types distincts de chromatine répressive et différant par leur contenu génomique. Ainsi, CS2 est l'état chromatiniens associé aux gènes réprimés, soit 23% des gènes totaux, et se caractérise par la présence des marques H3K4me₂ et H3K27me_{2/3}. Et les marques H3K9me₂, H3K27me_{1/2}, H4K20me₁ et la méthylation de l'ADN caractérisent l'état CS3, qui correspond à l'hétérochromatine composée d'éléments transposables. Enfin, l'état CS4 regroupe environ 10% des gènes totaux qui sont des gènes faiblement exprimés, ainsi que des régions intergéniques, et ne présente pas d'enrichissement significatif en aucune des marques épigénétiques étudiées (Roudier et al., 2011).

Chez les plantes, la présence de trois contextes de méthylation de l'ADN, l'existence de variants d'histones, ainsi que les modifications post-traductionnelles possibles des histones, à différents degrés et sur différents résidus simultanément, apportent de nombreuses combinaisons de marques possibles. Ceci témoigne du haut degré de complexité des mécanismes qui permettent l'établissement et l'interprétation de ce réseau épigénétique.

(IV) Héritabilité de la méthylation de l'ADN

Parmi ces marques épigénétiques qui modulent la condensation et l'accessibilité de la chromatine pour modifier l'activité transcriptionnelle du génome, une partie peut être transmise de génération en génération. Des changements d'ordre épigénétique peuvent apparaître de manière spontanée ou induite et être transmis à la descendance comme une mutation génétique, générant ainsi des variants épigénétiques. Ces variants reposent sur des « allèles épigénétiques », appelés également épiallèles. Par définition, un épiallèle est un allèle qui ne varie pas génétiquement mais qui présente différents états de méthylation de l'ADN ou de modifications des histones.

1) Épiallèles naturels

Peu d'épiallèles naturels ont été identifiés chez les plantes et ceux qui ont été identifiés sont des variants de méthylation de l'ADN. Puisque l'état de méthylation d'un gène n'est pas systématiquement accompagné d'une modification de son activité transcriptionnelle, il est d'autant plus difficile de les identifier. De plus, la plupart des épiallèles décrits dans la littérature sont les conséquences de variations génétiques, comme des mutations, des remaniements chromosomiques ou des insertions de transposons.

Historiquement, le premier épiallèle a été découvert chez *Linaria vulgaris*. En effet, il a pu être observé chez cette espèce une modification spontanée de l'organisation de la fleur qui passe d'une structure radiale à une structure bilatérale. Cette modification appelée « peloric » est liée au gène *Lcyc* dont l'expression est réprimée. Il a été montré que le phénotype est directement corrélé au niveau de méthylation du gène *Lcyc*. Une réversion du phénomène peut pourtant parfois être observée. Elle s'accompagne alors d'un retour à la normale de l'état de méthylation de *Lcyc* (Cubas et al., 1999). Le variant « peloric » est considéré comme un vrai épiallèle car aucune variation de séquence n'a pu être mise en évidence. Cependant, une régulation en *trans* ne peut pas être exclue pour expliquer la cause de cette variation de méthylation. La cause de l'épiallèle est également inconnue dans le cas du locus *CNR* (*Colorless Non-Ripening*) chez la tomate, pour lequel il n'y a aucune variation de séquence dans un intervalle de 13 kb autour du gène. Il n'est pourtant pas exclu que cette variation

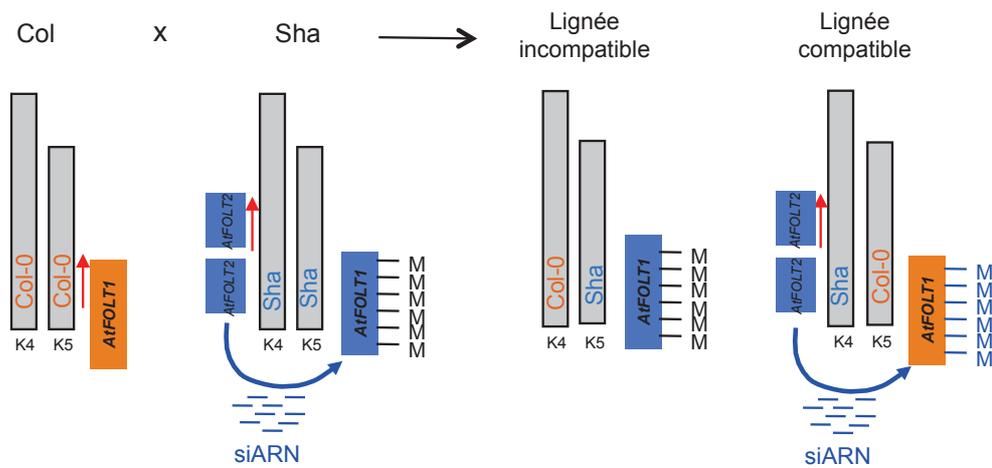


Figure I-18 : Représentation de l'incompatibilité FOLT chez *Arabidopsis*.

Cette figure illustre les locus FOLT aux chromosomes 4 (K4) et 5 (K5), schématisés sous forme de bâtons gris, pour les lignées incompatibles et compatibles provenant du croisement entre les accessions Col (en orange) et Sha (en bleu). Les boîtes oranges et bleues correspondent aux différentes copies de FOLT de Col et Sha respectivement. Près de ces boîtes, les flèches rouges indiquent les copies qui sont transcrites et les M schématisent les copies méthylées. Une des copies FOLT2 de Sha au chromosome 4 génère des siARN, représentés par des traits bleus, qui vont cibler au chromosome 5 la copie FOLT1 Sha (chez le parent Sha) ou Col (chez la lignée compatible) pour méthyler le locus. D'après Durand et al. 2012.

épigénétique soit due à l'influence du rétrotransposon *COPIA*-like dans la région promotrice de *CNR* (Manning et al., 2006).

A l'inverse, l'épiallèle de *WIPI* chez l'espèce *Cucumis melo* est un exemple d'épimutation connue pour être due à un élément causal génétique. L'expression du gène *CmWIPI*, responsable de la détermination sexuelle du melon, peut en effet être réprimée par l'insertion d'un élément transposable à proximité (Martin et al., 2009).

L'exemple du locus *FLOWERING WAGENINGEN (FWA)* chez *Arabidopsis* illustre également bien la cause génétique de la variation épigénétique observée. Le locus *FWA* a été identifié comme responsable du délai de floraison, avant que des analyses de séquence montrent que le phénotype était lié à l'expression ectopique de *FWA* due à une hypométhylation du locus (Koornneef et al., 1991; Soppe et al., 2000). Dans un fond génétique sauvage, *FWA* ne s'exprime pas dans les tissus végétatifs, mais il est exprimé dans l'endosperme et soumis à empreinte parentale (Kinoshita et al., 2004). Le gène *FWA* possède des répétitions en tandem autour de son site d'initiation de la transcription et il a été montré que l'épiallèle *fwa* induit par *ddm1* est stable et lié à la méthylation de ces éléments répétés (Kinoshita et al., 2006).

Pourtant, il est parfois difficile d'identifier l'élément causal à l'origine d'un épiallèle. C'est le cas par exemple pour la variation épigénétique de *FOLT1*, codant un transporteur de folates, entre deux accessions Columbia (Col) et Shahdara (Sha) d'*Arabidopsis*. En effet, dans l'accession Sha, le gène *FOLT1* est réprimé et méthylé mais son paralogue *FOLT2*, localisé à un autre chromosome, est actif. Dans l'accession Col en revanche, *FOLT1* est exprimé alors que *FOLT2* est inexistant. Le croisement entre les deux accessions aboutit à la génération de plantes qui portent à fois l'épiallèle *FOLT1* de Sha, inactif car méthylé, et pas de copie *FOLT2*. Ces plantes, qui n'ont plus de transporteurs de folates, présentent une fertilité très réduite (Durand et al., 2012) (**Figure I-18**).

Le cas de *PAI (PHOSPHORIBOSYLANTHRANILATE ISOMLERASE)* chez *Arabidopsis thaliana* est un exemple très semblable à celui des gènes *FOLT*. Dans certaines accessions comme Col-0, *PAI* est présent en trois copies qui partagent une grande identité de séquence : *PAI1*, *PAI2* et *PAI3*. Dans d'autres accessions comme Wassilewskija (Ws), *PAI1* possède une copie dupliquée et inversée, *PAI4*, l'ensemble *PAI1/PAI4* formant une région répétée inversée

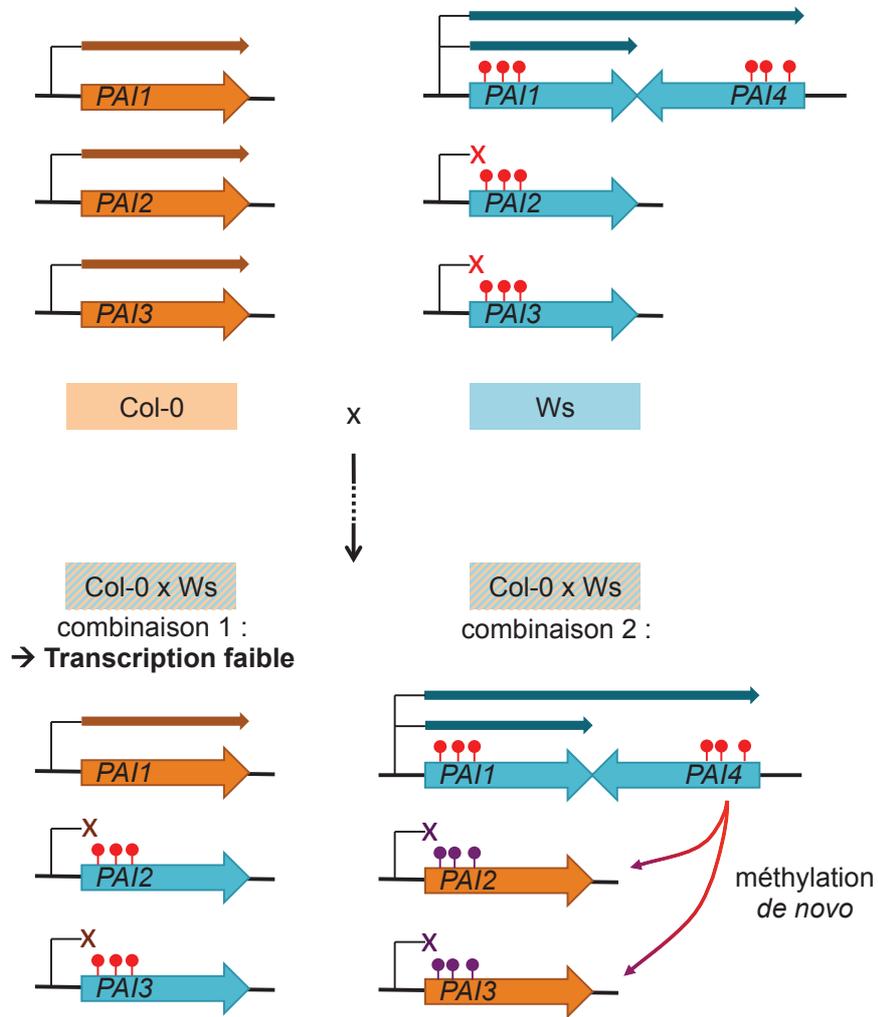


Figure I-19 : Interactions entre les locus *PAI*.

Dans la partie haute sont schématisés les parents Col-0 et Ws, et dans la partie basse sont schématisées deux des combinaisons issues de la F1 Col-0 x Ws. Les 4 copies *PAI1-4* sont représentées par des flèches épaisses en orange pour Col-0 et en bleu pour Ws. Au dessus des gènes est illustré l'état de transcription de la copie : par une flèche lorsque la copie est transcrite, par une croix lorsque la copie est silencieuse. La méthylation présente au niveau de certaines copies sont représentées par des boules rouges ou violettes.

(Melquist et al., 1999). Dans le cas d'un croisement Col x Ws, cette région inversée entraîne l'extinction des autres copies (*PAI2* et *PAI3*), par l'intermédiaire de la voie du RdDM, ce qui diminue l'abondance des transcrits *PAI* (**Figure I-19**) (Melquist and Bender, 2003). La méthylation des différents locus *PAI* est donc un autre exemple d'épiallèle naturel, comme pour les gènes *FOLT*.

Le gène *QQS* (*Qua-Quine Starch*) d'*Arabidopsis* est connu pour être impliqué dans le métabolisme de l'amidon. Des analyses portant sur 36 accessions naturelles différentes, représentatives de la diversité génétique de l'espèce, montrent que ce gène est hypervariable épigénétiquement dans ces populations naturelles (McKhann et al., 2004; Silveira et al., 2013). Il a ainsi été montré que l'expression de *QQS* est négativement corrélée avec la méthylation de l'ADN observée au niveau des séquences répétées insérées dans la région 5' du gène (Silveira et al., 2013). A ce titre, l'exemple de *QQS* illustre bien l'importance de la méthylation de l'ADN comme source de variabilité.

En comparaison à des variants génétiques causés par des mutations de l'ADN, les épiallèles sont beaucoup plus difficiles à identifier sur une base uniquement phénotypique. Ainsi, de nombreuses études se sont intéressées à la variabilité épigénétique, afin de déterminer si elle a, comme la variabilité génétique, des conséquences phénotypiques importantes.

2) Variabilité épigénétique naturelle

La diversité inter- et intraspécifique est majoritairement contrôlée par la variabilité génétique. De cette manière, *Arabidopsis thaliana* s'est adapté à de nombreux environnements grâce à des mutations qui ont pu être identifiées par séquençage de nombreuses accessions (Cao et al., 2011). Pourtant, les quelques exemples de variants phénotypique liés à des causes épigénétiques montrent que celles-ci contribuent aussi à la diversité des espèces.

Les analyses globales d'expression des gènes dans des mutants impactant la méthylation de l'ADN, comme *met1*, montrent que les changements apparaissent plutôt au niveau des éléments transposables que dans les gènes (Lippman et al., 2004; Zhang et al., 2006; Zilberman et al., 2007). Une autre alternative pour étudier la variabilité des profils de méthylation est de s'intéresser à des accessions dont les épigénomes diffèrent naturellement.

Par exemple, la comparaison des profils de méthylation au chromosome 4 entre les accessions Col et Landsberg *erecta* (Ler) a permis de montrer que d'une part, les éléments transposables sont très méthylés dans les deux accessions et que d'autre part, la méthylation peut apparaître dans les gènes, mais également être perdue, générant ainsi une plus grande diversité épigénétique entre populations. Même si les différences observées en terme de méthylation dans les gènes sont héréditaires, elles ne sont pas forcément corrélées à des différences d'expression (Vaughn et al., 2007). Une autre étude s'est intéressée au polymorphisme épigénétique entre accessions. Ainsi, l'analyse des accessions Col et Vancouver (Van) a permis de mettre en évidence une variabilité épigénétique importante entre ces deux accessions qui contribue potentiellement à des variations d'expression de certains gènes (Zhang et al., 2008).

Dans le but de comprendre dans quelle mesure des variations de méthylation de l'ADN peuvent apparaître naturellement, des lignées isogéniques provenant des autofécondations successives d'un même parent ont été générées et étudiées. Après 30 générations, la comparaison des lignées ancestrales et des descendants montre que des régions différenciellement méthylées (DMR) peuvent apparaître ou disparaître d'une génération à l'autre. Il a été montré que ces DMR sont plus nombreuses lorsque les descendants sont comparés entre eux plutôt qu'avec la lignée ancestrale. De plus, cette variabilité est moins importante au niveau des transposons, la méthylation étant transmise de manière plus stable pour maintenir leur répression (Schmitz et al., 2011). Une étude similaire s'est intéressée à l'apparition de variations épigénétiques dans des lignées issues d'un même individu, comparées après 30 générations. Elle a pu confirmer ainsi que l'apparition de méthylation sur des cytosines isolées est fréquente alors que l'apparition de DMR est un phénomène plus rare et plus stable (Becker et al., 2011). Ces deux études ont permis de calculer un taux de changement d'état de méthylation d'une cytosine donnée par génération et ont ainsi pu en conclure que ce taux d'épimutations est plus important que le taux de mutations par génération (Becker et al., 2011; Schmitz et al., 2011). Il est de l'ordre de $4,46 \cdot 10^{-4}$ changements d'état de méthylation par génération pour un site CG donné alors que celui des mutations est estimé autour de $7 \cdot 10^{-9}$ substitutions de base par site et par génération (Ossowski et al., 2010; Schmitz et al., 2011).

L'utilisation de lignées d'*Arabidopsis* qui diffèrent en terme de méthylation de l'ADN permet d'étudier l'hérédité de ces épimutations. Désormais, les techniques de séquençage haut-

débit des génomes et des épigénomes peuvent contribuer à étudier les variations de méthylation dans le but d'identifier les éléments causaux de cette variabilité épigénétique.

3) Epiallèles induits par mutation

L'intérêt d'étudier la variabilité épigénétique est de comprendre dans quelles mesures elle peut être transmise et quelles en sont les conséquences phénotypiques, donc de pouvoir estimer la part de variabilité liée aux modifications épigénétiques, au sein d'une population. Récemment, deux études ont permis d'apporter quelques éléments de réponse à la question de la stabilité et de l'héritabilité de cette variabilité épigénétique. Toutes deux reposent sur la génération de populations de lignées recombinantes épigénétiques dites « épiRIL », induites par des mutations responsables d'une réduction de la méthylation de l'ADN et permettent ainsi d'étudier la variabilité épigénétique en réduisant la variabilité génétique (Johannes et al., 2009; Reinders et al., 2009).

La première s'intéresse à une population qui dérive d'un croisement entre une plante sauvage Col et un mutant *met1* hypométhylé. La F2 est générée par autofécondation de la F1, puis les plantes fixées pour l'allèle sauvage *MET1* sont autofécondées jusqu'en F8. Certaines lignées F8 testées montrent un taux de méthylation entre 5 et 6%, contre 6,8% pour le parent Col et 2,2% pour le parent *met1*. Le fait que les taux de méthylation des lignées F8 soit plus proche du parent Col que du parent *met1* suggère que seulement certaines régions hypométhylées sont transmises ou qu'il y a établissement *de novo* de la méthylation de l'ADN (Reinders et al., 2009). La population générée dans le cadre de la seconde étude est issue d'un croisement entre *ddm1* et Col. La F2 provient d'un backcross de la F1 avec le parent sauvage Col, puis les plantes sont autofécondées jusqu'en F8 également. L'état de méthylation à plusieurs locus choisis pour leur différence de méthylation dans les lignées parentales a été déterminé pour plusieurs lignées F9. Certains allèles hypométhylés induits par *ddm1* sont transmis de manière stable pendant au moins 8 générations alors que d'autres retrouvent une méthylation à un niveau sauvage après quelques générations (Johannes et al., 2009).

Ces résultats similaires ne sont pas les seuls points communs mis en évidence entre les deux populations. En effet, il a été également montré que ces populations d'épiRIL présentent une variabilité phénotypique importante. Certaines de ces lignées ont des phénotypes variés, qui

apparaissent parfois même au cours de la génération des populations, comme par exemple des phénotypes de délai de floraison ou de croissance. Ces deux populations présentent toutefois des différences, notamment en terme de fertilité. En effet, certaines lignées issues de la population d'épiRIL *met1* n'ont pas pu être conduites jusqu'en F8. Cela concerne 30% des lignées alors que dans la population d'épiRIL *ddm1* 1% seulement n'a pas pu être propagé. Ceci peut être dû à une sélection à l'encontre de certains locus qui ont un fort impact lorsqu'ils sont déméthylés (Johannes et al., 2009; Reinders et al., 2009).

La population d'épiRIL induite par *ddm1* est un outil qui a permis par la suite de générer une carte de 126 DMR stables d'une génération à l'autre, couvrant plus de 80% du génome, et qui ségrègent de façon mendélienne (Colome-Tatche et al., 2012). Cette carte a pu être utilisée, comme une carte génétique classique, pour cartographier des caractères complexes comme la date de floraison ou la taille de la racine (Cortijo et al., 2014).

Les résultats obtenus à partir de ces populations d'épiRIL ont permis d'avoir un aperçu de la variabilité épigénétique qu'il est possible d'induire en utilisant des mutations. Certains des épiallèles générés dans ce cadre sont stables sur plusieurs générations, suggérant que les profils parentaux de méthylation ne sont pas spontanément réétablis, et ont pu être utilisés comme des marqueurs génétiques. Pourtant, la stabilité de ces épiallèles n'est pas figée et ils peuvent être également perdus. L'analyse plus approfondie de la stabilité des épiallèles et de leurs conséquences phénotypiques pourrait apporter des outils supplémentaires pour déterminer la part de variabilité épigénétique dans la variabilité des populations naturelles.

(V) Sujet et objectifs de la thèse

L'équipe dans laquelle j'ai réalisé ma thèse s'intéresse particulièrement à la variabilité naturelle épigénétique. Mon projet de thèse fait suite à l'identification, au laboratoire, d'un QTL de croissance foliaire qui a permis de mettre en évidence une mutation ponctuelle apparue spontanément dans le parent Col utilisé pour générer la population de lignées recombinantes étudiée. Le gène portant cette mutation a été nommé *SG1* pour « *SHOOT GROWTH1* » et code une protéine de fonction inconnue au début de ma thèse. Les domaines prédits de cette protéine ainsi que le phénotype du mutant correspondant, à la fois pléiotrope et hétérogène et s'aggravant au cours des générations, suggèrent l'implication de *SG1* dans des mécanismes de régulation épigénétiques.

Les objectifs de ma thèse étaient d'étudier le rôle de la protéine *SG1*, d'effectuer la caractérisation fonctionnelle de la protéine, et de déterminer quelle était son implication dans les mécanismes de régulation épigénétiques. J'ai ensuite initié la recherche de partenaires protéiques de *SG1*, ainsi qu'un crible suppresseur du phénotype du mutant. Enfin, j'ai cherché à identifier des épimutations stables induites dans un fond mutant *sg1* qui ont des conséquences phénotypiques.

La première partie de mon projet fait suite à un travail initié au laboratoire et auquel j'ai contribué notamment par des analyses d'expression de *IBM1* et de certaines de ces cibles, d'immuno-précipitation de la chromatine et de complémentation de mutants. Ce travail a permis de montrer que la transcription de *IBM1* est sous le contrôle de *SG1* et que par cet intermédiaire *SG1* régule indirectement l'épigénome d'*Arabidopsis*.

Dans un deuxième temps, nous avons pu, en collaboration avec l'équipe de Geert de Jaeger au VIB de Gent, déterminer les partenaires protéiques qui forment un complexe avec *SG1*. J'ai par la suite essayé de déterminer le rôle de ces différents interactants, notamment en m'intéressant à leur implication dans la régulation de la transcription de *IBM1* et à leur impact en terme de méthylation au niveau de locus cibles, afin de mieux comprendre le mécanisme dans lequel est impliqué *SG1*. J'ai également réalisé un crible par mutagenèse à l'EMS du mutant *sg1* qui m'a permis d'identifier un suppresseur, tant d'un point de vue moléculaire que phénotypique.

Enfin, mon projet concernant la détection d'épimutations induites par *sgl* m'a permis de mettre en évidence un certain nombre d'épimutations qui sont maintenues de génération en génération et j'ai pu identifier des gènes candidats pour un phénotype choisi.

Ainsi, mes travaux de thèse, qui seront détaillés au cours des trois parties qui vont suivre, ont permis de préciser le rôle de SGI et ont contribué à la compréhension des mécanismes sous-jacents.

Partie II : La protéine SG1 permet le maintien de l'épigénome d'*Arabidopsis thaliana* en régulant *IBM1*

(I) Introduction : identification du locus *SG1*

Le gène *SG1* que j'ai cherché à caractériser au cours de ma thèse a été initialement identifié par l'équipe de Olivier Loudet qui s'intéresse à la variabilité naturelle chez *Arabidopsis thaliana*, en particulier à la variabilité de réponse au stress entre les différentes accessions. Leur thématique de recherche s'appuie sur des populations de lignées recombinantes ou RIL (*Recombinant Inbred Lines*) disponibles au Centre de Ressources Biologiques de Versailles.

L'étude d'une population de RIL issue du croisement Bur-0 x Col-0 dans le cadre d'une analyse QTL de croissance foliaire a permis de mettre en évidence un phénotype particulier lié à un locus appelé *SG1* pour *SHOOT GROWTH1* (**Figure S1A** de (Coustham et al., 2014)). Dans cette population de RIL, une réduction de la croissance apparaît lorsque ce locus *SG1*, cartographié en haut du chromosome 5, est fixé à l'état homozygote pour l'allèle Col-0 (**Figure S1B** de (Coustham et al., 2014)). Plusieurs étapes de cartographie fine ont permis, à l'aide de populations de HIF (*Heterogeneous Inbred Family*), de réduire le locus à un intervalle de 8,5 kb contenant un seul gène, dont la fonction était encore inconnue et qui a été par la suite nommé *SG1*. En parallèle, une analyse QTL de croissance racinaire dans une population de RIL issue du croisement Ct-1 x Col-0 a également permis d'identifier un locus appelé *RG1* pour *ROOT GROWTH1* avec un effet majeur sur la croissance des racines (**Figure S2** de (Coustham et al., 2014)). La réduction de l'intervalle correspondant au locus *RG1* a mis en évidence la même région, puis le même gène, *SG1*.

En l'absence de polymorphisme commun pouvant expliquer le phénotype, dans cet intervalle entre les accessions Bur-0 et Ct-1, et la séquence de Col-0, une analyse de séquence des cDNA a été entreprise et a permis de révéler la présence d'une mutation ponctuelle dans la séquence de l'allèle Col-0. En effet, cette mutation, qui est une substitution d'un C par un T entraîne l'apparition d'un codon stop précoce au niveau de la protéine prédite. La recherche

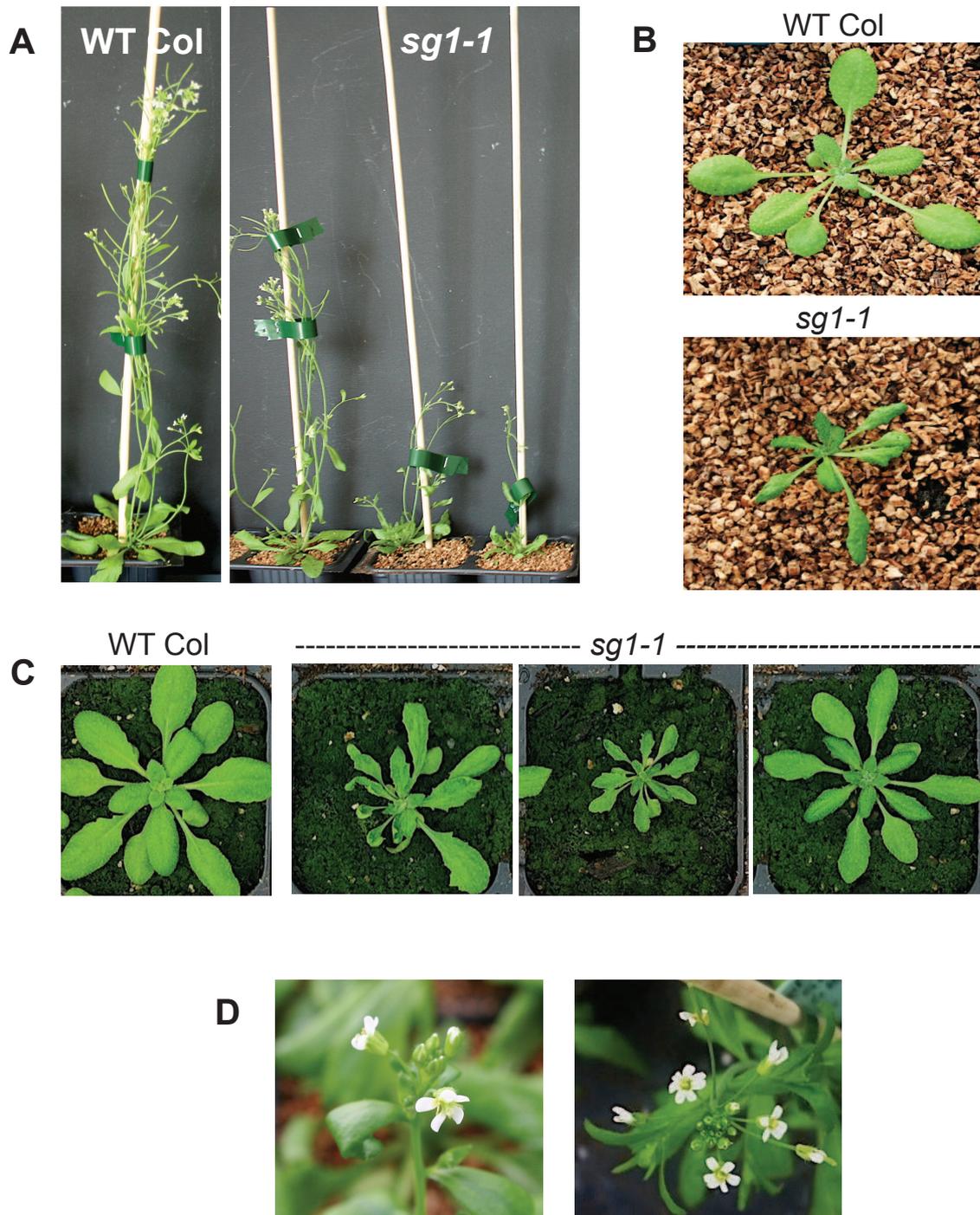


Figure II-1 : Illustration de défauts caractéristiques du mutant *sg1*.

A: Comparaison de la croissance de plantes *sg1* avec Col WT. **B et C:** Illustrations des différences de taille et de forme foliaire entre *sg1* et Col. **D:** Fleurs malformées dans un fond mutant *sg1*.

de l'allèle Col-0 muté sur plusieurs générations a révélé qu'il s'agit d'une mutation apparue spontanément dans le parent Col-0 utilisé pour générer les populations de RIL. C'est à partir de ce parent, de fond génétique Col-0, que cette mutation, nommé *sgl-1*, a été étudiée par la suite.

Les observations du mutant *sgl-1* ont permis de mettre en évidence un phénotype caractéristique. Au delà de la réduction de croissance foliaire qui a été montrée précédemment, d'autres défauts développementaux variés peuvent être observés au niveau des plantes de fond génétique mutant *sgl-1*. En effet, les feuilles du mutant présentent un phénotype particulier. Elles sont petites, plus foncées, avec des bords crénelées et des défauts de phyllotaxie. De plus, les racines sont courtes, plus fines et les organes floraux sont anormaux, à la fois en nombre, et en termes de morphologie et d'identité des tissus (**Figure II-1**).

Le phénotype du mutant *sgl-1* est stochastique et les défauts développementaux sont plus ou moins sévères d'une plante à l'autre au sein d'une même génération. En effet, toutes les plantes issues d'un même parent homozygote pour la mutation ne présentent pas la même sévérité de phénotype (**Figure II-2**). Par ailleurs, ce phénotype s'aggrave d'une génération à l'autre, jusqu'à aboutir à une stérilité complète après plusieurs générations. Les phénotypes observés pour deux autres allèles mutés disponibles, qui sont des mutants d'insertion appelés *sgl-2* (SAIL_310B06) et *sgl-3* (GK_045A07), présentent les mêmes caractéristiques que *sgl-1*, ce qui confirme que le phénotype développemental est lié à la mutation du gène *SGI*. Ce phénotype pléiotrope qui tend à s'aggraver au cours des générations rappelle ce qui est observé chez des mutants affectés dans des voies de régulation de la chromatine, comme les mutants *met1* ou *ddm1* (Kakutani et al., 1996; Mathieu et al., 2007).

D'après les prédictions de l'outil ScanProSite (de Castro et al., 2006) à partir de la séquence protéique disponible sur TAIR10, le gène *SGI* code une protéine contenant un domaine BAH (*Bromo-Adjacent Homology domain*) et un domaine RRM (*RNA Recognition Motif*). Le domaine RRM, qui est retrouvé dans un très grand nombre de protéines, est connu pour être impliqué dans des liaisons à l'ARN ou à d'autres protéines. En revanche, le domaine BAH est partagé avec un nombre restreint de protéines qui jouent le plus souvent un rôle dans des modifications de la chromatine, comme des ADN méthyltransférases, des chromométhylases,



Figure II-2 : La stochasticité au sein d'une génération, autre aspect du phénotype du mutant *sg1*.

des facteurs de transcription ou des protéines liées à la réplication de l'ADN (**Figure S6** de (Coustham et al., 2014)).

Les caractéristiques phénotypiques du mutant *sg1*, très similaire à un mutant *ibm1*, ainsi que les domaines protéiques prédits sont deux arguments en faveur de l'implication de SG1 dans des mécanismes de régulation épigénétique. Ainsi, nous avons cherché à déterminer, à travers des approches de séquençage haut débit du méthylome de *sg1*, d'immuno-précipitation de la chromatine et d'analyses d'expression, la fonction de SG1 et la nature de son implication dans des mécanismes épigénétiques.

(II) Article 1 : « SHOOT GROWTH1 maintains *Arabidopsis* epigenomes by preventing ectopic DNA and histone methylation of gene bodies » PLoS ONE, 2014.

SHOOT GROWTH1 Maintains Arabidopsis Epigenomes by Regulating *IBM1*

Vincent Coustham^{1,2^{‡a}}, Daniela Vlad^{1,2^{‡b}}, Aurélie Deremetz^{1,2}, Isabelle Gy^{1,2}, Francisco A. Cubillos^{1,2^{‡c}}, Envel Kerdaffrec^{1,2^{‡d}}, Olivier Loudet^{1,2*}, Nicolas Bouché^{1,2*}

1 INRA, UMR1318, Institut Jean-Pierre Bourgin, RD10, Versailles, France, **2** AgroParisTech, Institut Jean-Pierre Bourgin, RD10, Versailles, France

Abstract

Maintaining correct DNA and histone methylation patterns is essential for the development of all eukaryotes. In Arabidopsis, we identified SHOOT GROWTH1 (SG1), a novel protein involved in the control of gene methylation. SG1 contains both a Bromo-Adjacent Homology (BAH) domain found in several chromatin regulators and an RNA-Recognition Motif (RRM). The *sg1* mutations are associated with drastic pleiotropic phenotypes. The mutants degenerate after few generations and are similar to mutants of the histone demethylase INCREASE IN BONSAI METHYLATION1 (IBM1). A methylome analysis of *sg1* mutants revealed a large number of gene bodies hypermethylated in the cytosine CHG context, associated with an increase in di-methylation of lysine 9 on histone H3 tail (H3K9me2), an epigenetic mark normally found in silenced transposons. The *sg1* phenotype is suppressed by mutations in genes encoding the DNA methyltransferase CHROMOMETHYLASE3 (CMT3) or the histone methyltransferase KRYPTONITE (KYP), indicating that SG1 functions antagonistically to CMT3 or KYP. We further show that the *IBM1* transcript is not correctly processed in *sg1*, and that the functional *IBM1* transcript complements *sg1*. Altogether, our results suggest a function for SG1 in the maintenance of genome integrity by regulating *IBM1*.

Citation: Coustham V, Vlad D, Deremetz A, Gy I, Cubillos FA, et al. (2014) SHOOT GROWTH1 Maintains Arabidopsis Epigenomes by Regulating *IBM1*. PLoS ONE 9(1): e84687. doi:10.1371/journal.pone.0084687

Editor: Sue Cotterill, St. Georges University of London, United Kingdom

Received: August 29, 2013; **Accepted:** November 26, 2013; **Published:** January 3, 2014

Copyright: © 2014 Coustham et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: VC was supported by the BAP INRA division and a Marie Curie European Reintegration Grant (FP7-PERG08-GA-2010-277068) to NB and OL. The authors acknowledge funding from ANR (Project 11-JSV7-0013) to NB. DV was supported by funding from the European Commission Framework Programme 6, Integrated Project 'AGRON-OMICS/LSHG-CT-2006-037704 to OL. EK was supported by funding from ANR grant 'ARABRAS'/ANR-06-ERAPG-003-01 to OL. FC was supported by funding from the European Commission Framework Programme 7, ERC Starting Grant 'DECODE'/ERC-2009-StG-243359 to OL. 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.

* E-mail: Olivier.Loudet@versailles.inra.fr (OL); Nicolas.Bouche@versailles.inra.fr (NB)

^{‡a} Current address: INRA, UR83 Recherches Avicoles, Nouzilly, France

^{‡b} Current address: Department of Plant Sciences, University of Oxford, Oxford, United Kingdom

^{‡c} Current address: Departamento de Ciencia y Tecnología de los Alimentos, Universidad de Santiago de Chile (USACH), Santiago, Chile

^{‡d} Current address: Gregor Mendel Institute, Austrian Academy of Sciences, Vienna, Austria

Introduction

Methylation of cytosines and histones are epigenetic modifications found in many eukaryote species, including plants and animals that influence gene expression and transposon mobilization. In contrast to mammals for which DNA methylation occurs predominantly in the CG cytosine context, with the exception of embryonic stem cells [1], plant DNA methylation is found in both symmetrical (CG and CHG where H is A, T or C) and non-symmetrical (CHH) contexts. These three types of cytosine methylations are maintained by different specific pathways in Arabidopsis [2]. The DNA methyltransferases METHYLTRANSFERASE1 (MET1) and CHROMOMETHYLASE3 (CMT3) are required for the maintenance of CG and CHG methylation, respectively [2]. Maintaining CHH methylation depends on both CHROMOMETHYLASE2 (CMT2) [3] and the small RNA-directed DNA methylation (RdDM) pathway, which involves plant specific RNA polymerases (Pol IV and V) and the production of 24-nucleotide small RNAs controlled by the RNA-DEPENDENT RNA POLYMERASE2 (RDR2) together with DICER-LIKE3 (DCL3) [2]. Histones can be methylated in several ways, one of them being at lysine 9 on histone H3 tail (H3K9me), a hallmark of

silent chromatin in several organisms, including plants [4]. Methylation at H3K9 can be classified into three different types: mono- di- or tri-methylated, but di-methylation (H3K9me2) is the most common in Arabidopsis. H3K9me2 is catalysed by three histone methyltransferases, KRYPTONITE/SUVH4 (KYP), SUVH5, and SUVH6 [5].

Locus specific CHG and H3K9me2 methylations are intimately connected. Indeed, CMT3 physically interacts with H3K9me2 marks through both its chromo and BAH domains [6] while KYP binds CHG-methylated cytosines through its SRA domain [7]. Altogether CMT3 and KYP participate in a self-reinforcing loop between DNA and histone methylation that is essential for transposons and repeated sequences silencing. Another evidence for an interplay between DNA and histone modifications was the identification of INCREASE IN BONSAI METHYLATION1 (IBM1), a Jumonji C (JmjC) domain protein belonging to the JHD2 family of H3K9 demethylases that is specific to H3K9 mono- and dimethylation [8–10]. *IBM1* was identified in a screen for increased methylation at the *BONSAI* (*BNS*) locus since cytosines of *BNS* in the CHG context are highly methylated in *ibm1*. In both *ibm1/cmt3* and *ibm1/kyp* double mutants, CHG methylation at *BNS* returns to a WT level, showing that *IBM1*

counteracts CMT3 and KYP activities at this locus. In fact, epigenome-wide analyses of *ibm1* identified thousands of other genes that are ectopically enriched for both CHG and H3K9 methylations in *ibm1* mutants [9]. Thus, IBM1 prevents genes from being targeted by the machinery that silences repeated elements and transposons. Additionally, IBM1 acts on components of the RdDM pathway itself through direct targeting of *RDR2* and *DCL3* genes [11].

Here, we report the identification of a novel player, named *SG1*, in the regulation of DNA and histone methylations. Initially identified as a major QTL candidate for shoot and root growth in several segregating populations, the *sg1* mutant exhibited pleiotropic developmental defects with a higher penetrance in successive generations similarly to previously reported mutants involved in chromatin modifications. *SG1* encodes a protein with two putative domains, one of them, a Bromo-Adjacent Homology (BAH) domain being found in several chromatin binding proteins of eukaryotes. A genome-wide methylation analysis in the *sg1* mutant background using bisulfite sequencing showed that a large number of gene body sequences were enriched in CHG methylation, similarly to previous reports on *ibm1*. We also found an increase in H3K9me2 levels within the gene bodies of all CHG hypermethylated candidates tested. We confirmed through a genetic approach that the *sg1* mutant phenotype is suppressed by *cml3* or *kyp* mutations, but not *ibm1*. We analysed the expression of some targets of SG1 and in particular genes involved in the RdDM pathway, but we found no correlation between an increase in gene body CHG methylation and a lack of function. Finally, we show that the accumulation of full-length functional *IBM1* mRNAs is compromised in *sg1*, which can be complemented by the introduction of a transgene producing the long version of *IBM1* transcripts. Altogether, our data reveals a new player in the maintenance of Arabidopsis DNA and histone methylations with an essential role in epigenome stability.

Results

Identification and Characterization of *sg1* Mutants

In a previous report we described a quantitative genetic screen to investigate natural genetic variation for shoot growth in *Arabidopsis thaliana* [12]. Among four significant QTLs mapped in the Bur-0 x Col-0 recombinant inbred line population, one QTL named *SHOOT GROWTH1* (*SG1*) was mapped at the top of chromosome 5 (Figure S1 in File S1). Plants carrying a Col-0 allele at *SG1* locus showed drastic shoot growth defects with respect to plants carrying the Bur-0 allele, as confirmed in heterogeneous inbred families (HIF). Fine-mapping analysis using recombinant HIF mapped *SG1* to a locus localized in a 8.5 kb interval comprising a single gene, At5g11470. In parallel, we fine-mapped a QTL detected in the Ct-1 x Col-0 RIL population with a major effect on root growth (named *ROOT GROWTH1*; *RG1*) down to the same region and gene (Figure S2 in File S1).

A sequence analysis on *SG1* did not evidenced obvious DNA polymorphisms common to stock accessions Ct-1, Bur-0 and Col-0, however, different gene model predictions for the *SG1* gene were found in the databases (Figure S3 in File S1). Hence, we decided to sequence the *SG1* cDNAs from the RILs themselves to exclude any possible mRNA mis-splicing that could account for the phenotype. Surprisingly, we found a single C to T mutation in the *SG1* cDNA from plants of the mapping populations carrying Col-0 alleles, compared to the sequences of parental accessions. Analysis of the cloned *SG1* cDNA, that matches EuGene prediction [13] with slight variations in regard to TAIRv10 prediction (Figure S3 in File S1), suggested that this *sg1-1* mutation

identified in the RILs occurs in the sixth exon of the gene (Figure 1A). By analyzing the original seeds of the Col-0 parent used to produce the RILs, we confirmed the presence of the *sg1-1* mutation compared to the Col-0 reference sequence. We concluded that *sg1-1* is a non-silent spontaneous mutation that appeared in our Col-0 parental line before we developed our RIL sets.

In the original Col-0 accession carrying the *sg1-1* mutation at the homozygous state, we observed that the phenotype was not fully penetrant since we identified both plants with no or subtle phenotype in addition to plants with reduced shoot size and a phenotype that increases throughout generations (Figure 1B and Figure S4 in File S1). The shoot growth reduction phenotype was associated with dark green plants and pleiotropic developmental abnormalities including altered leaf morphology (narrow and serrated), and flowers with more petals, some being fused or with an altered morphology (Figure 1C). After at least four generations at the *sg1*-homozygous state, plants exhibited a terminal phenotype with small deformed leaves associated with strong sterility. We then isolated two T-DNA knockouts (*sg1-2* and *-3*) with insertions localized within the *SG1* coding sequence (Figure 1A). Both mutants were similar to *sg1-1*: the severity of the phenotype varied between individuals and became stronger in successive generations to reach a complete sterility from the 4th–5th generation (Figure S5 in File S1).

SG1 (At5g11470) encodes an unknown protein with two predicted domains: an RNA Recognition Motif (RRM) and a Bromo-Adjacent Homology (BAH) domain (Figure 1A). The *sg1-1* mutation results in a Gln to STOP codon, leading to a potentially truncated protein with no RRM domain. While the Arabidopsis genome encodes a unique SG1 protein, we found a total of 21 proteins that contain a BAH domain (Figure S6 in File S1), including well characterized DNA methyltransferases like MET1 and CMT3, pointing towards a function for SG1 in chromatin regulation.

Gene Body CHG Methylation and H3K9 Marks are Impaired in *sg1*

The pleiotropic and generational nature of the phenotype is reminiscent of those observed in mutants disrupted in chromatin modification pathways such as *ibm1* [10], *met1* [14] or *ddm1* [15]. Thus, we analysed the methylome of *sg1-1* mutant through Illumina deep sequencing after bisulfite conversion, in comparison to the Col-0 background. The coverage of sequenced Arabidopsis methylomes was 30X with an error rate of 0.4% (Table S1 in File S2). First, we calculated the percentage of methylation for cytosines in either one of the three contexts CG, CHG or CHH. Genome-wide, we observed no significant differences between Col-0 and *sg1-1* for cytosines methylated in CG or CHH, while *sg1-1* seemed to contain more methylated CHG than Col-0 (Figure 2A; Table S2 in File S2). This was further confirmed when we calculated the percentages of methylation in 200 bp-windows (with a 50 bp overlap) to scan the whole-genome. The *sg1-1* mutant contained a large number of regions enriched in methylated-CHG compared to Col-0, while no such large-scale contrast was observable for cytosines in the CHH or CG contexts (Figure S7 in File S1). To determine whether CHG regions highly represented in *sg1-1* corresponded to genes or other parts of the genome, we calculated the percentage of cytosine methylated per gene and we found that many genes of *sg1-1* were enriched in CHG methylation compared to Col-0 (Figure 2B). Gene body-methylation in the CG context was similar between Col-0 and *sg1-1* (Figure S8A in File S1), while -overall- genes seem to be slightly over-methylated in the CHH context in *sg1-1* (Figure S8B in File

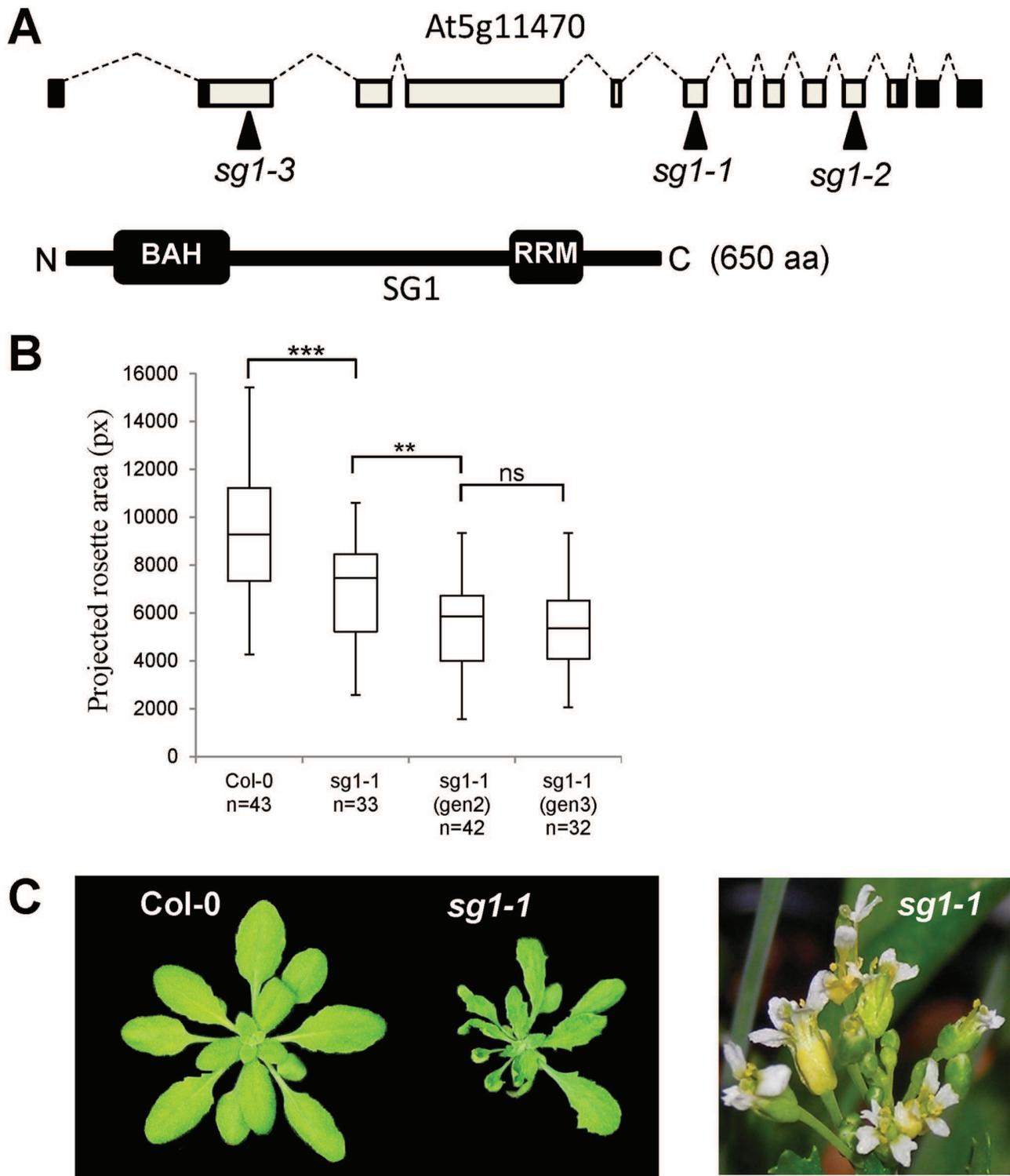


Figure 1. *SG1* encodes a protein with domains involved in chromatin modification. (A) Top: *SG1* gene structure showing UTRs (black boxes), exons (clear boxes), introns (dashed lines). The position of *sg1-1* mutation (cytosine at position 2,954 relative to ATG) and *sg1-2* or *sg1-3* T-DNA insertions are indicated by plain arrows. Bottom: protein structure showing the two putative BAH and RRM domains (black boxes). (B) Analysis of the projected rosette area (pixels) of 15 days-old seedlings grown *in vitro*. *sg1-1*(gen2) : progeny of *sg1-1*; *sg1-1*(gen3) : progeny of *sg1-1*(gen2). n: number of individuals analyzed for each genotype. ** $p < 0.01$, *** $p < 0.001$, ns: not significant. (C) Left panel: *sg1-1* phenotype of a 25 days-old plant (right) compared to WT Col-0 plants. Right panel: *sg1-1* floral defects. Note the aberrantly shaped petals, incorrect organs number. doi:10.1371/journal.pone.0084687.g001

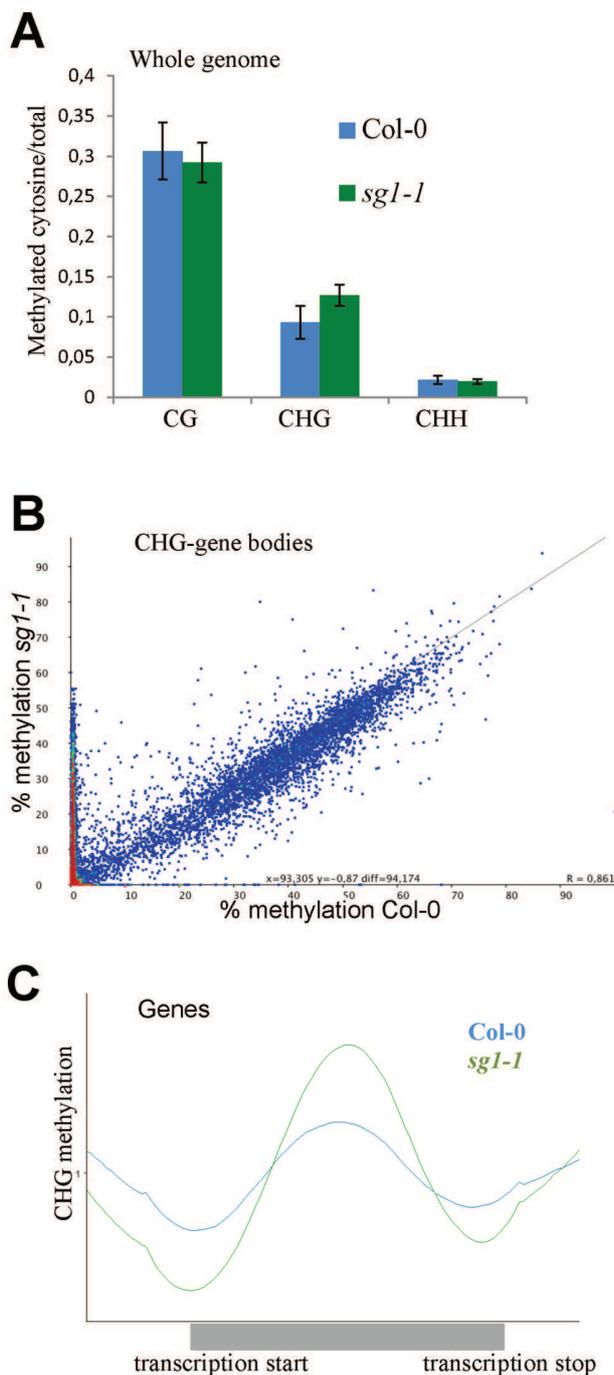


Figure 2. *sg1-1* methylome analysis. (A) Whole genome analysis of cytosine methylated in all three contexts of methylation (CG, CHG and CHH) in Col-0 and *sg1-1* mutants. Error bars indicate the SD between two biological replicates. (B) Scatter plot representing the percentage of CHG methylation calculated for each gene in *sg1-1* and Col-0. Hypermethylated genes in *sg1-1* are visualized by the dense red area. (C) Graphic representation of the average distribution of CHG methylation in gene bodies.
doi:10.1371/journal.pone.0084687.g002

S1). The average distribution of CHG methylation over genes indicates that genes are more methylated in their core bodies (Figure 2C). We found no systematic contrast in methylated-

cytosine content between *sg1-1* and Col-0 transposable elements (Figure S9 in File S1).

We fixed a minimal methylation ratio of two from Col-0 to *sg1-1* to classify genes as hyper- (2x increase in *sg1-1*) or hypomethylated (2x decrease in *sg1-1*; see Materials and Methods). Under these criteria, we found less than 100 genes that were hypomethylated in either one of the three contexts, one gene hypermethylated in the CG context, 0 in CHH and over 3,300 in the CHG context (Table S3 in File S2). Genes found to be CHG-hypermethylated were equally distributed among the five Arabidopsis chromosomes (Figure S10 in File S1). We selected six genes from this list to analyze in more details the distribution of the methylated cytosines within gene bodies. We confirmed that all these genes were significantly enriched in CHG-methylated cytosines in *sg1-1* mutants compared to Col-0 throughout the gene sequence (Figure 3), while no differences were observable for CG methylated cytosines, and a slight increase for some of them in the CHH contexts (Figure S11 in File S1). Altogether, our results indicate that SG1, very similarly to IBM1, protects certain protein-coding genes from ectopic CHG methylation. Notably, as observed in *ibm1-1* mutants [10], CHG and -to some extent- CHH methylation, extend from an adjacent transposon into the *BONSAI* locus in *sg1-1* (Figure 4A).

Since *IBM1* encodes a histone demethylase catalysing the removal of H3K9 marks, a process tightly linked to CHG methylation, we determined H3K9me2 contents by ChIP analyses in the subset of genes that present elevated levels of CHG methylation (Figure 5). Using several couples of oligos along gene bodies, we observed in both *sg1-1* and *sg1-2* mutant alleles as well as in *ibm1-1*, a significant enrichment of H3K9me2 localized within the center of genes, and to a lesser extent in UTRs (Figure 5). Similarly, the *BNS* gene was enriched in H3K9me2, particularly in regions adjacent to the transposon nearby (Figure 4B). Altogether, our results suggest a function for SG1 in the protection of genes from H3K9me2 and CHG DNA methylation, similarly to that described for IBM1.

Recently, Jacobsen and collaborators revealed the DNA methylome landscapes of several Arabidopsis mutants involved in gene silencing and chromatin remodeling, including components of the methylation pathways [16]. We retrieved the *ibm1* and the corresponding WT methylome sequences to identify the genes differently methylated in *sg1-1* versus *ibm1-6* (SALK_006042). Interestingly, we found that the vast majority of gene bodies differentially CHG methylated in *sg1-1* were also targeted by IBM1, but we also found that many genes more CHG methylated in *ibm1* were unchanged in *sg1-1* (Figure 6). We verified that the CHG methylomes of the WT samples were similar between the two experiments (Figure S12 in File S1).

CMT3 and KYP act Antagonistically to SG1

IBM1 and KYP have opposite effects on H3K9 marks, and this has a direct impact on CMT3 activity and CHG methylation. Thus, IBM1 restricts CMT3 and KYP activities to transposons and repeated sequences, protecting genes from elevated levels of CHG methylation and H3K9me2. To investigate the relationships between *SG1*, *IBM1*, *KYP* and *CMT3* we crossed *sg1-2* with mutants corresponding to these genes. Both *sg1/cmt3* and *sg1/kyp* double mutants showed a WT phenotype, indicating that either *cmt3* or *kyp* mutations alone can suppress the severe phenotype of *sg1* (Figure 7) and *ibm1* [8]. On the other hand, the double *sg1/ibm1* mutant is phenotypically similar to the *sg1* or *ibm1* mutants (Figure 7). This suggests a function for SG1 in the protection of genes from H3K9me2 and CHG DNA methylations, together with IBM1, and antagonistically to that of CMT3 and KYP.

CHG methylation

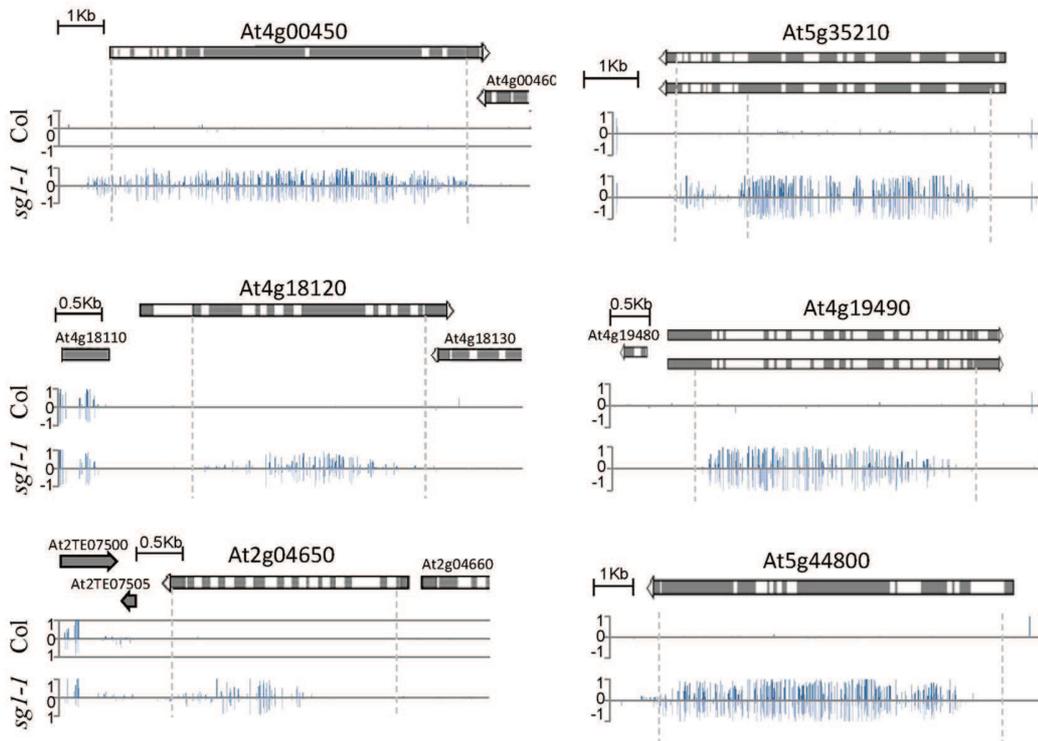


Figure 3. Gene-body CHG hypermethylation in *sg1-1*. Methylation on top (positive values) and bottom strand (negative values) across the coding sequence of the genes indicated is shown. Data are based on methylome results. The corresponding gene models are shown according to TAIR v10.

doi:10.1371/journal.pone.0084687.g003

CMT3 is the methyltransferase that maintains CHG methylation and MET1 is involved in the maintenance of CG methylation. To understand whether perturbing one of the other types of methylation in *sg1* would have some consequences, we combined the *met1* mutation with *sg1*. In the F2 population, we obtained only three plants fixed for both mutations, on a total of 96 plants (Figure S13A in File S1). The F3 progenies of one of these *met1/sg1* double mutants had a strong phenotype (Figure S13B in File S1) and were all sterile. Therefore, suppressing the CG-methylation in both genes and transposons (i.e. *met1*), while adding silencing marks in the core of genes (i.e. *sg1*) has a deleterious immediate effect on plant development.

Previous genetic analysis showed that some components of the RdDM pathway, including RDR2, NRPD1A and AGO4, are dispensable for *ibm1*-induced DNA hypermethylation defects, suggesting that de novo DNA methylation of the *IBM1* targets is subject to siRNA-independent regulation [9]. Therefore, we investigated if hypermethylation in *sg1* was also siRNA-independent, similarly to what was reported for *ibm1*. When we combined *sg1* with a mutant lacking the AGO4 function that is crucial for RdDM processes [2], we found that the double *sg1-2/ago4-2* mutant was phenotypically very similar to *sg1-2*, suggesting that *sg1*-induced hypermethylation does not depend on components of RdDM (Figure S14 in File S1).

Levels of mRNAs in *sg1*, Methylation and Consequences on Gene Function

One of the most intriguing questions regarding the methylation marks that appear in gene bodies of *sg1* or *ibm1* mutants is whether they have direct consequences on gene expression, and ultimately on the function of these affected genes. Recently, *IBM1* was suggested to act indirectly on the RdDM pathway by targeting both *RDR2* and *DCL3*, reducing their expression, correlating with a final up-regulation of some RdDM-targets [11]. We analyzed the H3K9me2 enrichment at *RDR2* and *DCL3* locus, by ChIP analysis. We found that, like *ibm1*, *sg1* mutants contain higher levels of H3K9me2 and cytosines methylated in the CHG context at both locus compared to the Col-0 WT plants (Figure 8A); this correlated with more subtle modifications of CHH cytosine methylation but not with CG methylation (Figure S15 in File S1). We confirmed that both *RDR2* and *DCL3* gene expression is reduced by half in *ibm1* and we found similar results for the two *sg1* alleles tested (Figure 8B). We also found other members of the RdDM pathway with a CHG-methylation status affected in *sg1*, including *NRPD1A*, one of the polIV subunits (Table S4 in File S2). However, even if some of the RdDM key players are impacted by *sg1*-induced methylation defects (at CHG, CHH and H3K9 levels), the functional consequences on the RdDM pathway activity remain to be determined.

We tested the expression of several other targets having gene bodies enriched in both H3K9me2 and CHG, and found genes (At4g00450 and At5g35210, see Figures 3 and 5) whose expression was not modified between *sg1-1* and WT plants (Figure S16 in File

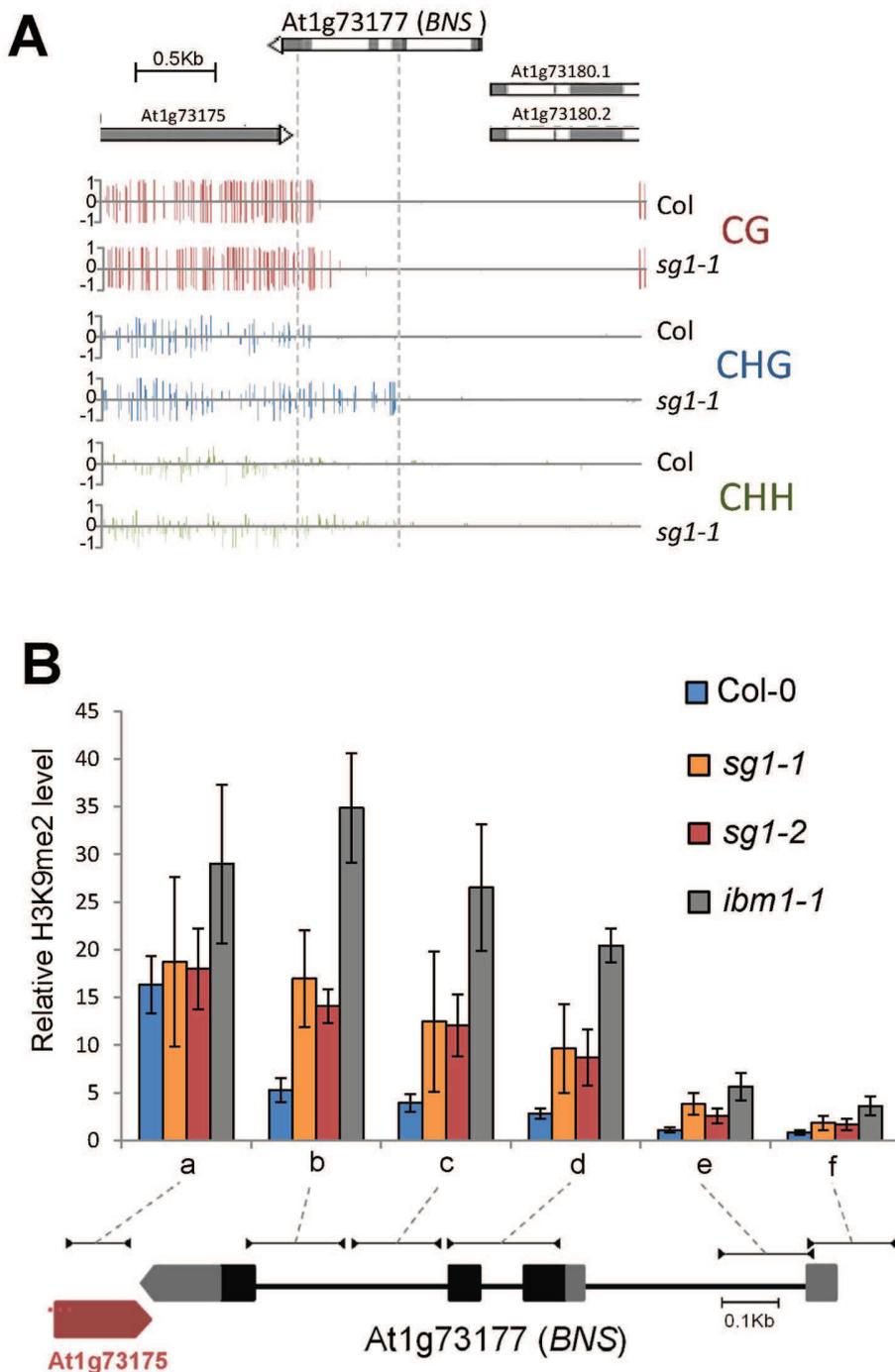


Figure 4. Methylation analysis at *BONSAI* in *sg1-1*. (A) Methylation in the three cytosine contexts near *BNS*. The spreading of DNA methylation from the adjacent LINE element (At1g73175) is visible in *sg1-1*. (B) Histone H3K9me2 accumulation in *BNS* determined by ChIP analyses. The correspondence of genomic regions tested by ChIP is shown on a gene schematic based on TAIR v10. Error bars are SEM from at least three biological replicates.

doi:10.1371/journal.pone.0084687.g004

S1). We concluded that *sg1*-induced methylation defects are not necessarily correlated with changes in the final quantity of transcripts detected, or that *sg1* may induce other changes such as transcript processing alteration that were not detected in our gene expression assays.

Since *sg1* and *ibm1* phenotypes are very similar, both molecularly and morphologically, we thought that the function of *IBM1* could be compromised in *sg1* mutants. The *IBM1* gene encodes two different transcripts and only the longest one (*IBM1-L*) is functional and can complement an *ibm1* mutant [17]. We first tested whether the *sg1* phenotype could be the result of *IBM1*

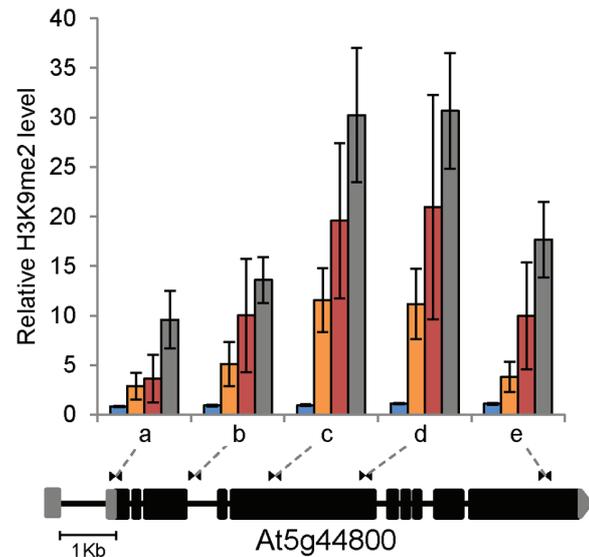
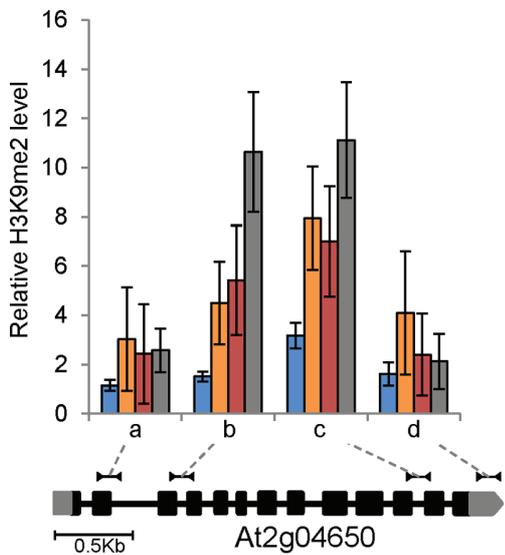
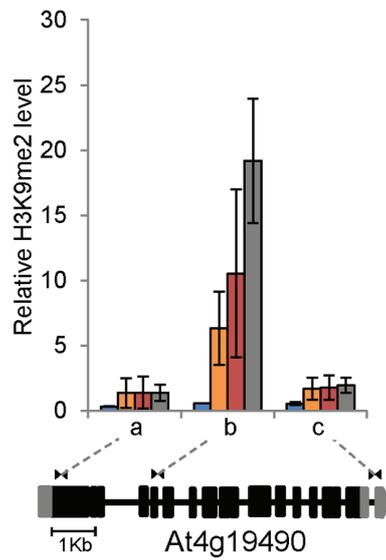
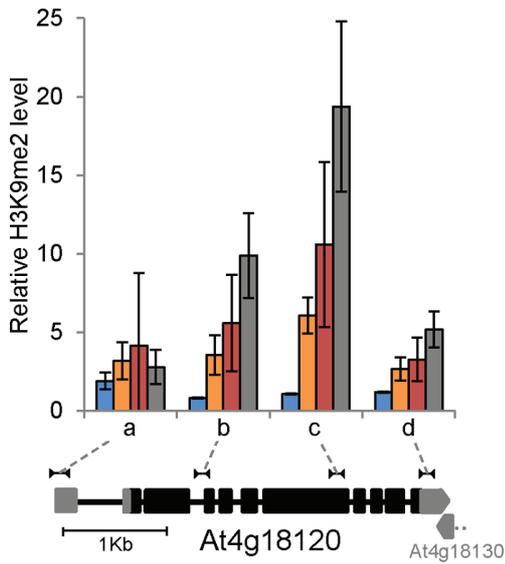
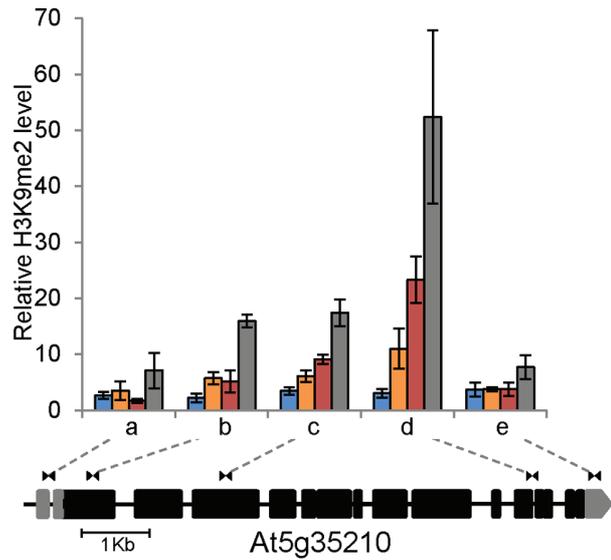
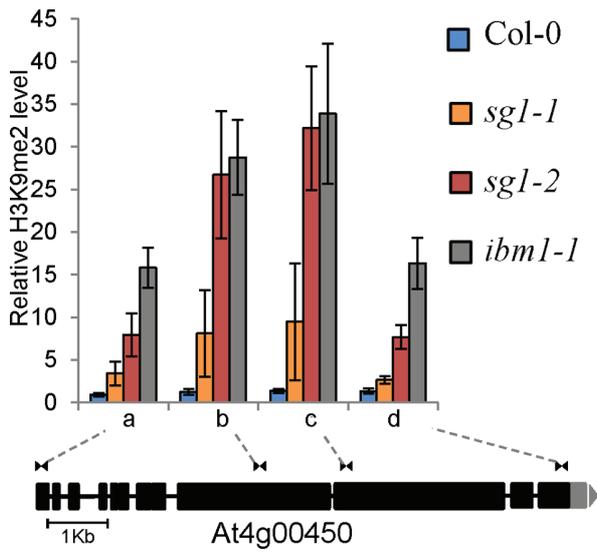


Figure 5. Gene-body H3K9me2 accumulation in *sg1-1* and *ibm1* mutants. H3K9me2 profile across six genes mentioned in Figure 3, in WT Col-0, *sg1-1*, *sg1-2* and *ibm1-1* plants. The correspondence of genomic regions tested by ChIP is shown on a gene schematic based on TAIR v10. Error bars are SEM from at least three biological replicates.
doi:10.1371/journal.pone.0084687.g005

transcript misregulation. To this end we examined by qRT-PCR the proportion of the *IBM1-L* mRNA product relative to total *IBM1* mRNA product (as defined by the amount of transcript upstream of the large intron heterochromatic region). Interestingly, we found a stronger reduction in the *sg1-1*/WT mRNA ratio corresponding to the *IBM1-L* product relative to the total amount (Figure S17 in File S1). This suggested that a significant fraction of the mRNA produced at the *IBM1* locus in a *sg1* mutant may correspond to the short, non-functional *IBM1-S* product. We then transformed *sg1* mutants with a transgene carrying a full-length functional *IBM1-L* cDNA construct [17]. Both *ibm1-1* and *sg1-1* were complemented by *IBM1-L*, indicating that the phenotype of *sg1* mutants most likely results from the compromised function of *IBM1* in *sg1* (Figure S18 in File S1).

Discussion

SG1 and IBM1 act in the Same Pathway

In an effort to map molecular determinants quantitatively controlling growth, we identified a new gene, *SG1* (At5g11470), coding a protein with domains found in chromatin regulators. We identified the *sg1-1* allele corresponding to a recent spontaneous point mutation in the *SG1* coding sequence from the specific Col-0 parent line used to build our mapping population. We also isolated two additional T-DNA alleles (*sg1-2* and *-3*) allowing us to determine that *sg1*-related phenotypes became apparent from the second homozygous generation. All *sg1* mutants present similar phenotypes appearing stochastically, namely growth defects, abnormally shaped organs (including flowers and leaves) and ultimately leading to sterility. The penetrance and severity of the phenotype increase across generations, similar to phenotypic

defects previously reported for *ibm1* [10]. Since *IBM1* is controlling H3K9 methylation, with consequences on the levels of cytosines methylated in the CHG context, we performed a high-throughput methylome sequencing of *sg1-1* mutants, showing that gene bodies are CHG-hypermethylated in the mutant background (Figure 2B and C) while transposable elements are not affected (Figure S9 in File S1). This was correlated with an enrichment of the heterochromatin mark H3K9me2 at gene bodies of all CHG hypermethylated candidates tested in our study (Figures 3, 4, 5 and 8A). These results strongly suggest that *SG1* is working closely to *IBM1*, indeed, similar enrichments within gene-bodies have been reported in an *ibm1* mutant context [9]. Altogether, these observations suggested a function for *SG1* in the maintenance of genome integrity, likely at the chromatin regulation level, together with *IBM1*. Recently, two other *sg1* alleles were identified, namely *ibm2* and *asi1* [18–19]. In agreement with our hypothesis for *SG1* function, the authors showed that *SG1/IBM2/ASI1* is required for the processing of long transcripts over intronic heterochromatic marks. Given the presence of an intragenic heterochromatin region in the large intron of *IBM1*, this conclusion is supported by findings that *IBM1* full-length transcript was decreased in the *sg1* mutant compared to the short transcript (Figure S17 in File S1) and that a functional *IBM1* transcript complement *sg1* (Figure S18 in File S1). *SG1/IBM2/ASI1* might indeed function by promoting the use of distal polyadenylation sites over proximal ones located in large introns [18–19].

SG1 contains two predicted domains, one of them, the BAH domain, being found in various chromatin regulating factors. In plants, one of the two BAH domains of *MET1* for instance can physically interact with the C-terminal region of the histone deacetylase *HDA6*, [20]. Interestingly, BAH domains can also

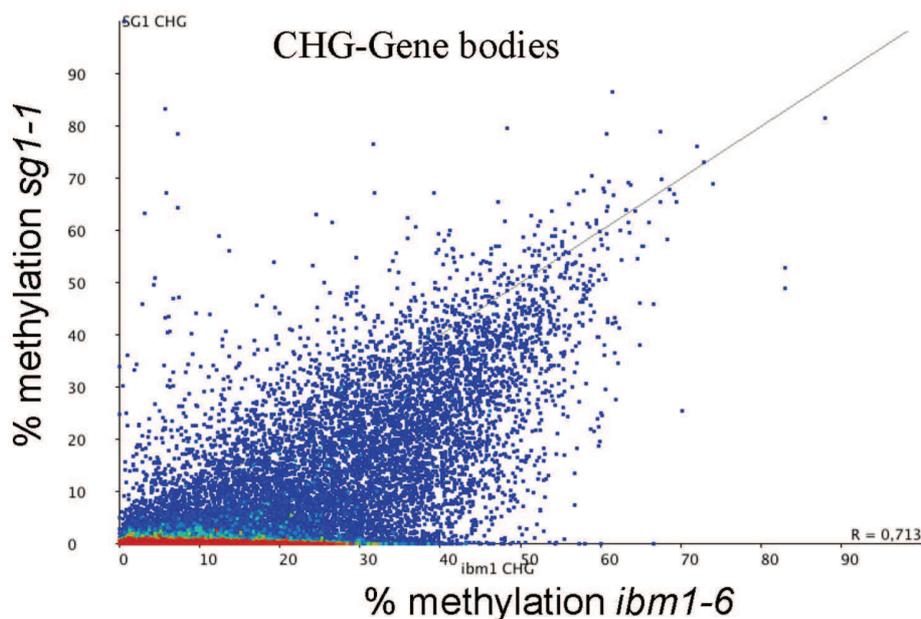


Figure 6. Comparison of *sg1-1* and *ibm1-6* CHG methylated genes. Scatter plot representing the percentage of CHG methylation calculated for each gene in *sg1-1* and *ibm1-6*, according to the data publicly available [16]. Genes solely methylated in *ibm1* are visualised by the dense red area.
doi:10.1371/journal.pone.0084687.g006

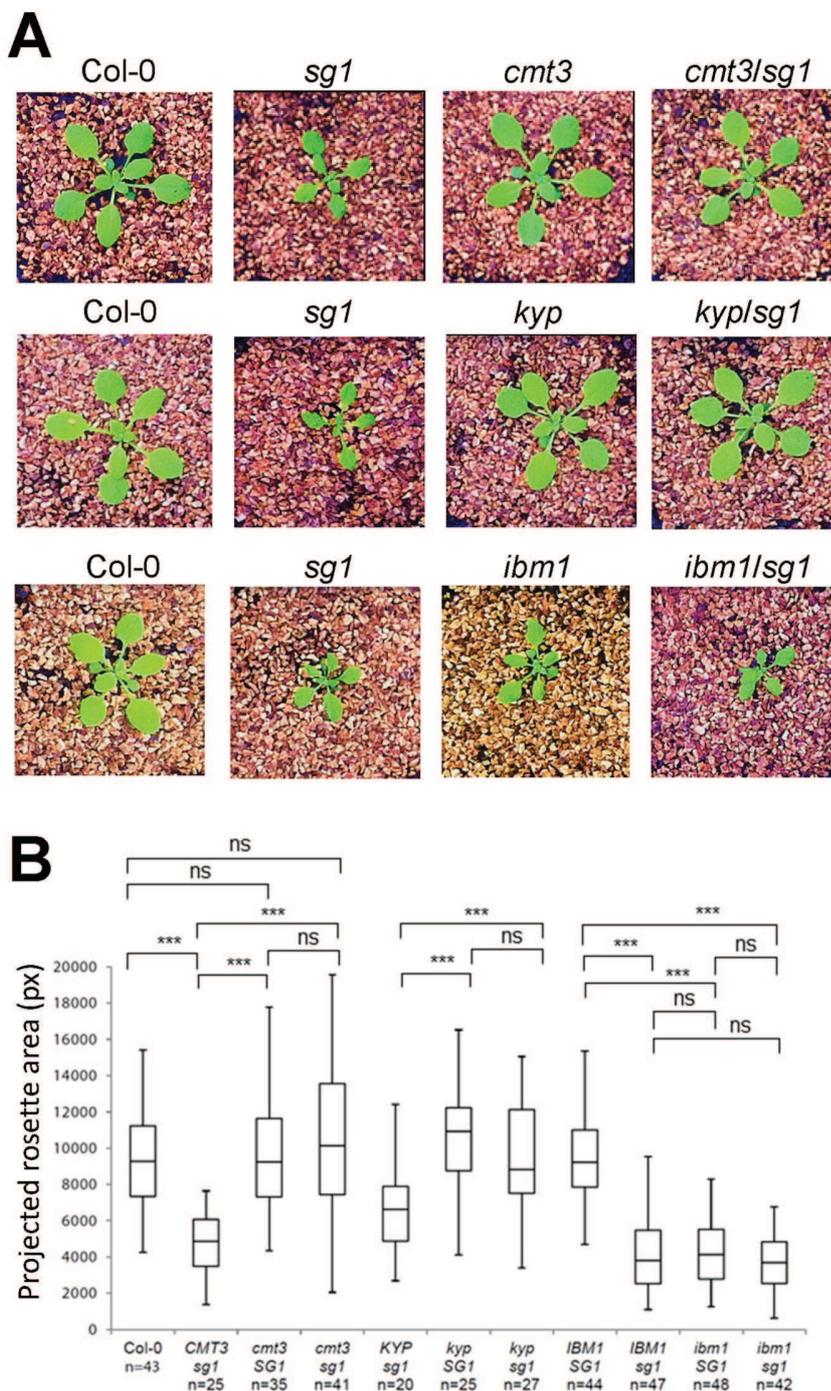


Figure 7. *sg1* phenotype is suppressed by *cmt3* or *kyp* mutations, but not *ibm1*. (A) Pictures of typical 15 days-old seedling corresponding to homozygous double mutants (F4) resulting from the cross between *sg1-2* and *cmt3-11*, *kyp* (line SALK_069326) or *ibm1-1* mutants. Contrast was enhanced for picture visibility. (B) Analysis of the projected rosette area (pixels) of 15 days-old double mutant seedlings (as described in A) grown *in vitro*. n: number of individuals analyzed for each genotype. *** $p < 0.001$, ns: not significant. doi:10.1371/journal.pone.0084687.g007

directly recognize histone marks. A component of the Origin of Replication Complex-1 (ORC1) can bind in a specific way H4K20me2 marks that are enriched at replication origins in mammals [21]. In maize, the homolog of CMT3 binds H3K9me2 marks via a dual recognition through both the BAH domain and the chromodomain of the protein [6]. Therefore SG1

may be recruited to the chromatin via its BAH domain. In addition, SG1 contains a RRM motif shared by a large number of RNA-binding proteins that could also mediate the targeting of SG1 to particular loci. Accordingly, Wang *et al.* showed *in vitro* that SG1/IBM2/ASI1 has RNA-binding properties, and that SG1/IBM2/ASI1 is enriched at intronic heterochromatin regions

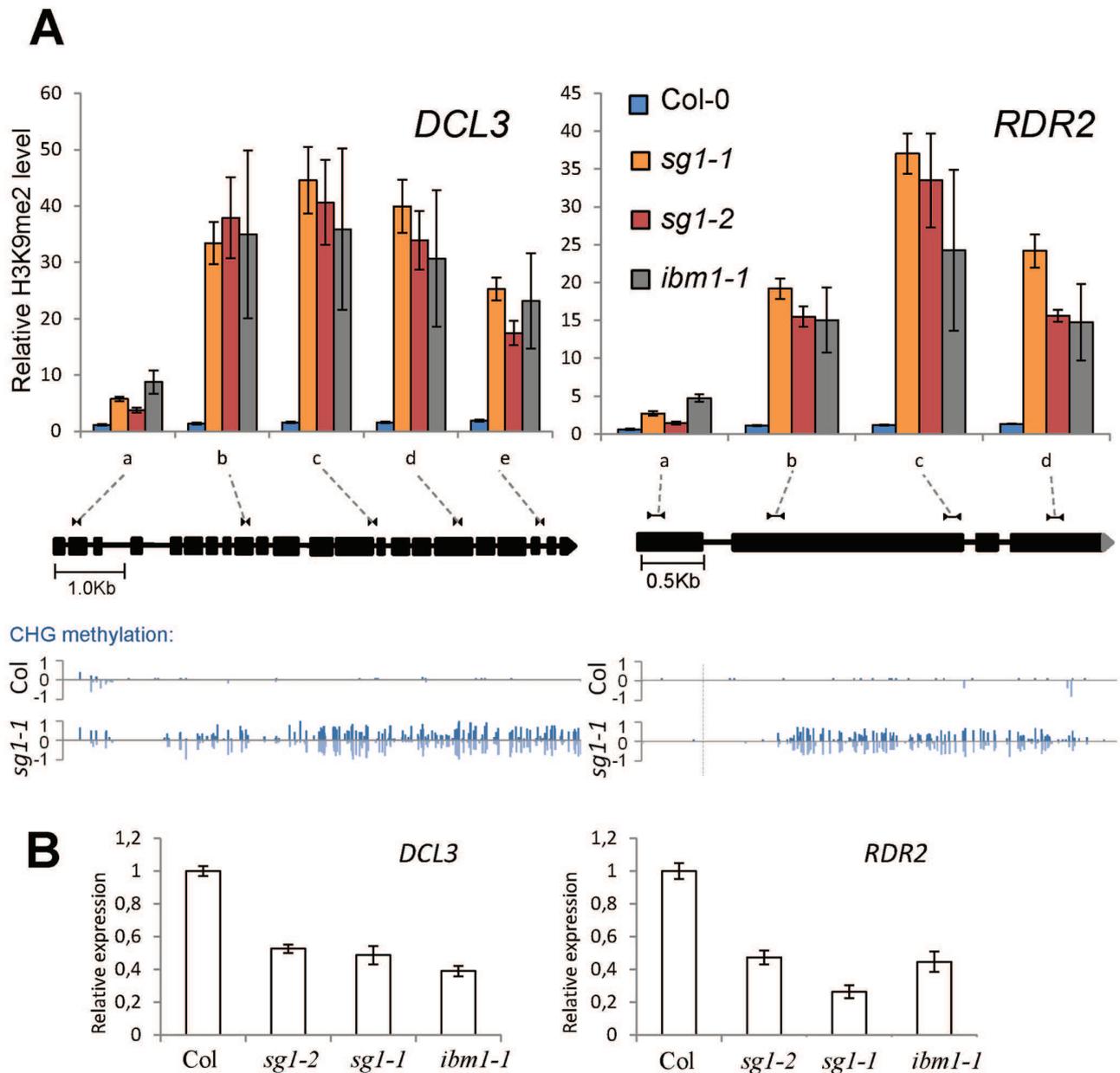


Figure 8. *sg1* and key players of the RdDM pathway. (A) H3K9me2 profile across *RDR2* and *DCL3* genes in WT Col-0, *sg1-1*, *sg1-2* and *ibm1-1* plants. The correspondence of genomic regions tested by ChIP is shown on a gene schematic based on TAIR v10; CHG methylation is presented below. Error bars are SEM from at least three biological replicates. (B) Transcript accumulation of *RDR2* and *DCL3* in *sg1* and *ibm1* determined by qPCR analyses. Error bars are SEM from three replicates. doi:10.1371/journal.pone.0084687.g008

[19]. The authors proposed a model in which SG1/IBM2/ASI1 is binding to intronic heterochromatin regions and nascent pre-mRNA via its BAH domain and RRM motif respectively, but whether SG1/IBM2/ASI1 BAH domain is directly involved in heterochromatin recruitment remains to be determined.

By comparing our methylome data with the one published recently for *ibm1* mutants [16], we observed that most IBM1 target genes are also affected in *sg1* mutants, while IBM1 has additional targets (Figure 6). However, this does not necessarily mean that IBM1 has a broader range of target genes compared to SG1. This observation might be the consequence of a generational effect since the phenotypes of *sg1* and *ibm1* mutants become more

deleterious after several generations. We suspect that the *ibm1* mutant used in the recent study of Jacobsen and coworkers [16] had been accumulating gene methylation for more generations than the *sg1* mutant used in our methylome analysis. In regions where both *ibm1* and *sg1* mutations affect the same genes, we observed that the CHG methylation in *sg1* seems to be more localized to the center of gene bodies. Contrariwise, the CHG methylation in the *ibm1* mutant, while being also localized to gene bodies, seems to be more extended toward the UTRs (Figure S19 in File S1). It is tempting to speculate that generations after generations, the methylation that first invade specifically certain regions of the genes slowly spread to adjacent regions in plants

lacking IBM1 or SG1 function. We also detected genes that are methylated in *ibm1* while being almost free of CHG methylation in the *sg1* background (Figure S20 in File S1). Since the phenotype is strongly correlated with the number of generations, it is possible that secondary targets, not impacted in the first generations, are gradually becoming methylated in the progenies, increasing the deleterious effects of the *ibm1* mutation. We currently do not know how methylation acts on mRNA accumulation, but the lethal effects observed after several generations indicate that the constant and possible increasing invasion of the methylation have strong consequences on plant development. To further confirm these hypotheses, methylome analyses of both *sg1* and *ibm1* mutants will have to be compared for the same generations.

Interconnection between Different Methylation Contexts of Gene Bodies

We have shown that the maintenance DNA methyltransferase CMT3 or the H3K9 methyltransferase KYP are epistatic to *SG1*. Both *cmt3* or *kyp* mutations suppress the phenotype of *sg1*, similarly to what was reported for *ibm1* and suggesting that SG1 functions antagonistically to CMT3 or KYP. In addition to KYP, two other methyltransferases are found in Arabidopsis (SUVH5 and 6), but since the suppression of KYP can restore the WT phenotype in a *sg1* background, it is likely that ectopic H3K9me2 marks deposition in gene bodies of *sg1* mostly involves KYP. The double *ibm1/sg1* mutant displayed similar phenotypes to a single *sg1* or *ibm1* mutant, confirming that IBM1 and SG1 act in the same pathway. Altogether, the interplay between the self-reinforcing CMT3/KYP loop and SG1/IBM1 maintains the Arabidopsis epigenome integrity to ensure correct expression of genes. How SG1 functions to prevent CMT3 and KYP activities at a given locus remains unclear, as well as the initial starting signal that first promotes their activities in genes. But the self-reinforcing loop and the subsequent spreading of the methylation generations after generations seem to be key elements to explain the increase invasion of genes by silencing marks. Recently, DDM1, encoding a chromatin remodeling factor, was shown to restrict CMT3/KYP activities at *BNS* and many other locus. The methylation at *BNS* in all cytosine contexts in a *ddm1* background depends on KYP and CMT3 since the corresponding mutants individually suppress this methylation [22]. In *ddm1*, residual methylation from the adjacent LINE element spreads to *BNS* after several generations, and through the repeated action of both CMT3 and KYP, exactly as in *sg1* or *ibm1* mutants.

The activity of IBM1 seems to be more prone to target actively transcribed genes [9] and the functions of IBM1 and SG1 are essential to keep gene bodies free of CHG methylation. Paradoxically, these genes are not exempt of methylation in the WT, but in a CG-context (Figures S11 and S15 in File S1). The presence of CG-methylation in actively transcribed genes has been described for years, with a role still undetermined at present. Several hypotheses are being examined to discover possible functions for that methylation, some converging towards a potential role in regulating splicing [23]. SG1 could possibly target CG-methylated genes in WT plants, and such targeting could rely on active transcription, or CG-methylation of gene bodies, or both, by direct binding of SG1 to proteins involved in these processes. Intriguingly, even if methylation invades some genes in *sg1* or *ibm1*, it does not prevent their transcription, or at least it does not change the level of transcripts detected by qPCR which is only tested through small transcript sections of 200 bp. The links between transcription or splicing and methylation (CG, CHG or H3K9me2) remains to be investigated further but *sg1*

mutants could be appropriate candidates to clarify that point in plants.

Another intriguing observation is that certain CHG-enriched gene bodies of *sg1* seem to be also slightly enriched in CHH methylation, compared to the WT (Figures S11 and S14 in File S1); an observation also true for *ibm1* mutant plants. Part of the CHH methylation depends on the RdDM pathway and the production of sRNAs, but the ectopic CHH methylation in *sg1* is unlikely due to a stimulation of the RdDM pathway and a production of sRNAs from genic regions: first, *ago4/sg1* double mutants are phenotypically similar to *sg1*, indicating that disrupting the RdDM pathway has limited consequences on *sg1* plants and, second, it is unlikely that *sg1* mutants and WT plants have genome-wide dissimilar contents of sRNAs, since *ibm1* and WT plants have not [11]. However, recent evidences point toward the crucial role of CMT2 in maintaining the CHH methylation that is independent on RdDM [3]. The authors hypothesized that CMT2 might bind H3K9me2 marks, exactly like CMT3, to target DNA and methylate cytosines in the CHH context. Genes enriched in H3K9me2 methylation are then also probably targeted by CMT2 in *sg1* and *ibm1*, explaining their enrichment in CHH. H3K9me2 marks could therefore be linked to two contexts of methylated cytosines, CHH and CHG. Further work is required to shed the light on the mechanisms of interplay between SG1 and IBM1, as well as to understand the functional consequences of genic non-CG methylation.

Materials and Methods

Plant Material and Growth Conditions

A. thaliana RILs and HIFs were obtained from the INRA Versailles collection (<http://publiclines.versailles.inra.fr/>) and described before [24–25]. *sg1* T-DNA mutants were obtained from the Arabidopsis stock center: *sg1-2* is SAIL_310B06 and *sg1-3* corresponds to GK_045A07. In addition, the following Arabidopsis mutants were used: *ago4-2* [26], *cmt3-11* [27], *kyp* (SALK_069326) [28], *met1-1* [14], and *ibm1-1* [10]. Projected area analyses of leaves were performed as previously described [12].

Whole-genome Bisulfite Sequencing and Analyses

Col-0 and *sg1-1* homozygous seeds were grown in long day conditions for 25 days in culture rooms and 5 to 25 μ g of genomic DNA were extracted using a CTAB phenol-chloroform extraction protocol followed by RNase treatment (Qiagen). Bisulfite treatment, library preparation and whole genome sequencing were performed at BGI (Shenzhen, China) using Illumina's recommendations and HiSeq technology for paired-end 91 bp reads. We sequenced the methylome of two biological replicates for the WT Col-0 and two for *sg1-1*. The sequencing results are described in Table S1 in File S2.

For sequencing analysis low-quality reads ($q < 30$) were discarded and clean reads were mapped to the Col-0 *Arabidopsis thaliana* TAIR 10 reference genome using the Bowtie and Bismark softwares with options '-q -n 3-non directional'. Subsequent analyses were done with SeqMonk (Babraham Bioinformatics - SeqMonk Mapped Sequence Analysis Tool by Simon Andrews). Only cytosines covered by at least 5 unique reads were retained. For genome-wide analyses, we defined windows of 200 bp overlapping by 50 bp intervals to calculate the percentages of methylation along the genome. To determine differentially methylated genes, a Fisher's exact test was performed comparing methylated and unmethylated counts. We then corrected nominal *P*-values and estimated the corresponding *q*-values to retain genes with *q*-values < 0.05 . For *ibm1-6* and the corresponding WT Col-0,

we retrieved the raw data from NCBI (GEO query: GSE39901) to align it as described for the data obtained in the present study.

Chromatin Immunoprecipitation (ChIP) and Real-time Quantitative PCR (qPCR) Analyses

ChIP assays were performed as previously described [29] on the aerial parts of 25 days-old seedlings using anti-histone H3 dimethyl lysine 9 antibody from Abcam (Catalog No. ab1220) and anti-histone H3 antibody from Abcam (Catalog No. ab1791). All ChIP experiments were quantified by qPCR using primers listed in Table S5 in File S2. *SHOOT MERISTEMLESS* (*STM*) was used as the internal control for the ChIP experiments. Data are represented as the ratio of (H3K9me2 gene/H3 gene)/(H3K9me2 *STM*/H3 *STM*). Experiments were repeated in at least three biological replicates.

Gene Expression Analysis

Total RNA was isolated from the aerial parts of 15 days-old seedlings using the RNeasy Plant Mini kit (Qiagen) followed by a DNase treatment (Fermentas). RT-PCR was performed on 500 ng of total RNAs with the M-MLV reverse transcriptase (Fermentas) and cDNAs were diluted 10 times. 5 μ l were used for qRT-PCRs using a CFX96 real-time PCR machine (BioRad) with a SYBR solution (Eurogentec) using the primers listed in Table S5 in File S2. Expression levels were normalized against the Arabidopsis *UBC21* gene (At5g25760). Experiments were repeated in three biological replicates.

Supporting Information

File S1 Contains: Figure S1: *SG1* fine mapping. (A) QTL mapping for shoot size in the Bur-0 x Col-0 RIL set identifies a major QTL on chromosome 5 [12]. (B) Fine-mapping of the major-effect locus to 8 kb around At5g11470 on chromosome 5. **Figure S2: *RG1* fine mapping in the Ct-1 x Col-0 maps to *SG1*.** QTL mapping for root growth (total root length) in the Ct-1 x Col-0 RIL set (A) reveals *RG1*, a QTL confirmed *in vitro* 10 days after germination (days after germination) (B) and in the greenhouse 20 days after germination (C). Fine-mapping and sequencing revealed that *sg1-1* mutation is responsible for this QTL. **Figure S3: *SG1* gene models.** *SG1* gene models according to TAIR v10, EuGene and our cDNA sequencing data is shown with UTRs (black boxes), exons (clear boxes) and introns (dashed lines). Left, 5' end; right, 3' end. **Figure S4: *sg1* phenotype at the flowering stage.** Picture of a tray of 40 days-old WT Col-0 plants (left) and *sg1-1* plants (middle, right) derived from a parent exhibiting (a) or not (b) a phenotype. The flowering delay can be seen for *sg1-1* (a). **Figure S5: *sg1-2* and *sg1-3* mutants.** (A) Phenotype of plants homozygous for *sg1-2* T-DNA (bottom) and WT Col-0 (top) in two successive generations after fixation at the homozygous state. (B) Analysis of the projected rosette area (pixels) of 15 days-old seedlings grown *in vitro*. *sg1-2*(gen2): second homozygous generation. Each following generation (gen3, gen4 and gen5) descends from the previous one. *** $p < 0.001$, ns: not significant. (C) Picture of a mature *sg1-2* plant. The arrow points one sterile bud that failed to develop into a silique. The frame shows an enlargement of a rosette leaf showing the over-serrated shape. (D) Segregating plants (left) originating from a unique *sg1-3* homozygous parent illustrate the stochasticity of the *sg1* phenotype compared to segregating plants from a unique WT Col-0 parent (right). **Figure S6: Proteins containing Bromo adjacent homology (BAH) domain (IPR001025) in Arabidopsis.** The different domains (according to the InterPro database) are indicated: Agenet (IPR008395),

ATPase AAA core (IPR003959), ZnF PHD-finger (IPR013083), ZnF PHD = finger (IPR019787), Chromodomain (IPR023780), TFIIIS centre (IPR003618), TFIIIS/elonginA/CRSP70 N (IPR017923), C5 DNA meth (IPR022702) and RRM RNP (IPR012677). **Figure S7: Comparison of the percentages of methylation in 200 bp-windows throughout the genome of *sg1-1* and Col-0.** Each 200 bp-window is represented by a dot. Windows were designed to overlap by 50 bp and only cytosine position covered by at least 5 reads were considered in the calculation. Reads that were not matching at unique locations were discarded. The percentage of methylation in CG (A) and CHH (C) contexts are similar between Col-0 and *sg1-1*, while CHG methylated regions are over-represented in *sg1-1* (B). **Figure S8: Scatter plots comparing the percentages of methylation in gene bodies of *sg1-1* and Col-0.** Each gene according to the TAIR10 annotations is represented by a dot. Only cytosine position covered by at least 5 reads were considered in the calculation. Reads that were not mapping to unique locations were discarded. The percentage of methylation in CG (A) are similar between Col-0 and *sg1*, while CHH-methylated genes seem to be slightly more methylated in *sg1-1* (B). **Figure S9: Comparison of the percentages of methylation in transposable elements (TEs) in *sg1-1* and Col-0.** Each transposable element according to the TAIR10 annotations is represented by a plot. Only cytosine position covered by at least 5 reads were considered in the calculation. Reads that were not matching at unique locations were discarded. The percentage of methylation in CG (A), CHG (B) and CHH (C) are similar between Col-0 and *sg1-1*. **Figure S10: Genomic distribution of the CHG differentially-methylated genes identified in *sg1-1*.** Only genes that present a 2x increase in CHG methylation are represented (see Mat&Met). One hypermethylated gene corresponds to a black line drawn on one of the five chromosomes of Arabidopsis. Gene-poor centromeres are indicated by a 'c'. The Chromosome Map Tool of TAIR (www.arabidopsis.org) was used to map the different genes on chromosomes. **Figure S11: Gene-body methylation in *sg1-1*.** Data is based on methylome results. CG, CHG and CHH methylations are shown on top (positive values) and bottom strand (negative values) across the genes indicated. The corresponding gene models are shown according to TAIR v10. **Figure S12: Comparison of data samples for *sg1-1* and *ibm1-6* methylome analyses.** (A) Distribution of the percentages of CHG methylation of genes in different samples: *sg1-1*, *ibm1-6*, the WT Col-0 used in this study and one of the three Col-0 samples (sample #3) used by Jacobsen and colleagues [16]. (B) Scatter plots comparing the percentages of methylation in gene bodies of the Col-0 WT used in this study and one of the three Col-0 samples (sample #3) used by Jacobsen and colleagues [16]. Each gene according to the TAIR10 annotations is represented by a dot. The percentage of methylation in CHG are similar between the two samples. **Figure S13: Phenotype of the double *met1-1/sg1-2* mutant.** (A) Genotyped F2 plants from the *met1-1/sg1-2* cross. We found only three plants fixed for the two mutations (circled in red). (B) F3 plants (descending from a F2-genotyped *met1/sg1* plant shown in A) were grown in the greenhouse and pictured after three weeks. **Figure S14: Phenotype of the double *ago4-2/sg1-2* mutant.** (A) Phenotype of the *sg1-2/ago4-2* mutant grown in greenhouse. (B) Analysis of the projected rosette area (pixels) of 15 days-old double mutant seedlings grown *in vitro*. n: number of individuals analyzed for each genotype. *** $p < 0.001$, ns: not significant. **Figure S15: Gene-body methylation of *RDR2* and *DCL3* in *sg1-1*.** Data is based on methylome results. CG, CHG and CHH methylations are shown on top (positive values) and bottom strand (negative

values) across the genes indicated. The corresponding gene models are shown according to TAIR v10. **Figure S16: qRT-PCR expression analysis of At4g00450 and At5g35210.** Error bars are SEM from three technical replicates. **Figure S17: qRT-PCR expression analysis of *IBM1*, expressed as a percentage of mRNA expression in *sg1* relative to WT.** A gene model representation shows the two mRNA regions assayed, either upstream or downstream of the heterochromatin region located in the large *IBM1* intron. Error bars are SEM from three technical replicates. **Figure S18: Complementation of *sg1-1* by *IBM1*.** *ibm1-1* and *sg1-1* were transformed by a construct carrying a functional *IBM1* cDNA [17]. T2 plants were selected on kanamycin, transferred to the greenhouse and pictured after three weeks of growth. **Figure S19: Examples of genes differentially methylated in *sg1-1* and *ibm1*.** Graphic representation is based on the publicly available data for *ibm1-6* and the corresponding Col-0 methylome [16]. The genome annotation is based on TAIR v10. The red box indicates the genes differentially methylated. **Figure S20: Examples of genes CHG-methylated in *ibm1* but not *sg1-1*.** Graphic representation is based on the publicly available data for *ibm1-6* and the corresponding Col-0 methylome [16]. The genome annotation is based on TAIR v10. The red box indicates the genes differentially methylated.

(PDF)

References

- Lister R, Pelizzola M, Downen RH, Hawkins RD, Hon G, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462: 315–322.
- Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Rev. Genet.* 11: 204–220.
- Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, et al. (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153: 193–205.
- Huff JT, Zilberman D (2012) Regulation of biological accuracy, precision, and memory by plant chromatin organization. *Curr. Opin. Genet. Dev.* 22: 132–138.
- Ebbs ML, Bender J (2006) Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. *Plant Cell* 18: 1166–1176.
- Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, et al. (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* 151: 167–180.
- Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, et al. (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr. Biol.* 17: 379–384.
- Inagaki S, Miura-Kamio A, Nakamura Y, Lu F, Cui X, et al. (2010) Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. *EMBO J.* 29: 3496–3506.
- Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, et al. (2009) An Arabidopsis *jmjC* domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J.* 28: 1078–1086.
- Saze H, Shiraishi A, Miura A, Kakutani T (2008) Control of genic DNA methylation by a *jmjC* domain-containing protein in Arabidopsis thaliana. *Science* 319: 462–465.
- Fan D, Dai Y, Wang X, Wang Z, He H, et al. (2012) *IBM1*, a *JmjC* domain-containing histone demethylase, is involved in the regulation of RNA-directed DNA methylation through the epigenetic control of *RDR2* and *DCL3* expression in Arabidopsis. *Nucl. Acids Res.* 40: 8905–8916.
- Vlad D, Rappaport F, Simon M, Loudet O (2010) Gene transposition causing natural variation for growth in Arabidopsis thaliana. *PLoS Genet.* 6: e1000945.
- Schiex T, Moisan A, Rouzé P (2001) EuGene: An Eucaryotic Gene Finder that combines several sources of evidence. *Comp. Biol. LNCS 2066: LNCS 2066.* pp. 111–125.
- Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, et al. (2003) Arabidopsis *MET1* cytosine methyltransferase mutants. *Genetics* 163: 1109–1122.
- Vongs A, Kakutani T, Martienssen RA, Richards EJ (1993) Arabidopsis thaliana DNA methylation mutants. *Science* 260: 1926–1928.
- Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* 152: 352–364.
- Rigal M, Kevei Z, Pélissier T, Mathieu O (2012) DNA methylation in an intron of the *IBM1* histone demethylase gene stabilizes chromatin modification patterns. *EMBO J.* 31(13): 2981–93.
- Saze H, Kitayama J, Takashima K, Miura S, Harukawa Y, et al. (2013) Mechanism for full-length RNA processing of Arabidopsis genes containing intragenic heterochromatin. *Nat Commun.* 4: 2301.
- Wang X, Duan CG, Tang K, Wang B, Zhang H, et al. (2013) RNA-binding protein regulates plant DNA methylation by controlling mRNA processing at the intronic heterochromatin-containing gene *IBM1*. *Proc Natl Acad Sci U S A.* 110(38): 15467–72.
- Liu X, Yu CW, Duan J, Luo M, Wang K, et al. (2012) HDA6 directly interacts with DNA methyltransferase *MET1* and maintains transposable element silencing in Arabidopsis. *Plant Physiol.* 158: 119–129.
- Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, et al. (2012) The BAH domain of *ORC1* links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* 484: 115–119.
- Sasaki T, Kobayashi A, Saze H, Kakutani T (2012) RNAi-independent de novo DNA methylation revealed in Arabidopsis mutants of chromatin remodeling gene *DDM1*. *Plant J.* 70: 750–758.
- Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Rev. Genet.* 13: 484–492.
- Loudet O, Gaudon V, Trubuil A, Daniel-Vedele F (2005) Quantitative trait loci controlling root growth and architecture in Arabidopsis thaliana confirmed by heterogeneous inbred family. *TAG* 110: 742–753.
- Simon M, Loudet O, Durand S, Berard A, Brunel D, et al. (2008) Quantitative trait loci mapping in five new large recombinant inbred line populations of Arabidopsis thaliana genotyped with consensus single-nucleotide polymorphism markers. *Genetics* 178: 2253–2264.
- Agorio A, Vera P (2007) *ARGONAUTE4* is required for resistance to *Pseudomonas syringae* in Arabidopsis. *Plant Cell* 19: 3778–3790.
- Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, et al. (2001) Requirement of *CHROMOMETHYLASE3* for maintenance of CpXpG methylation. *Science* 292: 2077–2080.
- Mathieu O, Probst AV, Paszkowski J (2005) Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. *EMBO J.* 24: 2783–2791.
- Angel A, Song J, Dean C, Howard M (2011) A Polycomb-based switch underlying quantitative epigenetic memory. *Nature* 476: 105–108.

Acknowledgments

We thank Hidetoshi Saze for providing seeds of *ibm1-1*. We thank Olivier Mathieu for the *IBM1-L* construct [17]. We thank Astrid Agorio for critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: VC DV NB OL. Performed the experiments: VC DV AD IG EK. Analyzed the data: VC DV AD IG FC EK OL NB. Wrote the paper: VC NB OL.

Figure S1

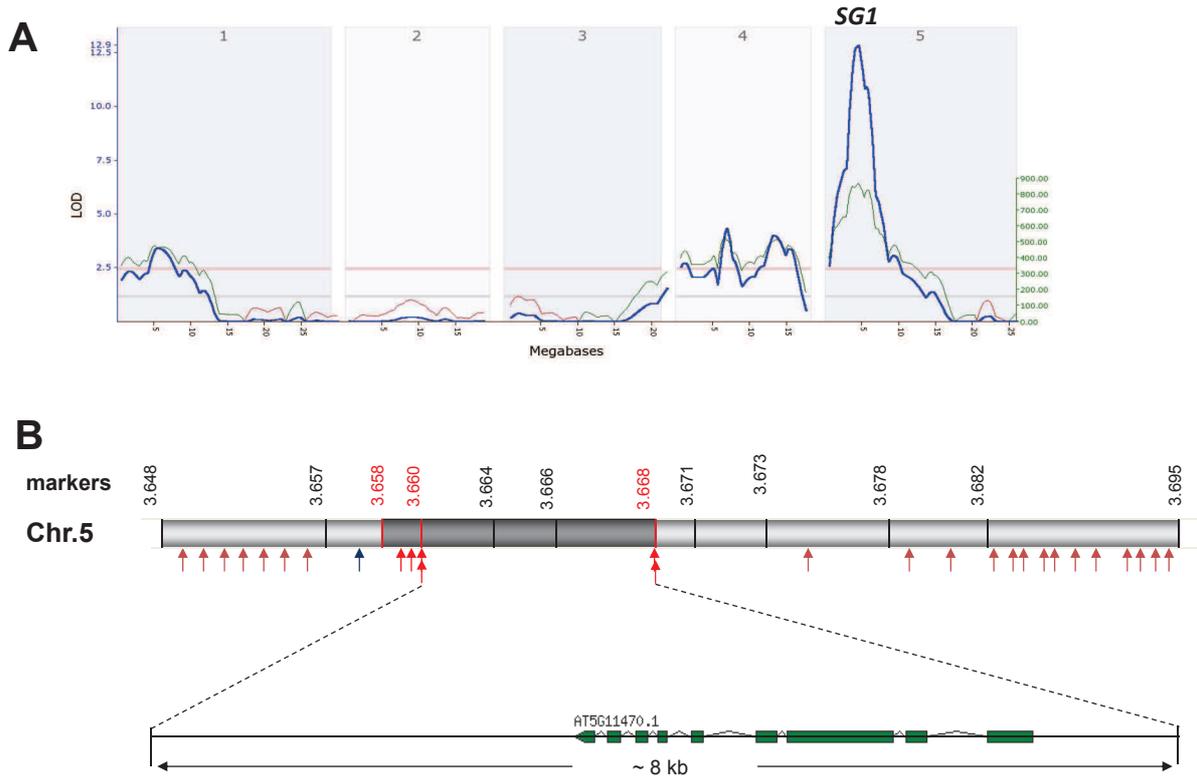


Figure S2

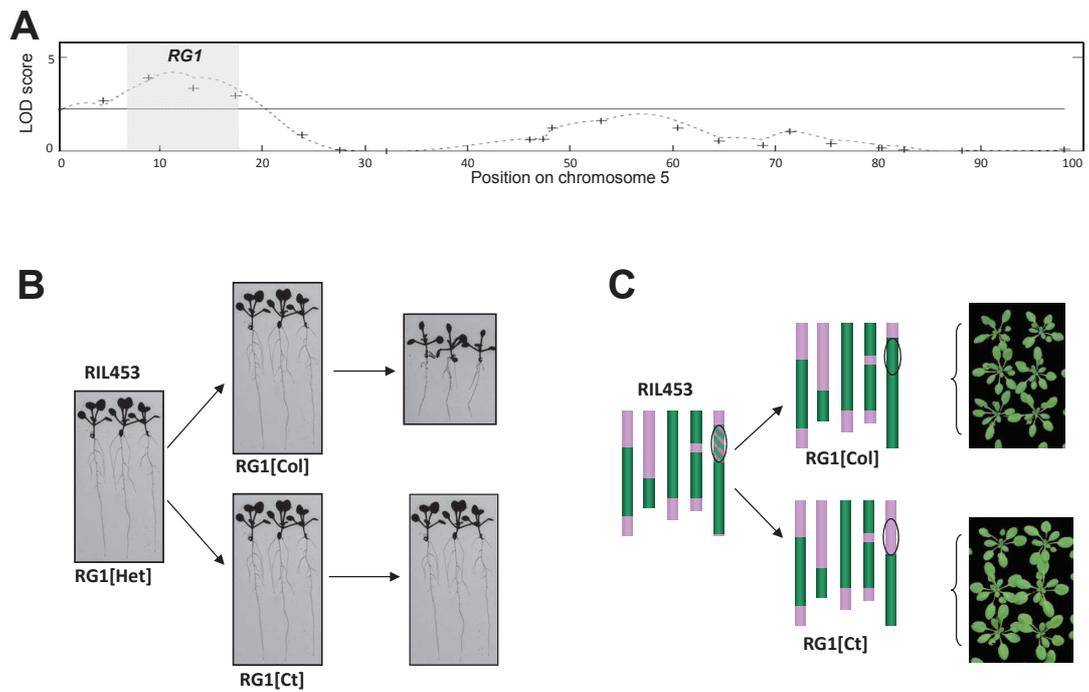


Figure S3

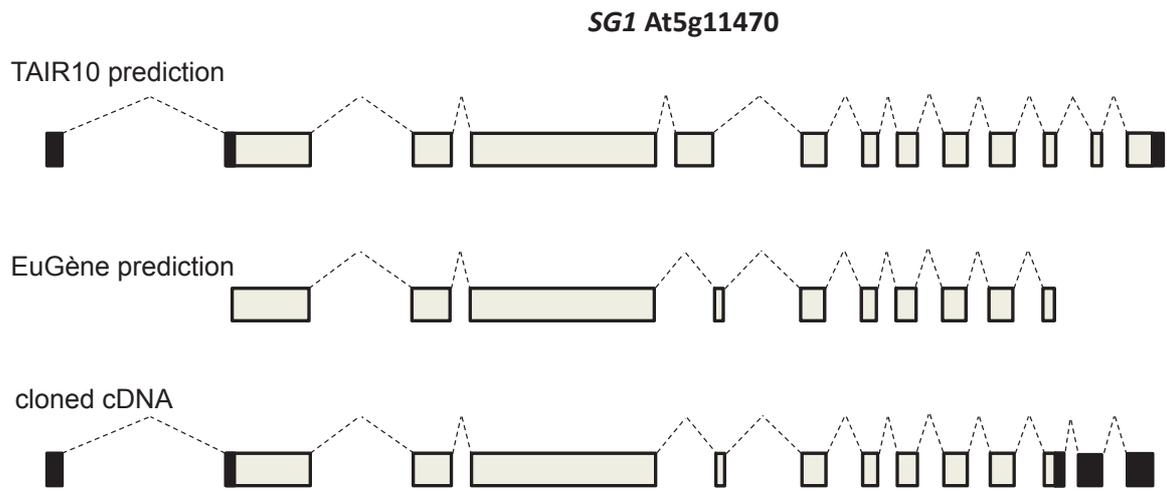


Figure S4

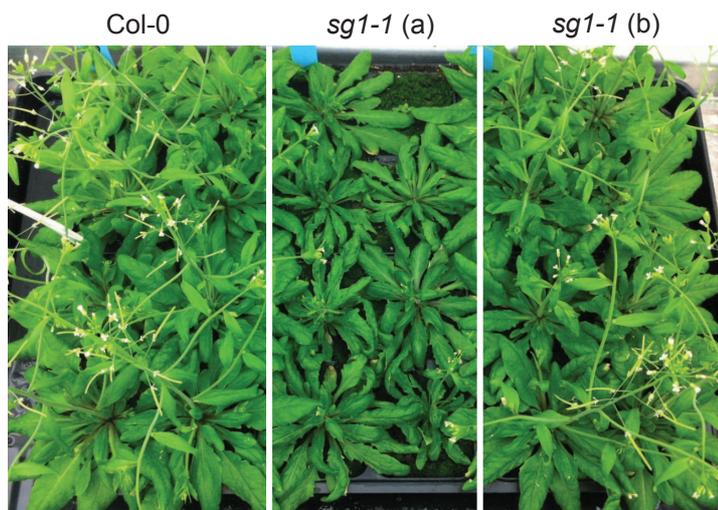


Figure S5

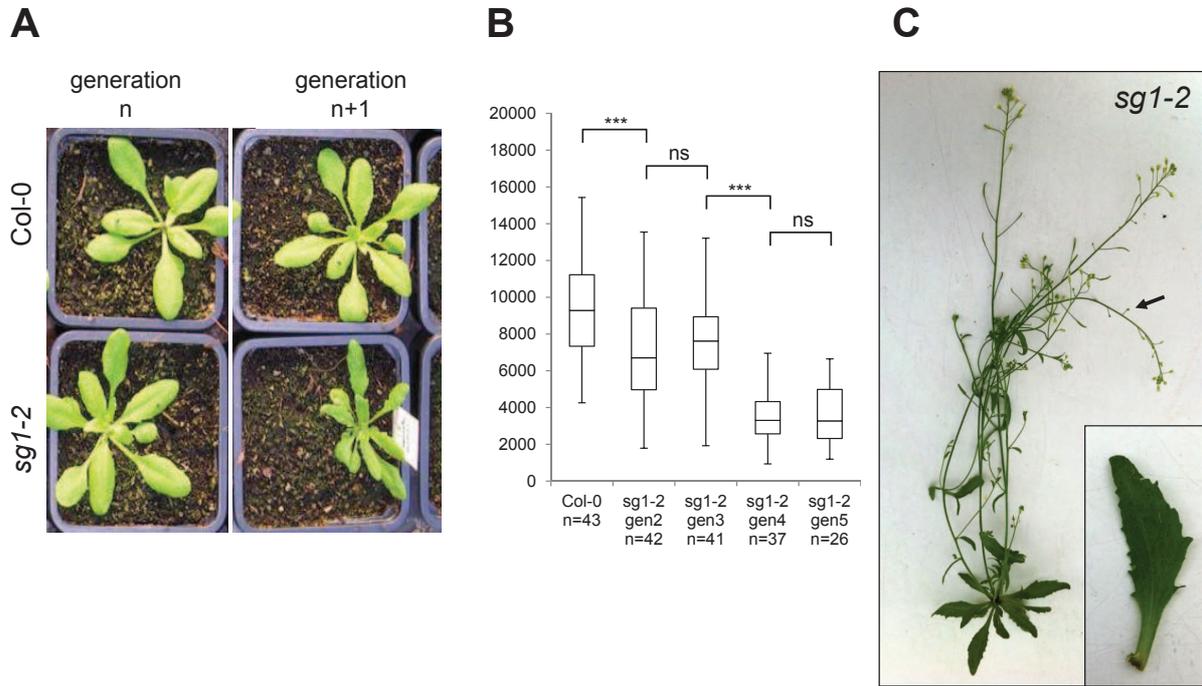


Figure S5 (continued)



Figure S6

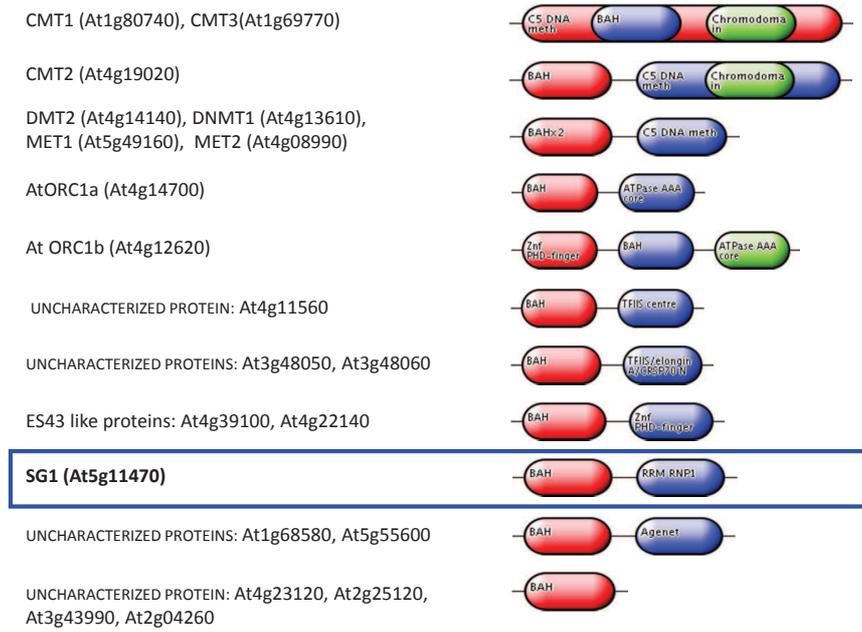


Figure S7

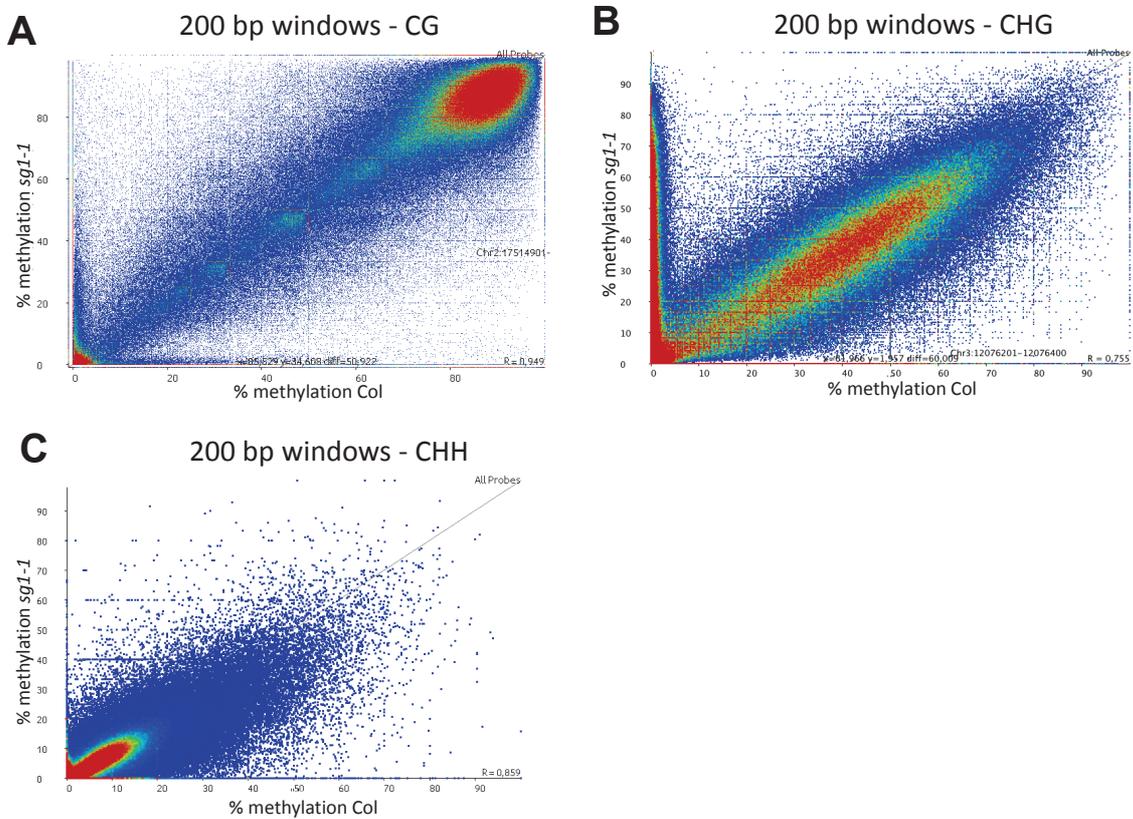


Figure S8

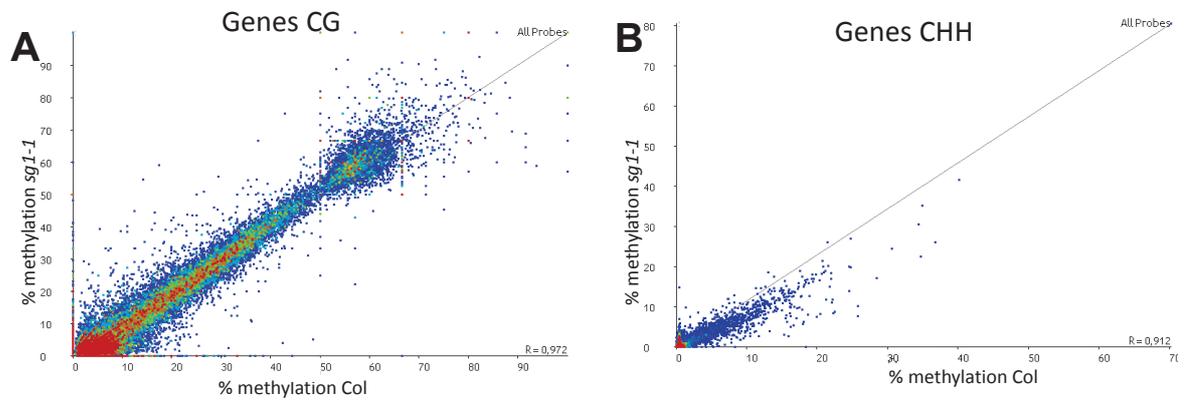


Figure S9

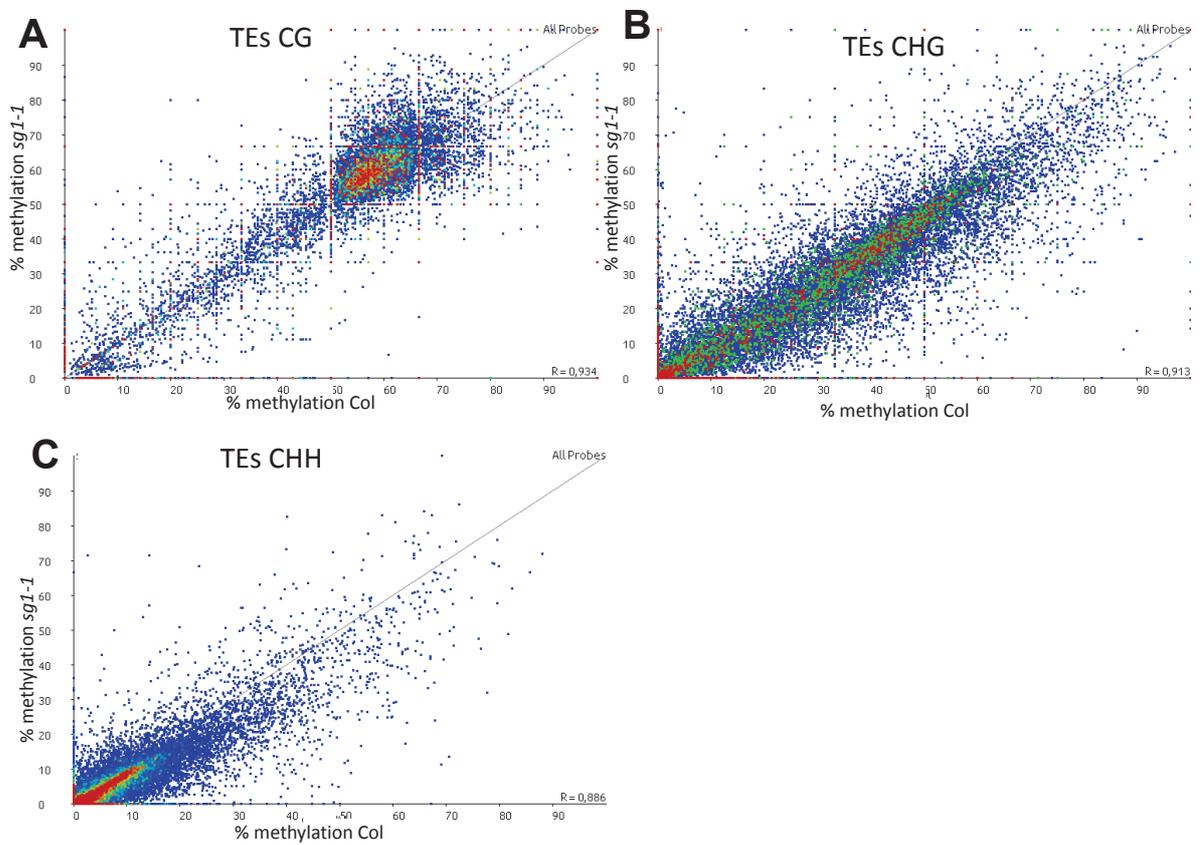


Figure S10

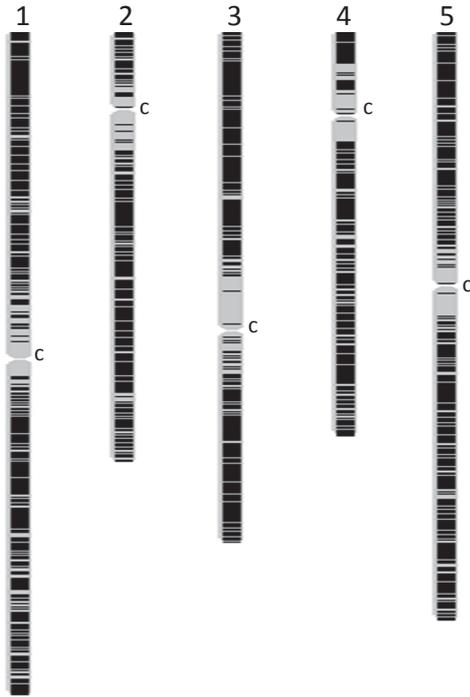


Figure S11

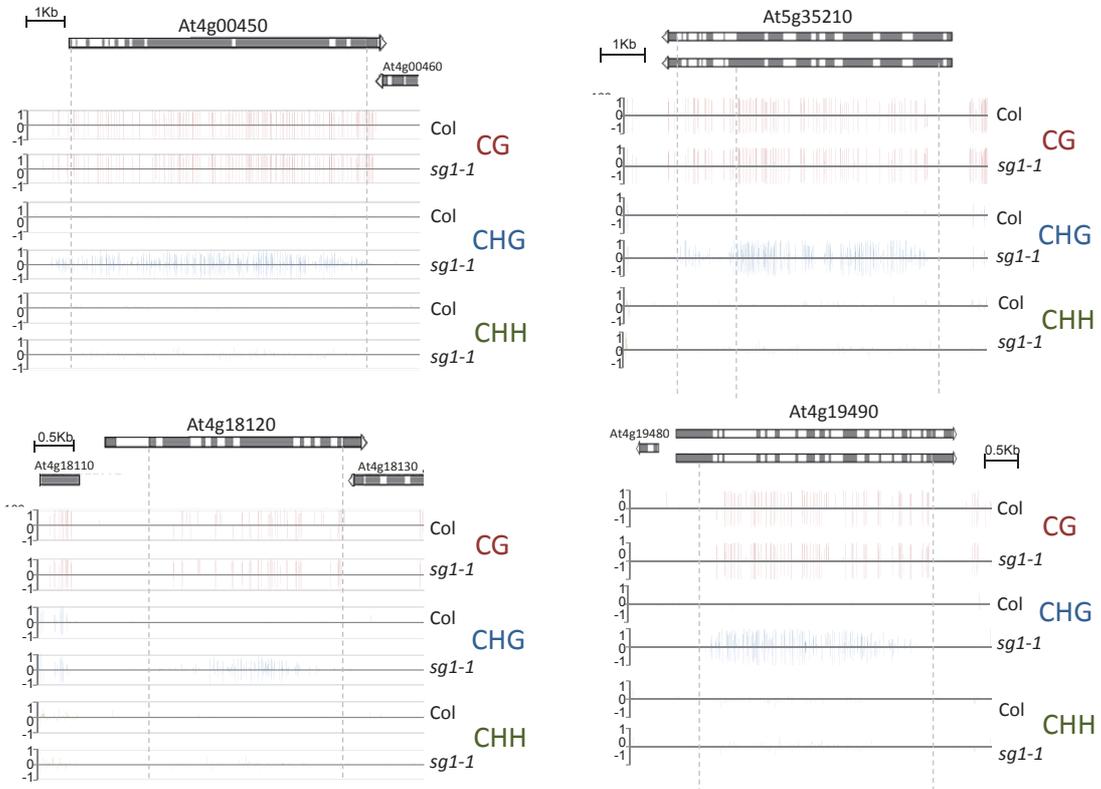


Figure S11 (continued)

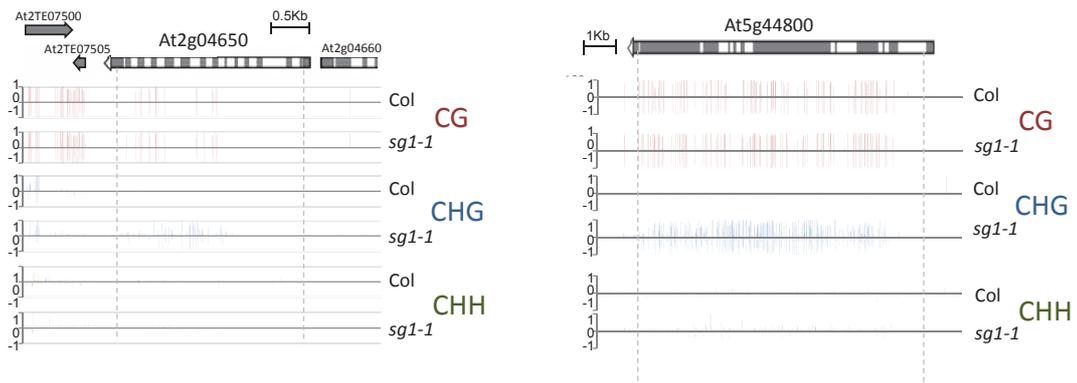


Figure S12

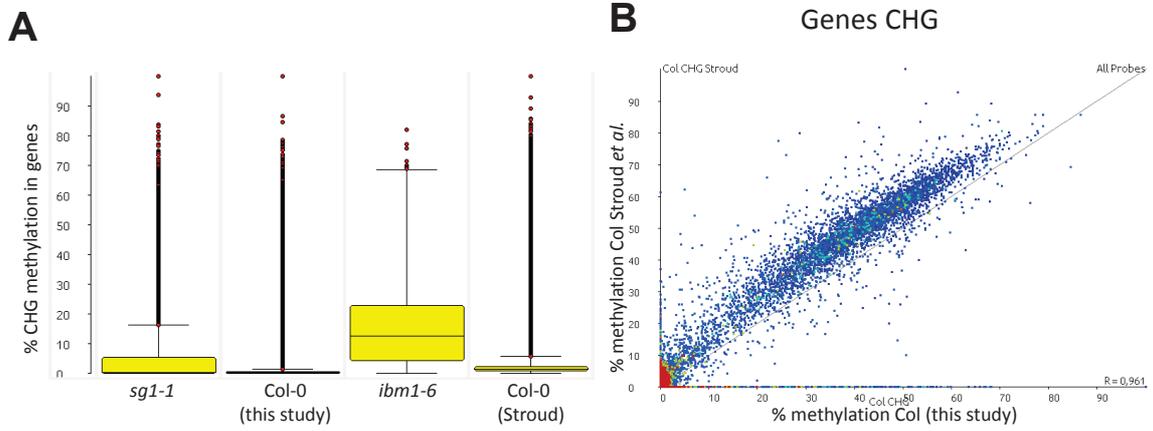


Figure S13

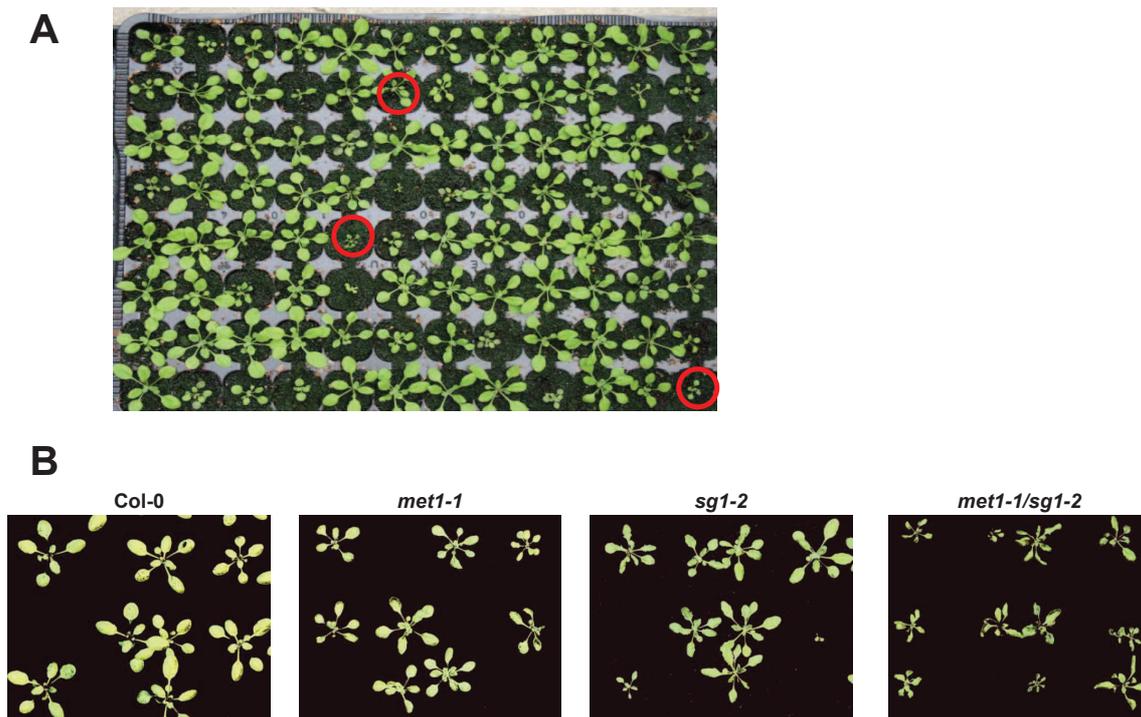


Figure S14

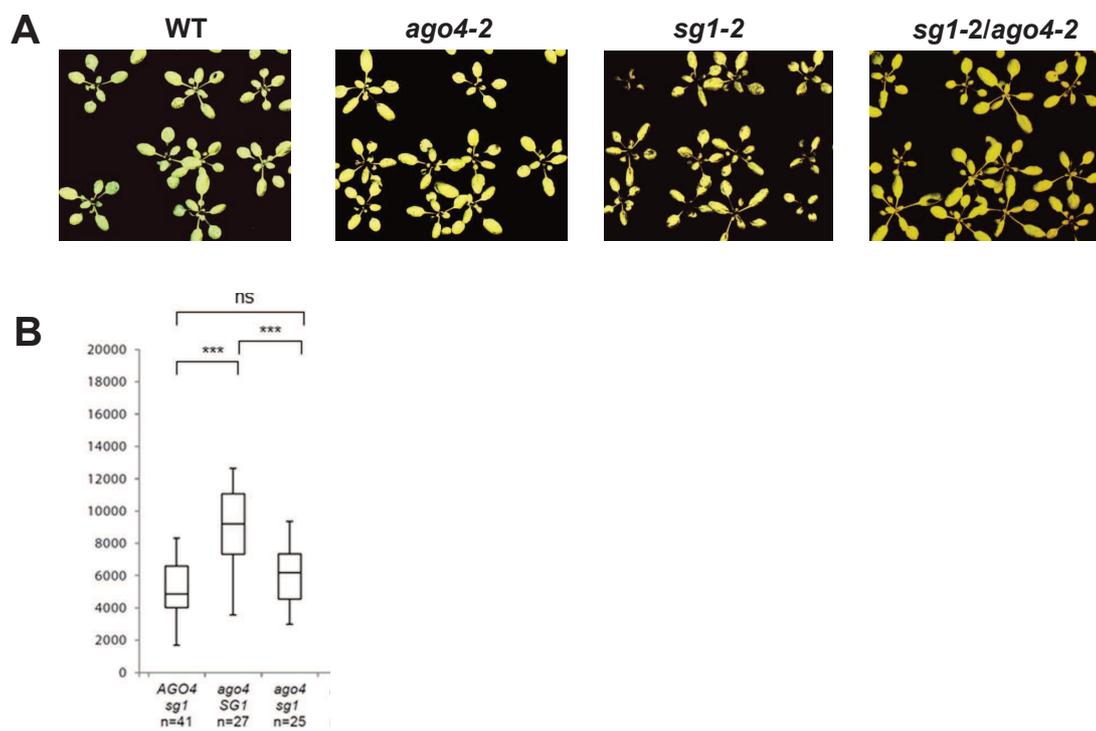


Figure S15

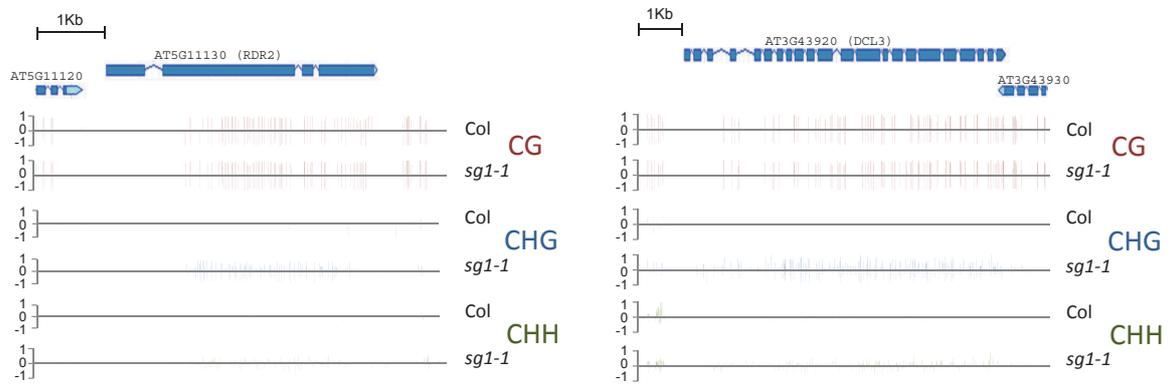


Figure S16

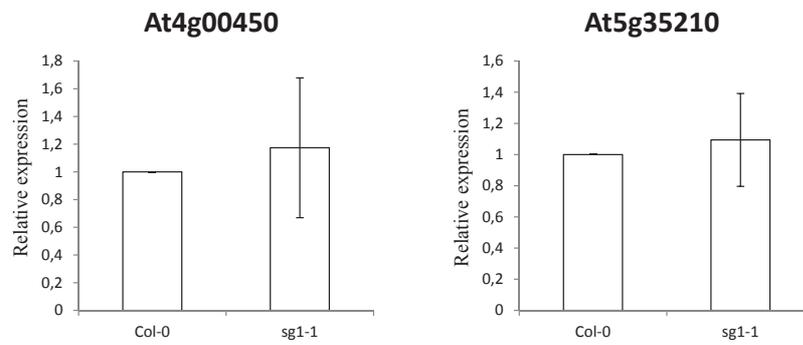


Figure S17

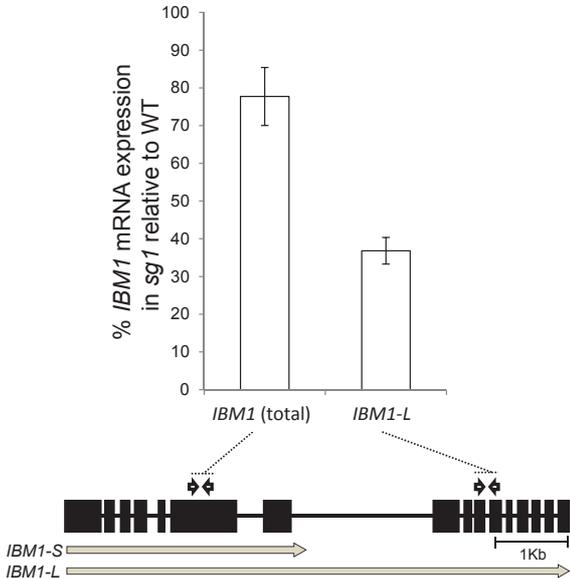


Figure S18

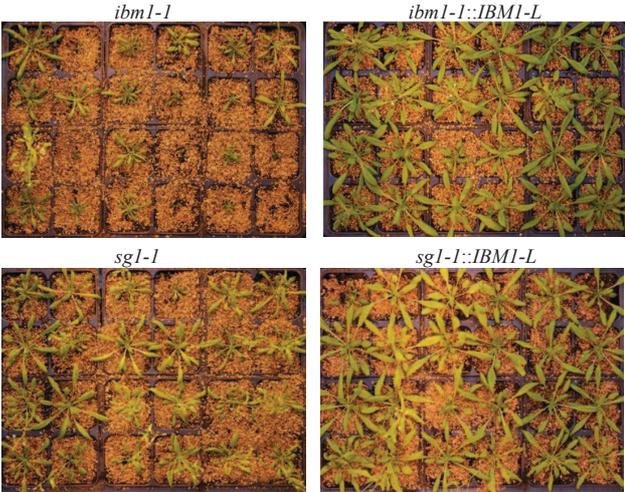


Figure S19

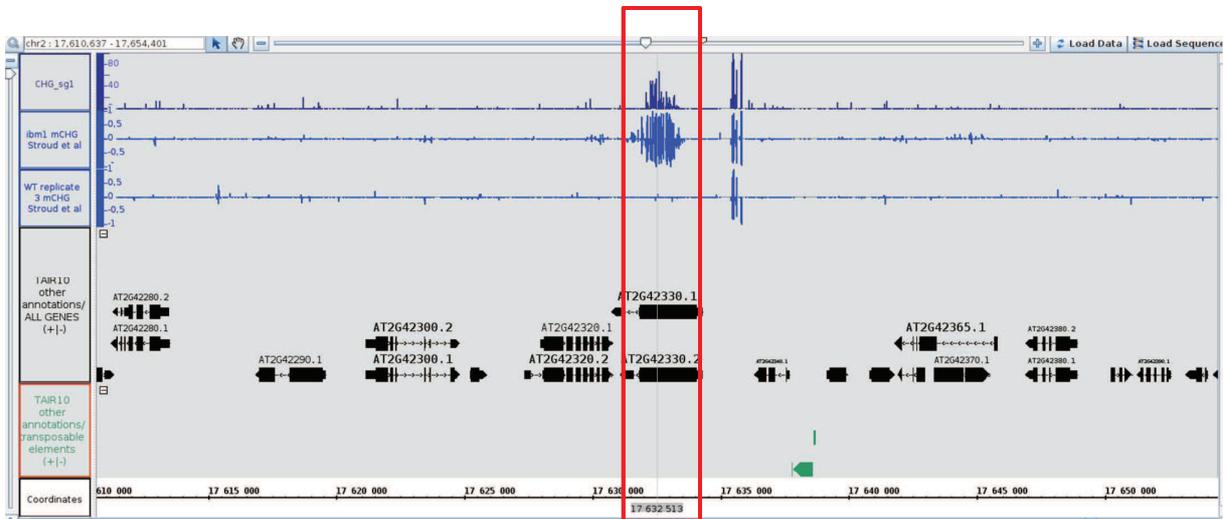


Figure S20

A

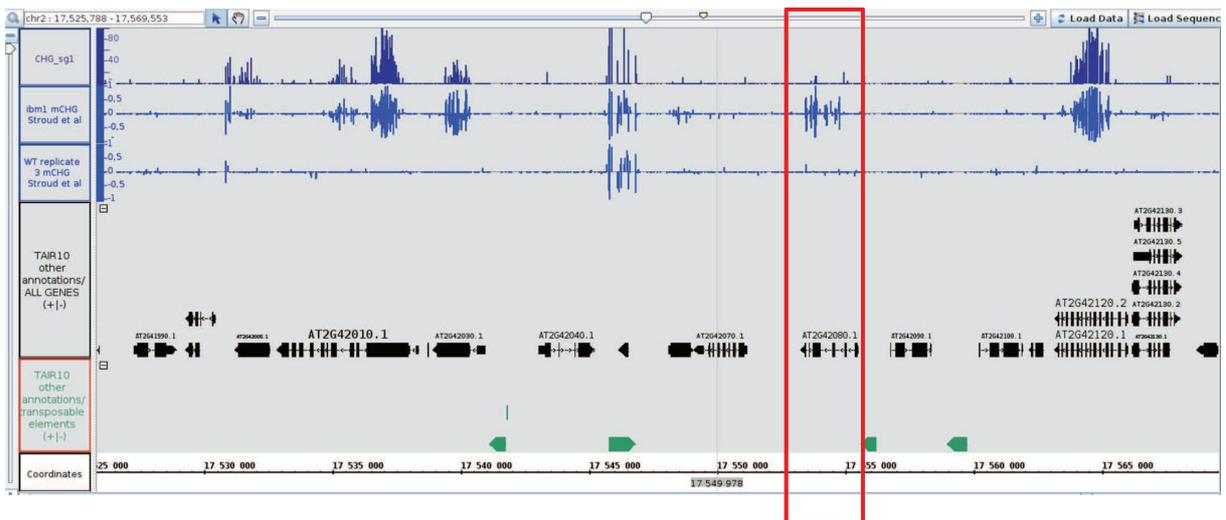
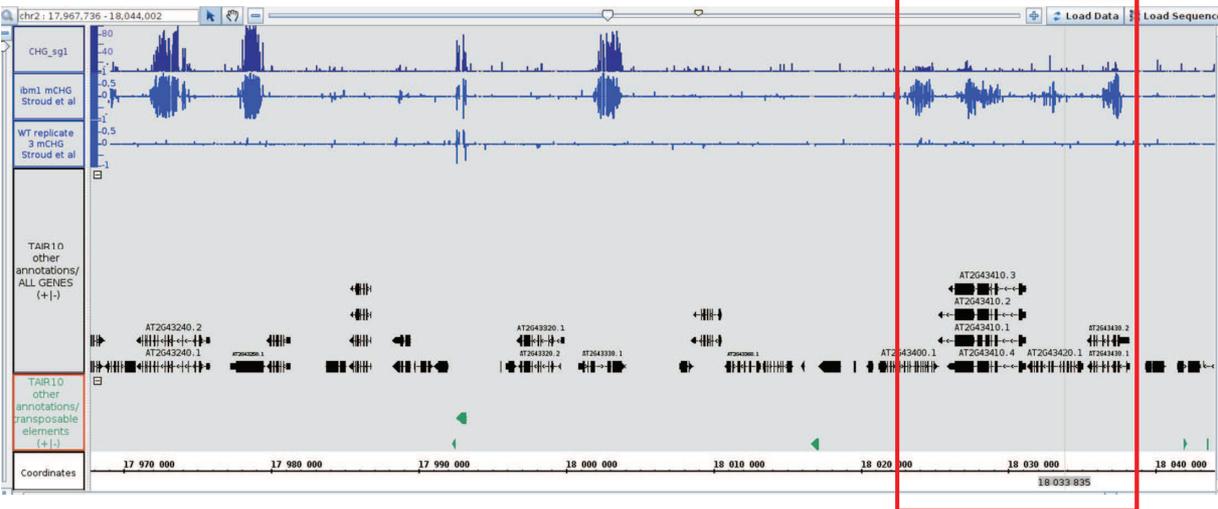


Figure S20 (continued)

B



Primers

ChIP			
gene	oligo set	FOR sequence 5'-3'	REV sequence 5'-3'
<i>BNS</i>	a	TGTCGTGAACAATCGACATAAC	ATCTCGGGATGGGTTAG
<i>BNS</i>	b	TGAGCTTTGTTTCAGAGGGTTT	AGATCACGCCAAGTTTCAGC
<i>BNS</i>	c	TTCTTTGCTTTGCTTTACAGTTTAC	AAAGTGACGTATGTCTCCATTAC
<i>BNS</i>	d	GGCGGAAGTAAGTCTTGGTATG	GAGCCACTGTTTCTTCAGTATC
<i>BNS</i>	e	GAGTACGGCTGCATTGTTCTAC	GTCGAGCAGTTGGTGAACATC
<i>BNS</i>	f	AGAAGTCGGTCGCTGTAGTG	GAAACAATGCAGCCGTAICTC
At2g04650	a	CGCCCGTTATCGTTCAATAC	GTATTGCTTGAGTCGTGAAAGTTG
At2g04650	b	TCTTTCCCTCTCAAGGCATAC	CCATATCTTCGATGTGCGTCTAC
At2g04650	c	TCGTGTTGGACTGGTGTAAAG	GATTACAATGACAGGCTTGGG
At2g04650	d	GGTAGGCCCGACATAATAC	ATTAGCCGGAACCGAAGTAG
At4g00450	a	GACTGTGGATTGGTCTTAGAG	CAACATGCCTGGAGTTTAGTG
At4g00450	b	CCGAGGCTGTGAGAACTTATTC	ATGTTTGCAGGTCCTTAC
At4g00450	c	TTTCTGGATCAGCAGCCTTAG	CCATTTCTGATCGCGTATG
At4g00450	d	TTGTTACGCCCTCTGGACTAC	GAAGCGTTGCTGTTGATAG
At4g18120	a	GCCGTTGTCCGATCATTAC	GTTTGTCTCCACGAATACAG
At4g18120	b	CCGTTCTGGCAGCAATATG	GGCCTTGAATGCTCGCATAG
At4g18120	c	CCGAGGAGCAAGTGGTAATATG	CGGCTTGGTTACCGTTACTATC
At4g18120	d	GCGAGAACAACATCAAAAGC	GATCATTGGCCAACCAGAC
At4g19490	a	TCGATCTCCGATCAAGTAG	CATGGAAGTCGGATCTGGTTAG
At4g19490	b	CGCCTTTCAGGTTAAGTTAG	TGCCCTGAATGAGCACTATG
At4g19490	c	ATTCCCTCGCTCCATTAC	CATCCGAGCCAAGAACTGTAG
At5g35210	a	CCAAGCGTTAAGGCAACATAC	GGGAATTGGACAAGGACATAAG
At5g35210	b	GAGTCGTCTGAGAGTGGAGATAAG	CACAGCCCTCTCTGGGATAC
At5g35210	c	GCCAGTTTGGCTGCTCTATC	CCGCTACAGCTGCATTTAAC
At5g35210	d	GCAACTCTGGGATATGCGTAAG	AGCATGAGCCAAGGCATAAG
At5g35210	e	TTTGACACTCTGCTGTATG	ITCGGATTGGCCCTAACTAC
At5g44800	a	TGAAGGTCTATCCGGTCTCTATC	CCGCTGCTTACTTGGTTAGAG
At5g44800	b	TTCAGAGATCAGAGATTGGTAGAG	TGCAGAAGATACTCAACATTAATACAG
At5g44800	c	AGAGCTTGGAGGAGCCTATTC	CCTTTCCCTTGTGGGATTAC
At5g44800	d	TGGCCTCGGAGATGTTTATC	GGAGCCGAGCATATCATTATC
At5g44800	e	CCATTCACCGTTTGTGATAC	ACAGCATTGTGGCTGCTACTAC
At4g11130	a	ACGAACTCTCCGGTCTTAG	GAGATTTGACTTCGAGGGTAGTG
At4g11130	b	CCGAGAAGACACTTTGAGTCTAAC	TGCGGTTTCCAAACTCATAAC
At4g11130	c	CAGAGCTACTTTGCAAGATTGATGA	CAAGAACTTAATATCTCCAGGGTGAA
At4g11130	d	AGCGAGGACACGGTAGCTTATG	CGTTAGCAGCTCCATAGTATATCATC
At3g43920	a	TTGAAAGTGTACGAGGTTGCTAAG	CAGAAGAACCATAGCTTTGATAAG
At3g43920	b	GAGAGAATTATAACTGCGAAAGTGATC	AAGAACAGGTAACCTTGCCATG
At3g43920	c	CGTAAGCGATTCTTCTGCTATGA	CAATCCATACCTGATTAAGGGGAA
At3g43920	d	TGGATTGTCTGCTTCTCCATATG	TTCTCTCAACTCTATGAGCTCATC
At3g43920	e	TGAGTTTGAAGAAGCATCTGTG	GACTCGGATACATGAGAGCAGA

Gene expression			
gene	oligo set	FOR sequence 5'-3'	REV sequence 5'-3'
At4g11130	RDR2	TGGCGAGAGATAACCGAGGTATG	CTTCTCATCGGATGGTTGGATTG
At3g43920	DCL3	GCCTACTTTCGATACCTCGGAAGA	GCATACATCACAGCCTCACGATTG
At2g13431	At2g13431	CTTTTACCAGGTGAGCAAAC	CAAGCTCTTAACGCAAATTCAGG
At3g50480	At3g50480	CTGTGGTTTGTACTAATGGTGTCTAC	TTGTATCTCACGTGTACCCCTTAC
At4g04293	At4g04293	GACGACTACTTCACCAACGCTG	CAGAATCTACGCAAGCACTCTAATG
At5g25760	UBC	CTGCGACTCAGGGAATCTCTAA	TTGTGCCATTGAATTGAACCC
At5g35210	At5g35210	TCAGCCTTACCCGGAAGTAG	TGCACCAAAGCTTCTAGTG
At4g00450	At4g00450	CCGAGGCTGTGAGAACTTATTC	ATGTTTGCAGGTCCTTAC
At3g07610	IBM1 (total)	AAGGGTAGGCCAAAGGGTAAG	CCAACCTCCGGAAGTAGTC
At3g07610	IBM1-L	GAACCTCTATCTGACGCGGTATC	ITGATGAGGGCAACTACAG

(III) Conclusion de l'article

La caractérisation du mutant *sg1*, d'une part, a révélé un phénotype développemental sévère, pléiotrope et qui s'aggrave au cours des générations. D'autre part, l'analyse du méthylome de *sg1* a permis de mettre en évidence des défauts dans les profils de méthylation en contexte CHG au niveau des corps des gènes. Nous avons pu montrer également, par une approche d'immuno-précipitation de la chromatine, que cette méthylation CHG ectopique des gènes s'accompagne de modifications de l'état de méthylation des H3K9.

Cette méthylation ectopique, à la fois en contexte CHG et H3K9me2, présente dans le corps des gènes est-elle suffisante pour éteindre leur expression ? Nous avons montré que, dans le mutant *sg1*, de nombreux gènes présentent une hyperméthylation en contexte CHG. Même si le phénotype développemental du mutant peut être sévère, il n'est pas systématiquement létal, ni stérile. En effet, les analyses d'expression réalisées sur les cibles hyperméthylées de SG1 montrent que la présence de méthylation CHG et H3K9me2 dans le corps du gène ne s'accompagne pas toujours de modification de l'expression. Par exemple, les gènes AT4G00450 et AT5G35210 ne présentent pas de modification d'expression dans *sg1* par rapport à Col-0 (**Figure S16** de (Coustham et al., 2014)) et sont pourtant lourdement méthylés en CHG (**Figure 3** de (Coustham et al., 2014)) et H3K9me2 (**Figure 5** de (Coustham et al., 2014)). A l'inverse, *DCL3* et *RDR2* qui sont aussi hyperméthylés montrent une réduction d'expression de moitié par rapport à Col-0. La méthylation en contexte CHG et la marque H3K9me2 sont connues pour être des marques répressives : elles sont absentes des gènes transcriptionnellement actifs, et sont retrouvées plutôt au niveau des éléments transposables et répétés (Cokus et al., 2008; Roudier et al., 2011), tout comme la méthylation en contexte CHH. La fonction précise de SG1 n'étant pas connue, il est également possible que la protéine SG1 soit impliquée dans d'autres processus qui permettent la prise en charge ou la maturation des transcrits.

Les similitudes que nous avons observées entre les mutants *ibm1* et *sg1*, en termes de phénotype et de méthylation, ainsi que l'absence d'aggravation du phénotype du double mutant *sg1 ibm1* en comparaison aux simples mutants, suggèrent que SG1 et IBM1 sont impliqués dans une même voie de régulation. La quantification par qPCR des transcrits *IBM1* a révélé une proportion plus faible du transcrit fonctionnel *IBM1-L* dans le mutant *sg1*

comparé à une plante sauvage. De plus, le fait que le transcrit fonctionnel *IBM1-L* soit capable de compléter le phénotype visuel du mutant *sg1*, comme cela a été montré pour *ibm1*, suggère que le phénotype développemental de *sg1* est dû à la dérégulation de *IBM1*. Pourtant, le mécanisme qui permet la régulation de la transcription de *IBM1* par SG1 reste encore inconnu.

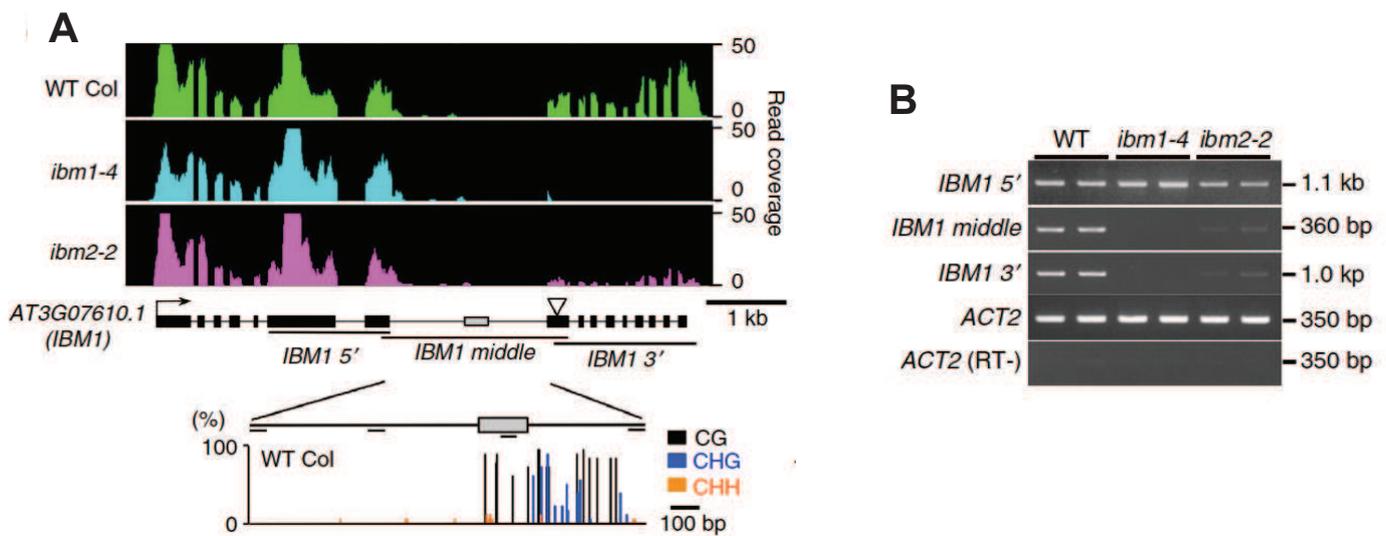


Figure III-1 : La transcription longue de *IBM1* est régulée par SG1/IBM2/ASI1.

A: Sur fond noir sont représentées les données obtenues par RNAseq pour Col, *ibm1* et *ibm2* (*sg1*). En bas, résultats de l'analyse bisulfite du septième intron de *IBM1* dans le fond sauvage Col. **B:** Résultats de RT-PCR sur trois régions de *IBM1* (indiquée sur le gène en A) pour Col, *ibm1* et *ibm2* (*sg1*) (Saze et al. 2013).

Partie III : La transcription de *IBM1* est régulée par différentes voies.

(I) Introduction

Comme cela a pu être démontré précédemment, la transcription de *IBM1* sous deux formes fonctionnellement différentes est un aspect clé qui doit être régulé sous peine d'aboutir à des conséquences développementales sévères. Ainsi, la compréhension du mécanisme qui régule la transcription de *IBM1* est essentielle, d'autant que le nombre de cibles qui sont contrôlées par ce mécanisme peut potentiellement être beaucoup plus large que seulement le gène *IBM1*. Deux autres équipes ont identifié SG1 par des approches différentes et ont ainsi permis de compléter nos résultats concernant la fonction de SG1 (Saze et al., 2013; Wang et al., 2013).

Au delà de cette régulation de *IBM1* par SG1, la première équipe a également montré que le septième intron de *IBM1*, relativement long (>2kb) par rapport à ceux rencontrés chez *Arabidopsis*, contient une région hétérochromatique qui correspond à une séquence homologue à un gène d'origine mitochondriale et chloroplastique et que celle-ci est lourdement méthylée (**Figure III-1**). Afin d'étudier le rôle de cette région hétérochromatique, le mutant *ibm1* a été complété avec des constructions portant la séquence d'ADN génomique de *IBM1*, à laquelle cette région a été enlevée. Le retrait de cette région hétérochromatique complémente le phénotype du mutant *sg1*, qu'ils ont nommé *ibm2*, en termes de méthylation en contexte CHG, de manière globale mais également au niveau de cibles de *IBM1* comme *BNS*. Ceci signifie qu'en l'absence de cette région hétérochromatique, la fonction de *IBM1* est restaurée. De plus, la production du transcrit *IBM1-L* devient indépendante de SG1 en l'absence de cette région (Saze et al., 2013). Ainsi la région hétérochromatique présente dans le long intron de *IBM1* semble être un élément clé pour permettre la transcription au delà de cet intron.

Il y a très peu de gènes chez *Arabidopsis thaliana* qui possèdent des introns de cette taille. L'étude plus approfondie de ces gènes a permis de montrer qu'une partie de ces introns longs

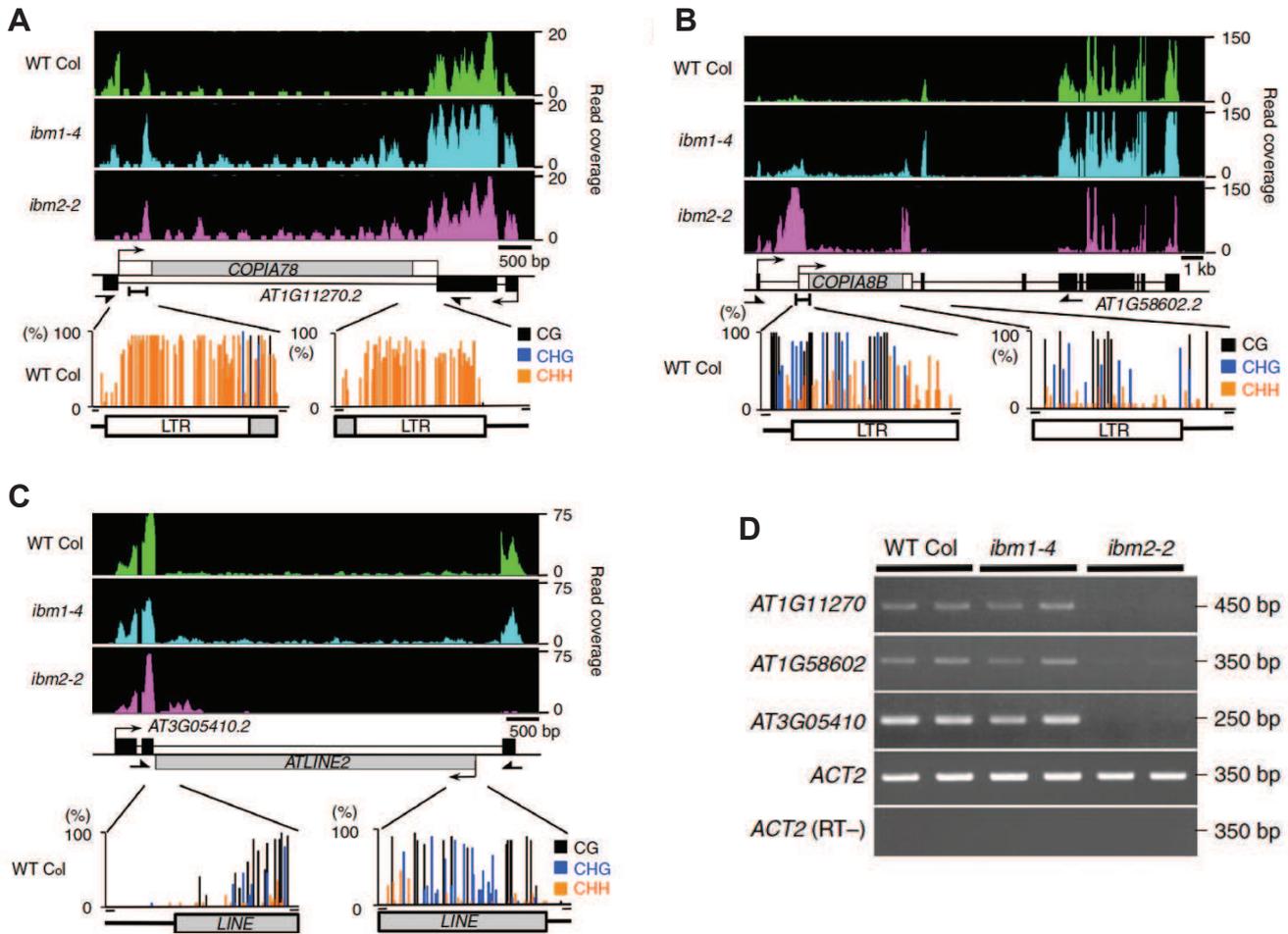


Figure III-2 : Les trois autres cibles identifiées de SG1/IBM2/ASI1.

A-C: Sur fond noir sont représentées les données obtenues par RNAseq pour les géotypes Col, *ibm1* et *ibm2* (*sg1*) aux locus AT1G11270, AT1G58602 (*RPP7*) et AT3G05410 respectivement. Pour chaque locus est indiqué, au-dessous, les résultats de l'analyse bisulfite des régions encadrant l'élément transposable dans le fond sauvage Col. **D:** Résultats de RT-PCR sur trois cibles (indiquées sur les gènes en A-C) pour Col, *ibm1* et *ibm2* (*sg1*) (Saze et al. 2013).

possède des insertions d'éléments transposables. Ainsi, il a été noté une corrélation entre la présence d'un élément transposable dans l'intron et sa méthylation d'une part, et entre la taille de l'intron et les conséquences en terme d'expression dans le mutant *sg1* d'autre part (Saze et al., 2013). Récemment, il a été mis en évidence que 3% des éléments transposables chez *Arabidopsis* sont situés dans des gènes, en particulier à l'intérieur d'introns. Ceux-ci sont généralement plus petits que les éléments transposables intergéniques, mais sont néanmoins ciblés par la méthylation, ce qui montre que les marques hétérochromatiques peuvent être maintenues au sein de gènes qui sont pourtant actifs. La présence de ces éléments transposable a un impact sur l'expression du gène hôte, mais il a été mis en évidence que le maintien des marques hétérochromatiques de ces éléments transposables est essentiel pour l'expression correcte du gène hôte (Le et al., 2015).

La seconde équipe a identifié SG1 nommé *ANTI-SILENCING1 (AS1)* au cours d'un crible suppresseur. Celui-ci repose sur la sélection sur un milieu riche en sucrose de plantes transformées avec une construction comportant un transporteur de sucrose, *SUC2*, sous la dépendance d'un promoteur fort 35S, et mutagénisées à l'EMS. La présence de ce transporteur entraîne une accumulation de sucrose dans les plantes, qui présentent alors un défaut de croissance racinaire. Ce crible s'intéresse aux plantes que la mutagenèse a rendu insensibles à la présence de sucrose dans le milieu de croissance dans le but d'identifier des facteurs qui préviennent l'extinction des gènes. Cette équipe, grâce à des analyses d'expression, ainsi que des données de méthylome et de séquençage des ARNm, a pu montrer que SG1 est capable de réguler les transcrits alternatifs qui contiennent des régions hétérochromatiques. L'hypothèse qu'ils ont formulée suggère que SG1 influencerait la transcription en favorisant l'utilisation du site de polyadénylation distal, permettant ainsi la production du transcrit dans son intégralité (Wang et al., 2013).

Ces deux études ont permis de mettre en évidence trois cibles de SG1/IBM2/AS1, autres que *IBM1*, qui sont :

- AT1G11270, codant une protéine à F-box de fonction inconnue.
- *RECOGNITION OF PERONOSPORA PARASITICA7 (RPP7 ; AT1G58602)*, locus connu pour être impliqué dans la résistance à certaines maladies.
- AT3G05410, codant une protéine de la famille PsbP (Photosystem II Reaction Center), impliquée dans le processus de photosynthèse.

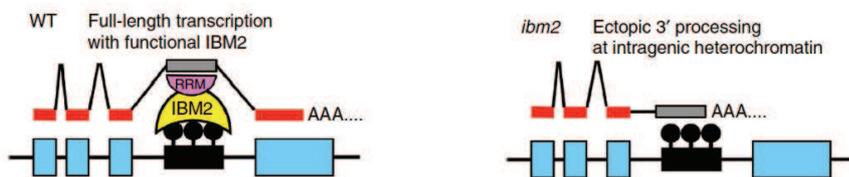


Figure III-3 : Modèle proposé par Saze et al. pour la régulation par IBM2.

Dans le sauvage (WT), IBM2 reconnaît la région hétérochromatique dans l'intron (en noir) et permet la production du transcrit (en rouge) sous sa forme longue alors que dans le mutant, l'absence de IBM2 offre le choix d'un site de polyadénylation alternatif qui génère des transcrits non fonctionnels. Les cassettes bleues représentent les exons. La cassette grise représente la partie du transcrit correspondant à la région hétérochromatique, reconnue par le domaine RRM de IBM2 (en rose) (Saze et al. 2013).

Ces trois cibles ont en commun un intron de taille importante, plus long que le septième intron de *IBM1*, et dans lequel est inséré un élément transposable. Il a été montré que, comme dans le cas de *IBM1*, la production du transcrite complet par delà l'intron contenant le TE est sous le contrôle de SG1 (**Figure III-2**) (Saze et al., 2013; Wang et al., 2013).

Les auteurs ont proposé un modèle fonctionnel qui suggère que SG1 influence le choix du site de polyadénylation au niveau de ses cibles. En effet, la protéine SG1 serait capable de reconnaître les régions hétérochromatiques présentes dans les introns de ses cibles par l'intermédiaire de son domaine BAH. Le recrutement de SG1 empêcherait l'utilisation du site de polyadénylation proximal, et faciliterait la production des transcrits longs. Ceci serait permis par l'interaction de SG1, par l'intermédiaire de son domaine RRM, avec le préARNm fonctionnel au cours de son élongation (**Figure III-3**) (Saze et al., 2013).

Le phénotype développemental caractéristique des mutants *sg1* et *ibm1* a également pu être observé dans des mutants affectés au niveau du gène *ENHANCED DOWNY MILDEW2* (*EDM2*) (Tsuchiya and Eulgem, 2013b). Le gène *EDM2* code une protéine qui possède deux domaines à doigts de zinc de type PHD (*Plant HomeoDomain*) qui suggèrent une capacité à se lier à d'autres protéines. Il a été montré que *EDM2*, tout comme SG1, est impliqué dans la production du transcrite fonctionnel de *IBM1* (Lei et al., 2014). De plus, *EDM2* est capable de reconnaître certaines modifications d'histones comme les H3K9me2 (Lei et al., 2014). *EDM2* est également connu pour réguler l'expression de *RPP7*, locus impliqué dans la résistance à certaines maladies (Eulgem et al., 2007). Le locus *RPP7* possède un intron dans son 5'UTR long de 2 175 pb qui contient un élément transposable et il a été montré que sa transcription est sous le contrôle de SG1 (Wang et al., 2013). SG1 et *EDM2* ont donc deux cibles communes, qui sont *IBM1* et *RPP7*, dont ils régulent la production des transcrits complets et fonctionnels en favorisant le choix du site de polyadénylation distal. Ceci suggère que SG1 et *EDM2* pourraient être impliqués ensemble dans un même mécanisme (Mathieu and Bouche, 2014).

Afin de mieux comprendre le ou les mécanismes qui permettent de réguler la production de transcrits longs au niveau des cibles de SG1 et de déterminer quels sont le rôle et le niveau d'implication de SG1 dans ces mécanismes, nous nous sommes intéressés à ses partenaires protéiques d'une part et à d'éventuels supprimeurs de la mutation *sg1* d'autre part.

(II) Article 2 : « Multiple pathways are involved in the processing of the functional *IBM1* transcript »

Multiple pathways are involved in the processing of the functional *IBM1* transcript

Aurélie Deremetz¹, Cécile Brousse¹, Astrid Agorio¹, Isabelle Gy¹, Eveline Van De Slijke^{2,3}, Kris Gevaert^{4,5}, Geert De Jaeger^{2,3}, and Nicolas Bouché¹

Affiliations:

¹ Institut Jean-Pierre Bourgin, UMR1318, INRA, 78000 Versailles, France

² Department of Plant Systems Biology, VIB, 9052 Ghent, Belgium

³ Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium

⁴ Department of Medical Protein Research, VIB, A. Baertsoenkaai 3, 9000 Ghent, Belgium

⁵ Department of Biochemistry, Ghent University, 9000 Ghent, Belgium

Corresponding author:

Nicolas BOUCHÉ

Tel: +33 1 30 83 31 71

E-mail: Nicolas.Bouche@versailles.inra.fr

Running title: Identification of an *ibm2* suppressor and characterization of the IBM2 protein complex.

ABSTRACT

Transposable elements (TE) are found in many higher eukaryotes genomes, including *Arabidopsis*, and are usually targeted by epigenetic marks like DNA and histone modifications to be silenced. Some genes are also containing TEs within introns but the mechanism, by which they are actively transcribed, despite the repressive marks, remains unclear. Here we show that FPA, revealed by a suppressor screen, regulates the processing of full-length transcripts of INCREASE IN BONSAI METHYLATION2 (IBM2) targets, including *IBM1*, by promoting the use of proximal polyadenylation sites. In *fpa* mutants, the transcription of *IBM1* becomes independent of the methylation found in the long intron. We also identified proteins that are involved in the IBM2 complex, including the histone methyltransferase SUVH6 and EDM2. We further demonstrate that SUVH6 partially controls the methylation of some IBM2 targets. Our results suggest that members of this complex are able to recognize chromatin modifications to influence the processing of the transcript. This study revealed an opposite regulation of *IBM1* transcript processing by IBM2 and FPA pathways, providing a model for intronic heterochromatin-containing genes transcription.

Keywords : Arabidopsis; FPA; INCREASE IN BONSAI METHYLATION (IBM1 & IBM2)

INTRODUCTION

DNA and histone methylations are epigenetic marks found in plants and animals that influence chromatin structure with a direct impact on gene function and transposon mobilization. Chromatin can be modified and remodeled in several ways. In plants, the Jumonji C (JmjC) domain-containing protein INCREASE IN BONSAI METHYLATION1 (IBM1) is a histone demethylase removing methylation on lysine 9 of histone H3 (H3K9me). The function of IBM1 is essential for plants because IBM1 prevents the deposition of these heterochromatic silencing marks at genes (1-3). H3K9me and CHG DNA methylation (where H = A, T or C) are tightly correlated. Indeed, CHG methylation is controlled by the DNA methyltransferase CHROMOMETHYLASE3 (CMT3) recruited to regions enriched in H3K9me with whom CMT3 directly binds (4,5). In a reciprocal way, H3K9me is catalysed by three histone methyltransferases, KRYPTONITE/SUVH4, SUVH5, and SUVH6 (6) and SUVH4 binds CHG-methylated cytosines through its SRA domain (7). Thus, CMT3 and SUVH4 participate in a self-reinforcing loop between DNA and histone methylation, essential for transposons and repeated sequences silencing, but deleterious to genes when IBM1 is absent. Consequently, *ibm1* mutants accumulate both H3K9me and CHG in coding regions with drastic consequences on development (2,3).

Two other categories of Arabidopsis mutants share the same developmental and molecular phenotype with *ibm1*. First, mutants of the *IBM2/AS11/SG1* gene, hereafter called *IBM2*, that we discovered with two other groups (8-10) and second, mutants of *ENHANCED DOWNY MILDEW2 (EDM2)* (11). The *IBM1* gene encodes two different transcripts but only the longest one is functional (12) and its production is controlled by both *IBM2* and *EDM2* in a way that remains to be further clarified. *IBM2* is a protein of unknown function containing a Bromo-Adjacent Homology (BAH) domain and an RNA-Recognition Motif (RRM) (8-10). *EDM2* contains several zinc-finger domains and a region similar to the active domains of certain methyltransferases (11). In addition to *IBM1*, *EDM2* and *IBM2* share another target, a disease-resistance gene, *RPP7* (9,10,13). Like *IBM1*, *RPP7* contains a heterochromatic domain within a long (>2 kb) intron, associated to H3K9me and DNA methylated in all cytosine contexts. Both *EDM2* and *IBM2* associate with these regions (9,10,13) to promote the production of the full-length functional transcript by an unknown molecular mechanism. One hypothesis is that both proteins function by promoting the use of distal polyadenylation sites over proximal ones located in large introns.

Polyadenylation is one key step of mRNA processing and the choice between alternative polyadenylation sites can largely impact the regulation of gene expression. FPA, an RNA-binding protein with three RRM domains, is involved in this process. FPA plays a major role in repressing floral transition by favoring the proximal polyadenylation site of an antisense of the *FLOWERING LOCUS C (FLC)* transcript (14,15). More than that, the role of FPA is widespread in regulating the 3'-end site choice of diverse mRNAs (16,17), including its own transcript (18,19). So far, FPA was not identified as a member of any splicing or polyadenylation complexes and therefore, the precise function of FPA and its mode of action are still unclear. *fpa* mutants were also identified in a genetic screen aimed at finding further components required for RNA-mediated chromatin silencing (20), but the role of FPA in chromatin silencing remains to be further determined. In addition, FPA was recently shown to be involved in plant defense response (21), pointing toward a more general role in stress responses, in addition to flowering.

In the present study, we identified *fpa* as a suppressor of the *ibm2* phenotype. The levels of functional *IBM1* long transcript, as well as other *IBM2* targets, are increased in an *fpa* background and in a double *fpa ibm2* mutant. We further show that *fpa* mutants are partially depleted in CHG methylated cytosine within the largest intron of *IBM1* providing evidence that mutating *fpa* has a direct impact on chromatin structure at *IBM1*. We also identified, by tandem affinity purification coupled to mass spectrometry (TAP-MS), proteins of the *IBM2* complex. We show that EDM2 and *IBM2* co-purify, together with MBD4, a protein containing putative methylated cytosine-binding domains, and RSA1, a Ca²⁺-binding protein required for tolerance to salt stress. In addition, we identified SUVH6 in the same protein complex. Consistently, genes corresponding to two of the previously identified *IBM2* targets are hypomethylated in a *suvh6* mutant background. Finally, we demonstrate that *IBM1* targets its own locus, providing evidence that a compromised *IBM1* function can impact the methylation status of *IBM1*, creating a new epiallele stable for at least three generations.

MATERIALS AND METHODS

Plant material and growth conditions

All plants are in the Col-0 background. The *ibm2-4* point mutation was previously named *sg1-1* and *ibm2-5* (SAIL_310B06) was *sg1-2* (8). The following mutants were previously described: *edm2-4* (SALK_142563; Ref#(22)), *fpa-3* (18), *ibm1-1* (3), *ibm2-2* (9), *suvh4/kyp* (SALK_069326; Ref#(23)), *suvh5-2* (SALK_074957; Ref#(6)), *suvh6-1* (SAIL_864_E08; Ref#(6)), *cmt3-11 ibm2-5* (8) and *suvh4 ibm2-5* (8). The double and triple *suvh* mutants were also described (6). The following T-DNA mutants were obtained from the Arabidopsis stock center: *rsal* (WiscDsLoxHs179_03A) and *mbd4* (SALK_042834).

For the EMS mutagenesis, seeds were incubated in water containing 0.1% ethylmethane sulfonate (EMS) for 15 hours at room temperature, washed several times with water and dried. Plants were then grown in greenhouses in long-day conditions, at 20°C.

Seeds of plants grown *in vitro* were first surface-sterilized and then sown on Gamborg B5 medium containing 1% Sucrose. Plants were cultivated in growth chamber at 21°C in long day conditions.

Whole-genome sequencing and bioinformatics analyses

ibm2-4 fpa-11 and *ibm2-4* were sequenced using the Illumina technology. Mutations were identified using the *MutDetect* pipeline described previously (24).

We retrieved the bisulfite-seq data for *suvh6* and the corresponding wild-type (repeats #2 and #3) from GEO (Ref GSE39901) (25). Data were trimmed with *Trim_Galore* (Babraham Bioinformatics) using the following options --non_directional --rrbs. Reads were then aligned to the Col-0 *Arabidopsis thaliana* TAIR10 reference genome using the Bismark software version 0.14.5 (Babraham Bioinformatics), with standard options (Bowtie2; 1 mismatch). Subsequent analyses were done with the R packages *bsseq* (26) and *DSS* using the *DSS-single* statistical method (27) to map Differently Methylated Regions (DMRs) without biological replicates. For genome-wide analyses, we defined windows of 200 bp overlapping by 50 bp intervals to calculate the percentages of methylation along the genome. Only cytosines covered by at least 5 unique reads were retained.

Bisulfite sequencing analyses

For each sample, 1 to 2 μg of genomic DNAs were extracted from leaves, using the NucleoSpin Plant II kit (Macherey-Nagel). DNAs were treated with bisulfite using the EpiTect Bisulfite Kit (Qiagen). Treated DNAs were amplified using primers listed in Table S1. PCR fragments were then cloned in pTOPO (Life Technologies) and sequenced individually. Results were analyzed with the Kismeth tool (28).

Chromatin immunoprecipitation (ChIP) and real-time quantitative PCR (qPCR) analyses

ChIP assays were performed as described (8) on aerial parts of 25 days-old seedlings using anti-histone H3 dimethyl lysine 9 antibody from Abcam (Catalog No. ab1220) and anti-histone H3 antibody from Abcam (Catalog No. ab1791). All ChIP experiments were quantified by qPCR using primers listed in Table S1. *SHOOT MERISTEMLESS (STM)* was used as the internal control for the ChIP experiments. Data are represented as the ratio of (H3K9me2 gene / H3 gene) / (H3K9me2 *STM* / H3 *STM*). Experiments were repeated in at least three biological replicates.

Gene expression analysis

Total RNA was isolated from the aerial parts of 21 days-old seedlings grown in vitro using the RNeasy Plant Mini kit (Qiagen) followed by a DNase treatment (Fermentas). RT-PCR was performed on 500 ng of total RNAs with the M-MLV reverse transcriptase (Fermentas) and cDNAs were diluted 10 times. 5 μl were used for qRT-PCRs using a CFX96 real-time PCR machine (BioRad) with a SYBR solution (Eurogentec) using the primers listed in Table S1. Expression levels were normalized against the Arabidopsis *UBC21* gene (AT5G25760). Experiments were repeated in at least three biological replicates with a minimum of two technical replicates per biological repeat. For the different targets of IBM2, qPCR were performed using oligos and conditions already described (9).

Raw RNA-seq data for Col-0 and *ibm2-2* were described previously (9). Reads were aligned with TopHat2 (29) using standard options and further analyses using SeqMonk (*Babraham Bioinformatics*). Numbers of reads were quantified for each exon. The data were normalized to calculate the read counts per million and filtered with an intensity difference test with a P-value below 0.05 compared to a group of 100 probes (*i.e.* exons).

Tandem Affinity Purification

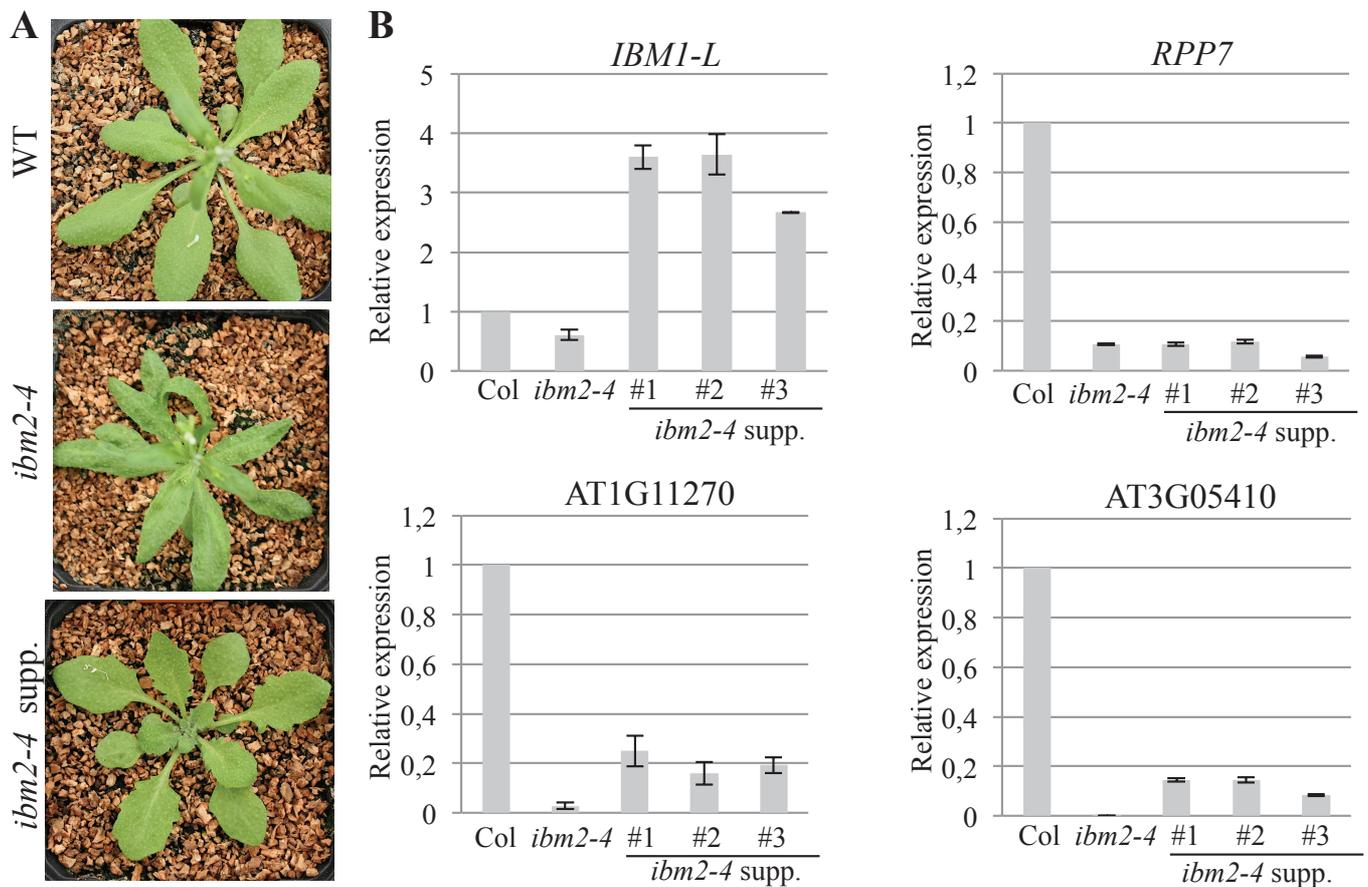


Figure 1. Phenotype of the *ibm2* suppressor

(A) Wild-type (WT; Col-0), *ibm2-4* and the *ibm2* suppressor (*ibm2-4* supp.) plants were grown in the greenhouse and pictured after 25 days. (B) Molecular phenotype of the suppressor. Expression analyses of IBM2 targets in three M3 plants (#1-3) retrieved from the *ibm2-4* suppressor screen. Primers are matching the last exons of the transcripts to amplify the long mRNA versions. Results were normalized to Col-0 (expression fixed at 1 for each experiment).

Cloning of transgenes encoding GS^{rhino} tag (30) fusions under control of the constitutive cauliflower tobacco mosaic virus 35S promoter and transformation of Arabidopsis cell suspension cultures (PSB-D) with direct selection in liquid medium were carried out as previously described (31). TAP experiments were performed with 100 mg of total protein extract as input as described in Van Leene *et al.*, 2015. Protein interactors were identified by mass spectrometry using an LTQ Orbitrap Velos mass spectrometer. Proteins with at least two matched high confident peptides were retained. Background proteins were filtered out based on frequency of occurrence of the co-purified proteins in a large dataset containing 543 TAP experiments using 115 different baits (30).

RESULTS AND DISCUSSION

fpa* suppresses the phenotype of *ibm2

To decipher new genes that could impact the function of IBM2, we did a forward genetic screen to isolate suppressors of *ibm2-4*, an allele that we previously called *sg1-1* (8). About 7,000 *ibm2-4* seeds were treated with ethyl-methyl sulfonate (EMS; Materials and Methods) and then grown in the greenhouse. We performed the genetic screen in two successive steps. First, since the phenotype of *ibm2* is related to a deficient production of the functional *IBM1* mRNA (8-10), we screened for plants with a wild-type phenotype aiming to select suppressors in which the function of *IBM1* was restored. We postulated that this pool of plants contained suppressors of *ibm2*, but also suppressors restoring only the *IBM1* function. Indeed, mutations like *cmt3* or *kyp* for instance are suppressing *ibm1* by preventing the accumulation of heterochromatic marks for a large range of *IBM1* targets (3). But these mutations are not restoring the transcription of *IBM2* targets like AT3G05410 or *RPP7* (supplementary Figure S1). Therefore, to isolate genetic suppressors of *ibm2*, we then did, in a second step, a molecular screen on these plants to determine the level of transcription of AT3G05410 (9,10) (supplementary Figure S2), a target of *IBM2* not targeted by *IBM1*. We isolated three M3 plants resembling wild-type plants with a partially restored expression of AT3G05410. We noticed that these suppressor plants were flowering later with leaves more serrated than the wild-type (Figure 1A). We determined more precisely, by qPCR, the expression level of several known *IBM2* targets in these suppressor plants. Levels of AT3G05410 and AT1G11270 mRNAs were restored to 20 % of the wild-type plant levels (Figure 1B; AT3G05410 and AT1G11270). Levels of *RPP7* (AT1G58602) were unchanged

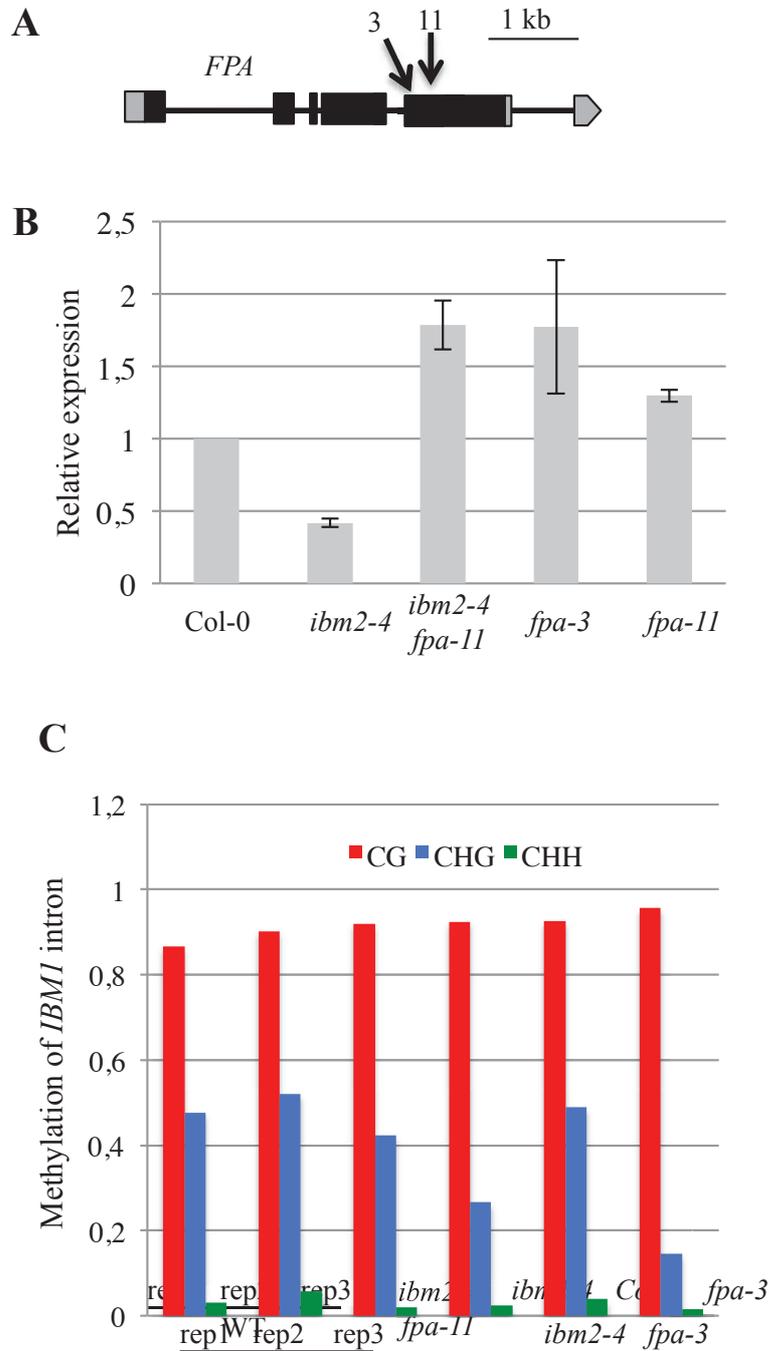


Figure 2. *fpa* suppresses the *ibm2* mutation

(A) Schematic representation of the *FPA* (AT2G43410) gene and position of the *fpa* mutants used in this work. *fpa-11* (CAA→TAA; Q586→STOP) is the new *fpa* allele isolated in our screen. *fpa-3* (18) was previously described. Exons are represented by boxes and 5'-UTRs are in grey. (B) Expression analysis of *IBM1-L* in *fpa* mutants. qPCR analysis of *IBM1-L* expression in *ibm2-4*, *ibm2-4 fpa-11*, *fpa-3* and *fpa-11* back-crossed to Col-0 two times. *IBM1-L* levels are increased by almost two times in both *ibm2-4 fpa-11* and *fpa-3*. Results were normalized to Col-0 (expression fixed at 1 for each experiment). (C) Methylation rate within the large intron of *IBM1*, in the region containing heterochromatic marks (chromosome 3, position 2,430,285 to 2,430,595). Data for the three wild-type (WT) repetitions were obtained from raw data publically available (25). Data for *ibm2-4 fpa-11* and *ibm2-4* were obtained by amplifying the same region after bisulfite conversion (Supplementary Figure S4).

(Figure 1B; *RPP7*). But interestingly, levels of the functional *IBM1* long transcript (*IBM1-L*) were increased by three times compared to wild-type (Figure 1B; *IBM1-L*). This indicates that the three plants isolated from the same M3 pool were likely progenies of a unique M2 plant carrying the suppressor mutation. Therefore, we compared their whole genomes, after sequencing, to retain mutations that were homozygous and common to the three plants, and absent from the original *ibm2-4* mutant that we also sequenced.

The sequencing revealed that the suppressor of *ibm2-4* carries a nucleotide change (C-to-T) in the fifth exon of AT2G43410 at position 586, creating a premature stop codon. AT2G43410 encodes FPA, an RNA-binding protein controlling the sites of 3'-end processing and transcription termination of various mRNAs; we therefore designated our *ibm2-4* suppressor as *fpa-11* (Figure 2A). FPA promotes proximal polyadenylation sites, and consequently, in its absence, read-through transcription leads to the transcription of many intergenic regions (16,17). Recent northern blot analyses revealed that *IBM1-L* accumulates in *fpa-7* (16), a result that we confirmed by qPCR on another allele, *fpa-3*, in addition to the one isolated in our genetic screen (Figure 2B). Altogether, our conclusion was that FPA promotes the use of the proximal polyadenylation site of *IBM1*. Accordingly, the longest *IBM1* transcript (*i.e.* the functional one) accumulates when FPA is absent and *fpa* is epistatic to *ibm2*.

FPA was previously identified in a mutant screen for genes required for the silencing of an inverted repeat (32). While no major genome-wide changes of DNA methylation patterns were observed in *fpa-7* (25), certain categories of low copy number transposons seem to be affected. For instance, the reactivations of a SINE retroelement (*AtSNI*), a DNA transposon (*AtMu1*) and an helitron transposable element (*At1TE93275*), are associated with DNA methylation changes in *fca fpa* (17,32). We thought to determine the methylation patterns within the large intron of *IBM1* in our newly identified *fpa-11* allele. We sequenced, after bisulfite conversion, the corresponding *IBM1* region in both *ibm2-4* and the double *ibm2-4 fpa-11* mutants. Compared to wild-type or to *ibm2* controls, we found that CHG methylation was reduced by almost half in *ibm2-4 fpa-11* mutants and even more drastically reduced in a single *fpa-3* mutant (Figure 2C and supplementary Figure S3). Similar CHG reduction in *fca fpa* double mutants was previously observed for the endogenous *PDS* locus silenced by an inverse repeat (32). Likewise, changes of methylation patterns for low-copy transposable elements in *fca fpa* is comparable to the ones we observed in the heterochromatic region of *IBM1*, which could be considered as an ancestral transposon. One explanation could be the constant high rate of transcription maintained in an *fpa* mutant background that might

Gene ID	Annotation	IBM2-Nterm		IBM2-Cterm		Total
		Exp#1	Exp#2	Exp#3	Exp#4	
AT5G55390	EDM2	1	1	1	1	4
AT5G11470	IBM2	1	1	1	1	4
AT2G22740	SUVH6	1	1	1	1	4
AT1G62690	unknown protein	1	1	1	1	4
AT2G03150	RSA1	1	1			2
AT1G05970	RNA-binding protein	1			1	2
AT3G63030	MBD4		1	1		2

Table 1. Proteins identified within the IBM2 complex by tandem affinity purification coupled to mass spectrometry.

Proteins identified with at least two high confident peptides and not present in the background list are retained. Identification details can be found in Supplemental Table S3.

interfere with chromatin factors involved in the deposition of CHG and/or H3K9 marks, depleting the *IBM1* intron from these marks.

cmt3 mutants are widely impaired in CHG methylation, and this also impacts the CHG methylation normally found within the intronic region of *IBM1* (12). Since *cmt3* mutants also contain lower levels of *IBM1* functional transcript, it was hypothesized that the CHG methylation is necessary to process the long *IBM1* transcript. Here we show that in the absence of FPA, the processing of this transcript becomes insensitive to the methylation found within the long intron of *IBM1*. Indeed, *fpa* mutants have low levels of CHG methylation, but their levels of long *IBM1* transcript are elevated compared to the wild-type. To understand the molecular links between the methylation level and the processing of the transcript, we thought to determine the nature of the protein complex binding the epigenetic marks localized in the intron of *IBM1*.

The IBM2 complex contains EDM2 and a stress-related protein

We performed tandem affinity purification coupled to mass spectrometry (TAP-MS) using IBM2 as bait, to identify proteins that belong to the IBM2 complex. Transgenic cultures were generated expressing IBM2 fused at its N- or C-terminus with the GS TAP tag (30). The retained list of co-purified proteins contained seven potential partners found in the same complex (Table 1).

First, we identified one protein of unknown function (AT1G62690) and a putative RNA-binding protein (AT1G05970). However, we were not able to recover any mutants publically available corresponding to these genes. We also found that the IBM2 complex contains the Methyl-CpG-Binding Domain protein 4 (Table 1; MBD4), one of the 13 Arabidopsis proteins carrying putative MBDs (33). MBD5, MBD6 and MBD7 are binding methylated CGs *in vitro* but MBD4 do not bind CG neither CHG sites (33). Our qPCR experiments show that the transcript levels of IBM2 targets are not significantly different between an *mbd4* mutant (Supplementary Figure S4) and the wild-type (Figure 3A). Therefore, the role of MBD4 within the IBM2 complex remains to be further clarified.

We also identified, among the proteins associated with IBM2, a Ca²⁺-binding protein localized in the nucleus and required for tolerance to salt stress and oxidative stress, namely SHORT ROOT IN SALT MEDIUM1 (RSA1) (34). qPCR analyses revealed that targets of IBM2 are not expressed at their wild-type levels in *rsa1* mutants (Figure 3A and

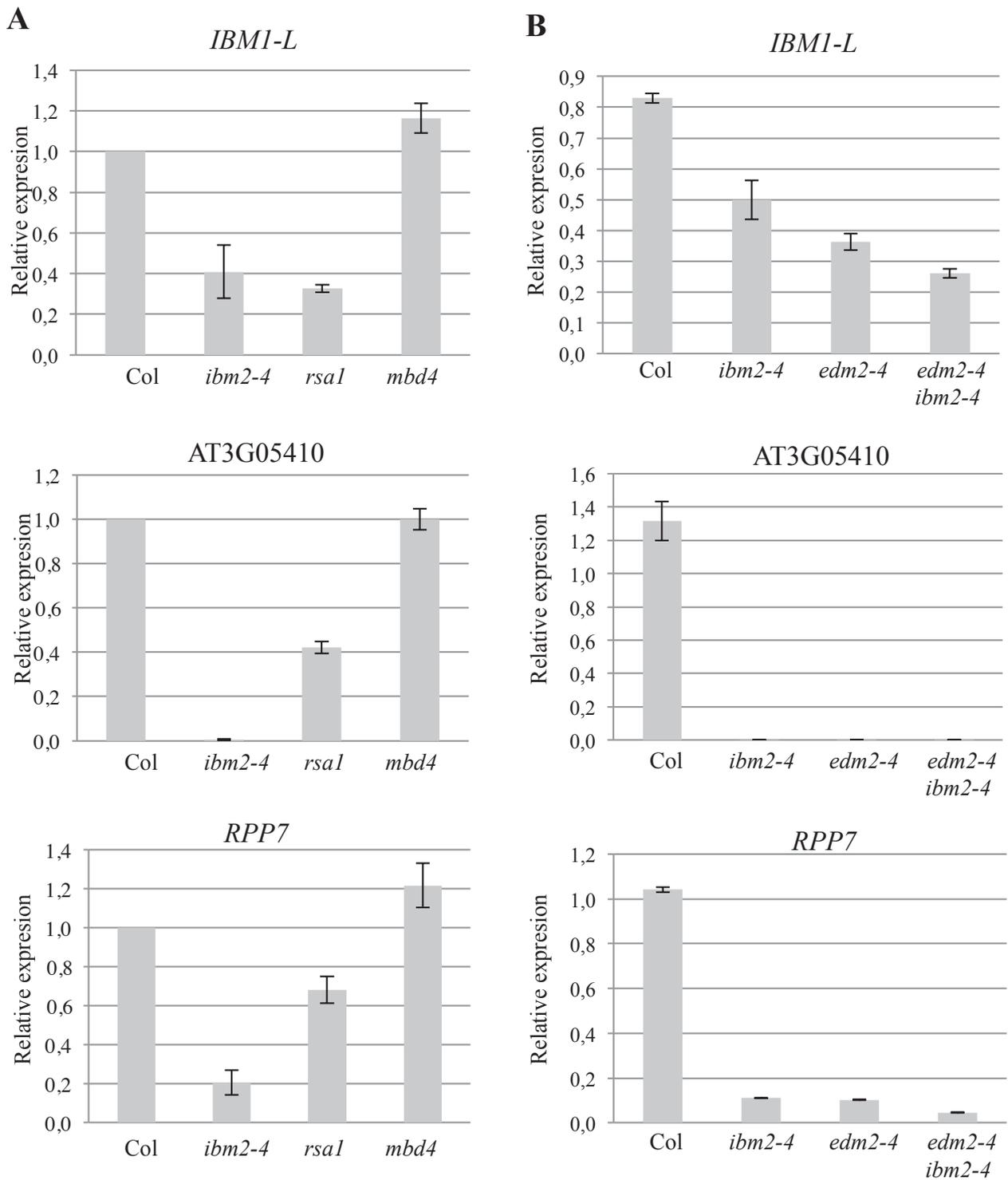


Figure 3. Expression analysis of IBM2 target in mutants of the IBM2 complex
(A) qPCR analysis of three EDM2 and IBM2 targets in *rsa1* and *mbd4*. **(B)** qPCR analysis of three EDM2 and IBM2 targets in *ibm2-4*, *edm2-4*, and the double mutant.

Supplementary Figure S4). This clearly indicates that RSA1 is involved in the transcription of IBM2 targets. The main interactor of the RSA1 protein (RITF1) identified previously in a yeast two-hybrid screen (34) was not found in our TAP experiment. This implies that the role of RSA1 within the IBM2 complex might be different from the one described previously (34). In support of this hypothesis, we note that many genes are differently expressed under standard conditions of growth in the *rsa1* mutant compared to wild-type (34), meaning that RSA1 do not play a unique role in salt tolerance. Indeed, RSA1, which contains an EF-hand domain, might decode calcium signals in the nucleus, in response to stress, and transduce these signals to different pathways, like calmodulin (CaM) for instance (35). Of particular interest in this context, we note that CaM is also involved in mRNA processing by regulating a subunit of the Cleavage and Polyadenylation Specificity Factor (CPSF) complex (AtCPSF30), in a calcium-dependent manner (36,37). Although the interaction of AtCPSF30 with CaM is not required for all processes controlled by AtCPSF30 (37), it is likely that some environmental stress linked to the Ca²⁺/CaM signaling pathway regulates the polyadenylation of many genes via AtCPSF30. This could also be the case for the IBM2 complex, through RSA1. To better understand the role of RSA1 in the processing of the *IBM1* transcript, we decided to test whether IBM1 is functional or not in an *rsa1* background. To this end, we examined the methylation state of a known target of IBM1, the *BNS* gene (3). When IBM1 is not functional (*i.e.* in *ibm1* or *ibm2* mutants), the methylation of a LINE element spreads into the adjacent *BNS* coding region. We sequenced, after bisulfite conversion, the *BNS* region of *rsa1*, and we found no changes of methylation patterns compared to the wild-type (Supplementary Figure S5). Therefore, even if the level of *IBM1* transcript is identically reduced in both *rsa1* and *ibm2* mutants, compared to the wild-type (Figure 3A), IBM1 is fully functional at *BNS* in *rsa1* only. In addition, we found that the phenotype of *rsa1* is not related to the one of *ibm1* mutants (Supplementary Figure S6). Thus, in *ibm2* mutants, the reduced level of *IBM1* transcript is not, by itself, impairing the function of IBM1. Our hypothesis is that IBM2 is promoting an essential post-transcriptional modification of the long *IBM1* transcript, to make it fully functional.

Our results also show for the first time that EDM2 and IBM2 co-purify together (Table 1). To decipher the relationships between these two proteins, we determined the level of expression of their targets in different mutant contexts. As we previously observed, the expression of *IBM1-L* was reduced by half in *ibm2-4* (8) and we show a similar reduction in *edm2-4* (Figure 3B), confirming results obtained with *edm2-6* (38). Likewise, the long *RPP7* transcript is

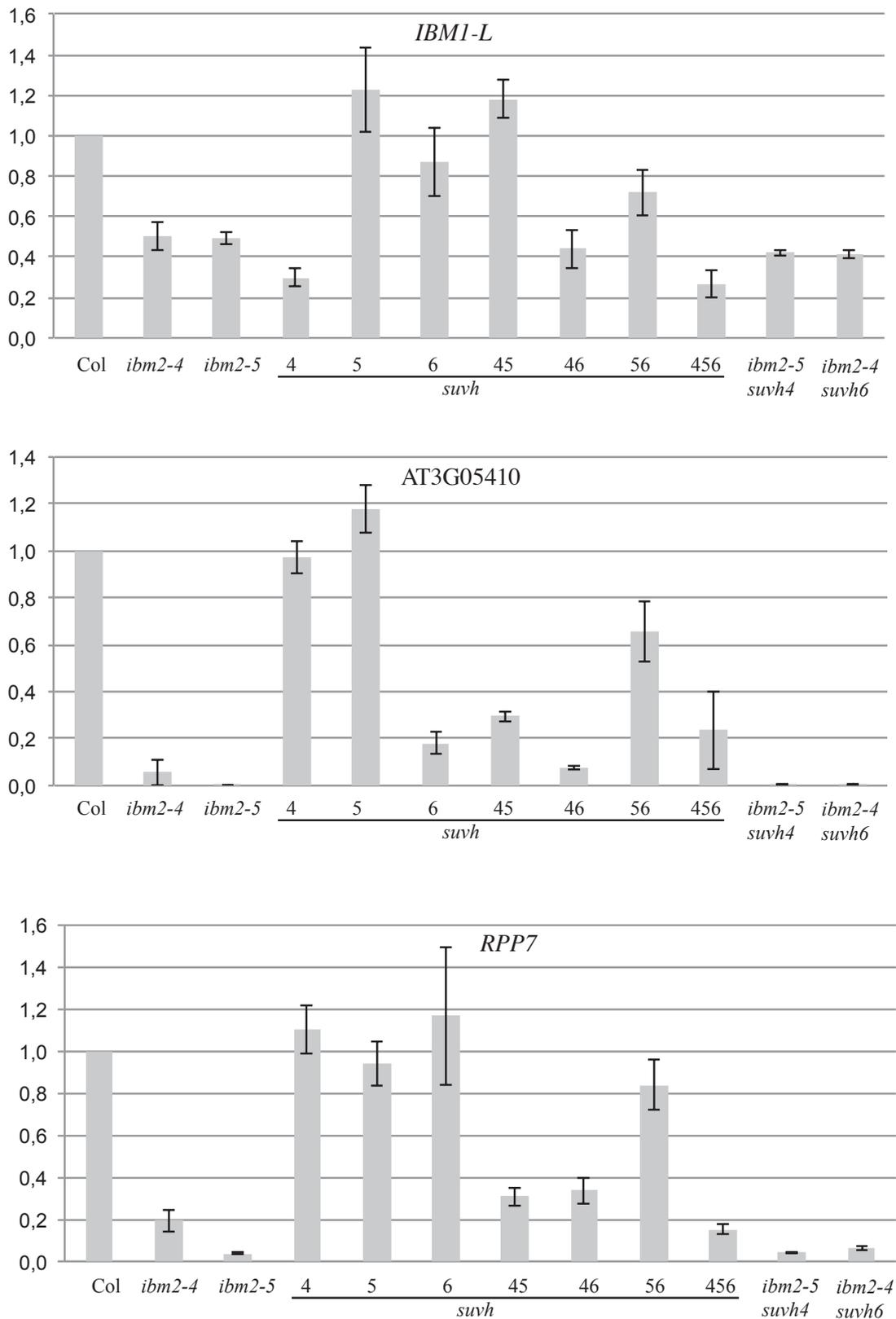


Figure 4. Expression of IBM2 targets in *svh* mutant backgrounds

Plants were grown *in vitro* for 21 days under standard long days conditions, and RNAs were extracted from bulks of 20 to 25 plants. qPCR experiments were conducted as follow: two biological repeats for *ibm2-5*, *ibm2-5 svh4*, *svh45*, 46 and 56; three biological repeats for *svh5* and *svh456*; four biological repeats for *ibm2-4 svh6*; five biological repeats for *ibm2-4*, *svh4* and *svh6*. Four technical replicates were done for each biological repeat. Results were normalized to Col-0 (expression fixed at 1 for each experiment).

absent from *ibm2-4* (Figure 3) and *edm2-4* (Figure 3B and (13)). Additionally, we further show similar results for AT3G05410 (Figure 3B), implying that IBM2 and EDM2 share an additional third target. In the double *edm2-4 ibm2-4* mutant, all targets tested have a level of expression similar to the ones observed in single *edm2-4* or *ibm2-4* mutants (Figure 3B), suggesting that none of the mutations can compensate the other ones. This indicates that IBM2 and EDM2 are both necessary for the transcription of their common targets, a result supported by the similar phenotypic alterations observed in *edm2*, *ibm1* and *ibm2* mutants, all resulting from the misregulation of *IBM1*. It remains unclear, however, whether IBM2 and/or EDM2 interact with proteins of the polyadenylation complex since none of them were identified by our TAP-MS approach.

Therefore, our results show that proteins potentially decoding the nuclear Ca^{2+} signal (RSA1), the methylation of DNA (MBD4) and histones (EDM2) are associated to IBM2 in the same complex. While we suspected that MBD4 and/or EDM2 could bind epigenetic marks localized within the large intron of *IBM1*, further evidences were needed to establish a molecular link between these marks and IBM2.

Histone methyltransferase are associated to IBM2

Among the proteins co-purified with IBM2, we also identified SUVH6, one of the three histone methyltransferases of Arabidopsis involved in histone H3K9 methylation, providing a direct additional molecular link between histone methylations and IBM2.

To dissect the functional links between SUVH proteins and IBM2, we analysed the expression of three different IBM2 targets in mutants of SUVH4 and its closest homologs, SUVH5 and SUVH6, as well as in all double or triple mutant combinations. For all targets, the complete loss of SUVH functions (*suvh456* triple mutant) strongly impairs the transcription of their full-length mRNAs (Figure 4). We noticed that *IBM1-L* is not properly expressed in *suvh4*, as previously observed (12), while the long transcript is fully expressed in *suvh5*, *suvh6* and *suvh45* (Figure 4). This last case indicates that SUVH6 can replace SUVH4 in its absence as already demonstrated for other loci (6). In the case of AT3G05410 (Figure 4), the full-length mRNA is not produced correctly in a *suvh6* mutant, contrarily to *suvh4* or *suvh5* mutants showing normal levels of transcripts. This implies that SUVH6 is the isoform responsible for the production of the AT3G05410 mRNA. Additionally, the production of the full-length transcript for *RPP7* is controlled by both SUVH4 and 6 (Figure 4). All together,

Position of exon				CG DMR		CHG DMR		CHH DMR	
Gene ID	Start	End	Diff	Diff Met	Dist (bp)	Diff Met	Dist (bp)	Diff Met	Dist (bp)
(A)									
AT1G36180	13546047	13546311	4.3	-0.35	0				
	13546515	13546897	3.1						
	13547263	13547424	4.1						
	13550721	13550908	2.0						
	13554315	13554643	2.2						
	13554739	13554976	1.6						
AT1G58400	21696165	21697870	3.9	0.35	0				
AT1G62560	23161093	23161248	2.2	-0.18	0				
AT1G64030	23753956	23754348	1.8	0.55	990	0.30	983	0.21	872
AT4G13820	8008535	8010694	1.7	-0.21	0				
AT4G16310	9220236	9220377	2.0	0.34	0				
AT5G20850	7072216	7072360	2.7	0.29	0				
AT5G38280	15293325	15294484	2.5	0.26	941				
AT5G52640	21353222	21353502	2.3	-0.22	0				
(B)									
AT1G58602 (RPP7)	21760167	21761055	2.8	-0.19	0	-0.22	0		
	21761146	21761292	4.2						
	21761385	21763765	1.9						
AT1G61500	22691624	22692881	1.8	0.27	0				
AT1G67210	25143434	25143774	1.7	-0.27	755				
	25143440	25143774	1.7						
AT2G25600	10894925	10895290	1.7	-0.21	0				
AT3G05410	1561144	1561456	6.4	-0.77	0	-0.68	0	-0.24	0
AT3G20210	7053752	7054010	2.2	0.29	0				
	7054088	7054294	2.6						
AT3G44250	15948505	15949119	2.9	0.26	0				
AT4G19520	10641003	10642285	1.5	-0.38	0				
AT5G01690	258500	258753	1.8	-0.17	0				
AT5G42250	16897172	16897450	2.3	0.24	0				
AT5G52400	21275084	21275509	1.6	0.30	0				
AT5G61680	24786666	24786904	2.4	0.18	0				

Table 2. Genes differently expressed between *ibm2* and Col-0 containing *suvh6*/Col-0 CHG DMRs

RNAseq reads were counted per exon (see Materials and Methods) and the ratio between *ibm2* and Col were calculated for each exon predicted by the TAIR10 version of the genome annotation. Differences of expression (*Diff*) below 1.5 were not considered significant. Genes associated with exons differently expressed are listed only when *suvh6*/Col DMRs were found within a maximum distance of 1kb around the coding sequence of those genes. DMR values represent the difference of cytosine methylations between *suvh6* and Col, within the genomic regions corresponding to these DMRs. Negative results indicate an hypomethylation of *suvh6* compared to Col and positive results indicate the opposite. The distance (*Dist*) between the DMR and the coding region is indicated. Targets of IBM2 previously identified (9) are in bold. The upper panel (A) shows genes that are overexpressed in *ibm2* compared to Col, and the lower panel (B) presents genes downregulated in *ibm2*.

our qPCR data show that SUVH4 and/or SUVH6 both control the processing of mRNAs targeted by IBM2. To confirm that the functions of IBM2 and SUVH4/SUVH6 are associated, we generated the double *ibm2 suvh4* (8) and *ibm2 suvh6* mutants and we determined the levels of expression of their common targets in these mutant backgrounds. In all cases, the production of the full-length mRNAs was strongly impaired (Figure 4), implying that IBM2 and SUVH4/6 proteins are necessary to process the full-length functional transcript at each target locus. This data confirm previous observations suggesting that heterochromatic marks of intronic transposable elements are essential for the transcription of their host genes (39), providing an hypothesis to sustain these results. Indeed, it is likely that IBM2 and possibly EDM2 as well, need to be associated in the same protein complex with SUVH4 or 6 to be functional.

Our qPCR data show that the transcription of AT3G05410 is controlled by SUVH6 (Figure 4), therefore, we thought to determine whether other regions targeted by SUVH6 are also targets of IBM2. Both SUVH4 and SUVH6 bind cytosine methylated in the CHG context (40), in addition to CHH for SUVH4 (5) and possibly SUVH6. Both enzymes methylate nearby histones, linking H3K9me2 to CHG DNA methylation, in a self-reinforcing feedback loop that involves the DNA methyltransferase CMT3 (40). Consequently, *suvh4* mutants have regions depleted in H3K9 methylation, but are also largely hypomethylated in the CHG cytosine context (40). We thus hypothesized that *suvh6* mutants are depleted of CHG methylation for regions corresponding to SUVH6 targets. We retrieved publically available data corresponding to the *suvh6* methylome (25) and we first calculated the percentages of methylation in 200 bp-windows (with a 50 bp overlap) to compare the whole-methylomes of *suvh6* and the corresponding wild-type controls. Genome-wide, we observed no significant differences between the wild-type and *suvh6* for all types of cytosine methylations (supplementary Figure S7), confirming previous results (25). We then identified Differently Methylated Regions (*DMRs*) between *suvh6* and the wild-type in all methylated contexts, using *DSS-single* a statistical method allowing to identify *DMRs* without biological replicates ((27); see Materials and Methods). We found a total of 220 *DMRs* (hyper- or hypomethylated) in the CHG context, 179 in the CHH and 2,380 in the CG context (supplementary Table S2). Second, using the RNA-seq data publically available (9), we calculated the number of reads for each exon in both *ibm2-2* and the corresponding wild-type. We then associated genes containing exons differently expressed with *suvh6*/wild-type *DMRs* colocalising within or around these genes. We found a very limited number of genes

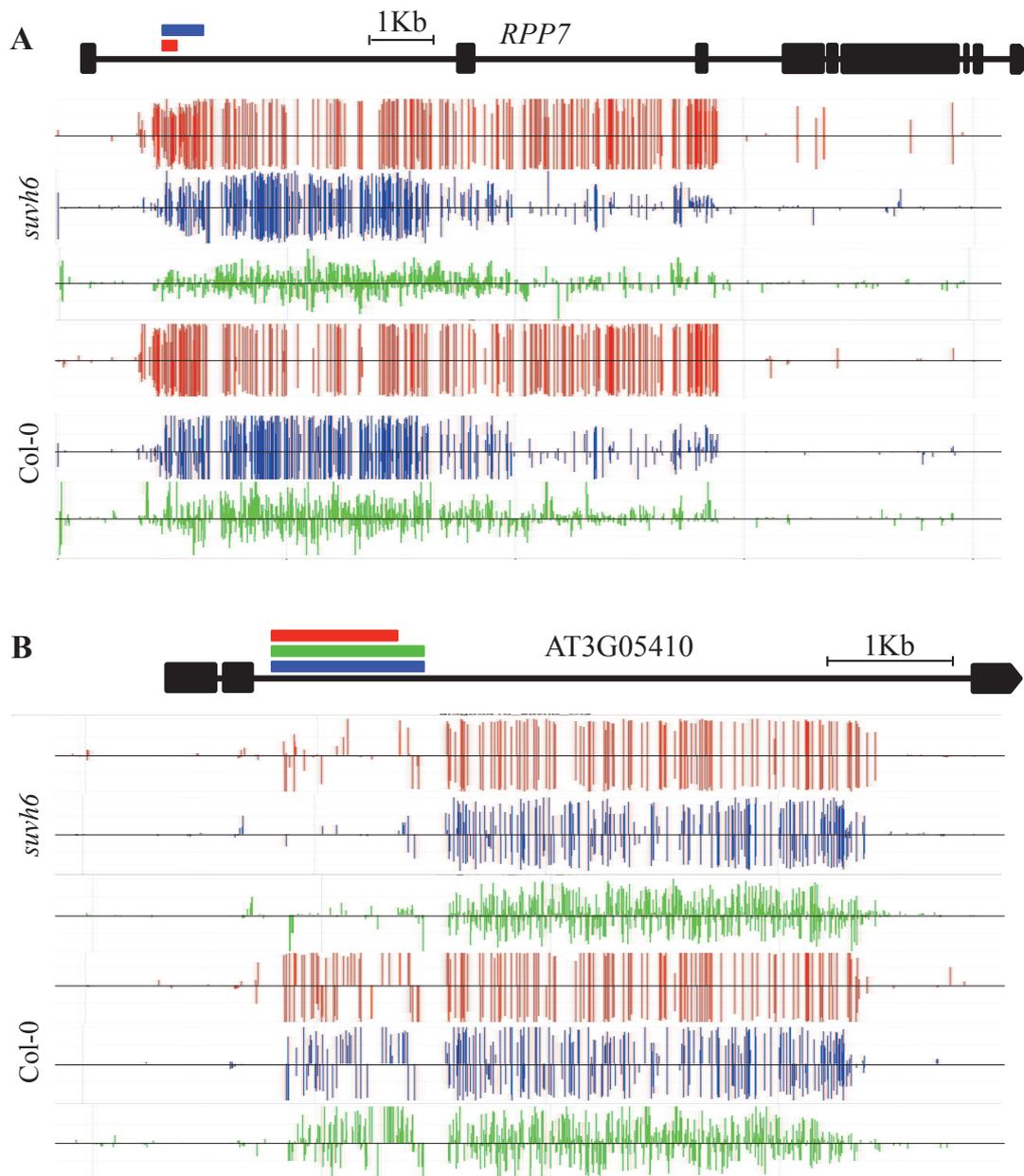


Figure 5. Methylation of IBM2 targets in *suvh6*

Methylation on top (positive values) and bottom strand (negative values) across the coding sequence of the genes indicated, namely *RPP7* (A) and *AT3G05410* (B). Data are based on methylome results (25). CG methylation is in red, CHG in blue and CHH in green. The corresponding gene models are shown according to TAIR10. DMRs (see Materials and Methods) are represented by boxes on top of the gene model.

differentially expressed in *ibm2* and containing *suvh6*/wild-type DMRs (Table 2). Interestingly, some of these genes are corresponding to IBM2 targets previously identified, further confirming the link between the two proteins. We then examined more precisely the DMRs associated with both AT3G05410 and AT1G58602. All of them are localized within the 5' region of the large introns (Figure 5). Moreover, DMRs are found in all contexts for AT3G05410, and both CHG and CG for AT1G58602. This implies that SUVH6 controls, indirectly, this types of methylations at these specific locus.

Acknowledgments

The authors wish to thank Judith Bender for providing seeds of the *svvh* mutants and Olivier Mathieu for the pENTR221-*IBM1-L* construct. We acknowledge funding from Agence Nationale de la Recherche (Project 11-JSV7-0013) and the PHC grant programs (Maimonide 2015/2016) to NB.

LEGENDS OF SUPPLEMENTARY MATERIALS

Sup Table S2. DMRs in *svvh6*

Data corresponding to the *svvh6* mutant and two wild-type repeats (repeats #2 and #3; (25)) were retrieved and aligned. DMRs were defined using the R packages *bsseq* and *DSS*, with a p-value fixed at 0.01.

Sup Table S3. Supplemental MS data

Protein identification details obtained with the LTQ Orbitrap Velos (Thermo Fisher Scientific) and Mascot Distiller software (version 2.5.0.0, Matrix Science) combined with the Mascot search engine (version 2.4.1, Matrix Science) using the Mascot Daemon interface (Matrix Science) and database TAIR10plus (30). Proteins and peptides headers used in the table are described below.

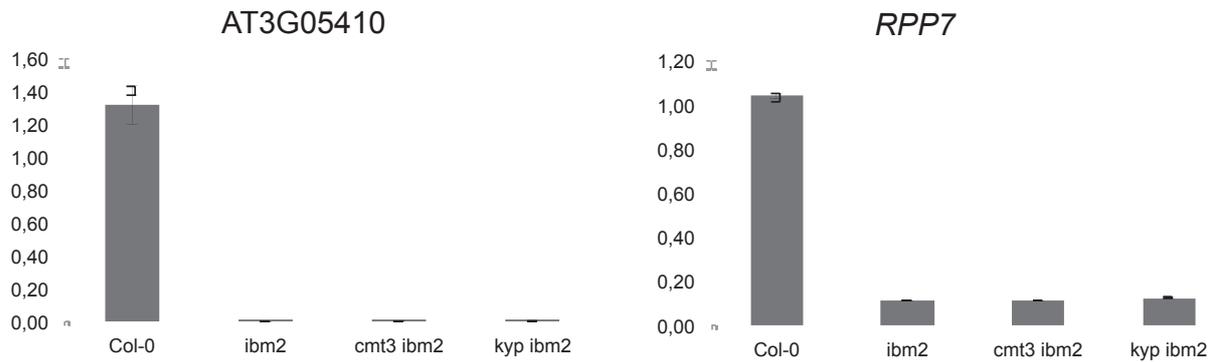
prot_score: protein score; **prot_mass**: protein mass; **prot_matches_sig**: number of assigned peptide matches above threshold (high confidence, $p < 0.01$); **prot_sequences_sig**: number of significant protein sequences above threshold (high confidence, $p < 0.01$); **prot_cover**: percentage of protein sequence covered by assigned peptide matches; **prot_len**: protein sequence length (AA); **prot_pi**: protein pI; **pep_query**: peptide query number; **pep_rank**: peptide rank number (1 for the top ranked peptide match); **pep_isbold**: peptide is in bold red. (Red and bold typefaces are used to highlight the most logical assignment of peptides to proteins. The first time a peptide match to a query appears in the report, it is shown in bold face. Whenever the top ranking peptide match appears, it is shown in red. Thus, a bold red match is the highest scoring match to a particular query listed under the highest scoring protein containing that match. This means that protein hits with many peptide matches that are both bold and red are the most likely assignments.); **pep_isunique**: peptide is unique to protein; **pep_exp_mz**: observed m/z value (precursor); **pep_exp_mr**: experimental relative molecular mass; **pep_exp_z**: observed peptide charge state; **pep_calc_mr**: calculated relative molecular mass; **pep_delta**: difference (error) between the experimental and calculated masses; **pep_start**: peptide start position in protein; **pep_end**: peptide end position in protein; **pep_miss**: number of missed enzyme cleavage sites; **pep_score**: peptide ions score; **pep_ident**: identity threshold score; **pep_expect**: expectation value for the peptide match (The number of times we would expect to obtain an equal or higher score, purely by chance. The lower this value, the more significant the result); **pep_res_before**: amino acid before peptide sequence; **pep_seq**: peptide sequence; **pep_res_after**: amino acid after peptide sequence; **pep_var_mod**: any variable modifications found in the peptide.

REFERENCES

1. Inagaki, S., Miura-Kamio, A., Nakamura, Y., Lu, F., Cui, X., Cao, X., Kimura, H., Saze, H. and Kakutani, T. (2010) Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. *EMBO J*, **29**, 3496-3506.
2. Miura, A., Nakamura, M., Inagaki, S., Kobayashi, A., Saze, H. and Kakutani, T. (2009) An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J*, **28**, 1078-1086.
3. Saze, H., Shiraishi, A., Miura, A. and Kakutani, T. (2008) Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana. *Science*, **319**, 462-465.
4. Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A., Terragni, J., Chin, H.G., Tu, A. *et al.* (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*, **151**, 167-180.
5. Du, J., Johnson, L.M., Groth, M., Feng, S., Hale, C.J., Li, S., Vashisht, A.A., Gallego-Bartolome, J., Wohlschlegel, J.A., Patel, D.J. *et al.* (2014) Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Mol Cell*, **55**, 495-504.
6. Ebbs, M.L. and Bender, J. (2006) Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. *Plant Cell*, **18**, 1166-1176.
7. Johnson, C., Bowman, L., Adai, A.T., Vance, V. and Sundaresan, V. (2007) CSRDB: a small RNA integrated database and browser resource for cereals. *Nucleic Acids Res*, **35**, D829-833.
8. Coustham, V., Vlad, D., Deremetz, A., Gy, I., Cubillos, F.A., Kerdaffrec, E., Loudet, O. and Bouché, N. (2014) SHOOT GROWTH1 maintains Arabidopsis epigenomes by regulating IBM1. *PLoS One*, **9**, e84687.
9. Saze, H., Kitayama, J., Takashima, K., Miura, S., Harukawa, Y., Ito, T. and Kakutani, T. (2013) Mechanism for full-length RNA processing of Arabidopsis genes containing intragenic heterochromatin. *Nat Commun*, **4**, 2301.
10. Wang, X., Duan, C.G., Tang, K., Wang, B., Zhang, H., Lei, M., Lu, K., Mangrauthia, S.K., Wang, P., Zhu, G. *et al.* (2013) RNA-binding protein regulates plant DNA methylation by controlling mRNA processing at the intronic heterochromatin-containing gene IBM1. *Proc Natl Acad Sci U S A*, **110**, 15467-15472.
11. Tsuchiya, T. and Eulgem, T. (2013) Mutations in EDM2 selectively affect silencing states of transposons and induce plant developmental plasticity. *Sci Rep*, **3**, 1701.
12. Rigal, M., Kevei, Z., Pelissier, T. and Mathieu, O. (2012) DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns. *Embo j*, **31**, 2981-2993.
13. Tsuchiya, T. and Eulgem, T. (2013) An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. *Proc Natl Acad Sci U S A*, **110**, E3535-3543.
14. Hornyik, C., Duc, C., Rataj, K., Terzi, L.C. and Simpson, G.G. (2010) Alternative polyadenylation of antisense RNAs and flowering time control. *Biochem Soc Trans*, **38**, 1077-1081.
15. Liu, F., Marquardt, S., Lister, C., Swiezewski, S. and Dean, C. (2010) Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. *Science*, **327**, 94-97.

16. Duc, C., Sherstnev, A., Cole, C., Barton, G.J. and Simpson, G.G. (2013) Transcription termination and chimeric RNA formation controlled by *Arabidopsis thaliana* FPA. *PLoS Genet*, **9**, e1003867.
17. Sonmez, C., Bäurle, I., Magusin, A., Dreos, R., Laubinger, S., Weigel, D. and Dean, C. (2011) RNA 3' processing functions of *Arabidopsis* FCA and FPA limit intergenic transcription. *Proc Natl Acad Sci U S A*, **108**, 8508-8513.
18. Hornyik, C., Terzi, L.C. and Simpson, G.G. (2010) The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. *Dev Cell*, **18**, 203-213.
19. Macknight, R., Duroux, M., Laurie, R., Dijkwel, P., Simpson, G. and Dean, C. (2002) Functional significance of the alternative transcript processing of the *Arabidopsis* floral promoter FCA. *Plant Cell*, **14**, 877-888.
20. Bäurle, I., Smith, L., Baulcombe, D.C. and Dean, C. (2007) Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing. *Science*, **318**, 109-112.
21. Lyons, R., Iwase, A., Gansewig, T., Sherstnev, A., Duc, C., Barton, G.J., Hanada, K., Higuchi-Takeuchi, M., Matsui, M., Sugimoto, K. *et al.* (2013) The RNA-binding protein FPA regulates flg22-triggered defense responses and transcription factor activity by alternative polyadenylation. *Sci Rep*, **3**, 2866.
22. Eulgem, T., Tsuchiya, T., Wang, X.J., Beasley, B., Cuzick, A., Tor, M., Zhu, T., McDowell, J.M., Holub, E. and Dangl, J.L. (2007) EDM2 is required for RPP7-dependent disease resistance in *Arabidopsis* and affects RPP7 transcript levels. *Plant J*, **49**, 829-839.
23. Mathieu, O., Probst, A.V. and Paszkowski, J. (2005) Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in *Arabidopsis*. *Embo j*, **24**, 2783-2791.
24. Girard, C., Crismani, W., Froger, N., Mazel, J., Lemhemdi, A., Horlow, C. and Mercier, R. (2014) FANCM-associated proteins MHF1 and MHF2, but not the other Fanconi anemia factors, limit meiotic crossovers. *Nucleic Acids Res*, **42**, 9087-9095.
25. Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V. and Jacobsen, S.E. (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell*, **152**, 352-364.
26. Hansen, K.D., Langmead, B. and Irizarry, R.A. (2012) BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol*, **13**, R83.
27. Wu, H., Xu, T., Feng, H., Chen, L., Li, B., Yao, B., Qin, Z., Jin, P. and Conneely, K.N. (2015) Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Res*.
28. Gruntman, E., Qi, Y., Slotkin, R.K., Roeder, T., Martienssen, R.A. and Sachidanandam, R. (2008), *BMC Bioinformatics*. BioMed Central Ltd, Vol. 9, pp. 371.
29. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, **14**, R36.
30. Van Leene, J., Eeckhout, D., Cannoot, B., De Winne, N., Persiau, G., Van De Slijke, E., Vercruyse, L., Dedecker, M., Verkest, A., Vandepoele, K. *et al.* (2015) An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of *Arabidopsis* protein complexes. *Nat Protoc*, **10**, 169-187.
31. Van Leene, J., Eeckhout, D., Persiau, G., Van De Slijke, E., Geerinck, J., Van Isterdael, G., Witters, E. and De Jaeger, G. (2011) Isolation of transcription factor complexes from *Arabidopsis* cell suspension cultures by tandem affinity purification. *Methods Mol Biol*, **754**, 195-218.

32. Baurle, I., Smith, L., Baulcombe, D.C. and Dean, C. (2007) Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing. *Science*, **318**, 109-112.
33. Zemach, A. and Grafi, G. (2003) Characterization of *Arabidopsis thaliana* methyl-CpG-binding domain (MBD) proteins. *Plant J*, **34**, 565-572.
34. Guan, Q., Wu, J., Yue, X., Zhang, Y. and Zhu, J. (2013) A nuclear calcium-sensing pathway is critical for gene regulation and salt stress tolerance in *Arabidopsis*. *PLoS Genet*, **9**, e1003755.
35. Bouché, N., Scharlat, A., Snedden, W., Bouchez, D. and Fromm, H. (2002) A novel family of calmodulin-binding transcription activators in multicellular organisms. *J Biol Chem*, **277**, 21851-21861.
36. Delaney, K.J., Xu, R., Zhang, J., Li, Q.Q., Yun, K.Y., Falcone, D.L. and Hunt, A.G. (2006) Calmodulin Interacts with and Regulates the RNA-Binding Activity of an *Arabidopsis* Polyadenylation Factor Subunit1[OA]. *Plant Physiol*, **140**, 1507-1521.
37. Liu, M., Xu, R., Merrill, C., Hong, L., Von Lanken, C., Hunt, A.G. and Li, Q.Q. (2014) Integration of developmental and environmental signals via a polyadenylation factor in *Arabidopsis*. *PLoS One*, **9**, e115779.
38. Lei, M., La, H., Lu, K., Wang, P., Miki, D., Ren, Z., Duan, C.G., Wang, X., Tang, K., Zeng, L. *et al.* (2013) *Arabidopsis* EDM2 promotes IBM1 distal polyadenylation and regulates genome DNA methylation patterns. *Proc Natl Acad Sci U S A*.
39. Le, T.N., Miyazaki, Y., Takuno, S. and Saze, H. (2015) Epigenetic regulation of intragenic transposable elements impacts gene transcription in *Arabidopsis thaliana*. *Nucleic Acids Res*, **43**, 3911-3921.
40. Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*, **11**, 204-220.



Sup Figure S1. Expression analysis of two IBM2 targets in *ibm2 cmt3* and *ibm2 suvh4* qPCR analysis of *RPP7* and *AT3G05410* expression in Col-0, *ibm2-4*, *cmt3-11 ibm2-5* (8) and *suvh4 ibm2-5* (8). Results were normalized to Col-0 (expression fixed at 1 for each experiment).

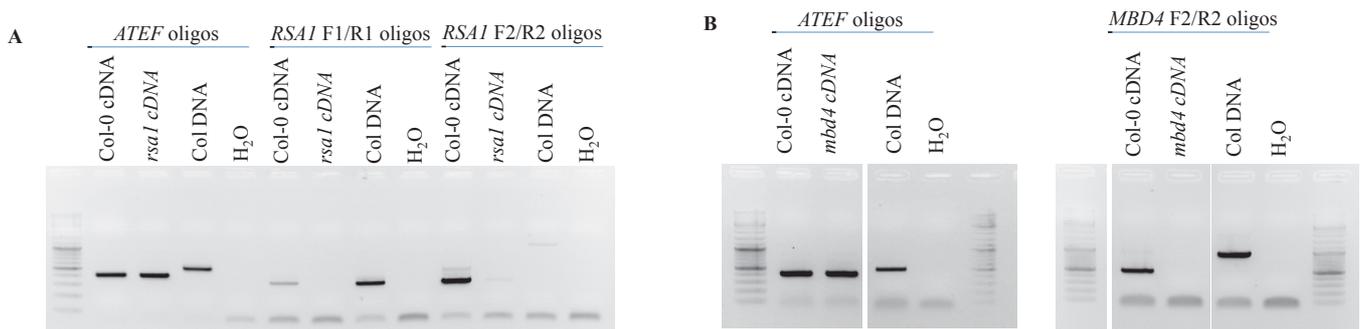


Sup Figure S2. *ibm2-4* suppressor screen

Molecular screen on bulks of suppressor plants with a wild-type phenotype. RNAs from leaves of 10 plants were extracted together and reverse transcribed to produce cDNAs. Oligos corresponding to a known IBM2 targets (A) (see supplementary Table S1) were used to screen the bulks and ATEFA cDNA amplifications (B) served as controls. Example of positive bulks are pool #3D, 7D and 10D. Amplifications on *cmt3 ibm2* and *kyp ibm2* confirm that both *cmt3* and *kyp* do not restore the molecular phenotype of *ibm2*.

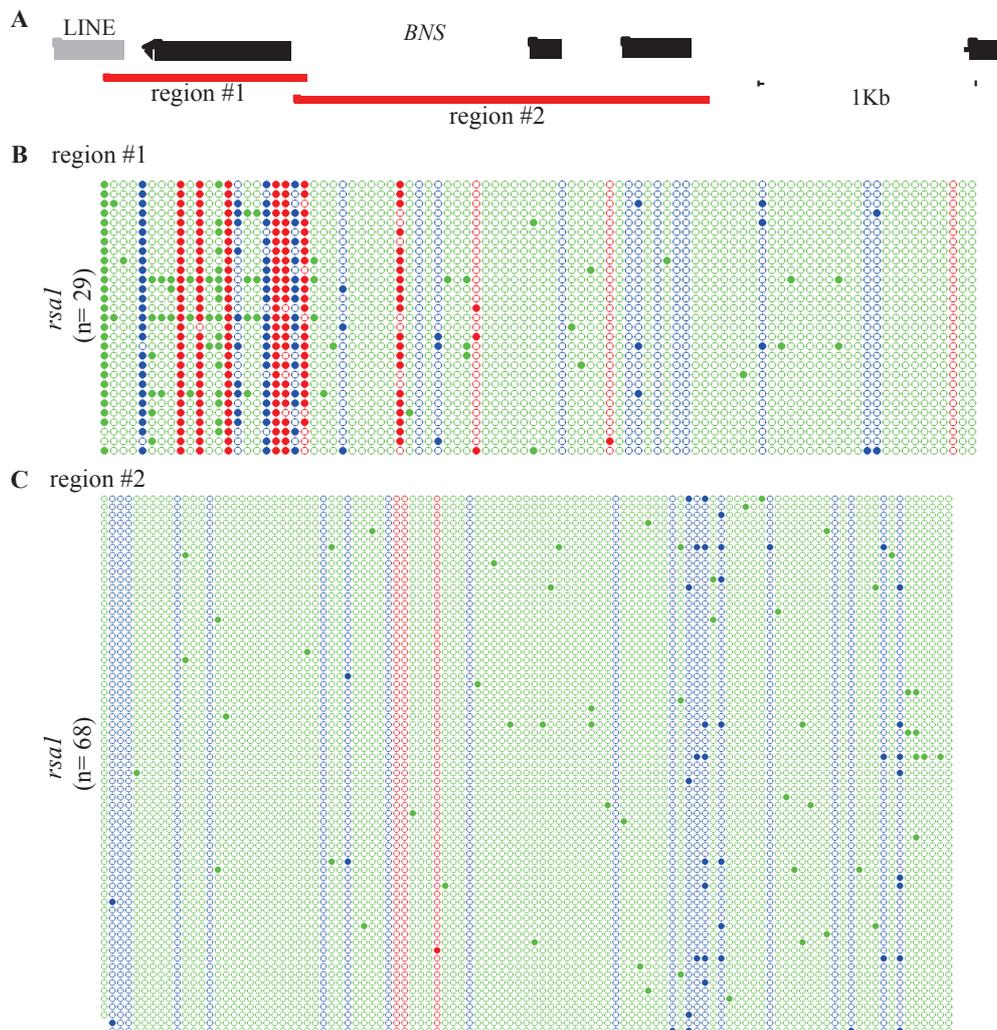


Sup Figure S3. Methylation patterns of the *IBM1* intron in *ibm2-4* and *ibm2-4 fpa-11*
(A) Schematic representation of the *IBM1* gene drawn to scale. The region (chromosome 3, position 2,430,285 to 2,430,595) corresponding to heterochromatic marks in the large intron is represented in red. **(B)** After bisulfite conversion of DNAs, the region indicated in (A) was amplified in the large intron of *IBM1* using primers described in Supplementary Table S1. The number of clones sequenced (n) is indicated. Sequences were aligned using the Kismeth tool (28) and methylation was quantified using the same tool.



Sup Figure S4. *mbd4* and *rsal* mutants molecular characterization

RNAs were extracted from bulks of plants homozygous for the DNA insertion in the *RSAl* or *MBD4* gene. *rsal* corresponds to the line WiscDsLoxHs179_03A and *mbd4* to the line SALK_042834. Both *RSAl* **(A)** and *MBD4* **(B)** were amplified (35 PCR cycles) using oligos specific of their cDNAs. Oligos *RSAl* F2/R2 were designed to amplify the cDNA only. *ATEF* cDNA amplifications (25 PCR cycles) served as controls.



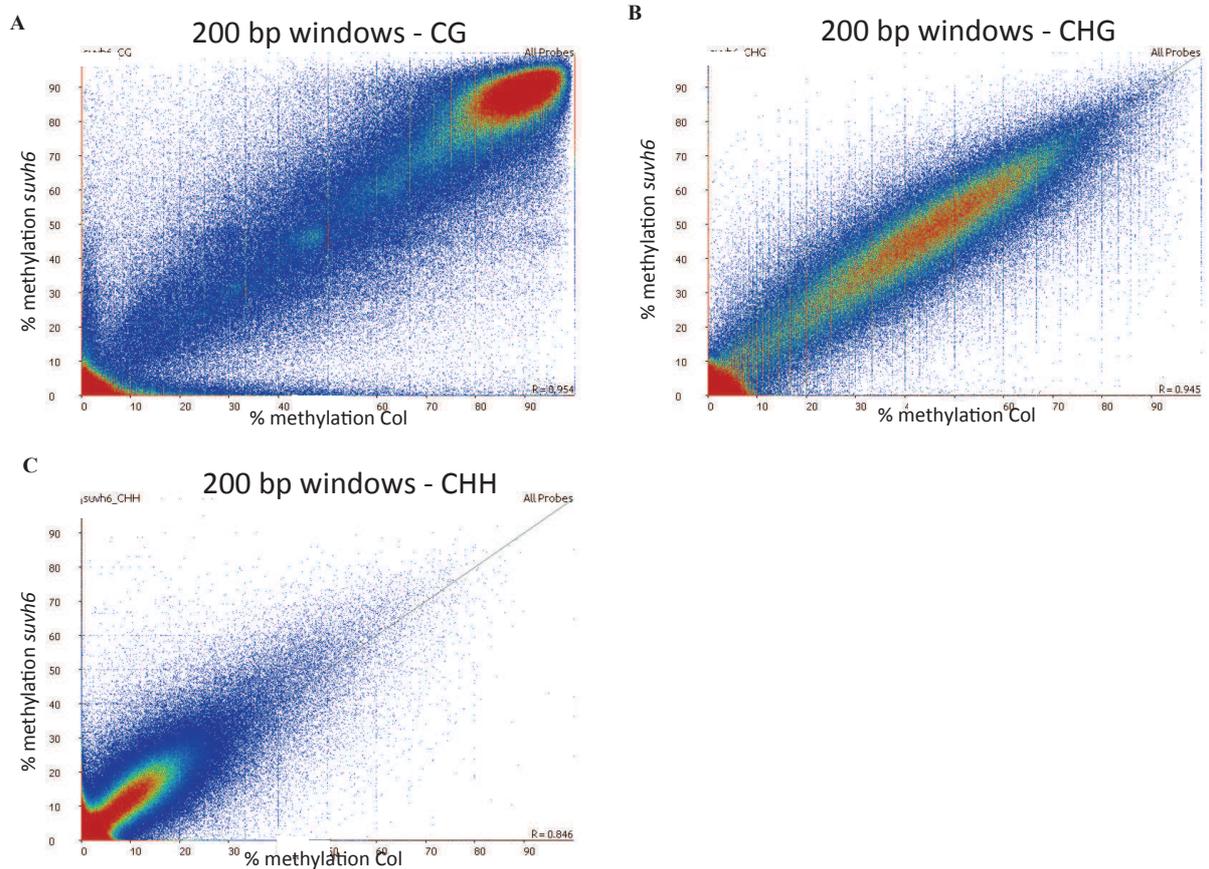
Sup Figure S5. Methylation patterns of the *BNS* gene in *rsal*

(A) Schematic representation of the *BNS* (AT1G73177) gene drawn to scale. The end of the adjacent LINE element (AT1G73175) is indicated by a grey box. Methylation of cytosines was determined in the two regions represented in red. (B) After bisulfite conversion of DNAs, the regions indicated in (A) were amplified in the *BNS* genes using primers described in Supplementary Table S1. The number of clones sequenced (n) is indicated. Sequences were aligned using the Kismeth tool (28) and methylation was quantified using the same tool.



Sup Figure S6. Phenotypes of *rsa1* and *ibm1-1*

Plants were grown 4 weeks in the greenhouse and then pictured.



Sup Figure S7. Comparison of the percentages of methylation in 200 bp-windows throughout the genome of *suvh6* and Col-0

The genome was artificially divided in 200 bp-windows, represented by a dot. Windows were designed to overlap by 50 bp and only cytosine position covered by at least 5 reads were considered in the calculation. Reads that were not matching at unique locations were discarded. The percentage of methylation in CG (A), CHG (B) and CHH (C) contexts are similar between Col-0 and *suvh6*. Raw data were retrieved from GEO at NCBI (25).

Sup Table S1. List of Primers

cDNA amplification

RSA1_R2	TCGCGAAGATAAAACCTTGC
RSA1_R1	TCTTGCCTACCTCCATCACC
RSA1_F2	GAGAACGAGAGCGTCAGAGG
RSA1_F1	AACGATGGAAAAGCAACTGG
AtEFA1	GCACTGTCATTGATGCTCC
AtEFA2	GTCAAGAGCCTCAAGGAGAG
MBD4_F2	GAGGAGGAGATCGGAAAACC
MBD4R2	CTCTTCTCTGAAGTAGTGGTAG

qPCR

IBM1_VC84	GAACTTCTATCTGACGCGGTATC
IBM1_VC85	TTGATGAGGGCAACCTACAG
AT3G05410_F	GAGCTTAATGTCAAGTGTTCG
AT3G05410_R	TGTGAGCGCAACAAACGGCTAC
AT1G58602_F	GAGTGGAGTCCTTGCGCGATC
AT1G58602_R	TGAAATAGCGTGTATTCATGGC
AT1G11270_F1	CCTTATCGGATAAATGATTAC
AT1G11270_R1	CAGCTATTTGACCTTGAAGC

bisulfite

IBM1BS_3F	AAAGTGTTTGTGATTGTTYGAGAAG
IBM1BS_3R	AATCAACTRCATCATCTCATAATC

(III) Conclusion de l'article

Le crible suppresseur de la mutation *sg1* que nous avons mis en œuvre a permis de mettre en évidence FPA, acteur déjà connu pour son implication dans la régulation du choix du site de polyadénylation de certains transcrits comme le transcrit antisens de *FLC* (Hornyk et al., 2010a; Liu et al., 2010). En effet, FPA supprime le phénotype *sg1* en favorisant l'utilisation du site de polyadénylation proximal de *IBM1*. De la même façon, FPA agit sur les autres cibles de SG1 puisque l'expression des transcrits longs de AT3G05410 et AT3G11270 est partiellement restaurée par rapport à ce qui peut être observé dans un mutant *sg1*. Ainsi, la régulation du choix du site de polyadénylation des cibles de SG1, comme *IBM1*, qui est un facteur déterminant pour la fonctionnalité du transcrit, se fait par l'intermédiaire d'au moins deux voies. D'une part, par SG1 qui régule la transcription de ses cibles en favorisant, par un mécanisme encre inconnu, l'utilisation du site de polyadénylation distal et par FPA qui favorise le site proximal.

Nous avons pu montrer que SG1 agit au sein d'un complexe, dont les différents membres suggèrent un lien entre SG1 et les marques chromatinienne présentes au niveau de ses cibles. En effet, SG1 co-précipite avec EDM2 et SUVH6, protéines connues pour reconnaître et déposer respectivement les marques H3K9me2 (Jackson et al., 2004; Lei et al., 2014). De plus, la présence dans le complexe SG1 de MBD4, membre d'une famille de protéines dont certaines sont connues pour leur capacité à reconnaître les cytosines méthylées, suggère une liaison possible à l'ADN.

Il est désormais évident que la présence de marques chromatinienne dans les introns ciblés est essentielle à la transcription correcte des cibles. Nous avons montré que SUVH6 fait partie du complexe SG1, mais l'analyse des simples et doubles mutants *suvh* suggère que SUVH4 est également impliqué, au niveau de certaines cibles.

De nombreux points restent encore à élucider, en particulier comment le complexe est-il recruté au niveau des cibles et comment est-il capable d'influencer le choix du site de polyadénylation ? Même si les résultats que nous avons obtenus suggèrent que la fonction de SG1 est en lien avec les marques hétérochromatiques trouvées au niveau des cibles et le choix du site de polyadénylation distal, aucun membre du complexe de polyadénylation n'a été co-

purifié avec SG1. La fonction précise de chacun des membres du complexe SG1 et leur rôle dans la production de transcrit longs sont encore à déterminer pour mieux comprendre le mécanisme sous-jacent, l'étendue de ses cibles et la fonction de la protéine SG1.

Partie IV : Discussion générale sur la fonction de SG1.

Les travaux réalisés au cours de ma thèse ont permis de mettre en évidence un complexe protéique composé de sept membres, dont SG1. Ce complexe intervient dans la production des transcrits longs au niveau de cibles qui possèdent des régions hétérochromatiques, comme des éléments transposables, au sein d'un intron. La production de transcrits complets et fonctionnels pour ces gènes est conditionnée par le site de polyadénylation choisi. En effet, les quatre cibles identifiées de SG1 présentent un site de polyadénylation proximal au niveau de l'intron qui contient la région hétérochromatique. Le choix de celui-ci pour la polyadénylation aboutit à la formation d'un transcrit court et non fonctionnel. Même si nos résultats suggèrent que le complexe SG1 jouerait un rôle crucial dans la régulation du choix du site de polyadénylation, le rôle précis de SG1 dans ce complexe n'a cependant pas encore été déterminé. La présence du domaine BAH dans la protéine prédite suggère que SG1 pourrait intervenir dans la modification de la chromatine en interagissant possiblement avec les marques chromatiniennes présentes au niveau des régions hétérochromatiques des introns cibles.

Les régions hétérochromatiques dans les introns des cibles de SG1 sont-elles un obstacle ou un signal pour la transcription longue ?

Le gène *IBM1* est l'une des cibles de ce complexe la mieux caractérisée. Elle a été la première à être identifiée dans la mesure où la dérégulation de la transcription de *IBM1* est responsable d'une grande majorité du phénotype visuel et moléculaire observé dans le mutant *sg1*. Comme les autres cibles de SG1 identifiées, *IBM1* possède un site de polyadénylation proximal situé dans un intron long de 1 975 pb et qui contient une région hétérochromatique d'environ 700 pb lourdement méthylée. Il a été montré que la méthylation de cette région est essentielle à la production de transcrits longs puisque des mutants qui présentent des défauts de méthylation comme *cmt3* ou *met1* s'accompagnent d'une diminution des transcrits *IBM1-L* (Rigal et al., 2012). Ce résultat laisse penser que les marques chromatiniennes présentes dans cette région forment une signature nécessaire pour permettent la transcription longue.

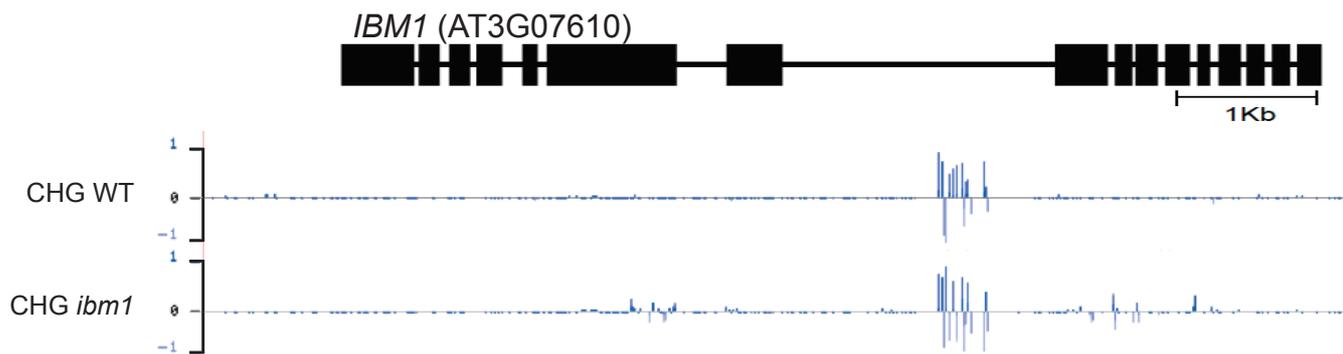


Figure IV-1 : Profil de méthylation du locus *IBM1*.

Sont représentés les profils de méthylation en contexte CHG pour les fonds Col (WT) et mutant *ibm1* au locus *IBM1*. A chaque position, la hauteur des pics est fonction du pourcentage de méthylation. D'après les données de Stroud et al. 2013.

Pourtant, la réinsertion, dans un fond mutant *ibm1*, de constructions correspondant au gène *IBMI* auquel a été retiré la région hétérochromatique de l'intron permet de restaurer le phénotype sauvage, suggérant ainsi que cette région hétérochromatique est une barrière à la transcription fonctionnelle de *IBMI* (Saze et al., 2013; Wang et al., 2013). Dans une autre hypothèse, puisque le retrait de cette région hétérochromatique réduit la taille de l'intron, il n'est alors pas exclu qu'ainsi réduit, la transcription de l'intron ne nécessite pas l'intervention d'un mécanisme permettant de favoriser la production du transcrit long. Ces régions hétérochromatiques présentes dans les introns des quatre cibles identifiées de SG1, si elles peuvent constituer une barrière à la production du transcrit long, ne constituent visiblement pas une barrière pour la machinerie de transcription. En effet, l'occupation observée de l'ARN Polymérase II, par immuno-précipitation de la chromatine, à ces locus n'est pas significativement différente dans un fond mutant *sg1* que dans le fond sauvage Col (Saze et al., 2013; Wang et al., 2013).

La perte de SG1 n'affecte pas la méthylation de l'intron long de *IBMI*, pourtant cela a un impact sur la production de transcrits longs (**Figure IV-1**). Il a été montré que SG1 se fixe davantage aux TE intragéniques qu'à ceux qui se trouvent en dehors des gènes (Saze et al., 2013; Wang et al., 2013). Cet enrichissement est observable également au niveau de l'intron de *IBMI* mais dans une moindre mesure, possiblement à cause de la taille de la région, relativement petite (Saze et al., 2013; Wang et al., 2013). Ces résultats montrent que SG1 agit en aval des marques hétérochromatiques présentes dans l'intron et qu'il serait recruté par ces marques. La méthylation de ces régions introniques pourrait alors être un signal pour le recrutement de protéines, comme SG1, pour réguler la production des transcrits alternatifs. Finalement, d'après les résultats qui ont été obtenus, ces régions hétérochromatiques présentes à l'intérieur des introns des cibles de SG1 constitueraient autant une barrière à franchir qu'un signal pour le recrutement des acteurs pour permettre la transcription au-delà de cet intron.

Quels sont les rôles respectifs des différents membres du complexe SG1 ?

Malgré l'étude et la caractérisation du mutant *sg1*, la fonction de SG1 et la manière dont ses cibles sont reconnues restent encore incertaines. L'identification de partenaires appartenant au même complexe protéique que SG1 a pu apporter quelques éléments de réponse à ces

interrogations. Parmi les protéiques impliquées dans le complexe SG1, MBD4, SUVH6 et EDM2 peuvent potentiellement être impliquée dans la reconnaissance de la signature hétérochromatique présente dans les introns.

Nous avons en effet mis en évidence que la protéine MBD4 co-précipitait avec SG1. MBD4 fait partie d'une famille décrite chez les plantes comme portant un domaine de liaison aux cytosines méthylées (Zemach and Grafi, 2007). Chez les mammifères, la fonction des protéines de la famille des MBD est variée, allant de la reconnaissance des cytosines méthylées à la réparation des mésappariements dus à la méthylation des cytosines, en passant par le recrutement d'autres facteurs, en particulier des protéines de modifications de la chromatine comme des histones méthyltransférases (Bogdanovic and Veenstra, 2009). Chez *Arabidopsis*, parmi les treize protéines MBD identifiées, trois seulement sont capables reconnaître spécifiquement les cytosines méthylés *in vitro*. MBD4 n'a cependant pas montré de capacité à lier davantage les cytosines méthylées que non méthylées (Zemach and Grafi, 2003). Pourtant, il est possible que MBD4 soit capable de reconnaître les cytosines méthylées à des locus spécifiques, dans les conditions *in vivo*. En effet, il n'est pas exclu que MBD4 ait besoin d'un contexte particulier ou du recrutement d'autres partenaires pour se lier aux cytosines méthylées, et que la protéine pourrait alors jouer un rôle dans la reconnaissance des régions hétérochromatiques dans les introns ciblés par le complexe SG1.

Il a été démontré que EDM2 est capable de reconnaître et de lier les marques hétérochromatiques présentes dans l'intron de *IBM1* par l'intermédiaire de son domaine PHD (Lei et al., 2014). De plus, EDM2 a été identifié, de la même manière que SG1, comme acteur déterminant de l'utilisation du site de polyadénylation distal de *IBM1*, ainsi que des autres cibles de SG1 (Lei et al., 2014). Le domaine PHD caractéristique de EDM2 est également capable de lier *in vitro* des marques chromatiniennes comme la méthylation H3K4, retrouvée dans les introns puisqu'ils sont transcriptionnellement actifs (Lei et al., 2014; Tsuchiya and Eulgem, 2014). Ainsi, EDM2 possède un domaine PHD capable de lier à la fois les marques H3K9me2 présentes dans les régions hétérochromatiques et les marques actives H3K4me1/2/3 attendues dans l'intron. Par conséquent, EDM2 pourrait, conjointement avec MBD4, reconnaître une signature épigénétique particulière, qui reste encore à déterminer, présente dans les régions hétérochromatiques des cibles de SG1 et recruter les autres partenaires du complexe à ces régions spécifiquement. Les marques H3K9me2 présentes au

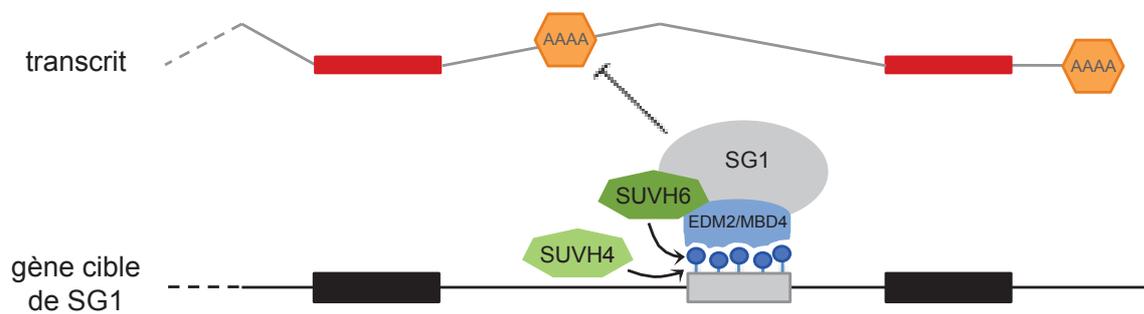


Figure IV-3 : Modèle schématique du fonctionnement de SG1.

Dans la partie basse est représenté un gène cible de SG1. Les cassettes noires correspondent aux exons et la cassette grise schématise la région hétérochromatique contenue dans l'intron qui porte des marque épigénétiques répressives, telles que la méthylation CHG et H3K9 (boules bleues). Le complexe SG1 serait recruté dans cette région grâce à la reconnaissance des marques chromatiniennes par MBD4 et/ou EDM2 et/ou SG1. Ces marques sont déposées et maintenues par SUVH6, membre du complexe, et/ou SUVH4. Ainsi, le complexe SG1 permet la production du transcrit (en rouge) sous sa forme longue en favorisant l'utilisation du site de polyadénylation distal plutôt que le site proximal, tous deux représentés en orange.

niveau de l'élément transposable inséré dans l'intron de *RPP7*, cible du complexe SG1, est réduite en l'absence de EDM2 et dans un triple mutant *svh456* (Tsuchiya and Eulgem, 2013a). Par conséquent, ces marques sont dépendantes à la fois de EDM2 et des protéines SUVH4, 5 et/ou 6. Ceci suggère qu'au locus *RPP7*, EDM2 et les SUVH agissent en coopération pour permettre la production de transcrit long.

Nous avons pu montrer que SUVH4 et SUVH6 jouent un rôle au niveau des cibles de SG1. En effet, nous avons observé que SUVH4 et/ou SUVH6, selon le gène cible, étaient impliqués dans la production des transcrits longs (**Figure 4, Deremetz et al**). Tous deux sont connus pour leur capacité à méthyler les H3K9 par reconnaissance des cytosines méthylées en contexte CHG (Ebbs and Bender, 2006; Jackson et al., 2004). Il a été démontré que les protéines SUVH4 et SUVH6 possèdent toutes les deux une activité histone méthylase qui peut être redondante au niveau de certains locus (Ebbs and Bender, 2006). L'étude des doubles mutants *sg1 svh4* et *sg1 svh6* montre que la fonction de SG1 et SUVH4/SUVH6 est nécessaire pour la transcription correcte des gènes cibles. Ainsi, il est probable que, tout comme EDM2, SG1 ait besoin d'être associé avec SUVH4 ou SUVH6 pour que le complexe puisse être fonctionnel et permettre la production de transcrits longs.

Pour résumer, les régions hétérochromatiques dans les introns des cibles identifiées du complexe SG1 sont très enrichies en marques épigénétiques répressives. Ces marques constituent une barrière à franchir, mais servent aussi d'ancrage pour le recrutement du complexe qui va permettre l'expression correcte et fonctionnelle des gènes cibles. La maintenance de ces marques est opérée par SUVH4 et/ou SUVH6, en lien avec le complexe. Ces marques seraient alors reconnues par des protéines du complexe, comme MBD4, EDM2 par l'intermédiaire de son domaine PHD, mais aussi potentiellement par SG1 lui-même grâce à son domaine BAH. Même si aucune protéine en lien avec le complexe de polyadénylation n'a été mise en évidence dans le complexe SG1, ce dernier permet la production des transcrits fonctionnels de ses cibles, en défavorisant, par un mécanisme qui reste encore à déterminer, l'utilisation du site de polyadénylation proximal qui se trouve dans l'intron (**Figure IV-2**).

Quelle est l'étendue des cibles de SG1 ?

Quatre cibles de SG1 ont pu être mises en évidence à ce jour, notamment grâce à la taille importante de leur intron et de l'élément transposable lourdement méthylé qu'il contient. Cependant, il existe très probablement d'autres cibles de ce complexe qui n'ont pas encore été identifiées. En effet, l'élément hétérochromatique dans l'intron de *IBM1* n'est pas un élément transposable à proprement parler, mais correspond à un fragment homologue d'un gène chloroplastique et mitochondrial (Saze et al., 2013). La présence de cette région hétérochromatique d'environ 700 pb est suffisante pour être reconnue par SG1. En effet, l'exemple de *IBM1* illustre bien le fait que SG1 n'intervient pas uniquement dans le ciblage des éléments transposables insérés dans les introns, mais pourrait reconnaître des insertions dues à des remaniements chromosomiques par exemple, ou des fragments provenant des génomes des mitochondriaux et chloroplastiques, augmentant ainsi le nombre de cibles potentielles.

L'analyse du méthylome du mutant *suvh6* ne révèle pas de différence significative de méthylation dans les trois contextes par rapport au sauvage Col (**Figure S7, Deremetz et al.**). En revanche, nous avons pu identifier 220 DMR entre *suvh6* et Col en contexte CHG. En associant ces résultats avec des données de séquençage des ARN entre Col et *sg1*, nous avons cherché à mettre en évidence des gènes qui étaient à la fois des cibles de SUVH6 et qui présentaient des différences d'expression au niveau de certains exons dans un fond mutant *sg1*. Ainsi, nous avons pu identifier 21 gènes dont au moins un exon présentait des différences d'expression entre Col et *sg1* : 12 gènes avec au moins un exon moins exprimé dans *sg1* que dans Col, dont AT3G05410 et AT1G58602 (*RPP7*) et 9 gènes avec au moins un exon moins exprimé dans Col que dans *sg1* (**Table 2, Deremetz et al.**). Ainsi, nous avons pu de cette façon mettre en évidence deux des quatre cibles de SG1, et ce résultat confirme le lien entre SG1 et SUVH6. Cependant, nous n'avons pas pu, par cette méthode, retrouver AT3G11270 et *IBM1*. En effet, le transposon contenu dans le gène AT3G11270 est très homologue à d'autres transposons, et les *reads* correspondantes ont été éliminées au cours de l'alignement car elles ne correspondaient pas à une région unique. De plus, il n'est pas surprenant que *IBM1* n'ait pas été identifié dans la mesure où nous avons montré que sa transcription longue est sous la dépendance de SUVH4 (**Figure 4, Deremetz et al.**). Le cas de *IBM1* montre qu'il est tout à fait possible que d'autres cibles de SG1, par exemple celles dont la transcription longue est dépendante de SUVH4, existent et n'aient pas encore été identifiées.

Comment le complexe SG1 influence-t-il le choix du site de polyadénylation ?

Le fait que le complexe SG1 se lie préférentiellement aux éléments transposables qui se trouvent dans les introns implique une reconnaissance spécifique de ses cibles (Saze et al., 2013; Wang et al., 2013). Les raisons de cette reconnaissance des régions hétérochromatiques restent à déterminer, d'autant plus que l'ensemble des cibles du complexe SG1 est encore inconnue. Il a été montré que la fonction de IBM1 est liée à l'activité transcriptionnelle. En effet, IBM1 semble cibler en particulier les gènes actifs d'un point de vue transcriptionnel et il a également été montré que les transposons qui sont réactivés dans les fonds mutants *cmt3* ou *kyp*, appauvris en marques répressives, sont alors ciblés par IBM1 (Inagaki et al., 2010). Cet exemple montre qu'il existe des mécanismes et/ou des signatures épigénétiques particulières qui permettent la perception de l'état transcriptionnel. Le même genre de processus pourrait permettre d'expliquer comment le complexe SG1 reconnaît ses gènes cibles. Comme mentionné précédemment, la capacité de la protéine EDM2 à reconnaître à la fois les marques de transcription active et les marques répressives, en fait un bon candidat pour la perception des gènes cibles.

La recherche de suppresseurs de la mutation *sg1* a permis de mettre en évidence le rôle de FPA dans la régulation de la polyadénylation des cibles du complexe SG1. Même si le rôle joué par FPA dans le choix du site de polyadénylation d'un certain nombre de gènes, y compris *IBM1* et *FPA* lui-même, a été démontré, aucun lien avec des composants connus des machinerie d'épissage ou de polyadénylation n'a pu être mis en évidence (Hornyik et al., 2010b). Ainsi, la manière dont la protéine FPA est impliquée dans la régulation de la polyadénylation reste encore inconnue.

Nous avons pu remarquer qu'au moins au locus *IBM1*, la méthylation non-CG de la région hétérochromatique était réduite dans un mutant *fpa-3* ainsi que dans un double mutant *fpa-11 sg1-1* (**Figure S3, Deremetz et al**). Or, il a été montré d'une part que la méthylation de cette région hétérochromatique est essentielle à la transcription de la forme fonctionnelle de *IBM1* et d'autre part que le complexe SG1, qui joue un rôle crucial dans la transcription de *IBM1-L*, est capable de se lier à ce locus (Rigal et al., 2012; Saze et al., 2013; Wang et al., 2013). Comment la transcription de *IBM1-L* peut-elle se faire dans un double mutant *fpa sg1* alors que la méthylation de l'intron est fortement réduite ? Ce résultat suggère qu'en l'absence de

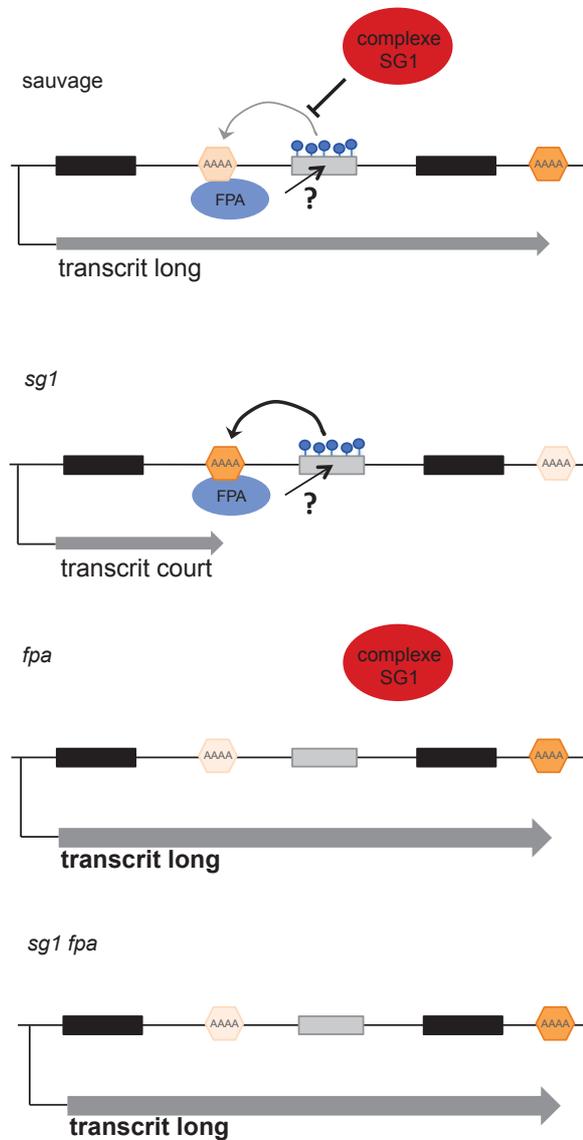


Figure IV-3 : Influence du choix du site de polyadénylation par les voies SG1 et FPA au niveau de gènes cibles dans les différents fonds génétiques.

Représentation pour les fonds génétiques wild-type, mutants *sg1* et *fpa* et double mutant *sg1 fpa* d'un gène cible de SG1 contenant une région hétérochromatique (en gris) dans un intron. Les exons sont représentés par des cassettes noires et les sites de polyadénylation sont schématisés en orange, plus foncés lorsqu'ils sont favorisés. Le lien direct ou indirect entre FPA et les marques chromatinienne est marqué d'un « ? » car sa nature reste à déterminer.

FPA, SG1 n'est pas essentiel pour permettre la production de transcrits longs, ce qui est cohérent avec les résultats obtenus concernant l'expression de *IBM1-L* dans les fonds mutant *fpa* et *fpa sg1* (**Figure 2A, Deremetz et al**). Ainsi, en l'absence de FPA et SG1, c'est le site de polyadénylation distal qui est utilisé (**Figure IV-3**). La force du signal de polyadénylation est sous l'influence des différents éléments qui le constituent et de leur variation par rapport à une séquence consensus (Mandel et al., 2008). Il est possible qu'au niveau des cibles de SG1, l'un des sites de polyadénylation soit un signal plus fort, c'est à dire qui est plus reconnaissable ou plus accessible à la machinerie de polyadénylation, ce qui expliquerait le choix de l'un plutôt que l'autre.

En l'absence de FPA, on peut envisager que la surproduction de la protéine IBM1 ait pour conséquence la disparition de marques H3K9me2 à des locus normalement éteints. Cela pourrait aussi concerner l'intron de *IBM1*, expliquant ce que nous avons observé dans le mutant *fpa* (**Figure S3, Deremetz et al**). Pourtant, à l'échelle du génome, il n'y a pas de changements importants en terme de méthylation dans un mutant *fpa-7* (Stroud et al., 2013), ce qui suggère que la surexpression de *IBM1-L* n'aurait de conséquences qu'à un nombre limité de locus. De plus, l'étendue des cibles de FPA étant inconnue, il est possible que FPA impacte la transcription de gènes qui sont impliqués, directement ou indirectement, dans le dépôt ou la maintenance des marques répressives présentes dans les introns des cibles de SG1. D'autre part, il a été montré que FPA jouait un rôle dans la régulation du maintien de la chromatine dans un état silencieux. En effet, certains locus, comme le rétroélément *AtSNI*, sont réactivés et présentent une diminution de méthylation en contexte non-CG dans un mutant *fpa* (Baurle et al., 2007). Cette perte de méthylation en l'absence de FPA pourrait avoir la même origine que celle observée dans la région hétérochromatique de *IBM1*.

Nous avons identifié la mutation *fpa* comme suppresseur du phénotype *sg1*. Le complexe SG1 semble favoriser, par un mécanisme qui est encore inconnu, l'utilisation du site de polyadénylation distal alors que FPA influence l'utilisation du site proximal présent dans l'intron des cibles de SG1. Ces deux voies de régulation semblent en effet s'opposer, mais partagent-elles pour autant les mêmes cibles. Il a été montré que FPA est impliqué dans la régulation de gènes comme *FLC* ou *SOC1* qui ne semblent pas être des cibles de SG1 (Duc et al., 2013). De plus, *fpa* ne complémente pas parfaitement la mutation *sg1* au niveau de toutes

les cibles. En effet, au locus *RPP7*, la production du transcrit long n'est pas différente de ce qui peut être observé dans *sg1* (**Figure 1, Deremetz et al**). Il est cependant possible que FPA ne soit pas le seul acteur qui permette le choix du site de polyadénylation proximal. Il a été montré par exemple que FPA a une activité partiellement redondante avec FCA au niveau de certains locus (Baurle et al., 2007). La protéine FCA est en effet impliquée dans la régulation la transcription d'un certain nombre de gènes, y compris lui-même, en influençant le choix du site de polyadénylation (Quesada et al., 2003). Même si les deux protéines ne partagent pas d'homologie de séquence, à part la présence d'un domaine RRM, le rôle de FPA et FCA montre une certaine redondance fonctionnelle selon les locus ciblés. Ainsi, dans la mesure où nous avons identifié FPA dans notre crible suppresseur, il est possible que FCA soit également impliqué dans la régulation de la polyadénylation au niveau de certaines cibles de SG1.

La fonction de FPA, comme cela a été montré, serait davantage de limiter la formation de transcrits provenant de locus qui doivent normalement être éteints, considérés alors comme aberrants, et le complexe SG1 serait plutôt impliqué dans un mécanisme permettant la transcription correcte et fonctionnelle de gènes dans lesquels se serait inséré un élément transposable, perturbant ainsi le choix du site de polyadénylation.

La question du lien entre SG1 et la polyadénylation reste complète. En effet, aucun composant de la machinerie de polyadénylation n'a été identifié dans le complexe. Comment le complexe SG1 peut-il influencer le choix du site de polyadénylation de ses cibles ? Plusieurs hypothèses peuvent être faites pour répondre à cette question. D'une part, il a été montré que la protéine EDM2 porterait un domaine putatif N6-Adenine Methyltransferase-like et pourrait donc jouer un rôle dans la méthylation des adénines des ARNm des cibles (Lei et al., 2014). Même si la méthylation des transcrits n'est pas un phénomène rare, la nature de son rôle régulateur reste encore mal connu (Zhong et al., 2008). L'étude de la distribution des adénines méthylées a cependant révélé un enrichissement dans la partie 3' du transcrit, suggérant ainsi que cette marque pourrait être impliquée dans des processus comme la terminaison de la transcription ou la polyadénylation des transcrits (Bodi et al., 2012). Ainsi, il est envisageable que cette marque joue un rôle régulateur dans la formation des transcrits longs, par l'intermédiaire de EDM2. D'autre part, le rôle des différents partenaires du

complexe SG1, en particulier ceux dont la fonction n'est pas connue, reste à préciser. Il est possible qu'un membre du complexe soit en effet un acteur en lien avec la machinerie de polyadénylation. Et finalement, le complexe SG1 qui regroupe pas moins de sept protéines différentes pourrait tout à fait masquer, par effet d'encombrement, le signal de polyadénylation proximal, favorisant ainsi l'utilisation du site distal au niveau des gènes cibles.

Des arguments suggèrent que la fonction des gènes cibles de SG1, comme *IBM1*, pourrait ne pas être uniquement déterminée par la régulation de la polyadénylation qui aboutit à la production de transcrits longs ou courts. En effet, il a été montré que *IBM1-L* est le transcrit fonctionnel, mais il n'est pas impossible qu'il existe un niveau de contrôle supplémentaire avec une régulation post-transcriptionnelle du transcrit long, en plus de la régulation de la polyadénylation. Dans le cas du mutant *rsal* par exemple, la quantité de transcrits *IBM1-L* est affectée (**Figure 3, Deremetz et al**). En effet, elle est réduite de plus de moitié par rapport à Col-0, tout comme dans un fond mutant *sgl*, ce qui suggère qu'en l'absence de RSA1 la production du transcrit long est impactée. Pourtant, le mutant *rsal* ne présente pas le phénotype développemental ni moléculaire (méthylation de *BNS*) caractéristique observé dans le mutant *sgl*. Ce résultat suggère qu'il existe d'autres niveaux de régulation de la fonctionnalité de *IBM1* au-delà de la seule production des transcrits longs et que SG1 pourrait participer également à un processus de régulation post-transcriptionnel et de maturation du transcrit *IBM1-L*.

Le travail réalisé sur ce projet dans le cadre de ma thèse a permis de décrypter en partie le mécanisme dans lequel est impliqué SG1, qui fait le lien entre les marques chromatiniennes et la production des transcrits. Il reste cependant encore beaucoup à démontrer et à comprendre sur la fonction précise de SG1 et de son complexe dans la régulation des gènes cibles, sur l'étendue des gènes qui sont régulés par ce mécanisme et sur la relation du complexe SG1 avec les machineries de polyadénylation.

Partie V : Caractérisation d'épimutations apparaissant dans un fond génétique *sg1*.

Dans les mutants *sg1*, le corps des gènes est envahi par des marques épigénétiques normalement retrouvées au niveau des transposons et des régions répétées (méthylation H3K9 et CHG). Le phénotype des mutants *sg1* se manifeste après plusieurs générations, évolue, de plus en plus sévèrement, dans les générations successives, et diffère d'un individu à l'autre, faisant penser à des épimutations. J'ai recroisé le mutant *sg1* avec une plante sauvage (Col-0) et j'ai constaté que les plantes F2 qui sont redevenues sauvage au locus *SG1* gardent des défauts phénotypiques qui peuvent être variés.

Mes objectifs dans le cadre de ce projet étaient de :

- déterminer si au moins un de ces phénotypes, sans doute liés à des épimutations, est transmis de façon stable aux générations suivantes ;
- déterminer si ce phénotype est maintenu après croisement avec une autre accession dans le but de cartographier les éventuelles épimutations correspondantes ;
- initier la cartographie de ces épimutations pour les identifier ;
- caractériser les méthylomes de ces épimutants pour identifier les régions ou locus pouvant être responsables des phénotypes.

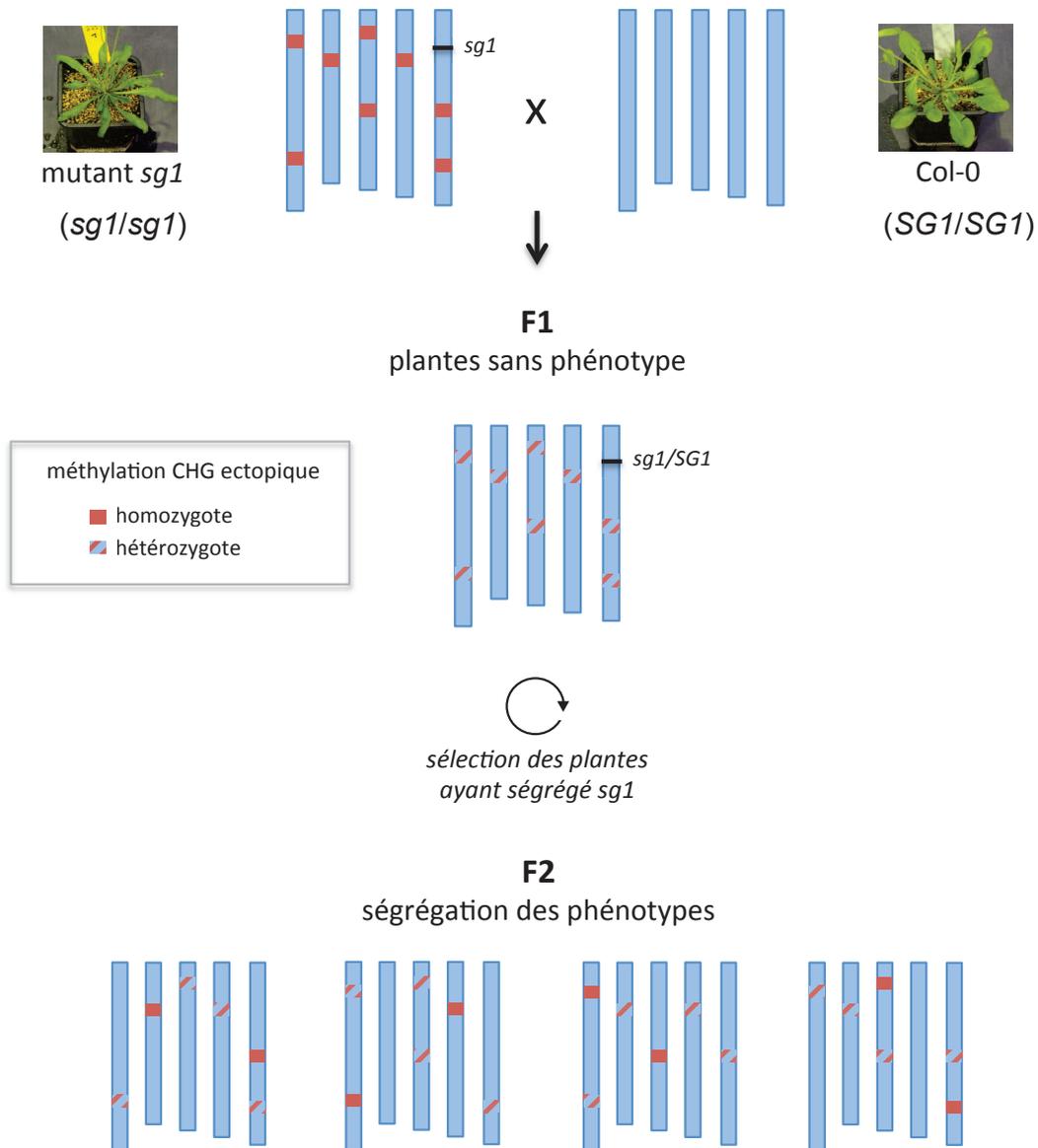


Figure V-1 : Fabrication des populations d'épimutants.

Pour chaque génération (les parents, la F1 et la F2), les cinq chromosomes d'*Arabidopsis* sont représentés par des barres bleues. Les plantes F1 sont issues d'un croisement entre un mutant *sg1* et une plante sauvage Col-0. La mutation *sg1*, représentée par une barre noire, se trouve à l'état hétérozygote en F1 et a été contre-sélectionnée dans les plantes F2 présentées ici. Les épimutations sont schématisées en rouge sur les chromosomes à l'état homozygote et en rouge barré de bleu à l'état hétérozygote.

(I) Génération des populations d'épimutants *sgl* induits

Dans le but d'identifier et de caractériser les épimutations induites par la mutation *sgl*, des populations d'épimutants ont été générées. Elles serviront d'outils pour la recherche d'épiallèles associés à un phénotype. Ceci pourrait permettre d'expliquer, dans notre cas, une partie du phénotype *sgl* en identifiant ses causes, et ainsi établir un lien entre la méthylation de l'ADN et par exemple des variations d'expression des gènes cibles, ou des modifications de la maturation de leurs ARN.

Pour générer ces populations (**Figure V-1**), une plante portant la mutation *sgl* dans un fond génétique Col a été croisée avec une plante sauvage Col-0. Ceci a été réalisé en parallèle avec deux allèles *sgl* : d'une part le mutant ponctuel *sgl-1* que nous avons identifié en caractérisant le QTL *SG1* (cf. **Partie II**), et d'autre part un mutant d'insertion T-DNA que nous avons appelé *sgl-2* (SAIL_310B06). Comme décrit précédemment, la fonction *SG1* est compromise dans ces deux mutants car ils présentent un phénotype morphologique fort associé à un phénotype moléculaire (*i.e.* invasion de nombreux gènes par de la méthylation en contexte CHG). Notre but était de déterminer si la modification des épigénomes observée dans les mutants *sgl* persiste lorsque *SG1*, et donc *IBM1*, redeviennent fonctionnels. Cette observation avait déjà été faite dans le cas de *ddm1* ou *met1*, conduisant à la création de lignées épIRIL basées sur ces mutations (Johannes et al., 2009; Reinders et al., 2009). Nous souhaitons donc étudier les lignées F2 ayant ségrégué la mutation *sgl* et retrouvé une protéine *SG1/IBM2* fonctionnelle. Dans ces lignées, nous nous attendons à ce que les épimutations ségrégent de la même manière, car elles sont en théorie à l'état hémiméthylés en F1. Ces plantes F2 portent alors les épimutations à l'état hétérozygote ou fixé, et parmi elles nous avons observé que certaines présentent toujours des caractéristiques phénotypiques du mutant *sgl* d'origine.

En tout, 1 645 plantes F2 ont été génotypées pour les deux populations issues du croisement avec les allèles *sgl*, 963 et 682 pour *sgl-1* et *sgl-2* respectivement, afin de sélectionner les plantes ayant ségrégué la mutation *sgl*. Comme attendu dans le cas d'une ségrégation mendélienne, environ un quart d'entre elles étaient homozygotes pour l'allèle sauvage *SG1*, à savoir 24,5% et 23,9% pour les F2 issues des croisements entre Col et *sgl-1* et *sgl-2* respectivement (**Tableau V-1**). J'ai ainsi sélectionné 399 plantes F2 de génotype *SG1/SG1*.

Croisement d'origine	Génotype			Nombre total
	<i>sg1/sg1</i>	<i>sg1/SG1</i>	<i>SG1/SG1</i>	
Col x <i>sg1-1</i>	224 23,3%	503 52,2%	236 24,5%	963 -
Col x <i>sg1-2</i>	167 24,5%	352 51,6%	163 23,9%	682 -

Tableau V-1: Ségrégation des mutations *sg1* dans les F2.

Dans ce tableau sont indiqués le nombre ainsi que le pourcentage observé pour chaque génotype au locus *SG1* en génération F2. En rouge est indiqué le pourcentage des plantes qui présentent l'allèle sauvage *SG1* à l'état homozygote.



Figure V-2 : Phénotype floral observé dans les F2 *SG1/SG1*.

Pour une très grande majorité des fleurs qui portent des pétales surnuméraires, cinq pétales sont dénombrés, mais des fleurs portant jusqu'à neuf pétales ont été observées.

(II) Etude du phénotype floral des populations d'épimutants

1) Suivi du phénotype floral

Parmi les phénotypes observés dans les F2 *SG1/SG1*, j'ai choisi de suivre un phénotype lié au développement de la fleur et qui se caractérise par un nombre anormal de pétales. En effet, dans le fond mutant *sg1*, il y a une augmentation du nombre de pétales au niveau de certaines fleurs (**Figure V-2**). Alors que la fleur d'*Arabidopsis thaliana* compte habituellement quatre pétales, nous avons observé des fleurs à majoritairement cinq pétales, mais pouvant aller jusqu'à neuf pétales. Ce phénotype est hétérogène au sein d'une même plante, mais également au sein d'une population.

Le choix de ce phénotype lié au développement de la fleur a été fait car nous l'avons observé fréquemment dans les populations d'épimutants *sg1*-induits, et également dans le fond mutant *sg1*. Il s'agit d'un phénotype visuel, relativement facile à quantifier et à suivre sur plusieurs générations.

A partir de la génération F2, certaines plantes présentent une variété d'autres phénotypes caractéristiques que j'ai pu distinguer au cours des générations, comme des phénotypes de croissance foliaire ou racinaire. Certaines données concernant le phénotype visuel de ces plantes ont été conservées et pourront être utilisées ultérieurement pour l'identification d'autres épiallèles, associés à ces phénotypes.

2) Sélection des plantes en génération F2

Afin de déterminer un seuil à partir duquel les plantes sont considérées comme présentant un phénotype, je me suis intéressée à l'apparition du phénotype floral dans une population de plantes sauvages. Ainsi, j'ai phénotypé une population de 59 plantes de fonds génétique Col-0 et déterminé que ces plantes présentaient de 0 à 7,9% de fleurs avec des pétales surnuméraires et en moyenne 2,8%. J'ai pu également déterminer que 83% de ces plantes présentaient moins de 5% de phénotype. Ce phénotype étant présent dans une population naturelle Col-0, j'en ai

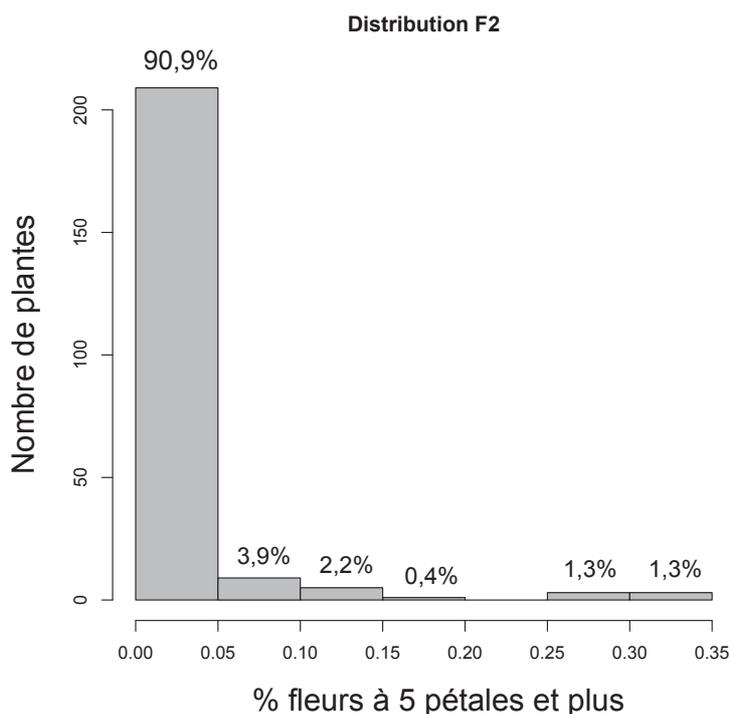


Figure V-3 : Distribution du phénotype floral dans la population F2 *SG1/SG1*.

Les plantes F2 sont issues du croisement Col x *sg1-1*. Parmi ces plantes 230 plantes *SG1/SG1* ont été phénotypées de manière quantitative. La distribution du phénotype floral est représentée en fonction de la fréquence du phénotype floral observé.

Génération	Individus	Phénotype	allèle du parent <i>sg1</i>
F2	1.58	majorité des fleurs avec phénotype	<i>sg1-2</i>
	8.19	beaucoup de fleurs avec phénotype	<i>sg1-2</i>
	5.32	beaucoup de fleurs avec phénotype	<i>sg1-2</i>
	7.52	majorité de fleurs avec phénotype	<i>sg1-2</i>
	5.15	31%	<i>sg1-1</i>
	3.39	30%	<i>sg1-1</i>
	5.38	35%	<i>sg1-1</i>
	15.11	32%	<i>sg1-1</i>
	18.45	26%	<i>sg1-1</i>
	22.2	26%	<i>sg1-1</i>
F3	1.58.4	4%	<i>sg1-2</i>
	5.32.18	38%	<i>sg1-2</i>
	3.39.21	36%	<i>sg1-1</i>
	18.45.18	14%	<i>sg1-1</i>

Tableau V-2: Phénotypage des individus épimutants F2 et F3 qui ont été étudiés.

Sont indiqués dans ce tableau les noms des 10 plantes F2 (4 issues du croisement avec *sg1-2* et 6 issues du croisement avec *sg1-1*) et des 4 plantes F3 qui ont été étudiées et phénotypées individuellement. Le phénotype, quand il a été quantifié, est exprimé en pourcentage de fleurs à 5 pétales ou plus.

donc conclu que, quelle que soit sa cause, il apparaît spontanément au niveau de certaines fleurs.

En tout, 230 plantes F2 résultant du croisement *sg1-1* x Col-0 ont été phénotypées : 209 d'entre elles, soit 90,9%, présentaient moins de 5% de fleurs anormales, un résultat comparable à celui obtenu pour le sauvage comme indiqué ci-dessus (**Figure V-3**). Parmi les 22 plantes F2 qui avaient plus de 5% de fleurs avec des pétales surnuméraires, six plantes présentaient un fort phénotype, c'est à dire au-delà de 25% de fleurs anormales. Ces plantes ont été gardées pour étudier leur descendance F3 et sont nommées 3.39, 5.15, 5.38, 15.11, 18.45 et 22.2 (**Tableau V-2 ; Figure V-4**).

Parmi les plantes F2 issues du croisement entre *sg1-2* et Col-0, j'ai également pu observer que 149 plantes, soit 91,4%, ne présentait pas ou peu de fleurs avec un nombre modifié de pétales et les plantes ayant le plus fort phénotype au niveau des fleurs ont été sélectionnées et gardées en génération F3. Ces quatre plantes sont nommées 1.58, 8.19, 7.52, 5.32 (**Tableau V-2 ; Figure V-4**).

3) Phénotypage des populations F3

Afin d'étudier l'héritabilité du phénotype floral observé chez certaines plantes ayant ségrégué la mutation *sg1*, les 10 lignées F2 (indiquées sur le **Tableau V-2** et la **Figure V-4**) ont été autofécondées puis phénotypées individuellement en génération F3.

Le phénotypage des populations F3 a permis de mettre en évidence l'hétérogénéité du phénotype floral au sein d'une même population. Les moyennes observées pour les différentes F3 varient de 8% à 20% de fleurs avec plus de quatre pétales, contre 2,8% pour la population sauvage Col-0 (comme déterminé ci-dessus), et les écart-types sont également très variables, allant de 8% à 21% (**Tableau V-3**). Ainsi, l'hétérogénéité du phénotype des individus en F3 est importante et variable d'une population à l'autre. Néanmoins, la comparaison de cette distribution par rapport à la population de plantes Col-0 montre que ces populations diffèrent significativement du sauvage pour le caractère observé (**Figure V-5**). Par exemple, la population F3 8.19, qui présente une moyenne de 20,3% de fleurs à cinq

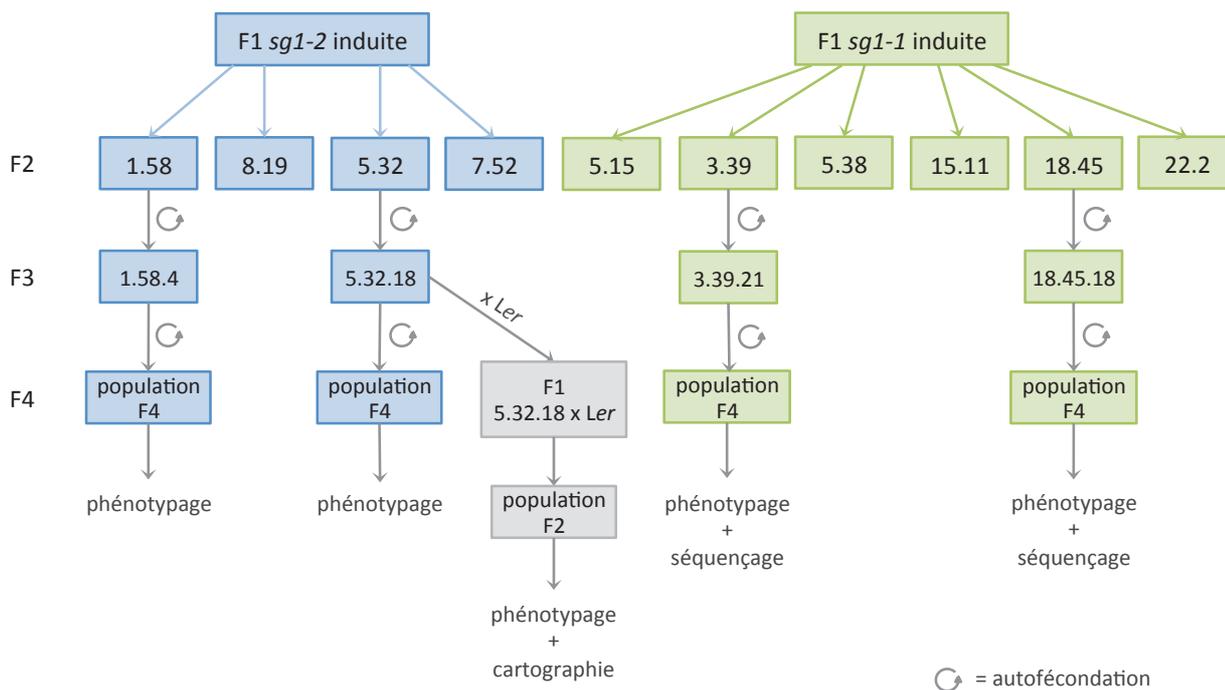


Figure V-4 : Généalogie des populations d'épimutants *sgI*-induites.

Sont représentées les plantes et les populations d'épimutants étudiées, en bleu celles issues du croisement de Col avec l'allèle *sgI-2* et en vert celles issues du croisement avec *sgI-1*. A partir de ces deux F1 ont été générées 10 F2 *SGI/SGI*. Quatre de ces F2 ont été étudiées en F3 puis en F4. Les populations de plantes ayant servi pour la cartographie et le séquençage sont indiquées.

Population	Individu Parent	Moyenne Phéno.	Phéno. min.	Phéno. max.	Médiane Phéno.
F3	1.58	17,9	0,0	73,3	9,8
	8.19	20,3	1,9	52,6	16,4
	5.32	20,0	0,0	100,0	9,5
	7.52	14,7	0,0	48,3	10,4
	5.15	9,0	0,0	33,3	7,6
	3.39	13,5	0,0	57,0	12,5
	5.38	19,4	0,0	66,7	17,4
	15.11	12,5	0,0	100,0	7,9
	18.45	10,8	0,0	50,0	5,9
	22.2	7,5	0,0	28,1	6,3
F4	1.58.4	2,4	0,0	6,7	2,1
	5.32.18	13,2	0,1	40,0	11,8
	3.39.21	21,1	0,0	81,6	18,2
	18.45.18	5,7	0,0	19,1	4,7

Tableau V-3: Phénotypage des populations épimutants F2 et F3 qui ont été étudiés.

Sont indiqués les noms des 10 population F3 et des 4 populations F4 qui ont été étudiées et phénotypées. Les valeurs de phénotype sont exprimées en pourcentage de fleurs anormales. Sont indiqués les pourcentages de fleurs anormales de la plante avec le plus faible et le plus fort phénotype au sein de la population, ainsi que la moyenne et la médiane de ces valeurs.

pétales, est constituée de plantes ayant un phénotype allant de 1,9% à 52,6 % de fleurs à cinq pétales. De même, la population F3 5.32 est constituée de plantes ayant en moyenne 20 % de phénotype : 0% de fleurs à cinq pétales pour la plante qui présente le moins de phénotype jusqu'à à 100% pour celle qui en présente le plus. Ainsi, pour deux populations qui présentent en moyenne la même proportion de fleurs à cinq pétales (autour de 20%), le pourcentage de phénotype floral des plantes les plus extrêmes (de 0 à 100% pour 8.19 et de 1,9 à 52,6% pour 5.32), et la répartition des individus entre ces extrêmes, varient de façon importante.

Cette comparaison illustre bien la variabilité du phénotype floral observé au sein d'une même population ou en comparant deux populations différentes (ici, chez les populations F3). La distribution des valeurs de phénotype mesurées, qui est continue d'un extrême à l'autre, suggère que le phénotype observé est un caractère complexe, et que par conséquent, il existerait plusieurs éléments causaux. En considérant l'hypothèse que ce phénotype serait lié à des épimutations induites par la mutation *sg1*, il y aurait alors potentiellement plusieurs épimutations causales. En effet, si le phénotype observé était lié à une seule épimutation, les plantes montreraient une proportion de fleurs à cinq pétales proche de celle observée pour la plante mère, ou inférieure dans le cas où l'épimutation, labile par définition, serait perdue. Or, ce n'est pas le cas pour les populations observées, puisque certaines plantes F3 ont plus de fleurs à cinq pétales que la F2 dont elles sont issues, et d'autres ont complètement perdu le phénotype.

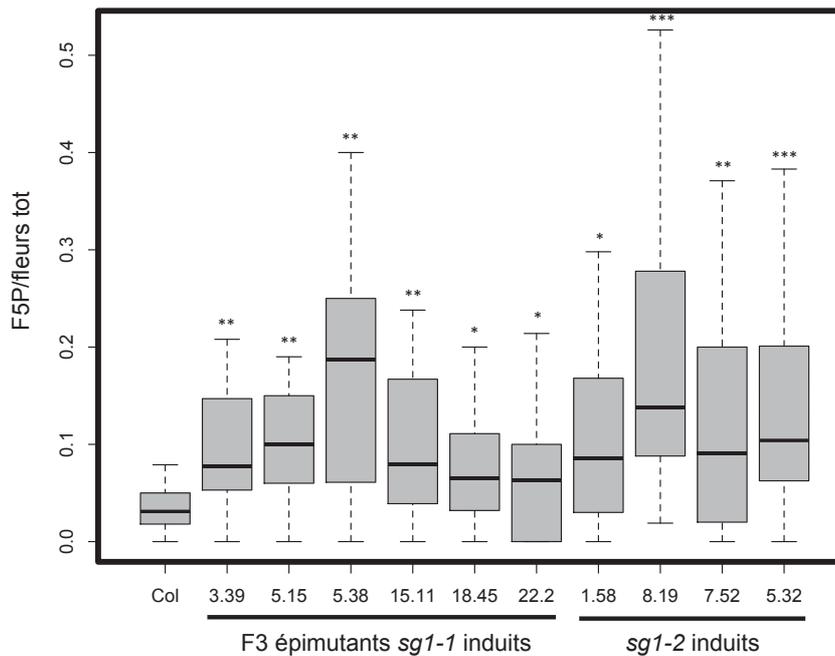


Figure V-5 : Phénotypage des populations d'épimutants en F3.

Le boxplot illustre la distribution des résultats du phénotypage des différentes populations d'épimutants, issus des deux allèles mutants *sg1*. Le test de Student réalisé montre que ces populations diffèrent de Col pour le caractère observé avec une probabilité de 95% (*), 99% (**) ou 99,9% (***). F5P = fleurs à cinq pétales ou plus.

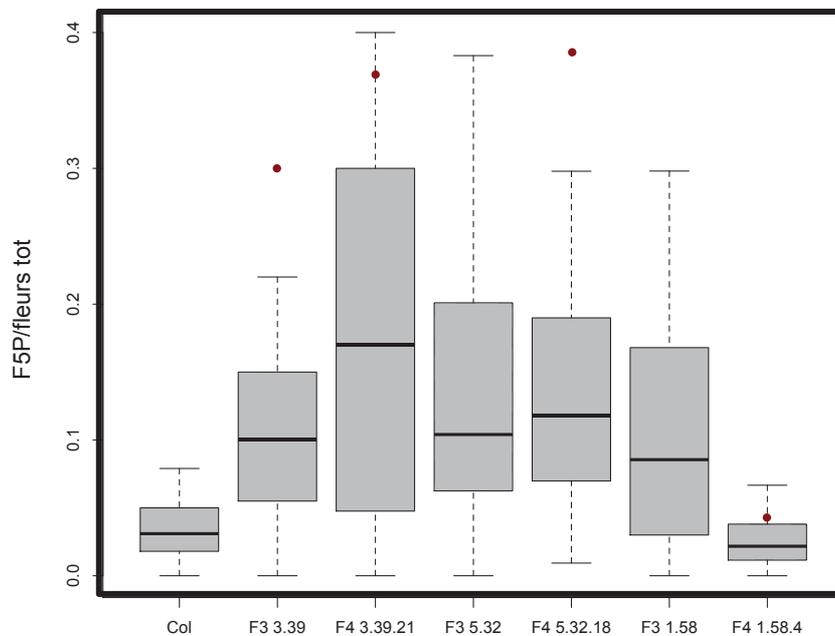


Figure V-6 : Suivi du phénotype entre les populations F3 et F4 d'épimutants.

Le boxplot illustre la distribution des résultats du phénotypage des populations d'épimutants en F3 et F4. Le point en rouge indique le pourcentage de fleurs à cinq pétales observé pour la plante mère dont la population est issue, quand celui-ci a été mesuré de manière quantitative.

(III) Identification des épimutations associées au phénotype floral

1) Stabilité du phénotype floral

Le prérequis pour pouvoir identifier ces épimutations est de s'assurer de leur stabilité, au moins pendant quelques générations, pour pouvoir les cartographier. Ceci permettra de pouvoir associer le phénotype à certaines régions du génome, comme dans le cas d'une cartographie classique. Dans ce but, des lignées présentant du phénotype en génération F3 ont été observées en génération F4.

J'ai montré que les descendances F4 issues des plantes F3 3.39.21 et 5.32.18 (qui ont respectivement 36% et 38% de fleurs à cinq pétales) présentent, en moyenne, 21% et 13% de phénotype floral (**Tableau V-3**). La moyenne des phénotypes observés en F4 est plus faible que le phénotype de la plante mère F3, mais la distribution des valeurs est très dispersée. Les valeurs du phénotype vont de 0 à 40% de fleurs à cinq pétales pour la population F4 5.32.18 et 82% de fleurs à cinq pétales pour la population F4 3.39.21 (**Figure V-6 ; Tableau V-3**). Ces résultats suggèrent que les éventuelles épimutations associées au phénotype se fixent puisque les valeurs observées sont davantage distribuées vers les extrêmes au cours des générations. Ainsi, pour certaines plantes, ces épimutations seraient stabilisées, au moins jusqu'en génération F4, soit deux générations après ségrégation de la mutation causale *sg1*.

Comme contrôle, la population F4 issue de la plante F3 1.58.4, qui présentait un phénotype faible avec 4% de fleurs à 5 pétales (ce qui est comparable aux valeurs obtenues chez le sauvage, voir paragraphe **(II) 2) « Sélection des plantes en génération F2 »**), a également été phénotypée. Les résultats de phénotypage de cette F4 montrent qu'elle ne diffère pas significativement de la population sauvage Col-0 en termes de phénotype floral (**Figure V-6**). Ceci suggère que, dans le cas de cette F4, les épimutations qui étaient responsables du phénotype en F2, puis F3 ont été perdues. Il est possible que cette perte se soit faite progressivement, et qu'elle soit liée au fait que la plante F3 1.58.4, qui a servi à générer la F4, présentait déjà un faible phénotype.

Il semblerait donc que la ou les épimutations associées au phénotype floral peuvent être perdues spontanément, mais que la sélection de plantes qui présentent un phénotype floral

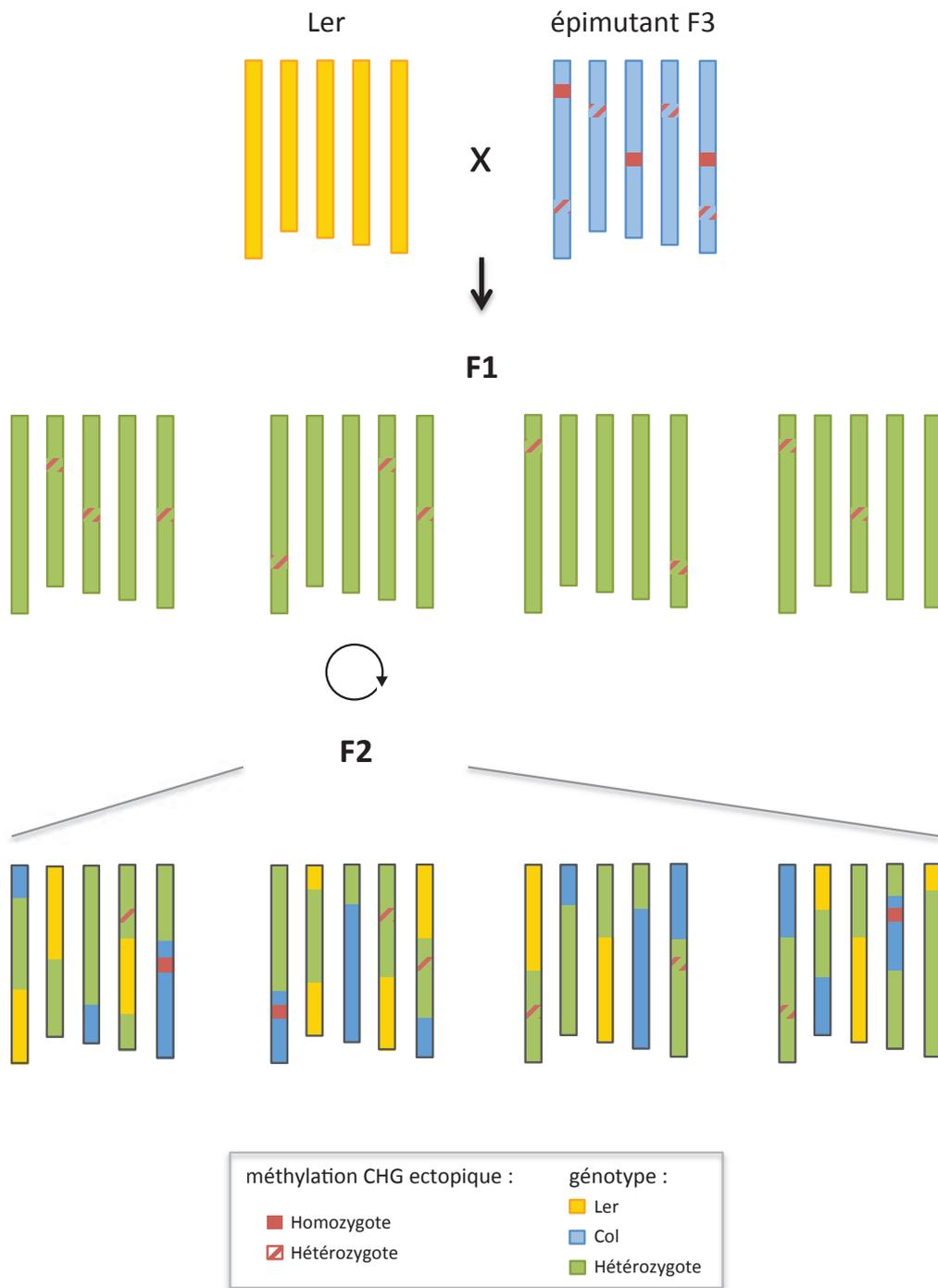


Figure V-7 : Génération des populations de cartographie.

Après un croisement entre *Ler* et la plante F3 qui porte les épimutations induites par la mutation *sgI*, celles-ci ségrègent. Les couleurs des chromosomes illustrent le génotype et les épimutations sont représentées en rouge, plein ou hachuré, selon la légende.

important permettrait leur stabilité de génération en génération. Ainsi, j'ai pu montrer que cette ou ces épimutations peuvent être fixées, au moins jusqu'en génération F4, ce qui permet en théorie de pouvoir les cartographier.

2) Cartographie des épimutations

Une des possibilités pour identifier ces épimutations est de les cartographier comme s'il s'agissait de mutations. Pour y parvenir, une plante de fond Col présentant le phénotype floral est croisée avec une plante ayant un fond génétique différent. L'idée est d'utiliser le polymorphisme entre les deux accessions pour cartographier les épimutations associées au phénotype. Ainsi, si après ce croisement le phénotype est retrouvé en F2, des marqueurs moléculaires classiques pourront être utilisés pour réaliser cette cartographie en associant ces marqueurs avec le phénotype.

Dans ce but, j'ai réalisé un croisement entre la plante F3 5.32.18, qui présentait un phénotype floral de 38% et était issue de la plante F2 5.32, avec une plante d'un fond génétique Landsberg *erecta* (*Ler*) (**Figure V-7**). Un phénotypage a été réalisé sur 1 000 plantes F2 issues de ce croisement. Parmi elles, 54 plantes qui présentaient une proportion importante de fleurs à 5 pétales ont été génotypées. La cartographie des épimutations a été réalisée sur ces plantes à l'aide de 26 marqueurs microsatellites répartis sur l'ensemble des chromosomes, tous les 5 à 7 Mb environ.

Les résultats du génotypage des F2 montrent qu'il existe 4 régions qui présentent un biais par rapport à la fréquence attendue en F2. En effet, de grandes régions aux chromosomes 1, 2 et 3, ainsi qu'une région plus réduite au chromosome 5, présentent une fréquence allélique significativement différente de la fréquence attendue dans le cas d'une ségrégation mendélienne. L'analyse des fréquences de chaque allèle montre que les régions aux chromosomes 1 et 5 sont en faveur de l'allèle *Ler*, alors que les régions aux chromosomes 2 et 3 sont en faveur de l'allèle Col (**Figure V-8**). Le fait que plusieurs régions aient pu être mises en évidence lors de la cartographie est un argument supplémentaire en faveur de l'implication de plusieurs locus dans le phénotype observé.

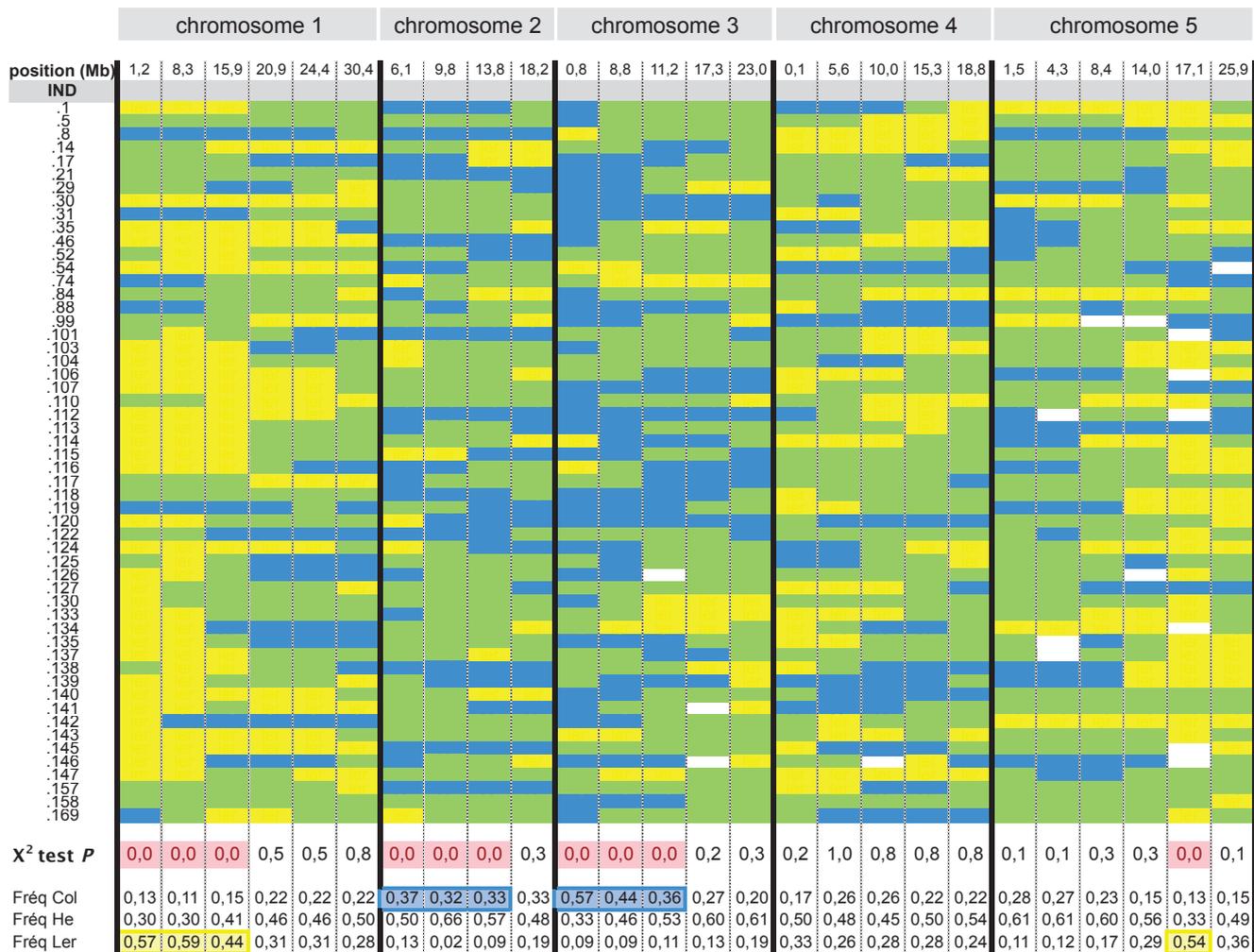


Figure V-8 : Génotypage des plantes F2 issues du croisement entre *Ler* et la plante F3 3.32.18, sur l'ensemble des chromosomes.

Les colonnes correspondent aux différents marqueurs utilisés pour le génotypage des F2 (Voir Figure V-7), et les couleurs correspondent au génotype associé : Col (bleu), hétérozygote (vert) ou *Ler* (jaune). Chaque ligne représente une plante F2 qui présente du phénotype au niveau des fleurs. Le seuil pour le phénotype a été fixé à 10% de fleurs avec un nombre de pétales surnuméraires. Les résultats du test du Chi2 réalisé sont surlignés en rouge lorsque la fréquence observée pour les génotypes est significativement différente de la fréquence attendue pour un marqueur donné, avec un risque d'erreur de 5%. Pour ces marqueurs qui présentent un biais, les fréquences sont entourées de couleur pour l'allèle favorisé.

Les épimutations ont été induites par la mutation *sg1* dans le fond génétique Col-0, ce qui laisse à penser qu'elles sont vraisemblablement portées par l'allèle provenant du parent Col-0. Ainsi, je me suis plus particulièrement intéressée aux régions situées au niveau des chromosomes 2 et 3. J'ai ajouté des marqueurs moléculaires supplémentaires autour de ceux qui mettaient en évidence le biais de ségrégation pour l'allèle Col-0 afin de délimiter plus précisément les régions candidates. Cinq marqueurs ont ainsi été ajoutés au chromosome 2 et un au chromosome 3.

Les résultats de cette nouvelle cartographie montrent que les marqueurs ajoutés présentent également un biais significatif en faveur de l'allèle Col-0 (**Figure V-9**). L'ajout de ces marqueurs a permis de réduire les régions candidates de 7,7 Mb au chromosome 2 et de 11,2 Mb au chromosome 3. Ces régions contiennent respectivement 2 155 et 3 393 gènes.

Malgré l'ajout de nouveaux marqueurs, je n'ai pas pu réduire significativement les régions candidates. Ceci est dû notamment au nombre relativement faible de plantes génotypées. D'après la taille de ces régions d'intérêt et le nombre important de gènes qu'elles contiennent, l'approche que j'ai réalisée par cartographie ne peut pas permettre à elle seule, avec le nombre de plantes que j'ai étudiées, l'identification d'épiallèles par une approche gène candidat.

3) Identification des épimutations par séquençage

Afin de compléter les résultats obtenus par cartographie et dans le but d'identifier les épimutations associées au phénotype observé, une approche par séquençage du génome après conversion bisulfite a été réalisée en parallèle.

a. Matériel biologique utilisé

Le séquençage a été réalisé sur des bouquets floraux issus de différentes plantes F4 à phénotype fort, mais également, comme contrôle, sur des bouquets floraux de différentes plantes F4 ayant un phénotype faible. J'ai réalisé deux ensembles indépendants, appelés P1 et P2, de plantes à phénotype fort (de 41,3% et 42,5% de phénotype en moyenne) provenant de la population F4 issue de la F3 3.39.21 (36% de phénotype). De la même façon, j'ai réalisé

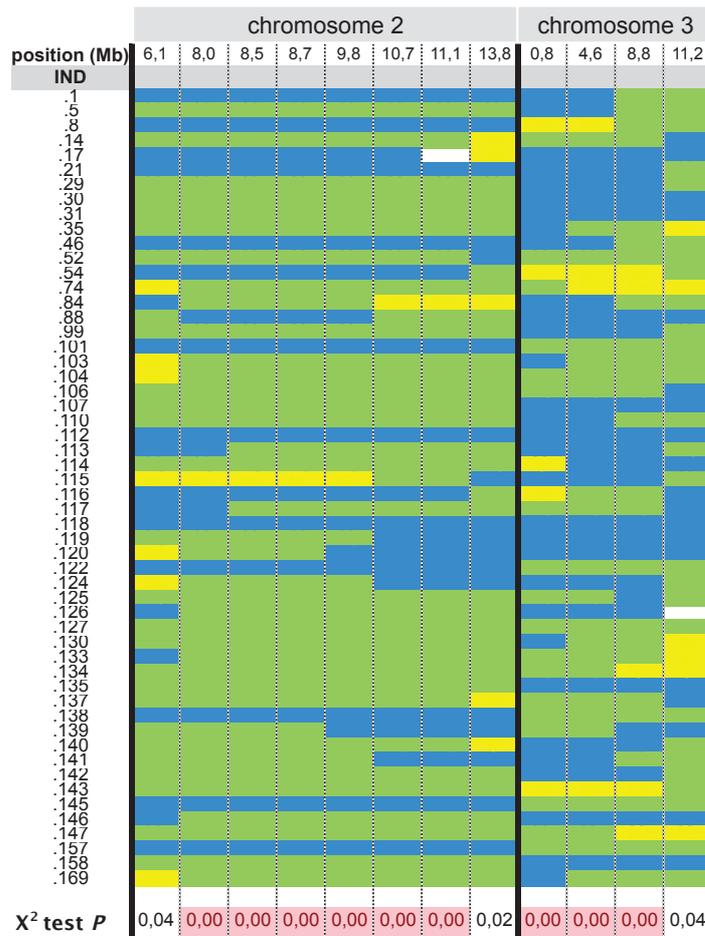


Figure V-9 : Cartographie fine aux chromosomes 2 et 3.

Pour le génotypage des F2 qui présentent un phénotype au niveau des fleurs, des marqueurs ont été ajoutés dans les régions des chromosomes 2 et 3 qui présentaient un fort biais en faveur de l'allèle Col (Figure V-8). Les résultats du test du X² réalisé sont surlignés en rouge lorsque la fréquence observée pour les génotypes est significativement différente de la fréquence attendue pour un marqueur donné, avec un risque d'erreur de 1%.

Nom du bulk	Moyenne phéno.	Phéno. min	Phéno. max	Médiane phéno.
N1	1,8	1,0	3,0	2,0
N2	1,6	0,0	3,0	1,5
P1	42,5	17,0	67,0	40,0
P2	41,3	30,0	82,0	34,0

Tableau V-4: Phénotypage des bulks de plantes F4 réalisés pour le séquençage.

Les valeurs de phénotype sont exprimées en pourcentage de fleurs anormales. Sont indiquées les pourcentages de fleurs anormales de la plante avec le plus faible et le plus fort phénotype, ainsi que la moyenne et la médiane de ces valeurs.

deux ensembles indépendants, N1 et N2, de plantes F4 à phénotype faible (de 1,8% et 1,5% de phénotype en moyenne) issues de la F3 18.45.18 (14% de phénotype) (**Tableau V-4**). P1 et P2, ainsi que N1 et N2, sont donc deux réplicats biologiques.

Le choix de procéder à des ensembles de plantes a été fait pour optimiser l'identification des épimutations responsables du phénotype observé. Même si la mutation *sg1* entraîne l'apparition d'une hyperméthylation CHG dans de nombreux gènes, nous ne nous attendions pas à avoir conservé autant d'épimutations non liées au phénotype floral. Ainsi, le fait de regrouper et analyser plusieurs plantes nous permettait d'enrichir l'ensemble en épimutations d'intérêt mais également de réduire les épimutations non liées au phénotype, qui sont non sélectionnées et donc qui ont plus de chance de ségréger. Ceci nous permettait de minimiser le bruit de fond causé par ces épimutations non liées au phénotype.

Des analyses d'expression de *IBM1* dans ces deux populations F4 3.39.21 et 18.45.18, qui sont pourtant différentes en termes de phénotype, montrent qu'elles ne diffèrent pas d'une population sauvage Col-0 en termes d'accumulation de transcrits *IBM1-L* (**Figure V-10**). Ce résultat indique que dans les populations F4 que nous avons séquencées, la fonction de *IBM1* est complètement restaurée. Les éventuels épiallèles identifiés lors de l'analyse du méthylome de ces épimutants en génération F4 sont des épiallèles établis lors des générations précédentes qui ont été maintenus et pas des régions hyperméthylées dues à un défaut de fonction de *IBM1*.

b. Séquençage et alignement

Les préparations d'ADN génomique de N1, N2, P1 et P2 ont été envoyées, selon leurs recommandations, au Beijing Genomics Institute (BGI) qui a réalisé le traitement pour la conversion au bisulfite, la préparation de la librairie et le séquençage du génome entier (avec une profondeur de 30X). Le système Illumina utilisé pour le séquençage consiste en un séquençage par synthèse sur HiSeq2000 après une amplification des fragments d'ADN dans les deux sens.

Les outils Bismark et Bowtie2 ont été utilisés pour l'alignement des reads. Bowtie2 permet l'alignement de séquences d'ADN de petite taille et Bismark permet de gérer les séquences

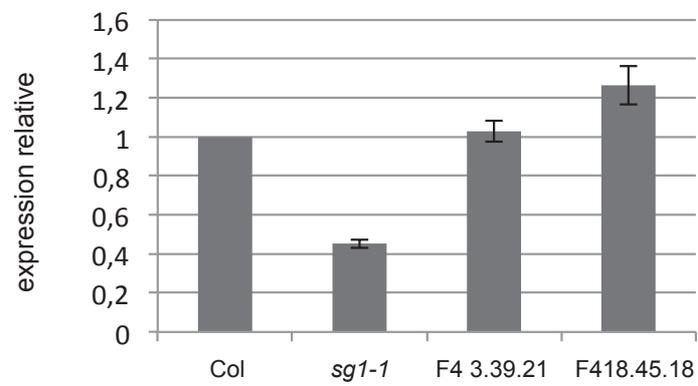


Figure V-10 : Analyse d'expression de *IBM1-L* dans les populations F4 d'épimutants.
Expression relative par rapport à Col du mutant *sg1-1* et d'un bulk de plantes (15 à 20 plantes) pour les des deux populations F4 3.39.21 et 18.45.18 par PCR quantitative.

converties au bisulfite pour les aligner et ainsi déterminer la méthylation des cytosines dans les différents contextes.

Les séquences des fragments d'ADN, appelés « *reads* » sont alignées grâce à ces deux outils qui procèdent à quatre étapes d'alignement. Deux de ces étapes consistent à aligner les *reads* tels qu'ils ont été séquencés à la fois sur le génome d'*Arabidopsis thaliana* disponible sur TAIR10 et sur ce même génome converti virtuellement au bisulfite, c'est à dire avec toutes les cytosines converties en thymines. Les deux autres étapes consistent à aligner sur ces deux génomes la version convertie de ces mêmes *reads*, ce qui signifie que toutes les cytosines séquencées des *reads* sont converties en thymines. Ainsi, chaque *read* séquencé est aligné quatre fois. Seuls les *reads* uniques sont conservés, c'est à dire ceux qui n'ont qu'un seul meilleur score à l'issue de ces quatre alignements et qui, par conséquent, ne s'alignent qu'à un seul endroit dans le génome. Lorsque plusieurs alignements pour un *read* donné présentent le même score ou avec un nombre identique de mésappariements, le *read* est éliminé.

Les fichiers des *reads* pour N1, N2, P1 et P2 générés par Bismark, ainsi que les données de méthylome de Col et *sgl-1* (cf. **Partie II**) ont été visualisés et analysés grâce à l'outil SeqMonk.

c. *Analyse des données de méthylome*

Dans la mesure où nous nous intéressons à des épimutations qui apparaissent dans le fond mutant *sgl* qui est responsable d'une hyperméthylation en contexte CHG dans le corps des gènes, nous avons regardé spontanément la méthylation CHG dans le corps des gènes.

La comparaison des pourcentages de méthylation CHG dans les gènes pour les groupes N1 et N2, ainsi que P1 et P2 deux à deux, montre qu'il n'y a pas de différence significative entre les deux répétitions biologiques (**Figure V-11, A et B**). Il en est de même pour la méthylation CG retrouvée dans le corps des gènes (**Figure V-11, C et D**). En effet, l'ensemble des gènes est globalement méthylé de la même manière, dans une répétition biologique comparée à l'autre, ce qui suggère qu'elles sont homogènes. Ensuite, les données qui correspondent aux différentes répétitions biologiques P1 et P2 ainsi que N1 et N2 ont été regroupés et celles-ci sont appelées respectivement P et N par la suite.

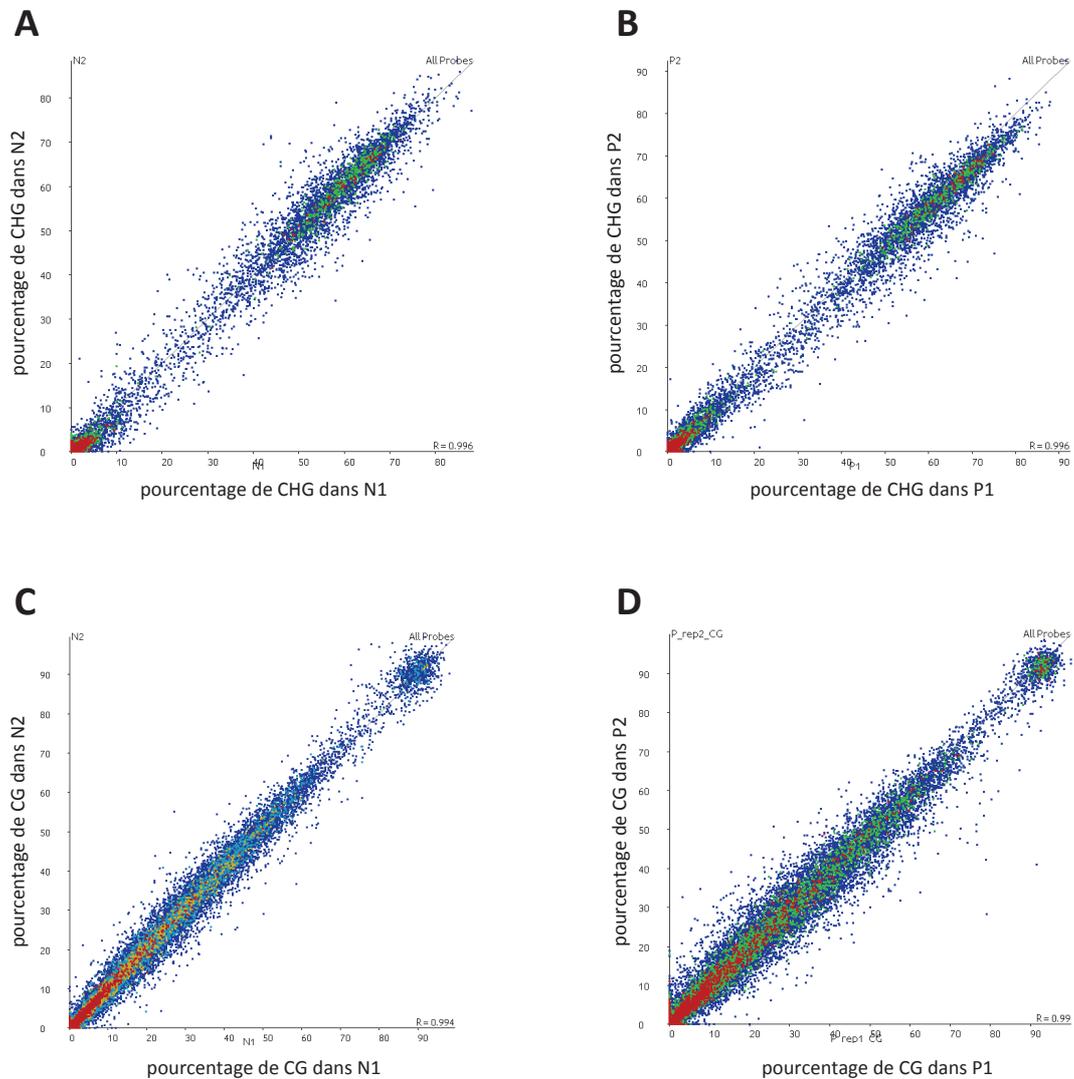


Figure V-11 : Comparaison des méthylomes entre répétitions biologiques.

Le scatter plot représente le pourcentage de méthylation de chaque gène dans les fonds génétiques N1 et N2, et P1 et P2. Chaque point représente un gène et les zones colorées correspondent à la densité de points dans cette zone, du bleu, au rouge pour la densité la plus forte. La diagonale matérialise la zone théorique où la méthylation est la même dans les deux fonds génétiques. A et B : Comparaison du pourcentage de méthylation CHG entre N1 et N2, et P1 et P2 respectivement. C et D : Comparaison du pourcentage de méthylation CG entre N1 et N2, et P1 et P2 respectivement.

La comparaison des méthylomes de N, qui correspond aux plantes avec un phénotype faible, et Col-0 d'un point de vue de la méthylation en contexte CHG dans les gènes montre que certaines épimutations ont été maintenues. En effet, le décalage par rapport à la diagonale du nuage de points sur le graphique correspondant indique qu'un certain nombre de gènes présente un pourcentage de méthylation plus élevé dans N que dans Col-0 (**Figure V-12, A**). En effet, l'analyse des données montre que 384 gènes sont plus méthylés dans N que dans Col-0, avec un seuil fixé à une différence de méthylation de cinq fois.

De la même manière, les résultats du méthylome de P, qui correspond à l'ensemble de plantes à phénotype fort, montrent que des épimutations ont été maintenues, comme l'illustre la comparaison avec le méthylome de Col-0 en contexte CHG (**Figure V-12, B**). Ceci suggère que la méthylation dans un contexte CHG a été maintenue dans les deux groupes de plantes N et P, en l'absence de la mutation *sg1*, et pendant au moins deux générations.

La comparaison du méthylome de P, ainsi que celui de N, avec le méthylome du mutant *sg1* montre qu'un grand nombre d'épimutations sont perdues par rapport au parent *sg1*. En effet, de nombreux gènes sont davantage méthylés dans le fond mutant *sg1* que dans les épimutants, comme l'illustre la densité de points observable sur l'axe des ordonnées et qui correspondent à des gènes méthylés dans le mutant *sg1* et non méthylés dans N ou P (**Figure V-13**). En tout, 4 791 et 3 899 gènes sont au moins cinq fois plus méthylés dans *sg1* que dans N et P respectivement. D'autre part, le décalage du nuage de points par rapport à la diagonale suggère un enrichissement en terme de méthylation de certains gènes dans N et P par rapport à *sg1* (**Figure V-13**). Les analyses des données de méthylome ont montré que 34 et 44 gènes sont davantage méthylés dans N et P respectivement, que dans *sg1-1*. Ceci suggère que certains épiallèles non parentaux ont pu apparaître.

Il n'y a pas de différence majeure en termes de méthylation en contexte CG des gènes entre N et P (**Figure V-14, A**). Ce résultat était attendu puisque la comparaison des méthylomes de Col et *sg1* ne montre pas non plus de variation significative de la méthylation CG au niveau des gènes (**Cf Figure S8A** (Coustham et al., 2014)).

Selon l'hypothèse qui a été faite que le phénotype floral observé est causé par une ou plusieurs épimutations induites par la mutation *sg1* et dans l'intention d'isoler et d'identifier

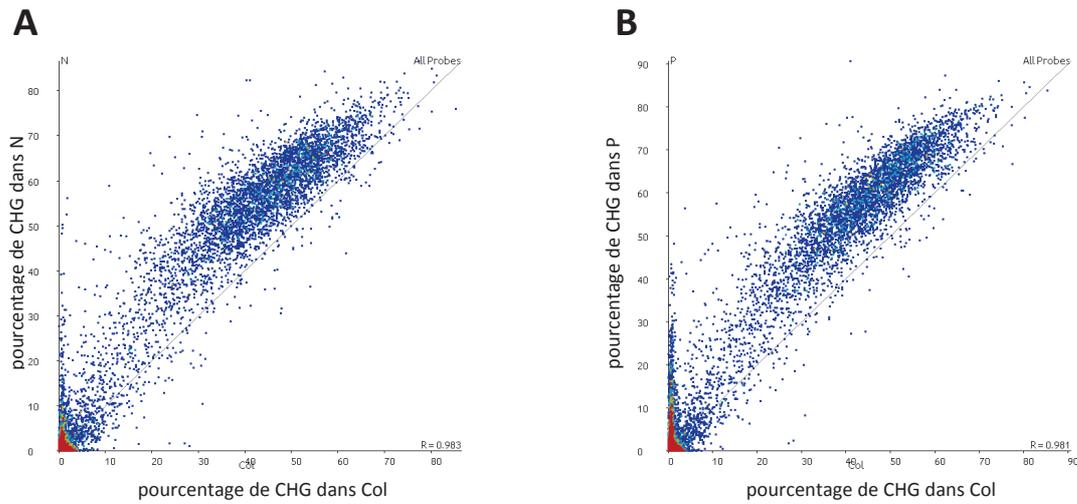


Figure V-12 : Comparaison des méthylomes des épimutants avec Col.

A : Comparaison du pourcentage de méthylation CHG entre N et Col. B : Comparaison du pourcentage de méthylation CHG entre P et Col.

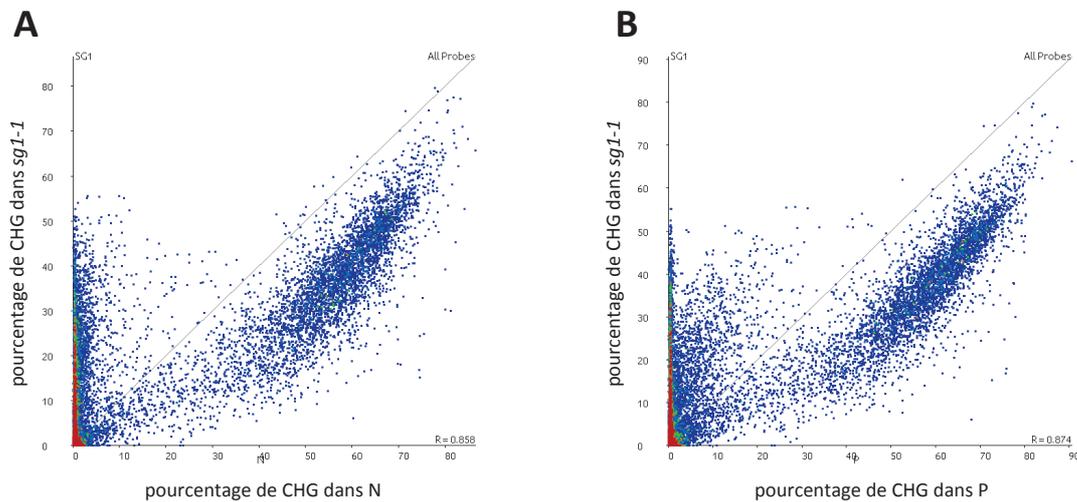


Figure V-13 : Comparaison des méthylomes des épimutants avec celui de *sg1-1*.

A: Comparaison du pourcentage de méthylation CHG entre N et *sg1-1*. B: Comparaison du pourcentage de méthylation CHG entre P et *sg1-1*.

ces épimutations, la comparaison de la méthylation des gènes en contexte CHG des deux groupes P et N, c'est-à-dire les plantes avec et sans phénotype, est la plus informative. Cette comparaison de P et N suggère qu'il n'y a pas d'enrichissement global en méthylation CHG dans l'un ou l'autre des ensembles (**Figure V-16**). En effet, comme illustré sur le graphique en nuage de points, une grande majorité des points sont regroupés autour de la diagonale. De plus, le compte des cytosines méthylées dans chaque contexte ne montre pas de différence importante à l'échelle globale entre les ensembles N et P (**Figure V-15**). Pourtant, quelques gènes semblent, sur le graphe, être davantage méthylés dans P que dans N en contexte CHG (**Figure V-16**).

d. Recouplement des données de cartographie/méthylome : approche gène candidat

En fixant un seuil d'une méthylation au moins cinq fois plus élevée dans P que dans N, 588 gènes ont pu être identifiés comme davantage méthylés dans P que dans N. Parmi ces gènes, 35 sont compris dans l'intervalle de 7,7 Mb du chromosome 2 et 229 dans l'intervalle de 11,2 Mb du chromosome 3.

J'ai réalisé une recherche à partir des annotations de ces 264 gènes disponibles d'après la dernière annotation du génome de *Arabidopsis* (TAIR10), et 8 d'entre eux sont identifiés comme jouant un rôle dans le développement de la fleur (**Tableau V-5**).

Parmi ces 8 candidats, le gène *PHOTOPERIOD-INDEPENDENT EARLY FLOWERING1* (*PIE1*) est annoté comme régulateur de la formation et du développement du pétale. Le gène *PIE1* code une protéine similaire à des remodeleurs de la chromatine de la famille SWI2/SNF2 qui ont pour fonction de changer l'architecture de la chromatine soit en modifiant la structure du nucléosome, soit en substituant certaines histones par des variants. Il a été montré que *PIE1* interagit avec d'autres protéines pour former un complexe impliqué dans l'incorporation du variant d'histone H2A.Z à certains locus (Choi et al., 2007). La protéine *PIE1* est connue pour réguler négativement la floraison par activation de *FLC* (Noh and Amasino, 2003). En effet, elle permet le remplacement de l'histone canonique H2A par l'histone H2A.Z nécessaire à l'expression de *FLC* (Deal et al., 2007). Divers phénotypes ont pu être observés dans un mutant *pie1* comme une floraison précoce, des feuilles aux bords

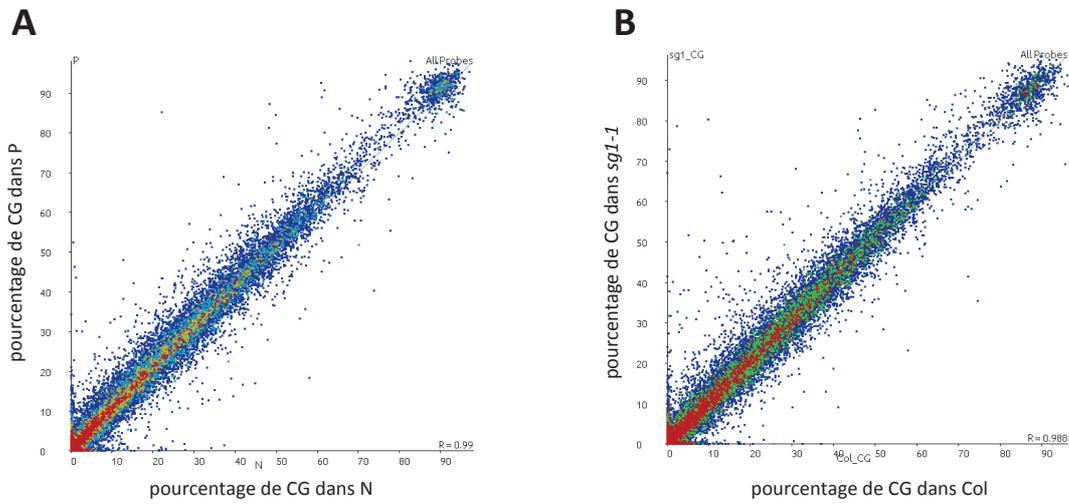


Figure V-14 : Comparaison des méthylomes en contexte CG.

A : Comparaison du pourcentage de méthylation CG entre N et P. B : Comparaison du pourcentage de méthylation CG entre Col et *sg1-1*.

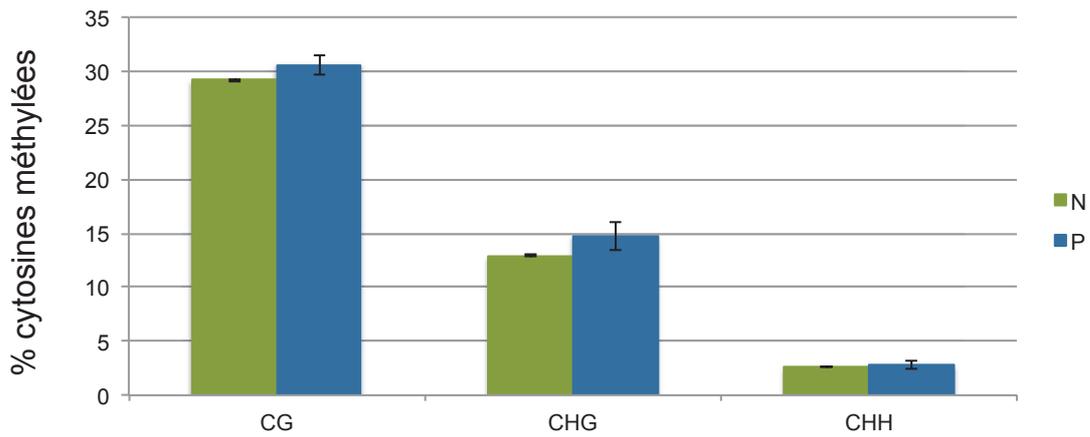


Figure V-15 : Pourcentage de méthylation des cytosines dans les trois contextes pour N et P.

L'analyse globale des cytosines a permis de déterminer le pourcentage de méthylation (nombre de cytosines méthylées/ nombre de cytosines totales) pour N (moyenne de N1 et N2) et P (moyenne de P1 et P2) en contexte CG, CHG et CHH.

dentelés, une faible dominance apicale et une production de pétales surnuméraires (Noh and Amasino, 2003).

A titre d'exemple, en s'intéressant plus particulièrement à la méthylation CHG au niveau du locus *PIE1* dans les différents fonds étudiés, j'ai pu observer qu'il fait partie des gènes hyperméthylés dans *sgl*. Par ailleurs, les profils de méthylation CHG dans N et dans P sont différents : La méthylation dans P, le groupe de plantes avec phénotype, ressemble à ce qui peut être observé dans *sgl* alors que le profil de méthylation de N, le groupe de plantes sans phénotype, est plus proche du profil parental Col-0 (**Figure V-17**). Le nombre de pétales observé dans un fond mutant *piel* et les profils de méthylation au niveau du gène pourraient en faire un bon candidat en tant qu'épiallèle causal du phénotype, mais d'autres phénotypes associés, comme la structure très ramifiée due à la perte de la dominance apicale, ne sont pourtant pas retrouvés dans les populations d'épimutants suivies.

L'étude plus approfondie de ces différents gènes candidats, notamment à travers des études d'expression et l'analyse de mutants en terme de phénotype, sera nécessaire avant de déterminer si l'un d'entre eux est un épiallèle associé au phénotype floral qui a été suivi.

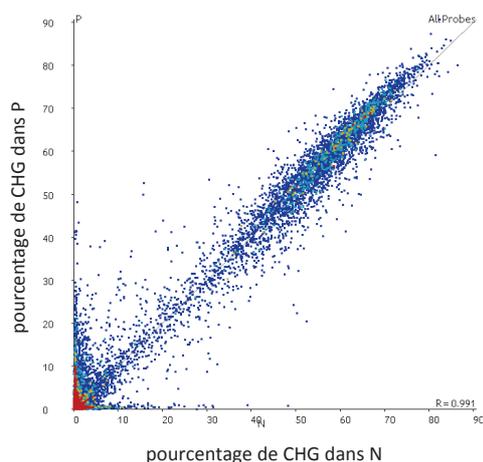


Figure V-16 : Comparaison des méthylomes des épimutants à faible et fort phénotype, N et P.

Comparaison du pourcentage de méthylation CHG des gènes entre N et P.

ID	Gène	GO	Rôle de la protéine	Phénotypes des mutants décrits
AT3G01460	METHYL-CPG-BINDING DOMAIN 9 (MBD9)	Photopériode, floraison	Réprime <i>FLC</i> et est impliqué dans la méthylation de l'ADN et l'acétylation de histones	Floraison précoce et importante ramification des parties aériennes.
AT3G02780	ISOPENTENYL PYROPHOSPHATE2 (IPP2)	Développement de la fleur		
AT3G05040	HASTY (HST)	Régulation négative du développement de la fleur	Membre de la famille des importines/exportines. Rôle dans le transport des microARN	
AT3G11540	SPINDLY (SPY)	Développement de la fleur	Impliqué dans la glycosylation de molécules de la voie de signalisation des gibbérellines.	
AT3G12810	PHOTOPERIOD-INDEPENDENT EARLY FLOWERING1 (PIE1)	Régulation négative du développement de la fleur. Formation, développement du pétale		Floraison précoce, feuilles crénelées, pétales surnuméraires et perte de la dominance apicale.
AT3G12980	HISTONE ACETYLTRANSFERASE OF THE CBP FAMILY5 (HAC5)	Développement de la fleur		
AT3G19980	FLOWER-SPECIFIC PHYTOCHROME-ASSOCIATED PROTEIN PHOSPHATASE3 (FYPP3)	Régulation négative du développement de la fleur		
AT3G24870	ATEAF1B	Régulation de la photopériode, floraison	inconnu	inconnu

Tableau V-5: Récapitulatif des gènes candidats dans la région du chromosome 3 réduite par cartographie.

Les huit gènes candidats ont été identifiés dans la région du chromosome 3 en fonction de leur pourcentage de méthylation (5 fois plus méthylé dans P, avec phénotype, que dans N, sans phénotype) et leur annotation comme gènes impliqués dans le développement de la fleur. Aucun gène candidat n'a pu être identifié de cette manière dans la région du chromosome 2. Ce tableau récapitule les indications sur le GO (gene ontology), le rôle et le phénotype des mutants décrits, d'après les données disponibles dans la littérature.

(IV) Discussion et perspectives

Ce projet portant sur l'étude de populations d'épimutants m'a permis, d'une part, de montrer que, pendant au moins quelques générations, des épimutations peuvent être maintenues et transmises à la génération suivante, indépendamment de l'élément causal. D'autre part, une proportion de ces épimutations induites dans un fonds génétique *sgl*, qui sont par définition labiles, sont perdues spontanément. Les approches que j'ai utilisées n'ont pas encore permis d'identifier clairement une ou des épimutations associées au phénotype observé, à savoir des pétales surnuméraires, mais j'ai pu identifier un certain nombre de gènes candidats qui devront, par la suite, être étudiés davantage avant de pouvoir les valider.

Des épimutations peuvent être induites par la perte de fonction de IBM1

La protéine SG1, avec ses domaines BAH et RRM prédits, n'a pas d'activité démontrée dans la modification directe de la chromatine. Ainsi, tout laisse à penser que les épimutations induites dans un fond génétique *sgl* sont provoquées par la dérégulation de *IBM1* dans le mutant *sgl*, d'autant plus que *IBM1*, avec son domaine JmjC, est connu pour modifier la chromatine en retirant les marques H3K9me2 au niveau de locus cibles. De plus, une construction contenant l'ADNc de *IBM1-L* peut compléter un mutant *sgl* (Coustham et al., 2014), ce qui implique que le phénotype de *sgl* est causé principalement par la perte de fonction de *IBM1*. Même si l'ensemble des cibles de SG1 reste encore inconnu, la comparaison des données de méthylome de *ibm1* et de *sgl* suggère fortement que les épimutations observées dans *sgl* sont uniquement causées par *ibm1*. Par conséquent, les populations d'épimutants que j'ai étudiées au cours de ce projet correspondraient à des populations d'épimutants induits par la mutation *ibm1*.

Dans l'hypothèse où ces épimutations sont induites par une perte de la fonction *IBM1* dans un fond mutant *sgl*, la question suivante se pose : est-ce que la ségrégation de la mutation *sgl* suffit à restaurer la fonction de *IBM1* et dans quel délai ? Autrement dit, les épimutations identifiées en F4 sont-elles stables ou sont-elles encore maintenues par un défaut de fonction de *IBM1* ? La mutation *sgl* étant récessive, lorsque *sgl* est à l'état hétérozygote, *IBM1* est fonctionnel. Ainsi, dès la F1 issue du croisement entre Col et *sgl*, la fonction de *IBM1* peut théoriquement être restaurée. J'ai pu montrer au cours de ce projet qu'au sein des populations

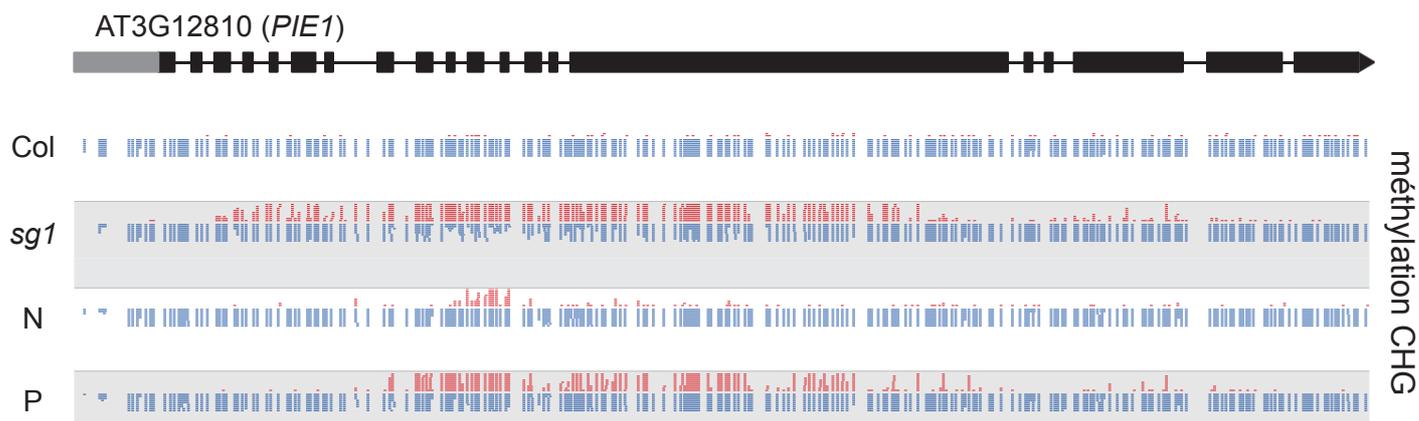


Figure V-17 : Profil de méthylation des cytosines au locus *PIE1*.

Est représentée la méthylation en contexte CHG pour les fonds Col, *sg1*, N (population d'épimutants sans phénotype) et P (population d'épimutants avec phénotype). Les points bleus correspondent aux cytosines des *reads* totaux (méthylés et non méthylés) qui ont été alignés et les points rouges correspondent aux cytosines des *reads* méthylés.

F4 d'épimutants, le niveau de transcrits fonctionnels *IBM1-L* est le même que celui du sauvage, indiquant donc que la fonction de *IBM1* est restaurée à cette génération. La génération F2 après le croisement entre les épimutants (en génération F3) et *Ler* se situe quatre générations après la restauration théorique de *IBM1*, et le phénotype peut encore être observé. Ceci est un argument en faveur de la stabilité des épiallèles, qui suggère qu'ils ne sont pas seulement maintenus par un défaut de fonction de *IBM1*. De plus, j'ai pu démontrer que dans les populations F4 3.39.21 et 18.45.18, le niveau de transcrit *IBM1-L* est de l'ordre de ce qui est observé dans une plante sauvage Col-0 (**Figure V-10**). Ce résultat suggère que trois générations après avoir ségrégé la mutation *sgl*, *IBM1* est pleinement fonctionnel. Il serait pourtant intéressant de déterminer s'il est à nouveau fonctionnel dès la F2 ou si pour certaines plantes la fonction de *IBM1* met plus de générations à se rétablir.

La pléiotropie du phénotype des mutants *sgl* et *ibm1* rend l'étude de la stabilité des épiallèles hyperméthylés en contexte CHG difficile. En ségrégant l'élément causal, il est possible de ségréger aussi les différents aspects du phénotype et par conséquent, les éventuels épiallèles sous-jacents.

Le phénotype floral suivi est-il lié à des épimutations ?

Le phénotype floral qui a été suivi est un phénotype que j'ai pu observer dans le fond *sgl*. Pourtant, des pétales surnuméraires sont plus rarement observables dans *sgl*, sans doute parce que le mutant présente un phénotype sévère, avec un grand nombre de défauts développementaux, dont des défauts de nombre et d'identité des organes floraux qui s'établit rapidement au cours des générations. Il est possible que ces malformations sévères masquent le défaut du nombre de pétales. Dans l'hypothèse où le phénotype pléiotrope du mutant est causé par de nombreuses épimutations, la réintroduction des épiallèles sauvages par croisement avec Col-0, permet, dès la F2 la ségrégation des épimutations et donc des potentiels phénotypes associés. Le phénotype mutant est ainsi décomposé. D'autant plus que l'analyse des méthylomes a permis de montrer que certaines épimutations, labiles, sont perdues spontanément. Ainsi, il n'est pas surprenant, dans la mesure où je me suis intéressée particulièrement à ce phénotype, que le phénotype « pétales surnuméraires » soit plus fréquemment observé dans les populations d'épimutants que le fond *sgl* puisque je l'ai sélectionné.

Dans l'hypothèse où le phénotype « pétales surnuméraires » est dû à une ou plusieurs épimutations qui ont pour origine le parent *sg1*, comment expliquer alors la présence de ce phénotype dans une population sauvage Col-0 ? Il est possible qu'une épimutation responsable du phénotype apparaisse spontanément dans une population sauvage, comme cela été montré dans des populations naturelles d'*Arabidopsis thaliana* (Becker et al., 2011; Schmitz et al., 2011). De plus, des liens possibles entre l'apparition du phénotype et les conditions environnementales ne sont pas à exclure.

Le fait que le phénotype observé puisse être transmis d'une génération à l'autre soulève une question essentielle. Comment affirmer que l'élément causal du phénotype observé est d'ordre épigénétique et pas seulement génétique, comme par exemple une mutation ou un élément transposable qui aurait été réactivé ? Les populations d'épimutants que j'ai générées sont issues de croisements indépendants avec deux allèles différents de *sg1*, à savoir un mutant ponctuel et un mutant d'insertion T-DNA. Puisque l'on observe le phénotype dès la F2, si celui-ci est d'origine génétique, cela signifie que l'événement causal est antérieur, soit parental, soit apparu dans la F1. La probabilité qu'un tel événement se produise dans deux populations indépendantes est très faible. De plus, les épimutations induites par *sg1* sont, en théorie, hyperméthylées en contexte CHG et essentiellement dans le corps des gènes, ce qui rend peu probable une remobilisation de transposons. En effet, il a été montré dans un mutant *ibm1* que la méthylation en contexte CHG au niveau des transposons, ainsi que la marque répressive H3K9me2, ne sont pas affectées (Miura et al., 2009; Saze et al., 2008). Si un des gènes candidats est validé, le séquençage de la région encadrant le gène dans différentes populations permettra d'exclure la cause génétique.

Il est difficile de déterminer combien de locus peuvent être impliqués dans l'apparition de pétales surnuméraires. En effet, l'organogenèse du pétale est sous le contrôle d'un réseau complexe de gènes (Irish, 2008). Cependant, la distribution des différentes populations en terme de phénotype et la ségrégation d'une génération sur l'autre, laissent suspecter un caractère quantitatif. La cartographie réalisée en croisant ces épimutants avec *Ler* montre qu'il y a des biais de ségrégation sur différents chromosomes. La taille importante de ces régions biaisées est due au faible nombre de plantes qui présentent un phénotype utilisées pour la cartographie. Pourtant, il n'est pas exclu que la réduction de l'un de ces intervalles puisse faire apparaître plusieurs régions qui présentent un biais de ségrégation, révélant ainsi la présence de plusieurs éléments causaux sous-jacents.

Concernant l'approche par cartographie, d'autres marqueurs polymorphes auraient pu être ajoutés, mais avec une population de seulement 54 plantes qui présentent un phénotype, cela n'aurait sans doute pas permis de réduire davantage les intervalles. Le meilleur moyen de réduire les régions qui ont montré un biais de ségrégation, et ainsi réduire le nombre de gènes candidats, serait d'ajouter des plantes à la population de génotypage, pour avoir par conséquent davantage de recombinants dans ces intervalles.

Héritabilité des épimutations induites

Le séquençage des populations d'épimutants après conversion bisulfite montre que de nombreuses épimutations sont maintenues en génération F4. En effet, il reste encore des gènes qui sont davantage méthylés dans les populations d'épimutants comparé à Col, que ce soit pour les plantes avec ou sans phénotype, et la majorité de ces épiallèles ne sont probablement pas liés au phénotype suivi. De plus, la comparaison des données de méthylome montre que 588 gènes sont plus méthylés dans P que dans N. Parmi ces épiallèles, tous ne sont pas non plus liés au phénotype observé. Les plantes en génération F2 montrent peu de phénotype ce qui pouvait suggérer un faible nombre d'épimutations conservées. Dans la mesure où des modifications de l'état de méthylation des corps des gènes en contexte CHG ne sont pas systématiquement corrélées avec des changements d'expression, comme cela a été montré dans le mutant *sgl*, il n'est pas surprenant d'avoir un tel nombre d'épiallèles qui sont conservés en génération F4 (Coustham et al., 2014). En effet, il est possible que ces épiallèles soient conservés parce qu'ils n'ont pas de conséquences phénotypiques.

Malgré un nombre non négligeable d'épiallèles conservés, la comparaison des méthylomes de P et N avec *sgl* montre bien qu'une partie d'entre eux reviennent spontanément vers un niveau de méthylation proche du sauvage. A quel moment ces épiallèles reviennent à leur état sauvage, comment y reviennent-ils, en une génération ou progressivement, et y reviennent-ils tous de la même manière, sont autant de questions qu'il reste à examiner. Ce qu'il se passe au niveau des gènes différentiellement méthylés en F1, lorsque les épigénomes sauvage et mutant *sgl* se rencontrent, est également important à déterminer pour comprendre comment ces épiallèles sont formés et maintenus. Une étude s'intéressant aux profils de méthylation dans les plantes issues d'un croisement entre les accessions d'*Arabidopsis thaliana* C24 et *Ler* montre que dans les hybrides F1 certains allèles méthylés provenant d'un parent peuvent

convertir les allèles non méthylés de l'autre parent, et inversement, et ces changements sont opérés par la voie du RdDM et les petits ARN de 24 nucléotides (Greaves et al., 2014). Il est alors possible que dans nos populations d'épimutants une partie des épiallèles d'origine parentale, qui reste à déterminer, soit fixée dès la génération F2.

Le phénotype que j'ai suivi, ainsi que les éventuelles épimutations sous-jacentes, peuvent être maintenus, comme je l'ai montré, au moins pendant quatre générations. Cependant, le phénotype est volontairement sélectionné. Dans les populations épIRIL basées sur les mutations *ddm1* ou *met1* qui ont permis de mettre en évidence des épiallèles stables, les populations n'étaient pas soumises à une sélection, seulement propagées sur huit générations après ségrégation de la mutation causale (Johannes et al., 2009; Reinders et al., 2009). Que se passerait-il alors au niveau des épiallèles que nous avons identifiés en l'absence de sélection ? Les générations de ces épIRIL et de nos populations d'épimutants ont en commun le fait d'avoir été induits par une mutation (*sg1* ou *ddm1* ou *met1*). Ainsi, même s'ils ont été induits par une mutation, ces études montrent qu'une mutation génétique qui apparaîtrait spontanément dans un gène impliqué dans des processus de méthylation peut générer une variabilité épigénétique plus ou moins importante, même dans une population naturelle, offrant ainsi davantage de possibilités adaptatives.

Les lignées d'épIRIL dérivées de *met1* et *ddm1* montrent une variation phénotypique importante dans certaines lignées en termes de retard de floraison ou de croissance par exemple. La combinaison aléatoire de différents épiallèles pourrait être à l'origine de cette variabilité phénotypique. Nous avons pu observer, dans nos populations F2 ayant ségrégré la mutation *sg1*, différents phénotypes tels que des phénotypes de croissance, de forme foliaire ou de délai de floraison, mais dans la mesure où les plantes ont été sélectionnées d'après leur phénotype, une telle variation phénotypique après quelques générations n'a pas pu être étudiée.

D'après les analyses de méthylome que nous avons réalisées, de nombreux gènes présentent un état de méthylation intermédiaire entre ce qui est observé chez les parents Col-0 et *sg1*, et quelques uns seraient transgressifs par rapport aux profils parentaux. De la même façon, dans les populations d'épIRIL, des profils de méthylation non parentaux, soit intermédiaires, soit transgressifs, ont été mis en évidence. En effet, selon les lignées, jusqu'à un tiers des polymorphismes de méthylation de l'ADN dans les épIRIL induites par *ddm1* seraient non-

parentaux. Ceci peut s'expliquer par une épi-hétérozygotie qui est maintenue, par l'acquisition d'une capacité à passer d'un état épiallélique à l'autre ou par l'acquisition d'un nouvel état chromatinien.

Les épiallèles mis en évidence dans ces épIRIL sont hypométhylés en contexte CG par rapport à des allèles sauvages, et certains de ces épiallèles sont perdus car certains locus sont activement reméthylés. Qu'en est-il en terme de stabilité pour nos épiallèles hyperméthylés en contexte CHG et pourquoi certains sont perdus ? La méthylation CHG est connue pour être une marque répressive. Le mutant *ibm1*, qui présente une hyperméthylation dans le corps des gènes en contexte CHG, a des défauts développementaux importants (Miura et al., 2009; Saze et al., 2008). L'augmentation de la méthylation CHG est susceptible d'avoir des conséquences phénotypiques en modifiant l'expression des gènes ou la maturation des transcrits, par épissage ou polyadénylation alternatifs par exemple. Le mutant *sg1* a un phénotype sévère alors que, dans les populations d'épimutants, les phénotypes que présentent certaines plantes sont modérés. Est-ce que ceci est dû à la ségrégation des différentes épimutations ou y a-t-il préférentiellement une perte des épimutations qui ont les conséquences phénotypiques les plus sévères ? Le fait que ces épimutations puissent se transmettre de génération en génération suggère qu'il n'y a pas massivement et spontanément de déméthylation active. Mais il est possible que, parmi les épiallèles perdus, certains locus aient été progressivement déméthylés, passivement, en l'absence de maintien de la méthylation à ces locus et que d'autres soient spécifiquement contre-sélectionnés et subissent une déméthylation active. La déméthylation active des gènes chez *Arabidopsis thaliana* est initiée par ROS1 qui est lui-même régulé par la méthylation CG. En effet, dans un mutant *met1* qui présente une perte de la méthylation en contexte CG, la méthylation CG du locus *ROS1* est perdue et l'expression de *ROS1* est réprimée (Mathieu et al., 2007; Zhang et al., 2006). Il a été montré que dans un mutant *sg1*, l'expression de *ROS1* n'est pas altérée, ce qui suggère que ROS1 est actif et donc que sa fonction dans la déméthylation active n'est pas altérée (Wang et al., 2013). Il serait intéressant d'étudier davantage le ou les mécanismes qui dirigent cette déméthylation afin de déterminer si la déméthylation de ces épiallèles est un phénomène aléatoire et progressif ou si elle résulte d'un mécanisme actif qui cible préférentiellement certains locus.

(V) Conclusion

L'étude des populations d'épimutants induits par la mutation *sg1* m'a permis de montrer que le phénotype floral auquel je me suis intéressée dans le cadre de ce projet est très probablement lié à plusieurs épimutations. J'ai pu ainsi montrer que ces épimutations qui sont liées au phénotype floral, même si elles restent encore à identifier, peuvent être transmises de manière stable d'une génération à l'autre. J'ai réalisé une cartographie dans le but d'identifier ces épimutations et l'ai recoupée avec les données de méthylome obtenues pour les épimutants en génération F4, avec et sans phénotype floral. J'ai pu ainsi délimiter des régions aux chromosomes 2 et 3 qui contiennent encore un grand nombre de gènes. Pour réduire davantage le nombre de gènes dans ces régions, l'ajout d'individu à la population de cartographie sera nécessaire. Pourtant, par une approche de gènes candidats j'ai pu identifier 8 gènes impliqués dans le développement de la fleur. Des analyses d'expression de ces gènes dans les fonds génétiques parentaux et épimutants, et l'étude de mutants correspondants à ces gènes seront nécessaires pour affirmer l'implication de l'un ou de plusieurs d'entre eux dans le phénotype floral qui a été suivi.

Il reste de nombreuses interrogations concernant l'apparition, le maintien et la perte de ces épimutations. L'étude des gènes candidats que j'ai pu identifier pourra peut-être permettre de valider un ou plusieurs épiallèles. Si ces épiallèles, liés au phénotype suivi, sont validés, ils pourront être utilisés pour répondre à un certain nombre de ces questions, en particulier concernant le maintien de ces épiallèles en l'absence de sélection, comme c'est le cas dans une population naturelle.

(VI) Matériels et méthodes

Préparation des échantillons d'ADN envoyés au séquençage.

Les extractions d'ADN ont été réalisées sur des bulks de bouquets floraux issues de 5 à 10 plantes F4, à phénotype floral fort pour les échantillons P1 et P2 et sans phénotype floral pour les échantillons N1 et N2. Le poids des échantillons P1, P2, N1 et N2 a été mesurée : 0,22g, 0,42g, 0,18g et 0,26g respectivement.

Les extractions ont été réalisées à l'aide du kit « Maxi Kit Macherey Plant II » (Macherey-Nagel). Sauf indication contraire, les réactifs utilisés sont ceux inclus dans le kit et le protocole utilisé diffère de celui fourni par le fabricant. Les valeurs indiquées correspondent aux quantités et volumes nécessaires pour l'extraction de 1g de matériel végétal.

Le matériel est broyé sous forme de poudre à l'aide d'un mortier, préalablement refroidi à l'azote liquide, avec 500 mg de PVPP (PolyVinylPolyPyrrolidone; non fourni dans le kit).

Dans un tube de 50 mL, 3,5 mL de réactif PL2 et 67 μ L de RNase A sont ajoutés à la poudre et le mélange est homogénéisé au vortex avant incubation à 65°C pendant une heure.

Puis, 467 μ L de réactif PL3 sont ajoutés et l'ensemble est mélangé vigoureusement avant d'être placé 10 min dans la glace.

Le lysat est alors transféré dans la colonne *NucleoSpin Filter Maxi* et centrifugée 10 min à 4500 g. Le surnageant est récupéré dans un nouveau tube, sans le culot formé.

10 mL de réactif PC sont ajoutés au surnageant puis mélangés au vortex pendant 30 secondes avant d'être transférés dans la colonne *NucleoSpin Plant II Maxi Column* et centrifugés pendant 2 min à 4500 g.

Le surnageant est éliminé et 4 mL de réactif PW1 sont ajoutés pour resuspendre le culot. Après centrifugation pendant 2 min à 4500 g, le surnageant est éliminé.

10 mL de réactif PW2 sont ajoutés au culot pour le resuspendre, puis centrifugés 10 min à 4500g. Le surnageant est éliminé et le culot est resuspendu dans 2 mL de réactif PW2, puis le tout est centrifugé 10 min à 4500 g.

Avec un nouveau tube collecteur, 1 mL de réactif PE préchauffé à 65°C est ajouté à la colonne. Après 5 min d'incubation à 65°C, la colonne est centrifugée 2 min à 4500 g. Ce premier éluat est conservé. De la même façon, une seconde élution a été réalisée, et le second éluat a été conservé.

Les éluats ont été séchés puis resuspendus dans 750 µL d'eau.

Pour chaque échantillon, 5 µg d'ADN ont été envoyés pour et séquençage après conversion au bisulfite, selon les recommandations du BGI.

Conclusion générale

Les travaux réalisés au cours de ma thèse ont permis de mettre en évidence l'implication de la protéine SG1 dans la régulation de la transcription de l'histone déméthylase IBM1. La dérégulation de *IBM1* dans un fond mutant *sg1* est responsable des perturbations observées au niveau de l'épigénome. En effet, le mutant *sg1* accumule de la méthylation CHG et H3K9me2 au niveau des corps des gènes cibles normalement protégés de ces marques répressives par IBM1.

L'étude des épimutations induites par la mutation *sg1* a permis de montrer, d'une part, que certains épiallèles peuvent être maintenus pendant plusieurs générations et d'autre part, que le phénotype de fleurs à cinq pétales observé dans un mutant *sg1* est très probablement lié à plusieurs épimutations. L'approche par cartographie combinée à l'analyse du méthylome de nos populations d'épimutants nous a permis de mettre en évidence des gènes candidats potentiellement responsables du phénotype floral. La confirmation d'un ou plusieurs de ces candidats pourra permettre de mettre en évidence un lien entre la méthylation dans le corps des gènes et des conséquences phénotypiques.

Le crible suppresseur que nous avons réalisé a abouti à l'identification de FPA, déjà connu pour être impliqué dans la régulation de certains transcrits en influençant le choix de leur site de polyadénylation. En effet, nous avons pu montrer que FPA était impliqué dans la régulation de la transcription des gènes cibles de SG1. L'identification des partenaires de SG1 par co-purification a permis de mettre en évidence des protéines qui suggèrent que la fonction du complexe SG1 est liée aux marques chromatiniennes. Les partenaires de SG1 que nous avons identifiés ne permettent pas, pour l'instant, de faire le lien entre le complexe et le choix du site de polyadénylation. Le mécanisme par lequel le complexe SG1 est capable d'influencer le choix du site de polyadénylation reste donc encore à déterminer. L'identification de la fonction précise des différents membres de ce complexe permettra de mieux comprendre le mécanisme qui permet de réguler la transcription de ses cibles.

Références bibliographiques

- Addepalli, B., and A. G. Hunt, 2007, A novel endonuclease activity associated with the Arabidopsis ortholog of the 30-kDa subunit of cleavage and polyadenylation specificity factor: *Nucleic Acids Res*, v. 35, p. 4453-63.
- Agius, F., A. Kapoor, and J. K. Zhu, 2006, Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation: *Proc Natl Acad Sci U S A*, v. 103, p. 11796-801.
- Alvarez-Venegas, R., 2010, Regulation by polycomb and trithorax group proteins in Arabidopsis: *Arabidopsis Book*, v. 8, p. e0128.
- Alvarez-Venegas, R., and Z. Avramova, 2005, Methylation patterns of histone H3 Lys 4, Lys 9 and Lys 27 in transcriptionally active and inactive Arabidopsis genes and in *atx1* mutants: *Nucleic Acids Res*, v. 33, p. 5199-207.
- Alvarez-Venegas, R., S. Pien, M. Sadler, X. Witmer, U. Grossniklaus, and Z. Avramova, 2003, ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes: *Curr Biol*, v. 13, p. 627-37.
- Ariyoshi, M., and J. W. Schwabe, 2003, A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling: *Genes Dev*, v. 17, p. 1909-20.
- Arteaga-Vazquez, M., L. Sidorenko, F. A. Rabanal, R. Shrivistava, K. Nobuta, P. J. Green, B. C. Meyers, and V. L. Chandler, 2010, RNA-mediated trans-communication can establish paramutation at the *b1* locus in maize: *Proc Natl Acad Sci U S A*, v. 107, p. 12986-91.
- Avramova, Z., 2009, Evolution and pleiotropy of TRITHORAX function in Arabidopsis: *Int J Dev Biol*, v. 53, p. 371-81.
- Bannister, A. J., and T. Kouzarides, 2011, Regulation of chromatin by histone modifications: *Cell Res*, v. 21, p. 381-95.
- Bannister, A. J., P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire, and T. Kouzarides, 2001, Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain: *Nature*, v. 410, p. 120-4.
- Barber, C. M., F. B. Turner, Y. Wang, K. Hagstrom, S. D. Taverna, S. Mollah, B. Ueberheide, B. J. Meyer, D. F. Hunt, P. Cheung, and C. D. Allis, 2004, The enhancement of histone H4 and H2A serine 1 phosphorylation during mitosis and S-phase is evolutionarily conserved: *Chromosoma*, v. 112, p. 360-71.

- Baumbusch, L. O., T. Thorstensen, V. Krauss, A. Fischer, K. Naumann, R. Assalkhou, I. Schulz, G. Reuter, and R. B. Aalen, 2001, The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes: *Nucleic Acids Res*, v. 29, p. 4319-33.
- Baurle, I., L. Smith, D. C. Baulcombe, and C. Dean, 2007, Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing: *Science*, v. 318, p. 109-12.
- Becker, C., J. Hagmann, J. Muller, D. Koenig, O. Stegle, K. Borgwardt, and D. Weigel, 2011, Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome: *Nature*, v. 480, p. 245-9.
- Belotserkovskaya, R., S. Oh, V. A. Bondarenko, G. Orphanides, V. M. Studitsky, and D. Reinberg, 2003, FACT facilitates transcription-dependent nucleosome alteration: *Science*, v. 301, p. 1090-3.
- Berg, M. G., L. N. Singh, I. Younis, Q. Liu, A. M. Pinto, D. Kaida, Z. Zhang, S. Cho, S. Sherrill-Mix, L. Wan, and G. Dreyfuss, 2012, U1 snRNP determines mRNA length and regulates isoform expression: *Cell*, v. 150, p. 53-64.
- Berger, S. L., 2007, The complex language of chromatin regulation during transcription: *Nature*, v. 447, p. 407-12.
- Bernatavichute, Y. V., X. Zhang, S. Cokus, M. Pellegrini, and S. E. Jacobsen, 2008, Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*: *PLoS One*, v. 3, p. e3156.
- Berr, A., R. Menard, T. Heitz, and W. H. Shen, 2012, Chromatin modification and remodelling: a regulatory landscape for the control of *Arabidopsis* defence responses upon pathogen attack: *Cell Microbiol*, v. 14, p. 829-39.
- Bodi, Z., S. Zhong, S. Mehra, J. Song, N. Graham, H. Li, S. May, and R. G. Fray, 2012, Adenosine Methylation in *Arabidopsis* mRNA is Associated with the 3' End and Reduced Levels Cause Developmental Defects: *Front Plant Sci*, v. 3, p. 48.
- Boeger, H., D. A. Bushnell, R. Davis, J. Griesenbeck, Y. Lorch, J. S. Strattan, K. D. Westover, and R. D. Kornberg, 2005, Structural basis of eukaryotic gene transcription: *FEBS Letters*, v. 579, p. 899-903.
- Bogdanovic, O., and G. J. Veenstra, 2009, DNA methylation and methyl-CpG binding proteins: developmental requirements and function: *Chromosoma*, v. 118, p. 549-65.
- Brink, R., 1958, Paramutation at the R Locus in Maize.
- Buratowski, S., and H. Zhou, 1993, Functional domains of transcription factor TFIIB: *Proc Natl Acad Sci U S A*, v. 90, p. 5633-7.
- Butenko, Y., and N. Ohad, 2011, Polycomb-group mediated epigenetic mechanisms through plant evolution: *Biochim Biophys Acta*, v. 1809, p. 395-406.

- Canzio, D., A. Larson, and G. J. Narlikar, 2014, Mechanisms of functional promiscuity by HP1 proteins: *Trends Cell Biol*, v. 24, p. 377-86.
- Cao, J., K. Schneeberger, S. Ossowski, T. Gunther, S. Bender, J. Fitz, D. Koenig, C. Lanz, O. Stegle, C. Lippert, X. Wang, F. Ott, J. Muller, C. Alonso-Blanco, K. Borgwardt, K. J. Schmid, and D. Weigel, 2011, Whole-genome sequencing of multiple *Arabidopsis thaliana* populations: *Nat Genet*, v. 43, p. 956-63.
- Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones, and Y. Zhang, 2002, Role of histone H3 lysine 27 methylation in Polycomb-group silencing: *Science*, v. 298, p. 1039-43.
- Cao, X., and S. E. Jacobsen, 2002, Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes: *Proc Natl Acad Sci U S A*, v. 99 Suppl 4, p. 16491-8.
- Cartagena, J. A., S. Matsunaga, M. Seki, D. Kurihara, M. Yokoyama, K. Shinozaki, S. Fujimoto, Y. Azumi, S. Uchiyama, and K. Fukui, 2008, The *Arabidopsis* SDG4 contributes to the regulation of pollen tube growth by methylation of histone H3 lysines 4 and 36 in mature pollen: *Dev Biol*, v. 315, p. 355-68.
- Chandler, V. L., W. B. Eggleston, and J. E. Dorweiler, 2000, Paramutation in maize: *Plant Mol Biol*, v. 43, p. 121-45.
- Charron, J. B., H. He, A. A. Elling, and X. W. Deng, 2009, Dynamic landscapes of four histone modifications during deetiolation in *Arabidopsis*: *Plant Cell*, v. 21, p. 3732-48.
- Choi, K., C. Park, J. Lee, M. Oh, B. Noh, and I. Lee, 2007, *Arabidopsis* homologs of components of the SWR1 complex regulate flowering and plant development: *Development*, v. 134, p. 1931-41.
- Choi, Y., M. Gehring, L. Johnson, M. Hannon, J. J. Harada, R. B. Goldberg, S. E. Jacobsen, and R. L. Fischer, 2002, DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*: *Cell*, v. 110, p. 33-42.
- Coe, E. H., 1959, A REGULAR AND CONTINUING CONVERSION-TYPE PHENOMENON AT THE B LOCUS IN MAIZE: *Proc Natl Acad Sci U S A*, v. 45, p. 828-32.
- Cokus, S. J., S. Feng, X. Zhang, Z. Chen, B. Merriman, C. D. Haudenschild, S. Pradhan, S. F. Nelson, M. Pellegrini, and S. E. Jacobsen, 2008, Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning: *Nature*, v. 452, p. 215-9.
- Colome-Tatche, M., S. Cortijo, R. Wardenaar, L. Morgado, B. Lahouze, A. Sarazin, M. Etcheverry, A. Martin, S. Feng, E. Duvernois-Berthet, K. Labadie, P. Wincker, S. E. Jacobsen, R. C. Jansen, V. Colot, and F. Johannes, 2012, Features of the *Arabidopsis* recombination landscape resulting from the combined loss of sequence variation and DNA methylation: *Proc Natl Acad Sci U S A*, v. 109, p. 16240-5.
- Corden, J. L., 1990, Tails of RNA polymerase II: *Trends Biochem Sci*, v. 15, p. 383-7.

- Cortijo, S., R. Wardenaar, M. Colome-Tatche, A. Gilly, M. Etcheverry, K. Labadie, E. Caillieux, F. Hospital, J. M. Aury, P. Wincker, F. Roudier, R. C. Jansen, V. Colot, and F. Johannes, 2014, Mapping the epigenetic basis of complex traits: *Science*, v. 343, p. 1145-8.
- Coustham, V., D. Vlad, A. Deremetz, I. Gy, F. A. Cubillos, E. Kerdaffrec, O. Loudet, and N. Bouche, 2014, SHOOT GROWTH1 maintains Arabidopsis epigenomes by regulating IBM1: *PLoS One*, v. 9, p. e84687.
- Cubas, P., C. Vincent, and E. Coen, 1999, An epigenetic mutation responsible for natural variation in floral symmetry: *Nature*, v. 401, p. 157-61.
- Czermin, B., R. Melfi, D. McCabe, V. Seitz, A. Imhof, and V. Pirrotta, 2002, Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites: *Cell*, v. 111, p. 185-96.
- de Castro, E., C. J. A. Sigrist, A. Gattiker, V. Bulliard, P. S. Langendijk-Genevaux, E. Gasteiger, A. Bairoch, and N. Hulo, 2006, ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins: *Nucleic Acids Res*, v. 34, p. W362-5.
- Deal, R. B., C. N. Topp, E. C. McKinney, and R. B. Meagher, 2007, Repression of flowering in Arabidopsis requires activation of FLOWERING LOCUS C expression by the histone variant H2A.Z: *Plant Cell*, v. 19, p. 74-83.
- Derkacheva, M., Y. Steinbach, T. Wildhaber, I. Mozgova, W. Mahrez, P. Nanni, S. Bischof, W. Gruissem, and L. Hennig, 2013, Arabidopsis MSI1 connects LHP1 to PRC2 complexes: *Embo j*, v. 32, p. 2073-85.
- Di Giammartino, D. C., K. Nishida, and J. L. Manley, 2011, Mechanisms and consequences of alternative polyadenylation: *Mol Cell*, v. 43, p. 853-66.
- Dong, G., D. P. Ma, and J. Li, 2008, The histone methyltransferase SDG8 regulates shoot branching in Arabidopsis: *Biochem Biophys Res Commun*, v. 373, p. 659-64.
- Du, J., L. M. Johnson, M. Groth, S. Feng, C. J. Hale, S. Li, A. A. Vashisht, J. Gallego-Bartolome, J. A. Wohlschlegel, D. J. Patel, and S. E. Jacobsen, 2014, Mechanism of DNA methylation-directed histone methylation by KRYPTONITE: *Mol Cell*, v. 55, p. 495-504.
- Du, J., X. Zhong, Y. V. Bernatavichute, H. Stroud, S. Feng, E. Caro, A. A. Vashisht, J. Terragni, H. G. Chin, A. Tu, J. Hetzel, J. A. Wohlschlegel, S. Pradhan, D. J. Patel, and S. E. Jacobsen, 2012, Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants: *Cell*, v. 151, p. 167-80.
- Duc, C., A. Sherstnev, C. Cole, G. J. Barton, and G. G. Simpson, 2013, Transcription termination and chimeric RNA formation controlled by Arabidopsis thaliana FPA: *PLoS Genet*, v. 9, p. e1003867.
- Durand, S., N. Bouche, E. Perez Strand, O. Loudet, and C. Camilleri, 2012, Rapid establishment of genetic incompatibility through natural epigenetic variation: *Curr Biol*, v. 22, p. 326-31.

- Ebbs, M. L., and J. Bender, 2006, Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase: *Plant Cell*, v. 18, p. 1166-76.
- Ehrlich, M., M. A. Gama-Sosa, L. H. Huang, R. M. Midgett, K. C. Kuo, R. A. McCune, and C. Gehrke, 1982, Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells: *Nucleic Acids Res*, v. 10, p. 2709-21.
- Erhard, K. F., Jr., and J. B. Hollick, 2011, Paramutation: a process for acquiring trans-generational regulatory states: *Curr Opin Plant Biol*, v. 14, p. 210-6.
- Eulgem, T., T. Tsuchiya, X. J. Wang, B. Beasley, A. Cuzick, M. Tor, T. Zhu, J. M. McDowell, E. Holub, and J. L. Dangl, 2007, EDM2 is required for RPP7-dependent disease resistance in Arabidopsis and affects RPP7 transcript levels: *Plant J*, v. 49, p. 829-39.
- Fan, J. Y., F. Gordon, K. Luger, J. C. Hansen, and D. J. Tremethick, 2002, The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states: *Nat Struct Biol*, v. 9, p. 172-6.
- Felsenfeld, G., and M. Groudine, 2003, Controlling the double helix: *Nature*, v. 421, p. 448-53.
- Filichkin, S. A., H. D. Priest, S. A. Givan, R. Shen, D. W. Bryant, S. E. Fox, W. K. Wong, and T. C. Mockler, 2010, Genome-wide mapping of alternative splicing in Arabidopsis thaliana: *Genome Res*, v. 20, p. 45-58.
- Filion, G. J., J. G. van Bommel, U. Braunschweig, W. Talhout, J. Kind, L. D. Ward, W. Brugman, I. J. de Castro, R. M. Kerkhoven, H. J. Bussemaker, and B. van Steensel, 2010, Systematic protein location mapping reveals five principal chromatin types in Drosophila cells: *Cell*, v. 143, p. 212-24.
- Fischle, W., B. S. Tseng, H. L. Dormann, B. M. Ueberheide, B. A. Garcia, J. Shabanowitz, D. F. Hunt, H. Funabiki, and C. D. Allis, 2005, Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation: *Nature*, v. 438, p. 1116-22.
- Foster, E. R., and J. A. Downs, 2005, Histone H2A phosphorylation in DNA double-strand break repair: *Febs j*, v. 272, p. 3231-40.
- Francis, N. J., A. J. Saurin, Z. Shao, and R. E. Kingston, 2001, Reconstitution of a functional core polycomb repressive complex: *Mol Cell*, v. 8, p. 545-56.
- Fransz, P., J. H. De Jong, M. Lysak, M. R. Castiglione, and I. Schubert, 2002, Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate: *Proc Natl Acad Sci U S A*, v. 99, p. 14584-9.
- Fransz, P., S. I. f. L. S. Nuclear Organization Group, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands, H. de Jong, and W. U. a. R. C. Laboratory of Genetics, Wageningen, The Netherlands, 2011, From nucleosome to chromosome: a dynamic organization of genetic information: *The Plant Journal*, v. 66, p. 4-17.
- Fransz, P., R. ten Hoopen, and F. Tessadori, 2006, Composition and formation of heterochromatin in Arabidopsis thaliana: *Chromosome Res*, v. 14, p. 71-82.

- Gehring, M., K. L. Bubb, and S. Henikoff, 2009, Extensive demethylation of repetitive elements during seed development underlies gene imprinting: *Science*, v. 324, p. 1447-51.
- Gehring, M., J. H. Huh, T. F. Hsieh, J. Penterman, Y. Choi, J. J. Harada, R. B. Goldberg, and R. L. Fischer, 2006, DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation: *Cell*, v. 124, p. 495-506.
- Gehring, M., V. Missirlian, and S. Henikoff, 2011, Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds: *PLoS One*, v. 6, p. e23687.
- Glover-Cutter, K., S. Larochele, B. Erickson, C. Zhang, K. Shokat, R. P. Fisher, and D. L. Bentley, 2009, TFIIF-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II: *Mol Cell Biol*, v. 29, p. 5455-64.
- Goll, M. G., and T. H. Bestor, 2005, Eukaryotic cytosine methyltransferases: *Annu Rev Biochem*, v. 74, p. 481-514.
- Gong, Z., T. Morales-Ruiz, R. R. Ariza, T. Roldan-Arjona, L. David, and J. K. Zhu, 2002, ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase: *Cell*, v. 111, p. 803-14.
- Greaves, I. K., M. Groszmann, A. Wang, W. J. Peacock, and E. S. Dennis, 2014, Inheritance of Trans Chromosomal Methylation patterns from *Arabidopsis* F1 hybrids: *Proc Natl Acad Sci U S A*, v. 111, p. 2017-22.
- Greaves, I. K., M. Groszmann, H. Ying, J. M. Taylor, W. J. Peacock, and E. S. Dennis, 2012, Trans chromosomal methylation in *Arabidopsis* hybrids: *Proc Natl Acad Sci U S A*, v. 109, p. 3570-5.
- Groszmann, M., I. K. Greaves, Z. I. Albertyn, G. N. Scofield, W. J. Peacock, and E. S. Dennis, 2011, Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor: *Proc Natl Acad Sci U S A*, v. 108, p. 2617-22.
- Haag, J. R., T. S. Ream, M. Marasco, C. D. Nicora, A. D. Norbeck, L. Pasa-Tolic, and C. S. Pikaard, 2012, In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing: *Mol Cell*, v. 48, p. 811-8.
- Hennig, L., and M. Derkacheva, 2009, Diversity of Polycomb group complexes in plants: same rules, different players?: *Trends Genet*, v. 25, p. 414-23.
- Herr, A. J., M. B. Jensen, T. Dalmay, and D. C. Baulcombe, 2005, RNA polymerase IV directs silencing of endogenous DNA: *Science*, v. 308, p. 118-20.
- Hornyik, C., C. Duc, K. Rataj, L. C. Terzi, and G. G. Simpson, 2010a, Alternative polyadenylation of antisense RNAs and flowering time control: *Biochem Soc Trans*, v. 38, p. 1077-81.
- Hornyik, C., L. C. Terzi, and G. G. Simpson, 2010b, The spen family protein FPA controls alternative cleavage and polyadenylation of RNA: *Dev Cell*, v. 18, p. 203-13.

- Hsieh, T. F., C. A. Ibarra, P. Silva, A. Zemach, L. Eshed-Williams, R. L. Fischer, and D. Zilberman, 2009, Genome-wide demethylation of Arabidopsis endosperm: *Science*, v. 324, p. 1451-4.
- Inagaki, S., A. Miura-Kamio, Y. Nakamura, F. Lu, X. Cui, X. Cao, H. Kimura, H. Saze, and T. Kakutani, 2010, Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome: *Embo j*, v. 29, p. 3496-506.
- Irish, V. F., 2008, The Arabidopsis petal: a model for plant organogenesis: *Trends Plant Sci*, v. 13, p. 430-6.
- Jackson, J. P., L. Johnson, Z. Jasencakova, X. Zhang, L. PerezBurgos, P. B. Singh, X. Cheng, I. Schubert, T. Jenuwein, and S. E. Jacobsen, 2004, Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana: *Chromosoma*, v. 112, p. 308-15.
- Jackson, J. P., A. M. Lindroth, X. Cao, and S. E. Jacobsen, 2002, Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase: *Nature*, v. 416, p. 556-60.
- Jacob, Y., S. Feng, C. A. LeBlanc, Y. V. Bernatavichute, H. Stroud, S. Cokus, L. M. Johnson, M. Pellegrini, S. E. Jacobsen, and S. D. Michaels, 2009, ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing: *Nat Struct Mol Biol*, v. 16, p. 763-8.
- Jacob, Y., H. Stroud, C. Leblanc, S. Feng, L. Zhuo, E. Caro, C. Hassel, C. Gutierrez, S. D. Michaels, and S. E. Jacobsen, 2010, Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases: *Nature*, v. 466, p. 987-91.
- Jacobsen, S. E., and E. M. Meyerowitz, 1997, Hypermethylated SUPERMAN epigenetic alleles in arabidopsis: *Science*, v. 277, p. 1100-3.
- Jeddeloh, J. A., T. L. Stokes, and E. J. Richards, 1999, Maintenance of genomic methylation requires a SWI2/SNF2-like protein: *Nat Genet*, v. 22, p. 94-7.
- Ji, Z., W. Luo, W. Li, M. Hoque, Z. Pan, Y. Zhao, and B. Tian, 2011, Transcriptional activity regulates alternative cleavage and polyadenylation: *Mol Syst Biol*, v. 7, p. 534.
- Jiang, D., N. C. Kong, X. Gu, Z. Li, and Y. He, 2011, Arabidopsis COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development: *PLoS Genet*, v. 7, p. e1001330.
- Jiang, D., Y. Wang, and Y. He, 2008, Repression of FLOWERING LOCUS C and FLOWERING LOCUS T by the Arabidopsis Polycomb repressive complex 2 components: *PLoS One*, v. 3, p. e3404.
- Johannes, F., E. Porcher, F. K. Teixeira, V. Saliba-Colombani, M. Simon, N. Agier, A. Bulski, J. Albuissou, F. Heredia, P. Audigier, D. Bouchez, C. Dillmann, P. Guerche, F. Hospital, and V. Colot, 2009, Assessing the impact of transgenerational epigenetic variation on complex traits: *PLoS Genet*, v. 5, p. e1000530.

- Johnson, L. M., M. Bostick, X. Zhang, E. Kraft, I. Henderson, J. Callis, and S. E. Jacobsen, 2007, The SRA methyl-cytosine-binding domain links DNA and histone methylation: *Curr Biol*, v. 17, p. 379-84.
- Kakutani, T., J. A. Jeddelloh, S. K. Flowers, K. Munakata, and E. J. Richards, 1996, Developmental abnormalities and epimutations associated with DNA hypomethylation mutations: *Proc Natl Acad Sci U S A*, v. 93, p. 12406-11.
- Kalyna, M., C. G. Simpson, N. H. Syed, D. Lewandowska, Y. Marquez, B. Kusenda, J. Marshall, J. Fuller, L. Cardle, J. McNicol, H. Q. Dinh, A. Barta, and J. W. Brown, 2012, Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*: *Nucleic Acids Res*, v. 40, p. 2454-69.
- Kanno, T., B. Huettel, M. F. Mette, W. Aufsatz, E. Jaligot, L. Daxinger, D. P. Kreil, M. Matzke, and A. J. Matzke, 2005, Atypical RNA polymerase subunits required for RNA-directed DNA methylation: *Nat Genet*, v. 37, p. 761-5.
- Kato, M., A. Miura, J. Bender, S. E. Jacobsen, and T. Kakutani, 2003, Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*: *Curr Biol*, v. 13, p. 421-6.
- Kawashima, T., and F. Berger, 2014, Epigenetic reprogramming in plant sexual reproduction: *Nat Rev Genet*, v. 15, p. 613-24.
- Kim, J., M. Guermah, R. K. McGinty, J. S. Lee, Z. Tang, T. A. Milne, A. Shilatifard, T. W. Muir, and R. G. Roeder, 2009, RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells: *Cell*, v. 137, p. 459-71.
- King, I. F., N. J. Francis, and R. E. Kingston, 2002, Native and recombinant polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro: *Mol Cell Biol*, v. 22, p. 7919-28.
- Kinoshita, T., A. Miura, Y. Choi, Y. Kinoshita, X. Cao, S. E. Jacobsen, R. L. Fischer, and T. Kakutani, 2004, One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation: *Science*, v. 303, p. 521-3.
- Kinoshita, Y., H. Saze, T. Kinoshita, A. Miura, W. J. Soppe, M. Koornneef, and T. Kakutani, 2006, Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats: *Plant J*, v. 49, p. 38-45.
- Klose, R. J., E. M. Kallin, and Y. Zhang, 2006, JmjC-domain-containing proteins and histone demethylation: *Nat Rev Genet*, v. 7, p. 715-27.
- Kooistra, S. M., and K. Helin, 2012, Molecular mechanisms and potential functions of histone demethylases: *Nat Rev Mol Cell Biol*, v. 13, p. 297-311.
- Koornneef, M., C. J. Hanhart, and J. H. van der Veen, 1991, A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*: *Mol Gen Genet*, v. 229, p. 57-66.

- Lachner, M., D. O'Carroll, S. Rea, K. Mechtler, and T. Jenuwein, 2001, Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins: *Nature*, v. 410, p. 116-20.
- Law, J. A., J. Du, C. J. Hale, S. Feng, K. Krajewski, A. M. Palanca, B. D. Strahl, D. J. Patel, and S. E. Jacobsen, 2013, Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1: *Nature*, v. 498, p. 385-9.
- Le, T. N., Y. Miyazaki, S. Takuno, and H. Saze, 2015, Epigenetic regulation of intragenic transposable elements impacts gene transcription in *Arabidopsis thaliana*: *Nucleic Acids Res*, v. 43, p. 3911-21.
- Lee, J. S., A. Shukla, J. Schneider, S. K. Swanson, M. P. Washburn, L. Florens, S. R. Bhaumik, and A. Shilatifard, 2007, Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS: *Cell*, v. 131, p. 1084-96.
- Lei, M., H. La, K. Lu, P. Wang, D. Miki, Z. Ren, C. G. Duan, X. Wang, K. Tang, L. Zeng, L. Yang, H. Zhang, W. Nie, P. Liu, J. Zhou, R. Liu, Y. Zhong, D. Liu, and J. K. Zhu, 2014, *Arabidopsis* EDM2 promotes IBM1 distal polyadenylation and regulates genome DNA methylation patterns: *Proc Natl Acad Sci U S A*, v. 111, p. 527-32.
- Levine, S. S., A. Weiss, H. Erdjument-Bromage, Z. Shao, P. Tempst, and R. E. Kingston, 2002, The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans: *Mol Cell Biol*, v. 22, p. 6070-8.
- Li, B., M. Carey, and J. L. Workman, 2007, The role of chromatin during transcription: *Cell*, v. 128, p. 707-19.
- Lindroth, A. M., X. Cao, J. P. Jackson, D. Zilberman, C. M. McCallum, S. Henikoff, and S. E. Jacobsen, 2001, Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation: *Science*, v. 292, p. 2077-80.
- Lindroth, A. M., D. Shultis, Z. Jasencakova, J. Fuchs, L. Johnson, D. Schubert, D. Patnaik, S. Pradhan, J. Goodrich, I. Schubert, T. Jenuwein, S. Khorasanizadeh, and S. E. Jacobsen, 2004, Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3: *Embo j*, v. 23, p. 4286-96.
- Lippman, Z., A. V. Gendrel, M. Black, M. W. Vaughn, N. Dedhia, W. R. McCombie, K. Lavine, V. Mittal, B. May, K. D. Kasschau, J. C. Carrington, R. W. Doerge, V. Colot, and R. Martienssen, 2004, Role of transposable elements in heterochromatin and epigenetic control: *Nature*, v. 430, p. 471-6.
- Lippman, Z., B. May, C. Yordan, T. Singer, and R. Martienssen, 2003, Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification: *PLoS Biol*, v. 1, p. E67.
- Lister, R., R. C. O'Malley, J. Tonti-Filippini, B. D. Gregory, C. C. Berry, A. H. Millar, and J. R. Ecker, 2008, Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*: *Cell*, v. 133, p. 523-36.
- Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A.

- H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker, 2009, Human DNA methylomes at base resolution show widespread epigenomic differences: *Nature*, v. 462, p. 315-22.
- Liu, F., S. Marquardt, C. Lister, S. Swiezewski, and C. Dean, 2010, Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing: *Science*, v. 327, p. 94-7.
- Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond, 1997, Crystal structure of the nucleosome core particle at 2.8 Å resolution: *Nature*, v. 389, p. 251-60.
- Luo, M., J. M. Taylor, A. Spriggs, H. Zhang, X. Wu, S. Russell, M. Singh, and A. Koltunow, 2011, A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm: *PLoS Genet*, v. 7, p. e1002125.
- Malagnac, F., L. Bartee, and J. Bender, 2002, An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation: *Embo j*, v. 21, p. 6842-52.
- Mandel, C. R., Y. Bai, and L. Tong, 2008, Protein factors in pre-mRNA 3'-end processing: *Cell Mol Life Sci*, v. 65, p. 1099-122.
- Manning, K., M. Tor, M. Poole, Y. Hong, A. J. Thompson, G. J. King, J. J. Giovannoni, and G. B. Seymour, 2006, A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening: *Nat Genet*, v. 38, p. 948-52.
- Marquez, Y., J. W. Brown, C. Simpson, A. Barta, and M. Kalyna, 2012, Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis: *Genome Res*, v. 22, p. 1184-95.
- Martin, A., C. Troadec, A. Boualem, M. Rajab, R. Fernandez, H. Morin, M. Pitrat, C. Dogimont, and A. Bendahmane, 2009, A transposon-induced epigenetic change leads to sex determination in melon: *Nature*, v. 461, p. 1135-8.
- Mathieu, O., and N. Bouche, 2014, Interplay between chromatin and RNA processing: *Curr Opin Plant Biol*, v. 18, p. 60-5.
- Mathieu, O., A. V. Probst, and J. Paszkowski, 2005, Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis: *Embo j*, v. 24, p. 2783-91.
- Mathieu, O., J. Reinders, M. Caikovski, C. Smathajitt, and J. Paszkowski, 2007, Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation: *Cell*, v. 130, p. 851-62.
- McKeown, P. C., S. Laouielle-Duprat, P. Prins, P. Wolff, M. W. Schmid, M. T. Donoghue, A. Fort, D. Duszynska, A. Comte, N. T. Lao, T. J. Wennblom, G. Smant, C. Kohler, U. Grossniklaus, and C. Spillane, 2011, Identification of imprinted genes subject to parent-of-origin specific expression in Arabidopsis thaliana seeds: *BMC Plant Biol*, v. 11, p. 113.
- McKhann, H. I., C. Camilleri, A. Berard, T. Bataillon, J. L. David, X. Reboud, V. Le Corre, C. Caloustian, I. G. Gut, and D. Brunel, 2004, Nested core collections maximizing genetic diversity in Arabidopsis thaliana: *Plant J*, v. 38, p. 193-202.

- McKittrick, E., P. R. Gafken, K. Ahmad, and S. Henikoff, 2004, Histone H3.3 is enriched in covalent modifications associated with active chromatin: *Proc Natl Acad Sci U S A*, v. 101, p. 1525-30.
- Melquist, S., and J. Bender, 2003, Transcription from an upstream promoter controls methylation signaling from an inverted repeat of endogenous genes in *Arabidopsis*: *Genes Dev*, v. 17, p. 2036-47.
- Melquist, S., B. Luff, and J. Bender, 1999, *Arabidopsis* PAI gene arrangements, cytosine methylation and expression: *Genetics*, v. 153, p. 401-13.
- Michaels, S. D., and R. M. Amasino, 2001, Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization: *Plant Cell*, v. 13, p. 935-41.
- Miura, A., M. Nakamura, S. Inagaki, A. Kobayashi, H. Saze, and T. Kakutani, 2009, An *Arabidopsis* jmjC domain protein protects transcribed genes from DNA methylation at CHG sites: *Embo j*, v. 28, p. 1078-86.
- Morales-Ruiz, T., A. P. Ortega-Galisteo, M. I. Ponferrada-Marin, M. I. Martinez-Macias, R. R. Ariza, and T. Roldan-Arjona, 2006, DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases: *Proc Natl Acad Sci U S A*, v. 103, p. 6853-8.
- Naumann, K., A. Fischer, I. Hofmann, V. Krauss, S. Phalke, K. Irmeler, G. Hause, A. C. Aurich, R. Dorn, T. Jenuwein, and G. Reuter, 2005, Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in *Arabidopsis*: *Embo j*, v. 24, p. 1418-29.
- Noh, Y. S., and R. M. Amasino, 2003, PIE1, an ISWI family gene, is required for FLC activation and floral repression in *Arabidopsis*: *Plant Cell*, v. 15, p. 1671-82.
- Nuthikattu, S., A. D. McCue, K. Panda, D. Fultz, C. DeFraia, E. N. Thomas, and R. K. Slotkin, 2013, The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs: *Plant Physiol*, v. 162, p. 116-31.
- Ohnacker, M., S. M. Barabino, P. J. Preker, and W. Keller, 2000, The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex: *Embo j*, v. 19, p. 37-47.
- Oki, M., H. Aihara, and T. Ito, 2007, Role of histone phosphorylation in chromatin dynamics and its implications in diseases: *Subcell Biochem*, v. 41, p. 319-36.
- Ortega-Galisteo, A. P., T. Morales-Ruiz, R. R. Ariza, and T. Roldan-Arjona, 2008, *Arabidopsis* DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks: *Plant Mol Biol*, v. 67, p. 671-81.
- Ossowski, S., K. Schneeberger, J. I. Lucas-Lledo, N. Warthmann, R. M. Clark, R. G. Shaw, D. Weigel, and M. Lynch, 2010, The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*: *Science*, v. 327, p. 92-4.

- Palmer, D. K., K. O'Day, M. H. Wener, B. S. Andrews, and R. L. Margolis, 1987, A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones: *J Cell Biol*, v. 104, p. 805-15.
- Park, S., S. Oh, and S. van Nocker, 2012, Genomic and gene-level distribution of histone H3 dimethyl lysine-27 (H3K27me₂) in Arabidopsis: *PLoS One*, v. 7, p. e52855.
- Pawlowski, W. P., 2010, Chromosome organization and dynamics in plants: *Curr Opin Plant Biol*, v. 13, p. 640-5.
- Penterman, J., D. Zilberman, J. H. Huh, T. Ballinger, S. Henikoff, and R. L. Fischer, 2007, DNA demethylation in the Arabidopsis genome: *Proc Natl Acad Sci U S A*, v. 104, p. 6752-7.
- Phatnani, H. P., and A. L. Greenleaf, 2006, Phosphorylation and functions of the RNA polymerase II CTD: *Genes Dev*, v. 20, p. 2922-36.
- Platero, J. S., T. Hartnett, and J. C. Eissenberg, 1995, Functional analysis of the chromo domain of HP1: *Embo j*, v. 14, p. 3977-86.
- Quesada, V., R. Macknight, C. Dean, and G. G. Simpson, 2003, Autoregulation of FCA pre-mRNA processing controls Arabidopsis flowering time: *Embo j*, v. 22, p. 3142-52.
- Rea, S., F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z. W. Sun, M. Schmid, S. Opravil, K. Mechtler, C. P. Ponting, C. D. Allis, and T. Jenuwein, 2000, Regulation of chromatin structure by site-specific histone H3 methyltransferases: *Nature*, v. 406, p. 593-9.
- Reinders, J., B. B. Wulff, M. Mirouze, A. Mari-Ordonez, M. Dapp, W. Rozhon, E. Bucher, G. Theiler, and J. Paszkowski, 2009, Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes: *Genes Dev*, v. 23, p. 939-50.
- Rigal, M., Z. Kevei, T. Pelissier, and O. Mathieu, 2012, DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns: *Embo j*, v. 31, p. 2981-93.
- Roberts, R. J., and X. Cheng, 1998, Base flipping: *Annu Rev Biochem*, v. 67, p. 181-98.
- Roudier, F., I. Ahmed, C. Berard, A. Sarazin, T. Mary-Huard, S. Cortijo, D. Bouyer, E. Caillieux, E. Duvernois-Berthet, L. Al-Shikhley, L. Giraut, B. Despres, S. Drevensek, F. Barneche, S. Derozier, V. Brunaud, S. Aubourg, A. Schnittger, C. Bowler, M. L. Martin-Magniette, S. Robin, M. Caboche, and V. Colot, 2011, Integrative epigenomic mapping defines four main chromatin states in Arabidopsis: *Embo j*, v. 30, p. 1928-38.
- Ryan, D. P., and T. Owen-Hughes, 2011, Snf2-family proteins: chromatin remodellers for any occasion: *Curr Opin Chem Biol*, v. 15, p. 649-56.
- Saleh, A., R. Alvarez-Venegas, M. Yilmaz, O. Le, G. Hou, M. Sadder, A. Al-Abdallat, Y. Xia, G. Lu, I. Ladunga, and Z. Avramova, 2008, The highly similar Arabidopsis homologs of trithorax ATX1 and ATX2 encode proteins with divergent biochemical functions: *Plant Cell*, v. 20, p. 568-79.

- Santos-Rosa, H., R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. Emre, S. L. Schreiber, J. Mellor, and T. Kouzarides, 2002, Active genes are tri-methylated at K4 of histone H3: *Nature*, v. 419, p. 407-11.
- Saurin, A. J., Z. Shao, H. Erdjument-Bromage, P. Tempst, and R. E. Kingston, 2001, A *Drosophila* Polycomb group complex includes Zeste and dTAFII proteins: *Nature*, v. 412, p. 655-60.
- Saze, H., and T. Kakutani, 2007, Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1: *Embo j*, v. 26, p. 3641-52.
- Saze, H., J. Kitayama, K. Takashima, S. Miura, Y. Harukawa, T. Ito, and T. Kakutani, 2013, Mechanism for full-length RNA processing of *Arabidopsis* genes containing intragenic heterochromatin: *Nat Commun*, v. 4, p. 2301.
- Saze, H., O. Mittelsten Scheid, and J. Paszkowski, 2003, Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis: *Nat Genet*, v. 34, p. 65-9.
- Saze, H., A. Shiraishi, A. Miura, and T. Kakutani, 2008, Control of genic DNA methylation by a *jmjC* domain-containing protein in *Arabidopsis thaliana*: *Science*, v. 319, p. 462-5.
- Schmitz, R. J., M. D. Schultz, M. G. Lewsey, R. C. O'Malley, M. A. Urich, O. Libiger, N. J. Schork, and J. R. Ecker, 2011, Transgenerational epigenetic instability is a source of novel methylation variants: *Science*, v. 334, p. 369-73.
- Schomburg, F. M., D. A. Patton, D. W. Meinke, and R. M. Amasino, 2001, FPA, a gene involved in floral induction in *Arabidopsis*, encodes a protein containing RNA-recognition motifs: *Plant Cell*, v. 13, p. 1427-36.
- Shahbazian, M. D., and M. Grunstein, 2007, Functions of site-specific histone acetylation and deacetylation: *Annu Rev Biochem*, v. 76, p. 75-100.
- Shen, Y., R. C. Venu, K. Nobuta, X. Wu, V. Notibala, C. Demirci, B. C. Meyers, G. L. Wang, G. Ji, and Q. Q. Li, 2011, Transcriptome dynamics through alternative polyadenylation in developmental and environmental responses in plants revealed by deep sequencing: *Genome Res*, v. 21, p. 1478-86.
- Shi, Y., F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, and R. A. Casero, 2004, Histone demethylation mediated by the nuclear amine oxidase homolog LSD1: *Cell*, v. 119, p. 941-53.
- Shu, H., T. Wildhaber, A. Siretskiy, W. Gruissem, and L. Hennig, 2012, Distinct modes of DNA accessibility in plant chromatin: *Nat Commun*, v. 3, p. 1281.
- Shuman, S., 1995, Capping enzyme in eukaryotic mRNA synthesis: *Prog Nucleic Acid Res Mol Biol*, v. 50, p. 101-29.
- Silveira, A. B., C. Trontin, S. Cortijo, J. Barau, L. E. Del Bem, O. Loudet, V. Colot, and M. Vincentz, 2013, Extensive natural epigenetic variation at a de novo originated gene: *PLoS Genet*, v. 9, p. e1003437.

- Simpson, G. G., P. P. Dijkwel, V. Quesada, I. Henderson, and C. Dean, 2003, FY is an RNA 3' end-processing factor that interacts with FCA to control the Arabidopsis floral transition: *Cell*, v. 113, p. 777-87.
- Sonmez, C., I. Baurle, A. Magusin, R. Dreos, S. Laubinger, D. Weigel, and C. Dean, 2011, RNA 3' processing functions of Arabidopsis FCA and FPA limit intergenic transcription: *Proc Natl Acad Sci U S A*, v. 108, p. 8508-13.
- Soppe, W. J., S. E. Jacobsen, C. Alonso-Blanco, J. P. Jackson, T. Kakutani, M. Koornneef, and A. J. Peeters, 2000, The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene: *Mol Cell*, v. 6, p. 791-802.
- Soppe, W. J., Z. Jasencakova, A. Houben, T. Kakutani, A. Meister, M. S. Huang, S. E. Jacobsen, I. Schubert, and P. F. Fransz, 2002, DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis, *EMBO J*, v. 21, p. 6549-59.
- Spies, N., C. B. Nielsen, R. A. Padgett, and C. B. Burge, 2009, Biased chromatin signatures around polyadenylation sites and exons: *Mol Cell*, v. 36, p. 245-54.
- Springer, N. M., C. A. Napoli, D. A. Selinger, R. Pandey, K. C. Cone, V. L. Chandler, H. F. Kaeppler, and S. M. Kaeppler, 2003, Comparative analysis of SET domain proteins in maize and Arabidopsis reveals multiple duplications preceding the divergence of monocots and dicots: *Plant Physiol*, v. 132, p. 907-25.
- Stroud, H., T. Do, J. Du, X. Zhong, S. Feng, L. Johnson, D. J. Patel, and S. E. Jacobsen, 2014, Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis: *Nat Struct Mol Biol*, v. 21, p. 64-72.
- Stroud, H., M. V. Greenberg, S. Feng, Y. V. Bernatavichute, and S. E. Jacobsen, 2013, Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome: *Cell*, v. 152, p. 352-64.
- The Arabidopsis Genome Initiative, 2000, Analysis of the genome sequence of the flowering plant Arabidopsis thaliana: *Nature*, v. 408, p. 796-815.
- Tie, F., T. Furuyama, J. Prasad-Sinha, E. Jane, and P. J. Harte, 2001, The Drosophila Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3: *Development*, v. 128, p. 275-86.
- Tsuchiya, T., and T. Eulgem, 2013a, An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication: *Proc Natl Acad Sci U S A*, v. 110, p. E3535-43.
- Tsuchiya, T., and T. Eulgem, 2013b, Mutations in EDM2 selectively affect silencing states of transposons and induce plant developmental plasticity: *Sci Rep*, v. 3, p. 1701.
- Tsuchiya, T., and T. Eulgem, 2014, The PHD-finger module of the Arabidopsis thaliana defense regulator EDM2 can recognize triply modified histone H3 peptides, *Plant Signal Behav*, v. 9.

- Turck, F., F. Roudier, S. Farrona, M. L. Martin-Magniette, E. Guillaume, N. Buisine, S. Gagnot, R. A. Martienssen, G. Coupland, and V. Colot, 2007, *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27: *PLoS Genet*, v. 3, p. e86.
- van Driel and Fransz, P., 2004, Nuclear architecture and genome functioning in plants and animals: what can we learn from both?, v. 296, p. 86–90.
- Vaughn, M. W., M. Tanurdzic, Z. Lippman, H. Jiang, R. Carrasquillo, P. D. Rabinowicz, N. Dedhia, W. R. McCombie, N. Agier, A. Bulski, V. Colot, R. W. Doerge, and R. A. Martienssen, 2007, Epigenetic natural variation in *Arabidopsis thaliana*: *PLoS Biol*, v. 5, p. e174.
- Vongs, A., T. Kakutani, R. A. Martienssen, and E. J. Richards, 1993, *Arabidopsis thaliana* DNA methylation mutants: *Science*, v. 260, p. 1926-8.
- Waddington, C. H., 1942, The epigenotype. 1942: *Int J Epidemiol*, v. 41, p. 10-3.
- Wang, B. B., and V. Brendel, 2006, Genomewide comparative analysis of alternative splicing in plants: *Proc Natl Acad Sci U S A*, v. 103, p. 7175-80.
- Wang, X., C. G. Duan, K. Tang, B. Wang, H. Zhang, M. Lei, K. Lu, S. K. Mangrauthia, P. Wang, G. Zhu, Y. Zhao, and J. K. Zhu, 2013, RNA-binding protein regulates plant DNA methylation by controlling mRNA processing at the intronic heterochromatin-containing gene IBM1: *Proc Natl Acad Sci U S A*, v. 110, p. 15467-72.
- Watson, J. D., and F. H. Crick, 1953, Genetical implications of the structure of deoxyribonucleic acid: *Nature*, v. 171, p. 964-7.
- Woo, H. R., T. A. Dittmer, and E. J. Richards, 2008, Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*: *PLoS Genet*, v. 4, p. e1000156.
- Wu, X., M. Liu, B. Downie, C. Liang, G. Ji, Q. Q. Li, and A. G. Hunt, 2011, Genome-wide landscape of polyadenylation in *Arabidopsis* provides evidence for extensive alternative polyadenylation: *Proc Natl Acad Sci U S A*, v. 108, p. 12533-8.
- Xu, L., Z. Zhao, A. Dong, L. Soubigou-Taconnat, J. P. Renou, A. Steinmetz, and W. H. Shen, 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*: *Mol Cell Biol*, v. 28, p. 1348-60.
- Yamada, K., J. Lim, J. M. Dale, H. Chen, P. Shinn, C. J. Palm, A. M. Southwick, H. C. Wu, C. Kim, M. Nguyen, P. Pham, R. Cheuk, G. Karlin-Newmann, S. X. Liu, B. Lam, H. Sakano, T. Wu, G. Yu, M. Miranda, H. L. Quach, M. Tripp, C. H. Chang, J. M. Lee, M. Toriumi, M. M. Chan, C. C. Tang, C. S. Onodera, J. M. Deng, K. Akiyama, Y. Ansari, T. Arakawa, J. Banh, F. Banno, L. Bowser, S. Brooks, P. Carninci, Q. Chao, N. Choy, A. Enju, A. D. Goldsmith, M. Gurjal, N. F. Hansen, Y. Hayashizaki, C. Johnson-Hopson, V. W. Hsuan, K. Iida, M. Karnes, S. Khan, E. Koesema, J. Ishida, P. X. Jiang, T. Jones, J. Kawai, A. Kamiya, C. Meyers, M. Nakajima, M. Narusaka, M. Seki, T. Sakurai, M. Satou, R. Tamse, M. Vaysberg, E. K. Wallender, C. Wong, Y. Yamamura,

- S. Yuan, K. Shinozaki, R. W. Davis, A. Theologis, and J. R. Ecker, 2003, Empirical analysis of transcriptional activity in the Arabidopsis genome: *Science*, v. 302, p. 842-6.
- Zemach, A., and G. Grafi, 2003, Characterization of Arabidopsis thaliana methyl-CpG-binding domain (MBD) proteins: *Plant J*, v. 34, p. 565-72.
- Zemach, A., and G. Grafi, 2007, Methyl-CpG-binding domain proteins in plants: interpreters of DNA methylation: *Trends Plant Sci*, v. 12, p. 80-5.
- Zemach, A., M. Y. Kim, P. H. Hsieh, D. Coleman-Derr, L. Eshed-Williams, K. Thao, S. L. Harmer, and D. Zilberman, 2013, The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin: *Cell*, v. 153, p. 193-205.
- Zhang, K., V. V. Sridhar, J. Zhu, A. Kapoor, and J. K. Zhu, 2007a, Distinctive core histone post-translational modification patterns in Arabidopsis thaliana: *PLoS One*, v. 2, p. e1210.
- Zhang, X., Y. V. Bernatavichute, S. Cokus, M. Pellegrini, and S. E. Jacobsen, 2009, Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana: *Genome Biol*, v. 10, p. R62.
- Zhang, X., O. Clarenz, S. Cokus, Y. V. Bernatavichute, M. Pellegrini, J. Goodrich, and S. E. Jacobsen, 2007b, Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis: *PLoS Biol*, v. 5, p. e129.
- Zhang, X., S. Germann, B. J. Blus, S. Khorasanizadeh, V. Gaudin, and S. E. Jacobsen, 2007c, The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation: *Nat Struct Mol Biol*, v. 14, p. 869-71.
- Zhang, X., S. H. Shiu, A. Cal, and J. O. Borevitz, 2008, Global analysis of genetic, epigenetic and transcriptional polymorphisms in Arabidopsis thaliana using whole genome tiling arrays: *PLoS Genet*, v. 4, p. e1000032.
- Zhang, X., J. Yazaki, A. Sundaresan, S. Cokus, S. W. Chan, H. Chen, I. R. Henderson, P. Shinn, M. Pellegrini, S. E. Jacobsen, and J. R. Ecker, 2006, Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis: *Cell*, v. 126, p. 1189-201.
- Zhang, Y., and D. Reinberg, 2001, Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails: *Genes Dev*, v. 15, p. 2343-60.
- Zheng, B., and X. Chen, 2011, Dynamics of histone H3 lysine 27 trimethylation in plant development: *Curr Opin Plant Biol*, v. 14, p. 123-9.
- Zheng, B., Z. Wang, S. Li, B. Yu, J. Y. Liu, and X. Chen, 2009, Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis: *Genes Dev*, v. 23, p. 2850-60.

- Zhong, S., H. Li, Z. Bodi, J. Button, L. Vespa, M. Herzog, and R. G. Fray, 2008, MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor: *Plant Cell*, v. 20, p. 1278-88.
- Zhong, X., J. Du, C. J. Hale, J. Gallego-Bartolome, S. Feng, A. A. Vashisht, J. Chory, J. A. Wohlschlegel, D. J. Patel, and S. E. Jacobsen, 2014, Molecular mechanism of action of plant DRM de novo DNA methyltransferases: *Cell*, v. 157, p. 1050-60.
- Zhu, J., A. Kapoor, V. V. Sridhar, F. Agius, and J. K. Zhu, 2007, The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in Arabidopsis: *Curr Biol*, v. 17, p. 54-9.
- Zilberman, D., X. Cao, L. K. Johansen, Z. Xie, J. C. Carrington, and S. E. Jacobsen, 2004, Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats: *Curr Biol*, v. 14, p. 1214-20.
- Zilberman, D., D. Coleman-Derr, T. Ballinger, and S. Henikoff, 2008, Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks: *Nature*, v. 456, p. 125-9.
- Zilberman, D., M. Gehring, R. K. Tran, T. Ballinger, and S. Henikoff, 2007, Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription: *Nat Genet*, v. 39, p. 61-9.